

Introduction to synthetic data for longitudinal data managers

Cristina Magder

Data Collections Development Manager,
UK Data Service

14 October 2024



Today's programme

- Introduction to Skills Development for Managing Longitudinal Data for Sharing Project.
- Introduction to Synthetic Data:
 - What is synthetic data?
 - What types of synthetic data are there?
 - For what purposes can synthetic data be used?
 - How can synthetic data be generated?
- [Q&A](#)
- Break (15 mins).
- Live coding demonstration.
- [Q&A](#).



What are the key challenges you face when managing longitudinal data for data sharing? List as many as you would like!

20 responses



Login to edit this Mentimeter

What are the key ways synthetic data could transform the management and sharing of longitudinal data?



21 responses



Login to edit this Menti

What are some key aspects you would like to learn of/about by the end of this workshop?

19 responses

How it replaces real data?

How to do it (and in R if possible)

The method of generating synthetic data

Best practices for synthetic data creation and sharing

how to build this into my workflow

Can it replace real data?

How to train researchers when to use it?

best practices

Existing validation on questions in this field. Thinking in quality and privacy

Tools and other courses, networking for

[Login to edit this Menti](#)

UK Data Service (UKDS)



Hosts UK's largest collection of social, economic and population research data.



Provide users with access, support, guidance and training to facilitate high quality social and economic research and education.



A partnership between UK Data Archive at University of Essex, Cathie Marsh Institute for Social Research at University of Manchester, Jisc, UK, EDINA from University of Edinburgh, and the Department of Information Studies and Centre for Advanced Spatial Analysis at University College London.



Support the development of best practices for data preservation and sharing standards.



ukdataservice.ac.uk

Stats about UKDS

~ **9,700** data collections (1169 cohort and longitudinal studies).

~ **250** new data collections and new editions added each year.

~ **48,000** registered users.

~ **130,000** downloads worldwide p.a.



UKDS data collections are accessed every 6 minutes



24 x 7 x 365

Population Research UK

Population Research UK (PRUK) is a national resource designed to maximise the use of, and benefits from, UK longitudinal population studies across social, economic and biomedical science.

PRUK is funded by the UK Research and Innovation (UKRI) Infrastructure Fund.



**UK Research
and Innovation**

Skills Development for Managing Longitudinal Data for Sharing

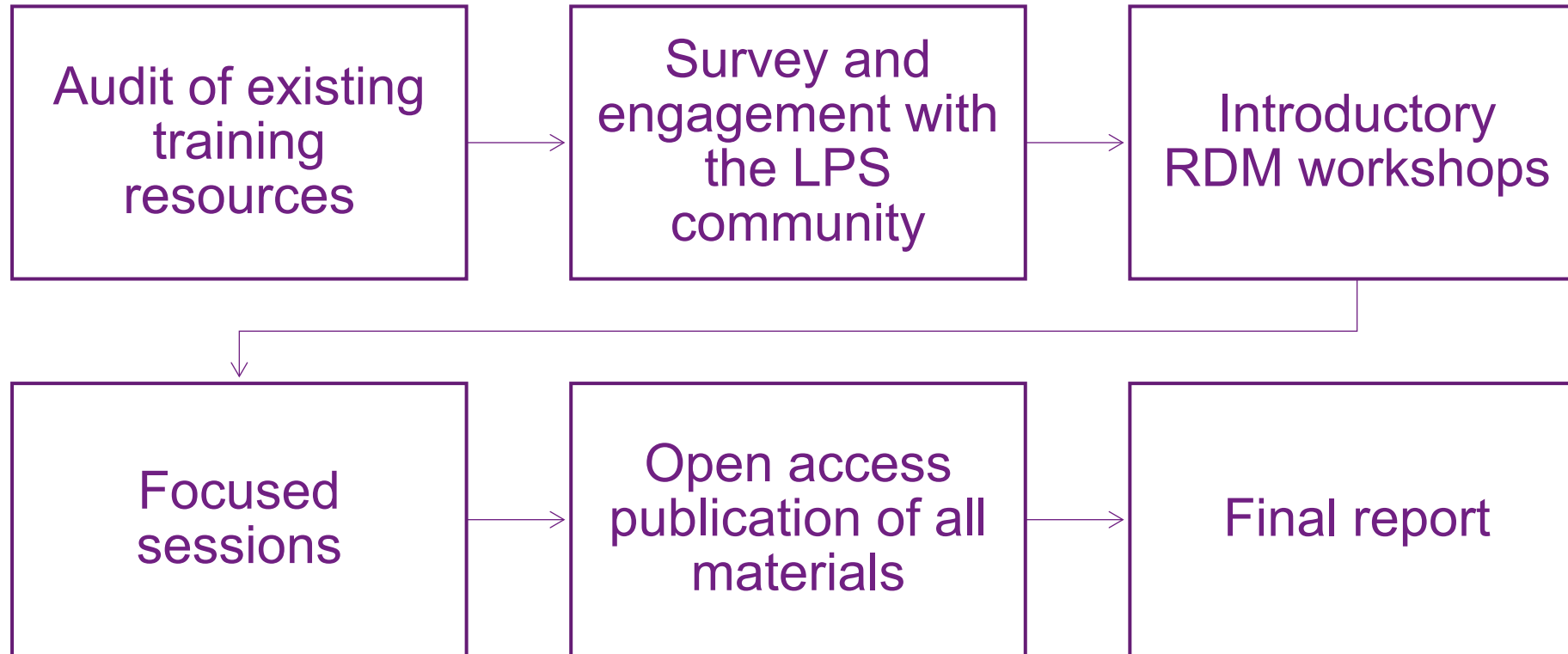
Skills development has been identified as a priority need within the wider Longitudinal Population Studies community.

UKDS has been commissioned to undertake an investigation into data management provision with a particular need to focus on wider data sharing via standardised training.

Visit the [project webpage](#) for additional information.



Project work packages



Introductory course

Enhancing longitudinal data sharing: Data management strategies and best practices

- Foundational research data management concepts.
- Ethical and legal considerations in handling longitudinal research data.
- Strategies and best practices for organising and formatting data.
- Effective methods and practices for documenting data.
- Techniques and guidelines for anonymising data.
- Data sharing strategies overview and recommendations.

All materials will be published by the end of November 2024, keep an eye on the [project webpage](#).

Why synthetic data?



SIPHER case study: synthetic population



SIPHER Consortium. (2023, June 15). *What is a synthetic population? SIPHER's synthetic data explained* [Video]. YouTube. https://www.youtube.com/watch?v=CkiORY7GSLc&t=49s&ab_channel=SPHSUUniversityofGlasgow
Visit the [SIPHER website](#).

SIPHER case study

The SIPHER Consortium is addressing health inequalities by linking social determinants (e.g. income, housing, education) with health outcomes.

Using spatial microsimulation, the SIPHER project created a synthetic population, reflecting the adult population in Great Britain based on Understanding Society and UK Census data. It offers a "digital twin" of these populations, supporting innovative research while protecting privacy.

- Enables simulation models to assess the impact of policies on different population subgroups.
- Helps policymakers evaluate public health interventions and outcomes across regions.

Data collection available through the UKDS [**SIPHER Synthetic Population for Individuals in Great Britain, 2019-2021 \(SN9277\)**](#) alongside the [replication package](#).

Read the [full case study](#) and explore the [interactive dashboard](#).

Thank you!

datasharing@ukdataservice.ac.uk

beta.ukdataservice.ac.uk/help