Using UK data portals to find and evaluate data for social science research

January 2021

National Centre for Research Methods Workshop

Scott Orford, Katie Dickson, Samuel Jones

Workshop programme

Part 1: Using data portals as data discovery tools for research:

– The benefits of using data portals to discover secondary datasets.

– An overview of the key data portals providing data for the UK.

Break

Part 2: How to use a data portal to discover data for a research project:

– Overview of the WISERD DataPortal project.

– Introduction and demonstration of the WISERD DataPortal's key functions.

Part 3: WISERD DataPortal workshop

– Practical workshop on using the WISERD DataPortal and Q&A session

WI/SERD

Part 1: Using data portals as data discovery tools for research

The benefits of using data portals to discover secondary datasets.

Data portals: Why do they exist?

- **Data are a key resource for research:**
  - Providing evidence to inform and support conclusions.
  - Data are expensive to collect and often underused.
  - Research Councils keen to encourage reuse and repurposing of existing data.

- **Issues with re-using secondary data**
  - Issues with discoverability of secondary research data.
  - Lack of awareness about what data exist.

- **Data portals try to mitigate these issues:**
  - Improve the discoverability and accessibility of secondary data.
  - Encourage reuse and repurposing of existing data.

W/SERD

Data portals: What functionality do they provide?

- **Meta-data discovery:**
  - Meta-data is descriptive information about a dataset.
  - Meta-data standards ensure catalogues hold useful information.
  - Catalogues may include: a description of the data, key themes, spatial extent, temporal extent, licensing information and how to access it.
  - Catalogues can often be filtered and searched.
- **Data access:**
  - Either as downloads, through a web-browser or via APIs
- **Data visualisation:**
  - Commonly as tables, charts or maps.
- **Data training and tutorials:**
  - Training materials, tutorials and guidance documentation.
  - Focused on using one particular portal, or more general guidance on making the most of data.

W/SERD

Data portals: Who hosts them?

- **Public sector:**
  - **Central government**: Data.gov.uk (UK),
  - **National government**: Statistics.gov.scot (Scotland), Northern Ireland Statistics and Research Agency (Northern Ireland) and StatsWales (Wales)
  - **Government departments**: ONS Portal, ONS Open Geography, DWP (Stat-Xplore), MHCLG (Energy Performance); NHS Digital; Local Government Data Unit (Infobase Cymru)
  - **Local government**: many, such as 'My Cardiff' (Cardiff Council).
- **Academic:**
  - Research Councils:
    - **UKDS**: UKDA, Infuse, Casweb, WICID, UK Borders, Understanding Society.
    - **ESRC's Big Data Network**: ADR-UK, CDRC, Urban Big Data Centre, Business and Local Government Data Research Centre.
    - **Research Centres:** WISERD DataPortal, CaCHE, Timescapes
  - Public sector collaborations:
    - NOMIS, Vision of Britain, Digimap

WISERD

Data portals: Who hosts them?

- **Third sector:**
    - Carnegie UK Trust: Understanding Welsh Places, Understanding Scottish Places
    - Welsh Council for Voluntary Action: The Data Hub
    - Shelter: The Housing Databank

- **Private sector:**
    - Local Data Company, CACI, plus many more…

Data portals: Who can access them?

- **Data licensing conditions:**
  - Open access, Restricted, Safeguarded, Controlled

- **Payment:**
  - May be necessary for data held by private companies

- **Data/data portal access prerequisites:**
  - Making formal applications
  - Training attendance
  - Institutional approval

- **Specialist skills or knowledge:**
  - Statistical packages
  - Statistical expertise

WISERD

Data portals: What information do they hold?

- **By theme, including:**
  - Demographic
  - Socio-economic
  - Economic, employment, work and industry
  - Education
  - Health
  - Housing
  - Transport
  - Language and identity
  - Community safety and crime
  - Agriculture and the environment

WISERD

Data portals: What information do they hold?

- **By type, including:**
  - Enumeration: Census, vital statistics
  - Survey: Cross-sectional
  - Survey: Longitudinal / panel
  - Survey: Qualitative interviews
  - Geographic boundaries and identifiers

  - Administrative / Routinely collected data
  - Sensor / automated data (e.g. environmental observations)
  - Crowdsourced (e.g. twitter)

W/SERD

Data portals: Choosing a secondary dataset

- **Trade-offs: by time, by place, by theme**
  - As granularity increases, the number of available datasets often decreases.
  - E.g. Seeking data for **England** by **year** on **demographic characteristics** will yield more datasets than looking for data for **Cardiff** by **month** on **student numbers.**

- **Addressing trade-offs:**
  - "New forms of (BIG) data" collected at record level (e.g. administrative data / sensor data)
  - Data linkage - enriching existing data sources by linking on additional information

WISERD

Any questions or comments?

Part 1: Using data portals as data discovery tools for research

An overview of the key data portals providing data for the UK.

UK data portals: Case Study 1: National Statistical Portals

- **Organisation(s):** UK, Scottish, Northern Irish and Welsh Governments
- **Key themes:** Agriculture and Environment, Economic, Demographic, Community safety and crime, Education, Health, Housing, Taxation, Tourism, Transport
- **Functionality:** Searchable meta-data catalogue, data downloads, data visualisations
- **Access requirements:** Open Access
- **Expertise required**: Basic
- **How to access:**
  - UK: https://www.ons.gov.uk/
  - Scotland: https://statistics.gov.scot/
  - Northern Ireland https://www.nisra.gov.uk/
  - Wales: https://statswales.gov.wales/

W/SERD

UK data portals: Case Study 1: ONS Portal

UK data portals: Case Study 1: ONS Portal

UK data portals: Case Study 1: ONS Portal

UK data portals: Case Study 1: ONS Portal

2. UK population growth

UK population growth slowest since 2004

In mid-2019, the population of the UK was 66.8 million (66,796,807 with a confidence interval of plus or minus 0.2%). Over the year to mid-2019, the population of the UK increased by 0.5%, or 361,000 people.

Decreasing numbers of births and net international migration have resulted in the slowest rate of growth that the UK has seen in 15 years, returning it to the level seen in mid-2004. Despite population growth slowing, Figure 1 shows this was the 37th consecutive year (since 1982) that the total UK population has increased.

Figure 1: UK population continues to grow in mid-2019, but at a slower rate than any year since mid-2004

UK, 1953 to 2019

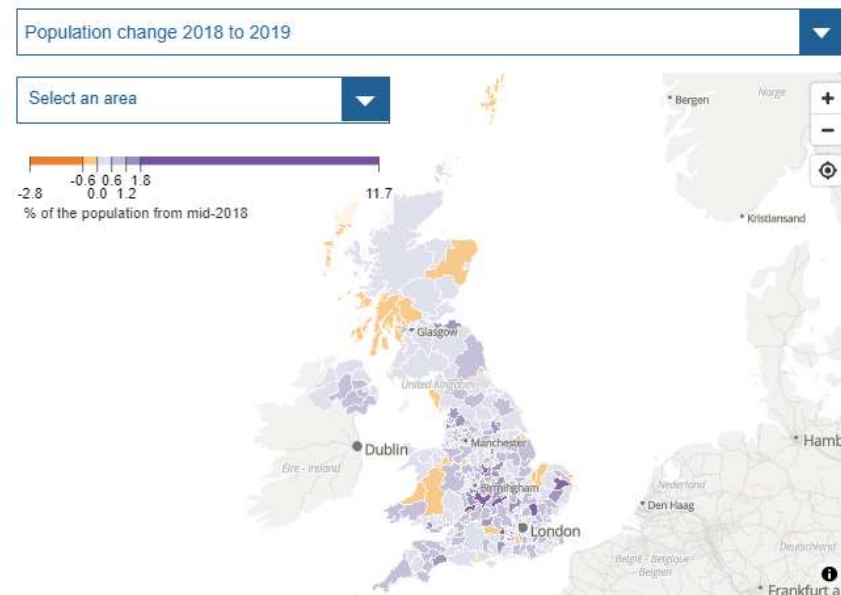# UK data portals: Case Study 1: ONS Portal

6. Local population change

Population increases in 86% of local authorities

Changes in local populations can be driven by international migration, internal migration, births and deaths. Taking account of each of these, in the year to mid-2019, the population increased in 326 (86%) of local authorities across the UK.

The interactive map in Figure 5 shows the overall change in population and components of change between mid-2018 and mid-2019 for each local authority in the UK. To use this interactive tool, select a component of change from the drop-down menu to view its impact on the mid-2019 population estimates at local authority level.

Figure 5: Population change and components of change, mid-2018 to mid-2019, local authorities in the UK

UK data portals: Case Study 1: ONS Portal

UK data portals: Case Study 1: ONS Portal

UK data portals: Case Study 1: ONS Portal

Office for National Statistics — Hierarchical Representation of UK Statistical Geographies
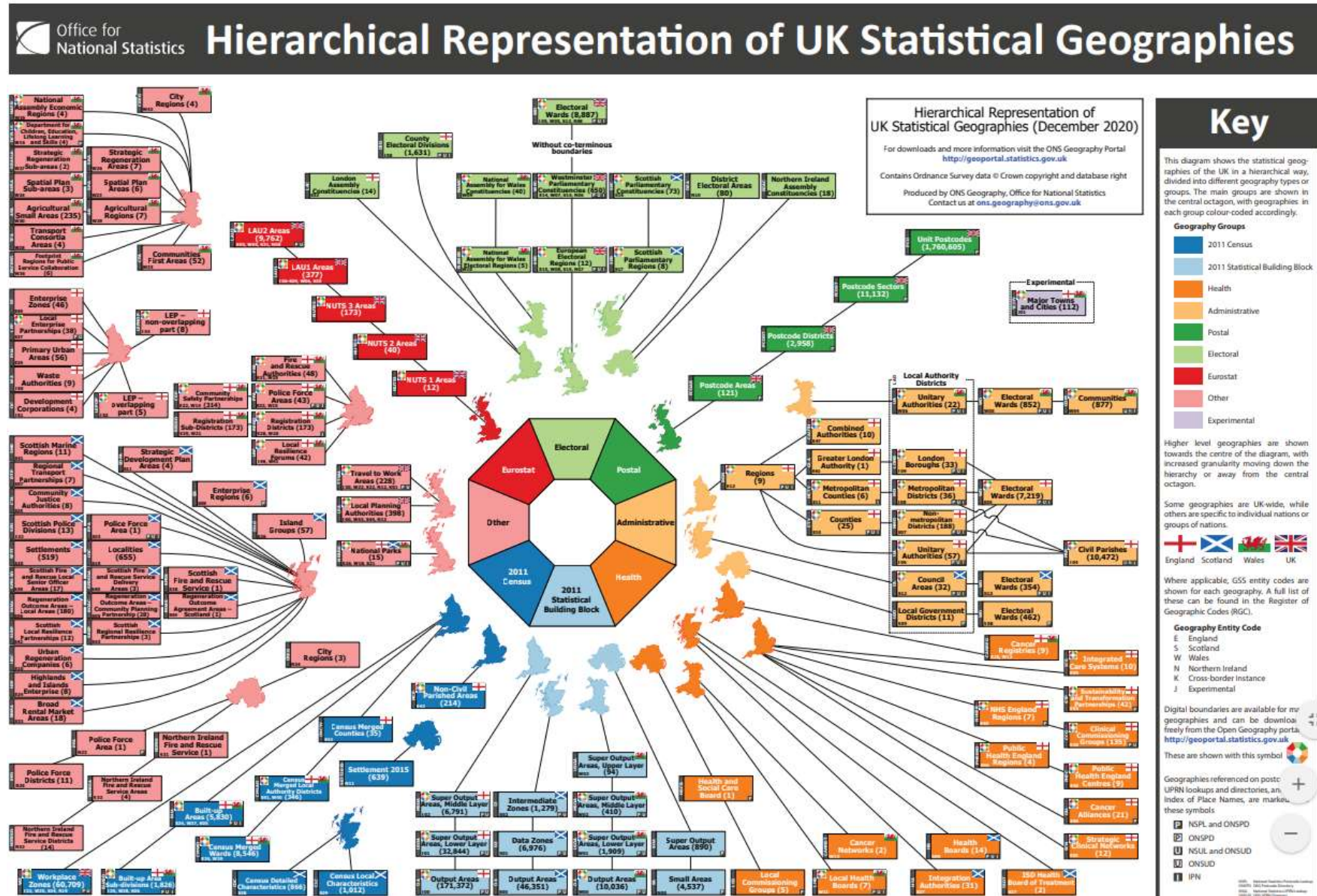
UK data portals: Case Study 2: UK Data Archive

- **Organisation(s):** UK Data Service (and associated institutions)
- **Key themes:** Social science data
- **Functionality:** Searchable meta-data catalogue, data downloads
- **Access requirements:** Catalogue is open access, data may have licensing restrictions
- **Expertise required**: Basic
- **How to access:**
  - UK Data Archive: https://beta.ukdataservice.ac.uk/datacatalogue/

UK data portals: Case Study 2: UK Data Archive

UK data portals: Case Study 2: UK Data Archive

Studies

Series

http://doi.org/10.5255/UKDA-SN-854030

Copy study DOI

Young people and the EU referendum 2016

Details | Access data

Details

Title:	Young people and the EU referendum 2016
Study number (SN):	854030
Access:	These data are safeguarded
Persistent identifier:	10.5255/UKDA-SN-854030
Principal investigator(s):	Pearce, S, Cardiff University Fox, S, Brunel University London

Sponsors and contributors

Depositor(s):	Cardiff WISERD, Cardiff University
Sponsor(s):	Economic and Social Research Council
Grant number:	ES/L009099/1

Topics

Thesaurus search on keywords

Abstract

Coverage and methodology

Citation and copyright

The citation for this study is:

Pearce, S., Fox, S. (2020). *Young people and the EU referendum 2016.* [data collection]. UK Data Service. SN: 854030, http://doi.org/10.5255/UKDA-SN-854030

Select citation format:

APA

XML citation formats:

CSL | EndNote

WISERD

UK data portals: Case Study 2: UK Data Archive

DATA DESCRIPTION (ABSTRACT)

This project conducted a cross-sectional survey of attitudes regarding the EU Referendum, the campaigns and political participation. Data were collected before the EU Referendum (March 2016), and after the referendum (June 2016). Sample sizes are as follows: March 2016 (5,120) and June 2016 (5,101).

Data creators:

Creator Name	Email	Affiliation	ORCID (as URL)
Pearce, Sioned	PearceS11@cardiff.ac.uk	Cardiff University	Unspecified
Fox, Stuart	Stuart.Fox@Brunel.ac.uk	Brunel University London	Unspecified

Sponsors: Economic and Social Research Council

Grant reference: ES/L009099/1

Topic classification: Politics

Keywords: YOUTH, POLITICS, EUROPEAN UNION MEMBER STATES, EUROPEAN UNION MEMBERSHIP, REFERENDUMS, YOUNG ADULTS

Project title: WISERD/Civil Society

Grant holders: Ian Jones, Rhys Davies, Christopher Taylor, Stephen Mark Davies, Philip Murphy, Elin Royles, Gary Higgs, Sally Power, Gareth Rees, Alison Green

Project dates:

From	To
1 October 2014	30 September 2019

Date published: 07 Feb 2020 14:00

Last modified: 07 Feb 2020 14:01

+ Coverage and Methodology

+ Access and Administration

AVAILABLE FILES

Data

+ 854030_June2016_YoungPeopleEUReferendum.sav

+ 854030_March2016_YoungPeopleEUReferendum.sav

Documentation

+ 854030_June2016_Codebook.xlsx

+ 854030_March2016_Codebook.xlsx

Read me

+ 854030_ReadMe.docx

WISERD

UK data portals: Case Study 2: UK Data Archive

UK data portals: Case Study 3: ADR-UK

- **Organisation(s):** Economic and Social Research Council (and associated institutions)
- **Key themes:** Administrative data
- **Functionality:** Searchable meta-data catalogue
- **Access requirements:** Controlled
- **Expertise required**: Expert
- **How to access:**
  - ADR-UK: https://www.adruk.org/

UK data portals: Case Study 3: ADR-UK

CATEGORIES

- ▲ SAIL Databank - (40)
  - Family Justice - (2)
  - **Administrative - (7)**
  - Health - (31)

ADMINISTRATIVE - (7)

The SAIL Databank brings together administrative datasets which include anonymised person level data on people who come into contact with government and public services. These organisations routinely gather information, when registering people or carrying out transactions, or for record keeping – usually when delivering a service. Linking administrative datasets enables researchers to study how different aspects of life interact with one another. These include national datasets (such as education records) and locality based datasets (such as from a local authority).

Show [10 ▼] entries   Search: [_____]

CAREERS WALES

Details of Careers Wales clients aged from 14, their school, employment status, qualifications and the interactions they have had with Careers Wales (2012-2019).

CENSUS 2011 - WELSH RECORDS (NEW)

All 2011 Census data for 'Welsh' records defined as those: · Currently resident in Wales · With a second address in Wales · With a previous Years Address in Wales · With a term-time address in Wales · Who work in Wales (but live in England) · In Armed Forces Establishments in Wales · Who are visitors in...

CHILDREN RECEIVING CARE AND SUPPORT (NEW)

The children receiving care and support census is based on the definition of eligible children who have care and support i.e children (under the age of 18) who have a care and support plan.

HIGHER EDUCATION STATISTICS AGENCY (HESA) STUDENT DATA

WISERD

UK data portals: Case Study 3: ADR-UK

- **Organisation(s):** Economic and Social Research Council (and associated institutions)
- **Key themes:** Administrative data
- **Functionality:** Searchable meta-data catalogue
- **Access requirements:** Controlled
- **Expertise required**: Expert
- **How to access:**
  - ADR-UK: https://www.adruk.org/

WISERD's data portal catalogue

- **Key UK Data Portals**
- **Downloadable from the WISERD events web page:**
  - https://wiserd.ac.uk/events/ncrm-training-using-uk-data-portals-find-and-evaluate-data-social-science-research
- **Contains information on:**
  - How to access it
  - Themes of data/metadata
  - Details of prerequisites

Part 1: Summary

- Data are a key resource for research.
- Data portals have been developed to improve data discoverability and reuse.
- A range of common functionality.
- Some portals are more accessible than others.
- There's no single 'go to' portal for all purposes – WISERD's catalogue can help find the correct portal.
- Familiarisation with some portals is essential for academics.

WISERD

Any questions or comments?

Break

Research funded by

Workshop programme

Part 1: Using data portals as data discovery tools for research:

- The benefits of using data portals to discover secondary datasets.
- An overview of the key data portals providing data for the UK.

Break

Part 2: How to use a data portal to discover data for a research project:

- Overview of the WISERD DataPortal project.
- Introduction and demonstration of the WISERD DataPortal's key functions.

Part 3: WISERD DataPortal workshop

- Practical workshop on using the WISERD DataPortal and Q&A session

WISERD

Part 2: How to use a data portal to discover data for a research project

WISERD DataPortal: Project overview

- **Purpose:**
  - Improve the discoverability and reuse of secondary data in Wales.
  - Develop a standards compliant meta-data catalogue and question bank.
  - Provide links to data feeds.
  - Allow simple data visualisation (in the form of maps)

WISERD

WISERD DataPortal: Functionality

Three key functions:

1. **Metadata discovery**
   - Search for data which may be relevant to your project.

2. **Access remote data**
   - Interrogate and map secondary data.

3. **Map My Data**
   - Map your own data from an Excel spreadsheet.

WISERD

WISERD DataPortal: 1. Metadata discovery

- **Metadata overview:**
  - Descriptive information about other data.
  - Standards compliant and searchable.
  - Thematic, temporal and spatial information.
  - Response rates, categories and variable names.
  - Where the data held.
  - Visualize content via word clouds.

- **Quantitative (questionnaire surveys):**
  - 150+ surveys, 13,000+ questions/responses
  - Detailed: down to question-response level

- **Qualitative (interview transcripts):**
  - 150+ surveys, 13,000+ questions/responses
  - Detailed: down to question-response level

W/SERD

WISERD DataPortal: Meta-data examples

- **Government:**
  - Census (1991, 2001, 2011)
  - Labour Force Survey 2010
  - Workplace Employee Relations Survey (1998 – 2011)
  - Living in Wales / National Survey for Wales (2004 – 2015)
  - Welsh Health Survey (2003 – 2011)

- **Academic:**
  - WISERD Multi-cohort Study (2012 – 2017)
  - WISERD Knowing Localities Stakeholder interview transcripts (2009 – 2012)
  - Wales Rural Observatory survey (2004 – 2010)

- **Third sector:**
  - Shelter Cymru

WISERD DataPortal: Demo #1: Meta-data discovery

1. Scope out a research project on Welsh Language

2. Search meta-data to find out what data may be available

WISERD DataPortal: Demo #1: Meta-data discovery

WISERD DataPortal: Demo #1: Meta-data discovery

Section: Household 11

Question Metadata

Question | Response | Response Table | Previous Question | View Survey

Question

Question Text	Which of the following best describes your ability in Welsh?
Question Number	Section: Household 11
Thematic Groups	Housing Language Citzenship Identity
Thematic Tags	Language
Question Type	SINGLE Question
Variable ID	N/A
Question Themes	ASK IF RESPONDENT CAN SPEAK WELSH (CODE 1 FOR PERSON 1 AT Q10a) SHOWCARD H SINGLE CODE ONLY

WISERD

WISERD DataPortal: Demo #1: Meta-data discovery

Section: Household 11

Question Metadata

Question   Response   **Response Table**

Previous Question   View Survey

Response Table

Display [All ▾] records

Filter : [_____] [✖]

Category	Variable	Number
I can only say just a few words	4	4
I can only speak a little Welsh	3	3
I can speak a fair amount of Welsh	2	2
I'm fluent in Welsh	1	1

Showing 1 to 4 of 4 entries

Previous   1   Next

WISERD

WISERD DataPortal: Demo #1: Meta-data discovery

Living in Wales 2008 Household Questionnaire

Survey Metadata

Dublin Core | Survey | Questions | Visuals

Dublin Core
Title
Creators
Subjects
Description
Publishers

Field	Value
Contributors	Welsh Government Ipsos MORI GfK NOP
Type	Dataset
Format	text
Source : URL	https://discover.ukdataservice.ac.uk/catalogue/?sn=6351
Source : DOI	http://dx.doi.org/10.5255/UKDA-SN-6351-1
Language	English
Relation : Same Collection	Living in Wales 2008: Property Survey
Relation : Different Collection	National Survey for Wales Enumeration Grid 2010-2010 National Survey for Wales Individual Questionnaire 0910 Living in Wales 2004: Property Survey National Survey for Wales 2014-15 questionnaire National Survey for Wales 2013/2014 Questionnaire. National Survey for Wales: Questionnaire 2012 Living in Wales 2004: Household Questionnaire Living in Wales 2006: Household Questionnaire National Survey for Wales Household Questionnaire 2009-2010 Living in Wales 2007: Household Questionnaire Living in Wales 2005 Household Final Questionaire National Survey for Wales Adult Questionnaire Pilot 09-10
Coverage : Spatial	Unitary Authorities (Wales)
Coverage : Start Date	2008-01-01T00:00:00
Coverage : End Date	2008-12-31T00:00:00
Rights	Welsh Assembly Government
Created	2011-03-31T14:57:58
Updated	2016-08-01T18:58:46.249391

WISERD

WISERD DataPortal: Demo #1: Meta-data discovery

WISERD DataPortal: Demo #1: Meta-data discovery
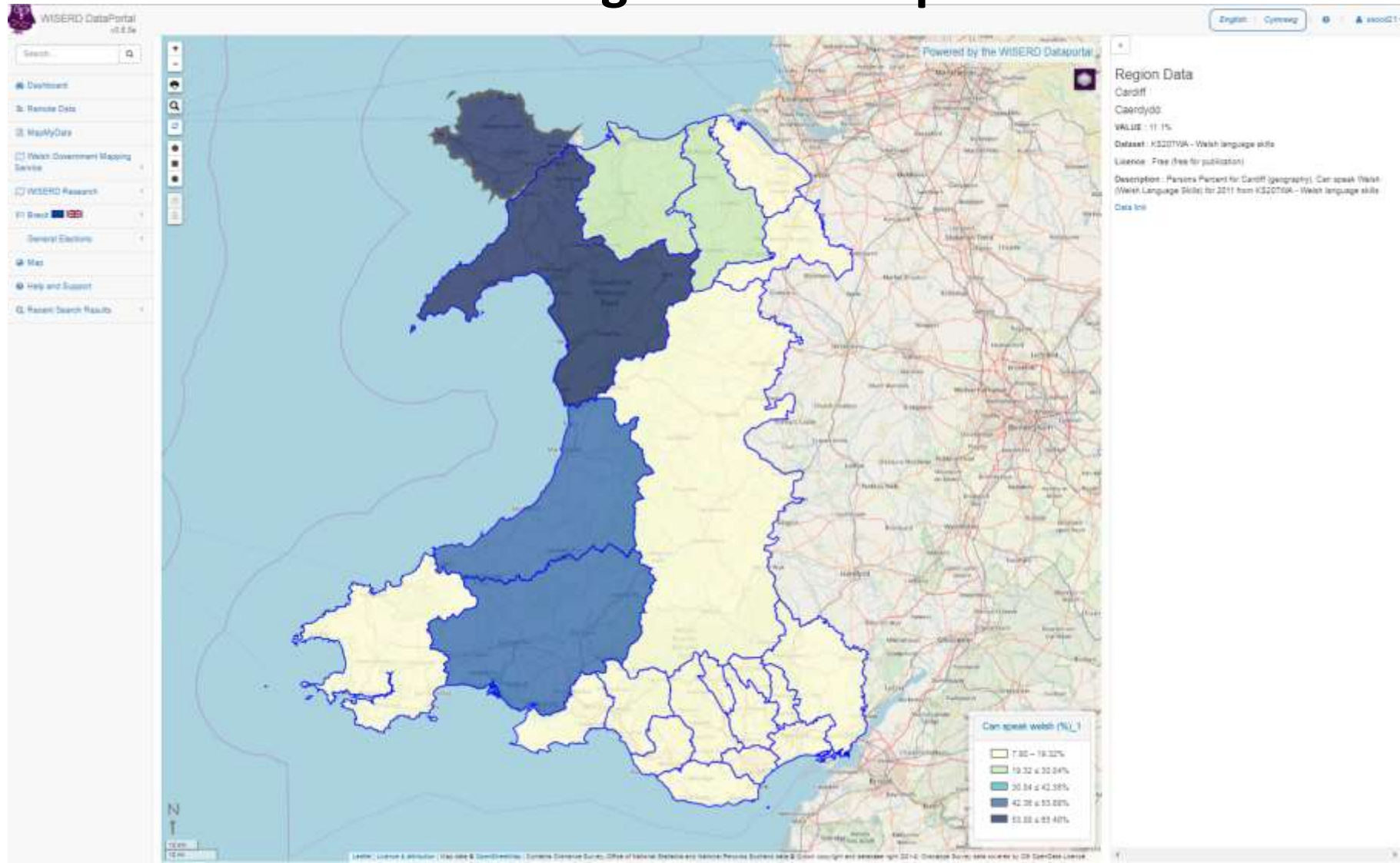
WISERD DataPortal: 2. Access Remote Data

- **Remote data:**
  - Secondary data made available through Application Programming Interfaces (APIs).
  - **Current source APIs**: NOMIS, Welsh Government Mapping Service

- **Datasets:**
  - Census, Annual Population Survey, Annual Survey for Hours and Earnings, Claimant Count, Life Events, Population Projections, WIMD, Service locations (schools, GP surgeries, hospitals, pharmacies etc).

- **Data visualisation (mapping):**
  - **Geographies:** LSOA, Unitary Authority, Parliamentary Constituency, Postal

WISERD

WISERD DataPortal: Demo #2: Remote data

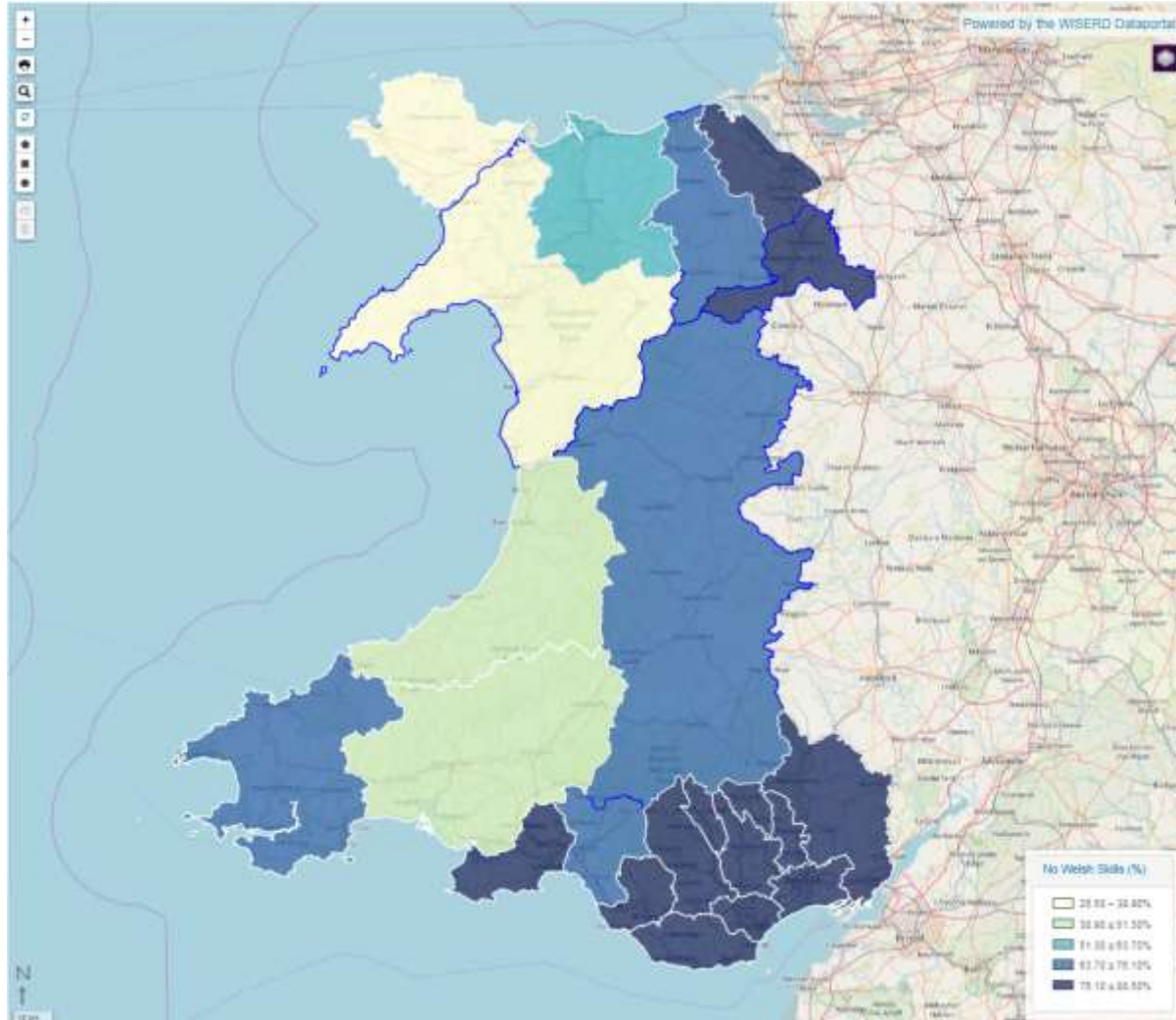1. Key statistics on Welsh speaking, reading and writing (KS207WA), Census 2011 (Source: NOMIS)

WI/SERD

WISERD DataPortal: Demo #2: Remote data

WISERD DataPortal: Demo #2: Remote data

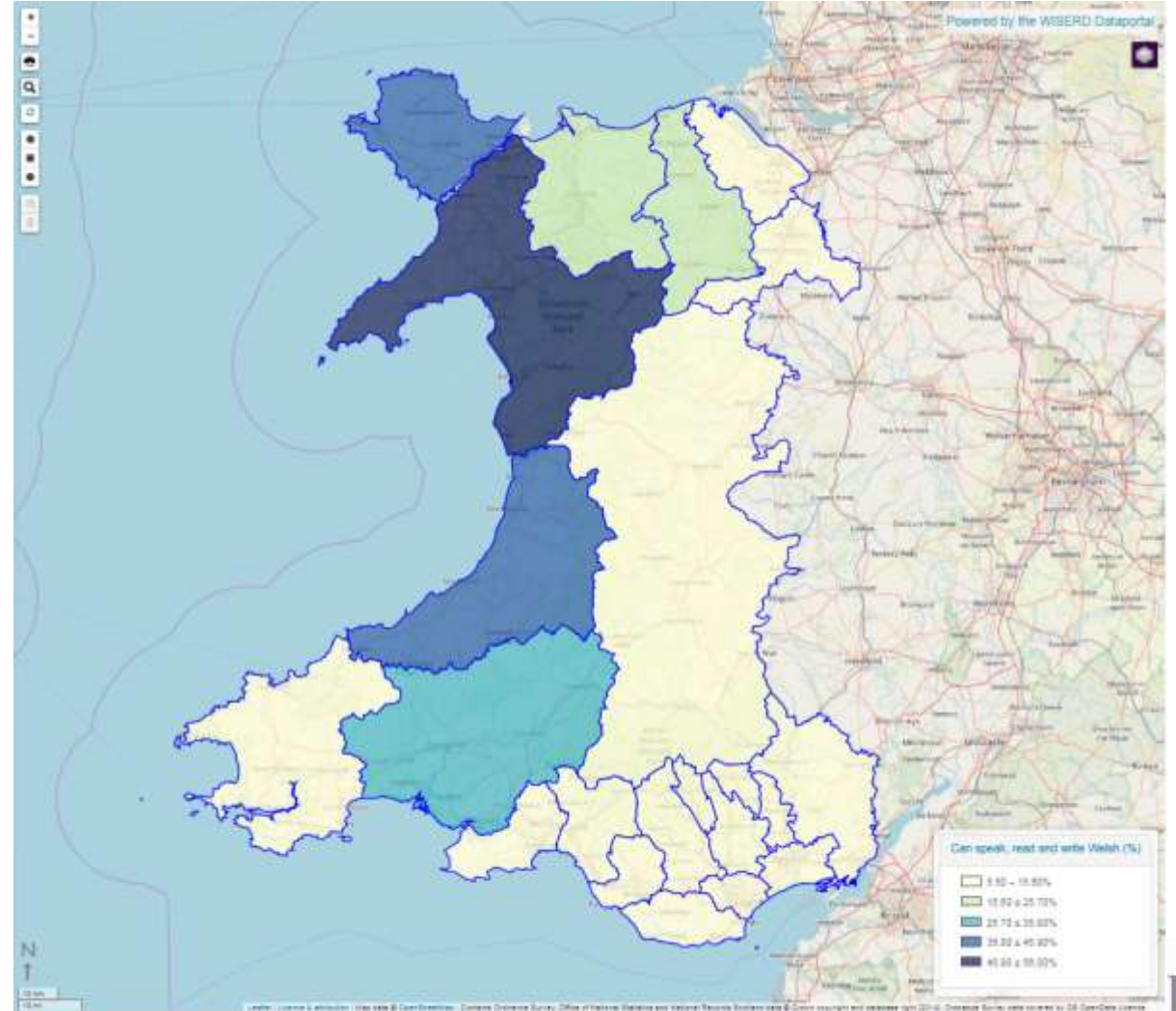**Percentage of Welsh speakers**

WISERD DataPortal: Demo #2: Remote data
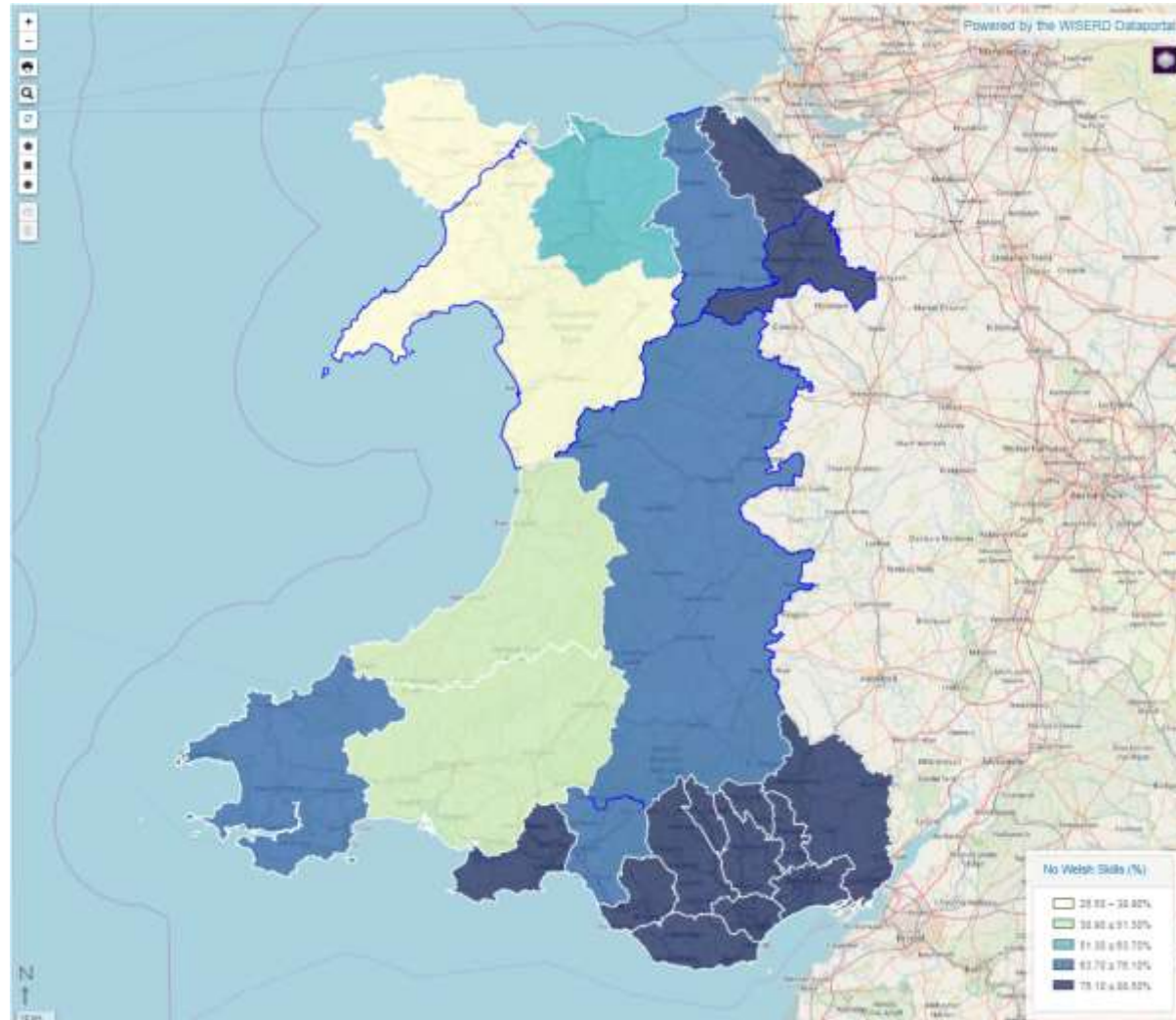
No Welsh skills (%)   **Speak, read and write Welsh (%)**
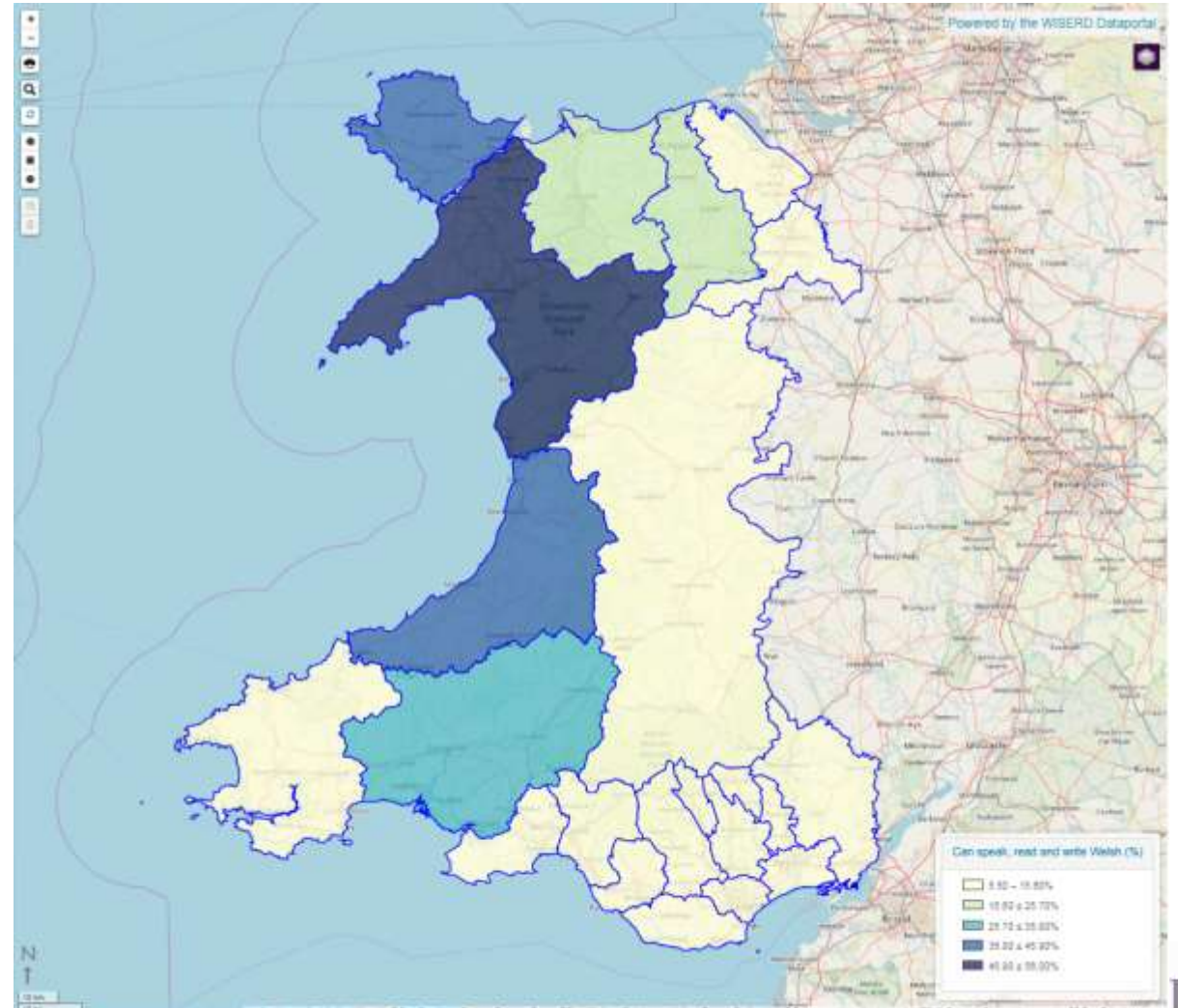
WISERD DataPortal: Demo #2: Remote data

No Welsh skills (%)

Speak, read and write Welsh (%)

WISERD DataPortal: 3. Map My Data

- **Map My Data:**
  - Import and map your own data from an Excel workbook/CSV file.
  - Secure - data does not leave your machine.
  - Point data as well as areal units.

- **Data mash-ups:**
  - Overlay your own data with remote data services (such as APIs).

W/SERD

Part 2: Summary

- **Overview of how the WISERD DataPortal could be used to:**
  - Scope out a research project on Welsh Language
  - Improve the discoverability and accessibility of data

- **Future developments:**
  - Importing WISERD's most recent datasets .
  - Updating the meta-data database.
  - Improving the catalogue functionality.
  - Improving the data visualisation element.

Any questions or comments?

Workshop programme

Part 1: Using data portals as data discovery tools for research:

- The benefits of using data portals to discover secondary datasets.
- An overview of the key data portals providing data for the UK.

Break

Part 2: How to use a data portal to discover data for a research project:

- Overview of the WISERD DataPortal project.
- Introduction and demonstration of the WISERD DataPortal's key functions.

Part 3: WISERD DataPortal workshop

- Practical workshop on using the WISERD DataPortal and Q&A session

WISERD

Part 3: WISERD DataPortal workshop

Available at: https://data.wiserd.ac.uk/

Any questions or comments?

Thank you

Scott Orford: OrfordS@cardiff.ac.uk

Katie Dickson: DicksonC5@cardiff.ac.uk

Samuel Jones: JonesS97@cardiff.ac.uk