

Challenges of creating and documenting derived variables in longitudinal studies

Centre for Longitudinal Studies (CLS)

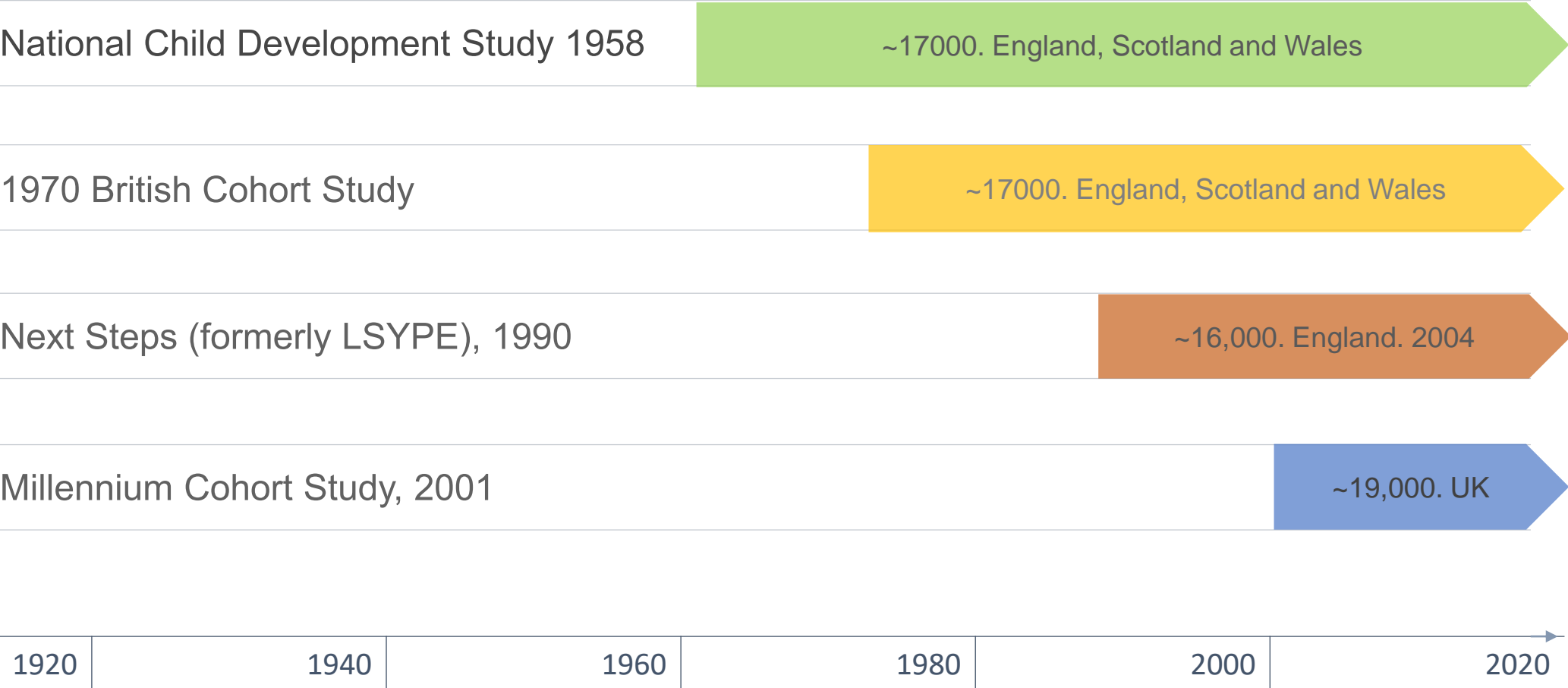
Aida Sanchez
Senior Data Manager




CENTRE FOR
LONGITUDINAL
STUDIES

Love Your Code

14th February 2020

Cohort Studies Timeline



<p>Birth </p>	<p>School years </p>	<p>Adult </p>
<p>Household composition Parental social class Obstetric history Smoking in pregnancy Pregnancy (problems, antenatal care) Labour (length, pain relief, problems) Birth (problems, sex, weight, length)</p>	<p>Household composition Parental social class Parental employment Financial circumstances Housing Health Cognitive tests Emotions and behaviour School Views and expectations Attainment</p>	<p>Household composition Employment Social class Income Housing Health (including biomarkers) Well-being and mental health Training and qualifications Cognitive tests Views and expectations Medication Alcohol use Diet Exercise</p>

Biomeasures data

Biosamples: blood, saliva, mill teeth

Blood pressure, pulse

Standing and sitting height

Weight, waist and hip circumferences

Lung function

Hearing thresholds

Depression and anxiety disorder (CIS-R)

Chronic widespread pain

Extensively assayed

Extensively genotyped

Externally linked data

Department of Education

NHS Digital

NHS Scotland

NHS Wales

ONS mortality

Geographical data

CLS data sharing: UK Data Service

[Abstract](#)

[FAQ's](#)

[Resources](#)

[Access data](#)

Access data

GN 33359

Millennium Cohort Study – Survey and Biomeasures Data



GN 33445

Millennium Cohort Study - Linked Administrative Data



GN 33557

Millennium Cohort Study - Geographical Data



CLS derived variables

- CLS researchers provide detailed description of DVs
- CLS data managers:
 - write the syntax from scratch (SIR PQL)
 - ingest into CLS research relational database
 - deposit DVs at the UKDS
- High level description of syntax in UKDS user guide
- Plan to release code via CLS GitHub (2020)

UKDS Derived Variables user guide

FDRLG00 - S6 DV Respondent: Religion - 7 category

Is main respondent's religion and is created by recoding and then combining the (merged) religion variables for England, Wales, Scotland, and Northern Ireland (FPRELE00, FPRELW00, FPRELS00, FPRELN00). If these questions were not asked at MCS6 answers from previous sweeps were used.

FPRELE00 and FPRELW00 are recoded into FDRLG00 as follows:

(1=8)
(2=1)
(3=6)
(4=3)
(6=2)
(7=4)
(8=7)
(9, 10, 11=0)

FPRELS00 and FPRELN00 are recoded into FDRLG00 as follows:

(1=8)
(2,3,4=1)
(5=2)
(7=4)
(8=5)
(9=3)
(10=7)
(11, 12, 13=0)

FDACT00 - S6 DV Respondents Economic Activity Status

If working (WKWK=1 or JBAW=1) and employee (EMPS=1), FDACT00=1
 If working (WKWK=1 or JBAW=1) and self-employed (EMPS=2), FDACT00=2
 If not working, recodes NWRK (non-working status) into FDACT00 as follows:

(3=3)
(4=4)
(5,6,7=5)
(8=6)
(2=8)
(1=7)
(95=9)

Millennium Cohort Study

MCS6 Derived Variables

MCS6(2015)

USER GUIDE

March 2017, Edition 1

Derived variables from external data users

Harmonisation – in place

- Consortium to deposit data at the UKDS

Users return their DVs to CLS – in discussion

- CLS checks, data management and UKDS deposit
- Very resource intensive
- Endorsement of external research

Users share their DVs via the UKDS – in discussion

- Syntax must be reproducible and well commented. No datasets.
- No involvement from CLS
- Checks for quality control purposes?
- UKDS location?

[Abstract](#)

[FAQ's](#)

[Resources](#)

[Access data](#)

Access data

GN 33359

Millennium Cohort Study – Survey and Biomeasures Data



GN 33445

Millennium Cohort Study - Linked Administrative Data



GN 33530

CLOSER Work Package 1: Harmonised Height, Weight and BMI in Five Longitudinal Cohort Studies



GN 33536

CLOSER Work Package 2: Harmonised Socio-Economic Measures in Four Longitudinal Cohort Studies



GN XXXX

Millennium Cohort Study: external user data outputs

Challenges of creating and documenting derived variables and harmonised datasets in longitudinal studies

Part 2

Dara O'Neill

CLOSER

Love Your Code 2020

Partner studies

Hertfordshire Cohort Study

1946 National Survey of Health and Development

1958 National Child Development Study

1970 British Cohort Study

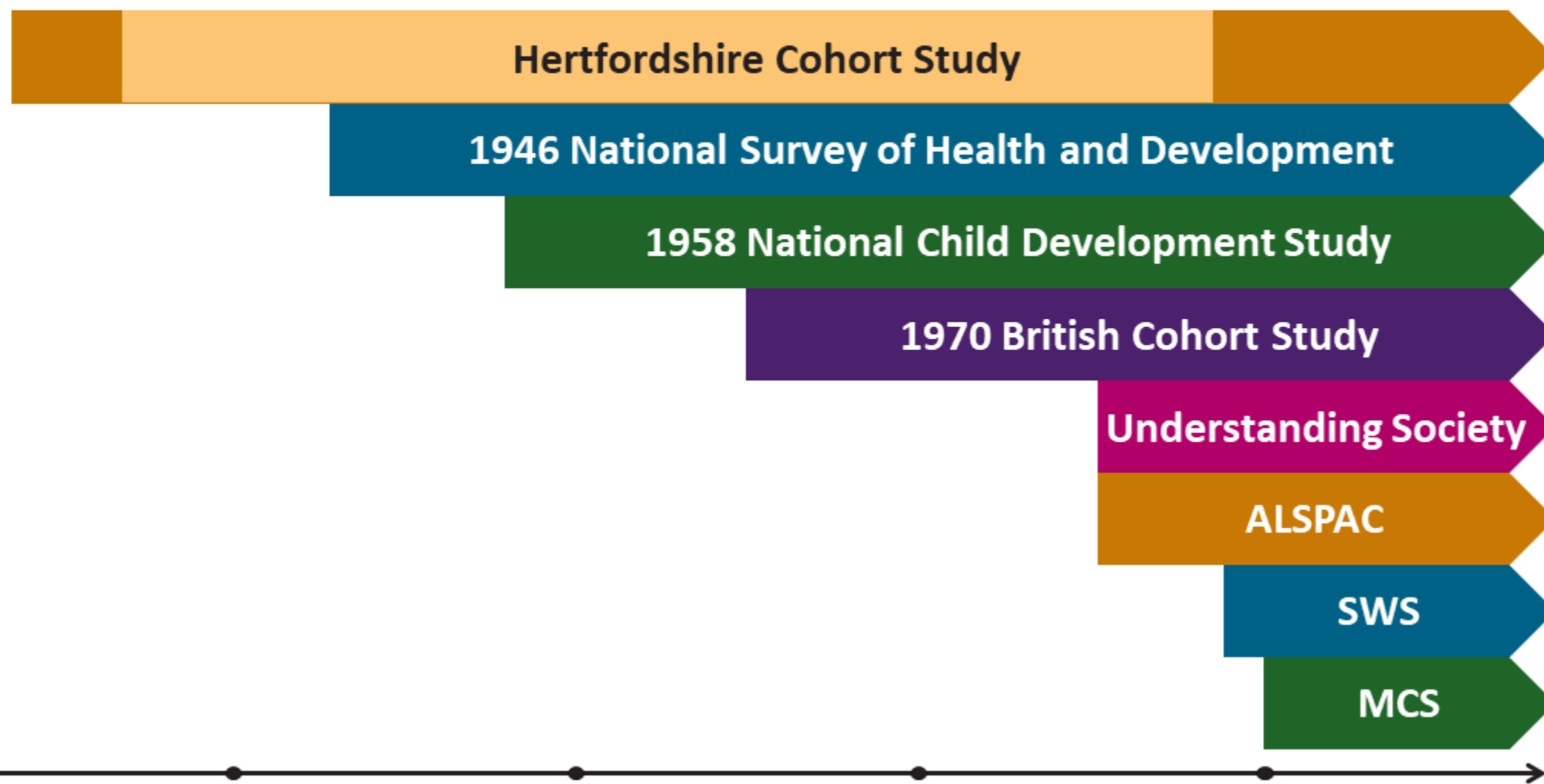
Understanding Society: The UK Household Longitudinal Study

Avon Longitudinal Study of Parents and Children (ALSPAC)

Southampton Women's Survey (SWS)

Millennium Cohort Study (MCS)

Partner studies



1920

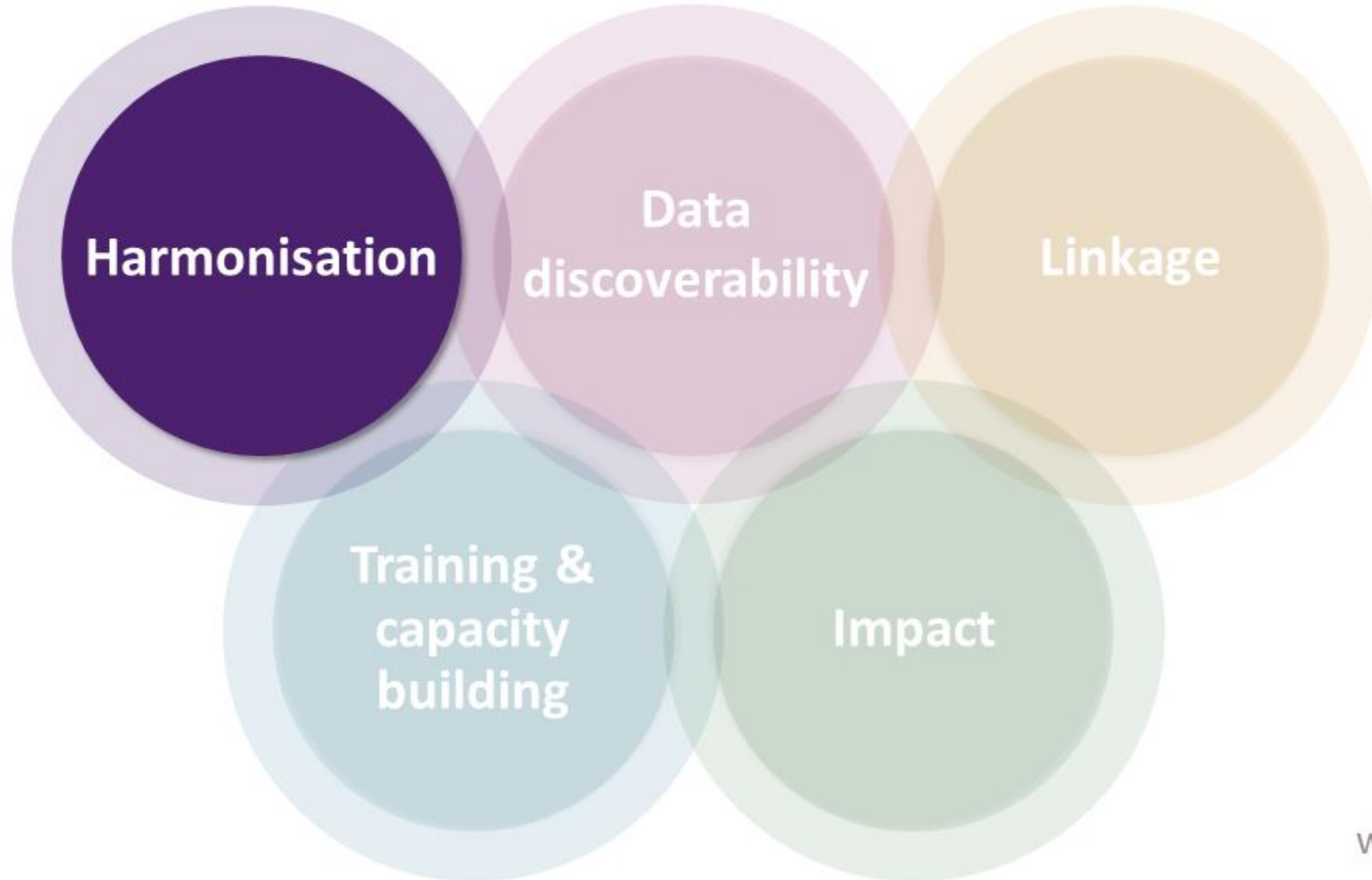
1940

1960

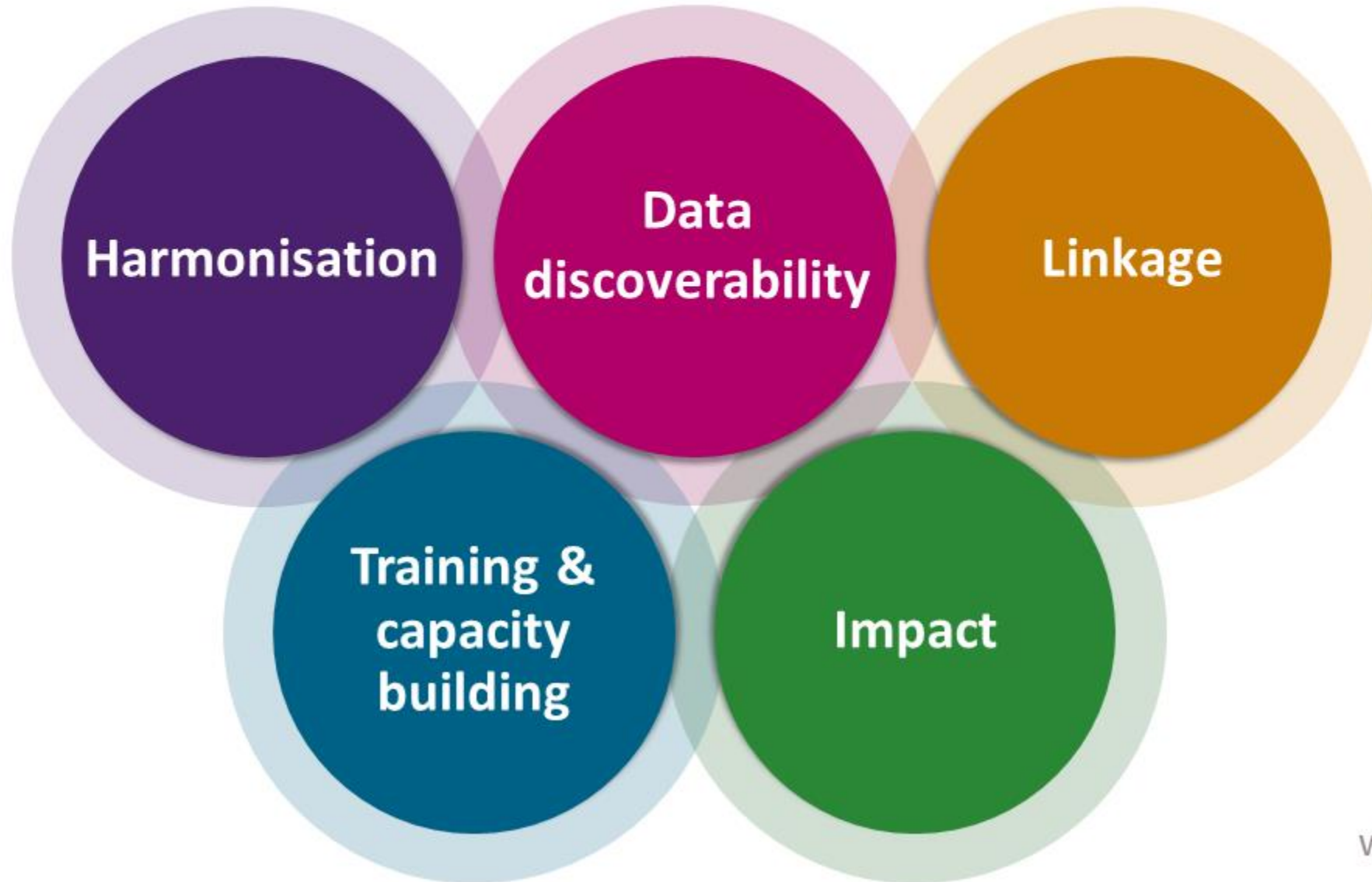
1980

2000

Our core areas of work



Our core areas of work



CLOSER harmonisation work packages

Socio-economic status



Education



Earnings & income



Household overcrowding



Childhood environment & adult mental wellbeing



Cognitive measures



Mental health measures



Mental health & wellbeing measure usage



Strategies for analysing biological samples



Cross-study biomarker data availability



DNA methylation



Visual functioning



Methods for determining pubertal status



Body size & composition



Physical activity measures



Dietary data



Process

Original study data

cross-sweep

cross-study



CLOSER harmonisation work packages

identification

evaluation

derivation

validation

documentation



Outputs

resource comparability reports

harmonised data releases

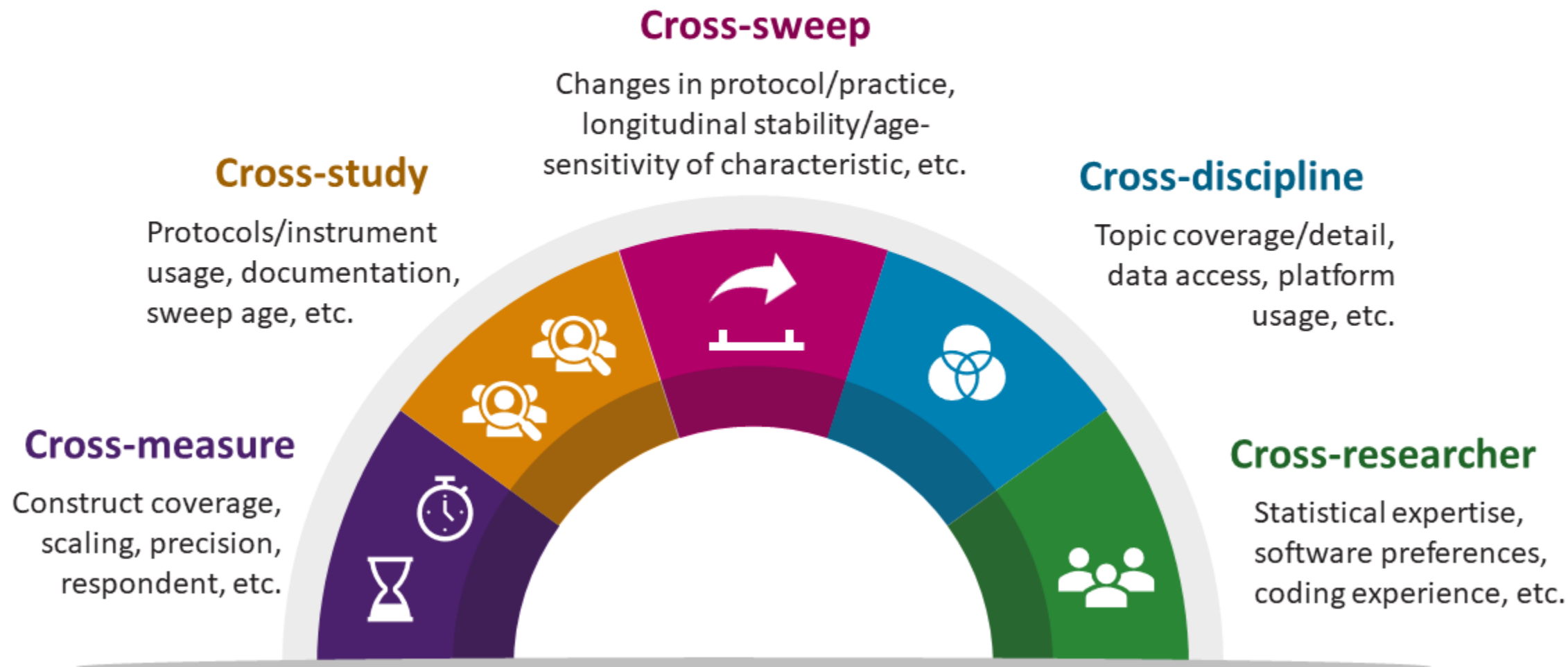


- measurement instruments
- assessment protocols
- data management

- datasets
- derivation code
- user guides



Process & documentation hurdles



Overcoming the obstacles

Data user guide templates



Standardised metadata documentation



Provenance

Question



Variable



Harmonised variable

Publishable lists



Code style consensus



Linking code to metadata: Parallel documentation of data transformation



Keep in touch



closer@ucl.ac.uk

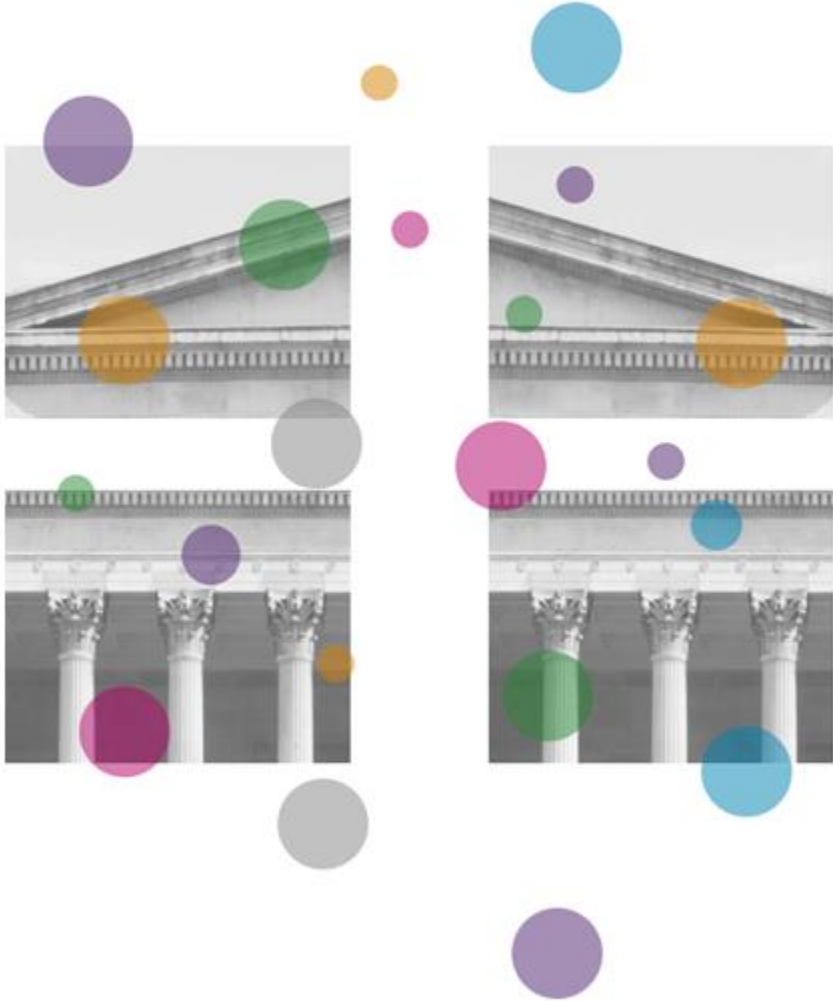


**Love longitudinal? So do we. Sign up for our email newsletters to get the latest from CLOSER:
closer.ac.uk**



@CLOSER_UK

Acknowledgements



Funder



Consortium partners



UK Data Service

