UK Longitudinal Linkage Collaboration:

The National Trusted Research Environment for Longitudinal Research

Research Protocol

Contents

1. Executive Summary

UK Longitudinal Linkage Collaboration (UK LLC) is the national Trusted Research Environment (TRE) for the UK's longitudinal research community. UK LLC supports Longitudinal Population Studies (LPS) by providing record linkage and TRE services; and supports the research community by providing external researchers with secure access to a research database of integrated data.

LPS are studies that follow the lives of participant volunteers over time; often over whole lifetimes and generations of families. Data collected include biological samples, genomic data and in-depth and self-reported measures of health and wellbeing. LPS therefore provide unique insights into population health, behaviours and wellbeing. The scientific opportunities of LPS are enhanced when participants' data are collated and linked to their health and non-health administrative data (e.g. education, employment, tax and benefits records) and environmental exposure data, and made available to researchers to access via a single application process with distributive review to data owners. This provides a valued and unique resource for UK-based researchers and policy makers.

UK LLC is led by the Universities of Bristol and Edinburgh, in collaboration with UCL, Swansea University, the University of Leicester, City St George's, University of London and many of the UK's most established LPS. The partnership also includes a strong public contribution, with members of the public and LPS participants being active in decision-making and informing the system's design. UK LLC was established in 2020 as part of the COVID-19 Longitudinal Health and Wellbeing National Core Study (LHW NCS), because previous linkage approaches – where LPS independently arrange data sharing agreements with data owners – have been varied in success, burdensome and expensive to all parties, and were insufficiently responsive to crisis situations. UK LLC integrates data from UK LPS, linked to health, environmental and non-health administrative data, into a centralised research resource available in a secure environment called a TRE. Co-locating many LPS' diverse samples into one location creates a highly heterogeneous UK-wide sample, increases statistical power and the number of 'rare' exposures/outcomes and includes sub-groups of people who tend to be harder to reach in terms of data collection. UK LLC is now viewed by the UK longitudinal research community – including funders and government departments – as the national TRE for longitudinal research using linked records.

Initially supported by HM Treasury to underpin high priority COVID-19 research questions, UK LLC is now funded by UK Research and Innovation (UKRI), Economic and Social Research Council (ESRC) and Medical Research Council (MRC) to provide a centralised service at a four-nation level (England, Northern Ireland, Scotland and Wales) for systematic linkage of LPS participants' data to their health, environmental and non-health administrative records and a mechanism for the secure pooled analysis of data to enable research in the public good. UK LLC is designed to scale to include most UK LPS and to provide research database services to very large numbers of research users.

UK LLC's rigorous application process and a TRE based on the SeRP UK infrastructure at Swansea University ensure that UK LLC follows the 'Five Safes' set of safeguards that represent best practice in data management. UK LLC works at the highest levels of data security: UK LLC's Information Security Management System (ISMS) is ISO 27001 certified (certificate number 21069); UK LLC is accredited by the UK Statistics Authority (UKSA) as a processing environment under the Digital Economy Act 2017 (DEA); and UK LLC completes the annual NHS Data Security and Protection Toolkit (Organisation Number: EE133799-LLC).

The preliminary work to design and build the required 'minimum viable product' infrastructure and governance framework for urgent COVID-19 research was completed in mid-2021. The UK LLC TRE is now available to researchers as a generic purpose database, hosting diverse and complex data. As a model for efficient and low-burden linkages, the UK LLC TRE forms part of a responsive UK data science capability, which can be used by approved external researchers to conduct any public good research and which is sufficiently responsive to support investigation of emerging policy questions and to meet future crises such as new pandemics, the impacts of climate change or economic shocks.

2. Lay Summary

The UK has a long-standing tradition of conducting 'longitudinal research' to inform decisions about health care and social policy making. Longitudinal Population Studies work by selecting a group of individuals or properties and then repeatedly collecting data on these people (or the people living in the properties). The groups of people are typically selected by having something in common, such as pregnant women or people living in a certain area or of a certain age range. The value of longitudinal research lies in collecting a broad range of data, and then repeating this data collection at regular intervals. This allows researchers to investigate the interactions between different things that occur in individuals' lives and how changes in this can lead to changes in their health, wellbeing or personal circumstances. It is estimated that between 2-3 million people in the UK take part in a Longitudinal Population Study, and many take part across a lifetime – for example, the National Study of Health and Development is still collecting data from babies born in one week of 1946.

The data collected from Longitudinal Population Studies are held within research databases. Researchers wanting to investigate a particular research question submit separate applications to access data from each of these databases. Researchers need to demonstrate how their research will improve the public good and how they will maintain appropriate standards and information security. For many years the data provided by Longitudinal Population Study participants have been enhanced by 'linking' data collected by these studies to data in their participants' health and other routine government records (such as school records). This provides new information which is hard to collect directly from people and which can improve the accuracy of research findings. It is also a way in which groups who find active study involvement difficult can be included in the research and therefore benefit from the value research brings.

UK LLC has developed a new centralised approach for linking Longitudinal Population Study participants to their routine records and provided a secure computing system where these data can be used in research. This means that the data from each contributing Longitudinal Population Study can be analysed with data from other studies. Combining the studies allows research to be more accurate and allows researchers to study diverse population groups that may not be possible with one study alone due to small population groups. All the data being used are first processed by the Longitudinal Population Study so that participants' identifiers (such as name, NHS ID) are removed. The data are then stored and analysed in a specially designed secure research computing system. No data can leave this system except by a process that checks no one individual can be identified from the analyses (graphs and tables), and researchers using the system are audited to make sure they are following the rules. There are governance processes in place to make sure that the data can only

be used to help improve the public good. Longitudinal Population Studies continue to control the data from their participants and have a role in the research approval process.

UK LLC was originally set up during the COVID-19 pandemic as part of the UK Government's COVID-19 research response. Now established as a model for efficient, low-burden linkages and streamlined researcher access, the UK LLC Trusted Research Environment forms part of a responsive UK data science capability, which can be used to conduct any public good research and to meet future crises such as new pandemics, the impacts of climate change or economic shocks.

3. Scope

This protocol provides a full summary of all UK LLC activity. It is recognised that the breadth of UK LLC activities, and thus the scope of this protocol, span the remit of a range of jurisdictions, data owner and regulatory authorities. It is anticipated that this protocol will be read by a wide range of stakeholders, of which, some review boards will only be able to consider a subset of UK LLC activity.

A key stakeholder is the Health Research Authority, whose Research Ethics Committee and Confidentiality Advisory Group will review this protocol. This section clarifies the UK LLC activities and documents that lie within the scope of the Health Research Authority remit. Wider UK LLC activity is included to provide context and to ensure transparency of data processing.

3.1 Activities

UK LLC activities detailed in this protocol that are **in-scope** include**:**

- The establishment and maintenance of UK LLC as a research database, providing TRE functionality to UK LPS and the research community

- The provision of participant identifiers (identifiable Personal Data) by the contributing LPS to NHS Digital Health and Care Wales (a processor of UK LLC) and then onto:

    o UK NHS authorities (NHS England, Public Health Scotland, Northern Ireland Business Services Organisation and DHCW/SAIL Databank) to establish record linkage to participants' NHS records and for the selection and processing of these records

    o UK Statistics Authorities (including Office for National Statistics) to establish record linkage to participants' non-health administrative records (e.g. education, employment, tax and benefits) and for the selection and processing of these records

    o UK LLC's geocoding and environmental modelling partners, e.g. City St George's University of London for the geo-coding of participant address information and the linkage to participants' environmental exposure, property and neighbourhood records and for the selection and processing of these records

- The ingest of de-identified (processed to the point where re-identification is not reasonably likely and thus not considered Personal Data) linked NHS, non-health administrative data and place-based records into the UK LLC TRE

- The processing of the de-identified data within the UK LLC TRE to prepare the data for research purposes

- The involvement of public and LPS participant volunteer contributors in UK LLC's Public and Participant Involvement & Engagement (PPIE) programme

- The generic approval for external researchers to access de-identified data to undertake scientific research, including for:

  - The process for reviewing and approving applications from external researchers to use the de-identified data within the UK LLC TRE (described in section 10.3)

  - The provision of access to the UK LLC TRE to approved external researchers

  - The scientific objectives of the UK LLC TRE (described in section 11).

Some dependent activities are managed by third parties and are **NOT in-scope of this protocol.** These include**:**

- Fundamental activities of LPS, including funding, data collection and data management

- The production and circulation of fair processing materials, including the collection of participant consent where appropriate, by each individual LPS. These activities are subject to separate review and approval which is managed by each LPS.

Note: UK LLC does not and is not intended to hold any human tissue. Data derived from analysing tissue may be provided to UK LLC by the LPS subject to LPS-specific ethical approval for tissue processing.

Note: UK LLC does not and is not intended to have any direct contact with any LPS participant, (outside the scenario where a participant volunteers to contribute to the UK LLC PPIE programme).

3.2   Documents

Documents that accompany this protocol and that are **in scope** are:

- Appendix 1: LPS that contribute data to the UK LLC research database

- Appendix 2: UK LLC data flows (including legal basis)

- Appendix 3: UK LLC data access review process

- Appendix 4: UK LLC external researcher ethical requirement process.

Any amendments to documents in scope (this also includes this protocol) will be shared with REC for approval prior to incorporation into UK LLC's Information Security Management System (ISMS).

All other documents are not in scope and amendments to them are managed by UK LLC's internal governance processes. All documents in UK LLC's ISMS are listed on UK LLC's Document Register. All are reviewed on an at least annual scale and the approval route for each is clearly stated. These documents are not included, because they document how UK LLC applies this protocol, rather than being an integral part of this protocol.

4. Objectives

The UK has a world-leading portfolio of LPS which have collected detailed phenotypic and biological information on an estimated 2-3 million members of the UK public across more than 200 separate LPS[1].

Note: UK LLC only contains data relating to LPS participants where a legal and ethical basis has been established for their inclusion.

UK LLC has three main objectives:

1) To provide record linkage services to UK LPS. This includes bringing together de-identified data from the UK's LPS and to systematically link these data to their participants' health, non-health administrative (e.g. education, employment, tax and benefits) and place-based records in the UK LLC TRE in a manner that is compatible with ethical and legal requirements, is publicly acceptable and is efficient to all parties.

2) To provide a secure generic research database – the UK LLC TRE – which will support efficient access from a large number of UK-based approved external researchers for approved projects, so they can conduct public good research using diverse data from one or more LPS and diverse linked participant records.

3) To offer the UK LLC TRE as a long-term research database for data linkage and secure analysis for the longitudinal community, in order to support a broad range of research studies (spanning all public good longitudinal research).

UK LLC's systematically linked data and large combined sample provides the data foundation for a very wide range of public good research. Its design provides novel mechanisms for:

- Life-course measurement of health exposures and non-health administrative data (e.g. education, employment, tax and benefits records), covariates and outcomes

- Identifying patterns of service use to understand care pathways and health/non-health administrative intervention take-up to inform policy evaluation and econometrics studies

- Capture of data not readily amenable to direct measurement (e.g. environmental exposures) and integration of participant self-report data on behaviours and aspirations which are not captured and recorded in routine health and non-health administrative records

- Participant follow-up independent of active LPS participation, enabling more inclusive research and the means to quantify and address attrition and attrition bias, as well as selection bias

- The study of rare events/outcomes and seldom-reached population sub-groups

- A responsive data platform capable of rapidly informing crisis response (i.e. to inform public health and wider government policy in response to an emergency situation).

See section 11 'Scientific Remit' for further details.

---

[1] Atlas of Longitudinal Datasets (Atlas - Home).

5. Information Governance

The UK LLC TRE hosts de-identified data relating to LPS participants. Safeguarding the anonymity and security of LPS participants' data is of paramount importance to UK LLC. Sufficient technical and governance controls have been implemented, meaning that for core UK LLC staff and research users, the data are de-identified to the point where the risk of re-identification of an LPS participant is not reasonably likely; these data are therefore not considered to be Personal Data.

5.1 UK LLC's Information Security Management System

To provide assurance to the public, LPS participants, the contributing LPS, the NHS and other national data providers, UK LLC has developed an **information security management system** (ISMS) that is certified and accredited to the highest standards of data security. An ISMS is the framework of policies and procedures that include all legal, physical and technical controls that an organisation has put in place to safeguard its information assets – the data held in the UK LLC TRE and all the associated documents and processes. See the UK LLC Information Security Policy for further details.

- UK LLC's ISMS is **ISO 27001 certified** by independent industry assessors (certificate number: 21069). Ongoing certification is subject to satisfactory annual external audit, with a full re-certification audit every three years. ISO 27001 is an internationally recognised best practice standard for an ISMS

- UK LLC completes the annual **NHS England Data Security and Protection Toolkit** (Organisation Code EE133799-LLC). NHS England DSPT enables organisations to measure their performance against the National Data Guardian's 10 data security standards

- UK LLC is accredited by the **UK Statistics Authority** to process non-health administrative routine records under the Digital Economy Act 2017 (DEA). Ongoing accreditation is subject to satisfactory annual external audit, with a full re-accreditation audit every five years – see List of Digital Economy Act Accredited Processing Environments – UK Statistics Authority

- UK LLC follows the 'Five Safes' governance framework (Safe people; Safe projects; Safe settings; Safe outputs; Safe data)

- UK LLC has a Secure Environments Policy that sets out the requirements for the UK LLC TRE – the physical environment, the associated processes and the involvement of stakeholders – that ensure the anonymity and security (confidentiality, integrity and availability) of LPS participants' data.

UK LLC's Information Security Team has designed and delivers a programme of training, internal audit and reporting that ensures the maintenance and continuous improvement of UK LLC's ISMS.

5.2 Working in partnership with LPS

UK LLC works in partnership with all the contributing LPS to set the design and operating rules of the UK LLC TRE – these are reviewed at regular LPS 'Vanguard' meetings (membership comprises UK LLC staff and LPS Data Managers) and LPS Principal Investigator (PI) meetings (membership comprises UK LLC staff and LPS PIs). Through this partnership, UK LLC ensures the system incorporates appropriate safeguards to respect each LPS's conditions regarding data use. See **Appendix 1** for the list of LPS that collaborate with UK LLC and contribute data to the UK LLC TRE.

The following key rules have been established:

- LPS retain oversight of data flows and database usage in order to effectively audit the data they are responsible for

- Participant consent/objections are managed by each LPS and updated prior to the first record linkage and subsequently on a quarterly basis. All objections/withdrawals are respected (see section 5.4.3)

- LPS review each proposed use of their data in the TRE and can implement a veto as to whether their data are used or not (based on their specific data reuse rules and participant assurances) – see section 10.3 and the UK LLC Data Access and Acceptable Use Policy

- The purpose of the data use must be for the public good

- Data in the UK LLC TRE are de-identified and researchers access them on an 'effective anonymisation'[2] basis. Outputs from the TRE are at an aggregated population level and are reviewed for disclosure risk before leaving the TRE – see section 9.3 and the UK LLC Output Review Policy.

5.3 Working in partnership with LPS participants and publics

UK LLC has developed a Public Involvement Programme to embed a public voice across its design and operations. The programme's objectives are to develop and maintain a sustained presence of both LPS participants and members of the public. This work sits across strategic, operational and communications functions with a view to engender public trust and to seek continuous improvement and understanding around public acceptability of data use for research. Through a variety of recruitment methods, the programme includes LPS participants, NHS service users, parents, carers, and people with experience of disability, neurodiversity and long-term health conditions. UK LLC's framework comprises two-year terms with options to extend or transfer within the programme and is supported by documentation including terms of reference, code of conduct and payment policy.

Across the programme, the public contributors help to inform and co-develop policies, procedures and materials relating to all aspects of UK LLC's functions. They also inform UK LLC's responses to emerging issues, e.g. changes in law and regulation. Public co-development is formalised through terms of reference and supported by training activities delivered by dedicated UK LLC public involvement staff. UK LLC is committed to knowledge exchange and insights from our public involvement programme are publicised through blogs, conference presentations and research publications. All outputs are made transparent to contributors who are given opportunities for co-authorship and involvement and, where this occurs, contributors are provided with fair processing and consent materials. The UK LLC Public Advisory Group helps ensure these materials are clear and appropriate.

---

[2] Anonymisation: managing data protection risk code of practice. Information Commissioner's Office. Available from: Anonymisation: managing data protection risk code of practice (ico.org.uk)

UK LLC groups or panels with public representation include:

- UK LLC Strategic Advisory Committee – lay members provide high-level input into UK LLC's strategic development

- UK LLC Public Advisory Group – lay members guide UK LLC in developing policies, methods and communications materials

- UK LLC Data Access Public Review Panel – lay members review researchers' applications to access UK LLC held data with a focus on evaluating likely public benefits and reviewing lay descriptions and public involvement strategies

- UK LLC Citizen Panel – lay members assessed the UK LLC data access process and provided key findings and recommendations

- UK LLC Confidentiality Due Diligence Panel – lay members review documentation involved in onboarding new partner LPS to make sure they meet agreed criteria for communicating with their participants

- UK LLC Public Involvement Network – lay members provide advice on strategic issues, risks and challenges.

5.4   Transparency and participant preferences

5.4.1   UK LLC Transparency

UK LLC commits to upholding best practice expectations for publicly transparent data science – see UK LLC Secure Environments Policy. UK LLC's public facing website includes pages developed specifically for the public (https://ukllc.ac.uk/public-involvement) and for LPS participants (https://ukllc.ac.uk/safeguards). The website information is designed using the Information Commissioner's Office recommended 'layered approach', using a variety of media (e.g. infographics, animations) and all the content is co-developed with the UK LLC Public Advisory Group.

The website materials explain UK LLC's approach and commitment to participants and includes a link to the UK LLC Data Use Register, which details all applications to access data in the UK LLC TRE, including which LPS are contributing data to each project. The aims of each application are explained in a short lay summary, the content of which is approved by UK LLC's Data Access Public Review Panel. When papers are published, the Data Use Register is updated with a short lay summary detailing the research findings, with a link to the paper provided on the Publications Page. The UK LLC website hosts a formal fair processing Privacy Notice. UK LLC also has a LinkedIn profile.

5.4.2   LPS Transparency

The contractual agreement between UK LLC (University of Bristol) and contributing LPS includes a requirement for each LPS to ensure that their involvement in UK LLC is made clear to their participants and that a mechanism to object is made available.

Contributing LPS are managed by a number of UK universities and NHS organisations under their legal and ethical frameworks and in line with UK-wide data protection legislation including GDPR. Fair processing materials – information detailing how LPS participants' personal data will be used in connection with the UK LLC TRE – are shared with LPS participants both directly by the LPS through their established communication channels (e.g. through newsletters, websites and direct interaction) and globally by UK LLC (via UK LLC's website and social media activities). These materials

explain how their data are to be shared with other organisations, the purpose of the data sharing and how they can opt-out or withdraw from their data being sent to the UK LLC TRE or linked with their routine records. Each LPS manages fair processing with their own participants, with UK LLC providing guidance on best practice. Fair processing updates will continue to be made available to LPS participants as the scope of UK LLC's processing activities change (e.g. through newsletters, privacy notice updates). The UK LLC Confidentiality Due Diligence Panel reviews the consent materials of all partner LPS – see section 6.2 for further details.

Specific LPS fair processing is out of scope for this protocol and constitutes part of the LPS' protocol and ethical review/approvals.

5.4.3   Participant preferences

LPS control which datasets their participants' data are linked to and all participants have the right to opt-out of UK LLC altogether or to opt out of all linkages, or only specific linkages – all participants' objections are upheld by UK LLC. As explained in section 9 in more detail, LPS send quarterly updates of Participant Identifiable Data (PID) or File 1s, containing participants' identifiers and their associated permission flags, to UK LLC's Trusted Third Party for linkage (NHS Digital Health and Care Wales, DHCW). Opting out means that no further data about that participant will flow into the UK LLC TRE (i.e. the LPS will not send additional data and the linkage to routine records will cease) and the participant's data will not be provisioned to new research projects. However, researchers who already have access to that individual's information will be permitted to retain that access until the end of the project, but they will not obtain any new data about that individual. Where using section 251 (s251 for English and Welsh participants), NHS National Data Opt Out will be applied (and has been applied for every UK LLC NHS England data extraction).

6. Legal Basis

UK LLC has conducted an assessment as to whether the data held and processed within the UK LLC TRE are identifiable or potentially identifiable and therefore Personal Data. The assessment determined that while the infrastructure, incorporating all data flows and all parties, should be considered as containing Personal Data and therefore subject to Data Protection Act 2018 (DPA18)/UK General Data Protection Regulation (GDPR) and Common Law requirements, data within the UK LLC TRE are considered de-identified to the point where re-identification is no longer reasonably likely and the data are not Personal Data. This means that UK LLC staff and external research users do not process Personal Data and their use of the data does not result in a breach of confidentiality.

6.1   Legal basis for UK LLC

UK LLC is owned by the University of Bristol. The University of Bristol's enacting legislation includes a remit and lawful basis to conduct research. The legal basis for UK LLC under UK GDPR and the DPA18 is: 1) Performance of a task carried out in the public interest (Article 6(1)(e) in the UK GDPR); and, for

the use of sensitive personal information[3], 2) Scientific or historical research purposes or statistical purposes (Article 9(2)(j) in accordance with Article 89(1)).This legal basis within UK GDPR and the DPA18 is separate to, and in addition to, the actions and legal basis of the collaborating LPS, which establishes the basis for LPS participants' data to be collected, processed and shared for research purposes.

6.2   Legal basis for flowing LPS data

The UK LLC TRE hosts data about LPS participants that are pseudonymous to the LPS staff and are therefore Personal Data to them. LPS that contribute data to the UK LLC TRE have their own legal basis for collecting, processing and sharing their participants' data for research purposes. This can generally be found in the privacy notice on each LPS's website or in their Participant Information Sheets. Each contributing LPS needs to provide evidence to UK LLC (UoB) that they have a legal basis to access these data and to provide participant identifiers to UK LLC's Trusted Third Party for linkage (NHS Digital Health and Care Wales) for linkage. The legal basis for LPS under UK GDPR and DPA18 will typically be: 1) Performance of a task carried out in the public interest (Article 6(1)(e) in the GDPR); and, for the use of sensitive personal information, 2) Scientific or historical research purposes or statistical purposes (Article 9(2)(j) in accordance with Article 89(1)). The LPS meet Common Law Duty of Confidentiality either through explicit consent, or through section 251 consent exemptions under the Health Service (Control of Patient Information) Regulations 2002 within England and Wales (Regulation 5 with support from the Health Research Authority's Confidentiality Advisory Group), or through undergoing public interest test assessments by the relevant UK devolved authorities.

UK LLC conducts an NHS England approved due diligence review of each LPS' legal basis. The UK LLC Confidentiality Due Diligence Panel includes senior sponsor representation (ethics, governance and legal data protection leads) from the University of Bristol and public contributors. The Panel advises whether each LPS has established a basis for addressing their Common Law Duty of Confidentiality appropriately and ensured compliance with the principle of 'no surprises'. The Panel reviews each LPS' consent materials (consent forms and participant information sheets) to identify if there are any statements that either limit or introduce ambiguity into the scope of an LPS' proposed data processing. Evidence from this process is provided to data owners (e.g. NHS England, Office for National Statistics) for their records and due diligence purposes.

Specific LPS legal basis is out of scope for this protocol and constitutes part of the LPS's protocol and ethical review/approvals.

6.3   Legal basis for flowing linked records

A valid legal basis will need to be in place to enable government departments – including the NHS in England and devolved NHS authorities – to share data with UK LLC for research purposes. This basis varies across departments, between health-related and non-health administrative records, and

---

[3] UK GDPR defines 'sensitive personal information' as information that reveals a person's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership; and the processing of genetic data or biometric data for the purpose of uniquely identifying a person; data concerning health or data concerning sex life or sexual orientation.

between the UK's four nations. All government departments will need to determine their appropriate legal basis under GDPR and DPA18 for flowing data into the UK LLC TRE. Below we summarise the ways in which UK LLC applies participants' preferences for the flow of their routine records.

6.3.1   NHS health records

6.3.1.1   NHS data from England

The Health and Social Care Act 2012 provides a statutory basis for the sharing of health data in England and the linkage and processing of health data in England and Wales is permitted under the NHS Act 2006. To address Common Law, UK LLC's partner LPS have a legal basis for enabling linkage, through either consent or Control of Patient Information Regulations (COPI) 2002. While some partner LPS use either consent or COPI as their legal basis, other LPS operate a blended consent/COPI model, because of the complexities surrounding seeking retrospective consent, to reduce bias and to improve inclusion in order to deliver benefits to vulnerable and marginalised communities. UK LLC has approval from the Secretary of State for the use of COPI support following a recommendation of the HRA Confidentiality Advisory Group (21/CAG/0044). UK LLC has established a framework process to conduct due diligence that partner LPS have an appropriate legal basis in place.

6.3.1.2   NHS data from Scotland

UK LLC has received conditional approval from the Scottish Public Benefit and Privacy Panel (PBPP) to flow Scottish health data to the UK LLC TRE. For those LPS that use consent as the basis for determining permissions for the data flow, there are already precedents in place for this approval (e.g. Generation Scotland). UK LLC will explore the mechanisms for accessing data for LPS that use s251 approval in England and Wales with the relevant authorities in Scotland.

6.3.1.3   NHS data from Wales

The linkage and processing of health data in England and Wales is permitted under the NHS Act 2006. NHS Wales flows de-identified data into the SAIL Databank (based on SeRP UK infrastructure within the Swansea University system) in such a way that it is not Personal Data whilst in the protective controls of the SAIL Databank TRE and SeRP infrastructure. DHCW extracts consented and s251 participants' NHS records from the SAIL Databank into the UK LLC TRE.

6.3.1.4   NHS data from Northern Ireland

It is UK LLC's intention to develop governance approvals to allow linkage to Northern Irish NHS records (Northern Ireland Business Services Organisation) and the flow of these into the UK LLC TRE. This work remains under discussion and is not included within this version of our protocol.

6.3.2   Non-health administrative records

The Digital Economy Act 2017 (DEA) section 64 provides an appropriate legal basis for non-health administrative (e.g. education, employment, tax and benefits) linkages and data processing. As detailed in section 5.1, UK LLC is accredited by UK Statistics Authority as a processor under the DEA. This permits UK LLC to flow non-health administrative data into the UK LLC TRE. UK Statistics Authority have provided guidance that data flowing under different gateways can be used together for research purposes (i.e. that participant data, NHS records and data flowing under DEA can be integrated into a research database).

7. A Secure Analytical Platform for UK LLC

7.1   Principles

UK Health Data Research Alliance has published principles and best practices for TREs[4]. The transition to TRE ways of working has been accelerated by the Goldacre Review[5] and the Department of Health and Social Care's Data Saves Lives policy[6], which emphasise the importance of the focused development of a limited number of TREs (also known as Data Safe Havens or Secure Data Environments).

Most UK TREs, including UK LLC, have adopted the 'Five Safes' governance framework for the design and management of TREs.

A Five Safes TRE must have:

1) **Safe data:** expert data curation and safeguards to effectively anonymise data, to respect participants' objections and to support accurate research.

2) **Safe projects:** an application review process ensuring compliance to Five Safes – applicants need to demonstrate that their proposed research and data request is appropriate, ethical and likely to deliver clear public good that will be disseminated.

3) **Safe people:** only researchers who are trained and authorised to use data safely can access the data.

4) **Safe settings:** a secure analysis environment where access and analysis of data can only take place within the system and where controls mean that only anonymous research findings can leave the system. Safe settings can be externally accredited to best practice security and regulatory standards, including ISO 27001, NHS England DSPT and UK Statistics Authority's DEA Processor Accreditation[7].

5) **Safe outputs:** outputs are screened and only those that are non-disclosive are approved for export from the TRE.

A TRE will only be considered legitimate, regulatory-compliant and trustworthy where it demonstrates all of these functions to the public and data owners.

The UK's data science community as a whole is increasingly adopting TRE ways of working; a process which is now driven firstly by the need for compliance with the DHSC Data Saves Lives policy, which sets the requirement for most research data processing to take place within TREs; and secondly by the need to ensure compliance with DEA 2017, which also requires TRE based data use. Aligning with both NHS and non-health administrative legislation enables support for research at the intersection of health and socio-economic factors (e.g. the investigation of health inequalities). The move to TRE-

---

[4] UK Health Data Research Alliance (2021). Building Trusted Research Environments: Principles and Best Practices; towards TRE ecosystems.

[5] Goldacre, B & Morley, J. (2022). Better, Broader, Safer: Using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care. Department of Health and Social Care.

[6] Department of Health and Social Care (2022). Data saves lives: reshaping health and social care with data.

[7] https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/better-access-to-data-for-research-information-for-processors/

University of BRISTOL   THE UNIVERSITY of EDINBURGH   UK Research and Innovation   Economic and Social Research Council   Medical Research Council

based working is also being driven by considerable public/participant feedback suggesting that the added safeguards that TREs deliver are needed for contemporary longitudinal data science to remain trustworthy.

This provides the rationale for UK LPS to move to using linked LPS and routine records within TREs even where consent exists. UK LLC is designed to provide the centralised infrastructure to enable this. While there are precedents for the use of TREs in longitudinal research[8] and some very large biomedical LPS are suitably resourced to maintain their own independent TREs[9], most UK LPS do not have TREs in place and nor do they have the resources to develop and maintain a bespoke TRE.

No existing TRE has the remit, basis or functionality to meet the needs of all LPS, which require broad linkages across health and non-health domains and across the UK's four nations. Further to this, the diverse and long-standing governance requirements of LPS-specific ways of working and past assurances to participants describing the ways in which data sharing would be conducted, has also been a barrier to establishing TREs. To meet this necessitates a bespoke governance framework tailored to the needs of the longitudinal community, a requirement which is distinct to the models operated by the NHS or ONS. UK LLC will seek accreditation to the NHS Secure Data Environment standard once this is published.

The combination of these factors made a compelling case to coordinate linkage by UK LLC within a centralised TRE. The most feasible option to implement this was to establish a collaboration between expert infrastructure/governance parties (Universities of Bristol, Edinburgh, Swansea (including SeRP UK) and UCL) and with contributing LPS. In this model the Universities of Bristol and Edinburgh manage the governance, data curation, data application and public involvement aspects of UK LLC (and lead the programme as a whole) with University of Bristol acting as Data Controller. SeRP UK, part of Swansea University, provides their existing and accredited technological secure research infrastructure, ensuring this is managed by specialist staff; with place-based modelling partners, e.g. City St George's, University of London providing place-based expertise; and, UCL providing expert interdisciplinary guidance in longitudinal research. This collaborative model makes best use of existing centres of expertise, ensuring sustainable funding models and taxpayer efficiency. The active involvement of the LPS in the co-design and operation of UK LLC ensures the system is designed to meet the collective and interdisciplinary needs of all contributing LPS.

7.2   UK LLC Trusted Research Environment

UK LLC contracted Swansea University to provide a Secure eResearch Platform (SeRP UK) infrastructure for the UK LLC TRE and their partner, NHS Digital Health and Care Wales (DHCW), to conduct trusted third party data linkage services. Both SeRP UK and DHCW are ISO 27001 certified and DHCW is also DEA accredited. Both are contracted to University of Bristol as Data Processors for UK LLC.

---

[8] For example, ALSPAC birth cohort study operates an ISO 27001 and NHS DSPT certified TRE based on SeRP UK infrastructure and a bespoke governance layer developed by ALSPAC. This approach incorporates participant requirements and also meets data owner expectations and those from ethical review panels.

[9] UK Biobank, Our Future Health and Genomics England are examples of LPS that will maintain their own approaches.

SeRP UK was designed to provide secure research computing facilities for data science[10], including LPS. SeRP UK provides the secure infrastructure for the SAIL Databank and hosts individual LPS such as ALSPAC[11] and the Millennium Cohort Study[12], and LPS consortia initiatives such as Dementias Platform UK[13]. SeRP UK was seen as the ideal TRE infrastructure to support the UK LLC TRE given that:

- It is recognised by key stakeholders as having a robust and independently accredited ISMS

- Its linkage capabilities enable the linkage of NHS records and other sources with diverse identifiers in a privacy-preserving manner; DHCW does not have access to any individual attribute data (beyond individual identifiers)

- The use of an independent organisation as a trusted third party to solely handle participant identifiers enables the central UK LLC TRE to be a fully de-identified environment where hosted data is effectively anonymous to all research analysts and the system administrators

- The SeRP UK security and governance advantages supports UK LLC accreditation requirements and is consistent with LPS reassurances to participants (e.g. where data are de-identified during analysis, or where LPS have made assurances that record linkage is not a means by which identifiable participant data are transferred to the government)

- SeRP UK also provides similar functionality for the Welsh national databank (SAIL Databank), and within Northern Ireland and can also provide federated access to the Scottish TRE. This provides a secure basis for the sharing and use of devolved non-health administrative data.

UK LLC has co-developed the UK LLC TRE with contributing LPS. The data flows have been designed with consideration for the non-health administrative, legal and jurisdictional boundaries of the devolved NHS and other government systems, while considering options to leverage UK-wide insights through proportionate and acceptable ways of working.

The UK LLC TRE provides the platform for LPS to upload data, to establish record linkages and to extract the routine records of participants, for UK LLC staff to conduct data management and integration (across LPS and linked NHS and other records) and then to provide approved research users with separately partitioned areas for their approved projects. All analysis of data takes place within the TRE and only research outputs checked for disclosure risk are permitted to exit the system (these are fully anonymous, even when placed in the public domain).

---

[10] Jones KH, Ford DV, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. The Lancet. 2016 Nov 1;388:S62.

[11] Cornish RP, John A, Boyd A, Tilling K, Macleod J. Defining adolescent common mental disorders using electronic primary care data: a comparison with outcomes measured using the CIS-R. BMJ open. 2016 Dec 1;6(12):e013167.

[12] Millennium cohort study data for Welsh residents has been deposited in the SAIL databank and linked to Welsh resident records (https://data.ukserp.ac.uk/Asset/View/52).

[13] Bauermeister S, Orton C, Thompson S, Barker R, Bauermeister J, Ben-Shlomo Y, Brayne C, Burn D, Campbell A, Calvin C, Chandran S. Data Resource Profile: The Dementias Platform UK (DPUK) Data Portal. BioRxiv. 2019 Jan 1:582155.

7.3   Data Access within the UK LLC TRE

Four groups of stakeholders may be permitted access to the UK LLC TRE, each for **distinct processing operations** (all external processors are covered by contracts with UoB):

1) **UK LLC University of Bristol staff**: curation of data (processing of data, data and linkage quality assurance, building of 'research ready data' including derived indicators and the archiving and reuse of project-based outputs); data documentation (including descriptive analysis, analysis of bias and representation); provisioning of minimised sub-sets of the data to approved researchers for approved purposes; conducting research feasibility assessments; and audit of researchers and their projects to check appropriate permissions are in place and that analyses fall within the scope of approved projects.

2) **Data processors at Swansea University**: maintaining the SeRP infrastructure and software (including the UK LLC TRE); ingest of data into the UK LLC TRE and documentation; and monitoring of analytical outputs for statistical disclosure control.

3) **Approved external research users** (UK legitimate research community): researchers are only able to access a read-only view of data minimised to the needs of their specific approved project and where the rights of objecting participants are also respected. Researchers are not able to access or alter the central UK LLC resource. They request outputs for presentations/publications that are statistical disclosure checked prior to release. They document and make available research tools (syntax, code lists, derived data variables) for reuse within the UK LLC TRE.

4) **Approved LPS staff** (various University and NHS organisations): LPS staff (as appropriate within LPS governance structures) from each contributing LPS have an operating area within their own LPS-specific demarcated area with access to both LPS collected data and linked records for their participants only. LPS staff use this area for data management (including data processing and cleaning, project feasibility assessments) and documentation (including descriptive analysis, analysis of bias and representation). The data provided are indexed using a re-encrypted individual-level ID which is specific to that LPS. LPS staff do not have access to the key to reverse the ID encryption back to their own participant identifiers. LPS staff are not able to access or alter the central UK LLC resource and cannot export data from the TRE. Any LPS-led applied research analysis is treated as a 'research project' and subject to the standard research project process (i.e. that described in group 3 – Approved external research users).

8. Types of Data

The UK LLC TRE hosts diverse and complex data about LPS participants. UK LLC integrates data from LPS and links to participants' health, place-based and non-health administrative records, where permissions allow. For all classes of data (LPS, NHS, non-health administrative and place-based data), only de-identified data flow into the UK LLC TRE and the data are de-identified at source by the data owner.

8.1   LPS data

The UK's LPS have collected an exceptionally broad and detailed array of data. Across the LPS contributing to UK LLC, these data cover most conceivable topics and include detailed information

about subjects that are discussed openly in society to those that are highly sensitive. Many LPS are designed as generic research databases to collect data for future (unspecified) research applications. These data have already been collected by the LPS for the purpose of research use by approved researchers, with UK LLC providing a means to facilitate this existing purpose.

For UK LLC to service this model, LPS deposit a very wide range of data into the TRE including:

- Broad ranging quantitative data on participants' demographic, socio-economic status (SES) and health status (physical and mental); family, SES and life-course indicators; and, information on diverse behaviours, aspirations and outcomes

- Summarised qualitative data on diverse topics collected directly from participants via interviews and focus groups

- Existing assayed biological information (e.g. blood group type, biomarkers such as serology)

- Genetic, metabolomic, proteomic and epigenetic information, including whole genome sequence data and other 'omic datasets

- Image data (such as brain or organ MRIs, DEXA bone density scans, retina images, routine cancer or other scans)

- Data about interactions with companies and devices, for example data extracted from participants' supermarket loyalty cards or smartphone applications

- Specific COVID-19 data collections (questionnaires and assayed biological data)

- Participant participation history and consent/opt-out status.

These are collected directly from participants via face-to-face or remote (online, postal) interviews/assessments or via linkage to routine records or novel data sources (such as social media posts, personal sensors, images).

Participant identifiers, including historical address data are used for record linkage purposes and to derive de-identified research indicators (e.g. encrypted geographical indicators from property level to postcode and electoral/census ward areas for clustered and multi-level analysis).

8.2   NHS data

To meet the needs of our current and anticipated research users, it is necessary to extract individuals' full life-course records (to support longitudinal assessments and to build health care utilisation pathways) and for the extract to be refreshed on a timely basis to ensure an accurate assessment of disease and outcomes. The data collection specification used across the UK nations will ideally be prospectively aligned to enable efficient data integration.

The intention is to extract as full coverage of participants from the UK population as possible. However, it is recognised that each contributing dataset will have its own inclusion/exclusion criteria, that some cases will be excluded from some data through setting patient objections (e.g. in England the primary care extract will respect Type 1 patient objections which block the sharing of records for purposes other than direct care), and that the NHS system has protective mechanisms to restrict the sharing of patient health records in certain sensitive situations. This may mean that some

vulnerable and marginalised groups are systematically excluded[14]. All data extracted under version 1 of the UK LLC protocol for COVID research purposes (the 'historical data') will be retained as they are directly relevant for the new research purpose and will support ongoing COVID research[15].

8.2.1 NHS England

NHS England provides access to English health records (Table 1).

Table 1 Health datasets in England

Controller	Dataset	Contents
NHS England	Demographics	Basic patient details such as date of birth, sex and entry and exit to NHS services.
	Civil Registration – Deaths (England & Wales)	Deaths registration data collected from The Registrar General for England and Wales. Record-level person data set, where a record represents one death registration.
	Cancer Registration Statistics (England)	Notifications on cancer status for people living in England who are diagnosed with malignant and pre-malignant neoplasms.
	Hospital Episode Statistics (HES) Admitted Patient Care, Critical Care, Outpatients, Accident & Emergency	Record-level patient data set of patients admitted for treatment at NHS hospitals in England.
	Emergency Care Dataset (ECDS)	The national data set for urgent and emergency care.
	Maternity Services Dataset (MSDS)	Provides information about the mother's demographics, booking appointments, admissions and re-admissions, screening tests, labour and delivery, along with the baby's demographics, admissions, diagnoses and screening tests.

---

[14] Boyd A, Thomas R, Macleod J. (2018). NHS Number and the systems used to manage them: an overview for research users. Bristol, UK: University of Bristol. Available from: https://www.closer.ac.uk/wp-content/uploads/CLOSER-NHS-ID-Resource-Report-Apr2018.pdf

[15] Where necessary, the purpose for which these historical data can be used will be restricted in line with NHS data controller conditions (e.g. the covid-specific datasets, including the pandemic Primary Care extract, will only be made accessible for COVID-19 relevant research). The alternative – to delete the historical extract and re-extract equivalent data – would be burdensome to UK LLC and the NHS and would have a high cost to the taxpayer and provide no meaningful benefits as changes in permission status are already managed through separate mechanisms and National Data Opt-Out has been applied to all UK LLC data extractions.

Controller	Dataset	Contents
	Mental Health Services Dataset (MHSDS)	Patient-level clinical and non-health administrative data for patients who are in contact with NHS funded mental health services in England.
	Talking Therapies for Anxiety and Depression (formerly Improving Access to Psychological Therapies)	Information about people in contact with adult psychological therapy services in England
	Community Services Dataset (CSDS)	National definitions for the extraction of data about children and adults: • Personal and demographic • Social and personal circumstances • Breastfeeding and nutrition • Care event and screening activity • Diagnoses • Scored assessments.
	Medicines Dispensed in Primary Care (NHSBSA)	Prescriptions for medicines that are dispensed or supplied by community pharmacists, appliance contractors and dispensing doctors in England.
	Primary Care GPES Data for Pandemic Planning and Research (GDPPR)	GP data that provides data to support vital planning and research into COVID-19.
	COVID-19 Vaccination Status	Data relating to COVID-19 vaccination status.
	COVID-19 Vaccination Adverse Reactions	Data relating to any adverse reactions which occur within the first 15 minutes after administration of the COVID-19 vaccination.
	COVID-19 Severe Acute Respiratory Infection (SARI-WATCH) surveillance system (formerly CHESS)	Data that relate to demographic, risk factor, treatment and outcome information for patients admitted to hospital with a confirmed COVID-19 diagnosis.
	COVID-19 Second Generation Surveillance System (SGSS)	Data relate to demographic and diagnostic information from Pillar 1 swab testing in PHE labs and NHS hospitals and Pillar 2 Swab testing in the community.
	COVID-19 UK Non-hospital Antigen Testing Results (also referred to as NPEX)	COVID-19 non-hospital Antigen Testing Results (Pillar 2) data.
	COVID-19 UK Non-hospital Antibody Testing Results (also referred to as iElisa)	Data relate to individuals who have undergone a finger prick test for antibodies from having had COVID-19.

8.2.2   Public Health Scotland

Public Health Scotland provide access to Scottish health records in conjunction with National Records Scotland for registry data (Table 2).

Table 2 Health datasets in Scotland

Controller	Dataset	Contents
Public Health Scotland (PHS) PHS/Albasoft	Demographics data	Basic patient details such as date of birth and sex.
	Scottish Morbidity Records	- Outpatient appointments (SMR00) - Inpatient/daycase activity (SMR01) - Maternity interaction/stay (SMR02) - Mental health stay (SMR04) - Scottish Cancer Registry (SMR06).
	Prescribing Information System (PIS)	Data on all medicines and their costs that are prescribed and dispensed in the community in Scotland.
	Unscheduled Care	Data from: - Accident and Emergency - GP Out of Hours - Scottish Ambulance Service - NHS 24.
	Scottish Immunisation and Recall System	All positive microbiology laboratory specimen results and a subset of antimicrobial susceptibility/resistance data in Scotland.
	The Scottish Intensive Care Society Audit Group (SICSAG)	A national database of patients admitted to adult general Intensive Care Units (ICU) in Scotland.
	Disease specific datasets	- Scottish Stroke Care Audit - Scottish Cardiac Audit Programme.
	COVID-19 specific datasets	Outcomes from citizen serology testing, COG-UK SARS-CoV-2 sequencing data, vaccination data, diabetes covariates.
	Deaths	All Registrations to the National Records of Scotland of deaths.

8.2.3   SAIL/NHS DHCW

Welsh NHS records for Welsh residents are centralised and stored within the Secure Anonymised Infrastructure for Linkage (SAIL) Databank in partnership with NHS Digital Health and Care Wales (DHCW) (Table 3).

22

Table 3 Health datasets in Wales

Owner	Data Set	Contents
SAIL Databank/DHCW	Welsh Demographics Service	Basic patient details such as date of birth and sex.
	Welsh Secondary Care records	• Critical Care Dataset • Emergency Department Dataset • Outpatient Dataset & Referrals • Inpatients (Patient Episode Dataset for Wales) • Maternity Indicators dataset • Welsh Cancer Intelligence & Surveillance Dataset • Welsh Ambulance Services Trust • NHS 111 Call data • Intensive Care National Audit and Research Centre.
	COVID-19 related datasets	• COVID-19 Test Results • COVID-19 Test Trace and Protect • COVID-19 Vaccinations • COVID-19 Shielded People List.
	Annual District Death Extract	Registration of deaths.
	Welsh Longitudinal GP Dataset	Primary Care Record.

8.2.4   Northern Ireland Business Services Organisation

See section 6.3.1.

8.3   Non-health administrative data

UK LLC is an interdisciplinary database supporting research into individuals' socio-economic outcomes (e.g. research to understand differential outcomes during key transitions, such as from education to employment) and to understand health/socio-economic interactions (e.g. ingrained health inequalities, occupational exposures and health outcomes, predictors for positive/adverse healthy and independent living in ageing populations).

UK LLC will link non-health administrative data via the Office for National Statistics (ONS) and in agreement with the source data owners. Data from the Department for Work and Pensions (DWP), HM Revenue and Customs (HMRC) and four nations education department records will flow into the UK LLC TRE because these provide key indicators of exposures and outcomes (and controlling/ mitigating factors).

8.3.1   HM Revenue and Customs

Data items relating to employment payments; workplace pensions; employment cessation payments; nature and source of income; pensions; share schemes.

- Start/end date of employment spells, and the timing of payments within each spell
- Employer information (type of payer (e.g. individual/partnership etc.); address; trade classification number, hours worked; turnover)
- Employment information (self-employment marker, SOC, hours worked)
- Tax (tax payable; tax relief; allowances; student loans)
- Tax linked benefits (e.g. child tax credit, working tax credit).

8.3.2   Department for Work and Pensions

Customer Information System (CIS); Benefits and Income Data (BIDS); National Benefits Database (NBD); Child Benefit Extract (CHB).

- Details of benefits and income including Universal credit, child benefit, housing benefit, tax credit, incapacity benefit, carers benefit
- Claims to key DWP benefits including, for example, type of claim, start and end dates, claimant characteristics such as DOB and partner status
- Information specific to Child benefit payments
- Some benefits are HMRC/DWP (e.g. child tax credit, working tax credit).

8.3.3   Four nations education records

National Pupil Database (NPD); Pupil Level Annual School Census (PLASC) ages 4-18 years.

- Key stage Attainment ages 5-18 years – Attainment data based on teacher assessments (early years) and teacher and externally marked tests including GCSE, A level and equivalent
- School census data collected in January from pupils in state schools. Includes background demographic characteristics and:
  - Special Educational Needs indicators – indicators of physical health conditions and neurodevelopmental status/conditions
  - Absences – information on pupil absences by term and annually by reason for absence
  - Exclusions – information on pupil temporary and permanent exclusions from school, including reason for each exclusion.

8.4   Place-based data

UK LLC has commissioned place-based modellers, including the University of Leicester and City St George's, University of London to model a number of environmental exposure estimates. These are based on taking existing environmental sensor readings and with other model inputs (e.g. traffic count data, weather pattern data) used to model pollution and other environmental exposures and to map these to participants' addresses (where LPS/participant preferences permit). UK LLC also collates and processes public domain datasets relating to a participant's address or neighbourhood (e.g. Lower Super Output Area)for inclusion in the UK LLC TRE. All geographic 'place' indicators

below the level of English 'region' and devolved nation[16] are encrypted prior to ingest to UK LLC TRE. The encrypted variables cannot be reversed by UK LLC staff or researchers and provide a consistent variable for clustered and multi-level modelling from the level of the property (e.g. household) through to neighbourhood.

8.4.1   Geospatial modelling and public domain datasets

These are assigned to a property, postcode or higher-level geography (e.g. a Lower Super Output Area or region). These are sourced from public-domain datasets which are processed to ensure they are de-identified prior to ingest into UK LLC. Public domain datasets are processed by UK LLC in line with a decision-making framework which acknowledges that anonymisation is heavily context-dependent and considers both data utility and the context in which the data is analysed to inform the controls and variable transformations that are needed.

- Modelled environmental exposure estimates of pollution, climate data (e.g. temperature and rainfall, pollen) and noise
- Modelled access to green and blue space, measures estimating the 'walkability' around a property or area
- Information about the neighbourhood (e.g. building density, land use characteristics, deprivation indices, crime rates, provision of services, availability of 'hazards' such as fast food outlets or gambling shops, ecological information about the local population such as those derived from census and local authority data – e.g. deprivation indices)
- Information about the property (e.g. building age, type – detached, semi, terrace, flat), building floor, sale dates, value, energy performance records)

8.4.2   Newly commissioned geographic data

- Place-based resources (e.g. mapping tools and geographies, national geocoded resources and derived socio-economic and neighbourhood/land use indicators) and exposure estimates (e.g. air pollution and noise exposure estimates)
- Contextual meta-data defining the COVID-19 context in which participants were living and reporting outcomes and behaviours (specific across time and location)
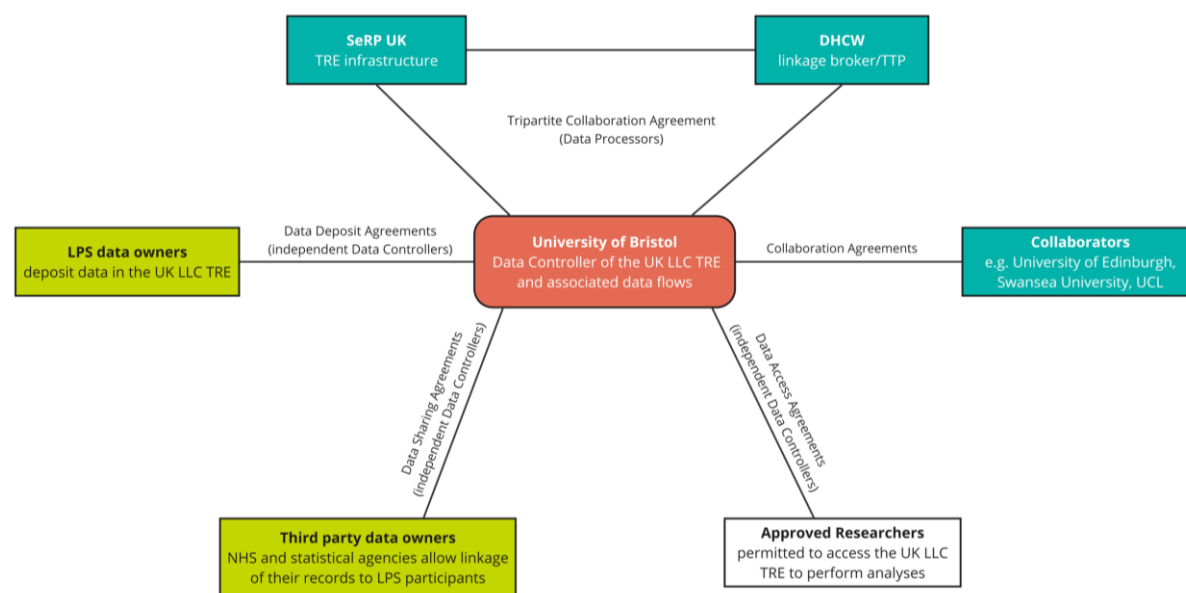- Data flow mapping.

9. Data Flows

9.1   Contractual framework

UK LLC's collaboration with LPS, data science infrastructure providers and coordinating bodies cuts through highly complex governance barriers which previously prevented progress. This breadth of agreement is unprecedented in the UK longitudinal and wider data science community (see figure 1). Data providers are engaged and supportive of UK LLC's approach because it provides a single point of contact and responsibility, thereby minimising data flows and provisioning costs.

---

[16] England is divided into nine regions, e.g. the South West; and along with the three devolved nations (Scotland, Wales, Northern Ireland) UK LLC provides 12 geographies that are identifiable to researchers in the TRE. These are used for reporting purposes and comparative policy assessments across the UK's four nations.

University of BRISTOL   THE UNIVERSITY of EDINBURGH   UK Research and Innovation   Economic and Social Research Council   Medical Research Council

Figure 1 Contractual arrangements in place with the University of Bristol for the TRE infrastructure, data processing and other collaborations

Each contributing LPS's institution enters into a Data Deposit Agreement (DDA) with the University of Bristol. This contract governs the flow of data, permitted uses and access mechanisms; this provides flexibility to enable the addition of new LPS and data flows into the UK LLC TRE.

Separately, University of Bristol has (or will have) Data Sharing Agreements (DSAs) with the owners of health, non-health administrative and place-based data[17]. These contracts establish the basis for data extract, the datasets shared and the mechanisms for project approvals and sharing of data with approved researchers (within the TRE).

The University of Bristol also has contracts with Swansea University (SeRP UK, owners of the UK LLC TRE infrastructure), the trusted third party (NHS Digital Health and Care Wales) for linkage and a number of collaborating institutes (including University of Edinburgh, UCL and Swansea University).

Data Access Agreements (DAAs) between the University of Bristol and an approved researcher's institution are signed before researchers are permitted access to the UK LLC TRE. The DAA binds users to the requirements of the UK LLC TRE and onwardly flows requirements agreed with third party data owners.

9.2   Data processing methodology

Figure 2 provides a high-level overview of the data processing methodology used to flow data into the UK LLC TRE. See **Appendix 2** for a more detailed flow diagram of UK LLC's data processing and legal basis.

---

[17] Data Sharing Agreements for place based and environmental data are often not necessary as these are free-to-access and in the public domain. Contracts are agreed where appropriate and where data owners impose usage restrictions.
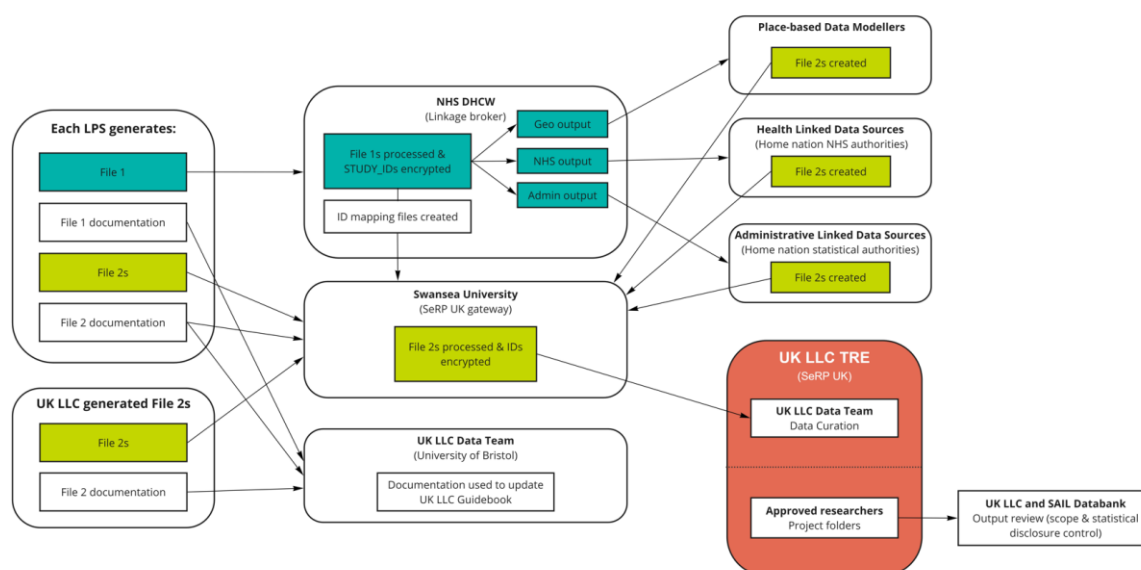
Figure 2 A high level overview of the processors and pathways used to send data to the UK LLC TRE – the University of Bristol, as Data Controller, is responsible for the majority of these data flows (at times, with other, independent, Data Controllers).

9.2.1   Split file approach

A 'split-file' anonymisation process developed by Swansea University for the SAIL Databank is used to securely transfer all datasets from contributing LPS into the UK LLC TRE. In this methodology, the LPS data are split by LPS data managers into a file of personal identifiers and an externally meaningless 'Link ID' (File 1). Separately, the attribute data are de-identified (direct and pseudo-identifiers are either dropped or transformed into less identifiable research variables) and indexed using the same 'Link ID' as the File 1 – this is called a File 2 (see figure 3).
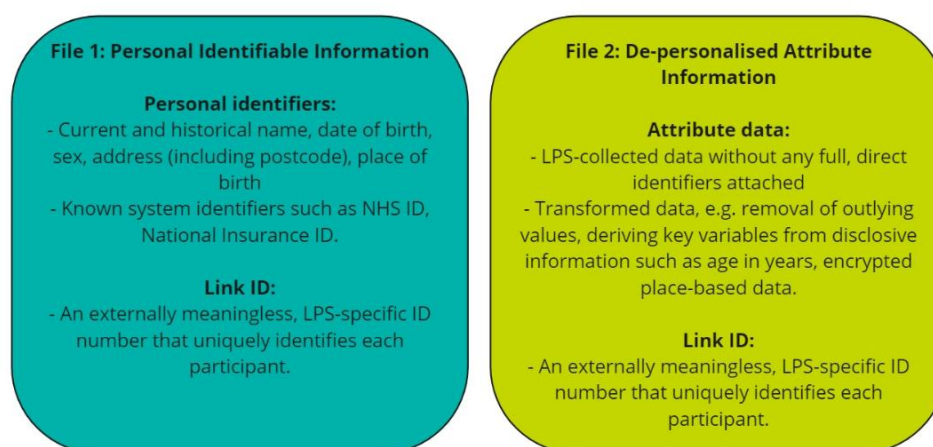
Figure 3 'Split-file' methodology: separating LPS participants' identifiers from their attribute data

Importantly, this process restricts the handling and management of LPS participants' personal identifiers to the contributing LPS, DHCW (the UK LLC's trusted third party for linkage) and the linked data owners and contracted place-based modellers. This means no one party or organisation can see personal identifiers and participant data (see Table 4). Data integration and management is conducted by a dedicated UK LLC Data Team within the UK LLC TRE.

Table 4 A table summarising the attributes of each group involved in processing data for ingest into the UK LLC TRE under the auspices of UK LLC (University of Bristol)

Attribute	Group					
	LPS	**DHCW**	**Linked data owners: NHS and statistical agencies**	**Place-based modellers**	**SeRP UK**	**UK LLC**
Organisation	Range of NHS and Universities	NHS Digital Health and Care Wales	NHS authorities and four nation statistical agencies, e.g. Office for National Statistics	e.g. City St George's, University of London	Swansea University	University of Bristol
Class of data accessed	Participant identifiers and permission flags; and de-identified attribute data	Participant identifiers and permission flags	Participant identifiers	Masked participant address data	De-identified attribute data*	De-identified attribute data
Processing activities	Provision of data for linkage, participant permission management and de-identification of data for research use	Record linkage and permission filtering	Health and non-health administrative record linkage and provision of de-identified data	Geo-coding and allocation of place-based data	Managing data ingest pipelines and encryption; TRE system administration	Integration of data, documentation, quality assessments, provision of data to researchers

*SeRP UK conduct encryption of some high-level place-based identifiers that have been determined to be de-identified by some data owners but are not permitted under UK LLC policy.

9.2.2   Linkage process

LPS Data Managers send File 1s to DHCW where they are split, processed, and Link IDs are encrypted. DHCW applies the preferences indicating which participants can be linked with which data sources. Participants' preferences are updated on a quarterly basis, enabling UK LLC to enact withdrawal/dissents. An ID mapping file containing the original Link ID (provided by the LPS) and the encrypted Link ID (created by DHCW) is created and sent to Swansea University (this enables Swansea to bring together the various attribute data for each participant). This process is overseen by both UK LLC and the LPS' institutions as independent Data Controllers.

DHCW then act as the linkage 'broker', facilitating linkages by sending reformatted and preference-filtered output files of unique identifiers and encrypted Link IDs to linked data owners (NHS Output and Admin Output) and contracted place-based modellers (Geo Output[18]) for processing. These organisations (see details below) create and send File 2s containing de-identified linked attribute data, including the encrypted Link ID, to Swansea University for re-encryption of the Link IDs and ingest into the UK LLC TRE. The File 2s sent to Swansea by LPS are processed in the same way.

The NHS Output is sent to the four NHS Authorities (NHS England, eDRIS/PHS/NRS in Scotland, NHS DHCW/SAIL in Wales and NHS NI BDO in Northern Ireland). The Admin Output is sent to the appropriate statistical agency, e.g. Office for National Statistics, who use optimal linkage methodology on behalf of government departments. The Geo Output is sent to the contracted place-based modellers.

LPS Data Managers send File 1 documentation to UK LLC and File 2 documentation to UK LLC and Swansea University.

9.2.3   A de-identified participant register and interactions across LPS

UK LLC will develop a de-identified participant index to enable researchers to understand which participants are active in multiple LPS and the relationship between different participants and also participants/properties. While this is not a new technology (e.g. many LPS work to ensure sample and uniqueness and family mappings), the cross-LPS approach and scale is novel.

UK LLC will need to address some new data challenges relating to the collation of multiple large LPS into a single environment. It will be necessary, using encrypted de-identified information from LPS (via DHCW) and data owners (e.g. NHS patient register), to define individual people and properties across the sum total of contributing LPS to create a de-identified participant register. This is crucial

---

[18] The Geo Output sent by DHCW to UK LLC's contracted place-based modellers contains only address data and so differs to the NHS Output and Admin Output, which also contain, e.g. names, dates of birth. The data sent includes participants' current and historical address data mixed in with 'masking addresses', i.e. addresses which are selected from whole UK look-up tables and therefore likely do not relate to LPS participants (at a 1:3 case/control ratio). These data are therefore Personal Data as all address data can be related to its occupant, but are not confidential as no information can be inferred from them. The place-based modellers link address data to AddressBase Premium (Ordnance survey data) to create a database of address coordinates, which they use to create derived variables, e.g. measuring the distance between home and the nearest GP surgery. The derived data are sent in File 2s to the UK LLC TRE. This process enables environmental exposures and characteristics to be quantified precisely, and linked to participants, without directly linking potentially disclosive location information with LPS/health data and therefore maintaining confidentiality.

because the joint analysis of data across multiple LPS will be conducted on a statistical assumption that the samples are independent, yet it is known that this is not the case.

The register will need to capture information across multiple dimensions (see figure 4):

- **Identifying unique individuals**: including the overlap of individuals across multiple LPS and where individual identity can be somewhat challenging to consistently determine (e.g. birth order amongst multiple births, where individuals have changed identities)

- **Annotating the roles/relationships between individuals**: within LPS it is common to track and annotate the roles and relationships between individuals, for example a mother-child pair, or that individuals are partners. This will supplement genetic 'pedigree' files that are generated to support genomic analysis and at a genetic level describe the extent of relatedness between individuals

- **Identifying clustering of individuals at household and other levels**: it will be important for some research questions to identify household composition and the temporal variation in this (e.g. students or renters returning to the family home). Household composition will be determined through using encrypted de-identified property reference data (unique property ID) provided by place-based modellers. This principle could be extended to other levels, for example neighbourhoods, schools, care homes or workplaces through de-identified information linked through routine records.
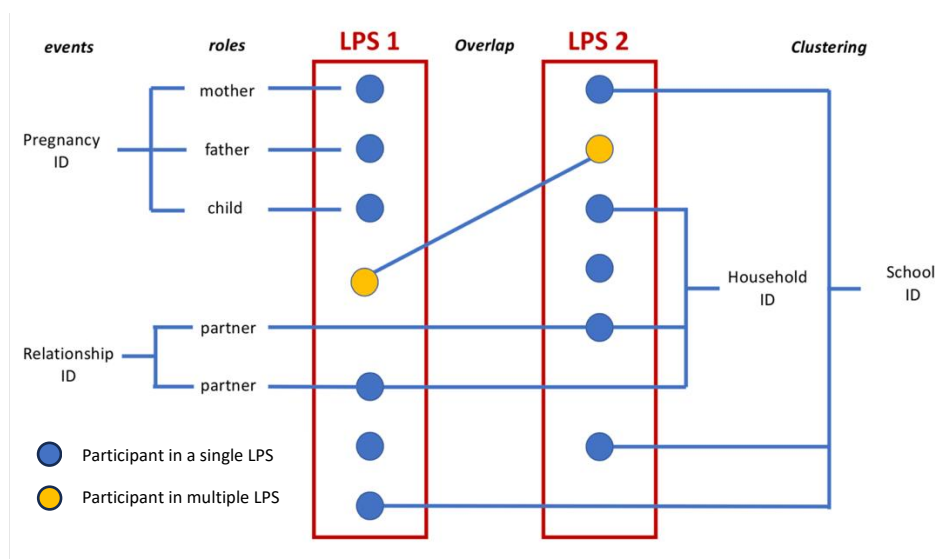
Figure 4 Illustrative examples of potential complexities of events, roles, overlaps and clusters within a matrix of LPS samples

The overlap of participant samples may introduce consent/permission ambiguities where permissions to link to and use routine records across different LPS are set in contrasting ways. UK LLC will work with LPS to establish the most effective methodology to overcome these challenges.

9.2.4   Data processing and cleaning

UK LLC takes a relatively light touch approach to data processing and cleaning, the rationale being related to expediency and because it is not possible to check individual values when making secondary use of records.

- The UK LLC TRE hosts data that have been collected by LPS and other data owners. The data owners undertake data quality checks and UK LLC cannot influence the quality of the original data or whether individuals can update their data, because University of Bristol is not the data owner

- University of Bristol UK LLC staff conduct quality and disclosure checks on data ingested to the UK LLC TRE

- A key objective of UK LLC is to co-locate multiple representations of similar data – sourced from LPS and routine records – to enable researchers to triangulate between different data sources, to assess quality and to address missingness through statistical processes (e.g. to address missing LPS data by using proxy information from routine records in multiple imputation models)

- UK LLC works with Health Data Research UK (HDR UK) and aligns with other TREs seeking to integrate complex health data across systems, nations and time. UK LLC will – where possible – adopt standardised approaches to shared data management challenges.

9.3   Outputs

Only anonymous aggregated information (e.g. statistical outputs, graphs and tables) is permitted to leave the UK LLC TRE. All analytical outputs are reviewed for disclosure risk by two independent members of the SAIL Databank Statistical Disclosure Control Team at Swansea University. SAIL Databank is ISO 27001 certified and accredited by UK Statistics Authority as a processor under the DEA. All users of the UK LLC TRE must hold valid ONS Accredited Researcher status; the training needed to gain this accreditation includes awareness of output disclosure risk. See the UK LLC Output Review Policy for further details.

9.4   Reproducible and reusable research

The UK LLC model promotes a culture of **reproducible** and **reusable** research where users generate discoverable research outputs (protocols, syntax, derived data) and deposit these back into the resource to inform other users' research, to drive efficiency and to enable follow-on or replication studies. UK LLC's Team Data Science approach is informed by the UK Reproducibility Network (ukrn.org) and the FAIR Guiding Principles for scientific data management and stewardship, which ensure all data and associated code and tools are Findable, Accessible, Interoperable and Reusable (FAIR).

UK LLC enables reproducible and reusable research because: (i) UK LLC **provides** reproducible research software in the TRE (e.g. Jupyter notebooks); (ii) UK LLC **requires** researchers to deposit reusable tools when they request data-out; and (iii) UK LLC is developing a new 'Data Output' API, **enforcing** upload of syntax into the 'Reusable Tool Library' (GitLab and GitHub) and the return of derived variables to the database for other researchers to apply to access. All project-datasets are archived to facilitate scientific challenge and reproducibility. Adopting these community standard open-source tools will facilitate cross-TRE working. See the UK LLC Reproducible and Reusable Research Policy for further details.

10.   Discovery and Access

The UK LLC TRE is a research database and therefore its primary objective is to be a resource for the wider UK research community. UK LLC, therefore, documents the resource, promotes its availability and considers applications to access the data.

10.1 Documenting the resource

The UK LLC denominator or sample is the sum product of all participants provided by the contributing LPS. The denominator is highly complex, dynamic and can be assessed at multiple levels, where:

1) The sample provided by each LPS will change as new participants join (e.g. babies being born in a birth cohort study, new individuals moving into an eligible property in a household panel study).

2) New LPS may join UK LLC and some LPS may withdraw.

3) The permission status for participants may change. This may be an objection to being included in the UK LLC TRE at all or an objection to one or more linkages, resulting in changes to data flows.

4) A new LPS joining UK LLC may result in participants in more than one LPS being identified as the same unique individual.

5) Participants' changing interactions with NHS and government services may result in their appearance in new datasets (e.g. an individual's first interaction with community mental health services will result in their first records being generated in these databases).

The denominator is fixed on a quarterly basis (linked to LPS refreshes of File 1s and the linkages based on these). Each quarter, UK LLC establishes a 'data freeze' of the UK LLC sample, which is critical to interpreting the resource, and provisions data to approved users based on this headline denominator – each Data Freeze is documented in UK LLC Guidebook and given a Digital Object Identifier (UK LLC sample — UK LLC Dataset Documentation). The data tables relating to each quarterly freeze are retained for archive purposes and scientific replication studies. The tables include numbers of participants from each LPS included in the UK LLC TRE and exclusion criteria; and linkage rates, including those excluded by permission status. UK LLC also periodically publishes a summary of the demographic characteristics of LPS participants in the UK LLC TRE, including their sex, gender, year of birth and ethnic group (UK LLC profile — UK LLC Dataset Documentation).

Some of the contributing LPS have population representative samples. In theory, this form of sampling approach enables the generation of accurate population inferences which can be generalised to the wider population. However, all LPS suffer from participant attrition and it is known that those continuing to participate tend to have different health and social characteristics to those lost to attrition. If uncontrolled, this can generate biased research findings and lead to situations where vulnerable sub-groups are excluded from the research and marginalised from the benefits of research (for example, groups whose health or social circumstances act as a barrier to continuing participation). To address and monitor this, UK LLC works with LPS to understand why individuals enrolled into a contributing LPS are not present in the UK LLC TRE and to identify ways to increase representativeness, e.g. use of section 251, to help policy makers and other users to draw accurate inferences from research findings.

UK LLC will also conduct statistical assessments of representation and use information from linked health and social records to develop statistical weightings to correct for the impact of attrition and changing population characteristics. Information from the linked records will be used to inform statistical procedures (inverse probability weighting, multiple imputation) to address missing data, to generate more accurate estimates.

This programme of work will be designed to fit to international research reporting guidelines (e.g. the RECORD statement[19] for reporting on the use of linked routine records in applied findings, the GUILD statement[20] for reporting on linkage process) and will be aligned with data quality assessment criteria being developed by stakeholders such as the ONS and increasingly used within UK government.

10.2 Discovery

UK LLC is designed to be a discoverable and accessible resource for public benefit research. UK LLC has developed a metadata catalogue and data selection portal called UK LLC Explore and more detailed data documentation and a user guide called UK LLC Guidebook. UK LLC Explore is populated with minimum metadata using existing automated metadata extraction software and is enriched through APIs to external sources, including HDR UK Gateway and the Catalogue of Mental Health Measures. This approach avoids duplication, minimises LPS burden and promotes interoperability and federation. The UK LLC TRE is promoted through the contributing LPS to their research users, through data science networks (including HDR UK), longitudinal study resources (such as the CLOSER consortium, www.closer.ac.uk) and funders.

10.3 Applying to access the UK LLC TRE

UK LLC has developed a bespoke online application management system called UK LLC Apply that handles the complexities of multiple data owners' terms and conditions and UK LLC's rigorous multi-stage application review process. The application process is based on the Five Safes and satisfies the needs of the contributing LPS; the third-party data owners; and includes a review by members of the public. Applicants need to demonstrate that they meet the Five Safes Framework criteria:

- **Safe people**: all researchers must hold valid ONS Accredited Researcher status, sign a UK LLC Data User Responsibilities Agreement (DURA), accept their project-specific data owners' terms and conditions and be part of an organisation with sufficient capacity to support good governance in research.

- **Safe project**: all researchers must demonstrate that their proposed research and data request is appropriate, ethical and likely to deliver clear public good that will be disseminated.

- **Safe setting**: all analyses take place within the secure UK LLC TRE.

---

[19] Benchimol EI, Smeeth L, Guttmann A, et al. The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10):e1001885

[20] Ruth Gilbert, Rosemary Lafferty, Gareth Hagger-Johnson, Katie Harron, Li-Chun Zhang, Peter Smith, Chris Dibben, Harvey Goldstein, GUILD: GUidance for Information about Linking Data sets, Journal of Public Health, Volume 40, Issue 1, March 2018, Pages 191–198, https://doi.org/10.1093/pubmed/fdx037

- **Safe data**: all organisations that provide data to the UK LLC TRE must ensure their data are de-identified. UK LLC performs disclosure checks throughout the data processing pipeline and only provisions the minimum required data to approved researchers for the approved purpose.

- **Safe outputs**: only anonymous population level outputs are permitted to leave the UK LLC TRE and these must go through statistical disclosure checks conducted by staff at SAIL Databank (Swansea University).

All applications are reviewed by LPS Data Access Committees (DACs) in Stage 3 against their data use criteria, so that LPS maintain control over how their data are used. UK LLC works with all third-party data owners to agree arrangements that work for each data owner, e.g. NHS England has delegated review to UK LLC against their data use criteria (Stage 3), while UK Statistics Authority requires all applications that include data accessed under the DEA legal gateway to be reviewed by their Research Accreditation Panel (Stage 4).

Appendix 3 details the UK LLC application process and the main stages are summarised below. All applications are listed on the publicly accessible UK LLC Data Use Register.

10.3.1   Stage 1

Stage 1 is an internal UK LLC review of expressions of interest (EoI) submitted by researchers. Incomplete or insufficiently justified EoIs are returned to applicants, who have the opportunity to revise and resubmit. At this stage UK LLC informs applicants of the ethics requirements. Applicants whose EoIs are approved by the UK LLC Applications Team are invited to submit a full application and data request form.

10.3.2   Stage 2

Stage 2 is an internal UK LLC review of full applications by UK LLC's Application Review Panel. This review assesses each application against the Five Safes Framework and checks that all requested fields are filled out appropriately. The Panel follows the ethical requirement checking process detailed in **Appendix 4** to assess whether ethical requirements have been met for each project. Only applicants that can provide evidence of favourable opinion project-specific Faculty REC or HRA REC, or who have completed the UK Statistics Authority's Ethics Self-Assessment Tool satisfactorily are permitted to move onto Stage 3. Any questions or concerns are sent back to the applicant(s) and resubmitted applications are re-reviewed. Applications that are declined are permitted to go through the appeal process once.

Note: ALL applications that include non-health administrative data will undergo further ethical review by the UKSA's Research Accreditation Panel at Stage 4.

10.3.3   Stage 3

The majority of applications are from researchers wishing to access LPS data linked with health (or other) data. However, a small number of researchers apply to access only LPS data. The review process at Stage 3 differs slightly for LPS only data, versus LPS and linked data.

Applications for **LPS data alone** are reviewed by:

1) The individual **LPS DACs** – each LPS DAC is responsible for reviewing the application against their LPS-specific application review framework. UK LLC collates and records the LPS' responses and shares this information, plus any stated conditions, with the applicant.

2) The **UK LLC Data Access Public Review Panel** – the panel focuses on the lay summary, assessing potential public good and public involvement sections of the application. Applicants have a five minute slot to present their research to the panel at an online meeting, with a short Q&A session thereafter.

Applications for **LPS data linked with health or other data** are reviewed by:

1) The individual **LPS DACs** (see above).

2) The **UK LLC Data Access Public Review Panel** (see above).

3) The **UK LLC Linked Data Review Panel** – the panel focuses on reviewing the application against the linked data owners' requirements (currently NHS England's).

Note: All applications that include requests for English and/or Welsh NHS data will have to demonstrate that the purpose will likely result in benefit to health and social care in England and Wales.

All approved applications that include non-health administrative data move to Stage 4 of the application review process. All approved applications that do not include non-health administrative data move to the completion of project governance stage (Data Access Agreement, Data User Responsibilities Agreement and Data Owners' Terms and Conditions).

10.3.4 Stage 4

All applications that include non-health administrative data will be sent to the UK Statistics Authority's Research Accreditation Panel (RAP) for ethical review and to give Government departments the opportunity to veto the inclusion of their data in applications.

10.3.5 Appeal process

Declined applications may go through the appeal process once. Applicants must address all comments, questions and criticisms fully for the appeal to be considered. Amended applications are then submitted for Stage 2 approval, and, if successful, progress to Stage 3 (and, if applicable, Stage 4).

10.3.6 DAA and DURA

Once an application is approved and before a researcher is allowed access to the UK LLC TRE, they must be covered by a Data Access Agreement (DAA) between their organisation and the University of Bristol. The DAA terms and conditions are non-negotiable. Where an individual is employed by multiple institutes, the DAA needs to be in place with the organisation sponsoring the research and taking accountability for it. Each applicant must also have completed a Data User Responsibilities Agreement (DURA), which makes clear a researcher's roles and responsibilities when working in the UK LLC TRE, and have accepted their project-specific Data Owners' Terms and Conditions. Lastly, if a researcher's organisation does not hold NHS England DSPT or ISO 27001 that covers the scope of the approved project, they must complete a System Level Security Policy (SLSP). Once the necessary documentation is completed, the UK LLC Data Team (University of Bristol) sets up approved users in the UK LLC TRE, provisions their approved data and manages their ongoing access.

11. Scientific Remit

The UK LLC research database has an explicit remit to support longitudinal public good research across the breadth of the sector, including supporting a broad range of both epidemiology and social science research. As such, its scientific remit is to establish and manage a generic research database open to all UK-based researchers.

The UK LLC's contributing LPS represent a broad spectrum of focal points and designs, each with an established research community. Examples include birth cohort studies aiming to understand health and social trajectories and outcomes at key life course transition points (such as early years, transition from education into the labour market, health ageing and independent living); occupational cohorts assessing workplace impacts on health and social outcomes; household panel studies assessing family dynamics and health and social outcomes and disease specific studies focused on conditions such as mental ill-health. **UK LLC is designed to service the scientific remit of all its contributing LPS and the patient groups/populations they include.**

11.1 Applied scientific programme

UK LLC is funded by UK Research and Innovation (through a Medical Research Council Partnership Award with Economic and Social Research Council co-funding). The primary goal of this funding, and the remit supported by UKRI following peer-review, is for UK LLC to be a generic research database supporting any public good longitudinal research.

Aligned with this, UK LLC funding is supported through a 'partnership' scheme where applicants describe the research they will conduct using the database. In our application the 24 Principal Investigators (of the original 24 contributing LPS) and leading epidemiologists and social scientists agreed to develop a programme of work that is indicative of UK LLC's use and potential. The focus of this partnership activity will be on the cross-cutting theme of health and social inequalities and assessment of outcomes in harder-to-reach groups and will include the following research topics:

- Long-term consequences of mental health conditions and disparities in care provision across socio-demographic groups

- Identifying modifiable risk factors of work-related mental ill-health to inform the Health and Safety Executive's national interventions

- Reducing chronic respiratory disease inequalities by identifying strategies that better predict, prevent and manage respiratory conditions

- Impact of air pollution on health

- Evaluating welfare policy impacts on health inequalities

- The role of common infections in neurodegenerative disease, examining the impacts of infections on long-term brain health

- Health and social consequences of early adversity as a key driver of adult health inequality

- Understanding changes in social mobility across generations

- Understanding when and why some individuals move onto less productive pathways

- Identifying the biological antecedents of multimorbidity

- Exemplar genome-wide and phenome-wide genetic association studies using electronic health records

- Intergenerational and family health and socio-genomics research

- Understanding how complex and interacting built environment factors influence modifiable risk factors for non-communicable diseases across childhood

- Identifying risk factors and patterns of interaction with health services for rare cardiovascular disease in ethnic minority groups

- Investigation of health inequalities in different socio-economic and demographic groups, in relation to the cost-of-living crisis.

This proposed research programme is **only indicative of the breadth of research that UK LLC could support as a generic research database** and is included to illustrate the breadth and interdisciplinary nature of the research potential. These projects will form only part of UK LLC's use and user base.

11.2 Ongoing COVID-19 scientific programme

UK LLC was established as a COVID-19 research platform. While the scientific remit will now extend to support generic public good longitudinal research, the platform remains available for ongoing COVID-19 research, including the investigation of medium to longer term outcomes and for future pandemic preparedness. The COVID-19 research programme covers the breadth of investigations relating to the pandemic (i.e. investigating the impact of COVID-19 as a virus, and the wider impact of COVID-19 mitigation measures on the population):

- Understand patterns and predictors of infection, (including re-infection) and disease outcomes (such as 'long covid'), and the role of antecedent and current health behaviours, health status, medication use, sociodemographic status, built and natural environmental factors, in impacting these outcomes.

- Explore population-level changes to physical and mental health – including hospital admission and mortality – and socio-economic status and social behaviours, in association with viral suppression measures and how these relate to changes in health behaviours and the extent to which behaviours and outcomes are returning, or not, to pre-pandemic patterns.

- Investigate the role of socio-economic and neighbourhood/environmental factors in determining population-level impacts to physical and mental health, to socio-economic outcomes, and to identify both groups at risk, and factors which offer resilience to adverse outcomes.

- Analyse changes in health care service use (using self-reported data and NHS health records) to determine if patterns in these have changed during the COVID-19 pandemic and post-pandemic period. Linked NHS records will also inform consideration of ongoing patterns of service use in relation to pre-pandemic health status and regular service interactions (e.g. screening, health reviews and routine service take up, such as annual seasonal flu vaccinations).

11.3 Methodological scientific programme

By pooling large numbers of LPS into a single generic research database, the UK LLC enables maximal use of the sample heterogeneity across these LPS. This will provide statistical power for

investigations of rare outcomes and of outcomes in population sub-groups. This has the potential to improve research equity through including harder-to-reach communities and to support the development of new analytical methodologies. To support this, the UK LLC research programme will include methodological research such as:

- Developing innovative statistical approaches to reducing bias in longitudinal research across pooled LPS data

- Understanding how to assess whether research using pooled LPS data is representative of the national (or regional/devolved) population and the extent to which findings are generalisable

- Developing mechanisms for UK LLC to provide a platform for replicating findings identified in other very large LPS (such as UK Biobank); to support international meta-analysis or replication studies (which does not involve access to individual level data from outside the UK); and to provide insights into the accuracy and completeness of whole population databases.

All methodological research will need to demonstrate potential public good, and where the use of NHS records are involved, likely benefits to the health and social care system. Methodological research is assessed through the same process as applied research and subject to the same policies and review criteria.

11.4 LPS functionality

UK LLC is established as a service function for the UK LPS community. Each LPS, once it has demonstrated it has the legal basis for this, will be able to access de-identified linked participant records for their LPS participants only, within the TRE. They will be permitted to conduct descriptive analysis of the data to inform research feasibility assessments and to document their LPS data and sample characteristics. This form of access is for data management purposes only and all LPS-led applied research requests will be subject to the standard UK LLC access rules and process.

12. **Strategic Development**

Over the next five years, UK LLC has the remit to integrate the breadth and depth of data across the UK's portfolio of LPS with national linked participant health, non-health administrative and environmental records, and make these data available to the research community to enable world class research.

UK LLC's four core objectives are:

1) **Maintain and consolidate UK LLC as a multidisciplinary data science asset of globally unique depth and breadth**. We will add new LPS data as LPS produce these and regularly update our linked records. We will expand LPS membership and increase our breadth of data, our sample size and diversity to support the wider LPS community and attract increasing numbers of research users across disciplines, UK Government departments and industry. We will build new functionality to enable multi-omics analyses and replicate key UK Biobank derived data and data formatting in our TRE to enable efficient cross-TRE replication analyses.

2) **Working with users we will co-produce user onboarding materials, training data and user support tools to establish code libraries** to enable transparent reproducible research and to iteratively improve ease of use and functionality.

3) **Working with LPS we will support a transition to working in TREs** by providing a linkage and governance 'expert hub' able to support LPS governance and fair processing communications (with the CLOSER LPS initiative). We will work with the funders and key national initiatives (e.g. Population Research UK and DARE UK) to model and implement pathways to sustainability, and we will facilitate cross-TRE interoperability by identifying and prioritising LPS-specific TRE needs, including enhancing the LPS' ability to be FAIR (Findable, Accessible, Interoperable and Reproducible).

4) **Working with public and participant contributors** to ensure the design and operation of UK LLC is co-developed with public input and that all activities of UK LLC are transparent to the public. To maximise the conditions for UK LLC to be perceived as trustworthy.

In summary, UK LLC will enable UK-wide analyses, allowing research results and effects on policies to be measured and compared. We will quickly be able to bring on new and existing LPS and linked data. As researchers use the resource, it will create an ever-increasing knowledge base to build and improve on. UK LLC will support a wide range of public good research and maintain a resource that is sufficiently responsive to enable the investigation of emerging policy questions and to meet future crises such as new pandemics, the impacts of climate change or economic shocks.