

# Metadata management in the real world: CLOSER Data Audit

---

Jon Johnson – Technical Lead, CLOSER

20 September 2023



# Housekeeping

- Please keep your mic and camera off unless speaking
- If you would like to contribute to the discussion or ask a question, please use the chat box function.
- This session is being recorded. Video and slides will be circulated as soon as possible post-event.
- Live captions are available under 'More options > Language and speech'
- Please help us keep improving our events by completing our short feedback form (link will be posted in the chat box).



# Overview

- About CLOSER
- Commissioning
- Survey and Follow-up
- Data Access
- Data Linkage
- Barriers
- Report Recommendations



# About CLOSER

The UK's partnership of social and biomedical longitudinal population studies, the UK Data Service and The British Library

UKRI ESRC-funded, based at UCL Social Research Institute

Mission = increase visibility, use and impact of longitudinal population studies, data and research

- Facilitate interdisciplinary research and encourage cross-study analysis
- Provide resources and training opportunities for researchers
- Mobilise evidence and insights to policymakers
- Represent LPS community on national and international stage



# CLOSER Discovery

**Search, explore** and  
**assess** data from multiple  
UK longitudinal population  
studies

[www.discovery.closer.ac.uk](http://www.discovery.closer.ac.uk)

**337**  
questionnaires



**131**  
data sweeps



**49,683**  
survey questions



**11**  
longitudinal  
population studies



**4,515**  
COVID-19 survey  
questions



**18,888**  
COVID-19 variables



# CLOSER resources

## Learning

Information and resources to help you better understand longitudinal population studies and how to use the data



## Evidence

Discover evidence and insights from UK longitudinal population studies on a range of topics including mental health and child obesity



## Innovations

Our data harmonisation and linkage projects help link data from our studies to administrative records and make the data more comparable.



## Society

Find out more about the historical and political backdrop to the lives of study participants using our resource - Our Changing Society



# CLOSER partner studies



# Data Audit Objectives

The purpose of the data audit was to provide a **baseline of information** to enable the ESRC and MRC to understand:

- the current state of the many activities that Longitudinal Population Studies (LPS) carry out;
- to better understand where gaps exist;
- where improvements may be possible and;
- to plan for future data collection and support for UK LPS.





# Survey and follow up

Survey with 77 questions covering:

- Study Information
- Data Collection
- Samples
- Data Linkage
- Data Management
- Governance
- Privacy and Disclosure

Follow up interviews with 8 studies



# Study profiles

	ESRC		MRC	
	No. Studies	Average No. Participants (baseline)	No. Studies	Average No. Participants (baseline)
Small	1	8,478	24	5,439
Medium	6	14,202	3	16,153*
Large	4	229,295	7	294,513**
Cross-sectional	4	6,901	-	-
Total	15	69,232	34	65,900

\*exc. \*\*Our Future Health & \*Early Life Cohort Feasibility Study



# Datasets being accessed

- Evidence base for understanding LPS data usage
- 36% growth across all LPS in last 5 years
- Growth in data usage occurs across all studies

Year	No. of data requests
2017	18430
2018	20210
2019	23998
2020	24222
2021	24963



# Where data are being accessed

	Data requests	Studies
Directly from study	65.7%	41
- British Election Study	53.7%	1
- Biobank	9.8%	1
- Other	4.0%	39
UKDS (EUL/SL)	25.9%	15
UKDS (Secure Lab)	0.4%	9
SERP Platforms (DPUK, SAIL, UKLLC)	3.1%	11
Other	3.1%	8

Total > 51 as Data accessed from multiple places



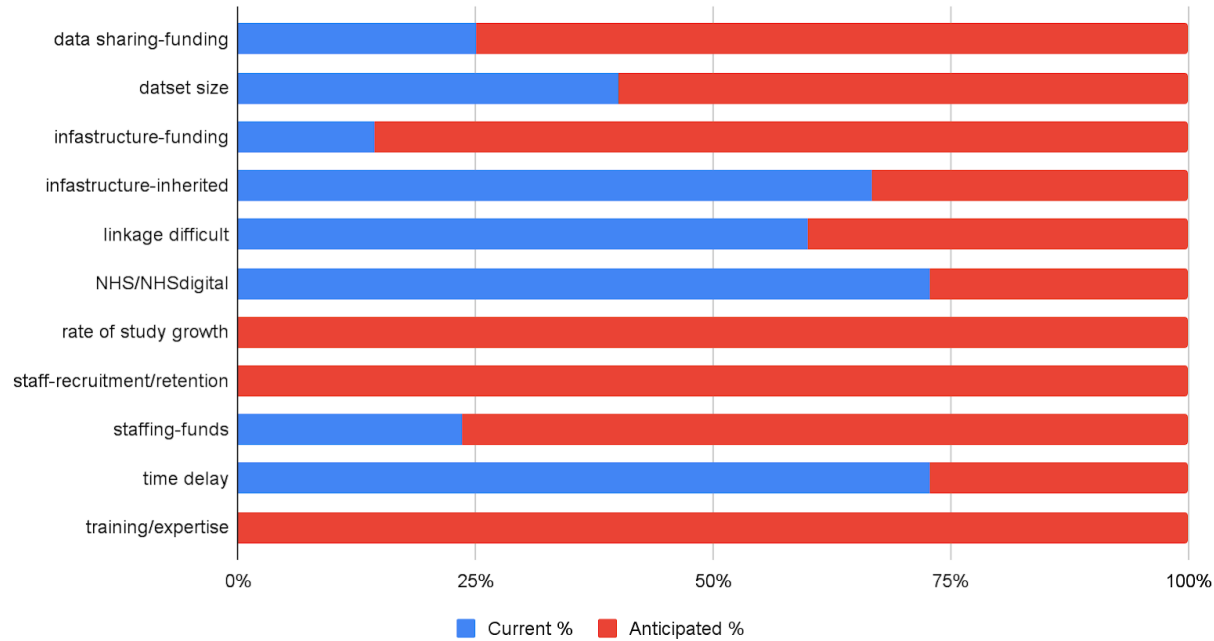
# Data linkages

- Doubling in planned and possible linkages
- Maintaining and refreshing existing linkages
- New applications in a changing linkage environment
- Multiple providers with different requirements

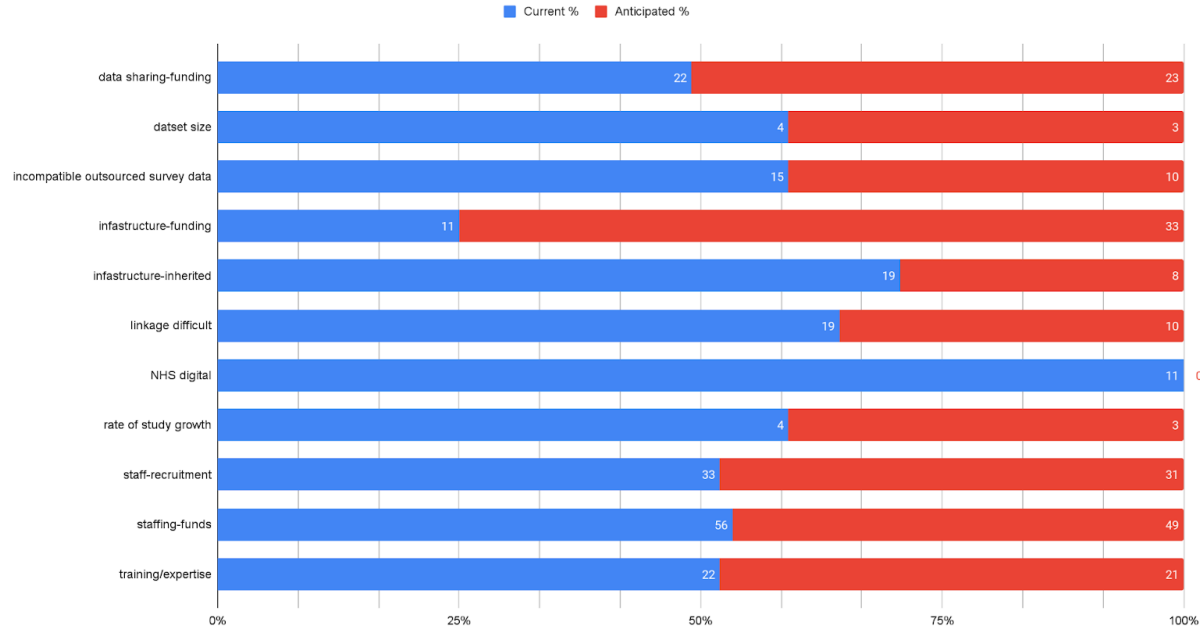
	Current	Planned	Possible
Health	47	24	16
Education	13	25	9
Employment & Income	8	6	27
Criminality	2	3	6
Spatial	49	17	3
Consumer	4	0	0
Other	12	1	6



# Barriers to data sharing



# Barriers to data management



# Metadata coverage

Figure A: Topical coverage by vocabulary

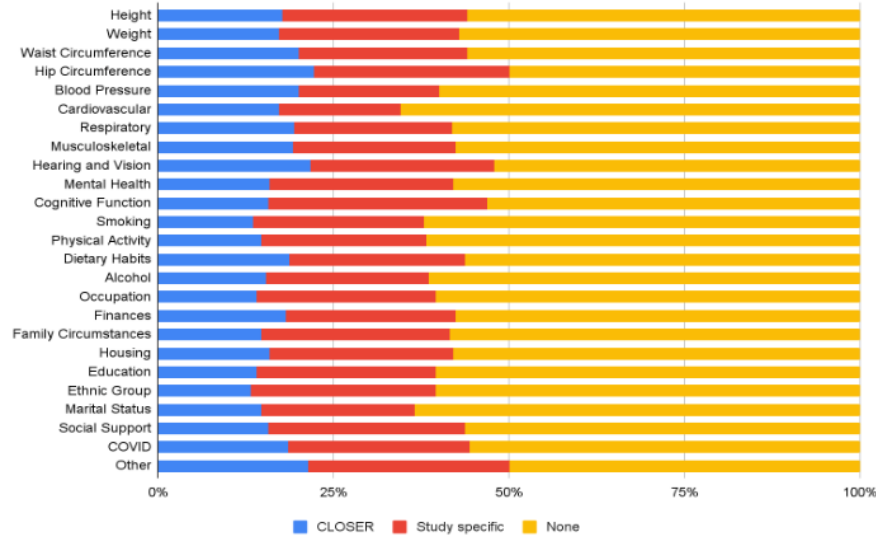


Figure B: Discovery by data usage

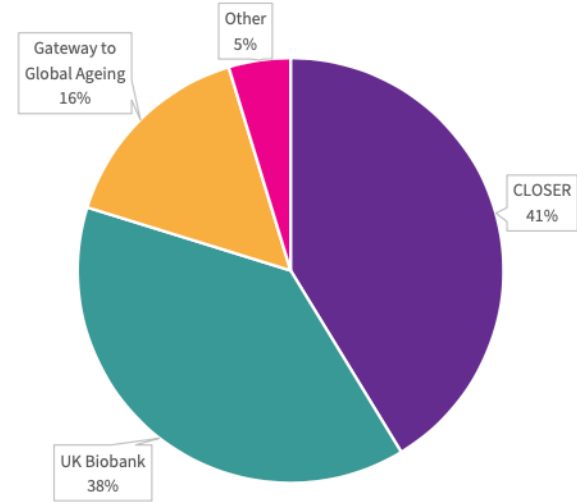


Figure A - % of studies. Figure B - % of all data requests





# Recommendations - Data sharing (1)

- *Coordinate and align access and governance policy with current and planned data holders outside of health and the ONS to meet the needs of LPS and to inform the technical and governance architecture of TREs.*
- *Develop a shared output disclosure strategy and processes.*
- *Promote the contribution of LPS in assessing the quality of administrative data.*
- *Support studies to improve and modernise data management processes and infrastructures, including incentives for using shared resources where possible.*
- *ESRC and MRC should investigate the development of guidance on best practice for data access and governance to ensure a consistent set of information is made publicly available.*



# Recommendations - Data sharing (2)

- Establish a network or community of practice to share experience, develop skills and best practice guidance. This would be welcomed by many, but especially the smaller studies.
- Investigate mechanisms for providing data to a single data provider, who would then serve as the 'canonical source' for data at other infrastructures.
- *Put in place better mechanisms for studies to interact with departments and thereby reduce the lag and burden of data linkage application and delivery (in light of the audit's findings that indicate a near doubling of planned linkage ..).*
- *Establish a scheme that would extend the common safe researcher training programme and accreditation aligned between ONS, UKDS and HMRC to data held in other TREs.*



# Recommendations - Samples

- Encourage studies to register samples at UKCRC to enhance the discoverability of samples in UK LPS.
- *ESRC and MRC should investigate the development of guidance on best practice for sample governance and sample depletion policy to ensure that a consistent set of information is made publicly available.*



# Recommendations - Metadata

- *Establish a metadata office to develop and oversee a metadata strategy which sets out clear guidance for studies and data dissemination infrastructures on minimum content and metadata format.*
- Create a single point of entry for these investments in the short-term to provide basic catalogue information and guide potential users to the 'right place' along the lines of an enhanced MRC cohort directory.
- *Develop a high-level catalogue, regularly updated and maintained as a central resource for high-level information about the coverage of topics, data and study profile of UK LPS.*
- *Investigate the possibility for federated discovery and interoperability across existing infrastructures.*





The home of longitudinal research

# Thank you

**Scan the QR code to sign up for our newsletters:**

- Longitudinal news
- Blog alerts
- Twitter highlights
- In the Loop



 @closer\_uk  @closer-uk  /closer-uk  closer.ac.uk

---