

GMC Analysis

Data Protection Impact Assessment

July 2025

V1.1



**NHS fraud.
Spot it. Report it.
Together we stop it.**

Executive Summary

This document contains information in relation to GMC data

The document is deemed OFFICIAL and any information viewed/obtained within it should be treated in the appropriate manner as advised and set out in the Government Security Classifications (May 2018).

More information in relation to this data classification, including the requirements for working with these assets can be found here:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/715778/May-2018_Government-Security-Classifications-2.pdf

In particular please note that:

ALL routine public sector business, operations and services should be treated as OFFICIAL - many departments and agencies will operate exclusively at this level. There is no requirement to explicitly mark routine OFFICIAL information. Baseline security measures should be enforced through local business processes. A limited subset of OFFICIAL information could have more damaging consequences (for individuals, an organisation or government generally) if it were lost, stolen or published in the media. This subset of information should still be managed within the 'OFFICIAL' classification tier, but may attract additional measures (generally procedural or personnel) to reinforce the 'need to know'. In such cases where there is a clear and justifiable requirement to reinforce the 'need to know', assets should be conspicuously marked: **'OFFICIAL-SENSITIVE'**

The document is subject to CROWN COPYRIGHT. It is provided in confidence under existing laws, regulations and agreements relating to the protection of data and shall be so protected. The information contained herein is proprietary and shall only be used for the purposes intended at release. It shall not be reproduced, adapted or used in, whole or in part, for any other purpose without the prior written consent of the Secretary of State for DEPT in the Government of the United Kingdom.

Nothing contained herein should be construed as endorsing any particular Technical Solution to any United Kingdom Government Invitation to Tender.

Table of contents

Executive Summary	2
Links & Dependencies	5
1. Data Protection Impact Assessment Requirement & Process	7
Introduction.....	7
Data Protection Impact Assessment.....	8
Ownership	33
2. DPIA Report	33
Section 1: Overview of Data Collection and Maintenance	33
Section 2: Uses of the Application and the Data.....	34
Section 3: Data Retention.....	35
Section 4: Internal Sharing and Disclosure of Data	35
Section 5: External Sharing and Disclosure of Data	36
Section 6: Notice/Signage	Error! Bookmark not defined.
Section 7: Rights of Individuals to Access, Redress and Correct Data .	Error! Bookmark not defined.
Section 8: Technical Access and Security.....	37
Section 9: Technology	37
3. Compliance Checks	37
DPA 2018 Compliance Check	38
The Privacy and Electronic Communications Regulations.....	38
The Human Rights Act.....	38
The Freedom of Information Act.....	38
Conclusion.....	39
Annex A - Definition of Protected Personal Data	40
Annex B - Data Protection Compliance Check Sheet	41

Version:	V1.0

Links & Dependencies

Document	Title	Reference	Date	POC
DPA	Data Protection Act	All	2018	HMG
EU GDPR	EU General Data Protection Regulation	All	2016	GDPR
FOI	Freedom of Information Act	All	2000	HMG
Government Security Classifications	Government Security Classifications	All	May 2018	Cabinet Office
HRA	Human Rights Act	All	1998	HMG
ISO/IEC 27000	Information security management systems Standards	ISO/IEC 27001:2013	Oct 2010	ISO
IS1P1 & P2	HM Government Infosec Standard No. 1, Part 1 – Risk Assessment and Treatment	P1 - Issue 3.5 P2 – Issue 3.5	October 2009 October 2009	CESG
IS2	InfoSec Standard 2	Issue 3.2	January 2010	CESG
PECR	The Privacy and Electronic Communications Regulations	All	2003	HMG

1. Data Protection Impact Assessment Requirement & Process

Introduction

1. The General Data Protection Regulation (GDPR) 2016 introduces a new obligation to undertake Data Protection Impact Assessments (DPIAs), before carrying out types of processing **'likely to result in high risk(s) to individuals' interests'**. DPIAs are now mandatory for certain types of processing and there are specific legal requirements for content and process. Where a DPIA identifies a 'high risk' that cannot be mitigated, the Information Commissioner's Office (ICO) must be consulted.
2. DPIAs provide a way to systematically and comprehensively analyse the intended processing and help to identify and minimise data protection risks. In addition to considering compliance risks, they should also consider broader risks to the rights and freedoms of individuals, including the potential for any significant social or economic disadvantage. The focus is on the potential for harm to individuals or to society at large, whether it is physical, material or non-material.
3. To assess the level of risk, a DPIA must consider both the likelihood and the severity of any impact on individuals. It does not have to eradicate the risks altogether, but should help to minimise them and assess whether or not remaining risks are justified. A DPIA may cover a single processing operation or a group of similar processing operations. For new technologies you may be able to use a DPIA done by the product developer to inform your own DPIA on your implementation plans.
4. A DPIA must consider 'risks to the rights and freedoms of natural persons'. While this includes risks to privacy and data protection rights, it can also affect other fundamental rights and interests:
 - a. "The risk to the rights and freedoms of natural persons, of varying likelihood and severity, may result from data processing which could lead **to physical, material or non-material damage**, in particular: where the processing **may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage**; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data¹..."
5. Under GDPR you must carry out a DPIA where for example you plan to:
 - a. process special category or criminal offence data on a large scale.
6. The ICO also requires a DPIA to be undertaken for example, where you plan to:
 - a. use new technologies;
 - b. match data or combine datasets from different sources;
 - c. collect personal data from a source other than the individual without providing them with a privacy notice ('invisible processing');
7. DPIAs are an essential part of the organisation's accountability obligations under GDPR and an integral part of the 'data protection by default and design approach'. An effective DPIA helps to identify and fix problems at an early stage, demonstrate compliance with data protection obligations, meet individuals' expectations of privacy and help avoid reputational damage which might otherwise occur.

¹ GDPR - Recital 75

8. Conducting a DPIA is a legal requirement for any type of processing. Failure to carry out a DPIA in required cases may leave the organisation open to enforcement action, including a **fine of up to €10 million**.

General Description

1. NHSCFA leads on a wide range of work to protect NHS staff and resources from crime. In particular, it has national responsibility for tackling fraud, as this has been identified a key activity that would otherwise undermine the effectiveness of the health service and its ability to meet the needs of patients and professionals.
2. To achieve this, NHSCFA collects data appropriate for preventing and detecting fraud within the NHS, remaining mindful that, where this includes personal data, the personal data is adequate, relevant and not excessive for the purposes for which it is processed.
3. Data is collected through the process of data sharing with the wider NHS, public sector bodies, professional regulatory bodies, and through NHSCFA's Fraud and Corruption Reporting Line and online fraud reporting tool.
4. In relation to this specific DPIA, NHSCFA is utilising access to General Medical Council Regulated Data for the purpose of preventing and detecting fraud and other criminal offences within the NHS. Data is collected through the process of data sharing with the GMC in alignment with the NHSCFA/GMC MOU and through a formalised Data Sharing Agreement.
5. The overriding principles and purpose of this data sharing remains the same - by utilising subsets of the identified General Medical Council Regulated Data, NHSCFA undertakes proactive analysis to confirm the presence, extent and characteristics of a range of recognised fraud risks that have been substantiated through intelligence processes and/or existing proactive analysis undertaken by NHSCFA in the past.
6. NHSCFA additionally intends to undertake a range of machine learning techniques which will identify previously undetected fraud risks concerning irregular GMC claims and dispensing activities through a range of supervised/unsupervised methods and the utilisation and development of fraud classifiers drawn from this data and, through a range of proactive detection. This is characterised by the NHSCFA's "Project Athena", although for all intents and purposes the considerations within this document are applied universally, and the distinction of Project Athena and Business as usual is considered under one operating model.
7. This DPIA has been carried out by the Data Acquisition Manager, with the support of the Information Governance Officer Information and Records Management Officer, in consultation with Analytical intelligence Lead and the Information Governance and Risk Management Lead, as well as the relevant Data Science elements within Project Athena
8. In addition to GDPR the NHSCFA's use of data, including that for GMC data analysis (which this DPIA pertains to) is also required to comply with other relevant HMG legislation including where applicable the Data Protection Act 2018, Human Rights Act 1998 and Freedom of Information Act 2000.
9. In order to demonstrate best practice, provide assurance to stakeholders and assist with evidencing compliance with the above statutes and all legal other requirements, it is necessary to record and review that the risks to personal data are identified and understood. Therefore, it is necessary to undertake a Data Protection Impact Assessment ("DPIA") which is broken down into the following stages:

Data Protection Impact Assessment

10. To ensure data received from GMC meets all legal requirements and the risks to personal data are identified and understood it is necessary to undertake a DPIA. This DPIA is based on the ICO's

recommended template² comprised of seven steps, this is to ensure that the data that will be used in the analysis complies with the organisation governance principles:

Step 1 - Identify the need for a DPIA.

Step 2 - Describe the processing.

Step 3 - Consultation process.

Step 4 - Assess necessity and proportionality.

Step 5 - Identify and assess risks.

Step 6 - Identify measures to reduce risk.

Step 7 - Sign off and record outcomes.

STEP 1: Identify the need for a DPIA

Explain broadly what the system/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.

² Version 0.3 (20180209)

NHSCFA utilises secured access to GMC datasets for the purpose of preventing and detecting fraud and other criminal offences within the NHS through the utilisation of data for proactive fraud detection. More specifically, the purpose of this proposed data sharing is to:

- A) Support proactive data analysis of various GMC activity in order to determine and identify outliers which could be indicative of fraud. E.g. Time sheet fraud, Prescription fraud, identity fraud & so on
- B) Determine the scale and extent of recognised fraud risks to direct and bolster the commencement of a variety of proactive fraud prevention activity and supporting process that can remove fraud risks by removing identified system weaknesses and other fraud risk loopholes.
- C) Pre-empt the potential commencement of a criminal investigation and other forms of civil and criminal action by informing and supporting subsequent data shares concerning any fraud that is believed to have occurred and by directing the proportionality of accessing more explicit information.
- D) Support recovery methods for the GMC where payments are felt to be inappropriate and financial recoveries can be sought without prejudice.
- E) Undertake continued measurement to quantify the success of the above steps to determine the success of these measures.

The above steps necessitate full transparency in all steps undertaken and the designation of any and all outliers. As such, this data will not be subjected to any form of artificial intelligence (AI) and any machine learning algorithms will be developed using NHSCFA's own in-house expertise, with all processes and findings subjected to human oversight and review within a defined ethical framework.

Because the nature of the data itself (which concerns GMC regulated data activity) and the array of software and tools that can be utilised in accordance with the above steps, it has been considered necessary to consider the above activities collectively within the DPIA framework as a single entity for review. This is because there are specific risks and mitigations that must be considered separately from the data sources themselves and warrant special consideration and recording within a DPIA format – for example, although this information concerns NHS staff, the nature of the staff (and the data) relates to that which is already published as part of their national registration and subject to public availability to confirm the authenticity of regulated roles of doctors, consultants etc. Thus there is a need process personal data, but there are additional considerations which may apply that require a DPIA, as cited below:

- a) The GMC data provided to NHSCFA will remain identifiable at the point of receipt and use, as personal data is required for fraud detection and cannot be anonymised in advance. While upstream providers (i.e. GMC themselves) apply data minimisation by removing unnecessary fields or limiting access to relevant records, these steps support data protection but do not render the dataset anonymous for NHSCFA's use.
- b) To reflect that the tool utilises new technologies and concerns data at scale
- c) To qualify the determination of findings as “outliers” as opposed to fraud and reflect the protection of individuals from automated findings
- d) To confirm the NHSCFA remit as a competent authority and the basis and provision for participation in a data share that concerns PID and sensitive personal data, whilst recognising the protections put in place.

STEP 2: Describe the processing

Describe the nature of the processing:

1. How will you collect, use, store and delete data?
2. What is the source of the data?
3. Will you be sharing data with anyone (consider using a flow diagram or other way of describing data flow)?
4. Why types of processing identified as 'likely high risk' are involved?

1. How will you collect, use, store and delete data?

The GMC will prepare a secure view of its registration dataset, applying minimisation to their full dataset to that which is published within their register. As part of maintain a population dataset and adding daily updates they make it available within their controlled access environment. Nominated NHSCFA analysts will access the dataset via a secure login and transfer the data directly into NHSCFA's secure cloud-based analytical platform.

Once transferred, NHSCFA will create a working dataset for analysis, where data scientists and intelligence analysts will:

- Explore and transform the data
- Join it with a range of datasets to support use
- Identify outliers or anomalies that may indicate fraudulent activity
- Produce analytical outputs such as dashboards, statistical reports, or written assessments
- Summarise risk findings, which may include aggregated data, sample records, or excerpts, depending on investigative need

In particular the following will be examples (non exhaustive) of relevant uses cases for GMC

- GMC registration and licence status: To verify that individuals are legally registered and licensed to practise during NHS employment periods.
- Role vs. registration match: Compare NHS job roles (e.g. GP, locum, consultant) with GMC records to detect mismatches or false claims.
- Employment metadata: Includes contract type, region, and linked roles to assess for dual working or undeclared practice.
- GMC numbers and registration flags: Used to confirm clinician identity and prevent fraudulent use of medical titles.
- Revalidation and fitness to practise data: Supports checking regulatory compliance

Where justified, results (e.g. suspected outliers) may be shared internally within NHSCFA or with stakeholders, however this is unlikely to include the GMC data itself (and would not include PID).

All data will be:

- Stored securely on NHSCFA's cloud servers within the UK
- Not transferred to other locations or internationally
- Managed in accordance with NHSCFA's Information Asset Register, Retention Schedule, and Information Governance policies
- Data will be retained only as long as necessary for the specified analytical or investigative purpose. With each update or iteration of the dataset, previous data will either be appended or retained if required (e.g. to track changes over time or support model refinement). Where older data is no longer needed, it will be securely deleted following NHSCFA's records management procedures.

- Retention will also account for cases where backdated changes to historical records occur, which may themselves be fraud indicators. NHSCFA will regularly review its holdings and remove any data no longer serving a clear, defined counter fraud purpose.

2. What is the source of the data?

Data utilised by NHSCFA for the purposes identified above is collected from the GMC, who act as Data Controllers and collect and process the data as part of their own remit for payment and management of GMC services. The data is accessed through a licence agreement that supports access to and extraction of GMC specific datasets (the public registers)

3. Will you be sharing data with anyone (consider using a flow diagram or other way of describing data flow)?

NHSCFA will create an analytical product from the data which will be used to derive outliers that can be indicative of fraud. Findings can then be applied to drive fraud prevention activity or used as the basis to form a separate explicit request for data to act as evidential data for the purposes of criminal investigations.

Should NHSCFA wish to extract additional identifiable GMC data following identification of outliers that may be indicative of fraud (for example to support commencement of a criminal investigation), permission will be sought through wholly separate means through an entirely separate, specific and focused data access request, subject to the specific investigative powers of the NHSCFA. This request and subsequent data share will be managed entirely separately from this agreement and the wider solutions outlined above.

Describe the scope of the processing:

1. What is the nature of the data and does it include special category or criminal offence data?
2. How much data will you be collecting and using?
3. How often?
4. How long will you keep it?
5. How many individuals are affected?
6. What geographical area does it cover?

1. What is the nature of the data and does it include special category or criminal offence data?

The data relates to licensed medical professionals registered with the General Medical Council (GMC) and includes identifiable information such as name, date of birth, GMC number, registration status, and fitness to practise history. This constitutes personal data. No special category or criminal offence data is within scope.

No encryption or pseudonymisation is applied prior to NHSCFA access, as identifiable data is required to carry out fraud detection and matching activities. All processing will take place within NHSCFA's secure environment under strict access controls.

The legal basis for this processing is set out Step 4 of this DPIA.

2. How much data will you be collecting and using?

The NHSCFA data request regards granular data regarding all published GMC registry data in England. This concerns the full extent of the registry data table which is summarised in Appendix A

It is estimated 350,000 doctors are registered in the UK with full, provisional, and specialist or GP registration and this therefore gives an estimate to the scale of the data

3. How often?

The purpose of this data share is to implement a dynamic dataset which updates monthly.

4. How long will you keep it?

Extracted Data will be retained only as long as necessary, up to a maximum period in accordance with the Data Protection Act 2018. The maximum period depends on whether fraudulent behaviour is detected and the extent of any criminal proceedings.

Retention is managed via the NHSCFA information register, and, through it, schedules are set for deletion and removal. NHSCFA also proactively review their data and will remove records once there is no longer a clear and specific purpose for their continued use.

Retention of the data will be monitored and discussed via regular touch point meetings with the GMC and NHSCFA to ensure that data is not retained for longer than necessary.

5. How many individuals are affected?

It is impossible to fully quantify the full extent of how many individuals there are as this will change with each dynamic upload. Given that the full extent of the data covers all GMC activity in England, this is believed to be extensive, and an estimation from the GMC identifies an estimated 350,000 doctors are registered in the UK with full, provisional, and specialist or GP registration and this therefore gives an estimate to the scale of the data

6. What geographical area does it cover?

This relates to GMC activity in England only.

Describe the context of the processing:

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

1. What is the nature of your relationship with the individuals?

The individuals in scope are licensed medical professionals registered with the General Medical Council (GMC). These include GPs, consultants, locums, and other doctors employed within the NHS.

NHSCFA does not have a direct relationship with these individuals in a clinical or contractual capacity, but processes their professional registration data solely for the purposes of fraud prevention and detection.

2. How much control will they have?

Data subjects (i.e. GMC-registered medical professionals) will not have control over the use of their data in this context, as the processing is carried out under a legal basis for the performance of a task in the public interest — specifically, the prevention and detection of fraud within the NHS. As such, certain rights such as erasure and objection to processing do not apply (see section 3 of the DPIA report which outlines this further), and the NHS National Data Opt-Out is not relevant. These exclusions are permitted under the UK GDPR and Data Protection Act 2018, where processing is necessary for crime prevention or detection.

Additionally, the GMC register is public to ensure transparency and maintain trust in the medical profession. It allows patients, employers, and regulators to verify a doctor's credentials, registration status, and fitness to practise. This openness helps protect the public and supports safe, informed healthcare decisions. The GMC is also legally required to publish this information under the Medical Act 1983.

3. Would they expect you to use their data in this way

The actual approach is not novel in other organisations within the public sector (for example, within HMRC or DWP), but it will be for NHSCFA due to the NHS/healthcare dynamic.

Both GMC and NHSCFA make extensive efforts to ensure a degree of awareness and transparency is maintained through publication of information in the public domain.

GMC provided information that the data would be disclosed for the protection of patients ([Disclosures for the protection of patients and others - professional standards - GMC](#)) which would include the types of fraud aligned with false representation, working without registration and other types of fraud this concerns. They also notify that data will be used to detect and prevent serious crimes like fraud in the [GMC Confidentiality - Good Practice In Handling Patient Information](#).

4. Do they include children or other vulnerable groups?

GMC data does not include children or other vulnerable groups. Data relates only to adult healthcare professionals registered with the GMC. No patient or public data is involved.

5. Are there any prior concerns over this type of processing or security flaws?

The nature of NHS fraud and the data that exists in relation to it (in any context) it would be expected that this would have potential to be contentious as it is likely to concern the treatment of patients and may require information about the recipient of treatments or NHS employees who provided or supported this work. For this reason the DPIA has been undertaken.

6. Is it novel in any way?

This would be dependent on each specific approach application of data sources being used and techniques being applied, as these would range from data matching to machine learning techniques. Many of these would be considered novel in so far as they have only recently become part of the NHSCFA capability.

7. What is the current state of technology in this area?**GMC**

The GMC maintains its registration and licensing data using established systems that support the secure storage and management of professional records. NHSCFA will receive structured data extracts from the GMC for fraud detection purposes, extracting subsets of data that is made available on their website through a collated dataset created specifically for the NHSCFA (and in line with the NHSCFA specifications). This dataset will be made available in the secure environment within MS Fabrics, at which point it can be accessed and extracted by NHSCFA. More specifically, the data will be stored and accessed via the Medical Registers Download Service, a GMC tool that allows authorised external organisations to download the entire medical register in a single data file. Data files are generated on a daily basis, and organisations have the option of downloading the entire medical register (full), or a file that contains only records that have been updated in the last 24 hours (delta).

The Medical Registers Download Service uses the GMC's Secure File Transfer System (SFTS). Secure File Transfer Systems (SFTS) are purpose-built data storage and access tools, used to facilitate the secure exchange of sensitive data across internal and external networks. These systems employ encryption, access controls, and automated workflows to ensure data confidentiality, integrity, and traceability throughout the transfer process. By incorporating audit logging and policy enforcement, SFTS platforms support compliance with key data protection regulations such as GDPR, making them essential for organizations handling personal or regulated information. Secure File Transfer Systems (SFTS) are also designed to meet internationally recognized security standards, including **ISO/IEC 27001**. This certification specifies requirements for establishing, implementing, maintaining, and continually improving an information security management system

NHSCFA

The originating The NHSCFA data platform makes use of a range of tools and their use and role will be specific to the project, and therefore are best summarised in the corresponding DPIA for this toolset itself. However for the purposes of outlining the current operating model, the following applies

a) Microsoft Fabric

Following its purchase in November 2024, NHSCFA are utilising Microsoft Fabric as the next generation software for use in NHSCFA's data science framework. Fabric is an all-in-one data analytics solution that covers everything from data movement to data science, real-time analytics, and business

intelligence. It offers a wide range of data products available to users, which fall under the four buckets: Data ingestion, data storage, data engineering and data science and business intelligence.

Alongside a range of optimisation and efficiencies, Microsoft Fabric offers several governance features to help manage, protect, and monitor NHSCFA data and provide sufficient governance, this includes:

- An admin portal to control the overall Fabric estate, including tenant, domain, and workspace settings and group and individual permissions
- Data security features like data loss prevention, information protection, and metadata scanning to secure sensitive information
- Data Discovery and Trust: Tools to encourage data discovery, trust, and reuse, such as endorsement and data lineage
- Business continuity and disaster recovery tools to prevent data loss and corruption
- Monitoring and insights capabilities to monitor data access and usage and to uncover insights, and act on them, ensuring continuous improvement and compliance.

Fabric meets several UK-relevant compliance standards to ensure data security and privacy, including:

1. **ISO/IEC 27001**: Information security management
2. **ISO/IEC 27017**: Cloud-specific security controls
3. **ISO/IEC 27018**: Protection of personal data in the cloud
4. **ISO/IEC 27701**: Privacy information management

The full list can be found [here](#)

These tools serve a similar function to the tools that NHSCFA are using currently, including those summarised above, but collectively the tool provides a range of benefits and efficiencies, in particular the “one lake”, which eradicates data silos and the need to create multiple copies of a dataset. Additionally, all (tabular) data in One Lake is stored in one format, called “Delta Parquet” which is an open file format. This solves the data integration problem and data storage sizing as it’s a compressed file format and allows. Data scientists, data engineers, data analysts to work with the same data, in the same format.

Finally, Fabric provides a unified user experience and one access control method & one security model. This ensures access control and security is simplified and effective, with one access control method and one security model, applied across all tooling and experiences. Access to resources is principally managed through Workspaces, which are a collection of Fabric items, and because all data is within One Lake, data governance and discoverability become a much easier task through data lineage option. Fabric comes with a single built-in Monitoring Hub, which monitors all Fabric activity and thus supports enhanced security for all elements within the data science process.

b) Databricks

NHSCFA has implemented Databricks in Q4 of 2024/25, designed to work alongside Fabric for the actual analytical tool. Databricks is a powerful solution for data analysis that offers several compelling features:

- A unified platform that integrates data engineering, data science, and machine learning. This means you can manage your entire data lifecycle, from ingestion to analysis, all in one place
- Scalability and performance to scale data processing capabilities. It leverages Apache Spark, to handle large datasets efficiently and perform complex computations at high speed and helps reduce costs associated with data processing and storage

- Collaborative notebooks where teams can work together to foster better communication and collaboration among data scientists, analysts, and engineers
- Databricks supports advanced analytics, including real-time data processing and machine learning. It is possible to build, train, and deploy models directly within the platform, making it easier to derive insights and make data-driven decisions
- Databricks integrates with various data sources and tools, including SQL databases, cloud storage, and BI tools. This flexibility ensures you can connect to your existing infrastructure and enhance your data workflows
- Databricks prioritizes data security and compliance, offering features like data encryption, and tranching and bespoke levels of access control.

Similarly to Fabric, Databricks meets several UK-relevant compliance standards to ensure data security and privacy, including

- **ISO/IEC 27001:2022:** This certification is for information security management systems (ISMS), ensuring robust security controls and management practices
- **ISO/IEC 27017:** Focuses on cloud-specific security controls, providing guidelines for both service providers and customers
- **ISO/IEC 27018:** Addresses the protection of personal data in the cloud, ensuring privacy and data protection measures

Other Compliance Standards

- **SOC 2 Type II:** Databricks publishes that they undergo regular audits to ensure compliance with the Trust Services Criteria for security, availability, and confidentiality
- **PCI DSS:** Databricks supports compliance with the Payment Card Industry Data Security Standard, ensuring secure handling of credit card information

UK-Specific Compliance

UK Cyber Essentials Plus (UKCE+): Databricks hold certification for protecting against common cyber threats, including enhanced monitoring and encryption

8. Are there any current issues of public concern that you should factor in?

NHS fraud is, in itself, a matter of public concern and the use of data to combat it, particularly when this might concern information about patients and other service users, whilst ensuring that it is applied in an appropriate and proportional manner, are all issues of public concern.

This exercise supports the mitigation of risks that are of significant public concern, including NHS fraud and patient safety. The use of GMC data helps ensure that only appropriately registered and qualified doctors are working within the NHS, particularly in settings where staff may move frequently between organisations. Fraud involving impersonation or unverified agency workers can have serious implications for patient safety, organisational reputation, and financial integrity. In some cases, such risks may also intersect with broader issues such as modern slavery or organised crime, making vigilance and appropriate data use essential.

Relevant public concerns also extend to the protection of data from loss or theft and the controls and oversight to ensure this does not occur.

9. Are you signed up to any approved code of conduct or certification scheme (once any are approved)?

NHSCFA codes of conduct are summarised in the following documents:

Information Governance:

- Information Governance Policy
- Information breach Reporting Policy
- Data Quality Policy
- GDPR – Data Protection Policy

Information Security:

- Information Security Policy
- NHSCFA Acceptable use Policy

The NHSCFA has ISO/IEC 27001:2022 certification, awarded in October 2024, concerning on information security which covers information processed within the NHSCFA network.

Describe the purposes of the processing:

1. What do you intend to achieve?
2. What is the intended effect on individuals?
3. What are the benefits of the processing, for you and more broadly?

1. What do you intend to achieve?

Broadly speaking, this activity concerns the gathering and use of GMC data that NHSCFA requires to undertake proactive fraud detection. This activity supports the NHSCFA's remit of preventing, detecting and investigating fraud, corruption and unlawful activities against or affecting the Health Service in England.

The main functions of this data are:

- Identifying irregular patterns which are indicative of fraud.
- Developing processes to identify the scale of fraud and the impact of counter fraud activity.
- Supporting the creation of fraud investigations and informing requirements for more extensive analysis where outliers are detected that may be indicative of fraud
- Identifying, and mitigating, through outlier analysis any system weakness which are identified, for example those which may cause vulnerability to fraud or an inability to detect it.

2. What is the intended effect on individuals?

The intended effect on individuals is minimal, as the processing concerns data that is already publicly available and is conducted solely for the purposes of fraud prevention and does not result in any automated decisions or direct actions without human oversight.

Unlike pseudonymised dental datasets, GMC data includes identifiable information such as name, GMC number, date of birth, and licence status — which is necessary to verify professional identity and match against NHS employment records. Identification is therefore intentional and required for accurate fraud detection.

However, any decision to investigate or escalate a case is subject to internal NHSCFA review, with safeguards in place to prevent unjustified conclusions. The goal of this processing is not to penalise individuals directly, but to detect indicators of fraud, such as false registrations or unlicensed practice, which are then validated by human experts before any formal action is taken.

In this context, the effect on individuals is limited to those whose data appears anomalous or high-risk, and only to inform where further scrutiny is warranted in line with NHSCFA's statutory remit.

Given these controls, the impact on individuals for the NHSCFA processing is deemed minimal.

2. What are the benefits of the processing, for you and more broadly?

As an individual, there are no benefits for this processing..

More broadly, processing serves the prevention and detection of crime. More specifically to the NHS remit of the NHSCFA, the benefits serve effective protection of public services and effective management of the public purse. This overwhelmingly falls within the public interest

STEP 3: Consultation process

Consider how to consult with relevant stakeholders:

1. Describe when and how you will seek individuals' views or justify why it's not appropriate to do so?
2. Who else do you need to involve within your organisation?
3. Do you need to ask your processors to assist?
4. Do you plan to consult information security experts or any other experts?

1. Describe when and how you will seek individuals' views or justify why it's not appropriate to do so?

Because the nature of this processing concerns cases of suspected fraud, as well as the methods and mechanisms used to detect it, it is often not possible or appropriate to engage with individuals. Nonetheless, NHSCFA has historically engaged with a range of stakeholders, including with the National Data Guardian within consultations and workshops that concerned the use of personal data for fraud detection within healthcare, which included representatives of patient groups.

NHSCFA also makes use of extensive online resources to outline the proposed use of data and the approach taken and ensure transparency in terms of the application of data, as well as the estimated extent of designated fraud risks.

2. Who else do you need to involve within your organisation?

The approach for GMC analysis is drawn from a wealth of in-house expertise concerning data science, supported by intelligence and operational knowledge that is linked to fraud investigation. Additionally, the technological, information security and wider IT considerations are supported by a Technology Team with sufficient knowledge. Additionally, there are robust mechanisms for oversight in terms of Information Governance, wider corporate governance mechanisms which necessitate compliance and, finally, the existence of oversight groups such as the Data Strategy Group which can take decisions and approve /reject submissions concerning the application of data and the mechanisms for recording their outcome.

3. Do you need to ask your processors to assist?

There are no identified circumstances where wider / non NHSCFA processors would directly assist in the handling and processing of data being utilised by NHSCFA's data platform - the extent of external parties involvement would be limited to indirect support, for example in terms of insight generation and domain expertise concerning the data and its usage.

4. Do you plan to consult information security experts or any other experts?

These are sourced in house through designated resources and expertise provided by the appropriate departments and deemed adequate, in particular due to the compliance mechanisms achieved with regulatory Information security standards such as the ISO/IEC 27000 family.

STEP 4: Assess necessity and proportionality

Describe compliance and proportionality measures, in particular:

1. What is your lawful basis for processing?
2. Does the processing actually achieve your purpose?
3. Is there another way to achieve the same outcome?
4. How will you prevent function creep?
5. How will you ensure data quality and data minimisation?
6. What information will you give individuals?
7. How will you help to support their rights?
8. What measures do you take to ensure processors comply?
9. How do you safeguard any international transfers?

1. What is your lawful basis for processing?

The GMC register is public to ensure transparency and maintain trust in the medical profession. It allows patients, employers, and regulators to verify a doctor's credentials, registration status, and fitness to practise. This openness helps protect the public and supports safe, informed healthcare decisions. The GMC is also legally required to publish this information under the Medical Act 1983, specifically Section 30, requires the GMC to maintain and publish a register of medical practitioners. This includes details such as names, qualifications, and registration status. The purpose is to protect the public by ensuring transparency and enabling verification of a doctor's right to practise in the UK.

However, because this remains personally identifiable data, the following is also considered:

NHSCFA's remit

The NHSCFA's remit of preventing, detecting and investigating fraud, corruption and unlawful activities against or affecting the Health Service in England is defined as follows:

The powers conferred by sections 8, 272(7) and (8), and 273(1) of the National Health Service Act 2006 are primarily used by the Secretary of State for Health ("SofS" henceforth) to issue directions and regulations concerning the operation and administration of NHS services in England. It confirms powers and also allows the SofS to give directions to NHS bodies (such as NHS England, GMC, Integrated Care Boards, and others) regarding the exercise of their functions. It is a broad enabling provision that ensures the Secretary of State can maintain oversight and control over how NHS services are delivered, whilst deferring responsibilities to appropriate bodies.

Section 273(1) of the act allows the Secretary of State to give directions to certain NHS bodies or other entities for the purpose of carrying out the provisions of the Act. It complements Section 8 by providing specific authority to issue directions that are legally binding.

Part 10 of National Health Service Act 2006 contains a power for the SofS to require provision of documents in connection with the exercise of the SofS's counter fraud functions. The term "documents" is interpreted broadly and includes data, especially in digital form; as confirmed by the DHSC themselves in "Accessing and using documents to counter fraud in the NHS: Code of Practice (Part 10, National Health Services Act 2006" which states *"the term 'documents' includes information held in any form, including electronic records, emails, databases, and other digital formats."*

Section 195(1) of the National Health Service Act 2006 also establishes the legal foundation for the creation of a Special Health Authority to deal with fraud and other unlawful activities affecting the NHS. Section 195(2) defines the scope of this further, in terms of the SofSs 'counter fraud functions' in relation to the health service, this means the power (by virtue of section 2(1)(b)) to take action for the purpose of preventing, detecting or investigating fraud, corruption or other unlawful activities carried out against or otherwise affecting the health service.

Section 197(2) of the NHS Act concerns the power and requirement for the production of documents, or information in a specified form or manner, where there are reasonable grounds for suspecting that any documents containing information relevant to the exercise of SofS's counter fraud functions are in the possession or under the control of any NHS body, statutory health body, health service provider or NHS contractor.

Through the "NHS Counter Fraud Authority (Establishment, Constitution, and Staff and Other Transfer Provisions) Order 2017", created at formation of NHSCFA in 2017, NHSCFA was given the express function of the prevention, detection and investigation of fraud, corruption and unlawful activities against or affecting the health service in England.

Using Paragraph 4 (1) of these directions, the SofS denotes that responsibilities for the health service in England, which include preventing, detecting, and investigating fraud and other unlawful activities affecting the NHS are transferred to NHSCFA.

Counter fraud functions in relation to the health service means the power (by virtue of section 2(1)(b)) to take action for the purpose of preventing, detecting or investigating fraud, corruption or other unlawful activities carried out against or otherwise affecting the health service. The NHSCFA purpose for seeking this data (to investigate possible fraud concerning professional registration and identify concerns) clearly falls within the purpose for which such data can be requested and thus is within this requirement.

This requirement is also supported and further clarified by written guidance circulated in 2017, at formation of NHSCFA. The SofS provided Directions to NHS Trusts and Special Health Authorities in respect of Counter Fraud that confirms:

- a. The Secretary of State for Health, in exercise of the powers conferred by sections 8 and 272(7) and (8) and 273(1) of the National Health Service Act 2006(a), made directions that all NHS Bodies must co-operate with the NHSCFA to enable the NHSCFA to efficiently and effectively carry out its functions,
- b. in particular (para 3(1)) of this guidance states that "*each NHS body must supply such information including files and other data (whether in electronic or manual form) as the NHSCFA reasonably requires for the purposes of the NHSCFA's counter fraud functions*".

In this context "NHS body" is defined (within the same document) as meaning "*a body which is a Special Health Authority (other than the NHS CFA) or an NHS trust*". GMC is a Special Health Authority and a Arm's Length Body of the DHSC and is therefore subject to directions under the Secretary of State's powers.

NHSCFA as a competent authority

The ICO guidance: [Law enforcement processing: Part 3 DPA 2018; and sharing with competent authorities under the GDPR and Part 2 DPA 2018 | ICO](#), identify a competent authority

NHSCFA is granted this basis by function, under section 30(1)(b) of the Act, the NHSCFA has statutory functions for the law enforcement purposes, as enshrined by the NHS Act and the transfer of powers from the SoS.

Despite not being explicitly listed in Schedule 7, NHSCFA is a competent authority by function, not by explicit listing. The above ICO guidance confirms that an organisation does not need to be listed in Schedule 7 of the Data Protection Act (DPA) 2018 to qualify as a competent authority. Instead, it can qualify - even if not listed in Schedule 7 - as a competent authority having legal power to process personal data for law enforcement purposes, referencing back to Section 30(1)(b) of the DPA 2018 as the basis for this provision

This means that if an organisation has statutory powers and performs law enforcement functions (such as investigating fraud), it qualifies as a competent authority even if not explicitly named in Schedule 7. NHSCFA overwhelmingly performs this function and is granted the appropriate powers (see above)

This data share itself is underpinned by the following legislation:

Basis for accessing personal data

The basis for processing personal data within this data share is supported by paragraphs 6 and 10 of Schedule 1, Part 2 of the Data Protection Act 2018, specifically:

6(1) This condition is met if the processing—

(a) is necessary for a purpose listed in sub-paragraph (2), and

(b) is necessary for reasons of substantial public interest.

(2) Those purposes are—

(a) the exercise of a function conferred on a person by an enactment or rule of law;

(b) the exercise of a function of the Crown, a Minister of the Crown or a government department.

10(1) This condition is met if the processing—

(a) is necessary for the purposes of the prevention or detection of an unlawful act,

(b) must be carried out without the consent of the data subject so as not to prejudice those purposes, and

(c) is necessary for reasons of substantial public interest.

(2) If the processing consists of the disclosure of personal data to a competent authority, or is carried out in preparation for such disclosure, the condition in sub-paragraph (1) is met even if, when the processing is carried out, the controller does not have an appropriate policy document in place (see paragraph 5 of this Schedule).

(3) In this paragraph—

- *“act” includes a failure to act;*

11. “competent authority” has the same meaning as in Part 3 of this Act (see section 30).

The specific UK GDPR articles NHSCFA are relying on for this data sharing are as follows:

Article 6:

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

Article 9:

(g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

2. Does the processing actually achieve your purpose?

Yes. This DPIA supports a new data flow that builds on NHSCFA's existing processes and experience in using regulatory datasets to identify fraud risks across the NHS workforce. Access to GMC registration and licensing data will enable NHSCFA to validate the credentials of individuals working in roles that require medical registration, detect potential misrepresentation or dual working, and support targeted investigations.

The processing is necessary and proportionate for achieving the stated counter fraud purpose and aligns with NHSCFA's statutory remit under the NHS Act 2006. It provides a reliable basis for identifying fraud indicators and supports system-level assurance and prevention activity across the NHS.

3. Is there another way to achieve the same outcome?

GMC, as data controllers, are responsible for maintaining and publishing registers. However, fraud is a difficult crime to address since it is by its nature hidden - a factor that differentiates it from other types of error, loss and inefficiency. Fraud therefore necessitates the level of expertise and focus that NHSCFA are able to provide, and sufficient access to data to carry out its official duties. Additionally, NHSCFA capability is enhanced and focused purely on fraud and has the benefit of operational and intelligence which cannot be shared with other stakeholders due to their own sensitivities.

NHSCFA needs to undertake the data analysis directly for a number of reasons

- NHSCFA needs to be able to collate the implications of our analytical undertakings for our own internal governance and to withstand external scrutiny. To demonstrate and evidence the financial impact of our work we are subject to our own oversight and must be able to quantify and evidence our fraud impact savings. This can include authorisation of our findings by third parties (for example the PSFA Prevention Panel for any declared fraud prevention savings), so being able to collate and present our methodologies and findings is essential, which is only possible if we can undertake analysis directly.
- For the data itself, NHSCFA's need to extract is for the purposes of transparency and to create an auditable process concerning all elements of data handling given that any element may be subject to scrutiny (potentially in a court of law) our activity with the data needs to be undertaken by NHSCFA in our secure area. This also removes GMC from being accountable for evidencing this if we extract and undertake it ourselves within our tenancy and ensures that the full extent of our proactive data use (across all organisations) is managed centrally and consistently and within the compliance standards we have to maintain for our work.

There is a clear distinction between error and loss and other inefficiencies and fraud, which is due to the nature of fraud as a deceptive activity. This warrants both a distinct approach and specific types of expertise and resources. NHSCFA is the national body for preventing fraud in the NHS and thus has expertise of fraud that GMC does not have. Indeed, NHSCFA was created in 2017, and separated from GMC where it was hosted as NHS Protect, to provide a more focused, independent, and strategically aligned approach to countering fraud in the NHS—something that is harder to achieve within the broader operational scope of GMC.

Additionally, the statutory powers and counter functions that make NHSCFA a competent authority for processing personal data for counter fraud (see 20b) do not necessarily apply to other bodies, who

despite having the right to process data for wider functions may not be considered a competent authority to process PID for the counter fraud function

Consequently we consider this data share necessary.

4. How will you prevent function creep?

The NHSCFA has a range of oversight tools, ranging from Project Boards to the centralised Data Strategy Group, and supporting governance and assurance processes, which can manage and mitigate this risk (this is also reflected in key roles within the organisation in terms of SRO's etc).

5. How will you ensure data quality and data minimisation?

GMC, in their role as data controller, are responsible for the accuracy of gathered data, for both their own records and additionally for information provided to NHSCFA within this data share. Because NHSCFA has no means to audit or review this data for accuracy, it is accepted that this must be used with caution.

Outlier detection, by its nature, identifies inaccuracies and this is intrinsic to the data science framework and the validation process necessary to determine whether fraud has occurred. Information which is subsequently used to support fraud investigations will be assessed against other records and forms of evidence for accuracy as part of the case management process for fraud investigations.

STEP 5: Identify and assess risks

Describe source of risk and nature of potential impact on individuals.	Likelihood of harm	Severity of harm	Overall risk
Include associated compliance and corporate risks as necessary	Remote, Possible or Probable	Minimal, Significant, or Severe	Low, Medium or High

<p>1. There is a risk that personal data will be used for purposes other than that which are stipulated in the business case</p>	Possible	Significant	Medium
<p>2. There is a risk that complex processing of data by participating authorities during the analysis process may lead to information being inadvertently disclosed.</p>	Remote	Significant	Medium
<p>3. There is a risk that data disclosed are not required for the purposes of fraud detection, or are excessive.</p>	Remote	Minimal	Low
<p>4. There is a risk that incorrect information may be disclosed by participating authorities during the pilot.</p>	Possible	Significant	Medium
<p>5. There is a risk that data is retained for longer than it is needed</p>	Possible	Minimal	Low
<p>6. There is a risk that an individual's rights under UK GDPR are violated.</p>	Remote	Significant	Medium
<p>7. There is a risk that an external attacker gains access to personal data.</p>	Remote	Significant	Medium
<p>8. There is a risk that information could be lost, released or shared inappropriately</p>	Remote	Significant	Medium

OFFICIAL

<p>9. There is a risk that processing is carried out internationally in a territory without appropriate personal data protection in place</p>	<p>Remote</p>	<p>Minimal</p>	<p>Low</p>
<p>10. There is a risk that individuals will be misidentified as a result of data processing</p>	<p>Possible</p>	<p>Significant</p>	<p>Medium</p>
<p>11. There is a risk that the quality of data will not be to a consistently high standard</p>	<p>Possible</p>	<p>Significant</p>	<p>Medium</p>
<p>12. Risk that wrong inferences are drawn for the data due to a lack of knowledge of the context leading to wrong decisions being made about groups of individuals such as fraud investigations which are down to genuine error or misunderstanding of the data</p>	<p>Possible</p>	<p>Significant</p>	<p>Medium</p>

STEP 6: Identify measures to reduce risk

Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5.	Effect on risk Eliminated, Reduced, Accepted	Residual risk Low, Medium, High	Measure approved Yes/No
(1) There is a risk that sensitive data will be used for purposes other than that which are stipulated in the business case	(1), (2), (3) Robust governance structure and project management techniques are used to make clear roles and responsibilities, timelines, data flows, and contingency plans. The data itself has been minimised to remove personal data from the structured data, ensure that all possible steps have been taken to prevent inclusion of personal data (beyond that included in unstructured data)	Low	Yes
(2) There is a risk that complex processing of data by participating authorities may lead to information being inadvertently disclosed.	(1), (2), (3) Robust governance structure and project management techniques are used to make clear roles and responsibilities, timelines, data flows, and contingency plans. The data itself has been minimised to remove personal data from the structured data, ensure that all possible steps have been taken to prevent inclusion of personal data (beyond that included in unstructured data)	Low	Yes -
(3) There is a risk that incorrect information may be disclosed by participating authorities during the pilot (leading to incorrect identification of fraud)	(1), (2), (3) Robust governance structure and project management techniques are used to make clear roles and responsibilities, timelines, data flows, and contingency plans. The data itself has been minimised to remove personal data from the structured data, ensure that all possible steps have been taken to prevent inclusion of personal data (beyond that included in unstructured data)	Low	Yes
(4) There is a risk that incorrect information may be disclosed to unauthorised parties	(4) Data quality review is undertaken by authorities sharing data to review individuals / organisations included and remove irrelevant ones from the matching process.	Low	Yes
(5) There is a risk that data is retained for longer than it is needed	(5) Each data source is subject to a range of mapping and corresponding retention schedule. Deletion is undertaken manually, with human oversight. Retention to be monitored and identified as part of regular touch point meetings between NHSCFA and GMC.	Low	Yes

OFFICIAL

<p>(6) There is a risk that individual's rights under UK GDPR are violated</p>	<p>(6) This risk is not considered to be increased beyond business-as-usual levels as a result of the considerations in this DPIA and those in corresponding DPIA's for individual datasets</p>	<p>Low</p>	<p>Yes</p>
<p>(7) There is a risk that an external attacker gains access to personal data</p>	<p>(7) NHS CFA have approved and assured methods of managing data and for information security and are ISO27000 compliant. The data analytical platform itself is password protected and secured on the Cloud.</p>	<p>Low</p>	<p>Yes</p>
<p>(8) There is a risk that information could be lost, released or shared inappropriately</p>	<p>(8) This risk can be mitigated with robust governance structures, as well as security accreditation and adherence to a common set of data standards, set out in the security statement and information sharing agreement.</p>	<p>Low</p>	<p>Yes</p>
<p>(10) There is a risk that individuals could be misidentified as a result of data processing.</p>	<p>(10) Any decisions made using matched data will be informed by the strength of the match and resulting outlier. The strength of all outputs from the Data Platform will be assessed to manage false positives and triage follow up investigations.</p>	<p>Low</p>	<p>Yes</p>

OFFICIAL

<p>(11) There is a risk that the quality of data will not be to a consistently high standard</p>	<p>(11) Data quality reviews are routinely undertaken as part of the analytical process and the impact of these findings will impact on the outputs produced. In terms of outlier detection, low data quality can sometimes be a useful factor in determining fraud risks.</p>	<p>Low</p>	<p>Yes</p>
<p>(12) there is a risk that wrong inferences are drawn for the data due to a lack of knowledge of the context leading to wrong decisions being made about groups of individuals such as fraud investigations which are down to genuine error or misunderstanding of the data.</p>	<p>(12) NHCFA have designated all findings as "outliers that could be indicative of fraud". NHSCFA will work closely with the data to ensure a good understanding of the data and review all findings with human oversight</p>	<p>Low</p>	<p>Yes</p>

STEP 7: Sign off and record outcomes

Item	Name/date	Notes
Measures approved by SMT Owner:		Confirmed approval to authorise
Residual risks Approved by SMT Owner:		Confirmed approval to authorise
DPO advice provided		DPO should advise on compliance, step 6 measures and whether processing can proceed.
<p>Summary of DPO advice: Having reviewed the DPIA I am satisfied that a comprehensive assessment of the GMC Analysis Exercise, with regard to the acquisition, storage, security and subsequent retention of personal data has been carried out. The GMC will prepare and apply 'minimisation' to their full registration dataset, making it available within their controlled access environment. Nominated NHSCFA analysts will access the dataset via a secure login and transfer the data directly into NHSCFA's secure cloud-based analytical platform. Access will therefore be fully auditable.</p> <p>The dataset will be made available in the secure environment within MS Fabrics which offers several features to help manage, protect, and monitor the data and provide sufficient additional tier of governance. The data will primarily be accessed by approximately 30 members of staff from NHSCFA, which includes the database administrators and Data Engineers.</p> <p>All data will be held and governed in accordance with current legislative requirements and handled in accordance with organisational best practice and its data retention policy. I am therefore satisfied with the with the organisational security measures employed.</p>		
DPO advice accepted or overruled by:	2025	29 th July If overruled, you must explain your reasons
Comments:		
Consultation responses reviewed by:		If your decision departs from individuals' view, you must explain your reasons
Comments:		
This DPIA will be kept under review by:		The DPO should also review ongoing compliance with DPIA

Ownership

The following table describes the roles and responsibilities

Table 1 - Roles and Responsibilities

Role	Responsibility
Information Asset Owner (IAO)	
Senior Information Risk Owner (SIRO)	
Application/Database Owner	
Data Protection Officer	

3. DPIA Report

Section 1: Overview of Data Collection and Maintenance

1. The impact level of this DPIA concerning NHSCFA GMC Analysis is assessed as OFFICIAL.
2. The following measures briefly describe what controls have been implemented to protect the Data Analytics Platform and the GMC data which will be accessed, stored and utilised within it. This Platform is subject to its own DPIA which provides greater detail. The combined considerations of this assessment are:
 - a. The data concerned by NHSCFA GMC analysis is primarily accessed by approximately 30 members of staff from NHSCFA, which includes the database administrators and Data Engineers. However, outputs and wider data disseminations are specific to individual projects and their designated outcomes.
 - b. The Data Analytics Platform upon which the GMC data is accessed and utilised has a separate DPIA. This platform has direct interconnections with other NHSCFA systems and applications, particularly in terms of data drawn in from other NHSCFA data sources. Each of these have their own access control measures and controls in place to mitigate any risk of unauthorised access.
 - c. The Data Custodian must comply with the data protection requirements. Examples include: regularly reviewing the business requirement to record the personal data; ensuring that the data is not excessive; it is being used for the purpose intended; that there is a deletion and disposal policy; that the application is registered on the NHSCFA register and the NHSCFA DPO is aware of its existence.
3. This activity does not rely on encryption or pseudonymisation prior to processing, as the ability to identify individuals is essential for the purposes of counter fraud detection. NHSCFA requires access to key identifiable fields such as GMC number, name, and date of birth to match professional registration records against NHS workforce data and detect potential misrepresentation or unauthorised practice. While patient-level pseudonymisation was appropriate in previous datasets, that approach is not suitable in this context, as identifying discrepancies between claimed employment and valid GMC registration requires direct identifiers. Nonetheless, all data will be stored and processed in a secure, access-controlled environment, and use will be strictly limited to fraud detection activity under NHSCFA’s statutory remit.

12. In this case, the data provided by the GMC will be identifiable by design, as it is necessary to confirm the validity of registration, licence status, and associated professional history. As such, no pseudonymisation or anonymisation is applied prior to NHSCFA access. The ability to directly identify individuals is critical to the effectiveness of fraud detection and cannot be removed without undermining the purpose of processing.

Despite this, robust technical and organisational controls are in place to minimise any privacy risks. These include secure storage within NHSCFA's Data Analytics Platform, role-based access controls, audit logging, and strict limitations on secondary use. Data is used exclusively for legitimate counter fraud purposes, in line with NHSCFA's legal powers and the terms of the relevant Information Sharing Agreement (ISA).

It is therefore assessed that there are no unmanaged residual privacy risks associated with NHSCFA's access and use of GMC data under this DPIA. All risks identified are proportionately mitigated through existing governance, oversight, and legal frameworks.

This DPIA will be reviewed:

If there are changes to the personal data being used

If the NHSCFA operating model changes

Or if there are any other developments that could affect the privacy rights of data subjects

Section 2: Uses of the Application and the Data

4. This concerns the gathering and use of GMC data that NHSCFA requires to fulfil their remit of preventing, detecting and investigating fraud, corruption and unlawful activities against or affecting the Health Service in England.
5. NHSCFA utilises secured access to GMC datasets for the purpose of preventing and detecting fraud and other criminal offences within the NHS through the utilisation of data for proactive fraud detection. More specifically, the purpose of this proposed data sharing is to:
 - A) Support proactive data analysis of GMC activity in order to determine and identify outliers which could be indicative of fraud.
 - B) Determine the scale and extent of recognised fraud risks to direct and bolster the commencement of a variety of proactive fraud prevention activity and supporting process that can remove fraud risks by removing identified system weaknesses and other fraud risk loopholes.
 - C) Pre-empt the potential commencement of a criminal investigation and other forms or civil and criminal action by informing and supporting subsequent data shares concerning any fraud that is believed to have occurred and by directing the proportionality of accessing more explicit information.
 - D) Support recovery methods for GMC where payments are felt to be inappropriate and financial recoveries can be sought without prejudice.
 - E) Undertake continued measurement to quantify the success of the above steps to determine the success of these measures.
6. Outputs concern the following:
 - a. Analysts will undertake analysis to determine outliers using a range of appropriate measures, this could range from rule based analysis or data matching to the development of machine learning models and algorithms, as dictated by the problem centric approach and insight that needs to be derived.
 - b. The creation of formal records that determine the activity and outcome, usually in the form of some type of written report that describes the extent of the fraud risk, the

methodology applied and the insight that has been possible from the activities above. This will necessarily utilise the data to demonstrate these points and highlight the appropriate considerations and reflects necessary (this may be through aggregation of datasets, or utilization of specific pieces of data to demonstrate outliers, provide examples and determine salient points for consideration)

- c. Where there is a need to formalise the outputs drawn from the findings, submission of a Technical Appendix document, which may include specific references to the data, will be submitted to the Data Strategy Group to allow oversight and approval of the outcomes.

11. The IAO must comply with data protection requirements. Examples include: regularly reviewing the business requirement to use the personal data; ensuring that the data is not excessive, it is being used for the purpose intended; that there is a deleting and disposal policy; that the application is registered and the DPO is aware of its existence.

Section 3: Data Retention

12. The source data within the GMC secure environment can be replaced with each iteration of updated Data. NHSCFA will be responsible for extraction and storage of any datasets which are required under the Data Sharing Agreement.

13. Retention is managed via the following formal documents which are maintained by NHSCFA, through these, datasets are mapped in terms of their content, origin and, schedules are set for deletion and removal. NHSCFA also proactively review their data and will remove records once there is no longer a clear and specific purpose for their continued use:

- Data Retention Schedule
- Information Asset Register
- Inventory of Information Transferred (IIoT)

Section 4: Internal Sharing and Disclosure of Data

14. The GMC data requirement is sourced to support fraud investigations and support fraud prevention activity, therefore findings will be shared internally with NHSCFA Fraud Investigators and other specialists in the following context

- Sharing findings (outliers) internally for the purposes of utilizing domain expertise for verification/validation.
- Using findings for intelligence, loss measurement and investigation and/or to support fraud prevention activity to mitigate identified fraud risks
- Analysts will utilize data to undertake analysis to determine outliers in order to determine and identify outliers which could be indicative of fraud.
- The outcomes of which would be summarized in a written report that describes the extent of the fraud risk and may contain excerpts or samples of data, as well as descriptive and analytical summaries of their content, format, application and results.
- Depending on the findings themselves, the basis for sharing and processing the data and the strength of the outliers, it is possible that the outliers may be shared with the GMC themselves or NHSCFA's own investigative services for further action.

Information Analysts produce a data-based output, for example an electronic dashboard or statistical physical report detailing output and findings, that is utilized to support a summary of the NHSCFA activity and outputs which may contain sample, aggregated or summarized excerpts of the data.

Section 5: External Sharing and Disclosure of Data

15. Depending on the findings themselves from the exercise concerning the data within this DPIA, the basis for sharing and processing the data and the strength of the outliers, it is possible that the outliers may be shared with the GMC in the pursuit of validation/verification of results. This is encapsulated in the Data Sharing Agreement and additionally, as GMC are both data controllers and originators of the data share, does not incur risk

The only wider reason information of this type would be shared externally is presented in a summary statistical format and/or presented in aggregate

16. Individual outliers may inform wider data shares for explicit data in specific circumstances where fraud is detected, if necessary for the administration of justice and/or in accordance with an appropriate Information Sharing Agreement/Memorandum of Understanding and supporting DPIA. Such considerations would be entirely separate to those associated with this programme of work.

Section 6: Notice/Signage.

17. NHSCFA's privacy policy on its website hosts separate sections in relation to data collection, retention and storage. This broadly covers all elements of the NHSCFA usage of data and provides contact details to contact NHSCFA in any capacity (or to make an information request)
18. Section 5b of Schedule 15 of GDPR negates the need to directly inform data subjects where a) the provision of such information proves impossible or would involve a disproportionate effort, or b) such notification may render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.
19. NHSCFA maintain a range of pages online that outlines our use of personal data and provides an in-depth mechanism for informing patients, service users and any stakeholders about the activity NHSCFA provides, the basis under which it acts and the standards NHSCFA holds itself to in terms of managing records. NHSCFA also have a mechanism for answering queries or concerns which is advertised on these pages.
20. The use of signage or other notifications to notify the public of the gathering and use of personal data for wider systems is not relevant to this Data Platform and therefore outside the scope of this DPIA.

Section 7: Rights of Individuals to Access, Redress and Correct Data

Right to Access

13. Individuals subject to certain exemptions, have the right to gain access to their own personal data. In the event an access request is directly or indirectly received by NHSCFA, we are required to provide the individual who has made the request with details of the personal data recorded about them. It is unlikely that many access requests will be received as the personal data recorded is primarily held by GMC.
21. In the unlikely event that information is identified as being incorrect, NHSCFA staff will take appropriate steps to correct the record where permissible. All NHS employees and members of the public have the right to request access, redress and correct personal data recorded about them, however this may need to be considered in light of the below

Right for deletion / redress/ correction

14. Corrections to GMC data will be managed as part of their usual process for maintaining the accuracy of their data and will be updated in subsequent iterations. This does not concern this data share.
15. Were GMC to be subject to a request for deletion by a data subject who wishes their data amended or deleted from that processed by NHSCFA for the purposes outlined, we note that 3b of Article 17 of the UK GDPR “the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller” would apply and be sufficient to deny this request.
22. The basis for this refusal concern A) the prevention and detection of fraud within the NHS overwhelmingly supports the public interests and B) the defined function of NHSCFA as an official authority, supported the NHSCFA Establishment Order and the Secretary of State for Health Directions. This is in line with guidance from the ICO which identifies the appropriateness of the above.

Right to object

23. Given that none of this data concerns direct marketing, the absolute right to object is not relevant. Nonetheless individuals have the right to object to the processing of their personal data at any time, even where a task is carried out in the public interest, in line with official authority and/or legitimate interest.
16. Again, the ICO have provided guidance on the Right to object | ICO which confirms that the applicant must give specific reasons why they are objecting. NHSCFA/GMC would need to determine if they have compelling legitimate grounds which override the interests of the use in detecting/preventing fraud. If an individual cites that processing is causing them substantial damage or distress (e.g. the processing is causing them financial loss), the grounds for their objection will have more weight. This, however, is very unlikely.
17. It would be necessary to document the considerations and outcomes and may also be possible to give assurance through the extent that GMC/NHSCFA already partly comply with this request through use of the masking techniques. However, the answers above provide a generic response, which may not encapsulate wider reasons that GMC has to consider requests of this nature, associated with their own basis for processing. In the interests of ensuring that specific considerations were made for individual circumstances

Section 8: Technical Access and Security

24. The security and technical access architecture of the Database/System is as explained in this DPIA. The application and the hosting infrastructure was assessed at Official-Sensitive and the hosting infrastructure is subject to the ISO27000 and ISO27001
25. Access to the system itself, with the exception of disseminated products, is restricted to internal staff only.
26. The technical controls to protect the database include:
 - a. Anti-virus protection;
 - b. Permission based access controls to shared drive.
 - c. Logging, audit and monitoring controls.
 - d. Vulnerability Patching Policy for the underlying infrastructure.

Section 9: Technology

27. The Database/System holds personal information obtained electronically and is located in the NHS Counter Fraud Authority cloud infrastructure, subject to the principles of use outlined in the following NHSCFA policies

4. Information Governance Policy
 5. Information breach Reporting Policy
 6. Data Quality Policy
 7. GDPR – Data Protection Policy
 8. Information Security Policy
 9. NHSCFA Acceptable use Policy
28. The Data Analytical platform has its own additional DPIA which provides further information specific to the technology and should be sourced for specific information relating to the software and how it is applied.

10. Compliance Checks

DPA 2018 Compliance Check

1. The DPO must ensure that the System, and the personal data that it records, and its business activities, are compliant and maintain compliance with:
 - a. The GDPR and the Data Protection Act in general;
 - b. The Data Protection Principles;
 - c. The interpretations of the Principles.
2. **This is not a recommendation but a requirement of law.**
3. The roles and responsibilities for the protection of personal data are described in the NHSCFA security policy.
 11. The system processes sensitive personal data and so a Data Protection Compliance Check Sheet has been completed (see Annex B) describing how the requirements of GDPR and the Data Protection Act 2018 have been complied with, See; also Annex A Category C

The Privacy and Electronic Communications Regulations

4. The Privacy and Electronic Communications Regulations is not applicable as personal data is not exchanged with external organisations for commercial purposes.

The Human Rights Act

5. The decisions and activities of the organisation are undertaken in compliance with the Human Rights Act, having due regard to appropriateness and proportionality to ensure compatibility with Convention rights.

The Freedom of Information Act

6. As a public authority we are compliant with the provisions of the Freedom of Information Act, in proactively publishing and making available upon request, certain recorded information held by the organisation subject to any relevant exemption(s). However, no personal information would be disclosed under the Freedom of Information Act as this would breach the data protection principles.

Conclusion

7. There are no residual privacy risks to the personal data that will be used in the production of management information. The controls described in this DPIA explain in detail how the data is protected and managed in accordance with the GDPR and Data Protection Act 2018. The DPO is responsible for ensuring that the controls are implemented through the life cycle of the system.

Annex A - Definition of Protected Personal Data

Personal data includes all data falling into Categories A, B or C below:-

A. Information that can be used to identify a living person, including:

Name;
Address;
Date of birth;
Telephone number;
Photograph, etc.

Note: this is not an exhaustive list.

B. Information which, if subject to unauthorised release, could cause harm or distress to an individual, including:

Financial details e.g. bank account or credit card details;
National Insurance number;
Passport number;
Tax, benefit or pension records;
DNA or fingerprints;
Travel details (for example, at immigration control or oyster records);
Place of work;
School attendance/records;
Material related to social services (including child protection) or housing casework.

Note: this is not an exhaustive list.

C. Sensitive personal data relating to an identifiable living individual, consisting of:

Racial or ethnic origin;
Political opinions;
Religious or other beliefs;
Trade union membership;
Physical or mental health or condition;
Sexual life
Commission or alleged commission of offences;
Proceedings relating to an actual or alleged offence.

Any data set containing this information must be processed in accordance with the GDPR and the Data Protection Act 2018 (DPA 2018).

Particular care must be taken with data in Category B and with any large data set. Information on smaller numbers of individuals may justify additional protection because of the nature of the individuals, source of the information, or extent of information.

There are additional, specific constraints within the provisions of GDPR and the DPA (2018) on the processing of data in Category C.

Annex B - Data Protection Compliance Check Sheet

PART 1: BASIC INFORMATION - New or existing Project, System, Technology or Legislation

1. Organisation and project.

Organisation	NHSCFA
Branch / Division	NHSCFA Performance, Projects and Analytics / Project Athena
Project	GMC Proactive Analysis

2. Contact position and/or name, telephone number and e-mail address.

(This should be the name of the individual most qualified to respond to questions regarding the DPIA)

Name, Title	
Branch / Division	
Phone Number	
E-Mail	

3. Description of the programme / system / technology / legislation (initiative) being assessed.

(Please note here if the initiative does not collect, use or disclose personal data*). If this is a change to an existing project, system, technology or legislation, describe the current system or program and the proposed changes.

NHSCFA utilises secured access to GMC datasets for the purpose of preventing and detecting fraud and other criminal offences within the NHS through the utilisation of data for proactive fraud detection by using the published registration information data to confirm the veracity, accuracy and appropriateness of regulated clinicians (those subject to GMC regulation) to deliver treatment.

4. Purpose / objectives of the initiative (if statutory, provide citation).

NHSCFA is utilising access to GMC datasets for the purpose of preventing and detecting fraud and other criminal offences within the NHS. Data is collected through the process of data sharing with the GMC in alignment with their publishing of data and an approved licence agreement, this DPIA represents an assessment to outline all considerations, given that this remains PID.

The overriding principles and purpose of this data sharing remains the same - by utilising subsets of the registry data, NHSCFA undertakes proactive analysis to confirm the presence, extent and characteristics of a range of recognised fraud risks that have been substantiated through intelligence processes and/or existing proactive analysis undertaken by NHSCFA in the past (i.e. false representation, working without registration etc.).

NHSCFA additionally intends to undertake a range of machine learning techniques which will identify previously undetected fraud risks concerning irregular activities through a range of supervised/unsupervised methods and the utilisation and development of fraud classifiers drawn from this data. and, through a range of proactive detection. This is characterised by the NHSCFA's "Project Athena", although for all intents and purposes the considerations within this document are applied universally, and the distinction of Project Athena and Business as usual is considered under one operating model.

NHSCFA remit from the 2017 Secretary Of State Directions supports data sharing in pursuit of fraud and supports NHS Bodies / Special Health Authorities in providing data. This approach is also supported by legislation [Data Protection Act 2018 \(DPA\) Schedule 8, sections 1 and 8](#) (i.e. necessary for enactment of rule of law, in substantial public interest, necessary for the purposes of preventing fraud or a particular kind of fraud,

As the act concerns the UK implementation of the GDPR principles, the specific GDPR articles we are relying on for this data sharing, which are as follows:

- a. Article 6: (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
- b. Article 9: (g) processing is necessary for reasons of substantial public interest

This DPIA has been carried out the Data Acquisition Manager, with the support of the Information Governance Officer, in consultation with Analytical intelligence Lead and the Information Governance and Risk Management Lead, as well as the relevant Data Science elements within Project Athena

5. What are the potential privacy impacts of this proposal?

Because the nature of the data relates to professional registration and licensing information for individual medical practitioners, and because NHSCFA’s purpose involves validating professional identity, it is not appropriate—or technically feasible—to anonymise the data prior to processing.

GMC data must remain identifiable to enable effective matching with NHS workforce, payroll, and clinical systems. As such, fields such as GMC number, name, and date of birth are essential and form the basis of analytical checks used to identify potential fraud.

This DPIA therefore treats the data, process, and use of analytical tools collectively, recognising that: The identifiability of the data is necessary and proportionate to the fraud prevention purpose
Robust governance and access controls are in place to manage the confidentiality risks
NHSCFA does not attempt to further minimise identifiability, because doing so would undermine the utility of the data for its intended purpose

We assess that while the data is identifiable, the overall privacy risk is appropriately mitigated through:
Secure NHSCFA systems (ISO27001-compliant data environment)
Role-based access controls
No automated decision-making (human review is always required)
Data is processed solely within NHSCFA under a legally defined counter-fraud function, with no wider disclosure or reuse

Unlike pseudonymised dental data (where identification would require reverse engineering of an encrypted key held only by GMC), the GMC data is used in its identifiable form, and so ICO anonymisation tests (such as the "motivated intruder" test or the "whose hands?" test) are not applicable in this context. Nonetheless, NHSCFA continues to apply appropriate safeguards in line with ICO guidance on proportionality, transparency, and accountability. The use of this data is covered by:

- A clearly defined legal basis (UK GDPR Article 6(1)(e), Article 9(2)(g))
- A formal Information Sharing Agreement (ISA)
- Transparent privacy notices from both GMC and NHSCFA
- Regular reviews of processing, including this DPIA, in line with changes to NHSCFA’s operating model or data use

Conclusion:

Because this dataset requires identifiability and is processed solely for lawful counter fraud purposes under strict governance, privacy risks to individuals are assessed as minimal and proportionately mitigated. The DPIA will be reviewed periodically and updated if the nature of processing or data content changes.

6. Provide details of any previous DPIA or other form of personal data* assessment done on this initiative (in whole or in part).

This is the see previous versions cited in version control, additionally there is a wider DPIA undertaken for the Data Analytics Platform that relates to the tools used for this activity DPIA carried out on the system.

IF THERE IS NO PERSONAL DATA INVOLVED, GO TO PART 3: DPA COMPLIANCE – CONCLUSIONS:

*IMPORTANT NOTE:
‘Personal data’ means data which relate to a living individual who can be identified:
(a) from those data, or

(b) from those data and other information, which is in the possession of, or is likely to come into the possession of, the data controller,

And includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

See Q5 (above) which outlines the mitigation of risk and additionally outlines how personal data risks are limited. All elements pertinent to this activity have therefore been considered and risk assessed and have found to be appropriate/proportional to the risk. Furthermore, the anonymisation undertaken of the data is proportional and any concerns that amalgamation of datasets could potentially lead to the identification of person identifiable information, is so remote that any risk would be considered negligible.