0. Proposal name (only use if filling the DMP outside of TFS)

The Airwave Health Monitoring Study of the British Police Forces (Airwave Study)

1. Description of the data

1.1 Type of study

Airwave is an observational study of current and former members of the British police service. It was setup to investigate effects of the Airwave radio system, if any, on users' health. We now address wider health issues within this interesting occupational group.

1.2 Types of data

Baseline and follow-up are available. See also https://police-health.org.uk/researchers/

CLINICAL: including, body composition, blood pressure, vascular health, and medical history. **Electrocardiogram**, with interpretation and trace images.

QUANTITIVE: Blood biochemistry and haematology, GWAS, methylation, NMR, mass-spectrometry, and proteomics. **Cognitive performance testing** (processing speed, episodic memory, working memory, attention, executive function). An extensive history of Airwave **Call Data Records** suitable for reconstructing participants' work-shift patterns.

SURVEYS: working arrangements, general health, work-stress, depression and anxiety, socioeconomic and demographic factors. **Diet** questionnaire (7-day).

MEDICAL RECORDS: cancers, all-cause mortality, demographics. **Absence from Work** supplied by employers.

1.3 Format and scale of the data

We collect Airwave datasets in whichever formats are best suited to the data collection protocol. Text and image-files are commonest, and we use industry-standard formats throughout. Typically, these are CSV, Excel, JPEG, TIFF (images) and PDFs.

We provide ongoing technical support for our datasets. We use open (not proprietary) formats to ensure longevity and we reissue datasets as standards change.

Complex datasets such as GWAS assays produce proprietary binary files. These are subject to data-reduction to yield tables tractable in commonly used statistical tools. We archive original files.

Our ORACLE-based **participant database** performs data-cleaning, linkage verification and transformation into common standards using self-developed applications. ORACLE ensures transaction safety, referential integrity, data preservation (backups) and confidentiality.

We manage participant identifiers separately from research data. They have no direct research value but are essential to maintaining a participant relationship. Research data reach analysts using only pseudonymous identifiers.

Quantity of Data: Our central database is ~50-GB. Multi-omic datasets are ~20-TB and includes some progressively enhanced versions of datasets.

2. Data collection / generation

2.1 Sources of data

(1) Follow-up via NHS registers. (2) Externally funded substudies

2.2 Data quality and standards

We are collaborating with partners at Dementia Platform UK (DPUK) to align the ontology (labelling, presentation, metadata) of our datasets with their standards.

2.3 Consent for data sharing and re-use

We have broad participant consent to retain data collected for as long as there is an infrastructure in place to support its safe use.

3. Data management, documentation and curation

3.1 Managing, storing and curating data

We store, process and curate using ORACLE on Windows, and LINUX servers for genetic datasets because of the software tools needed (PLINK). College ICT conducts server-level backups using enterprise-level software. We provide research extracts in industry-standard formats appropriate to the data's complexity and users' available statistical tools.

3.2 Metadata standards and data documentation

An overview document describing our datasets and the methods generating them is available here. We define each variable in Annexes, and work is ongoing.

3.3 Data preservation strategy and standards

Computing infrastructure has been in place since 2004, now based in our Trusted Research Environment (TRE). We archive original data and redundant versions of extracts supporting published research. The DPUK tenancy at UKSeRP holds current and past versions of pseudonymised research datasets. ISO/27001 is common to both infrastructures.

4. Data security and confidentiality

4.1 Formal information/data security standards

Our ISO27001 Certification Number is 16170 (awarded: 18th January 2023).

4.2 Main risks to data security and how they will be managed

We manage the risk of exfiltration of personal identifiers and research data into uncontrolled environments through rigorous procedures to protect our computing infrastructure. Users must first authenticate at institution level via enterprise class software checking usernames, password, and two-factor-authentication. We also enforce password complexity and frequency of change. Access to the TRE requires a second level of approval from senior research management and the use of specialist security software. Access to identifiable data requires approval for further database privileges.

The TRE separates its own network traffic from day-to-day university affairs and the Internet. Only operating system updates have any pathway through the TRE from outside. Privileged users may transfer aggregated research results out of the TRE. Automated processes preserve transferred datasets and email logs to the PI.

Formal training is mandatory for all research and support staff. Training ensures users understand the nature of their legal obligations under local policy and GDPR. We also verify the technical skills needed to prevent mistakes. Researchers' contracts forbid attempting to identify a study participant and require them instead to report an exception.

Disclosure Control Policy forbids describing any participant group comprising five or fewer persons in published research, thereby mitigating the risk of unintentional participant identification.

5. Data sharing and access

5.1 Suitability for sharing

We have collected in-depth lifestyle and phenome data on participants and have obtained substantial amounts of genomic and metabolomic data on a sizeable proportion of the cohort. Airwave can therefore address pressing health concerns such as diabetes, cardiovascular risk, metabolic disorders, cognitive decline, and the mental health of this important occupational group.

Participants granted broad consent for reuse of their data for research purposes. By communicating with the cohort via newsletters, surveys, PPIE groups and invitations to other research projects, we are confident of their ongoing support.

Our Disclosure Control Policy applies to research extracts during the data curation process, and we remove all personal identifiers. This means, for example, that we provide a merged rank for senior officers. Ages replace dates of birth; postcodes redacted to district.

177 applications to reuse Airwave data have been made. Researchers have published 90 journal papers so far.

5.2 Discovery by potential users of the research data

The Airwave asset appears on ten publicly accessible registries used by bona fide health research researchers.

At DPUK we maintain a copy of research data for third-party access. With 83 applications to date, Airwave is the 5th most frequently requested cohort.

We have been accepted into UK Longitudinal Linkage Collaboration (UK LLC), and in the term of this grant we shall link to NHS health data for the benefit of researchers applying there.

We have published the methods used to collect, clean, and export our research data. We provide extracts in the most user accessible formats, with version control and extensive metadata. We are collaborating with partners on a cross-cohort ontology.

We have used **persistent identifiers** for open-source, anonymised datasets (example) and will extend these during the next 5-years.

5.3 Governance of access

Our Access Committee (DAC) determines which research projects we support. It includes the PI, an epidemiologist, and a representative of Police Federation (lay member). DAC determines whether proposals are for the general good, have appropriate ethical approval, are worthwhile, achievable, and that any risks are manageable. Applicants start here**.**

5.4 The study team's exclusive use of the data

New data obtained by a research group and which they are actively analysing will be made available to the wider community after six months, or immediately if there is no active research requiring it.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

We respect obligation placed upon us in DSAs that disallow onward sharing. For example, our linked health records require an applicant to obtain their own approval from NHS for reuse.

5.6 Regulation of responsibilities of users

External researchers must be contractually bound to a third-party institution with whom we have an active Data Sharing and / or Data Transfer Agreement. This will require (i) a safe processing location (TRE); (2) prohibition from transferring record-level data out of the TRE without our permission; (3) Disclosure Control Policy; (4) users training; (5) proscribe participant identification; (6) A commitment to make available research results that may be useful to others; (7) Commitment to publish results.

5.7 Working with overseas collaborators or data users

International users access Airwave data on the TRE at DPUK. In principle, we may establish a DSA to hold Airwave data with an appropriate non-UK institution that operates within a trustworthy legal system which has adopted a version of the GDPR. No exemptions to data governance requirements exist for international users.

6. Responsibilities

The Airwave study's Database Manager established technical procedures for data management, metadata creation, quality assurance, and this DMP. A separate team jointly run by the School of Public Health and Imperial's central ICT team run the Airwave TRE and ensure security of the infrastructure.

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	**URL or Reference**
Data Management	Based on the ISO27001 and internally available on SharePoint.
Data Security Policy	Link
Data Sharing	Link
Institutional Information	Link
IG Framework	Link

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their **telephone & email contact details**

Andrew Heard, Database Manager. a.heard@imperial.ac.uk. 02075943842.