

ALSPAC Data Management Plan 2024- 2029

1. Description of the data

Type of study

ALSPAC is a multi-generation, pregnancy cohort following >14000 mothers recruited 1990-1992 and their partners (collectively G0), offspring (G1) and grandchildren (G2). ALSPAC adheres to FAIR Principles (www.go-fair.org/fair-principles/) to guide data access, management and use and this data management plan is in line with MRC and UKRI data sharing guidelines. This section is coordinated with the ethical and participant facing components described in sections 14. **Ethics and responsible research and innovation** and 18. **Research involving human participation**.

Types of data

Research data are obtained from online/paper-based questionnaires, in-person clinic-based assessments and biosample derived measurements. ALSPAC archives images (magnetic resonance, ultrasound, dual x-ray, computer tomography and specific retinal, face and body shape images) and other media, (video, audio and transcripts). Data are also obtained through linkage to administrative records including maternity, birth, child health, primary and secondary health care, education, and crime. Environmental data are connected through geographical information and digital footprint data via social media and transactional records.

Format and scale of the data

14541 mothers were originally enrolled and data have been collected from them and their families for >30 years. ~35000 participants have taken part at some point, providing >90000 variables (excluding omics, momentary/streaming and image data). Data are stored in multiple formats ranging from simple flat text files (GBs) through to complex formats such as array and image data (many TBs). Datasets are made available to researchers in all common formats.

2. Data collection / generation

Methodologies for data collection/generation

New data are generated from participant detail updates, questionnaires, clinic assessments, biosamples and record linkage. This is based on infrastructure and other research activity exploiting the ALSPAC platform.

Data quality and standards

All variables are documented following standard operating procedures. Online and clinic data are collected using REDCap with controls for logic and range checking. Paper based data are collected and digitised using character recognition software. Scanned data are verified by trained personnel. Clinic data are collected by trained fieldworkers according to ethically approved protocols and subject to regular audit. We encourage use of externally validated assessment scales and a sample (~3%) of participants are re-invited to clinics to validate measures, assess fieldworker bias and assist equipment calibration. Interview data from qualitative studies are collected and validated in real time and on encrypted laptops.

3. Data management, documentation and curation

Managing, storing and curating data

Data items are referenced and stored using a universal internal indexing and naming convention. All research data are stored separately from administrative data and linked through anonymised files accessible only to specified members of the team. Instrument specific and bespoke data formats are archived "as is" to ensure integrity of source material. All research data archived in University of Bristol (UoB) systems which are scalable, secure and backed-up routinely as detailed in UoB policy.

Metadata standards and data documentation

Metadata are collected as an integral process in order to (i) catalogue and index data, (ii) define the assessment tools, (iii) describe data collection, (iv) catalogue laboratory information and (v) assign appropriate geographical references. Research data are not made available until data items are clearly labelled and documented. Supporting documents and descriptions are all made freely available.

Data preservation strategy and standards

ALSPAC operates as a research resource and data is therefore maintained indefinitely. UoB infrastructure consists of real time mirroring of data across two geographically separate data centres. Data stored in UoB's Research Data Storage Facility are backed up to tape nightly in three geographically distinct storage locations, with backups kept for 30 days. Additionally, an annual deep freeze of all systems including management documentation is undertaken. UoB systems require multifactor authentication for access.

ALSPAC reviews data storage on a regular basis and will apply digital continuity methods where applicable to migrate data formats at risk of obsolescence. Where data disposal is required, this is done securely and in line with UoB information security policy. Primary source material are preserved – where possible electronically. The UoB library holds an extensive administrative archive (Special Collections ref:DM2616) of historical paperwork (grants, protocols, ethical approvals, participant information, keying/coding specifications).

4. Data security and confidentiality of potentially disclosive information

Formal information/data security standards

ALSPAC has been ISO27001 certified since 2012 (certificate number IS573207). The ALSPAC linkage team is compliant with the NHS Information Governance (IG) Toolkit Standard (ODS reference: 8J370). The UK Secure Research Platform (<https://serp.ac.uk/assurance-and-accreditations/>) which holds linkage data and also data hosted by UK LLC, is also accredited to ISO27001 and is an "accredited processor" under the Digital Economy Act 2017. The Bristol Bioresource Laboratory, which holds data related to ALSPAC biosamples, is certified to ISO9001 (certificate number: FS651591).

Main risks to data security

ALSPAC has certified process standards for data management and sharing, however there are inherent risks with cohort data, notably around disclosure, security breaches and data loss. These are addressed through the established data infrastructure and to minimise disclosure risk, the following are in place: complete dates are never made available; cell counts <5 are avoided; "split-stage protocols" are used with potentially disclosive individual data; sensitive data are stored in secure data environments; unique IDs are assigned to any data release or federation.

The access policy (www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf) outlines the consequences of security breaches and legally binding Data Access Agreement (DAA www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/alspac_data_access_agreement.pdf) are signed at an Institutional level before data release.

5. Data sharing and access

Suitability for sharing

It is the expectation that all ALSPAC data are suitable for sharing and all data generated are returned to ALSPAC in accordance with the legal agreements surrounding data use.

Discovery by potential users of the research data

Discoverability is driven by web material and a reputation for high quality data (www.bristol.ac.uk/alspac/researchers/our-data/). Data are described in an interactive timeline (www.bristol.ac.uk/alspac/researchers/our-data/timeline/) connecting assets to meta-data and also

in a variable search tool (<http://variables.alspac.bris.ac.uk>). ALSPAC actively promotes the use of data and resources through social media and open publication (<https://wellcomeopenresearch.org/gateways/alspac/about>). ALSPAC remains visible to all researchers through a collection of platforms serving different stakeholders. These include The Catalogue of Mental Health Measures (www.catalogumentalhealth.ac.uk), CLOSER Discovery (<https://discovery.closer.ac.uk>), UK Tissue Directory (www.biobankinguk.org), Dementias Platform UK (<https://portal.dementiasplatform.uk>), Birthcohorts.net (www.birthcohorts.net), Health Data Research UK Innovation Gateway (www.healthdatagateway.org). ALSPAC data are available and/or advertised through all but one of the access portals flagged by the Population Research UK prospectus and there is support for access to data via third party routes (www.bristol.ac.uk/alspac/researchers/our-data/external-sources/) and through simulated data for method development (www.bristol.ac.uk/media-library/sites/alspac/documents/ALSPAC-syntheticdata-checklist.pdf).

Governance of access

ALSPAC is committed to providing access to the resource directly and via indirect routes including secure data environments. Data and support are made available through an online proposal system and are regularly accessed (**Table 4** section 6. **Approach**). Approval is given data requested are available and release does not (i) disclosure participant identity; (ii) violate ethico-legal or other stipulations applicable; (iii) run the risk of harming the study or participants. Data are viewed as a non-finite resource and therefore access is not subject to formal scientific review. Datasets are provided upon completion of a DAA. Where we have concerns around security, we will complete a Data Privacy Impact Assessment.

Our material transfer agreement (www.bristol.ac.uk/media-library/sites/alspac/migrated/documents/appendix5b-hta-material-transfer-agreement.pdf) contains detail about the management of data associated with biosample use. This includes biosample manifests and information required for analysis (sample identifier, box location, co-variables). The MTA includes terms requesting that data generated must be returned to ALSPAC.

Requests for sensitive data, health and administrative records or from alternative governance backgrounds are subject to access constraints determined by ALSPAC and the original data owner. These constraints can involve seeking additional project clearance with the original data owner, the statistical modification of data to control for disclosure risks or explicit ethical approval.

The study team's exclusive use of the data

The ALSPAC team does not have exclusive access to research ready data. Where an external researcher has secured funding for the collection/analysis of new data, they may apply for up to 6 months exclusive access. The ALSPAC Executive will still consider other requests for access to this restricted data, but permission for collaboration is sought from the lead researcher.

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

Data sharing is only restricted if i) participants have withdrawn consent for sharing, ii) there are technical delays or iii) risks to information security or study integrity are identified.

Regulation of responsibilities of users

The access policy and DAA provide detailed information on the responsibilities of users with respect to maintaining privacy and onward sharing.

6. Responsibilities

The ALSPAC Executive takes ultimate responsibility for all aspects of data management. During the 2024-2029 funding period, the data team will be overseen by the Executive Director (Data) who has responsibility for new data released for research use and accompanying metadata. UoB takes responsibility for file storage.

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

Policy	URL
Data Management Policy & Procedures	www.bristol.ac.uk/research/environment/governance/research-data-policy/
Data Security	www.bristol.ac.uk/infosec/policies/information-security-policy-isp-01/ https://www.acrc.bris.ac.uk/acrc/RDSF_policy.pdf
Data Sharing	www.bristol.ac.uk/research/environment/governance/research-data-policy/
Institutional Policy	http://www.bristol.ac.uk/media-library/sites/infosec/documents/ISP-01v1.2.pdf
Other	www.bristol.ac.uk/research/environment/governance/research-data-policy/

8. Author of this Data Management Plan

The ALSPAC Executive
alspac-exec@bristol.ac.uk