

Airlock Policy

Document Key: 428	Version: 2.0	Approval Date: 16 Jul 2020 15:43	Effective Date: Two weeks after the approval date
Document Author: Anna Need	Document Approver: Nick Maltby	Status: Live	
Department: GeCIP		Document Type: Policy	
Read & Understood Required: Yes		Training Session Required: No	
Next Review Date: 16 Jul 2021		Total Pages: 9	

Contents

1. Document Version History	3
2. Scope and applicability.....	3
3. Dataset and cohort definitions.....	3
4. Other Definitions.....	4
5. Common policies for airlock.....	6
6. Policies for airlock export.....	7
7. Exported material.....	8
8. Policies for airlock import.....	9

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 2 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

1. Document Version History

Version <i>(newest on top)</i>	Date of Approval	Summary of main changes and reasons
2.0	As per cover sheet	Policy updated to reflect a) we would in some circumstances allow processing of data outside the RE b) we have multiple Res c) to use the correct wording around the National Genomic Research Library

Note: The document version history detailed above is effective for the last live version of this document only. Details of changes for prior versions are available in EQMS.

2. Scope and applicability

This policy applies to any research carried out on the Genomics England Dataset (**The Dataset**) by GeCIP, GEL Staff or Discovery Forum members within any Genomics England **Research Environment**. Airlock applications are not required for export of data to be used in clinical care (i.e. export of data by NHS GMCs or GLHs), such applications can be made to the via the Genomics England Service Desk. Furthermore, Airlock application is not required for import or export (including whitelisting of websites) by Genomics England Service Providers for the purpose of service provision or agreed service improvement activity. These requests can be made directly to the Genomics England Research Environment team (who may reject if they have security or support concerns) and must be covered as part of agreed contractual arrangements. However, a complete record of agreements must be kept to provide assurance to Airlock Review Committee on all data movements.

3. Dataset and cohort definitions

The Dataset - all data contained within the Genomics England **Research Environment** from cohorts that make up **The National Genomic Research Library** and **External data** sources.

Data types from cohorts include, but are not limited to:

- Clinical and personal data provided at registration, or as part of any longitudinal/surveillance monitoring specific to a cohort;
- Data from any tests or procedures carried out on the participant's samples given at registration or during follow-up;
- Any and all 'omics data (including all genomic, transcriptomic, metabolomic, proteomic data) derived from participant's samples.

National Genomic Research Library (The Library) - a secure national resource of genomic, health data and samples managed by Genomics England in partnership with NHS England (see Protocol: "The National Genomic Research Library"). Participants whose data and samples are held within **The Library** are recruited through a variety of mechanisms including the **NHS Genome Medicine Service**, **The 100,000 Genomes project**, **The COVID-19 project** and other cohorts and under different consents. Cohorts under different consents are managed through the principles of Airlock, taking into account differences in stringency or contractual rules that may affect handling of subsets of data.

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy	Page 3 of 9	
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

NHS Genome Medicine Service (GMS) cohort - WGS and all other genomic and health data data of those NHS patients who are treated though the NHS GMS and consent to research.

100,000 Genomes cohort - information associated with, or derived from, any individual’s participation in the 100,000 Genomes Project, or associated with, or derived from, the samples they provide over the course of the project.

COVID cohort- information associated with, or derived from, any individual’s participation in the Genomics England COVID-19 Project or associated with, or derived from, the samples they provide over the course of the project.

External data - data accessible from the **Research Environment** that was not collected as part of a National Genomic Research Library cohort (e.g. 100,000 Genomes Project, the COVID Research Project etc.). This includes, but is not limited to:

- additional data on 100,000 Genomes Project participants that was collected as part of a separate research project, and has subsequently been made available in the Research Environment;
- data from public data repositories (e.g. 1000 Genomes Project) that has been made available within a Genomics England Research Environment (either via direct data import, or via 'whitelisting' of web resources);
- data from restricted access datasets (e.g. dbGAP) that have been made available centrally within a Genomics England Research Environment;
- data from ‘private’ datasets, that are within a Genomics England Research Environment (either with restricted or unrestricted access).

4. Other Definitions

Airlock – the process by which information and data is securely moved in or out of any Genomics England Research Environment.

Airlock Policy Guidelines – an evolving document overseen by the Airlock Review Team that describes the various likely data types that will be submitted for airlock export, and the principles/rules-of-thumb that will guide the researcher during the preparation of their analyses, and the output checker during their review and decision.

Airlock Review Committee – a delegation of the Chief Scientist responsible for oversight of all airlock requests in accordance with the Airlock Policy and the group’s Terms of Reference. It must include the Senior Information Risk Officer (SIRO), the Caldicott Guardian and at least one User representative.

ARC approval – The current approval from the Genomics England Access Review Committee under which any analyses pertaining to an airlock request have been undertaken.

Dataset Owner – The person Genomics England considers to be the nominated contact person for a particular piece of External data (e.g. a nominated individual from the study PIs). For some external data, this responsibility may be delegated to the Office of the Chief Scientist.

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 4 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

Genomics England Publication Policy – the policy that all publications detailing analyses carried out the Dataset must adhere to.

Identifiable data - Data that can be used on its own, or with other easily obtained information, to identify, contact, or locate a single person, or that might identify an individual within a given context.

Individual-level data (aka “Row-level data”) - data, both raw and calculated, from the Dataset that is presented on a per-participant basis (regardless of whether the participant is anonymised or not), presented in written or graphical format.

Information - is distinct from data in this context in that information is not derived from participants in a research study (examples include a SNP list, chromosomal coordinates, or non-human cell line ‘omics data).

Output checker – an individual who assesses statistical data within an export request for any risk of disclosure, following the principles described in the Airlock Policy Guidelines, and who will recommend to the ART whether the data should or should not be approved for export.

Research - any research carried out on the Dataset by GeCIP, GEL Staff or Discovery Forum members within any Genomics England Research Environment.

Research Environment – Any Genomics England analysis environment that allows researchers to securely access **The Dataset** of de-identified genomic and other associated health data.

Research Registry – The catalogue of all research projects being undertaken on the Dataset.

Row-level data – see Individual-level data

Staging Environment – Any Genomics England analysis environment that allows researchers to develop and test software against synthetically generated test datasets which have the same structure as the data held in the Genome England Research Environments but which contain no real data from cohort participants.

Summary data - is defined as data calculated from multiple observations across multiple participants or measurements within **The Dataset**, presented in written or graphical format. Graphical summaries presenting individual data points (e.g. scatter plots) would not be considered summary data.

Whitelisted website - a website to which read-only access is possible from within a Genomics England **Research Environment** (an environment that otherwise cannot access the internet).

5. Common policies for airlock

- 5.1. Users of Genomics England Research Environments will comply with all reasonable requests associated with bringing data in and out of the research environment.

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 5 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

- 5.2. An airlock request will be assessed considering:
- its intended use
 - its compatibility with the relevant ARC approval;
 - its implications for the security of the Dataset;
 - the risk for participant identification;
 - the technical feasibility of the request;
 - any costs Genomics England may incur by fulfilling the request.
- 5.3. Airlock requests will not necessarily be processed in the order in which they are received. Succinct, clear, well-prepared requests consisting of a small number of items will be prioritised over more complex requests.
- 5.4. Genomics England will aim to perform an initial review of all requests within two working days, although at peak times (e.g. close to conference deadlines), this may take up to five working days. Certain types of requests will not need further review by the Committee (see [Guidelines](#)).
- 5.5. A decision on requests that have to be referred to the Airlock Committee will be made within 7 working days. Any delay to review beyond this will be communicated to the applicant in writing, stating the reason for the delay. This 7-day limit is suspended if Genomics England has requested further information from the Requestor and has not yet received it.
- 5.6. The decision of Genomics England will be communicated to the applicant within one working day of a decision being made.
- 5.7. Approved requests will aim to be actioned within one working day of a decision being made, though complex requests (e.g. import of software packages) may take longer.
- 5.8. Applicants will be expected to promptly reply to any questions or queries raised during the review process or make any required changes to the material in a timely manner. Where this is not the case, Genomics England reserves the right to reject the request due to insufficient information being provided or lack of engagement from the applicant. Requests will be closed if an applicant is asked a question and does not reply within 10 working days.
- 5.9. A request may be rejected if the meaning of the data is ambiguous and the researcher does not provide sufficient information to allow the Airlock reviewers to understand what exactly is being proposed for export/import.
- 5.10. The applicant may appeal any final decision by Genomics England and should do so in writing to the Chief Scientist within 25 working days of the decision, clearly stating how they believe the decision of the Airlock Review Team contradicts the terms of this policy.
- 5.11. The Chief Scientist will consider all appeals and respond with a final decision within 25 working days of the receipt of the appeal.
- 5.12. Breach of the airlock (for example by 'screen-shots' of a Genomics England

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 6 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

Research Environment or copying data long hand), or contravention of the policies of the airlock (for example, using exported materials for uses other than those within the original request) will be considered a serious breach of the terms of data access. Genomics England reserves the right to ban the researcher's institution, and all their researchers, from access to the Dataset and pursue legal or criminal action against both the individual and the institution.

6. Policies for airlock export

- 6.1. It is expected that all analyses will be carried out within a Genomics England Research Environment and that export will be of analysis results (summary data) only.
- 6.2. Export of data for use in further analysis (this does not include changing the format of the data) will only be permitted exceptionally if that analysis is well justified and cannot reasonably be carried out within the Research Environment. In cases where data is permitted for export for further analysis, the following must be adhered to:
 - a) The data files will not be used for any purpose other than the specific work stated in the export request.
 - b) Any external environment in which the data files are stored must be secure. Genomics England may impose specific technical criteria to areas where the data are to be stored/analysed.
 - c) Once the specified analyses are completed, the resulting data will be uploaded to the Research Environment for further analysis and any requests to export data outside the Research Environment will be subject to the usual airlock process. Download of data for analysis outside the Research Environment does not confer agreement to use or share the results of those analyses outside the Research Environment.
 - d) The data files will be deleted from all locations that they are stored outside the Research Environment once this analysis is completed.
 - e) Genomics England reserves the right at its discretion to refuse export requests for similar data for similar stated purposes in future as the capability of a Genomics England Research Environment improves such that similar work can be performed inside the Research Environment, or may refuse the request(s) for other reasons, which will be explained in full when each request is made.
- 6.3. Binary files will not be considered for airlock export.
- 6.4. Any data for export must have been generated in line with the applicant's current ARC approval and must have been generated as part of a specific approved project within the Genomics England Research Registry. Commercial companies engaged in pre-research may only export data which is necessary to assist them to formulate a research plan.
- 6.5. Generally, only summary data will be eligible for export. Export of individual-level data will be considered for, but not limited to, instances in which:

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 7 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

- a) it forms part of a graphical summary of the data (e.g. a scatter plot) that is sufficiently complex that individual values for the data point are not discernible.
- b) It allows reporting of phenotypes associated with rare genetic variants, and the associated phenotypic information is not sufficient to specify a subset of fewer than five participants. In cases where it is unclear whether the associated phenotypic data could be identifying, the Genomics England clinical team will be consulted.
- c) The data are identifying (e.g. because they are to be used in a Case Report) and the researcher can demonstrate that they have been in contact with the recruiting clinician and the participant or their guardian has given explicit consent for the publication.

6.6. While Genomics England may allow the export of individual-level data on a case-by-case basis, it will be considered a breach of the airlock for researchers to use this process to reconstruct individual-level datasets outside the Research Environment.

6.7. Any export should contain the minimum set of information required to fulfil its purposes i.e. no extraneous variables or data points should be included.

6.8. No export should contain any participant identifiers, or other labels that could be considered identifiable data.

6.9. Any export for the purposes of public sharing, e.g. presentation, publication, or other purpose covered by the Genomics England Publication Policy, must be submitted referencing an approved project within the Research Registry and be formatted in accordance with the publication policy.

6.10. Any data for export must have been prepared considering the relevant statistical disclosure control principles as outlined in the Genomics England Airlock Policy Guidelines. The review of statistical output will be conducted by output-checkers who will consider its suitability for export taking into account its disclosure risk given the intended use.

6.11. Where complex software packages are brought into the RE and modified within the RE, only the specific modifications may be exported, not the entire modified software package, as it is not feasible to review large software packages for export.

7. Exported material

7.1. Exported material

- a) can only be shared publicly if this was stated in the original airlock request;
- b) can only be used for those purposes stated, and agreed to, in the original airlock request.

7.2. Any intellectual property contained within, or produced from, the exported material will, unless otherwise agreed, be wholly owned by Genomics England. See Genomics England IP policy for further details.

7.3. The Research Environment contains External Data (for example Hospital Episodes

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy	Page 8 of 9	
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		

Statistics [HES]) which is subject to data sharing framework contracts and data sharing agreements between Genomics England and other parties that dictate how the data may be used and what can be exported. Where an export contains External Data, Genomics England will always apply the requirements placed on them as conditions of having access to the data. In some cases, particularly concerning the export of individual-level data, these will be more conservative than those applied to National Genomic Research Library cohort data alone.

7.4. Where agreed, an export that features External Data may need to be brought to the attention of the Dataset Owner, who may, if raised in sufficient time, veto the export.

8. Policies for airlock import

- 8.1. Import of any data, tools, or other material will only be considered if its intended use is in line with the applicant's ARC approval, the relevant ethical approvals and is part of an approved, registered research project.
- 8.2. Imported data may require a Data Sharing Agreement to be entered into. The importer is responsible for ensuring they have the necessary rights to import the data and any consents are consistent with its use in the research environment.
- 8.3. Where a researcher is paying for their own storage and compute within the infrastructure of a Research Environment, they are free to import software or data themselves for their use (subject to compliance with 28-29)
- 8.4. Where a researcher wishes Genomics England to cover the storage and management costs of hosting data or software tools, 32-35 also apply.
- 8.5. Import of licensed software will only be considered if:
 - a) there is no comparable software within the relevant Genomics England Research Environment;
 - b) there is no equivalent open source software package available;
 - c) it does not require excessive support, or maintenance by Genomics England, or does not require significant re-configuration of the Research Environment;
 - d) a license is provided by the applicant that clearly supports its deployment into the Research Environment.
- 8.6. Import of data will only be considered if relevant to the research aims of the project and there is justification for its use.
- 8.7. Genomics England may additionally charge for imported data to be held or analyzed within the Research Environment.
- 8.8. Where users of the Research Environment request import of data at the expense of Genomics England, the imported data may be made available to other researchers. Genomics England, at their discretion, may agree to facilitate a moratorium period whereby the use of the data is restricted to specific researchers.

Document Key: 428	Version: 2.0	Status: Live
Airlock Policy		Page 9 of 9
USER IS RESPONSIBLE FOR REFERRING TO THE LATEST VERSION OF THIS DOCUMENT UNCONTROLLED WHEN PRINTED		