UK Data Service

UKDA – Data Product Builder

Deirdre Lungley, Principal Developer

Thomas Gilders, Data Engineer

Copyright © 2022 UK Data Service. Created by DaRT team, UK Data Service.

Data Product Builder

What does it do?

✓ Row and column sub setting, linkages, disclosure risk analysis

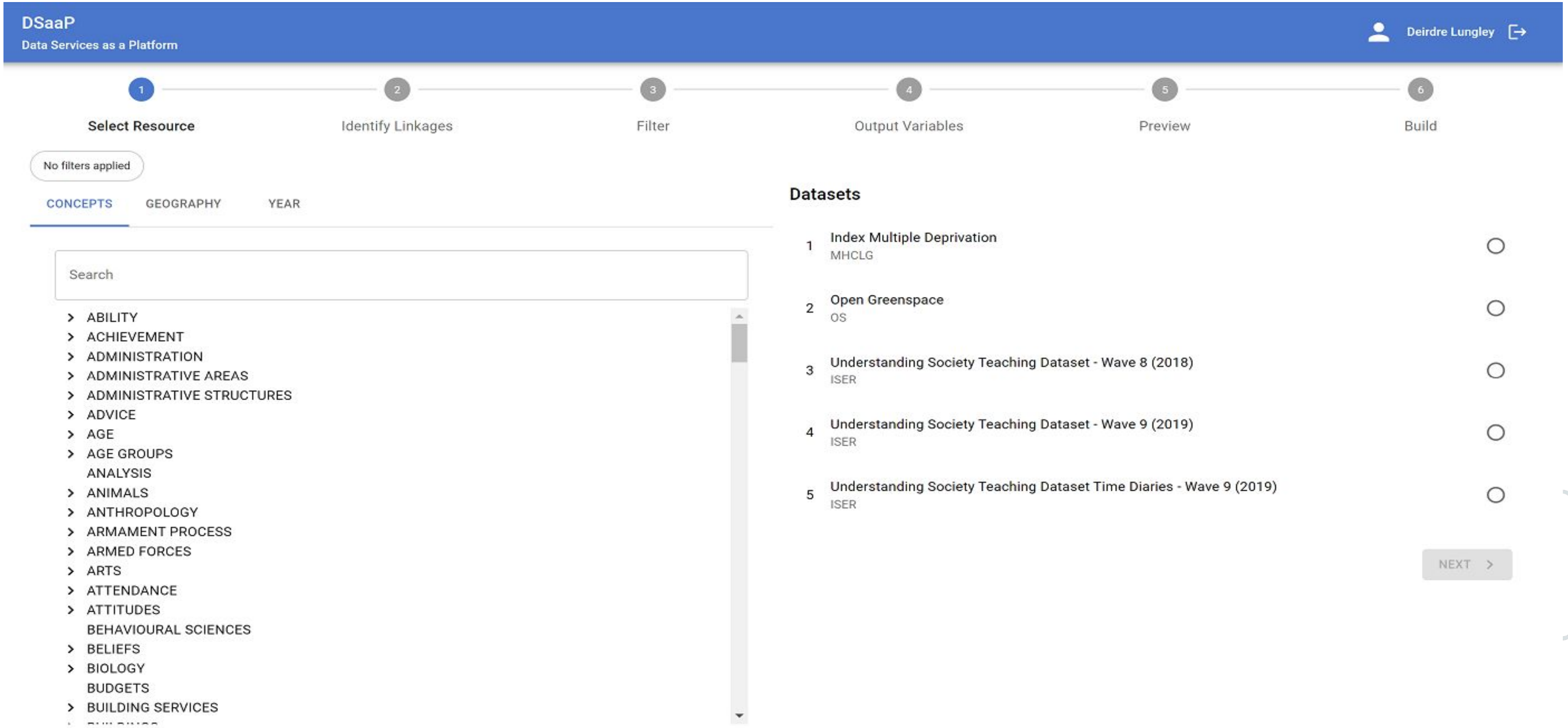
✗ Aggregations, crosstabs

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

Forward

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
NEXT >

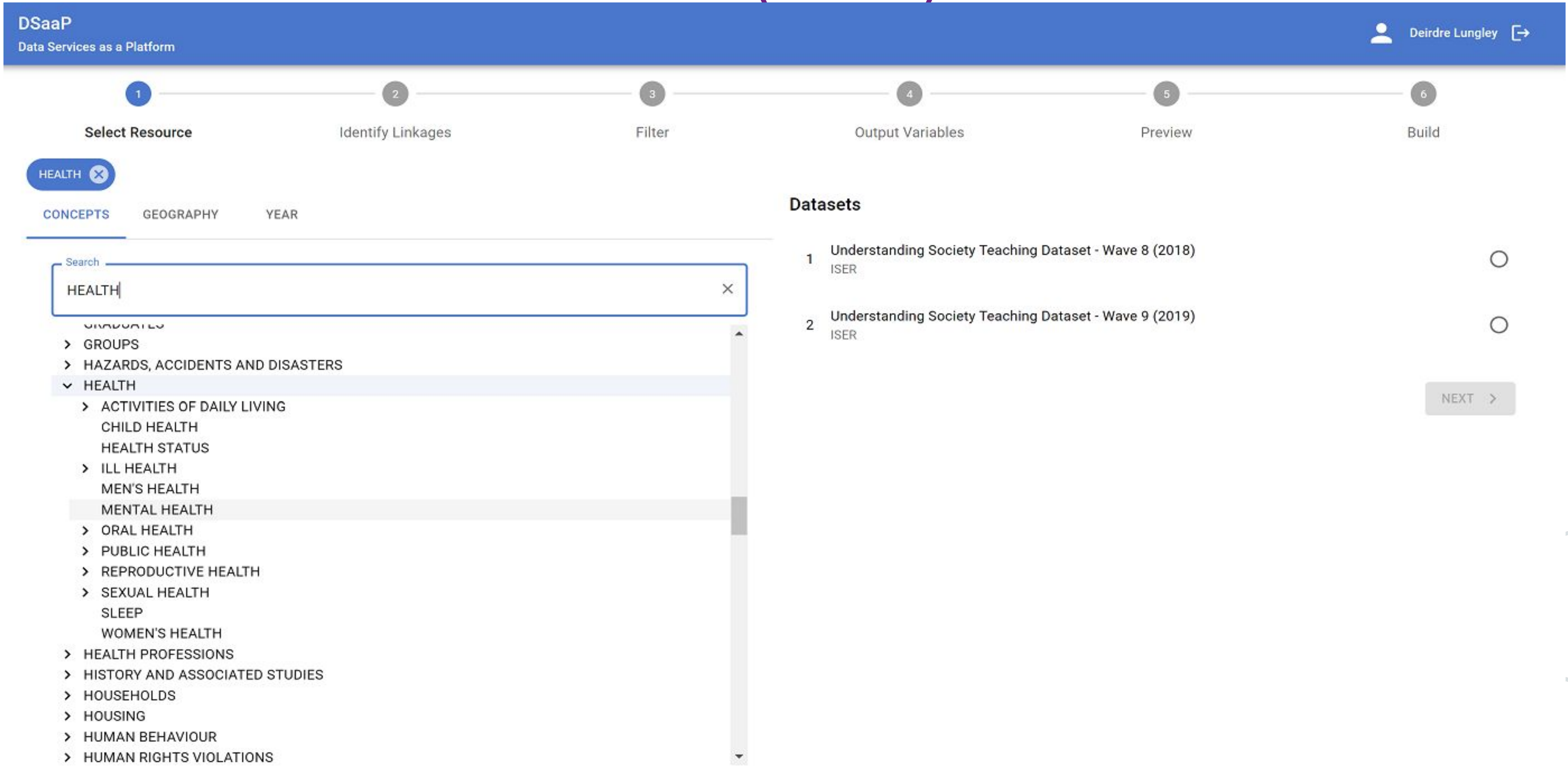
- › ABILITY
- › ACHIEVEMENT
- › ADMINISTRATION
- › ADMINISTRATIVE AREAS
- › ADMINISTRATIVE STRUCTURES
- › ADVICE
- › AGE
- › AGE GROUPS
- ANALYSIS
- › ANIMALS
- › ANTHROPOLOGY
- › ARMAMENT PROCESS
- › ARMED FORCES
- › ARTS
- › ATTENDANCE
- › ATTITUDES
- BEHAVIOURAL SCIENCES
- › BELIEFS
- › BIOLOGY
-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

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	○

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
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
NEXT ›

Search: HEALTH

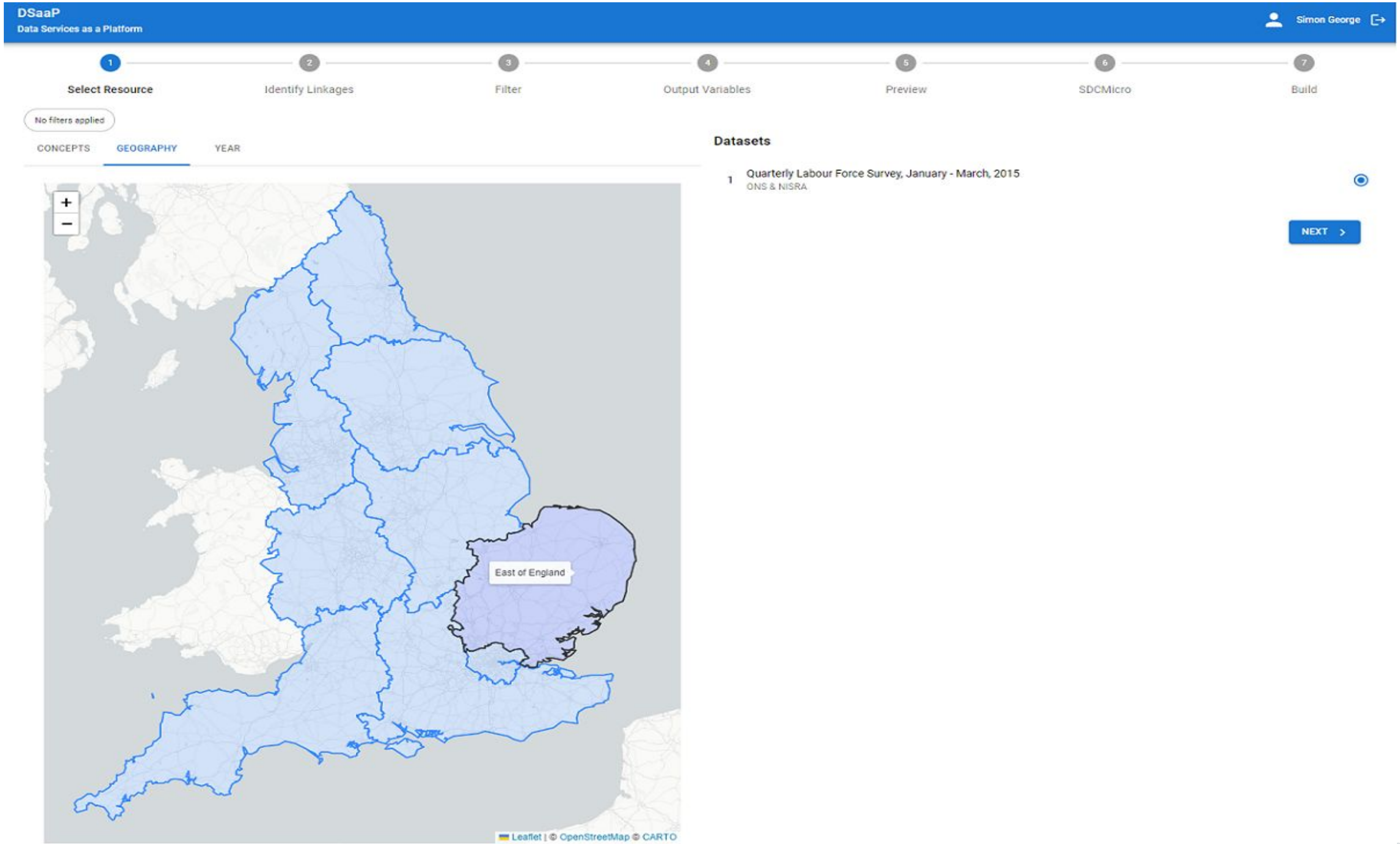
Datasets

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

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

Back

DPB – Dataset Selection (cont.)

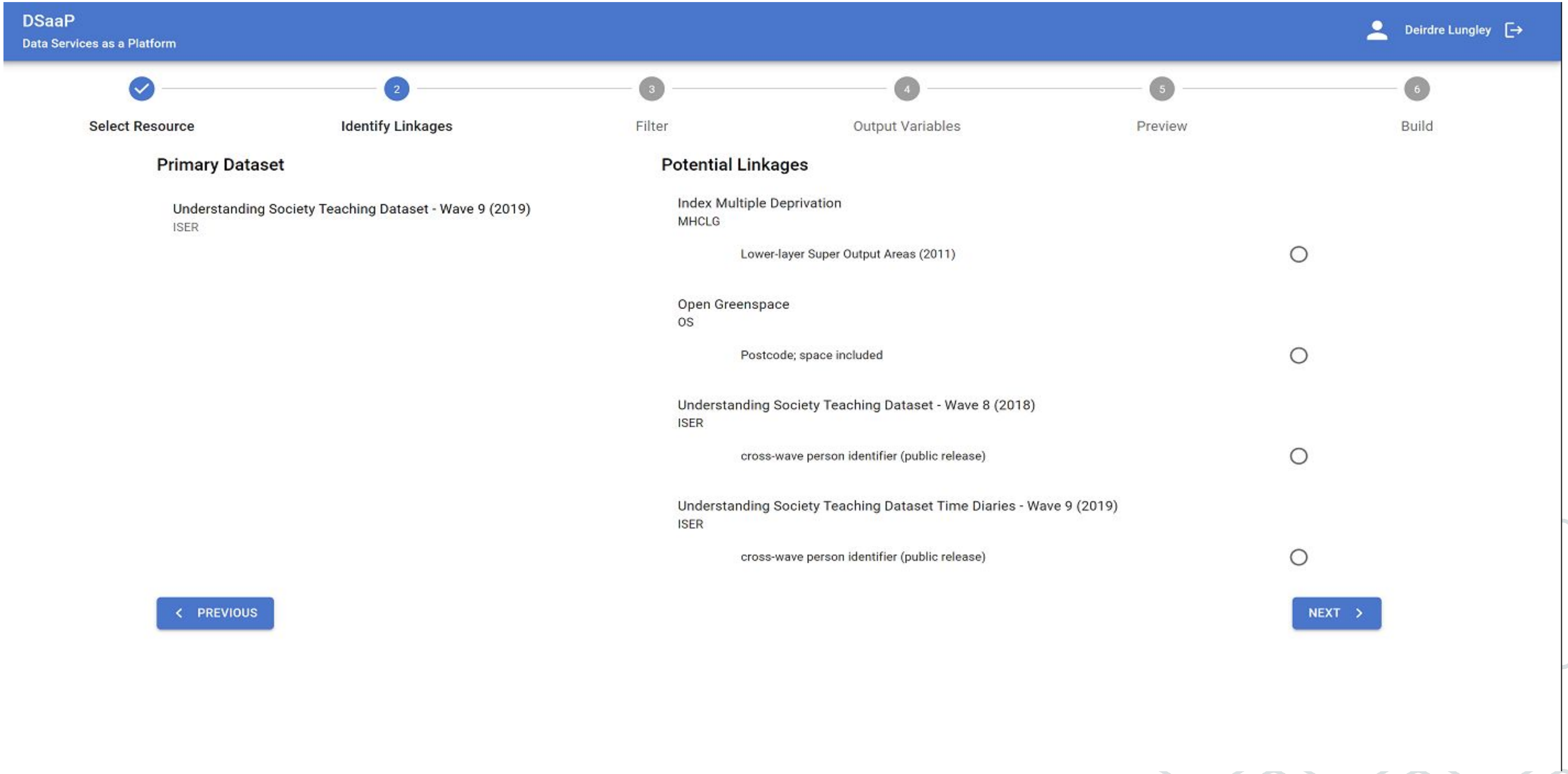

DSaaP
Data Services as a Platform
Simon George
1 Select Resource
2 Identify Linkages
3 Filter
4 Output Variables
5 Preview
6 SDCMicro
7 Build
No filters applied
CONCEPTS
GEOGRAPHY
YEAR
+
−
East of England
Leaflet | © OpenStreetMap © CARTO
NEXT >

Datasets

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

Back

DPB – Select Linkage


DSaaP
Data Services as a Platform
Deirdre Lungley
Select Resource
2
Identify Linkages
3
Filter
4
Output Variables
5
Preview
6
Build
PREVIOUS
NEXT

Primary Dataset

Understanding Society Teaching Dataset - Wave 9 (2019)
ISER

Potential Linkages

Index Multiple Deprivation
MHCLG
Lower-layer Super Output Areas (2011) ○

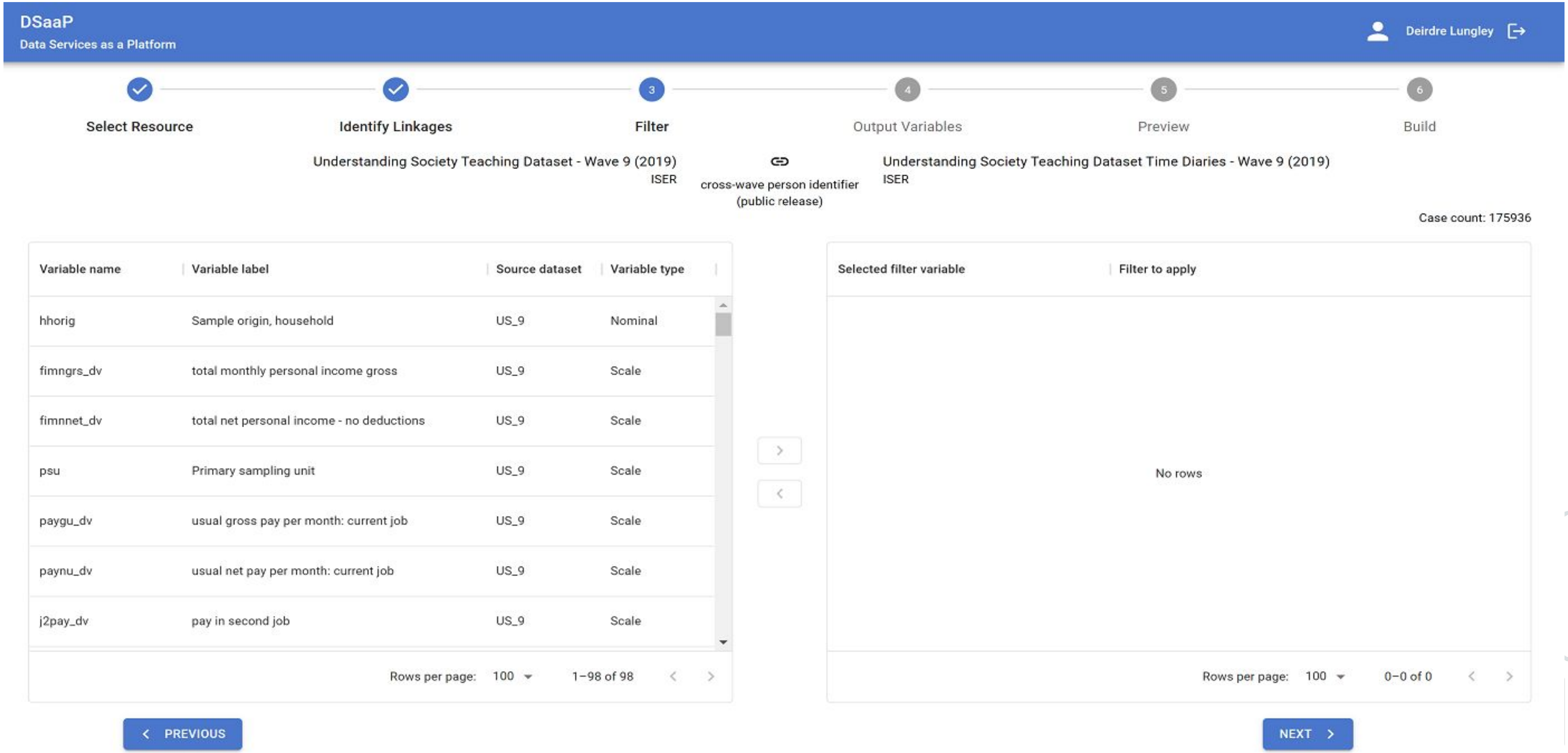
Open Greenspace
OS
Postcode; space included ○

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) ○

Back

DPB - Filterable Variables


DSaaP
Data Services as a Platform
Deirdre Lungley
Select Resource
Identify Linkages
3
Filter
4
Output Variables
5
Preview
6
Build
PREVIOUS
NEXT

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: 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

Rows per page: 100 1–98 of 98

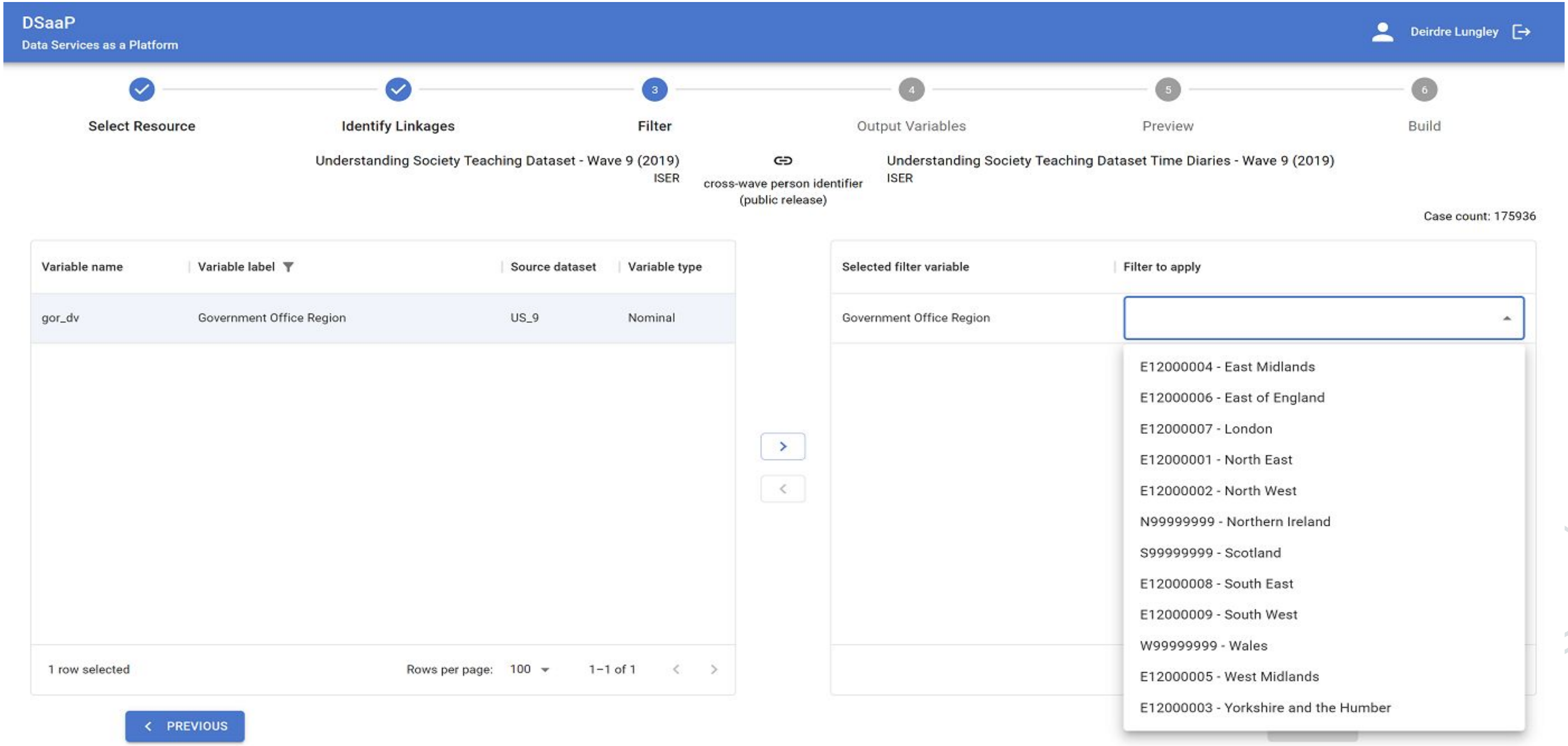
Selected filter variable	Filter to apply

No rows

Rows per page: 100 0–0 of 0

Back

DPB – Filter Responses by Category Selection



DSaaP
Data Services as a Platform
Deirdre Lungley
Select Resource
Identify Linkages
3
Filter
4
Output Variables
5
Preview
6
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: 175936

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

1 row selected

Rows per page: 100

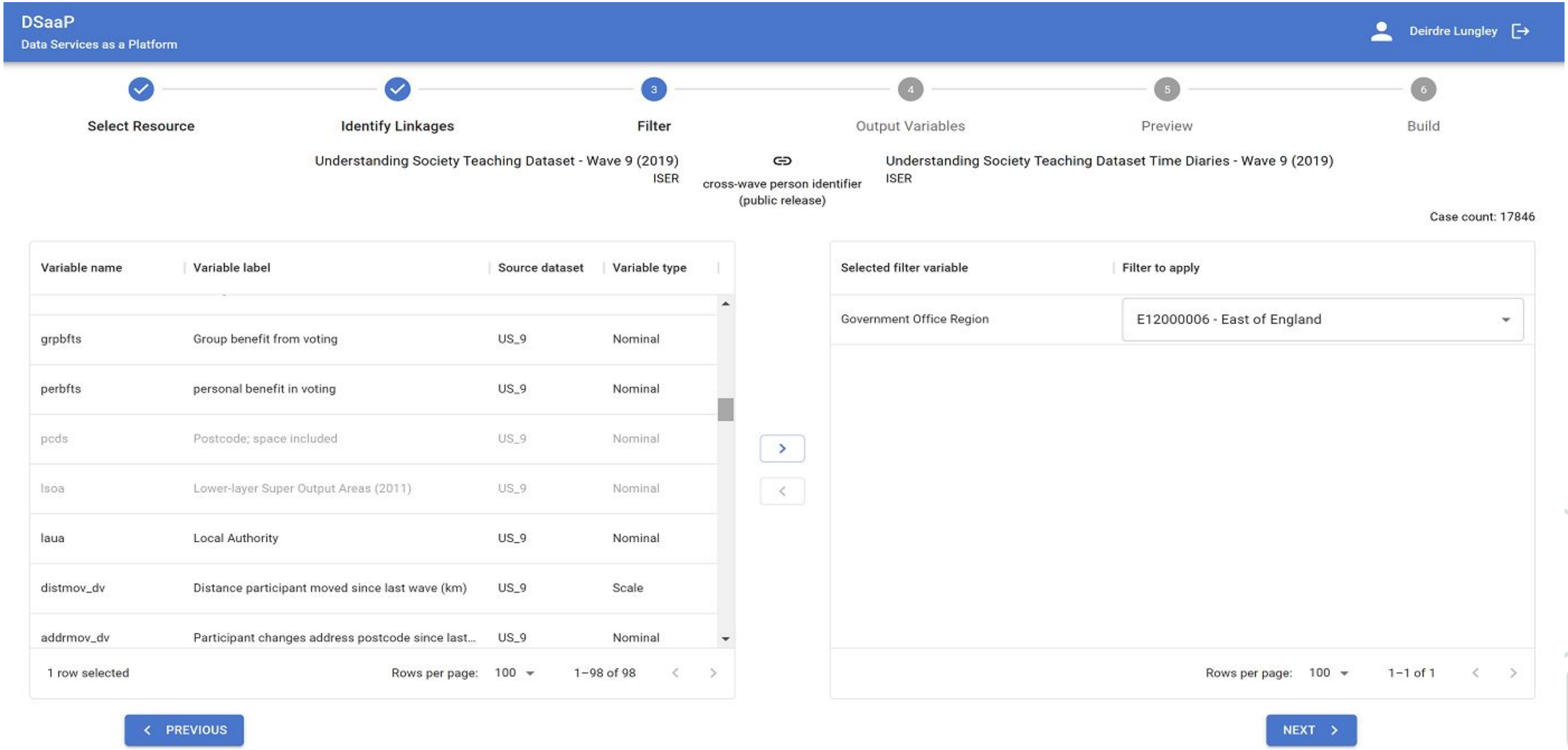
1–1 of 1

Selected filter variable	Filter to apply
Government Office Region	

- 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
Select Resource
Identify Linkages
3
Filter
4
Output Variables
5
Preview
6
Build
PREVIOUS
NEXT

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
grpbfts	Group benefit from voting	US_9	Nominal
perbfts	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

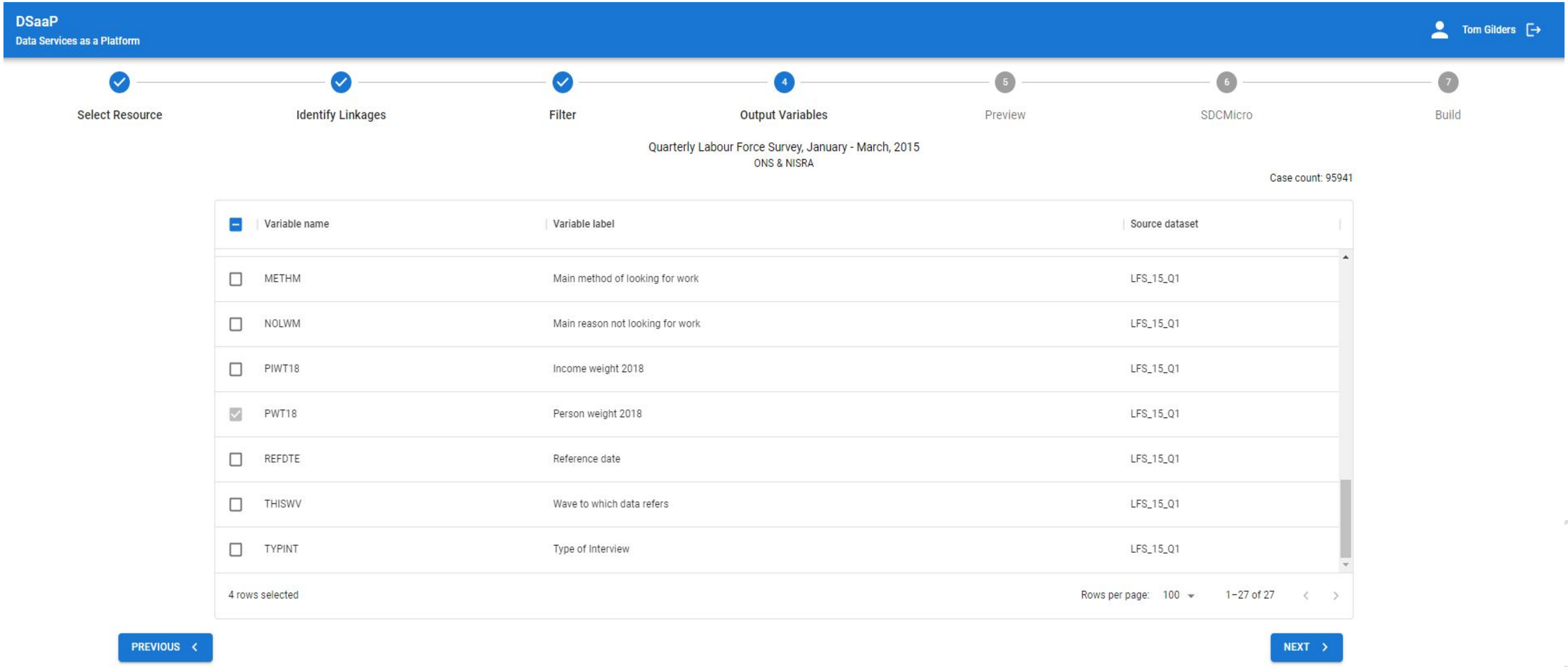
1 row selected   Rows per page: 100   1–98 of 98

Selected filter variable	Filter to apply
Government Office Region	E12000006 - East of England

Rows per page: 100   1–1 of 1

Back

DPB – Weight Variable


DSaaP
Data Services as a Platform
Tom Gilders
Select Resource
Identify Linkages
Filter
4
Output Variables
5
Preview
6
SDCMicro
7
Build
PREVIOUS
NEXT

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

Case count: 95941

▬	Variable name	Variable label	Source dataset
☐	METHM	Main method of looking for work	LFS_15_Q1
☐	NOLWM	Main reason not looking for work	LFS_15_Q1
☐	PIWT18	Income weight 2018	LFS_15_Q1
☑	PWT18	Person weight 2018	LFS_15_Q1
☐	REFDTE	Reference date	LFS_15_Q1
☐	THISWV	Wave to which data refers	LFS_15_Q1
☐	TYPINT	Type of Interview	LFS_15_Q1

4 rows selected

Rows per page: 100

1–27 of 27

Back

UK Data Service

Disclosure Risk Analysis Demo

UK Data Service

Thank you.

Deirdre Lungley
(dmlung@essex.ac.uk)