The UK Longitudinal Linkage Collaboration (UK LLC)
Linking Data for the Longitudinal Research Community

Robin Flaig
University of Edinburgh
January 2022
A collaboration to benefit public health

UK LLC led by the Universities of Edinburgh and Bristol

In collaboration with

A work package in the COVID-19 Longitudinal Health & Wellbeing National Core Study
UK Longitudinal Linkage Collaboration

A four-nations infrastructure for secure cross-cohort analysis

Align complex and divergent governance frameworks into a streamlined and predictable linkage/access route

Develop a sustainable four-nations data science infrastructure

A new way of working for record linkage

Watch our animation: https://youtu.be/zFkvpKD3jvs

UK LLC Design Decisions:
- Interdisciplinary
- Four nations approach
- Pre-pandemic & pandemic data
- Trusted Research Environment
- Five Safes access model
- De-personalised data only
- Distributed governance
- Transparent in its operations
- Public Involvement programme
A globally unique cross-cutting resource

20+ Longitudinal Population Studies
- COVID-19 collections (Questionnaires, Serology)
- SES & demographic
- Baseline physical & mental health
- Baseline family, SES and life-course indicators
- Genetics

NHS COVID-19 datasets
- GP Data
- Vaccinations
- COVID-19 test data (Pillar 1-3)
- Accident & Emergency
- CHESS

Wider NHS datasets
- Hospital Inpatient data
- Cancer & mortality registers
- Community Mental Health (MHSDS, IAPT)
- Prescribing data

Administrative records
- HMRC
- DWP
- DfE

Environmental & neighbourhood
- Air pollution
- Noise
- Greenspace
- SES & demographic
- Service provision

Negotiating data access
Being modelled
The Minimum Viable Product

20+ longitudinal studies
Novel NHS Digital data pipeline
Public contributors & application panel
Nine applications through beta-testing phase

Accessible data from:

ALSPAC: Avon Longitudinal Study of Parents and Children
1970 British Cohort Study
Born in Bradford
ELSA: English Longitudinal Study of Ageing
EPIC-Norfolk: The European Prospective Investigation into Cancer (EPIC) Norfolk Study
EXCEED: Extended Cohort for E-health, Environment & DNA
The Fenland Study
Generation Scotland
GLAD: Genetic Links to Anxiety and Depression Study
The Millennium Cohort Study
NCDS58: 1958 National Child Development Study

Next Steps
NICOLA: Northern Ireland Cohort for the Longitudinal Study of Ageing
NSHD46: MRC National Survey of Health and Development Cohort/1946 Birth Cohort
TRACK-COVID Study
TwinsUK
Understanding Society - the UK Household Longitudinal Study

UKLLC_CLOSERConference_January 2022
Public Involvement

4 touchpoints for involving study participants and members of the public:

- Strategic Advisory Committee
- Data Access Public Review Panel
- Public Advisory Group
- Public Involvement Network

- Strategy and future collaborations
- Involve in data access process
- Assess remit and scope of public involvement
- Public involvement policy development
- Mutual ways of working
- Access wider public involvement networks
- Acknowledge and communicate public involvement
- Measure impact
- Training and guidance
UK LLC Trusted Research Environment (TRE)

Split File & Trusted Third Party Data Processing

Study

Personal Identifiers LLC ID

De-identified NHS records & LLC ID

Data Linkage & Extraction

SeRP Linkage Brokerage

Study ID & LLC ID

ID Swap & Encryption

NHS Records & eLLC ID

Study Data & eLLC ID

Functionally Anonymous TRE
UK LLC: efficient and predictable data access for analysts

Data linkage made easy and predictable

UK LLC ‘Five Safes’ Application Framework

- UK LLC distributes applications to contributing studies
- UK LLC has delegated review authority from NHS Digital
- COVID-19 research only, UK based analysts
- Public benefit with public review
- Requirement for reproduceable research

Apply for access via portal at the HDRUK Gateway (https://www.healthdatagateway.org/)

- All researchers need to be ONS Approved Researchers & trained
- Data Access Agreement
- Light touch research user IT security framework as heavy lifting done at the UKLLC (ISO27001)
User experience & reproduceable research

All processing within the TRE

NHS Records & eLLC ID

Study Data & eLLC ID

Permissions

Project Data View

Code Lists & Code Mappings

Script Library

Project Analysis Space (Jupyter & GitHub)

Code Lists & Derived variables

SeRP

Output Disclosure Checks

Research Outputs

Project Jupyter & GitHub

Safe Outputs

Reproduceable Research
How we built a new research resource

JUN '20
- Wellcome COVID-19 questionnaire

AUG '20
- Scoping for National Core Studies
- Name, protocol, CAG Approval
- 3 x part-time staff working up

OCT '20
- LPS Agreement
- Participation, application and data provision
- Building blocks in place to deliver minimum viable product
- Website build
- Public Involvement

APR '21
- NHS Digital Application
- 6 months funding
- 18 month funding extension
- Deliver minimum viable product
- Flexibility and learning from the Vanguard LPS experience was crucial
Main Challenges

- Short time-frame and short-term funding structure
- Working across 20+ different LPS and across four nations - all with different requirements
- Data processing complexities – split data method = no one can see the whole process
- Applicants unable to select data as we couldn’t see it = made access process challenging
- Fast-paced evolving landscape posed challenges, required new innovations to be created and tested
- Lack of capacity at various points in the pipeline & volume of COVID-19 research required and underway
- Challenging to find the balance with communications about data to the public – clarity vs accuracy
What we learned

**Good communication is essential**
Regularly talk with funders, researchers, data providers, LPS, public contributors.
Working closely with the public on planned use of data, and how to communicate it is crucial.

**Don’t make assumptions**
Many LPS are set up in similar ways but not the same – there are many differences.

**Good Documentation**
Create process flows and SOPs that are tested and reviewed regularly

**Test and Learn**
The creation new systems is an iterative process.
Be willing to change the process as a result of testing, if things don’t work don’t keep them.

**Don’t reinvent the wheel**
Talk to other projects who face similar challenges and learn from their experience – DPUK, BHF Data Science Centre and SAIL.

**Think ‘future’**
Even with short-term funding, try to future-proof what you set up in the event longer-term funding might be available.
Future plans
(above consolidating minimum viable product)

User experience & reproducible research
Enhance with administrative records
Efficient and predictable data access for analysts