



Towards Machine-assisted Disclosure Assessment with DDI-CDI, DPV and sdcMicro

Deirdre Lungley, Principal Developer

Thomas Gilders, Data Engineer

Darren Bell, Associate Director, IT Services

Artiom Rumiancev, Data Engineer



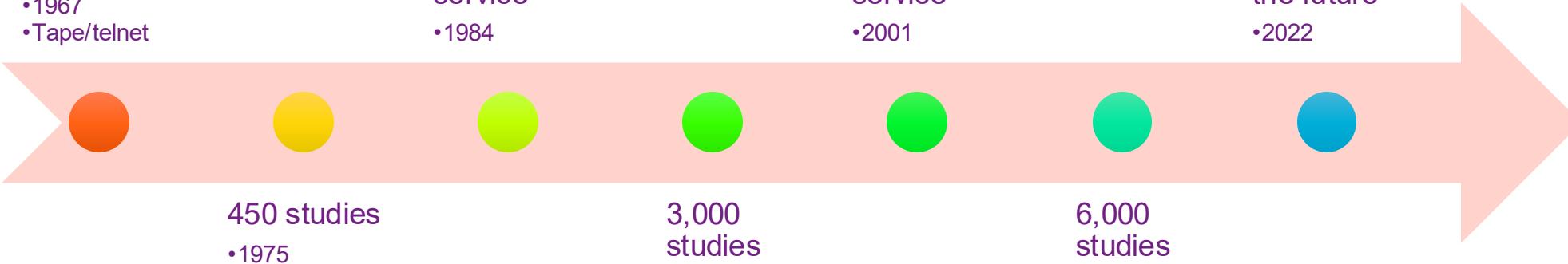
Data Dissemination at the UKDA

SSRC Data
Bank
created
•1967
•Tape/telnet

Floppy disk
service
•1984

Download
service
•2001

Prototyping
the future
•2022



450 studies
•1975

3,000
studies
•1993

6,000
studies
•2007

[Find data](#)[Deposit data](#)[Learning hub](#)[Training and events](#)[About](#)[News](#)[Impact](#)[Help](#)[Contact](#)

Data catalogue



Search our data catalogue...



Search

[Search tips](#) > [Variable and question bank](#) > [QualiBank](#) >[STUDIES \(9004\)](#)[SERIES \(79\)](#)

Date from: Date to:
[Refine date](#)

Topic:

Data Type:

Access:

Country:

[Reset filters](#)

Displaying 1 - 10 of 9004 results

Page 1 of 901

Results per page:

10

Sort by:

Date of collection (newest)

SN 852508 | [NEWETHPOP - Ethnic population projections for UK local areas 2011-2061](#)

Wohland, P, Hull York Medical School | Rees, P, School of Geography, University of Leeds |

Norman, P, School of Geography, University of Leeds | Lomax, N, School of Geography, University of Leeds |

Clark, S, School of Geography, University of Leeds

SN 855012 | [Food Demand Scenarios in 70 World Regions, 2010-2060](#)

Vinuales, J, University of Cambridge | Pollitt, H, Cambridge Econometrics | Salas, P, University of Cambridge

SN 5023 | [Eurostat New Cronos, 1950-2060](#)

Eurostat

SN 6777 | [Ethnic Population Projections for the United Kingdom and Local Areas, 2001-2051](#)

Wohland, Pia | Norman, Paul | Boden, Peter | Rees, Philip

SN 8544235 | [Maritime trade projections under four climate related socio-economic development scenarios 2010-2050](#)

SCC Consortium, , SCC Consortium

SN 8544436 | [Demand for ports to 2050: Climate policy, growing trade and the impacts of sea level 2010-2050](#)

Hanson, S, University of Southampton | Nicholls, R, University of East Anglia



Data Dissemination in 2022

Challenges



Data is more complex, coming from more sources, and more quickly



Increased expectations from researchers for access turnaround times



Funders, industry and government increasingly interested in research which combines data from different arenas



Data Product Builder

What does it do?



Row and column sub setting, linkages, disclosure risk analysis



Aggregations, crosstabs



DPB Tool Powered by DDI-CDI (examples)

CDI Class	Class Association	Class Attribute	Function
ConceptualVariable	uses Concept		Enable dataset selection via semantic, geographical and temporal concepts
RepresentedVariable			Shared RepresentedVariables used for finding linkages between datasets
WideDataStructure	has MeasureComponent		Specify dataset filterable variables
RepresentedVariable	takesSubstantiveValuesFrom CodeList		Filter responses by category selection
ConceptualVariable	uses Concept		DPV concept - used to specify sensitive variables
InstanceVariable		variableFunction	Specify weight variable
ConceptualVariable	uses Concept		DPV concept - used to specify key variables to feed SDC analysis

DPB – Dataset Selection

DSaaP
Data Services as a Platform

Deirdre Lungley 

1 Select Resource 2 Identify Linkages 3 Filter 4 Output Variables 5 Preview 6 Build

No filters applied

CONCEPTS **GEOGRAPHY** YEAR

Search

- > ABILITY
- > ACHIEVEMENT
- > ADMINISTRATION
- > ADMINISTRATIVE AREAS
- > ADMINISTRATIVE STRUCTURES
- > ADVICE
- > AGE
- > AGE GROUPS
- ANALYSIS
- > ANIMALS
- > ANTHROPOLOGY
- > ARMAMENT PROCESS
- > ARMED FORCES
- > ARTS
- > ATTENDANCE
- > ATTITUDES
- BEHAVIOURAL SCIENCES
- > BELIEFS
- > BIOLOGY
- BUDGETS
- > BUILDING SERVICES
- > CUMULUS

Datasets

- 1 Index Multiple Deprivation
MHCLG
- 2 Open Greenspace
OS
- 3 Understanding Society Teaching Dataset - Wave 8 (2018)
ISER
- 4 Understanding Society Teaching Dataset - Wave 9 (2019)
ISER
- 5 Understanding Society Teaching Dataset Time Diaries - Wave 9 (2019)
ISER

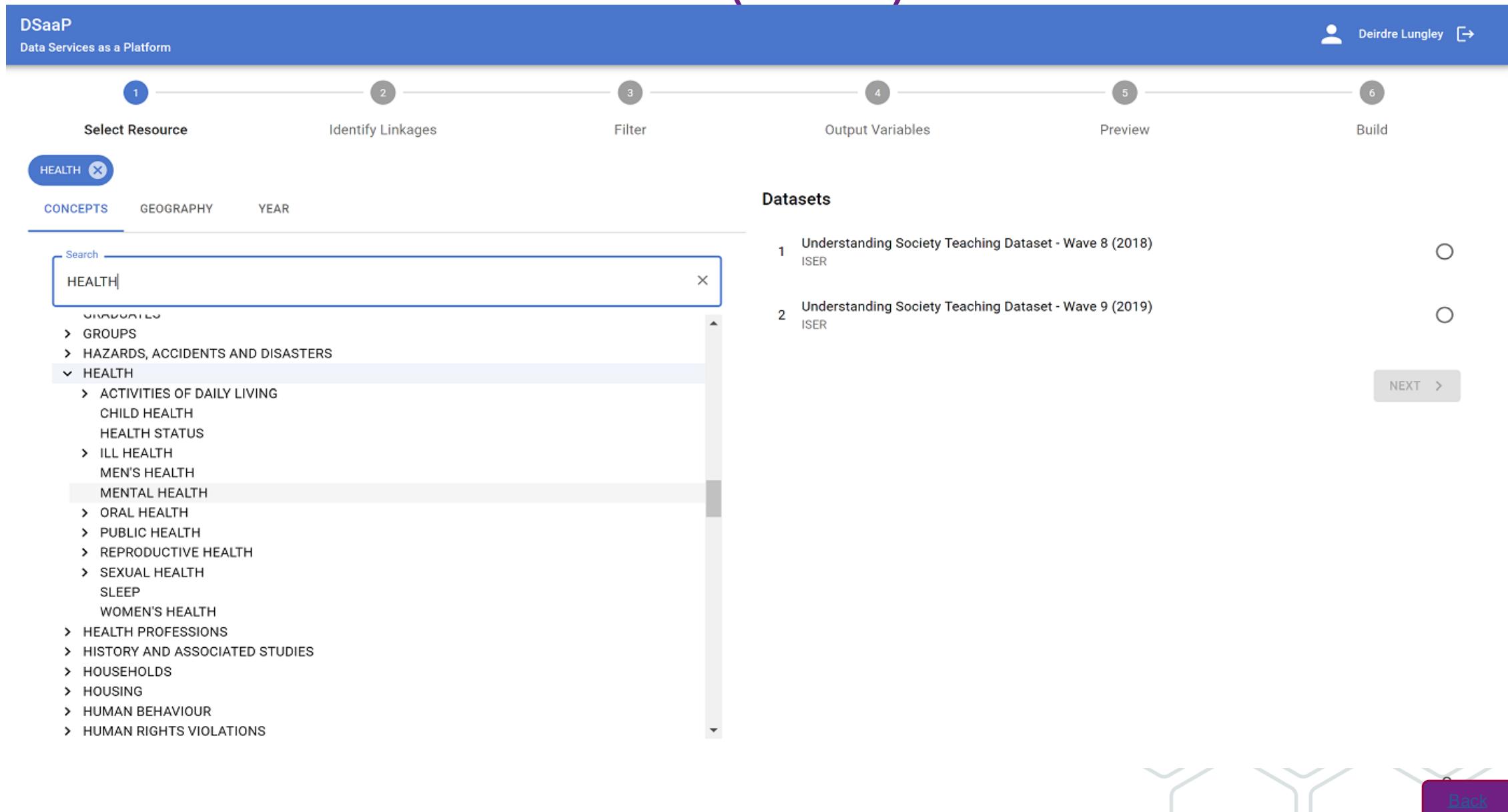
NEXT >

Back

DPB – Dataset Selection (cont.)

DSaaP
Data Services as a Platform

Deirdre Lungley 



1 Select Resource 2 Identify Linkages 3 Filter 4 Output Variables 5 Preview 6 Build

HEALTH 

CONCEPTS GEOGRAPHY YEAR

Search 

HEALTH

GRADUATES

> GROUPS

> HAZARDS, ACCIDENTS AND DISASTERS

< HEALTH

> ACTIVITIES OF DAILY LIVING

CHILD HEALTH

HEALTH STATUS

> ILL HEALTH

MEN'S HEALTH

MENTAL HEALTH

> ORAL HEALTH

> PUBLIC HEALTH

> REPRODUCTIVE HEALTH

> SEXUAL HEALTH

SLEEP

WOMEN'S HEALTH

> HEALTH PROFESSIONS

> HISTORY AND ASSOCIATED STUDIES

> HOUSEHOLDS

> HOUSING

> HUMAN BEHAVIOUR

> HUMAN RIGHTS VIOLATIONS

Datasets

1 Understanding Society Teaching Dataset - Wave 8 (2018)
ISER 

2 Understanding Society Teaching Dataset - Wave 9 (2019)
ISER 

NEXT >

Back

DPB – Dataset Selection (cont.)

DSaaP
Data Services as a Platform

1 Select Resource 2 Identify Linkages 3 Filter 4 Output Variables 5 Preview 6 SDCMicro 7 Build

No filters applied

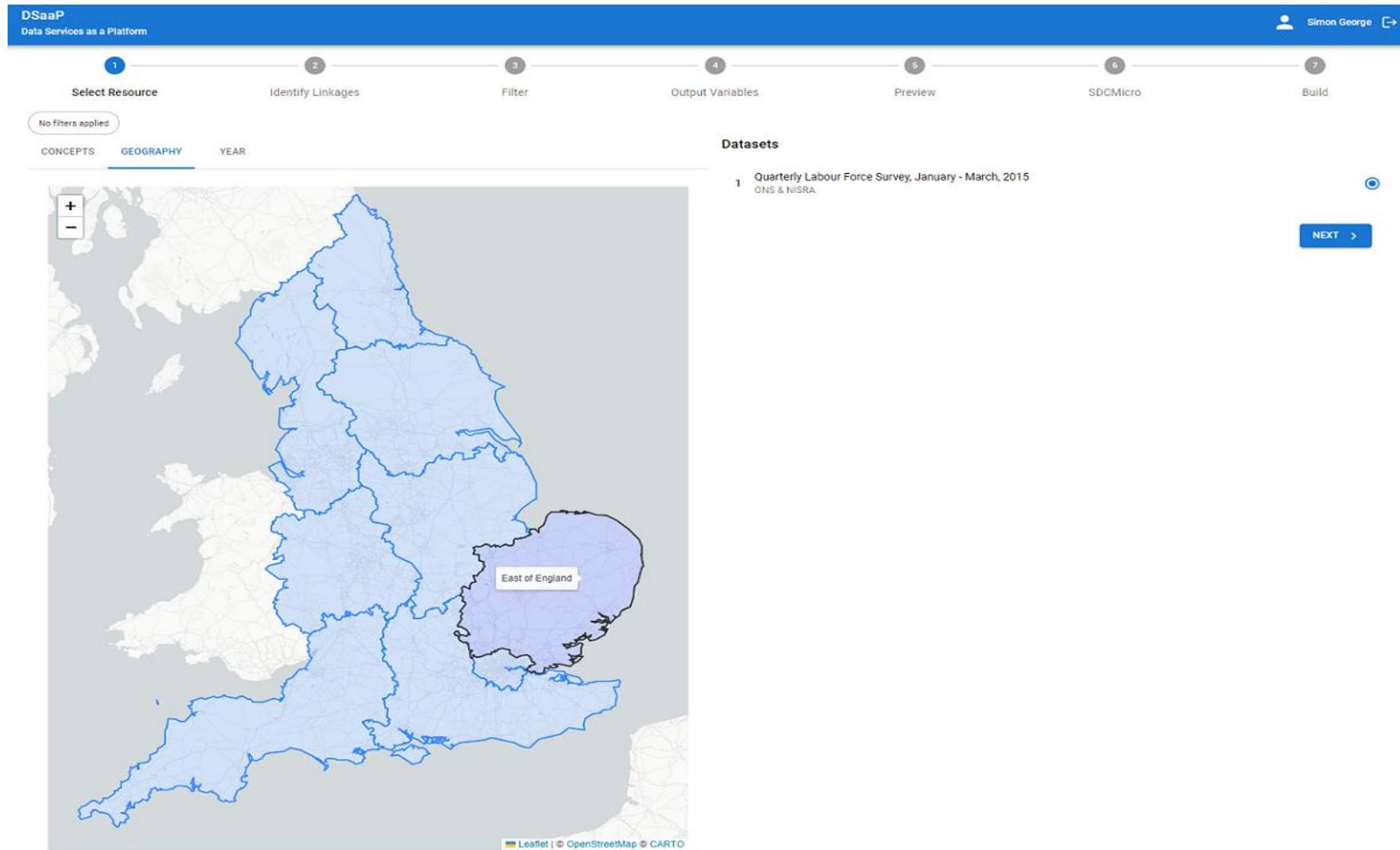
CONCEPTS GEOGRAPHY YEAR

Geography selection map showing the United Kingdom with various regions highlighted in blue. A callout box labeled "East of England" points to the highlighted area.

Datasets

1 Quarterly Labour Force Survey, January - March, 2015
ONS & NISRA

NEXT >



Back

DPB – Select Linkage

DSaaP
Data Services as a Platform

Deirdre Lungley 

1 2 3 4 5 6

Select Resource Identify Linkages Filter Output Variables Preview Build

Primary Dataset

Understanding Society Teaching Dataset - Wave 9 (2019)
ISER

Potential Linkages

Index Multiple Deprivation
MHCLG

Lower-layer Super Output Areas (2011)

Open Greenspace
OS

Postcode; space included

Understanding Society Teaching Dataset - Wave 8 (2018)
ISER

cross-wave person identifier (public release)

Understanding Society Teaching Dataset Time Diaries - Wave 9 (2019)
ISER

cross-wave person identifier (public release)

< PREVIOUS   NEXT >

Back

DPB - Filterable Variables

DSaaS
Data Services as a Platform

Deirdre Lungley 

1 2 3 4 5 6

Select Resource Identify Linkages Filter Output Variables Preview Build

Understanding Society Teaching Dataset - Wave 9 (2019)
ISER

ISER
cross-wave person identifier
(public release)

Understanding Society Teaching Dataset Time Diaries - Wave 9 (2019)
ISER

Case count: 175936

Variable name	Variable label	Source dataset	Variable type
hhorig	Sample origin, household	US_9	Nominal
fimngrs_dv	total monthly personal income gross	US_9	Scale
fimnnet_dv	total net personal income - no deductions	US_9	Scale
psu	Primary sampling unit	US_9	Scale
paygu_dv	usual gross pay per month: current job	US_9	Scale
paynu_dv	usual net pay per month: current job	US_9	Scale
j2pay_dv	pay in second job	US_9	Scale

Selected filter variable Filter to apply

No rows

Rows per page: 100 < >

1–98 of 98

PREVIOUS   NEXT  Back

DPB – Filter Responses by Category Selection

DSaaP
Data Services as a Platform

Deirdre Lungley

1 2 3 4 5 6

Select Resource Identify Linkages Filter Output Variables Preview Build

Understanding Society Teaching Dataset - Wave 9 (2019)
ISER

GD cross-wave person identifier (public release)

Understanding Society Teaching Dataset Time Diaries - Wave 9 (2019)
ISER

Case count: 175936

Variable name	Variable label	Source dataset	Variable type
gor_dv	Government Office Region	US_9	Nominal

1 row selected Rows per page: 100 1-1 of 1 < >

< PREVIOUS

Selected filter variable	Filter to apply
Government Office Region	<input type="text"/> E12000004 - East Midlands E12000006 - East of England E12000007 - London E12000001 - North East E12000002 - North West N99999999 - Northern Ireland S99999999 - Scotland E12000008 - South East E12000009 - South West W99999999 - Wales E12000005 - West Midlands E12000003 - Yorkshire and the Humber

Back

DPB – Sensitive Variables

DSaaP
Data Services as a Platform

Deirdre Lungley ➔

1 2 3 4 5 6

Select Resource Identify Linkages Filter Output Variables Preview Build

Understanding Society Teaching Dataset - Wave 9 (2019)
ISER

cross-wave person identifier
(public release)

Understanding Society Teaching Dataset Time Diaries - Wave 9 (2019)
ISER

Case count: 17846

Variable name	Variable label	Source dataset	Variable type
grpbfst	Group benefit from voting	US_9	Nominal
perbfst	personal benefit in voting	US_9	Nominal
pcds	Postcode; space included	US_9	Nominal
lsoa	Lower-layer Super Output Areas (2011)	US_9	Nominal
laua	Local Authority	US_9	Nominal
distmov_dv	Distance participant moved since last wave (km)	US_9	Scale
addrmov_dv	Participant changes address postcode since last...	US_9	Nominal

Selected filter variable Filter to apply

Government Office Region E12000006 - East of England

Rows per page: 100 ▾ 1-1 of 1 < >

PREVIOUS < NEXT >

Back

DPB – Weight Variable

DSaaP
Data Services as a Platform

Tom Gilders 

1 2 3 4 5 6 7

Select Resource Identify Linkages Filter Output Variables Preview SDCMicro Build

Quarterly Labour Force Survey, January - March, 2015
ONS & NISRA

Case count: 95941

Variable name	Variable label	Source dataset
<input type="checkbox"/> METHM	Main method of looking for work	LFS_15_Q1
<input type="checkbox"/> NOLWM	Main reason not looking for work	LFS_15_Q1
<input type="checkbox"/> PIWT18	Income weight 2018	LFS_15_Q1
<input checked="" type="checkbox"/> PWT18	Person weight 2018	LFS_15_Q1
<input type="checkbox"/> REFDT	Reference date	LFS_15_Q1
<input type="checkbox"/> THISWV	Wave to which data refers	LFS_15_Q1
<input type="checkbox"/> TYPINT	Type of Interview	LFS_15_Q1

4 rows selected Rows per page: 100 1–27 of 27

PREVIOUS < NEXT >

Back

DPB – Disclosure Risk Analysis

DEMO



Back

Our Vision for Sensitive Interoperability

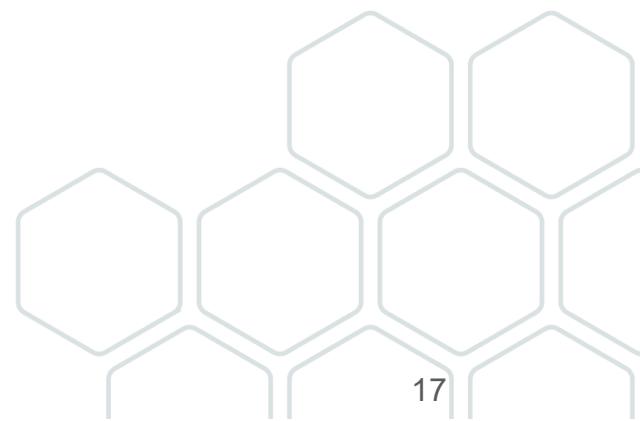
Break the current binary utility/risk trade-off of Secure Access or Open Access

Reduce the time to access data

How?

- Study metadata, enhanced with Data Privacy Vocabulary (DPV) concepts
- Feed sdcMicro risk function which drive machine-assisted disclosure assessment
- In turn feed a decision tree of real-time access outcomes
- Provide researchers with more flexible choices based on real-time mitigations of key variable sensitivity e.g. top/bottom recoding

Any questions?





Thank you.

