



# Office for National Statistics (ONS) and Welsh Government joint work plan on coherence of Welsh language statistics

## Introduction

This document outlines the work that the Office for National Statistics (ONS) and the Welsh Government are planning to undertake during 2023-24 and beyond to improve our understanding of the main survey and administrative data sources used to produce statistics about the Welsh language.

This work will contribute to the continued publication of high quality, trusted and valuable statistics about the Welsh language by both the ONS and the Welsh Government (including those funded as part of [Administrative Data Research \(ADR\) Wales](#)).

## Background

The census collects data on Welsh language ability for people aged three or older in Wales. The Welsh Government considers the census to be the authoritative source of information about the Welsh language ability of the population in Wales. [Data on Welsh language from the 2021 Census](#) was published in December 2022.

The Welsh Government's Welsh language strategy, [Cymraeg 2050: A million Welsh speakers](#), outlines its aim of a million Welsh speakers by 2050. The strategy states clearly that the census will be used to monitor progress against this aim. The statistical trajectory to a million Welsh speakers developed and published alongside the strategy is also based on census data.

The census is also used by the Welsh Government to measure progress of the [national well-being indicator and milestone on the number of Welsh speakers](#).

The Welsh Government publishes a range of other statistics about the Welsh language from various data sources, from both surveys and administrative data sources. For example:

- The Welsh Government's [National Survey for Wales](#) provides statistics about the Welsh language ability of adults aged 16 years or older on an annual

basis. It also includes statistics about how often Welsh speakers speak the language, and how fluent they consider themselves to be.

- The ONS' [Annual Population Survey](#) (APS) provides statistics about the Welsh language ability of people aged three years or older. It also includes statistics about how often Welsh speakers speak the language. These statistics are available every three months.

While these data sources should not be compared directly with census data due to differences in the way that the information is collected for example, they are used to monitor trends in Welsh language ability between censuses.

## **Census 2021**

The ONS is responsible for the census in Wales and in England. A question about Welsh language ability has been included in the census in Wales since 1891.

Census 2021 estimated that there were over 538,000 people aged three years or older living in Wales able to speak Welsh, or 17.8% of the population on Census Day (21 March 2021). This was a decrease of almost 24,000 people since Census 2011, or 1.2 percentage points lower than Census 2011.

Meanwhile, the APS estimated that there were around 884,000 people aged three years or older able to speak Welsh in the year ending 31 March 2021, or 29.2% of the population. This was an increase of almost 134,000 people since the year ending 31 March 2011, or 3.4 percentage points higher.

Differences in the estimates of Welsh language ability between the census and household surveys such as the APS are longstanding, and both the [ONS](#) and the [Welsh Government](#) have explored possible reasons for some of these differences in the past.

While household surveys typically provide us with higher estimates of Welsh-speaking ability, this is the first time that the census has estimated declining numbers of Welsh speakers and the APS has estimated increasing numbers of Welsh speakers.

Following the publication of Census 2021, we are prioritising work to examine the differences between these data sources in more detail. This work includes the exploration of innovative approaches such as data linking (via the new cross-government Integrated Data Service (IDS); the Secure Anonymised Information Linkage Databank (SAIL)), to ensure we have a coherent evidence base which can be used for decision making.

While this workplan progresses, the Welsh Government will continue to publish the latest quarterly Welsh language estimates from the Annual Population Survey, providing clear advice to users on how these should be used and interpreted in light of the estimates from Census 2021.

## Workplan

The high level workplan set out in this section is intended to explore the divergence that has been identified in reported Welsh language skills. The workplan below will allow for:

- A fuller and more comprehensive understanding of the reasons for this divergence.
- A clear assessment of the quality of different data sources and how this might be improved.
- Closer collaboration between ONS and Welsh Government to further the respective strategic aims of [Statistics for the Public Good](#), and [Cymraeg 2050: A million Welsh speakers](#).
- A showcase for the benefits and value of the [Integrated Data Service \(IDS\)](#) and [Secure Anonymised Information Linkage \(SAIL\) Databank](#).

In order to improve our understanding of the main survey and administrative data sources that are used to produce statistics about the Welsh language, the ONS and the Welsh Government will address the following three main research questions.

- What are the differences in the estimates of Welsh language ability between the main data sources about the Welsh language?
- Why are they different, and why are the differences between these data sources changing over time?
- How can we improve the collection of information about Welsh language ability that will support public policymaking about the Welsh language for future generations?

The sections below outline the work that both the ONS and the Welsh Government intend to undertake to address each of these research questions in turn. More detailed delivery plans will be developed for each of the projects, but where possible interim steps have been set out for some projects in this plan.

**What are the differences in the estimates of Welsh language ability between the main data sources about the Welsh language?**

Project number	Research question	Lead(s)	Detail of project proposal	Timescales
1	What are the differences between Census 2021 and ONS household surveys in relation to the estimates of Welsh language ability?	ONS and the Welsh Government (WG)	<p>This project will use the ONS' Integrated Data Service to analyse a dataset that links Census 2021 data with the Labour Force Survey (LFS) at the individual-level. This will allow us to understand the characteristics of people who report their Welsh language ability differently between Census 2021 and the LFS. If the sample proves to be too small for meaningful results, this project may be extended to link Census data to the APS (a boosted version of the LFS) which includes additional respondents, specifically to provide better estimates for local areas.</p> <p>Interim steps</p> <ul style="list-style-type: none"> <li>• Undertake an exploratory project to agree requirements for variables and process for accessing IDS. (Complete)</li> <li>• Submit application for project approval to UKSA Research Accreditation Panel by early May</li> <li>• Undertake research to assess whether LFS sample size is sufficient to allow for meaningful analysis.</li> <li>• Undertake analysis and produce summary report on the characteristics of</li> </ul>	August 2023

			people who report their Welsh language ability differently between Census 2021 and the LFS	
2	What are the differences between the Pupil-Level Annual School Census and other sources in relation to the estimates of Welsh language ability of children aged 5 to 15 years old in local authority maintained schools?	WG	<p>This project will initially explore the feasibility of linking Census 2021 data with Pupil-Level Annual School Census data at the individual-level using the Secure Anonymised Information Linkage (SAIL) databank. The project could also be expanded to other Welsh language sources within SAIL such as historical teacher assessments, GCSE and A level entries and data from the Lifelong Learning Wales Records (LLWR).</p> <p>Subject to the outcome of project 1, it could also be expanded to include linkage with the APS.</p> <p>Interim steps</p> <ul style="list-style-type: none"> <li>• Submit application for project approval to SAIL Information Governance Review Panel by August.</li> <li>• Undertake analysis and produce summary report on the characteristics of people who report their Welsh language ability differently between PLASC, Census 2021 and other sources included in the project scope.</li> </ul>	January 2024
3	What are the differences between the National Survey for Wales and other sources in relation to the	WG	This project will initially explore the feasibility of linking Census 2021 data with National Survey for Wales data at the individual-level using the	Work to follow from project 2

	estimates of Welsh language ability of adults aged 16 years or older?		Secure Anonymised Information Linkage (SAIL) databank. Subject to the outcome of project 1, it could also be expanded to include linkage with the APS	
--	---	--	---	--

## Why are they different, and why are the differences between these data sources changing over time?

Project number	Research question	Lead(s)	Detail of project proposal	Timescales
4	What are the survey mode and survey design effects on the collection of information about Welsh language ability?	ONS and WG	<p>Both the ONS and the WG have previously looked at the possible survey mode and survey design effects on the collection of information about Welsh language ability.</p> <p>This project will bring together these various strands of research and will also include evidence from question design and testing from Census 2021 and Census 2011, as well as evidence from census quality surveys and other household surveys.</p> <p>Areas to explore include:</p> <ul style="list-style-type: none"> <li>• differences between interviewer-led questionnaires and self-response questionnaires</li> <li>• differences between paper, online and telephone collection</li> <li>• differences between statutory and non-statutory surveys</li> <li>• impact of proxy responses</li> <li>• order of questions in questionnaires</li> </ul> <p>Interim steps</p> <ul style="list-style-type: none"> <li>• review and summarise evidence from Census 2011 and Census 2021 development and testing.</li> </ul>	August 2023

			<ul style="list-style-type: none"> <li>• review and summarise wider academic and international literature</li> <li>• produce a report of key findings.</li> </ul>	
<b>5</b>	Why do the differences between the data sources about Welsh language ability appear to be changing over time?	ONS and WG	<p>This project will use findings from each of the previous projects.</p> <p>Response rates for household surveys are decreasing in general, do we know whether these decreasing survey response rates are having an impact on Welsh language changes over time or creating a bias within the data?</p>	October 2023



**How can we improve the collection of information about Welsh language ability that will support public policymaking about the Welsh language for future generations?**

Project number	Research question	Lead(s)	Detail of project proposal	Timescales
6	What data already exists to understand more about people who are able to speak Welsh but live outside Wales?	ONS	<p>This project would entail documenting the needs for this data, for public policy making, a need which is growing with a more transient population.</p> <p>A scoping exercise would be required to look at different options or methods for collecting this data and improving coherence of the data on this topic. The outcome of this project will contribute to work already underway at ONS on the transformation of the LFS.</p> <p>Interim steps</p> <ul style="list-style-type: none"> <li>• document the user requirements for this data.</li> <li>• review data currently collected on people able to speak Welsh who live outside Wales.</li> <li>• consider options for collecting such data with a recommendation on the preferred way forward.</li> </ul>	August 2023

## **Governance and reporting arrangements**

1. A steering group will be established, comprising Welsh Government and ONS officials, to oversee this work plan. The group will meet monthly to monitor progress. Projects may also establish their own task-and-finish groups. The task force will provide regular updates, as required, to Welsh Government Ministers and the National Statistician.
2. Progress and findings of the projects will be published through a combination of formal statistical or research publications and informal information sharing such as blog posts.

## **Contacts for Welsh Government statistics**

### **General statistics enquiries**

Email: [stats.info.desk@gov.wales](mailto:stats.info.desk@gov.wales)

### **Subject specific**

[Contacts for statistics](#)

**Twitter** 

[StatisticsWales](#)

### **Welsh Government email newsletter and alert service**

[News and alerts](#)

### **Admin queries**

Email [kasadmin@gov.wales](mailto:kasadmin@gov.wales)