UK Longitudinal Linkage Collaboration

Population Health Sciences

Bristol Medical School

Canynge Hall

39 Whatley Road

Bristol BS8 2PS

UK Longitudinal Linkage Collaboration (UK LLC)

REPRODUCIBLE AND REUSABLE RESEARCH POLICY

PUBLIC

Version 1

15 August 2023

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: Policies

Page **1** of **9**

Policy number:	POL-DAT-014	Version:	V1
Author:	Katharine Evans, Governance & Policy Manager	Date:	20/07/2023
Authorised by:	Andy Boyd, Director	Date:	15/08/2023
Date published:	15/08/2023	Date to review:	15/08/2024
Permission to edit this policy must be provided by:	Senior Data Manager; Governance & Policy Manager		

Review History

Version:	Review Date:	Reviewed by:	Section(s) amended:	Authorised by:

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: Policies

Page **2** of **9**

Contents – UK LLC Reproducible and Reusable Research Policy

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: Policies

Page **3** of **9**

1. INTRODUCTION

1.1 Background

The UK Longitudinal Linkage Collaboration (UK LLC) organisation is led by the University of Bristol (UoB) and University of Edinburgh (UoE) and operated in collaboration with Swansea University, University of Leicester and UCL. The UK LLC manages the collation, curation and access to de-identified data about Longitudinal Population Study (LPS) participants held in the UK LLC Trusted Research Environment (TRE).

The UK LLC model promotes a culture of **reproducible** and **reusable** research where users generate discoverable research outputs (e.g. research protocols, syntax, derived data, code lists, documentation) 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).

> **All UK LLC research users must deposit derived data, syntax, documentation and other reusable research outputs into the UK LLC resource.**

1.2 Purpose

This policy sets out UK LLC's requirements for the deposit of reusable research outputs for future researchers to use. This policy will help UK LLC to be transparent in its operations and to enable reproducible and efficient research for public good.

This policy will be reviewed to respond to any changes in the UK LLC risk assessment or risk treatment plan and at least annually.

2. SCOPE

This policy applies to all approved researchers and the UK LLC Data Team working on approved projects in the UK LLC TRE.

This policy applies to the following reusable research outputs:

1. Derived (or returned) data.
2. Documentation, including protocols, syntax and code lists.

This policy does **not** apply to statistical or analytical outputs (i.e. research findings, including tables and graphs of anonymous population-level aggregate data). The process researchers must follow to export analytical outputs from the UK LLC TRE is covered in the [UK LLC Output Review Policy](#).

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: [Policies](#)

Page **4** of **9**

3. ABBREVIATIONS

LPS   Longitudinal Population Study

TRE   Trusted Research Environment

UK LLC   UK Longitudinal Linkage Collaboration

UoB   University of Bristol

UoE   University of Edinburgh

4. DEFINITIONS

Application – all research use within the UK LLC TRE is embedded within an approved application. See the UK LLC Data Access and Acceptable Use Policy for further details.

Project – an application is referred to as a project at the point where the application is approved and data are prepared for provisioning within the UK LLC TRE.

Reusable research output – any output comprising derived data or documentation, e.g. data processing/analysis syntax, user guide material, code lists (groups of medical codes used to define a particular condition or disease in electronic health records) that is:

- necessary for other researchers to understand and reproduce the research and/or
- useful to others for reuse within their own project and/or
- necessary for external stakeholders (including participants and the public) to understand how data are being used in the UK LLC TRE.

The **inter-relationship** between **applications, projects** and **reusable research outputs** is defined in figure 1.

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: Policies
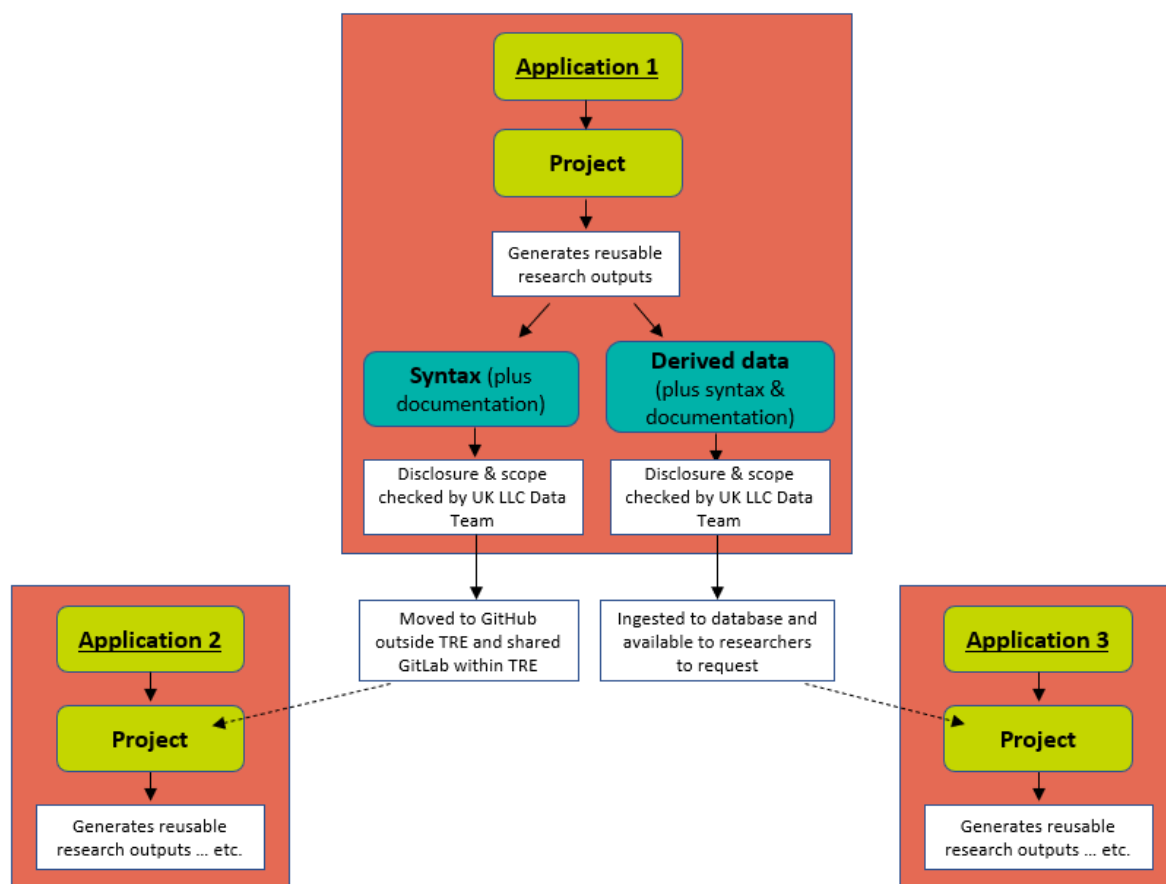
Page **5** of **9**

Figure 1 The inter-relationship between applications, projects and reusable research outputs in the UK LLC TRE
Note: the dashed arrows illustrate the re-use of research outputs from one project to another (with prior approval(s) by LPS and/or UK LLC Linked Data Review Panel in the case of derived data)

5. ROLES AND RESPONSIBILITIES

Who	What & Why
UK LLC Applications Team (UoE)	1. Sets up the GitHub repository for each project, setting access to private. 2. Coordinates review by LPS Data Access Committees and UK LLC Linked Data Review Panel of requests for derived data. 3. Hands over approved data requests to Data Team for provisioning.
UK LLC Data Team (UoB)	1. Triages all output review requests and disclosure and scope checks all reusable research outputs. 2. Sets permissions for GitLab repositories within the TRE and makes approved syntax/documentation available to all TRE users in the shared GitLab.

Who	What & Why
	3. Runs the 'derived data process', which includes disclosure checking, adding data to the TRE database and making this selectable to other researchers.
Researchers	1. Label file-out requests for Git repositories to indicate they contain reusable syntax and documentation.
	2. Make GitHub repository publicly accessible as soon as they have added first content.
	3. Document all reusable research outputs appropriately as per the UK LLC TRE User Guide.

6. REUSABLE RESEARCH OUTPUTS

6.1 Key points

- Researchers must make their reusable research outputs **accessible and understandable to future users**

- **Derived data cannot leave the UK LLC TRE.** Following a disclosure risk assessment, returned data files are written as a table to the UK LLC database and are then available for request by other researchers (following the standard UK LLC application or amendment process)

- **Documentation, syntax and code lists** can be held **outside the UK LLC TRE** following a disclosure risk assessment

- **All researchers must follow the** UK LLC Naming of Projects and Data-related Outputs Policy for labelling reusable research outputs**.** This ensures consistent naming conventions, enabling discovery and attribution

- **The UK LLC system ensures that researchers receive full attribution for the reusable content they generate**

- UK LLC **cannot guarantee the quality of reusable research outputs**. This applies to all outputs including documentation, syntax and derived data. It is the **responsibility of the researcher** to check reusable research outputs prior to inclusion in analyses and subsequent interpretation.

6.2 UK LLC Git repositories

To facilitate reproducible and reusable research, UK LLC maintains Git repositories where, for each project, researchers add cleaned and documented reusable research outputs. There are **three distinct UK LLC Git spaces**:

1. Each approved project has its own **Project GitLab within the UK LLC TRE**. This space is only accessible to the researchers named on the project and to UK LLC Data Team staff.

2. There is a **Shared GitLab within the UK LLC TRE** for sharing code between projects. This space is accessible to all researchers active in the UK LLC TRE and to UK LLC Data Team staff.

3. Each approved project has its own **Project GitHub** (within the UK LLC GitHub) **outside the UK LLC TRE**. Following review of documentation, syntax or code lists, researchers add their outputs to the UK LLC GitHub and set the visibility to 'Public', thereby making them publicly accessible.

This document was printed on Fri, 15 Sep 2023 and may be out of date. The current version is found here: Policies

Page **7** of **9**

The table below summarises the read/write/access permissions for these three UK LLC Git spaces that are applicable to each potential user group.

User group	GitLab – within UK LLC TRE		GitHub – Public
	Project GitLab	**Shared GitLab**	**Project GitHub**
UK LLC Data Team	Read/Write	Read/Write	Read/Write
Researcher within their project space	Read/Write	Read only	Read/Write
All researchers	No access	Read only	Read only
Public	No access	No access	Read only

6.3   Derived data

- All files must be named following the UK LLC Naming of Projects and Data-related Outputs Policy

- The file path of the datafiles containing derived outputs must be shared with the UK LLC Data Team via support@ukllc.ac.uk. This email must also supply the project number and GitLab repository containing the accompanying syntax and documentation used to create and document the derived data – **derived data without syntax and documentation will NOT be processed**

- The contents of the data files and accompanying syntax and documentation **must not increase identifiability risk of UK LLC participants** and should contain the following metadata:

  o   Name of dataset

  o   Dataset description

  o   Full name of dataset creator

  o   Up to six keywords

  o   Whether the dataset contains sensitive information e.g. mental/sexual health

  o   Data sources (this only needs to include the names of the LPS that have contributed data and/or the linked data owners (e.g. NHS England), rather than the names of the individual datasets)

  o   Value and variable labels

- **Failure to make derived data available will risk future access to the resource.**

6.4   Syntax and documentation

- All files must be named following the UK LLC Naming of Projects and Data-related Outputs Policy

- All output review requests for syntax and documentation must be submitted via the SeRP Portal and must be marked with '**For sharing via GitLab/GitHub**'

- The syntax and contents of the accompanying documentation **must not increase identifiability risk of UK LLC participants**

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

Public

- Documentation must be sufficiently detailed for other researchers to understand the syntax. Syntax should follow good coding practices, which includes meaningful naming, appropriate level of comments to give context and well-structured

- All approved reusable research outputs should be uploaded by researchers to their Project GitHub on the UK LLC GitHub. As soon as content is added, researchers should switch the setting of their GitHub from 'Private' to 'Public'. The UK LLC Data Team will check that each project's GitHub corresponds to the File Out Log

- **Failure to make syntax and documentation available will risk future access to the resource.**

7. RELATED DOCUMENTS

- POL-ISM-001 UK LLC Information Security Policy

- POL-ISM-003 UK LLC Data Access and Acceptable Use Policy

- POL-DAT-006 UK LLC Naming of Data and Data-related Outputs Policy

- POL-ISM-012 UK LLC Output Review Policy

- DOC-DAT-040 UK LLC TRE User Guide

- SOP-DAT-020 Requests to Add/Export Files To/From the UK LLC TRE

- SOP-DAT-022 Depositing Reusable Data

- SOP-DAT-023 GitLab in the TRE – curation and reproducibility.