Genomics England

IG Data Access and Acceptable Uses Policy

Programme	**Genomics England**	**Document Record ID Key**	
Sub Programme	IG		
Programme Director	Grant Stapleton	**Status**	Final
Document Owner	Caroline Moth	**Version**	2.7
Document Author/s	Grant Stapleton	**Version Date**	13/11/2015

Author: Grant Stapleton

Document history

Version history

Version	Date	Description
0.1	01/07/2014	Initial draft
0.2	01/09/2014	Updated draft in support of the draft Protocol
0.3	13/11/2014	Consolidated policy in-line with Protocol version submitted to REC
04	14/11/2013	Updated following comments from reviewers
0.5	19/11/2013	Final version
0.6	25/11/2013	Updated following review from all stakeholders
2.2	28/11/2014	Text on informatics updated
2.3	02/12/2014	Head of Bioinformatics and Programme Director for Informatics
2.4	19/12/2014	Amendments after review by Ethics Advisory Group
2.5	16/01/2015	Updated after comments from General Counsel and NHSE
2.6	26/05/2015	Update following peer review
2.7	13/11/2015	Updated following a review by IG lead

Reviewers

This document must be reviewed by the following:

Name	Title
Grant Stapleton	Head of Service Management/ SIRO
Ed Stafford	Director of Informatics
Nick Maltby	General Counsel
Graham Colbert	Chief Operating Officer
Dr Tom Fowler	Director of Public Health and Caldicott Guardian
Laura Riley	Ethics Programme Manager
Jim Davies	Chief Technology Officer
Data Advisory Committee Membership	Members of the Data Advisory Committee

Approvers

This document must be approved by the following:

Name	Responsibility	Date	Version
Nick Maltby	General Counsel, Genomics England	13/11/2015	2.7
Graham Colbert	Chief Operating Officer	13/11/2015	2.7
Dr Tom Fowler	Director of Public Health and Caldicott Guardian	13/11/2015	2.7

The approvers will act as representatives for the following committees:

- Scientific Advisory Committee
- Information Governance Steering Group
- Genomics England Programme Board

Formal changes to the document must be controlled by the Genomics England Programme Board (the Board) or as delegated by the Board. The latest version of this document is available from the Genomics England Programme Management Office. Printed copies of this document are not controlled.

Related Polices

This document should be read in conjunction with the following documents:

Title
Genomics England Information Quality and Records Management Policy

Key terms

This document may include the following key terms:

Abbreviation	Expansion and Definition
ARC	**Access Review Committee** – The ARC are responsible for reviewing all requests for data access. The ARC analyses and reviews the appropriateness of access requests, looking at the specifics of the request and is able to suggest amendments.
Caldicott Review	The original **Caldicott** Report, published in 1997, resulted in the establishment of seven principles for NHS bodies (and parties contracting with such bodies) to adhere to in order to protect patient information and confidentiality.
DAC	**Data Advisory Committee –** The DAC is accountable to the Board and responsible for maintaining an overview and driving the broader IG agenda within Genomics England
Data Protection Act (DPA)	The ***Data Protection Act 1998*** **(DPA)** is an Act of Parliament of the United Kingdom of Great Britain and Northern Ireland which defines UK law on the processing of data on identifiable living people. It is the main piece of legislation that governs the protection of personal data in the UK.

Abbreviation	Expansion and Definition
IAs	**Information Assets** – IAs are information which is central to the efficient running of an organisation. IAs will include computer systems, network hardware and software which are used to process this data.
IAO	**Information Asset Owner** - IAOs are accountable to the Information Governance Lead (IG Lead) for ensuring that all information assets are recorded in the Information Asset Register (IAR), and to provide assurance to the SIRO that associated risks are managed effectively.
IAR	**Information Asset Register** – An IAR is a mechanism for understanding and managing an organisation's assets and the risks to them.
ICO	**Information Commissioner's Office** – The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.
IG Lead	**Information Governance Lead** – the role of the IG Lead is to coordinate, publicise and monitor the standards of information handling throughout the Genomics England. The role will also include leading on Caldicott, Data Protection and Freedom of Information issues ensuring that employees are fully informed of their own responsibilities for maintaining the standards and that information about the standards is made available to the public where appropriate.
IG Toolkit	**Information Governance Toolkit –** The IG Toolkit is an online system which allows NHS organisations and partners to assess themselves against Department of Health Information Governance policies and standards.
GMC	**NHS Genomic Medical Centre -** GMCs are responsible for identifying and enrolling participants, as well as collecting samples and providing engagement with and feedback to participants.
PCD	**Personal Confidential Data** is a term used in the Caldicott Information Governance Review and describes personal information about identified or identifiable individuals, which should be kept private or secret and includes dead as well as living people. The review interpreted 'personal' as including the *Data Protection Act 1998* definition of personal data, but included data relating to the deceased as well as living people, and 'confidential' includes both information 'given in confidence' and 'that which is owed a duty of confidence' and is adapted to include 'sensitive' as defined in the Data Protection Act. Examples of identifiable data are: name, address, postcode, date of birth and NHS number.
Personal data ***(as per the DPA 1998)***	**Personal data**, as defined in the *Data Protection Act 1998*, is information which can directly identify a person – in which the person is the focus of the information and which links that individual to details that would be regarded as private (e.g. name and private address, name and home telephone number etc.). This definition also includes members of staff and contractors.

Abbreviation	Expansion and Definition
	It is important to note that, where the ability to identify an individual depends partly on the data held and partly on other information (not necessarily data), the data held will still be "personal data".
Sensitive personal data ***(as per the DPA 1998)***	**Sensitive personal data**, as defined in the *Data Protection Act 1998*, is personal data consisting of information relating to the data subject with regard to racial or ethnic origin; political opinions; religious beliefs or other beliefs of a similar nature; trade union membership; physical or mental health or condition; sexual life; the commission or alleged commission by the data subject of any offence; or any proceedings for any offence committed or alleged to have been committed by the data subject, the disposal of such proceedings or the sentence of any court in such proceedings
SAR	**Subject Access Request** – when a person (data subject) makes a request for their own information (data) under the *Data Protection Act 1998*.
SIRO	**Senior Information Risk Owner** – the SIRO has overall responsibility for ensuring that effective systems and processes are in place to address the IG agenda. The SIRO is responsible for information risk within the organisation and the provision of written advice to the Executive Board on the content of the Governance Statement in regard to information risk.

Contents

1.0 Introduction and Objective

Data is defined in the Genomics England Protocol as either:

i) the data held within the Genomics England data centre about participants within the 100,000 Genomes Project (service data); or

ii) the data held by Genomics England relating to its business, its customers and users, its staff etc. (corporate data).

Paragraph 9.2.2 of the Genomics England Protocol (service data sharing), states that:

All data held within the Genomics England data centre that relates to 100,000 Genomes Project data subjects is considered to be service data.

Service data will be shared, within an agreed governance framework, in two ways:

i) With internal staff (under substantive contract of employment) within Genomics England, in order to check data quality and integrity; to undertake agreed *cleansing* activities; to deliver initial bioinformatics analytics on the data, and to undertake other monitoring, audit and service improvement activity. It is noted that internal access to all data held shall be controlled, monitored and reported upon. Access to identifiable data, which is held separately, must be explicitly logged and approved by the nominated Genomics England Information Governance Lead (**IG Lead**).

ii) With external users and consumers, as follows:

a. In the form of proactive publication of anonymised, aggregated information as summary metadata about the work of Genomics England, according to an Information Commissioner's Office (**ICO**) approved schema that meets Freedom of Information requirements on transparency and has a very low risk of re-identification. Further information can be found in the *Genomics England Freedom of Information Publication Scheme.*

b. Delivery of access (within the Genomics England data centre) to agreed datasets for approved users, including scientists undertaking research into whole genome data. Data will be anonymised, based on the ICO code of practice, unless a legal gateway exists for access to identifiable information.

Delivery of limited access to agreed datasets for those requesting information under the *Freedom of Information Act 2000* (FOI request) or the *Data Protection Act 1998* (**SAR request**), or under any other regulatory or legal basis for disclosure. Data will be anonymised, based on the ICO code of practice, unless a legal gateway exists for access to identifiable information.

Further information can be found in the *Genomics England Data Access and Acceptable Uses* policy.

2.0 Scope

This document outlines the requirements for:

- Controlled access to service data and granting of access to certain sets of data, to approved users.

- Use of the dataset by external users, including data processors engaged by Genomics England, and researchers.

This policy provides a statement of how Genomics England will work in regards to data access, and defines basic principles for decision making. It is not a standard operating procedure (**SOP**) because it does not define ways of working. Subsequent SOPs may be created that adhere to the policy and ensure organisational consistency, coherence and compliance. This document is key to decision-making on data access and sharing, and should be used in conjunction with other applicable Genomics England policies, based on the type of request at hand.

Where possible we have utilised or adapted the approaches used by the UK Biobank Access Sub-Committee. Where feasible and appropriate this policy has been harmonised to corresponding policies for NHS England, Health Education England, Health and Social Care Information Centre (HSCIC) and Public Health England, as well as the Department of Health.

3.0 Key Principles

1. All principles expressed in this document are based on the current version of the 100,000 Genomes Project Protocol and may be amended as the programme evolves.
2. Activity within the 100,000 Genomes Project is based on the donation of samples and associated consent for sample storage, DNA sequencing, and other analysis of the participants' tissue and clinical data for the purposes of processing by research and commercial users.
3. Genomics England will process and deliver decisions on data sharing and access requests keeping in mind the public interest, scientific utility, the corresponding consent, and wider Genomic England policies.
4. Genomics England intends to offer comprehensive access to enable leading edge outputs from analysis of the data that are expected to have both scientific research value and clinical value particular to the participants whose data is involved.
5. For each request, Genomics England will determine the appropriate cohort of genomic and associated data to make available within the approved trusted environment.
6. Data access will only be granted to users validated by Genomics England, using traceable IP addresses, who have a data-sharing contract or are approved staff of Genomics England.

7. Genomics England will have procedures for monitoring user activities and interaction with the data, and for the management and escalation of suspected non-compliance.

8. Decisions relating to data access and data sharing will be made considering the policy provisions given in this document, as well as further detailed in corresponding SOPs.

9. This policy and associated SOPs shall be made available to all Genomics England staff; and corresponding training and awareness activity relevant to their role shall be provided as appropriate.

10. Genomics England shall create and operate an Access Review Committee (**ARC**). This shall include an independent chair, and appropriate independent experts in areas such as genomic medicine and ethics. It will receive advice and input as needed from internal committees of Genomics England such as the Ethics Advisory Committee and the Data Advisory Committee. A member of Genomics England staff will support the Committee including a Senior Information Risk Owner who may call upon the advice of a Caldicott Guardian for all data held.

11. Genomics England will process access applications and provide decisions via the independent ARC on behalf of the Genomics England Executive Board.

12. Genomics England will utilise best practice guidance and privacy enhancing technologies to provide de-identification of data, meet anonymisation standards, and minimise the risks of inadvertent disclosure.

13. Compliance with the policy shall be audited and verified as required. Non-compliance shall be reported to the Genomics England Programme Board and appropriate contractual and legal action taken. Actionable decisions in the case of a breach will be under the remit of the Caldicott Guardian.

14. Arrangements for archiving of the data when it is no longer required will be set out in the Genomics England Information Quality and Records Management Policy.

15. Parties will sign legally binding data access contracts with Genomics England that outline the terms of access and processing information, which includes as a breach of contract any attempts to re-identify participants.

16. Users will have a facility to introduce new data sets into the Data Centre for wider research purposes as defined in Section 8 of the Genomic England Protocol document. This data will then be subject to the same rigorous data access, sharing and acceptable uses procedure as set out in this document. Genomics England will ensure compliance with governance requirements across the pipeline regardless of data source to safeguard data confidentiality and participant privacy.

3.1 Trusted Environments

The Project will establish trusted environments whereby users will be able to interact safely with the gathered genomic and other associated data.

Data within the secure virtual data centres will be de-identified. Data users will be able to declare and bring tools and data into the environments, with the user organisation and users having explicitly committed not to attempt to re-identify any participant.

Data intended for extraction will be through an 'air-lock' where the data will be verified as appropriate for export, either by meeting the Anonymisation Standard ISB 1523 or as appropriate for the needs of the research purpose and the potential scientific/clinical value.

3.2 Types of Acceptable Use

Acceptable uses for external users are defined in the Genomics England Programme Protocol section 9.3.1.

No.	Definition	Explanation
1	Clinical care	Use of data to directly diagnose/treat a participant.
2	Hypothesis driven research and development in health and social care – observational	Hypothesis driven review and analysis of data collected within the Genomics England Data Centre.
3	Clinical trials feasibility	Queries on the broad location and numbers of participants eligible to participate in interventional research project.
4	Hypothesis driven research and development in health and social care – interventional	Hypothesis driven review and analysis of agreed cohorts of participants within a research study.
5	Non hypothesis driven R&D - health	Review and analysis to identify associations between data and variables - to improve our understanding of the causes of disease.
6	Non hypothesis driven R&D - non health	Use of the data by data scientists to test the effectiveness of tools and models for data analysis.
7	Public health purposes	Work to gain a population level understanding of the prevalence of disease and corresponding health protection actions where possible.
8	Other health use - clinical audit	Use of the data to assess whether clinical standards are met.
9	Other health use - commissioning/commissioning policy	Use of Genomics England data to inform decision-making by any relevant commissioning body/organisation developing commissioning policy.

		Trust use to undertake local planning and administrative tasks.
10	Tool evaluation and improvement	Use of the data to validate, improve and deliver new annotation tools for WGS data.
11	Interpretation and validation of the Genomics England Knowledge Base	Use of the data to interpret annotation findings and validate results for use in a clinical context.
12	Education and training of health and public health professionals	NHS professionals and others learning how to analyse, interpret, and apply genomic medicine.
13	Deeper phenotyping	Gaining further information on participants with particular genetic results.
14	Non hypothesis driven R&D - health	Review and analysis to identify associations between data and variables - to improve our understanding of the causes of disease.

3.3 Excluded Uses

The types of purposes that will typically be refused outright following consideration by the Genomics England Informatics Team and the Genomics England ARC are defined in and can be referred to in the Genomics England Protocol Document in Section 9.3.2. Examples of excluded uses include:

- Paternity testing and level of relatedness testing (though participants can request their data and as such if several family members make this request and take the data to a third party to analyse it is conceivable that the data can be used in this way).
- Parental request for child's data once the child reaches 16 where either the participant assents or dissents.
- Uses requiring linkage to employment records, tax records, benefits records or personal life insurance records for non-scientific or non-healthcare related purposes.
- Uses requiring linkage for personal insurance or forensic purposes.
- Purposes that might lead to discrimination against a person or a group of people, based on genetic/genomic characteristics.

3.4 Complex Cases

In cases where disclosure is not covered by our consent, but may be required by law, requests fall under the category of complex cases.

If these types of purposes are proposed – they will either be refused outright, or, at the discretion of the Genomics England ARC, referred to the Genomics England Programme Board for consideration and approval on a case-by-case basis. Additionally, all requests for identifiable data that are required by court order will be notified to the Genomics England General Counsel

and the Executive management team. All disclosures required by court order should be referred to Genomics England's General Counsel as promptly as possible, so that all representations may be made to the court, for example to limit the information requested.

The Genomics England decision-making process on all occasions shall be documented and transparent, and will involve external advice.

4.0 Access Review Committee (ARC)

The ARC will analyse and review the appropriateness of access requests and responding based on the specifics of the request, with the possibility of suggesting amendments. It is modelled upon the tried and trusted procedures adopted by UK Biobank and therefore the ARC will be composed of an independent majority representation. In addition, the ARC will include membership of participants or participant representatives from each of the disease groups in scope (rare disease, cancer and infection).

This Committee is expected to support the widest possible access to the data providing it accords with the terms of informed consent[1]. This Committee is intended to ensure the access to 100,000 Genomes Project data accords with the principles outlined in the *Genomics England Acceptable Uses and Data Sharing* policy. For example, non-compliance with the terms of access will be flagged for the ARC to review whether or not this may have been a reasonable. Further details regarding this Committee can be found in the ARC Discussion document and the ARC Terms of Reference. The Committee will operate a rigorous but streamlined process where possible to provide timely approvals. We anticipate much of the committee's business will be done by email or teleconference.

A summary of decisions and recommendations will be notified to applicants and published on the Genomics England website. This transparency should assist the participants, the public and applicants with information on ongoing or recently approved applications for access to the 100,000 Genomes Project dataset.

Key operational relationships of the ARC are described in below.

[1] Taking informed consent means that a trained health professional has explained what participation involves. They are responsible for making sure that the potential participant has a clear understanding of the risks and benefits of taking part and what they are being asked to take part in. This information is included in the Participant Information Sheets which form the basis of the discussion between the potential participant and the healthcare professional. All participants in the 100,000 Genomes Project will have been through this process.

4.1 Organisational Roles and Responsibilities

This document will be approved by the Data Advisory Committee (**DAC**) and administered by the IG Lead but may be delegated to other relevant committees/boards in support of the ARC. Compliance with this policy will be monitored by various work streams within Genomics England as outlined in our Standard Operating Procedures, when adopted.

Section 8.2 of the 100,000 Genomes Project Protocol, 4th paragraph states:

> *Genomics England takes data security and participant privacy extremely seriously, and will take action where a breach or near breach is proven. Penalties for non-compliance include Genomics England's revocation of user access, organisation access, and reporting the offending activity to the Information Commissioner. We will where possible notify individual participants affected if we discover a data breach notifications enabling them to either exercise their 'right to object' to further processing of their personal data.*

The Caldicott Guardian will be responsible for decisions relating to penalties over non-compliance.

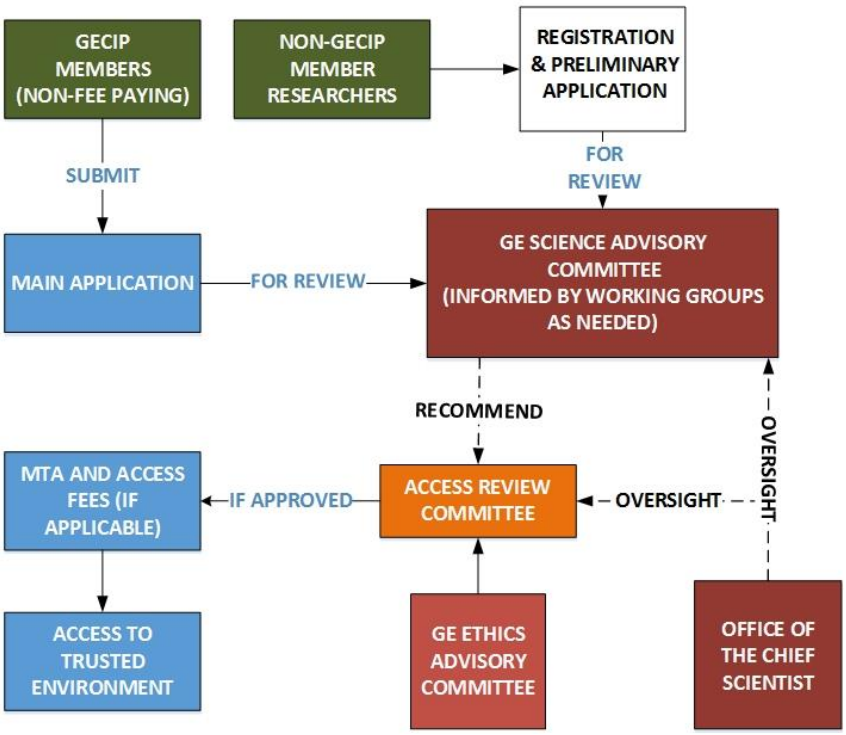
The publication of the results obtained from analysing the data within the 100,000 Genomes Project will be undertaken by individual researchers or groups of researchers, NHS healthcare teams, and trainees.

The work to identify, log, classify and commission responses to requests for data access shall be undertaken by the ARC Secretariat.

5.0 Procedural Requirements

Genomics England will administer this policy and related procedures. There will be a log of all requests, and tools to monitor and report on compliance to the conditions of data sharing. Additionally, the Programme Board will ensure that the Chief Technology Officer and Chief Operating Officer establish adequate procedures are in place for documentation, tracking, monitoring, and reporting for the overall process of data sharing.

DATA ACCESS APPLICATION AND REVIEW PROCESS



6.0 Audit and Monitoring Criteria

This policy and adherence to it will be audited regularly will be monitored. Compliance with this policy, framework and procedures are undertaken through the IG Toolkit annual assessment and an audit undertaken on the level of assurance provided by the submitted evidence.

The table below sets out how we will monitor implementation and utilisation of this Policy.

Document Audit and Monitoring Table	
Monitoring requirements "What in this document do we have to monitor"	We will ensure that staff are aware of this policy, the constituent aspects of the IG framework, and abide by legal, technical and mandatory IG requirements. Performance in the IG Toolkit and the completion of Data Flow Mapping and Information Asset Register (**IAR**).
Monitoring Method	IG Toolkit annual assessment
Monitoring prepared by	IG Lead
Monitoring presented to	Programme Board, DAC and Operational Management Team
Frequency of presentation	Quarterly internal review and presentation of procedure status

	Annual independent audit, submissions in line with Department of Health Guidance