Thursday 4th June 2026

A letter from the Chair of UK Biobank's Oversight Committee

UK Biobank's first responsibility is to keep our participants' data safe. The offer for sale of de-identified UK Biobank participant data in China, discovered in April 2026, was a serious breach of UK Biobank's Material Transfer Agreement with the research institutions involved (the Incident).

Only de-identified participant data are supplied to researchers. It contains no directly identifiable information – such as your name, address, email, mobile phone number or NHS number – nor any indirectly identifiable information[1]. For the avoidance of doubt, your personal identifying information is stored by UK Biobank in encrypted form. It is only available to a very restricted number of UK Biobank staff. It is never supplied to researchers.

This Incident caused the Board of UK Biobank to establish an Oversight Committee to investigate the Incident and to recommend to the Board any improvements that can be put in place to minimise the chance of any such future incidents or other forms of breach.

The Incident in April 2026 showed that UK Biobank had not moved quickly enough in the policing of researchers' use of participants' de-identified data, for which we sincerely apologise. While we celebrate the research breakthroughs that UK Biobank has, thanks to the generosity and trust of our participants, brought to the world since we began work in 2003, there is more to be done, which is set out in our report.

The review was conducted by the Oversight Committee, comprising the following members (and assisted and advised by an Independent Risk Expert):

- Bernard Taylor (Chair)
- Sir Michael Ferguson (chair of the Access Committee)
- Sir Anthony Finkelstein (external member of the Information Governance Committee and past Chief Scientific Adviser for National Security to the British Government)
- Nicola Perrin (chair of the Ethics Advisory Committee)
- Sir Adrian Smith (external member of the Information Governance Committee and past Chief Executive of the Alan Turing Institute and past President of the Royal Society)

Secretary to the Committee:

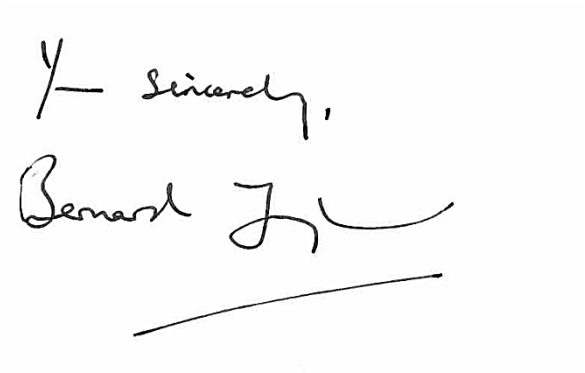
- Jonathan Sellors (General Counsel and Secretary)

[1] such as a rare occupation or free text fields in clinical notes.

Our report sets out the facts around what happened in the Incident, the immediate steps taken in response, and identifies our nine recommendations that UK Biobank will now implement. This report has been accepted and endorsed by the Board of UK Biobank.

I will continue to lead this Committee as it oversees the execution of a detailed plan of action by the leadership team of UK Biobank. We are determined that the organisation learns and fully takes on board the lessons from this Incident. Participants' continuing trust and confidence in UK Biobank is vital to our mission of supporting health-related research that makes a difference and must be maintained.

I hope this report answers your questions about what happened and reassures you of UK Biobank's commitment to the long-term security of all our participants' data.



Bernard Taylor CVO, CBE, DL
Chair of the Oversight Committee, Chair UK Biobank Audit and Risk Committee and of UK Biobank Information Governance Committee

Executive Summary

The Committee undertook to produce this report as quickly as practically possible after it was discovered that de-identified participant data had been removed from UK Biobank's cloud-based Research Analysis Platform (UKB-RAP) and listed for sale on a consumer website in breach of the Material Transfer Agreement with the researchers and their institutions ("Incident"). The Committee believes this report represents an appropriate, forensic and thorough investigation of what we know about the Incident. The Committee has made nine explicit recommendations related not just to this particular Incident but also to the security of the participants' data more broadly – and it strongly recommends that these are implemented in full.

The Committee recognises that UK Biobank's first responsibility is to keep the participants' data safe. It would also like to re-iterate the UK Biobank apology in this report. The Committee truly regrets that this breach occurred and for the understandable concern that this has caused. The Committee also recognises that breaches like this impact participant trust, not just for UK Biobank but for other major health research projects seeking to make scientific advances in the public interest. The Committee believes that the recommendations, which are set out below, represent the steps that UK Biobank needs to take to address these issues and rebuild any loss of trust.

What happened?

The initial notification that de-identified UK Biobank participant-level data were being offered for sale on the Xianyu platform came from an unnamed researcher - presumed to be based in China and who requested anonymity - via email. In reviewing all listings on the Xianyu consumer marketplace, owned by Alibaba Group, UK Biobank identified two further listings offering access to UK Biobank data. It is believed no data were sold from any of these listings.

- Instance 1 related to an approved project being undertaken by the Second Xiangya Hospital in China. The Principal Investigator (the researcher in charge) has confirmed that a postgraduate researcher, not registered on the project, had used the credentials of a colleague – who was a registered researcher on the project – and had inappropriately exported data before offering this for sale via the Xianyu listing.

- The investigation into Instance 2 involves both Beijing Chao-Yang Hospital and the China-Japan Union Hospital (and other collaborators). The parties

were approved in the project application and both Beijing Chao-Yang Hospital and the China-Japan Union Hospital (the primary institutions) have been banned from further access.

- Instance 3 was linked back to an approved application at Tongji Hospital in China. The Principal Investigator (the researcher in charge) of this application stated that a temporary employee of their institution – not a registered researcher on the project – was confirmed to have gained unauthorised access to the UK Biobank Research Analysis Platform (UKB-RAP) via a registered researcher's account and exported data.

- In all three cases, UK Biobank believes the individuals involved modified and then exported de-identified participant-level data (in this case tabular data) outside of the UKB- RAP to their own environment. The data were then made available via a Xianyu listing. This is in breach of UK Biobank policy, and a serious breach of the Material Transfer Agreement between UK Biobank and the research institution concerned. The investigators and institutions involved have been banned from any continued access to the UK Biobank resource.

- With support from both the UK and Chinese governments, Alibaba removed these listings immediately. Alibaba has also put in place automated searches to remove any further listings related to UK Biobank. We are grateful to Alibaba for their help in this matter.

- UK Biobank has written to every participant, either by email or post, drawing their attention to the Incident and so they can make an informed decision on remaining in the study.

Further information, including details about reporting to participants, the Board, funders, UK Government and regulators, and an assessment from the committee and an independent risk expert are provided in section 2 of the report. This includes discussion about the prevailing uncertainties about the Incident, and the limited opportunity for independent verification.

Taking the overall circumstances into account and given the pseudonymised nature of the dataset and the lack of confirmed sales, the risk assessment based on the investigation so far is that the Incident has been contained rapidly. At this time, the number of participant withdrawal requests received remains low. Notwithstanding this assessment, UK Biobank recognises the significant concern caused to some participants and the possible loss of trust in UK Biobank, which this report seeks to address with contrition and humility.

How the datasets were exported: risks and exposures for UK Biobank

UK Biobank opened for researcher access in 2012. The access model then was based on allowing approved researchers to download de-identified data (see Annex 3 below for a more detailed description of what de-identification involves) for analysis within their own environment, which was the prevailing means of providing research data to researchers at the time. The researchers signed a Material Transfer Agreement, which set terms of use and required that they delete the de-identified participant data at the end of their research project.

As the size of the UK Biobank research data resource grew and technology developed, UK Biobank procured a platform to allow researchers to use de-identified data in a secure and controlled environment. The UKB-RAP was implemented in 2020, allowing access to large-scale whole genome sequencing data through the platform. Other de-identified data were still available to download as before.

In mid-2024, UK Biobank implemented a 'Platform-by-default' policy. With a few limited exceptions, all approved researchers could only access UK Biobank data via the UKB-RAP. Policy and technical controls were in place to restrict researchers downloading de-identified participant data. However, researchers could continue to download their research analyses and results data from the UKB-RAP. There is not currently an 'output check' or airlock, to screen all outputs and check to ensure that a researcher has not extracted any participant-level data in the outputs.

Further, as a result of the Incident, the UKB-RAP is closed. It will not be re-opened until an Output Checking System (OCS) has been put in place. This is likely to be a manual system at first.

Due to the scale of use of UK Biobank, manual checks will be very resource intensive. Consequently, UK Biobank was already developing an automated OCS to provide these additional download restrictions to follow on from the manual system. This work was already underway before the Incident and because, for the avoidance of doubt, it had already been agreed by UK Biobank, NHS England and the Government in 2025 that GP data would not be available through the UKB-RAP until such checks were in place. The development and implementation of the automated OCS is in the final stages of public procurement, with implementation scheduled for early 2027. Once an OCS (manual or automatic) has been implemented, the occurrence of the Incident should no longer be possible.

Further details are set out in sections 2.3, 3 and Annex 2, which provide important context. The Committee has also considered the ongoing risks associated with de-identified data that were downloaded before 2024, which are considered in detail

in section 3. For any research resource, there is a balance to be struck between enabling access to data (where a certain level of risk is inevitably involved) and restricting access (such that the risk is mitigated but so is the research utility). However, the Committee does have a clear view that UK Biobank's approach should be minimal tolerance for risk in relation to data security, and the Committee's recommendations are prepared with that in mind.

Recommendation 1 – Internal reporting and governance

The governance and reporting of such incidents within UK Biobank have been reviewed. The internal protocols themselves have worked reasonably well but need to go further and faster.

Action: *A standing committee of the Board has been established so that appropriate risks/incidents can be brought to their attention as soon as possible and within 24 hours of an incident first being recognised and then prioritised and escalated to the Board as necessary.*

Recommendation 2 – Communication to participants

It has taken too long to contact the participants.

Action: *UK Biobank will seek to obtain the contact details for all its participants and procure an email service and postal service which can email/hard-copy mail all participants on the same day and be on permanent standby.*

Recommendation 3 – Security review of UK Biobank's data and systems

Although this was not a factor in the Incident, a number of healthcare organisations have experienced external hacking and other cyber threats from State, quasi-State and non-State actors. The security of UK Biobank's data storage, use and infrastructure needs to be reviewed and, if necessary, further strengthened (and demonstrated to be properly robust).

Action: *An external security review of UK Biobank's systems and data management will be commissioned immediately.*

Recommendation 4 – Review of the access procedures and oversight of the access process for researchers to use participant de-identified data

To evaluate the access policy and procedures. Establish where there is reliance on functionality, reporting, policy or legal compliance. Review the legal framework in which UK Biobank operates and the oversight of applications. Review the use of more extensive sanctions on researchers.

Action: *Commence forthwith an end-to-end review, with such external assistance as may be required, of the access procedures and related controls, including the effectiveness of any monitoring controls.*

Recommendation 5 – Establishment of internal proactive cyber and data security capability within UK Biobank

To continually assess and review access, security protocols, and the effectiveness of monitoring in order to identify and mitigate examples of researcher or third-party misuse or revealing the existence of UK Biobank data in public-facing on-line environments. This will need significant input and advice from the external security consultants to establish a state-of-the-art capability.

> **Action**: *Set up a robust proactive capability, which will include the establishment of a dedicated security team within UK Biobank.*

Recommendation 6 – Protocols for dealing with already downloaded data (before and after the introduction of the UKB-RAP by default policy)

It is necessary to remove – as far as possible – downloaded datasets in the possession of researchers. This will require confirmation, and audit thereof (which is allowed by the terms of the Material Transfer Agreement), of the deletion of the downloaded data from completed or terminated projects. The latter will involve incentives to move all research projects onto the UKB-RAP.

> **Action:** *Operate protocols for removing these data rapidly with a risk-based audit of the results.*

Recommendation 7 – Preventing downloads of data from the UKB-RAP

This requires the prevention of any future downloads of participant-level de-identified data from the UKB-RAP. The Department of Health and Social Care (DHSC) are soon due to set out guidelines for a Secure Data Environment (SDE) and the UKB-RAP will need to meet these requirements when they become available.

> **Action**: *Immediate evaluation and implementation of the functionality needed for the planned manual data airlock, in conjunction with a review of the specification for the automated data airlock (which should also include, as soon as available, controls over data ingress ability). Ensure the UKB-RAP meets SDE relevant requirements when these are known.*

Recommendation 8 – Evaluation of the re-identification risk

Although this topic is widely covered in scientific journals, there is a lack of hard evidence and probabilistic rigour, and a wide variety of measures used for assessing and measuring risk.

> **Action**: *This issue impacts on all research projects involving consented participant and/or patient data and it is proposed to set up a collaborative review – in conjunction with other research resources – to commission external research into the risk of re-identification and the measures that can be deployed to reduce the risk (such as generalisation, differential privacy,*

randomisation and use of homomorphic encryption). This will need to take into account both current and prospective technology (particularly in the context of next generation AI models).

Recommendation 9 – Risk assessment review

This Incident should prompt the Board to review its appetite for risk, focusing in the short term on data exposure or loss.

Action: *The Audit and Risk Committee shall refresh the strategic risk description and examine what new controls are needed to give the Board the assurance that the risks are within (or will become within) the proscribed appetite reviewed and agreed by the Board.*

The Oversight Committee presented this report to the UK Biobank Board on Monday 1st June 2026, and the Board granted its endorsement and approval to proceed with the implementation by the Executive of the above recommendations overseen by and under the direct governance of the Oversight Committee. The Board and the Oversight Committee have agreed on the importance of publishing this report in order to demonstrate UK Biobank's ongoing commitment to openness and transparency.

Table of Contents

1. Introduction

This report was commissioned following a serious data incident where de-identified participant data, which had been made available to researchers at academic institutions in China, were listed for sale on a consumer website, Xianyu platform, owned by Alibaba Group. With support from both the UK and Chinese governments, Alibaba removed those listings. It is believed that no data, relating to these listing, were sold. However, the Oversight Committee (which has been set up to produce this report) is clear that this was a serious data incident that affected the security of participants' de-identified data and that this required a detailed review of UK Biobank's security measures and operational processes to ensure that this does not happen again.

UK Biobank's first responsibility is to keep the participants' identified data safe and then to ensure the responsible use of the de-identified data (see Annex 3 below for a more detailed description of what de-identification involves) UK Biobank allows approved researchers to access for approved projects. UK Biobank is truly sorry that this breach occurred involving de-identified data and for the understandable concern that this has caused. UK Biobank has emailed or written to every participant personally to share what happened, apologise, set out the immediate actions that have been taken to keep their data secure and used responsibly, and offer to answer any questions they may have.

UK Biobank's participants generously agreed to share their identified data and other information about themselves, and for it to be de-identified to enable thousands of discoveries that are already leading to improvements in the prevention and treatment of many different diseases. They have a right to expect that UK Biobank will do everything possible to protect both their identified and de-identified information. UK Biobank also recognises that breaches like this impact participant trust and participation rates, not just for UK Biobank but for other major health research projects seeking to make scientific advances in the public interest. UK Biobank also wants to apologise to the government and to the research community, and it commits to sharing lessons from this review.

The Oversight Committee was set up with the Terms of Reference, which are attached in Annex 1. As this report sets out, UK Biobank did not get everything right in the run up to or during this Incident. The Committee has also identified clear areas of future risk that UK Biobank needs to address in short order. Specifically, the Committee has made a number of accompanying recommendations relating to UK Biobank's security measures and operational processes that need to be initiated straight away under the continued oversight of the Committee. The Committee and UK Biobank are committed to engaging transparently and constructively with their stakeholders on this vital work, which is UK Biobank's top priority.

The Committee has concentrated on two priority items:

- The Incident itself and the reasons behind it, which are set out in section 2; and
- The risks and exposures for other potential incidents in the future, which are still directly relevant for UK Biobank. These are set out in section 3, along with the mitigations needed to be initiated immediately, so that any attendant risks/exposures can be corrected as soon as practically possible.

The Oversight Committee considers that the evaluation of these risks/exposures should begin immediately under the direction of the Committee. The Board of UK Biobank has strongly endorsed and approved this approach. The Committee recognises that, due to the unique scale of UK Biobank, these recommendations (which address distinct areas of activity) each require proper consideration, planning and implementation.

Related to the implementation of the recommendations, the Oversight Committee will also conduct an evaluation of UK Biobank's prevailing organisational culture and the breadth of necessary skills and experience of the Executive Team.

In preparing this report, the Committee has interviewed each member of the Executive Team at UK Biobank, as well as two participants identified by the Head of Participant Engagement, in order to inform its understanding of events and inform the resulting recommendations. In assessing the circumstances of the Incident, the Committee has also engaged the services of a leading independent risk expert, who has provided input and analysis of the Incident (summarised in section 2.6).

The main recommendations and actions are set out in section 3 of this report. This report is the first step in the process to ensure that lessons are learned from this Incident, and the necessary changes are implemented throughout the organisation to mitigate against this and any future incidents.

2. Review of the Incident

2.1 Timeline of the Incident

On Tuesday 14 April 2026, UK Biobank was notified via email by an unnamed researcher – presumed to be based in China, who requested anonymity – that UK Biobank de-identified participant-level data were being offered for sale on a Chinese website, the Xianyu platform (which is the consumer-to-consumer part of the Alibaba group). Between Tuesday 14 April and Friday 17 April 2026, UK Biobank investigated listings on the Xianyu website – which ran to over 160 in number from ~100 discrete sellers – to establish if they appeared to be offering UK Biobank data for sale. The Xianyu website is a consumer marketplace where sellers can offer goods or services. The website does not contain nor can be used to directly distribute data, and fulfilment is provided via other physical or online channels (with the use of Quark Cloud Drive, an online data storage service provided by Alibaba, being noted in some of the listings).

The vast majority of the identified listings were not offering to sell data, instead offering access to analysis support/consultancy services, or access to publicly accessible summary statistics data (typically made available in conjunction with a journal publication). In addition to references to UK Biobank, such listings also included references to the China Kadoorie Biobank, FinnGen, and All of Us. UK Biobank subsequently informed these other resources of the issue and also requested support from colleagues at the China Kadoorie Biobank in Oxford to use their Chinese language services. Three listings, in particular, did appear to offer de-identified UK Biobank data for sale (because UK Biobank only provides de-identified data for research use). The data associated with the three listings were confirmed as being UK Biobank data, as UK Biobank is able to cross reference participant-level data with UK Biobank's internal data systems.

In each case, when the UK Biobank Data Analyst team had confirmed the source of the data, UK Biobank wrote to each identified and linked institution to require the listing to be removed and to explain how the data had come to be listed. By Friday 17 April 2026, UK Biobank had written to three Chinese institutions (and has subsequently written to further institutions as part of its investigation). UK Biobank further wrote to the published contact email address listed within the take-down notification section of Alibaba including details of all the listings that had been observed.

UK Biobank engaged an external Chinese law firm to monitor the listing URLs and to issue take-down notices for any which remained. In order to help trace the source of any data being listed, UK Biobank instructed a third-party service already engaged by UK Biobank to routinely monitor the internet and dark web ("Third-Party Service Provider") to procure data from at least one of the listings to provide confirmation (and determine the extent to which) UK Biobank data were being offered through the Xianyu website.

UK Biobank has continued to investigate each of the instances related to the unauthorised distribution of UK Biobank data. These investigations are further expanded on in the sections below.

On Monday 20 April 2026, UK Biobank sought assistance from the UK government (Department of Science, Innovation and Technology (DSIT)) to engage with Alibaba to secure support for the rapid removal of these listings. As a result, Alibaba implemented keyword searching on their website to ensure that any listings related to UK Biobank are immediately removed. It is believed that no sale of any UK Biobank de-identified participant-level data, relating to the listings, has taken place. UK Biobank is very grateful to the UK Government and Alibaba for their prompt help in this matter.

At the request of UK government, UK Biobank immediately suspended all access to the UK Biobank Research Analysis Platform (UKB-RAP), a cloud infrastructure-based service hosted in the UK where researchers can undertake their analyses of UK Biobank data. The UKB-RAP will not be re-opened until appropriate (automatic or manual) technical controls are in place that prevent the download of de-identified participant-level data.

The key events that have happened following the Ministerial Statement to the House of Commons and following this suspension of the UKB-RAP are noted below:

- UK Biobank's Third-Party Service Provider has been able to obtain five datasets related to listings on the Xianyu website. In four of the five listings, the data being offered were publicly available summary statistics data, freely downloadable from US or European websites. There was one dataset which did appear to relate to the selling of UK Biobank pseudonymised participant-level data, and this is detailed further in section 2.3 under 'Instance 3'.
- The Incident response team, supporting the senior Executive Team meetings and Gold team (the strategic-level Incident management group responsible for overall coordination during a major incident or emergency – see section 2.2 below) have continued to meet regularly to progress the investigation and ensure participant and media-related issues are addressed. This has included support from the UK Biobank Legal team (and the Data Analyst team) working with its external legal counsel in China and collaborating with researchers at the China Kadoorie Biobank (and their colleagues) to obtain further information as needed and expand the scope of enquiry where required.
- UK Biobank continues to correspond with the Principal Investigators (PIs) and research institutes relating to where it is believed the data may have emanated. The current summary of these investigations is set out in section 2.3.

- There have been several exchanges with NHS England and other health outcome data linkage providers (such as SAIL in Wales and Public Health Scotland). UK Biobank have advised that its current focus is on supporting the Board-led review and will follow-up with the respective organisations in a timely way. UK Biobank has responded to the questions received from the Information Commissioners' Office ("ICO") relating to this Incident by their requested deadline. On 29 April 2026, UK Biobank received a separate enquiry from the ICO from the section that deals with criminal investigations – noting this is not an investigation into UK Biobank but rather an investigation into whether an offence has been committed by a third-party contrary to section 170 of the Data Protection Act 2018. UK Biobank's external counsel is supporting their enquiry.
- UK Biobank has now received backup archive logs from its Platform Provider and has commenced a review of this information. These are sizeable audit logs dating back to mid-2024 and the review will take some time and noting that certain limitations and constraints may affect the completeness of the audit logs provided by the Platform Provider.
- There have been additional calls and requests for follow-up information from DSIT/DHSC throughout this process, to which UK Biobank has responded promptly and transparently.

2.2 Reporting

Internal reporting

An information security Incident was raised within an hour of the email notification from the unnamed researcher, presumed to be based in China. Under existing UK Biobank policies and protocols, an Incident response team was established and met in the following hour to agree initial steps to investigate, including verifying the legitimacy of the email and the provided listing URL. Several actions were agreed, including contacting the seller named on the listing to obtain more information to inform enquiries, and to extend searches across the Xianyu website to identify whether there were other listings of concern. The investigations continued through the evening and into the morning of the next day. It should be noted that the account below relies on information that has been provided by the institutions involved in the Incident, which information UK Biobank cannot independently verify.

On Wednesday 15 April 2026 at around midday, once the listings had been reviewed and their veracity confirmed, the UK Biobank Communications team were made aware of the Incident, and subsequently the Legal team met with the General Counsel to provide an update on the progress being made into the investigation. At 3:40pm, the General Counsel alerted the UK Biobank Chief Executive to the Incident. He noted that there were several listings on the Xianyu website referencing UK Biobank, with several appearing to be trying to resell UK Biobank data. Further, there were references to the

availability of China Kadoorie Biobank data. It was noted that several listings contained published guides to UK Biobank, and that the Data Analyst team were investigating each of the listings as a matter of priority.

As a result of the work of the Information Governance Committee, UK Biobank maintains a Crisis Communications playbook to help manage incidents that may have a significant operational and reputational element. This playbook sets out the formation of a Gold team that includes members of the Executive Team, Sanctuary Counsel (a communications consultancy retained by UK Biobank) and relevant subject matter experts to urgently look into the issue and formulate a response to participants, researchers, staff, external stakeholders such as the government, and the media.

As the investigation was in its initial stages, with technical teams still gathering preliminary information, it was agreed that the first Gold meeting would be scheduled after the Incident response team had convened at 9am on the next day to review their overnight findings.

On Thursday 16 April 2026, the Gold meeting convened at 11:30am to ascertain whether this was a crisis and, if so, to determine the required operational and communications steps. The Gold team met twice on Friday 17 April 2026, and then again on Monday 20, Wednesday 22, Thursday 23 and Thursday 30 April 2026 (and later meetings the following week), with meetings of the Incident response team taking place in-between, using twice daily stand-ups and use of MS Teams to share relevant information and updates to the Gold team.

The aim of the initial Gold meeting was threefold:

- establish what was known and what was not known (regarding the whistleblower and the veracity of their claims, the source of the data, the type of data, and UK Biobank's legal/reporting requirements);
- assess the options to have the data taken down and prevent this happening again; and
- prepare communications to participants, researchers, and stakeholders (as well as a holding statement in the event the media picked up the story before participants and other key stakeholders were informed).

The reporting to the government is detailed below and, leading up to and directly after the Ministerial Statement, the focus of the Communications team became providing supporting material for the government, finalising a statement to participants, preparing answers to questions from participants, researchers and staff, and drafting briefing letters to key parliamentarians.

UK Biobank Board, including funders and Board Committees

On Monday 20 April 2026, the matter was discussed in a previously scheduled meeting of a UK Biobank Board Nominations Committee (with the Chairman, the Chair of the Access Committee, and the Chair of the Ethics Advisory Committee in attendance). The decision was taken to notify UK government (specifically, DSIT as the department responsible for this policy area) through Minister Vallance (detailed further below).

On Tuesday 21 April 2026, around 12 noon, the full UK Biobank Board was informed of the data Incident (including for the first time, the Chairman of the Information Governance Committee). The membership of UK Biobank's Board includes representatives from UK Biobank's key funders: Wellcome, Medical Research Council, Cancer Research UK, British Heart Foundation and the National Institute for Health and Care Research.

Further updates were provided by email on Wednesday 22 and Thursday 23 April 2026. An extraordinary meeting of the Board was held virtually at 3:30pm on Friday 24 April following the Ministerial announcement in the House of Commons on Thursday 23 April 2026 to review the situation and set up a Board-led review into the matter. On Monday 27 April 2026, the Board Committees (including external members) were informed of the data Incident.

Reporting to UK government

As previously noted, on Monday 20 April 2026, the matter was discussed in a UK Biobank Board Nominations Committee. The decision was taken to notify UK government (specifically, DSIT) through Minister Vallance. A briefing note was sent to the Minister in the afternoon of Monday 20 April 2026, including a list of URLs relating to Xianyu website listings.

Late in the evening of Monday 20 April 2026, the UK Biobank Chief Executive received an email from the DHSC with questions relating to the Incident and a request for a call to be arranged, with responses to questions requested by Tuesday 21 April 2026. UK Biobank immediately commenced a more thorough review of the listings to help answer these questions.

By the morning of Tuesday 21 April, it seemed that all the listings that had been sent to Minister Vallance had been removed, although some new ones had appeared (which were also removed). Around 1pm, UK Biobank received email communication from the Office of Life Sciences (OLS, a part of DSIT). It requested a single point of contact to direct communications, and for an urgent call to be arranged with policy officials at DSIT and DHSC to obtain the requested information needed to brief Ministers. The call was arranged for 4pm with attendance by the UK Biobank Chief Executive, Deputy Chief Executive and General Counsel, and UK Biobank provided follow-up to additional information and actions later in the evening.

On Wednesday 22 April, a further call was arranged with representation from DSIT and DHSC, ahead of a meeting between Ministers Vallance, Murray and Ahmed with the UK Biobank Board Chair and UK Biobank Chief Executive later that afternoon. The purpose of the meeting was for UK government to set out its intended response and where it was communicated that a statement would be made to the House of Commons the next day. A further call was held with DSIT and DHSC later that evening, with additional email exchanges to provide information to inform the Ministerial statement.

It was suggested by UK Biobank and agreed that the following operational steps would be taken:

- UK Biobank would make no public announcements or pre-brief any media prior to the statement;
- Participants would be informed contemporaneously with (and not before) the Ministerial statement via an announcement on UK Biobank's website;
- UK Biobank would inform all UK Biobank participants individually as soon as possible (by a combination of email and post);
- UK Biobank would shut down access to its Research Analysis Platform to all researchers;
- Researchers would be informed that the platform is suspended, and UK Biobank would bear the continuing cost of data storage – and compute for analyses that were terminated mid-flight;
- The platform would only re-open when appropriate download controls and safeguards could be introduced; and
- UK Biobank's Board would initiate a rapid review by the Oversight Committee.

Regulatory reporting

On Friday 17 April 2026, UK Biobank filed a notification with the ICO within 72 hours of the Incident. A request for further information was subsequently received on Tuesday 21 April 2026 and UK Biobank has responded to the ICO's request by the deadline of Tuesday 5 May 2026. The UK Biobank General Counsel (the Data Protection Officer) has also spoken directly to the ICO at their request.

On Friday 24 April 2026, a serious incident report was filed with the Charity Commission using their online reporting form. Acknowledgement of receipt of the report was received the same day.

On Tuesday 28 April 2026, the Incident was reported to the North West Haydock Research Ethics Committee (REC), part of the Health Research Authority, and responsible for oversight of UK Biobank's REC approval.

On Wednesday 29 April 2026, the Incident was also reported to the Human Tissue Authority (HTA), whilst emphasising that this related to a specific data incident and did not impact upon the HTA codes of practice for the biological samples held by UK Biobank.

Further, on 29 April 2026, UK Biobank received a separate enquiry from the ICO from the section that deals with criminal investigations – noting that this is not an investigation into UK Biobank, but rather an investigation into whether an offence has been committed by a third-party contrary to section 170 of the Data Protection Act 1998. UK Biobank's external counsel is supporting their enquiry.

UK Biobank has also informed its providers of linked health record data: NHS England, Public Health Scotland and the Welsh authorities (noting there are no participants from Northern Ireland).

2.3 Assessment of the Incident

Following the initial notification via email by an unnamed, but presumed to be based in China, researcher that (de-identified) UK Biobank participant-level data were being offered for sale on the Xianyu platform, UK Biobank's Incident response team has been investigating what has happened and how it happened. This section should be read in conjunction with Annex 2 (UK Biobank Access).

Annex 2 provides further background to UK Biobank and the basis of its access policies and procedures that have been in place since 2012 and that prescribe how researchers can apply to use the resource in support of health-related research in the public interest. It describes how the access model was originally based on allowing approved researchers to download de-identified data for analysis within their own environment under a Material Transfer Agreement (MTA), and how these access policies and procedures aligned to the Five Safes framework – safe data, safe projects, safe people, safe settings and safe outputs.

Annex 2 also describes the early work in 2020 to implement the UKB-RAP in support of allowing researcher access to large-scale whole genome sequencing data, and the continuing platform development needed to support the move to a 'Platform-by-default' access model in August 2024. It sets out the technical and policy controls implemented in support of the revised access model, and the further technical development underway to fully implement technical download restrictions based on an 'airlock' that will use automation to interrogate data files that researchers want to take out of the platform to ensure that legitimate research results files do not contain participant-level data.

The circumstances surrounding the original listing brought to UK Biobank's attention are described below (Instance 1). This section outlines the actions UK Biobank has taken to establish the source of the data, identify who created the listing, and prevent any further UK Biobank data from being listed by this individual. Based on additional searches of the Xianyu website, two further listings were identified where it seemed UK Biobank data were being offered for sale, detailed below (Instance 2 and Instance 3). Similarly, details are provided on what is known about where the data came from, who listed the data and steps that have been taken, and highlights current gaps in understanding whilst the investigation continues.

The Committee wishes to highlight a number of prevailing uncertainties about the Incident, namely:

- The description below is heavily reliant on assurances/accounts from the institutions and the Principal Investigators (PIs), with limited opportunity for independent verification on behalf of UK Biobank;
- There is no understanding of the standing or motivation of the sellers of the data, and who stands to benefit from this Incident; and
- UK Biobank's understanding of the Incident may well change over time if more information comes to light.

Instance 1 – Relating to the listing brought to UK Biobank's attention by an unnamed researcher, presumed to be based in China

The listing described access to tabular record data for all 500,000 UK Biobank participants, including coded health record data (such as ICD diagnosis codes).
UK Biobank sought to verify the legitimacy of the listing by engaging with the seller as a potential buyer, based on seeking further information before making a purchase. Through these interactions, the seller provided screenshots that contained a subset of UK Biobank data alongside a limited set of data fields. UK Biobank Data Analyst team were able to determine that the data being offered for sale related to an approved project being undertaken by the Second Xiangya Hospital in China.

UK Biobank wrote to the PI to request the listing be immediately removed and to explain how the data came to be listed. In correspondence, the PI confirmed that a postgraduate researcher – name supplied and not a registered researcher on the project – had admitted to accessing the UKB-RAP. The individual had used the credentials of a colleague – who was a registered researcher on the project – and had inappropriately exported data before offering this for sale via the Xianyu listing.

The authorised researcher confirmed that the individual had been required to delete all UK Biobank related data. The listing was immediately removed and, separately, Alibaba has put in place automated searches to remove any further listings related to UK Biobank. The Second Xiangya Hospital has been banned from any continued access to the UK Biobank resource. The PI on the identified application has been required to delete all UK Biobank data they hold (this project had been running for several years and had previously downloaded data to their environment) and such confirmation has now been provided.

UK Biobank will be instructing the institution that all approved applications must now be brought to a close, and confirmation will be required that all UK Biobank data have been deleted. The Second Xiangya Hospital is part of the Central South University (who are undertaking their own approved projects). Accordingly, UK Biobank will be seeking assurance from the University that all required actions are undertaken by the Second Xiangya Hospital. Note: this institute was one of the three named institutes by Minister Vallance in the House of Lords statement.

Instance 2 – Relating to a further listing on Xianyu identified by UK Biobank as part of the investigation

In reviewing all listings on the Xianyu consumer marketplace, UK Biobank identified this further listing as offering access to UK Biobank data. The listing described access to code for extracting protein measures and included a screenshot of tabular data with a subset of UK Biobank pseudonymised identifiers. The UK Biobank Data Analyst team were able to use this screenshot to trace the data back to an approved project being undertaken by the Beijing Chao-Yang Hospital, which was collaborating with the China-Japan Union Hospital on their research project.

UK Biobank wrote to the Beijing Chao-Yang Hospital, copying the China-Japan Union Hospital as a collaborator, and following correspondence, the PI at the Beijing Chao-Yang Hospital indicated that they had engaged with the seller to obtain details of the dataset (the listing has since been removed) to assist their own investigation. The PI provided a detailed response to UK Biobank, explaining that the dataset they reviewed was not the same data as had been provided to their approved project.

UK Biobank requested a copy of the dataset to progress its own investigation. Following analysis, the UK Biobank Data Analyst team were able to confirm that the dataset had been made available to a different approved application being undertaken by the Shanghai Ninth People's Hospital with collaborators at the Shanghai Jiao Tong University, Zhengzhou University, and the Longhua Hospital Shanghai University of Traditional Chinese Medicine. UK Biobank has written to

the Shanghai Ninth People's Hospital, copying each of the collaborating institutions, as part of its investigation.

Whilst it seems the dataset that Beijing Chao-Yang Hospital had reviewed did not come from their approved project, there is currently no clear explanation as to how a screenshot of pseudonymised identifiers (that have been confirmed as being provided to their project) appeared with the Xianyu listing.

The investigation into this instance involves both Beijing Chao-Yang Hospital and the China-Japan Union Hospital and they have been banned from further access. This may be revisited once the investigation concludes. Note: both of these institutes were named by Minister Vallance in the House of Lords statement.

Instance 3 – Relating to a further listing on Xianyu identified by UK Biobank as part of the investigation

This instance relates to a further listing identified by UK Biobank as part of its investigation as appearing to be offering UK Biobank data for sale. The narrative detail contained within the listing did not give sufficient information to allow UK Biobank to identify what data may be being offered, or where the data may have come from. Consequently, UK Biobank requested its Third-Party Service Provider to obtain an extract of the data for examination.

The extract of the dataset obtained contained a subset of tabular data including limited health outcome measures. The UK Biobank Data Analyst team were able to link this back to an approved application at Tongji Hospital in China.

Following correspondence, the PI of this application confirmed that a temporary employee of their institution – not a registered researcher on the project – was confirmed to have gained unauthorised access to the UKB-RAP via a registered researcher's account. They confirmed the individual had exported data and had offered this via the Xianyu listing (the listing has been removed). The PI confirmed that a full inventory of all locally stored data had been reviewed and that all storage devices used by the individual had been inspected and no other UK Biobank data had been found. This institution has been banned from further access to the UK Biobank resource and steps are being taken to bring all existing projects at this institution to a close.

As a result of the publicity arising from the data Incident, UK Biobank subsequently received an email on Saturday 25 April 2026 from a further unnamed researcher presumed to be based in China relating to a website appearing to offer certain analytical services. The website is not a consumer marketplace and rather offers educational services. The concerned researcher brought the matter to UK Biobank's attention due

to a concern the website may be involved in providing unauthorised access to UK Biobank data (Instance 4). This instance 4 below was referred to by Minister Vallance in his statement to the House of Lords.

Instance 4 – Relating to an email from a further unnamed researcher presumed to be based in China highlighting concerns

> UK Biobank received an anonymous report on Saturday 25 April 2026 alleging that MUMDATA, a China based medical data analytics and training platform, was involved in the unauthorised use of UK Biobank participant-level data, supported by chat records and a sample dataset.
>
> The informant provided a dataset that could not be conclusively linked to a specific UK Biobank project data, and investigations continue to determine whether MUMDATA's involvement extends beyond training materials to broader data handling or redistribution.
>
> Open source MUMDATA training materials have been found with reference to UK Biobank; analysis of a screenshot within this content was linked to a project being undertaken by Soochow University. UK Biobank has issued formal correspondence to this institution requiring immediate removal of UK Biobank data from MUMDATA materials, confirmation of secure data deletion, and an explanation of the relationship with MUMDATA and how the data were used.
>
> UK Biobank notified the ICO on Tuesday 28 April 2026 specifically in relation to MUMDATA.

How the datasets were exported from the UKB-RAP

As described in Annex 2, UK Biobank opened for researcher access in 2012. At that time, pseudonymised research data were made available to approved researchers by way of download (as this was the prevailing means of providing research data to researchers). The researchers used these data in their own environment in accordance with the terms of the Material Transfer Agreement (which, amongst other things, required that they delete the pseudonymised participant data at the end of their research project).

As the size of the UK Biobank research data grew, UK Biobank sought to procure a platform whereby researchers would use the pseudonymised data in a secure and controlled environment, and in 2020, following a public procurement exercise, and with funding from Wellcome as part of the Whole Genome Sequencing project, UK Biobank developed the UKB-RAP.

The UKB-RAP became available for researchers to use in 2021. In terms of the ability to download data from the UKB-RAP:

- larger datasets such as the whole genome sequence data (WGS) were not available for download;
- other datasets such as assessment centre data, health outcomes and web-based questionnaires could be downloaded from the UKB-RAP (in addition to being available for download directly from UK Biobank's existing download systems); and
- researchers could download their research from the UKB-RAP, which included their analyses of the data, including summary statistics and other results data, which scientific journals require for publication.

In mid-2024, UK Biobank made the decision to implement the 'Platform-by-default' policy. On 5 July 2024, this policy change was communicated to all researchers confirming that all new research projects would be required to work on the UKB-RAP and that further downloads of pseudonymised participant data were only available in very limited circumstances.

The technical controls that were in place to protect the WGS data were extended to all other UK Biobank provisioned pseudonymised research data such that:

- all UK Biobank data (WGS and now others) were only available via the UKB-RAP and that researchers may no longer download pseudonymised participant data (with certain controls in place to prevent this); researchers could continue to download their research from the UKB-RAP which included their analyses of the data, including summary statistics and other results data, which scientific journals require for publication; and
- researchers that had previously downloaded data to their own secure environments could continue to work with these data until the end of the three-year period of their existing approved project (when any downloaded participant pseudonymised data would need to be deleted).

In extending its technical controls, UK Biobank changed its underlying systems that control data permissions such that any new project workspace created on UKB-RAP would be dispensed with all UK Biobank data marked as non-downloadable (meaning these data files could not be taken out of the platform). UK Biobank engaged its Platform Provider to develop functionality to retrospectively change all previously deployed project workspaces (of which there were thousands), to systematically mark all UK Biobank data files as non-downloadable. The engineering work and download permission changes for these existing projects was completed during August 2024.

It remains a policy control – supplemented by training and education before access to the UKB-RAP is granted – that researchers are required to ensure that any UK Biobank provisioned participant-level research data are not modified in a way or designated as analysis or results data. However, there is not yet a technical solution which

automatically checks to ensure that a researcher has not breached UK Biobank policy to create participant-level outputs that are then classified as derived results data and thereby enabling download. UK Biobank is developing an automated airlock (an Output Checking System, OCS) to provide these additional download restrictions. The development and implementation of the OCS is now in the final stages of public procurement, with implementation scheduled for early 2027.

UK Biobank's risk assessment of the Incident and its impact

The datasets identified as being offered for sale included health data, but the data were pseudonymised with all direct identifiers removed. Each dataset contained only de-identified data, and the re-identification methods are held securely within UK Biobank internal systems and remain entirely separate to the datasets made available to researchers.

To date, UK Biobank is not aware of any incidence where a participant has been re-identified (<1 in 500,000 over a period of 14 years) by a researcher or other third-party without the participant's active cooperation (which was the example referred to in the related recent Guardian article). Re-identification is not impossible, because the combination of UK Biobank data with other publicly available data could (but not easily) enable a 'motivated actor' to re-identify a participant. The use of pseudonymisation de-identification techniques – with the direct and indirect identifiers removed (see Annex 3 for more detail) – significantly reduces both the likelihood of participants being identified and the severity of potential harm arising from potential disclosure, even though the data concerned are health-related data.

In the current cases, Alibaba swiftly and helpfully removed these listings. It is believed that no sales of data, relating to these listings, have in fact been made. In addition to swiftly removing these listings, UK Biobank has suspended all access to the UKB-RAP and the platform remains unavailable. This was reported in the Ministerial statement to the House of Commons, and the Incident has also been widely reported in the media.

Taking the overall circumstances into account, and given the pseudonymised nature of the dataset, the lack of confirmed sales, the unlikely chance of individuals being re-identified the risk assessment based on the investigation so far is that the Incident has been contained rapidly without further complications. UK Biobank has now written to every participant, either by email or conventional post, drawing their attention to the Incident and so they can make an informed decision on remaining in the study. At this time, the number of withdrawal requests received remains relatively low (and this is dealt with in more detail below). Notwithstanding this assessment, UK Biobank recognises the significant concern caused to some participants and the possible loss of trust in UK Biobank, which this report seeks to address with actions and contrition and humility.

Impact on UK Biobank and beyond

The Committee would note that:

- although there has been no loss of identified data, there has been unauthorised use of de-identified data which may cause considerable concern to participants. The erosion of trust with participants (and members of the public and stakeholders) is a serious issue;
- UK Biobank has suffered some reputational damage amongst its key stakeholders, including its funders, the government, and a consequential impact on the ability of researchers to carry out their vital work and on the general landscape for the conduct of clinical research in the UK and abroad; and
- UK Biobank has suffered an impact on internal morale and experienced substantive organisational stretch, as well as there being a direct and serious financial impact of the Incident.

2.4 External communications and reactions

Ministerial statement

The statement was made by Minister Murray at about midday on Thursday April 23 April. It was repeated in the House of Lords on Tuesday 28 April 2026 (normally statements in the Lords are repeated on the same day but sometimes scheduling does not allow this).

Press and social media coverage

Initially, the media responded factually to the statement made in the House of Commons, and this was led by a piece in The Times. However, some coverage contained incorrect information, with media outlets using phrases such as 'hacked', so the UK Biobank communications team and Sanctuary Counsel worked to have the articles corrected so as to reduce possible participant concerns.

Beyond the announcement posted to the UK Biobank website, UK Biobank agreed to key interviews with the UK Biobank Chief Executive and Chief Scientist across print and broadcast outlets to help ensure as many participants (and other stakeholders) heard from UK Biobank as rapidly as possible whilst direct communications could be arranged. This included BBC News, BBC Breakfast, BBC Radio 4 Today, Sky News, ITV News, FT, Observer, and The Times. The communications aimed to convey key messages: an apology to participants and acknowledgement of the seriousness of the situation; no personal identifiers were involved and the high unlikelihood of re-identification; swift removal; no evidence that data had been sold; paused access to the platform; working closely with the government.

Coverage declined after the first 48 hours, and further media requests were responded to with a holding statement to help reduce further media coverage, and to allow focus on arranging direct communications to participants. Additional stories arose following a second Ministerial statement in the House of Lords (The Guardian) and Polly Toynbee commentating (also in The Guardian). As of 4:30pm Friday 1 May 2026, the UK Biobank Communications team had logged more than 80 pieces of coverage, including national, broadcast and trade media.

During this period, the UK Biobank Communications team monitored social media, responding to posts from participants and directing them to the UK Biobank website statement and responding to the misinformation where appropriate. The majority of social media posts were focused on topics adjacent to UK Biobank, such as Palantir and identification cards.

Participant communication

On Thursday 23 April 2026, at about midday, Minister Murray made a statement in the House of Commons. It had been agreed that UK Biobank would contemporaneously (and no earlier) publish its own statement to the UK Biobank website to inform participants, and that UK Biobank would write to all participants as soon as possible to make sure they were aware of what had happened.

The majority of inbound enquiries and withdrawal requests arrived within the first 48 hours of the parliamentary announcement and subsequent media coverage. After a weekend dip on 25 and 26 April, a second wave built from 29 April because of the email campaign to participants informing them of the Incident. A third wave was anticipated as a result of the postal campaign, with letters sent to participants from 1 May. Table 2.1 sets out the communications received by date, noting the majority of participants have contacted UK Biobank via email rather than by telephone.

Date	Total inbound communications	Total inbound calls	Total inbound emails	Data Incident related inbound calls	Data Incident related inbound emails
Thu 23 Apr	322	94	228	38	172
Fri 24 Apr	283	124	159	46	138
Sat 25 Apr	33	0	33	0	23
Sun 26 Apr	11	0	11	0	11
Mon 27 Apr	151	95	56	22	47
Tue 28 Apr	158	59	99	13	39
Wed 29 Apr	237	79	158	20	106
Thu 30 Apr	272	70	202	31	130
Fri 1 May	213	85	128	28	97
Sat 2 May	35	0	35	0	25
Sun 3 May	32	0	32	0	15
Mon 4 May	30	0	30	0	15
Tue 5 May	127	96	31	30	21
Wed 6 May	99	65	34	26	21
Thu 7 May	100	76	24	37	10
Fri 8 May	105	78	27	44	6
Sat 9 May	22	0	22	0	5
Sun 10 May	8	0	8	0	1
Mon 11 May	230	157	73	93	20
Tue 12 May	176	127	49	89	16
Wed 13 May	136	98	38	58	12
Thu 14 May	107	82	25	45	6
Fri 15 May	96	64	32	21	10
Sat 16 May	20	0	20	0	2
Sun 17 May	18	0	18	0	3
Mon 18 May	123	87	36	31	1
Tue 19 May	71	56	15	26	5
Wed 20 May	54	44	10	18	1
Thu 21 May	56	42	14	14	2
Fri 22 May	68	46	22	17	9
TOTALS	**3393**	**1724**	**1669**	**747**	**969**

Table 2.1 – Participant communications received by the UK Biobank Participant Contact Centre

On Monday 27 April 2026, a meeting was held with UK Biobank's Participant Advisory Group to discuss the situation and its communications with participants, including creating additional content on the UK Biobank website that would be updated to support participants as they received their emails and letters over the next few days. This process began on Tuesday 28 April with an email campaign, followed by a postal campaign to those participants for whom UK Biobank does not hold an email address.

On Tuesday 28 April 2026, a webinar was held with participants who had contacted UK Biobank over the preceding days and who wanted more information to address their concerns. The webinar was hosted by the UK Biobank Head of Participant Engagement and participants were able to ask questions of the UK Biobank Chief Executive and Deputy Chief Executive. The webinar was recorded and it is intended to convene subsequent webinars – which would be available to all participants – following completion of the email and postal campaigns.

UK Biobank initiated the outbound communication campaign to ~425,000 contactable participants via email and post. This figure is less than the ~500,000 participants that joined UK Biobank because: about 57,000 participants have passed away since joining the study; about 10,000 participants have asked UK Biobank not to contact them anymore (but agree their samples and data can continue to be used for research); about 9,000 participants for whom UK Biobank no longer has valid contact information (despite continuing efforts to keep these updated); and about 2,000 participants have asked to fully withdraw from the study since baseline recruitment.

Two communication campaigns were developed following input from the Participant Advisory Group, and both included a personal statement from the UK Biobank Chief Executive, Professor Sir Rory Collins.

An email campaign to ~313,000 participants began on 28 April 2026. The email send volumes were staggered across the week (see Table 2.2) for two reasons: to protect email domain reputation with internet service providers (whilst UK Biobank's email infrastructure has the theoretical ability to send up to 340,000 messages per day, this has not been tested at scale and is against best practice in order to avoid emails being marked as 'spam' and not being received), and to allow the Participant Contact Centre to maintain a quality service and be able to respond to the anticipated volume of participant queries as a result of the campaign. Additionally, there are technical restrictions in the IT systems that prevent a campaign being defined and initiated on the same day and so, whilst the campaign was constructed following the Participant Advisory Group on Monday 27 April, the sending of emails could not start until the next day.

Date	Tue 28 Apr	Wed 29 Apr	Thu 30 Apr	Fri 1 May	Total
Emails sent	10,000	80,000	125,000	~98,000	~313,000

Table 2.2 – Email campaign: email send volumes Tuesday 28 April to Friday 1 May 2026

All undelivered emails – as a result of a 'bounce-backs' due to email addresses no longer in use – will be followed up with a postal communication to ensure all participants receive the information and, if possible, obtain new data to allow UK Biobank to update the email address records.

A postal campaign to the remaining ~112,000 participants for whom UK Biobank does not hold a valid email address on file was initiated on Friday 1 May 2026 (see Table 2.3).

Date	Fri 1 May	Tue 5 May	Wed 6 May	Thu 7 May	Total
Letters sent	13,000	20,000	30,000	~49,000	~112,000

Table 2.3 – Postal campaign: letter send volumes Friday 1 May to Thursday 7 May 2026

Participants are free to withdraw from UK Biobank at any time and without providing a reason. Withdrawal requests since Thursday 23 April are tracked against the three withdrawal types available to participants, ranging from no further contact to full removal of consent for future research and deletion of data and biological samples. The withdrawal rates have so far remained low in absolute terms relative to the contactable cohort (see Table 2.4). A proportion of participants who initially contacted UK Biobank to request withdrawal have chosen to remain following the chance to speak with the Participant Contact Centre and understanding more about the nature of the Incident.

Channel	Total received	Data Incident related		Withdrawal requests	Withdrawals confirmed
Email	1,669	969	58%	217	146
Telephone	1,724	747	43%	59	12
TOTALS	**3,393**	**1,716**	**51%**	**276**	**158**

Table 2.4 – Participant communications and the number of participant withdrawals (as at 22 May 2026)

The process the Participant Contact Centre follows for withdrawal requests is to acknowledge the request and, where it is unclear if the participant has received any UK Biobank communication or published statement on the Incident, participants are offered these materials to allow an informed decision. If a participant subsequently wishes to withdraw, they are asked to confirm the level of their withdrawal and UK Biobank will immediately commence the withdrawal process. In some cases, participants change their minds and agree to remain in the study because of receiving additional information. Where a participant is explicit about their wish to withdraw, or has already received communication from UK Biobank, their withdrawal is commenced immediately.

As of 22 May 2026, there have been 276 withdrawal requests, which are at different stages of the process (see Table 2.5):

- Participants where UK Biobank has replied and provided further information and is now awaiting further instruction from the participant (categorised as “Pending / undecided”);

- Participants who have explicitly stated their request to withdraw, including withdrawal level (categorised as "Withdrawals confirmed"); and
- Participants who have decided not to withdraw and to remain in the study after receiving additional information (categorised as "Reversed decision").

Stage	Participants	% of enquiries
Withdrawal requests	**276**	**100%**
Withdrawals confirmed	158	57%
Reversed decision after PCC contact	9	3%
Pending / undecided	109	40%

Table 2.5 – Number of participant withdrawal requests by stage

The majority of the 158 participants that have confirmed their withdrawal request have selected No Further Use (i.e. full withdrawal) rather than the partial alternatives, which is to be expected given the nature of the Incident. To put the current withdrawal figures into perspective, the previous negative reporting (in The Observer and The Guardian) relating to use of the UK Biobank resource by insurance companies and, separately, by researchers in China led to 31 and 166 confirmed withdrawals respectively.

It should be noted that the timing of communication to participants was the single most powerful driver of participant dissatisfaction. Participants who learned of the Incident through media coverage, before hearing directly from UK Biobank, experienced a 'trust shock' that characterised the entire initial response period. The subsequent direct communication from UK Biobank produced a measurable and positive shift in tone in the participant reaction.

Researcher communication

On 23 April 2026, all researcher access to the UK Biobank Research Analysis Platform was suspended and all in-flight analyses and computational work was terminated. UK Biobank's Platform Provider updated their operational status page to confirm the platform was suspended for all users until further notice, with guidance to visit the UK Biobank website for further information. The process for suspending the platform commenced at 10:45am to ensure the necessary operational steps to take the platform down would be completed immediately ahead of the Ministerial statement, which was anticipated to take place at 11:15am.

The immediate focus of the UK Biobank Communications team was to update the UK Biobank website to inform participants of the data Incident. At 12:15pm, an email was sent to 28,582 researchers to inform them that the platform had been suspended using text that had been agreed with DSIT and DHSC. The same message was posted to the UK Biobank Researcher Community and the UK Biobank Access Management System:

> "UK Biobank is suspending access to the Research Analysis Platform with immediate effect. This suspension will remain in place until a technical restriction on the download of data is implemented. For further information, please see here.
>
> We expect the Research Analysis Platform to be available again in the first part of May. UK Biobank regrets the sudden requirement to introduce this suspension."

This latter statement has proven to be unrealistic in terms of timeframe – as the UKB-RAP will likely be closed for longer – and will need to be updated. An amendment was subsequently made to the Researcher Community page and the Access Management System to clarify that researchers would not be charged for data storage costs that would accrue during this time, or for the compute jobs that were terminated mid-run. A further communication was sent to all researchers on Thursday 14 May to reaffirm this and to advise that further communications on the timetable for re-opening would be provided in early June.

In the period to 7 May 2026, UK Biobank received 97 ticket requests from researchers related to the data Incident, with a further 1182 ticket requests related to 'business-as-usual' queries. There has been little noticeable change in the usual number of weekly ticket requests, where the Researcher Engagement team typically receive 750 ticket requests per week. The majority of ticket requests relating to the data Incident were received on Thursday 23 April when the platform was suspended, with a further small spike in tickets requests on Monday 27 April.

The researcher correspondence reflects a predominantly negative yet professional sentiment. In a usual week, researchers predominantly contact UK Biobank regarding the following topics: application and registration queries (35%), MTA and payment queries (17%), data-related queries (15%), and platform-related queries (8%). Communications from researchers following the suspension of the platform has not materially changed the proportional distribution of the support ticket topics.

Regarding the tickets associated with the suspension of the platform, the number of ticket requests have remained small, with researchers reporting the obvious inability to log in to the platform but also sharing their concerns on the impact to their research. Users have been citing missed or threatened thesis deadlines, conference submissions, annual reports and raising concerns about paid project timelines and funded research milestones. These concerns have also been echoed in private communications to senior UK Biobank staff by researchers from both academia and industry who, while expressing an understanding of the situation, have raised concerns about the potential knock-on consequences of the platform suspension.

Qualitative analysis of ticket requests suggests that respondents understand and support the need for the suspension, but patience is becoming strained due to:

- Lack of clear timelines for restoration;
- Loss of data access mid-analysis, including terminated compute instances; and
- Perceived financial and productivity losses during paid access periods.

Despite a certain level of frustration, the tone remains respectful, collaborative, and compliant but with researchers seeking timely communication, predictable resolution pathways, and mitigation options to prevent further academic, financial, and reputational impact.

2.5 Commentary from the Oversight Committee

The Oversight Committee has appointed an independent risk, data and security professional ("Independent Risk Expert") to review the Incident.

The Chair of the Oversight Committee (Bernard Taylor) and Professor Sir Mike Ferguson – on behalf of the Oversight Committee – interviewed the Executive Team (Rory Collins, Gareth Gregory, Mark Conway, Naomi Allen, Ed Sykes, Mark Effingham and Jonathan Sellors) and two participants over the course of the day on Wednesday 6th May.

Overall, the Executive Team were open, helpful and dedicated to resolving the issues necessary to allow UK Biobank to re-open to researchers. They expressed deep regret for the concern the Incident has caused to the UK Biobank participants and also to the possible long-term outlook for clinical health research in the UK.

It is clear that the crisis management procedures (Gold and Silver teams) that were put in place (by the Information Governance Committee) worked well but fell short of escalating the issue fast enough to the Board. The Executive Team also recognised that it took too long to communicate with the participants following the Minister's announcement in the Commons. It is also clear the UK Biobank's workforce are deeply concerned.

UK Biobank is a world class research resource, but its focus must now turn to the complex data issues highlighted in this report and to the implementation of this report's recommendations in full.

2.6 Independent commentary from the Independent Risk Expert appointed by the Oversight Committee

The Independent Risk Expert attended the interviews with the members of the Executive Team and the two participants on Wednesday 6th May and their independent commentary on the Incident and on the response is as follows:

Detecting the Incident

UK Biobank were alerted to this set of Incidents by an anonymous tip-off. Had UK Biobank not received this information it may not have been aware of the data being sold via Xianyu. UK Biobank has in place a number of controls and activities to detect public exposure of de-identified participant data, but these did not alert the organisation to any of the listings mentioning UK Biobank. It is challenging to build a capability to deliver a high confidence of detection of exposure of such data, but it should consider reviewing what the organisation's current capability offers and whether it should be enhanced (in line with Board risk appetite for the exposure of de-identified data). There should also be clear expectations placed on core suppliers and partners to ensure that UK Biobank has the necessary controls to spot new and emerging threats to the UK Biobank participant de-identified data, as well as a defined operating model for how these controls and triggers work and what coverage and assurance they offer.

The Incident and the initial response

This was a fast moving, developing series of events. The circumstances are unusual and may be unprecedented (the attempted series of sales of UK de-identified participant health data via an online marketplace). There was and remains considerable uncertainty regarding the facts of the Incident and UK Biobank has been heavily reliant on information from third parties, some of which is difficult to verify. The organisation, as a result, found it challenging to establish the facts quickly and some early assessments (which supported briefings to seniors and to government/other stakeholders/participants) needed to be regularly revised. There may be merit in establishing a daily situation report, produced by trained assessment staff, to support decision makers during future incidents.

The initial response was fast and effective and drew on existing incident management plans. The UK Biobank teams worked well together and under significant pressure and worked diligently and innovatively to understand the facts and then assess, and manage, the risks. Having said that, the teams were stretched thin and some staff had to cover more than one role or substitute into vacant roles. Information flows within the organisation were slower and less efficient than decision makers needed (caused at least in part by dated infrastructure), and in some cases actions and progress was made possible only by individuals being flexible and working long hours.

Stakeholder engagement

There were a number of stakeholder relationships brought into focus by these events (particularly participants, researchers, government, UK Biobank staff). The organisation found it difficult to describe the impact of these events on different stakeholders, in part because of an evolving understanding externally of the risk and consequences of this de-identified data being made public. A lessons-learned exercise on stakeholder engagement would likely identify opportunities for how UK Biobank engages more effectively with participants to give them the information they need at the right time, and develop an approach to engaging with government such that incident managers are able to focus on responding to the incident, and seniors are equipped with the right information and assessments to provide government with what it needs.

Risk and governance model

The events, while unusual in nature, fall under UK Biobank's previously identified highest strategic risk (data being lost/made public) and as such there is clear evidence that the organisation has developed controls to reduce this risk. Further controls, such as downloading restrictions and monitoring, and the suspension of UKB-RAP access, have been introduced in response to the events. UK Biobank has acted quickly to reduce the risks identified by this Incident and has also sought to use this Incident as an opportunity to review and strengthen its overall risk model. The Incident has demonstrated that there is an opportunity for UK Biobank to refresh Board appetite for the risk relating to the exposure of de-identified and identified data, and to enhance existing controls.

Board members were informed of the Incident later than was required for them to fully support the executive in delivering a strategic response. The Executive Team and Board acted quickly to establish a standing committee to oversee the response to the Incident and this will now be in place for future incidents. Gold and Silver command and committees operated well.

3. Risks/exposures for UK Biobank & plans for mitigation

3.1 Assessment of relevant UK Biobank risks/exposures

This Incident has raised questions about not only the circumstances of the breach, and the lessons UK Biobank should learn from it, but also what other key risks UK Biobank needs to address forthwith to mitigate the risk of other incidents occurring. The Committee would like to re-iterate that keeping participants' data safe and retaining their trust is critical to UK Biobank's licence to operate.

For any research resource, there is a balance to be struck between enabling access to data (where a certain level of risk is inevitably involved) and restricting access (such that the risk is mitigated but so is the research utility). However, the Committee does have a clear view that UK Biobank's approach should be minimal tolerance for risk in relation to data security, and these recommendations are prepared with that in mind.

The Committee proposes that the overall approach to risk should be reviewed by UK Biobank's Audit and Risk Committee in light of the current Incident. The Committee would note that the current risk register does highlight the risk of an incident such as the one covered by this report. This assessment should take into account that the participant identifiers are never (and should never be) made available to researchers, but only the de-identified data are made available to researchers. The Committee's view is that UK Biobank should have a minimal risk appetite for the following:

- Loss, leak or corruption of any of the identifiers;
- Failure to ensure the deletion of pre-2024 downloaded de-identified data from projects where the MTA has expired;
- Failure to ensure the deletion of pre-2024 downloaded de-identified data from projects where the MTA is current (and which need to be transferred to the UKB-RAP);
- Failure to ensure the deletion of post-2024 downloaded de-identified data from projects which have been granted an exemption (and which need to be transferred to the UKB-RAP);
- Failure to put in place suitable download restrictions on the UKB-RAP.

The Committee recognises that, with respect to the itemised instances above, there will need to be a transition from the current risk exposure to the desired risk appetite.

This section focuses on the summary findings of the Committee about the Incident itself and how it was dealt with and then goes on to evaluate in more detail further risks and exposures and their respective mitigations.

In this section of the report, the Committee makes a series of key recommendations for implementation as soon as practically possible. The objective in the case of each recommendation is to:

- confirm the nature of any risk and the potential mitigation steps;
- develop an implementation plan to address each recommendation and mitigate the potential risk; and
- implement the plan.

The closure of the UKB-RAP for a period of time provides the necessary space to implement the plans prior to the re-opening of the UKB-RAP. The implementation of these plans will be overseen by and under the direct governance of the Oversight Committee.

The majority of these actions can be undertaken within UK Biobank, with assistance from such external experts as may be required.

3.2 The Incident itself

3.2.1 The Committee acknowledges that the facts of the Incident are complex and, in some cases, not fully clear (please refer to previous comment on the subject).

3.2.2 However, on reviewing the materials and two of the members of the Committee meeting with the senior Executive Team in person, the Committee has made two immediate findings relating to the Incident:

- although the internal investigation proceeded in line with the protocols on incident investigation and internal reporting, the Incident itself was not reported sufficiently promptly or effectively by the Executive to the Board; and
- the communication to participants, by a combination of email and post, has taken too long.

Recommendation 1 – Internal reporting and governance

The governance and reporting of such Incidents within UK Biobank have been reviewed. The internal protocols themselves have worked reasonably well but need to go further and faster.

Action: *A standing committee of the Board has been established so that appropriate risks/incidents can be brought to their attention as soon as possible and within 24 hours of an incident first being recognised and then prioritised and escalated to the Board as necessary.*

Recommendation 2 – Communication to participants

It has taken too long to contact the participants.

Action: *UK Biobank will seek to obtain the contact details for all its participants and procure an email service and postal service which can email/hard copy mail all participants on the same day and be on permanent standby.*

3.3 Security of the UK Biobank data

3.3.1 The security risks within UK Biobank cover both cyber security around UK Biobank's operation systems and the security of participant data.

3.3.2 The security risk with UK Biobank's IT infrastructure covers:

- risks from external hacking into the UK Biobank infrastructure (and the related systems provided by contracted partners); and
- other types of cyber attack, including phishing, ransomware and similar.

Recommendation 3 – Security review of UK Biobank's data and systems

Although this was not a factor in the Incident, a number of healthcare organisations have experienced external hacking and other cyber threats from State, quasi-State and non-State actors. The security of UK Biobank's data storage, use and infrastructure needs to be reviewed and, if necessary, further strengthened (and demonstrated to be properly robust).

Action: *An external security review of UK Biobank's systems and data management will be commissioned immediately.*

3.4 Access to/use of UK Biobank de-identified data for research

3.4.1 **Out of scope access**: this could happen either by research institutions that should not have been approved in the first place or by individual researchers who should not have been registered. An overview of UK Biobank's access policy and procedure is set out in Annex 2.

3.4.2 **Unlawful, out of scope and out of contract use**: researchers are permitted to use UK Biobank data securely for the purposes of their research project.

- Unlawful use by researchers: researchers are not permitted to use the UK Biobank data for another purpose, nor are they permitted to sell, transfer or otherwise make the data available to anyone else. ***This is what happened in this particular Incident.***

- Out of scope use: In the access process, researchers describe their research project (which may be hypothesis testing or hypothesis generating) and this is reviewed by the UK Biobank Epidemiology team as part of the access process. A summary of all research projects is made available on the UK Biobank website along with the resultant publication. There are instances where:
    - Researchers use the data for another purpose which would still qualify as 'health-related research in the public interest';
    - Researchers use the data for another purpose which would not qualify as 'health-related research'; and/or
    - Researchers use the data for a subsidiary purpose, which may be legitimate 'health-related research' in their jurisdiction, but is disallowed under UK Biobank's Material Transfer Agreement.
- Out of contract use: There are instances where the deletion of UK Biobank data has not been confirmed by researchers, notwithstanding they have been reminded to do so and are indeed required to do so. This needs to be dealt with in a robust and auditable manner.

Recommendation 4 – Review of the access procedures and oversight of the access process for researchers to use participant de-identified data

To evaluate the access policy and procedures. Establish where there is reliance on functionality, reporting, policy or legal compliance. Review the legal framework in which UK Biobank operates and the oversight of applications. Review the use of more extensive sanctions on researchers.

Action: *Commence forthwith an end-to-end review, with such external assistance as may be required, of the access procedures and related controls, including the effectiveness of any monitoring controls.*

Recommendation 5 – Establishment of internal proactive cyber and data security capability within UK Biobank

To assess and review access, security protocols, and the effectiveness of monitoring in order to identify and mitigate examples of researcher or third-party misuse or revealing the existence of UK Biobank data in public-facing on-line environments. This will need significant input and advice from the external security consultants to establish a state-of-the-art capability.

Action: *Set up a robust proactive capability, which will include the establishment of a dedicated security team within UK Biobank.*

3.4.3 **Careless or negligent use by researchers**: this is where UK Biobank participant-level data is posted (inadvertently) to accessible websites, public code repositories or other platforms. The most common instance of this occurring is in relation to

public code repositories (GitHub being by far the largest) where researchers post their analysis code for their research, as commonly required by scientific journals, and the code then drags up underlying data. Once the data are on a public code repository they can then be copied by others (whether known to UK Biobank or not). The GitHub problem (and this also covers similar but less frequently used repositories, such as Software Heritage, Zenodo, Gitee and others) is actively pursued by UK Biobank.

Underlying this problem of careless use by researchers is the legacy issue that UK Biobank data have already been extensively downloaded – with the consent and knowledge of all relevant parties – by a significant number of research groups in the following ways:

- Before the instruction of the 'Platform-by-default' approach in August 2024 (and where all newly approved access projects would be required to use the UKB-RAP), there are a significant number of research institutions (~1,500) who were allowed to download the data within the scope of their MTA with UK Biobank;
- There are a number of institutions who have failed to confirm (~700) that they have deleted the data. These MTAs are now out of term, albeit the obligation to delete the data remains;
- There are number of institutions with valid, post UKB-RAP (~200) exemptions from the 2024 download restriction; and
- There are a number of institutions who may have downloaded the data from the UKB-RAP, notwithstanding the policy on no downloads of participant-level data.

The solution to the careless/negligent use issue is to ensure that the number of downloaded UK Biobank datasets which are in the possession of researchers (both in and out of contract) is fully minimised. If the researcher posts their code to GitHub (or similar) and they do not have any underlying data to drag up (because it is contained within the UKB-RAP), then the issue resolves itself.

3.4.4 **Legitimate use:** there are circumstances where:

- researchers have a pre-2024 MTA – before the introduction of the 2024 'Platform-by-default' policy – and they are entitled to retain the data until the end of their three-year term;
- there are also circumstances where researchers have – post the introduction of the 'Platform-by-default' policy – been granted an exemption and they are entitled to retain the data until the end of the exemption period (typically this has been granted to the end of 2026).

Recommendation 6 – Protocols for dealing with already downloaded data (before and after the introduction of the UKB-RAP by default policy)

It is necessary to remove – as far as possible – downloaded datasets in the possession of researchers. This will require confirmation, and audit thereof (which is allowed by the terms of the Material Transfer Agreement), of the deletion of the downloaded data from completed or terminated projects. The latter will involve incentives to move all research projects onto the UKB-RAP.

Action: *Operate protocols for removing these data rapidly with a risk-based audit of the results.*

3.4.5 **Preventing downloads of participant-level data from the UKB-RAP:** The functionality of the UKB-RAP needs to be upgraded to incorporate a manual or automated airlock so that whatever is egressed from the UKB-RAP does not incorporate any participant-level data. The former would be available sooner than the latter. There should be consideration given to ingress requirements.

Recommendation 7 – Preventing downloads of data from the UKB-RAP

This requires the prevention of any future downloads of participant-level de-identified data from the UKB-RAP. The Department of Health and Social Care (DHSC) are soon due to set out guidelines for a Secure Data Environment (SDE), and the UKB-RAP will need to meet these requirements, where relevant, when they become available.

Action: *Immediate evaluation and implementation of the functionality needed for the planned manual data airlock, in conjunction with a review of the specification for the automated data airlock (which should also include, as soon as available, controls over data ingress ability). Ensure the UKB-RAP meets SDE relevant requirements when these are known.*

3.5 Assessment of re-identification risks from UK Biobank data

This risk can be divided into the following categories:

- Albeit that no UK Biobank participant (without their cooperation) has to date been identified (as far as UK Biobank is aware) there needs to be a forensic statistical analysis of the risk, taking into account the latest AI models (such as Anthropic's Claude and the new Anthropic model Mythos), rare and unusual data points (such as rare disease, remote post code), participants' social media habits, and other public postings and generally available data;
- The identifiers themselves. As above, the security around these identifiers needs to be reviewed as part of the review of UK Biobank's IT systems; and

- AI models generally. Where AI models are allowed to train their weights on UK Biobank participant-level data, there is a reasonable prospect that these trained weights could be used – in conjunction with the AI model – to prompt the trained model to reproduce the underlying training data (namely UK Biobank participant-level data). This has been a matter of current focus but needs expedited review as part of the re-identification risk analysis.

Recommendation 8 – Evaluation of the re-identification risk

Although this topic is widely covered in scientific journals, there is a lack of hard evidence and probabilistic rigour, and a wide variety of measures used for assessing and measuring risk.

> **Action**: *This issue impacts on all research projects involving consented participant and/or patient data and it is proposed to set up a collaborative review – in conjunction with other research resources - to commission external research into the risk of re-identification and the measures that can be deployed to reduce the risk (such as generalisation, differential privacy, randomisation and use of homomorphic encryption). This will need to take into account both current and prospective technology (particularly in the context of next generation AI models).*

Recommendation 9 – Risk assessment review

This Incident should prompt the Board to review its appetite for risk, focusing in the short term on data exposure or loss.

> **Action:** *The Audit and Risk Committee shall refresh the strategic risk description and examine what new controls are needed to give the Board the assurance that the risks are within (or will become within) the proscribed appetite reviewed and agreed by the Board.*

Annexes

Annex 1 Terms of reference: UK Biobank Oversight Committee

Annex 2 UK Biobank Access

Annex 3 UK Biobank data de-identification protocols

Annex 1 - Terms of Reference: UK Biobank Oversight Committee

For the recent Incident involving UK Biobank data being offered for sale on the Chinese Xianyu platform.

1. Purpose

The Oversight Committee is established as a separate stand-alone committee of the UK Biobank Board which shall be convened for the purposes set out below:

- to review and evaluate the background and cause(s) – direct and indirect - of the recent Incident involving the (ostensible) sale of UK Biobank's participant-level de-identified data;
- to review and evaluate UK Biobank's actions and responses, in conjunction with those of other third-parties, to the Incident;
- to comment on the adequacy and fitness-for-purpose of UK Biobank's infrastructure, systems, culture and personnel;
- to make recommendations as to what improvements – in light of the review - can be made to minimise the chances of any such future incidents.

The output shall be a summary written report which will be presented to the Board. The Oversight Committee shall recognise the seriousness and urgency of the situation. The summary conclusions from the report should be for publication.

2. Membership

The Oversight Committee will have the following membership:

- Bernard Taylor (Chair)
- Sir Michael Ferguson (chair of the Access Committee)
- Sir Anthony Finkelstein (external member of the Information Governance Committee past Chief Scientific Adviser for National Security to the British Government)
- Nicola Perrin (chair of the Ethics Advisory Committee)
- Sir Adrian Smith (external member of the Information Governance Committee and past Chief Executive of the Alan Turing Institute and past President of the Royal Society)

In addition:

- Secretary of the Review: Jonathan Sellors (General Counsel and Secretary)

The Oversight Committee shall meet as often as required and the quorum will be the Chair and at least two of the other members.

3. Protocol

The Oversight Committee will have the following powers, including the right to:

- request any UK Biobank staff, executive or external advisers to attend a meeting, or provide evidence and information in writing;
- review any external advice and engage such external advisors and experts as they see fit;
- seek a viewpoint/opinion from such external parties as they see fit.

All Committee discussions, materials, and decisions are strictly confidential.

Annex 2 - UK Biobank Access

Background to UK Biobank

UK Biobank recruited its 500,000 participants between 2006 and 2010, with each participant accepting an invitation to attend an assessment centre where they spent time answering questions about their health and lifestyle, provided physical measures and biological samples, and gave consent to have their health followed through linkage to health records and for their data to be used by researchers around the world on a non-differential basis, whether the researchers be UK or international, academic or industry, for health-related research in the public interest. Further details and information on UK Biobank can be found on the website: https://www.ukbiobank.ac.uk

UK Biobank's Access Policy and Access Procedures

UK Biobank's Access Policy was established prior to the commencement of recruitment, and it is set out in the original 2007 governance document, the Ethics and Governance Framework (EGF). The EGF incorporates an aggregation of the information that was provided to participants both prior to recruitment and during recruitment (including the consent form itself).

The Access Procedures were developed between the end of recruitment and the opening of the resource for access. The standard three-year Material Transfer Agreement (MTA), which all research institutes are required to execute, is available here.

UK Biobank's approach to access

UK Biobank adopts the Five Safes framework (originally published by the UK Office for National Statistics in 2003 as the 'four safes' security model and evolved through the 2010s to the current framework as it stands today). This framework provides a set of principles to enable safe research access to data, ensuring data protection whilst balancing the demands for open science and transparency. In recent years, the framework has been adopted by the Secure Data Environment community and is based on the following 'five safes':

- Safe data: data provided to researchers are de-identified;
- Safe projects: access only by approved projects for health research in the public interest;
- Safe people: all researchers are vetted;
- Safe settings: data were provided under an MTA with strict requirements for data security; and
- Safe outputs: all researchers' publications are reviewed by UK Biobank.

In 2012, the software and compute infrastructure platforms that form the basis of Secure Data Environments in use today were not available. Hence, the working practices underpinning the access policies and procedures – vetted researchers, defined projects, de-identified data, contractual provisions for data use and security, and checking of researcher publications – were assessed to mitigate potential risks and to achieve the right balance between accessibility and data protection.

It should also be noted that UK Biobank has had interaction and discussions with DBT (previously BEIS), the Research Collaboration Advice Team (RCAT), the Office for Life Sciences and the National Protective Security Authority (NPSA) relating to certain categories of applicants.

Summary of key access developments

2010-2012: Implementation of access policy and procedures

- Developed in accordance with funder requirements.
- Board-led Access Committee to review project applications.
- "Hypothesis-testing" approach: specific datasets for pre-specified question.
- Standard download approach for providing data to researchers:
    - Bona fides of project PI and associated researchers checked.
    - Institution signs MTA (commitment to keep data secure and not share).
    - De-identified datasets provided by download to researchers.

  Note: the de-identification protocol is in accordance with current ICO guidance.

2016-2017: Change driven by generation of genotype data

- Access Committee agreed change to data access policy.
- "Hypothesis-generating" approach: non-specific datasets for exploratory analyses.
- Substantial increase in international access applications (US in particular).
- Standard download approach for providing de-identified data continued.

2020-2021: Further change driven by generation of genome sequence data

- Large-scale of sequence data made it impractical to download to researchers.
- UK Biobank Research Analysis Platform (UKB-RAP) provided by a US platform software company hosted on UK cloud infrastructure pursuant to a public procurement and a six-year contract.
- Focus on restricting access to the de-identified data for approved researchers (but not on preventing the continued downloading of data,

with the exception of whole genome sequence data, which has never been downloadable).

Note: To encourage UK Biobank use by Chinese researchers, a meeting was held in Beijing in 2019, attended by the Executive Chair of MRC and the Head of Wellcome Population Studies.

2020-2024: Extension of capabilities of UKB-RAP for wider range of researchers

- Such platforms typically focused on genetic analyses by specialised users.
- The Platform Provider extended capabilities for scale and range of UK Biobank users.
- Standard download approach for providing de-identified data continued unchanged.

  Note: Consequently, participant data had been downloaded over a period of 12 years to tens of thousands of researchers working on several thousand projects worldwide.

2024 Q2: Decision taken in May 2024 to transition to 'Platform-by-default' access model

- UKB-RAP considered to be sufficiently mature to support most types of analysis.
- Access to all data for projects approved from May 2024 onwards solely on UKB-RAP.
- Existing projects retain previously downloaded data, but no further data downloads (~3,000 such projects operating under a (up to) three-year contract that retain downloaded data; i.e. all will have ended by April 2027).
- Block on downloading de-identified participant data extended from sequence to all data (reinforced by policy, but not an airlock), but the system allowed analyses of data to be downloaded.

  Note: Large numbers of researchers continued to hold downloaded de-identified data, and each project could involve more than one institution, each with a separate MTA.

2024 Q3-4: Process for transition of researchers to doing analyses on UKB-RAP

- Negotiations with the Platform Provider to support much larger numbers of researchers and to develop the capacity of the UKB-RAP to support a much wider range of researchers.
- International imaging researchers objected to moving to Platform-only

approach due to lack of tools for image analysis, leading to development of exemption process.

- Access Team developed mandatory UKB-RAP training programme for researchers, which included MRC "Confidentiality and Data Protection in Health Research" course.
- Wellcome provided a grant to support additional Platform Provider costs, computing credits for researchers, and development of an automated output checking system.

2025-2026: Continued process of transitioning to 'Platform-by-default' data access model

- PIs of projects required to confirm deletion of previously downloaded data at end of three-year contract (with ~1,200 such projects in April 2026 still to complete by April 2027).
- Decision to extend current agreement with UK Biobank's Platform Provider for up to two years to focus on supporting researcher transition, and on developing automated output checking system (which is a requirement of NHSE for UK Biobank to receive primary care data).
- Legal advice obtained for this extension (noting procurement rules), and successful negotiations to extend the agreement with the Platform Provider, and with the cloud infrastructure provider continuing to donate data storage credits.
- Development of specification for automated output checking system:
    - March-May 2025: An external consultancy worked with UK Biobank to develop an outline technical specification for the system.
    - April 2025-January 2026: Input from Informatics Working Group (experts from Genomics England, All of Us, Our Future Health, Pharma/Tech industry, etc.).
    - May 2025-February 2026: Specification finalised for tender process.

2026: Tendering for development of automated output checking system

- Rationale for developing 'first-in-class' automated system:
    - Unprecedented large-scale use of UK Biobank renders implementation of manual airlock very labour intensive (requiring 80+ FTE staff).
    - More reliable and consistent approach to assessing files at scale, providing an accurate triaging system that balances security and efficiency appropriately.
- Procurement divided into two parts:

- Airlock framework: system into which files egress for checking.
- Decision engine: AI-enabled automated checking system.

- Airlock framework:
  - February 2026: procurement initiated.
  - May 2026: responses evaluated and contract to be awarded.
  - October/November 2026: development completed and framework implemented.
- Decision engine:
  - February 2026: procurement initiated.
  - May-June 2026: responses evaluated and contract to be awarded.
  - April-June 2027: development completed and decision engine integrated.

  Note: The UKB-RAP would then be ready to be assessed by NHSE for accreditation as a Secure Data Environment (SDE) by the end of 2027.

These timelines are visually depicted in figure A3.1 (an extract from UK Biobank Board papers, September 2025):

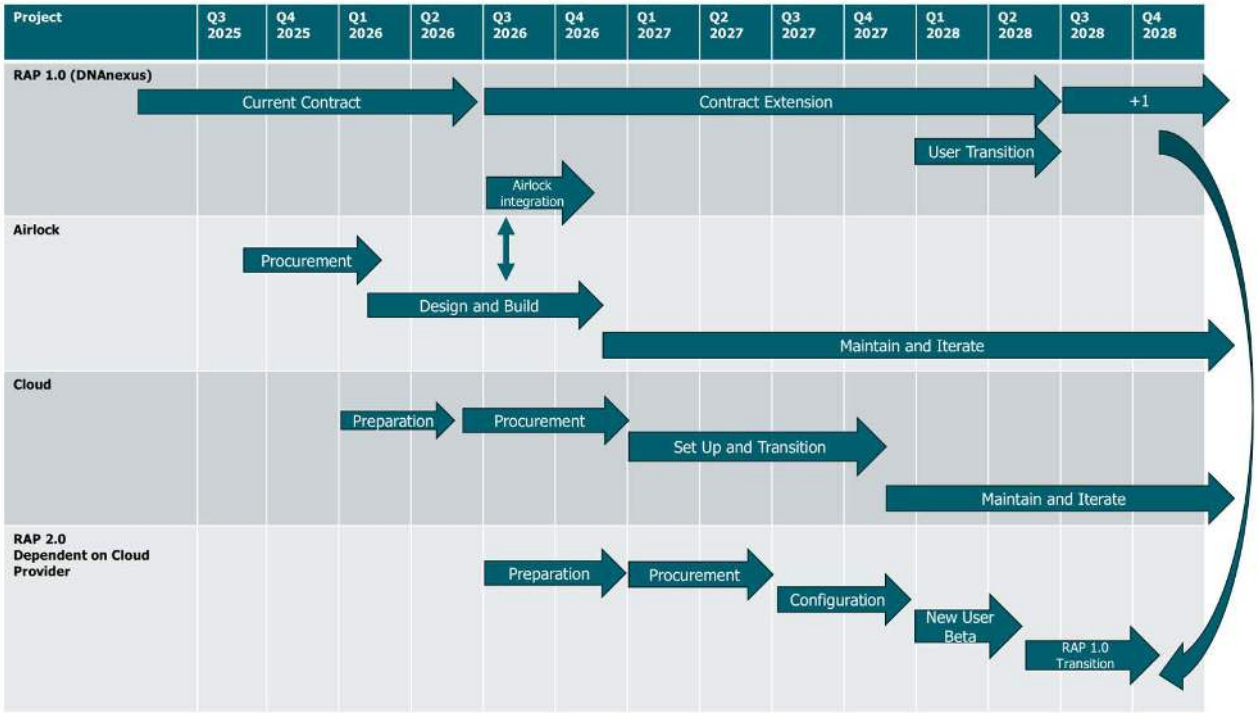


Figure A3.1 – Timeline for UKB-RAP extension, airlock development and future re-procurements

Additional background on the development of the UKB-RAP

In 2019, when UKRI/Wellcome and four industry parties funded the whole genome sequencing of all 500,000 participants, funding was provided to UK Biobank (£20m) by Wellcome to establish the UKB-RAP given the size of the sequence data that would be generated (which would be petabytes of data in size).

The funding was committed by Wellcome in 2019, and development of UKB-RAP began in 2020. UK Biobank appointed a US-based platform software company as its Platform Provider hosted on cloud infrastructure located in the UK.

The UKB-RAP was initially made available in late 2020 for use by the four industry parties generating the whole genome sequencing data. It was subsequently launched for use by all researchers in September 2021. At this time, the UKB-RAP worked alongside the existing download model – where researchers could continue to download de-identified data to support their project – except that the whole genome sequencing data could only be analysed within UKB-RAP.

In mid-2024, at which time the UKB-RAP was thought to be able to sustain most types of research (apart from some types of image and other large compute analyses), UK Biobank moved to the 'Platform-by-default' model with no further downloads (except in very limited exceptions) to any existing or any subsequently approved studies.

Technical controls that stopped whole genome sequencing data being directly downloadable were extended to all other UK Biobank data. However, the platform still allowed researchers to download data they had generated in order to allow them to egress their analysis results files.

UK Biobank also put in place training that all researchers must do before they were allowed to access the data on the UKB-RAP. Whilst reminding researchers of their obligations, the training served to help them become quickly effective in their use of UKB-RAP and to protect the platform given the rapid expansion by predominantly new users. This required UK Biobank to establish a number of specific modules and implement training of the tens of thousands of researchers involved in existing studies during the second half of 2024.

UK Biobank's Platform Provider does monitor platform usage and alert us to suspicious activity (including unusual downloads). However, this monitoring is after the event. As a result of these alerts, security improvements have been introduced such as the implementation of multi-factor authentication to further secure access to the platform.

UK Biobank recognised the need to develop an airlock that deploys a significant proportion of automated checking using machine learning or other artificial intelligence methods. The specification was developed in conjunction with an external consultancy and an Informatics Working Group (with representation from All of Us, Genomics England, NHSE, industry and others).

The automated airlock is under public procurement, which procurement is split into two parts: the first part is for the airlock framework (i.e. the system into which files requested

for data egress will be passed for interrogation and decision) and the second part is for the decision engine (i.e. the technology that provides the added intelligence to automatically interrogate data files to make an informed decision and reduce the need for manual human review). UK Biobank's estimate for the implementation of the automated airlock is Q2 2027. A manual airlock could be implemented faster.

UK Biobank's access checks

The access process

The access process is set out in detail on the main UK Biobank website and on the UK Biobank researcher community portal. This is a summary of the checks that are run by the Access Team.

The access test is that UK Biobank data are available for use by eligible bona fide researchers from academic, charity, government and commercial organisations from around the world, for health-related research that is in the public interest.

- A research institute needs to be approved and in order to do so it must:
  - Demonstrate a track record of legitimate health-related research; and
  - Operate from a country that complies with international regulations and is not subject to UK, US or EU sanctions.
- Individual researchers need to be affiliated with a recognised research organisation. The Access team at UK Biobank carry out background checks on any researcher who applies to access the data to ensure that they meet these eligibility requirements. All research institutes and individual researchers are run through the sanctions-checking software platform prior to approval/registration.
- Applications are reviewed by Ethox http://www.ethox.ox.ac.uk/ and Professor Mike Parker (the Director of Ethox) attends Access Committee meetings.
- Some applications are also referred by the Access Committee to the Ethics Advisory Committee and/or the Participant Advisory Group for their input and comment.
- Once approved, a research institute can make an application setting out:
  - A summary of the research they intend to conduct;
  - Confirmation that the research they are intending to undertake is health related and in the public interest;
  - The UK Biobank data tier they require;
  - A summary description of any new data or derived variables the research will generate; and
  - The findings of the research must be published and the results of the

research returned to UK Biobank.

- UK Biobank requires approved researchers to:
  - Provide annual project reports about their research;
  - Notify UK Biobank of any publications reporting on research connected to UK Biobank data;
  - Return results to UK Biobank so that they are available for other researchers to use; and
  - Keep relevant information such as contact details up to date.
- When an application is approved, the research institute must sign a legal MTA before any data can be accessed. The MTA is not negotiable and must be executed prior to data release.
- The research organisation does not need to apply for separate REC approval for their research: it is covered under the auspices of UK Biobank's Research Tissue Bank approval.

UK Biobank monitors compliance with researchers' contractual obligations through compulsory annual reporting and formal governance controls. Each institution named on an approved project is required to submit an annual compliance report via the Access Management System (AMS). This includes:

- confirmation of project status and progress;
- verification that individuals with access to UK Biobank data are correctly registered in AMS;
- declaration of any affiliated organisations (part of the same corporate group) or third-party data processors; and
- formal attestation that the institution is fully complying with the applicable Material Transfer Agreement(s), including all amendments, updates, and annexes.

These declarations provide a real-time mechanism for UK Biobank's ongoing oversight and assurance. Failure to comply leads to suspension of the institution's ability to access UK Biobank data.

As part of the annual report and the final project report (completed at the project end date), UK Biobank also:

- compares the publication (and patents) with the scope of the approved application to ensure that the researchers did do the research that they were approved for; and
- checks that the authors on the publication include the researchers who are formally listed on the approved application (it may be that there are authors on

the publication who had no contact with the UK Biobank data).

If, in either instance, that is not the case then UK Biobank contacts the institution for an explanation.

Summary of the phases and dealing with the legacy issues (towards 'Platform-only' model)

2012-2024: MTA with institutions allows researchers to download de-identified data and requires them to keep the data secure and delete them at the end of three-year projects.

- Access to UK Biobank data is provided under an MTA with the institutions of approved researchers, which requires them to keep the data secure, not to share them, and to delete them at the end of the approved three-year project.
- Between 2012 and mid-2024, researchers working on more than 4,000 separate projects had been provided with downloads of participant data that had been de-identified in accordance with the ICO's guidance for pseudonymisation.

2024: MTA supplemented by no further data downloads, and requirement not to take data off the UKB-RAP (with a block on removing raw sequence data extended to all data).

- By mid-2024, when the decision was made that all future access to participant data would only be on the UKB-RAP ('Platform-by-default'), ~3,000 ongoing projects still held downloaded data under their MTA.
- These ongoing projects were allowed to retain the previously downloaded data, but any further data were only available for them on the UKB-RAP (with limited exceptions).

2024-2027: Ongoing process of deletion of previously downloaded data as projects come to the end of three-year contracts, and introduction of automated 'air lock' in 2027.

- By April 2026, there were still ~1200 projects with downloaded data that had not yet completed, with the last of them scheduled to end by April 2027, and when projects are contractually required to delete any downloaded data that they hold.

Annex 3 - UK Biobank data de-identification protocol

UK Biobank takes steps to de-identify the data to preserve the anonymity of participants as far as practically possible. All data provided are pseudonymised in a process consistent with the ICO's current requirements and released to researchers in accordance with UK Biobank's published De-Identification Protocol.

No personally identifiable information, such as name, address, date of birth or NHS number, is shared with researchers. Each participant is assigned a unique Participant Identifier known as a PID, which is algorithmically derived to avoid collision within the same numbering space, and cannot be reverse engineered by third-parties. The PID is never shared with researchers or any other third-party. Each dataset released to a researcher cannot be linked back directly to the PID, which is securely stored by UK Biobank.

In addition, UK Biobank does not release certain data fields obtained from linked healthcare data that could increase re-identification risk, including admission of a participant to a particular hospital, dates that could expose exact date of birth, and unedited free-text fields. From the data that UK Biobank collect directly from participants, potentially identifying codes (such as rare occupations) are not released. Only brain MRI images that have been 'defaced' (i.e. blurred to an extent that they do not contain potentially identifiable features) are made available to researchers, by default.

Any data considered for release are scrutinised to consider the potential for re-identification, with granularity of the data reduced if re-identification is considered a risk. For example, location data is only made available on request and is provided at an aggregated level of 1km square grid reference or LSOA (equivalent to the first three digits of a postcode). UK Biobank does not release both variables for the same project owing to the risk of re-identification at the borders of these areas. If a participant lives in a sparsely populated area, location data is further aggregated to a 10km square grid reference. Requests for more granular data are considered on a case-by-case basis. If it is not possible to generate a bespoke variable for the researcher (for example, a binary flag of participants that self-reported having a managerial job at recruitment, or born before or after a change in education policy) then the data are released in a way that ensures they cannot be linked to the main dataset, via a separate project with unique EIDs. This allows researchers to derive the required data (for example, pollution measures derived from more granular location data). These derived data are also aggregated to remove any unique rows prior to incorporation into the main resource.

There are also certain data items which are inherently unique to a participant, such as genetic sequence data. However, the re-identification risk posed by this type of data is, in practice, relatively small. For example, a researcher in possession of sequence data (or a collection of single nucleotide polymorphisms (SNPs) or tandem repeats) would

have to possess a comparable genetic sequence from another source which also identified the participant.

Policy guidance

UK Biobank requires each research institution to enter into a Material Transfer Agreement (MTA) – a legal contract that sets out the basis on which the research institution (and approved researchers) can use the UK Biobank data. The MTA provides, among other things, that the research institution can only use UK Biobank data to conduct their approved research and that they shall not sub-licence, disclose, transfer, sell, gift or supply the data to any unauthorised third-party.

UK Biobank also take steps to reduce the risk of re-identification from published results (including peer-reviewed, pre-print or personal blogs/social media posts). Guidance is also provided to researchers to ensure that participants cannot be re-identified, including by the participants themselves, including:

- Collapsing categories to reduce the sparsity of the data;
- Aggregating data (e.g., over a greater period, or a larger geographical area);
- Rounding to a specific base to avoid very small numbers (no less than five – reported totals should use the rounded values to avoid the possibility of reverse-engineering the raw counts);
- Suppressing very small numbers (case and reported totals should be derived excluding the suppressed counts, to avoid the possibility of reverse-engineering);
- Web-based browsers that present summary statistics s (e.g., GWAS-PheWAS browsers) should contain a minimum number of 100 participants within a cell.

In addition, UK Biobank advises participants to be careful about what they share about themselves online and on social media, as well as taking part in other large scale genetic databases, to further reduce the risk of re-identification. There are pre-existing plans to undertake more detailed participant communication on staying safe online.