



Llywodraeth Cymru
Welsh Government

PUBLICATION

Welsh linguistic infrastructure policy

Our policy on how we will coordinate and develop Welsh linguistic infrastructure.

First published: 27 June 2023

Last updated: 27 June 2023

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html> for the latest version.

Get [information on copyright](#).

Contents

Ministerial Foreword

Glossary

Executive summary

Introduction: why have the responsibility for coordinating these elements in one place?

Areas of development

Epilogue

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html> for the latest version.

Get [information on copyright](#).

Ministerial Foreword

When we speak of ‘linguistic infrastructure’, we mean the resources which help us use the Welsh language from day to day, such as dictionaries, terminology resources and corpora, and all the research and standardization work that enables these resources to grow and develop.

There are several resources available but, rather than making it easier to find an answer, there are times when that becomes a problem. It means that it’s not always obvious which resource is the most appropriate one, and we find ourselves jumping from one website to another to find a word or term. And then not every resource will have the answer to every query, or an answer in one place will be different to an answer in another. I know that some people are increasingly turning to websites like Bing and Google Translate, and although these can be useful in some situations, the answers we get from them aren’t always reliable.

We have many projects of the highest calibre here in Wales. Work on Geiriadur Prifysgol Cymru (the standard historical Welsh dictionary), for instance, started in 1921, and is by now the basis for all other work on Welsh words and terms. Other resources, such as Bangor University’s Porth Termau, and the Welsh Government Translation Service’s BydTermCymru, have arisen more recently in response to gaps in the provision and the demand for up to date terms. But this area of work remains reactive, with terms only commissioned if there’s great demand for them, and no one keeping a high level, strategic overview of future needs.

The question arises, therefore – is this good enough? That’s what we asked in our consultation on the draft version of this Welsh Linguistic Infrastructure Policy. We received 88 responses, and I’d like to thank everyone who contributed. We’ve taken those responses into consideration when drafting this final policy.

We've already started taking some of the things in this policy forward. The main one, perhaps, is setting up a new unit in the Welsh Government to be responsible for linguistic infrastructure. The unit will have a role that extends across linguistic infrastructure resources, as well as policy areas within the Welsh Government.

Coordinating resources is essential as wider Welsh language policy continues to develop. For instance, we've recently consulted on a White Paper which includes proposals that will form the basis of a programme of work which includes the Welsh Language Education Bill. At the core of the White Paper proposals is improving linguistic outcomes for learners between the ages of 3 and 16, but it also suggests expanding the role of the National Centre for Learning Welsh to be a specialist organisation supporting the acquisition and learning of Welsh for learners of all ages in Wales. Of course, ensuring that coordinated, easy to use resources are available to learners of all ages, as well as teachers, school pupils and parents, is essential to the success of these proposals, and the success of the Curriculum for Wales in its entirety.

We've also established a new company called Adnodd, that will maintain an overview of resources for learning and teaching, and will commission appropriate resources for the Curriculum for Wales and the new qualifications. These will be suitable to be used by teachers and learners in the classroom, as well as in the home for independent study and revision. These education resources need consistent terminology so that they can be published simultaneously in Welsh and in English. The relationship between the unit, the Language Technologies Unit at Bangor University who are responsible for the Termiadur Addysg (Education Terminology), and Adnodd will be vital in this regard.

The unit will also have one additional responsibility to what was proposed in the draft policy, namely our commitment to safeguard Welsh place names, which derives from our Programme for Government and the Cooperation Agreement with Plaid Cymru. Recent announcements by Eryri and Bannau Brycheiniog

National Parks concerning the use of Welsh names for their organisations are positive statements in terms of the status of Welsh place names. The unit outlined in this policy will focus on house names, the names of topographical features, and historic names. Our initial steps in this regard are outlined in our Welsh Language Communities Housing Plan, and we'll announce more detailed steps on the basis of research that will be completed by the end of the year.

Welsh isn't the only language facing such challenges. Many languages across Europe – be they minority languages or major world languages – face the issues described in this document, and are exploring how to solve them. In that regard, we aren't unique. But it's fair to say that the main gap in Wales is the element of coordination of work and resources, and the part that government agencies play in this respect, so that we can more strategically disseminate and give access to linguistic resources. By doing that better, we will succeed in improving the provision for everyone who wants to use Welsh.

Jeremy Miles MS
Minister for Education and Welsh Language

Glossary

Corpus (plural: corpora)

A large collection of linguistic data (for example, words, terms, phrases) assembled following certain principles. Corpora are used to create digital resources and to conduct linguistic research in order to develop terminology databases or dictionaries. They can be based on sound recordings, written texts or human gestures (for example, sign language).

Dictionary

Traditional dictionaries usually describe the language and record its vocabulary. They can also suggest synonyms. This is different to terminology, where a term corresponds to one concept only.

Term

Terms are words or phrases which have specific meanings in specific contexts or areas. These meanings can be different to the meaning given to them in everyday language.

Terminology

A collection of terms relevant to specific subject areas (for example, health, education). A terminology shows which terms are the standard ones to use in a specific context, for example, when writing for a specific sector or customer. Terminology work is different to that on a dictionary, as it means studying concepts, conceptual systems and their labels, rather than studying words and their meanings, as happens when developing a dictionary.

Executive summary

During the summer of 2017, we published a Welsh language strategy called **Cymraeg 2050: A million Welsh speakers** which introduced a plan for doubling the number of Welsh speakers to a million and increasing the use of Welsh. The strategy entailed a significant change to Welsh language policy.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html> for the latest version.

Get [information on copyright](#).

In the strategy we stated that, in order to reach the target of a million speakers, we need more people to speak Welsh and more people to use it. But we also said that we need to create favourable conditions (things like linguistic resources, a prosperous economy, or ensuring people's support for the language). Dictionaries, terminologies and corpora (things we refer to here collectively as 'linguistic infrastructure') are an extremely important part of this.

As increasing the number of speakers means including everyone, from all backgrounds, we want to see these different resources being coordinated effectively so that they are consistent, fit well together, and are easy to find. We want to do this so that everyone can use them easily wherever they are on their journey with the Welsh language – be they new speakers, school-children and their parents, students, people who use Welsh in the workplace, professional translators, social care and health professionals, lawyers, the police and emergency services, journalists, teachers, and many more.

Our priorities

With so many resources available to people who want to use Welsh, people continue to say they don't know where to look for a word or term. The truth is that these resources are scattered around (some of them on websites or apps, and some in the form of books) and people don't always know where to find answers or guidance. That's possibly why many of the consultation responses on our draft policy supported the idea of creating a website to include information about the different resources available, and which would allow users to search through all the resources in one place. We therefore want to create a new and better way of coordinating the main resources and providing them in one place, so that everyone can easily find answers.

With this in mind, this document outlines our policy for the following six areas of work:

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html> for the latest version.

Get [information on copyright](#).

1. Dictionaries
2. Terminology
3. Corpora
4. Standardisation (for example, orthography, spelling, place-names, terms)
5. A central website
6. A new unit to coordinate Welsh linguistic infrastructure

Our aim is to make sure that all the elements work together smoothly, in order to make it easier for more people than ever to use Welsh.

Since we published the draft policy, we have started to implement some of the initial steps in it. These include:

- creating a new unit in the Welsh Government to begin the work of making the different elements work better together
- taking initial steps to create an easily accessible, easy to use website that will be available to everyone who wants to use the Welsh language – a website that explains and offers easy access to resources like the main dictionaries and termbases, as well as resources like corpora, all in one place
- establishing a Welsh Language Standardisation Panel to begin to solve linguistic problems, focusing at the outset on longstanding orthographical problems
- alongside Bangor University, coordinating the process of standardising terminology in the field of race and ethnicity, consulting with key stakeholders and individuals to create a contemporary list of terminology in this area

Introduction: why have the responsibility for coordinating these elements in one place?

Making it easier for the user

People who want to use Welsh express a number of frustrations. One of them is that it can be difficult to find resources and answers. Another is that the number of terminology resources and dictionaries available can be confusing, and they find it difficult to know which one has priority. This suggests that people are unclear about where to find answers to linguistic questions. Creating one unit to be responsible for making the different resources work together better, and providing guidance on their use, and access to them through one website, is a way of getting to grips with this.

We will help to coordinate resources with the aim of providing more clarity for the user. In that regard, we will consider solutions and work towards a situation where the user can find many of the main linguistic resources (be they words or terms, corpora or linguistic advice) through one searchable portal.

Creating a consistent experience for the user

A number of nations and languages have official bodies or academies of different models, which are responsible for coordinating work on linguistic infrastructure, and which are an obvious starting point to anyone with a question about the language. This creates a unity of experience, where a national and historic institution gives users confidence that they're using an appropriate word or term.

In the case of Welsh, a person's experience will depend on what the user already knows about the resources which are available. For instance, a

professional translator will know which source, from among many, to choose; some Welsh speakers will have some knowledge, but will often be confused or unaware of a resource which fits the bill; and often new speakers or people who don't speak Welsh won't know where to start.

We will aim to create a consistent experience for everyone who uses Welsh linguistic resources. In order to achieve that, we will look again at the way resources are commissioned and coordinated, and market them with the aim of giving users confidence that the answers they find are authoritative.

Avoiding duplication and filling gaps in provision

Duplication can not only cause confusion, but also lead to inefficient use of resources and specialist skills.

At present, there are several dictionaries and terminology databases available with different people and organisations responsible for producing them, with no specific coordination between them. The Welsh Government already provides funding for several of these projects, so we see this as an opportunity to play a coordinating role in order to further develop provision.

Horizon scanning

One obvious gap in the present provision is the lack of strategic planning concerning future requirements. For instance, if it becomes evident that work is being planned in a new policy area, or a certain industry will be locating to Wales and Welsh terms will be needed for the workforce, who is responsible for ensuring that terms are standardised in time? We saw something similar happening recently with attention worldwide to the equalities agenda in relation to race and ethnicity resulting in highlighting gaps in standardised terminology in Welsh. This is the result of a systemic failure, as no single organisation in Wales

is responsible for having a strategic oversight of Welsh language terminology. If we look at the two main centres which standardise terms in Wales, the Welsh Government's Translation Service respond to Welsh Ministers' plans and use of terms through the BydTermCymru website, while Bangor University's Terminology and Language Technology Unit develops new terminology dictionaries for their Welsh National Terminology Portal, in tandem with the organisations who commission the work. These are their roles, and they do the work to an extremely high standard.

So the linguistic infrastructure unit within the Welsh Government (see Area 6) will be responsible for scanning the horizon for policy developments and large public projects which are in the pipeline, and for ensuring that appropriate experts are responsible for proactively producing and standardising the necessary terminology, without duplicating efforts. This is especially important in the case of technical terms, to ensure that they're conceptually correct and appropriate for specialist use, be that in education, science, medicine or any other field.

Responding to the need for urgent terminology

As well as planning for the long term, the unit will also establish a way of helping with high profile terms that arise without warning, but for which there is a pressing need. We want to reduce the number of situations where individuals and organisations have to urgently come up with new terms without wider consultation or guidance. That can lead to duplication, as different bodies or agencies, because of the pressing nature of their work, set out to create new terms in isolation, often leading to several organisations creating and disseminating a number of terms for the same concept. We believe that helping to coordinate terminology in cases such as these will lead to more consistency, and make it easier for users to recognise, understand and begin to use new terms.

Coordinating for the benefit of all

Although there is some collaboration between the organisations and teams responsible for different elements of Welsh linguistic infrastructure, there is no structure to ensure that developments and lessons learned from one area are shared with others as a matter of course.

Much of the present work on linguistic infrastructure is publicly funded, therefore establishing a new way of coordinating it will be a way of ensuring better value for money, with the aim of reducing duplication and maximising the funding, resources and expertise available. Ensuring sufficient and stable investment in this area of work, as well as ensuring that projects which have already received public funding remain current and valuable, will also be essential.

In the short term, one possible challenge may be a paucity of individuals with the requisite expertise to do the work – there are only so many people in Wales with experience of developing dictionaries and terminologies, for instance. It will be the unit's responsibility, therefore, to plan for the longer term, and to explore ways of fostering expertise and expanding the pool of experts who can do the work.

The unit will work with translation services of all kinds, as well as Cymdeithas Cyfieithwyr Cymru (the association of Welsh translators and interpreters). Close collaboration will ensure that the different areas are joined-up in an appropriate way, so that any developments in one area are considered in others, whilst maximising the benefits for all.

Areas of development

In this section, we outline the six areas we will develop in order to strengthen

and coordinate Welsh language linguistic infrastructure. We believe that doing this will help increase use of the language.

Area 1: Dictionaries

“What’s the Welsh word for...?” is a question we hear all the time. On the radio or television, in the classroom, on social media or in the workplace – questions about Welsh words are very common. And often, without anyone realising it, the answer already exists in a dictionary, usually within easy reach online and through apps.

It’s easy to take dictionaries for granted. They play an important role in making sure that information about the definition of words, and their translations, are available to their users. In the modern world, different kinds of dictionaries can record the historical use of a language, reflect every day usage, or provide for specific audiences, such as new speakers or children.

Safeguarding the future of dictionaries like these, and making sure they’re easily available to all, will remove barriers and enable more people to use Welsh. Dictionaries are also a strong foundation for developing all sorts of other useful linguistic resources for everyone who wants to use Welsh.

What resources are available now

There are several Welsh dictionaries available. Below we focus on those that are available online.

Geiriadur Prifysgol Cymru

This is the standard historical Welsh dictionary, originally prepared between 1921 and 2002 by the University of Wales, and which continues to be updated. It

is similar in status to the Oxford English Dictionary. This is the foundation for all other reference books for the Welsh language, for example, dictionaries, terminology work, grammar books, decisions on orthography (for example, how words are spelled) and so forth. This resource is also available in the form of an app.

The Welsh Academy English-Welsh Dictionary

An English-Welsh dictionary, edited by Bruce Griffiths and Dafydd Glyn Jones. Before its publication, there was no comprehensive dictionary for use by the translation profession. It has also been an extremely important resource to help people work through the medium of Welsh in bilingual organisations, and in the field of education. The Welsh Language Board, and more recently the Welsh Language Commissioner, have been responsible for a project to convert the dictionary to digital format and publish it online, and further work is needed to amend and update the contents. The contents have therefore been static to all intents and purposes since the second print edition in 2003 – a period of swift change for the language – and as a result there are many more recent words which have not been included, or which need to be updated.

Geiriadur Prifysgol Bangor

Another comprehensive Welsh-English and English-Welsh dictionary prepared by the Terminology and Language Technology Unit at Bangor University. It is especially useful to new speakers. It includes links to other dictionaries and technical terminology resources. It is also available free-of-charge on the app Geiriaduron app.

University of Wales Trinity Saint David's Welsh Dictionary

A comprehensive Welsh-English and English-Welsh dictionary, prepared by the Welsh Language Department at the University of Wales Trinity Saint David in Lampeter. This dictionary hasn't been updated for several years, but it remains

an extremely useful resource.

The Gweiadur

A Welsh-English and English-Welsh online dictionary, which includes definitions and pronunciations and is part of a suite of multimedia resources for Welsh language users of all ages and abilities. It can be used free-of-charge, but users need to register to do so. The website also includes useful resources such as a Verb Conjugator and a Thesaurus.

Other resources: we also need to acknowledge that many people use resources like Bing Microsoft Translator and Google Translate to find words or terms. These can be useful and easily accessible tools that facilitate the use of the Welsh language in some situations, but they don't always provide reliable answers.

Objectives

Foster consistency between different dictionaries, where appropriate, by encouraging and facilitating work to update them on the basis of the same research, so that they also enrich each other.

Making it easy for anyone to find out where to get the definition or translation of a Welsh word.

Ensuring that resources that are funded by the Welsh Government are regularly updated, so that they reflect Welsh as a living language.

Ensuring that a modern Welsh>English dictionary, which is regularly updated, is available for everyone to use digitally.

Making the most of the expertise that already exists, and further develop it.

How we will do it

- Help to plan consistent and complementary dictionary resources, so that people who use Welsh have access to comprehensive provision.
- As part of this, encourage resuming the work of amending and updating The Welsh Academy English-Welsh Dictionary.
- Ensure that users can access the main dictionaries through a central website (see area 5) which is marketed effectively so that people know where to turn when the question “What’s the Welsh word for...?” arises.
- Explore ways of developing the training available for those who wish to be lexicographers, in order to increase the number of people able to do the work.

Area 2: Terminology

Just like the words in a dictionary, we also depend on terms in our everyday lives – whether we realise it or not.

Terms are used to denote specific concepts in specific fields and, unlike words in more general usage, standardisation processes are required to ensure one term for each concept. In that regard, terminology dictionaries are different to descriptive dictionaries like the ones we speak of in Area 1. But they’re just as important.

For instance, we can’t use a computer without a mouse and keyboard, or inputting a password – these are all terms that are connected to the field of information technology. Mouse has more than one meaning, but by including it in terminology for the field of technology, we know that it refers to the tool that helps us work on a computer, rather than the animal. In collecting terms that are connected to one field together in one place, we create a list of terms or, in other words, a terminology. It’s essential that standardised sources of terminology

exist to make it easier to use Welsh in every aspect of day to day life, including in specialist fields like technology, law, health and education.

Work to standardise Welsh terms has been happening since the mid-twentieth century (for example, the volume *Termau Technegol*, University of Wales Press, 1950). In the field of education, work to standardise terms started in 1993 when the Schools Curriculum Authority awarded a contract to Bangor University to standardise Welsh terms for schools in Wales. This led to the creation of *Y Termiadur Addysg*, which continues to be an extremely important resource. But the need to create and standardise terms never goes away. Recently, for instance, we saw the need to update terminology in relation to race and ethnicity. By working with Bangor University and various language specialists, as well as members of different ethnic communities, a list of Welsh terms was created that's suitable for everyone to use. The aim is to make it easier for people to have important and valuable discussions confidently in Welsh about this important field.

Terminology dictionaries are based on concepts rather than words, and follow a standardisation process set out by the ISO/TC37 family of international standards. This means it can take a little more time to standardise terms and include them in terminology resources. It's important, therefore, that terminology work is coordinated and planned thoroughly to make sure there are no obvious gaps in the provision.

We will work to solve two main issues in relation to terminology.

1: that there are so many resources, some standardised and others not, meaning that people don't always know which ones to use, and which can occasionally lead to inconsistencies between different resources

Although there are many online terminology resources of the highest standard

available for Welsh – be they terminology databases by terminology experts, or online lists kept for ease of reference by different public bodies – it's evident that people don't always know where to look for a term, which terms have been standardised following recognised rules, or which resource is best for their needs.

As the work of standardising terms takes place in a number of contexts, by individuals and organisations, in order to express a technical concept, this can lead to a situation where a number of different sources suggest terms, some which have been standardised, and others that haven't. To those who work in a specific field, it can be obvious which source is the most reliable one, and which term to use. But what about the rest of us?

2: prioritising which technical terms to standardise, as well as a better process for creating new terms which are needed urgently and ensuring they are widely adopted

Broadly speaking, the work of standardising terms can be separated into two activities:

- Filling gaps in the language's vocabulary, for instance so that words and terms are available for new concepts as they arise.
- Standardising packages of technical terms, so that those terms are appropriate for specialist use in specific fields, for example, education, science, medicine, food hygiene.

Filling gaps in the language's vocabulary

From time to time, a new term will urgently need to be created, as a result of something that's happened in the news, for instance, or because new information needs to be given to the public. In Wales, as is the case for many

languages, this usually happens to try to convey in Welsh concepts that have initially been expressed in English. Recent examples have included terms concerning Brexit, medical terms because of the COVID-19 pandemic, and equality terms in the field of race and ethnicity.

Standardising packages of technical terms

On other occasions, terms won't be needed quite so urgently, but it's still obvious that a package of new terms will be needed in the near future, for example, if the Government intends to develop a new policy in a certain field, if a specific subject garners public attention (for example, equality in the field of race and ethnicity), or if a new industry comes to Wales and new Welsh terms are needed. With two large projects creating terms (the Terminology and Language Technology Unit at Bangor University, and BydTermCymru in the Welsh Government's Translation Service), and both working independently on their areas, the new unit will plan provision when a series of terms arise in a new field, to weigh up which sets of terminology it would be beneficial to prioritise.

In both cases, if more than one resource creates a different term for the same technical concept, there needs to be a way of deciding which one has priority, and of ensuring consistency between different resources where that is appropriate (we accept that having different terms are unavoidable in some situations due to variations with regard to audience or context).

What resources are available now

There are several sources of Welsh terminology, below we list the main ones:

Welsh National Terminology Portal

This terminology resource was developed by the Terminology and Language

Technology Unit at Bangor University. The Welsh National Terminology Portal includes collections of specialist terminology which have been standardised in collaboration with a number of specialist external bodies, for instance:

- Y Termiadur Addysg (terms for education)
- The Coleg Cymraeg Cenedlaethol's Terms for Higher Education
- Social Care Wales terms
- Justice Wales Network terms
- Edward Llwyd Society nature dictionaries
- Food Standards Agency terms

Some of these resources can also be accessed through Cysgeir (part of the Cysgliad software package) and the Ap Geiriaduron app.

TermCymru

A searchable collection of terms that Welsh Government translators use in their everyday work. It is regularly updated with a view to providing a comprehensive database of standardised terms which reflect current usage in different aspects of the Government's work. As well as being able to search the database, you can also download it in its entirety under open licence, which gives more flexibility than offering it in searchable form only.

Y Termiadur Addysg

A Welsh Government-sponsored resource, developed by the Terminology and Language Technology Unit at Bangor University, it provides standardised terms for the education sector. These are the standardised terms used in Welsh-medium exam papers and assessments, as well as resources of all sorts for teachers and students. This terminology can also be searched through the Welsh National Terminology Portal.

The Coleg Cymraeg Cenedlaethol's Terms for Higher Education

A database of terms for Welsh-medium Higher Education, standardised by the Terminology and Language Technology Unit at Bangor University. This is an online resource for students and staff to facilitate their studies, education and research through the medium of Welsh at University level. This terminology can also be searched through the Welsh National Terminology Portal, and through the Ap Geiriaduron app.

Other terminology collections: various other bodies also draw up and maintain their own lists of terms, occasionally publishing them online. In most cases, these won't have been standardised using recognised processes, or been fed into the resources above. In the absence of other sources, however, they can be very useful (examples include lists kept by Stonewall and National Museum Wales).

Objective

Ensuring that everyone can easily find a term, and have confidence that it's the right one for them.

How we will do it

What we have done

- Set up a new unit (see area 6) to coordinate terminology work, in order to avoid duplication and work towards more consistency. The unit will also be responsible for drawing up annual work programmes to prioritise terms to be standardised.

What we will do

- In the short term, we will develop a website to ensure that everyone knows where to turn to get answers about terminology. This will include marketing the provision effectively, raise awareness of the resources available, and offer advice about where best to find answers, depending on the user.
- Starting immediately, but coming to fruition in the long term, we will work on a solution to enable users to search through the main Welsh terminology databases (for example, BydTermCymru and Welsh National Terminology Portal) in one place, through an easy-to-use website (see area 5). This will mean that users don't have to search through a number of different websites for a suitable source of terminology, and give them confidence that the term suggested is approved and standardised.
- We will also provide links to lists of terms by various bodies that could help people find specialist terms, explaining that these haven't necessarily been standardised through recognised processes.
- We will help users to know what resource is appropriate for their needs, for example, if they're a school pupil, using Welsh at work, or a new speaker. We will monitor usage patterns, to find out who the different audiences are, how they use the resources, and under which circumstances they're likely to search for a term.
- We will encourage different projects to ensure that terms are consistent, where appropriate, to try to ensure that there is no inconsistency between terms from one resource to the next with regard to, for example, orthography or meaning.
- We will set up an online enquiry service to collect evidence around which terms are a priority for the public and therefore need urgent answers.
- Occasionally, we will evaluate to what degree terms are accepted and used, thereby learning lessons and feeding the findings back into the process.
- We will help to increase the number of professional terminologists, for example, through a combination of training, marketing the opportunities available in the field, and apprenticeships.

Objectives

In the case of gaps in the language's vocabulary, resulting from concepts that arise without warning and where a solution is urgently required: we will establish a quicker process for creating new terms, in a way which responds to events as they arise.

In the case of technical terms it can be foreseen will be needed in the near future (for example, because of a policy in development, or as a large employer wishes to open a new workplace and terms are needed to meet demand in a particular sector) we will create annual work programmes to coordinate this important work.

In both instances, we will facilitate the work of ensuring that standardised terms are consistent across the different resources.

How we will do it

What we have done

- Set up a unit (see area 6) to identify gaps in terminology that need to be rectified urgently in the short term, and to scan the horizon for terms that will be needed in the medium-term.
- In conjunction with Bangor University, we have coordinated the process of standardising terms in the field of race and ethnicity equality, which included consulting with key stakeholder and individuals to create an contemporary list of terms. The lessons learnt from this experience will set a precedent for standardising other packages of terms in the future.

What we will do

- In the case of urgently needed terms, alongside the main providers in the field, we will agree on a fast track standardisation process so that these terms have status and are used more consistently. We will facilitate disseminating the term without delay to the people most likely to need it.
- Regarding terms where there is less urgency, where there are gaps we will commission a group of providers to lead on the work of standardising packages of terminology.
- We will encourage that the terms agreed upon are disseminated and included in other appropriate terminology databases, so that anyone searching for a Welsh term can rest assured they're using the appropriate term.
- We will give publicity to new terms which are standardised, so that users are aware of them.

Area 3: Corpora

A 'corpus' (plural: corpora) can mean several things. But what we have in mind in this document is a database of linguistic information collected following specific principles, which is used to conduct linguistic research and create resources. Corpora such as these can be based on written text or recordings of oral language, or a combination of both.

Linguistic work of all kinds is done on the basis of corpora – for example, a corpus can provide evidence for people developing a dictionary, a terminology, or a language technology resource. How the data in a corpus is collected depends on what we want to get out of it, for example, sound and transcription corpora will be needed to train Welsh speech technologies and artificial intelligence.

We can also use them to develop resources for different kinds of Welsh speakers. This includes new speakers and school children (as they can show users how certain words or expressions are used in a sentence, and how often they appear), as well as publishers, lexicographers (for example, people who develop dictionaries), researchers and the media.

As a result, it's important to make sure that a variety of tagged corpora are available for everyone to use, whether by individuals or technology. If we can also ensure that they're available through open licence, which will mean structured data is available to those who wish to develop different resources, for example, Welsh language systems, speech technologies or artificial intelligence. We will place an emphasis on this in future, in order to ensure that resources, especially those developed using public funds, are available to all.

Corpora are especially important for creating and maintaining a language's digital infrastructure. The aim of our Welsh Language Technology Action Plan is for technology to support the language, so that it's used in as many situations as possible. We're aware that technology develops quickly, and we want to make sure that Welsh moves with it. Our proposals in this new linguistic infrastructure policy will mean, for instance, that corpora are developed (including those that feed dictionaries and terminology databases) that will provide more structured data to support the Welsh Language Technology Action Plan, and enrich the experience of those using technology in Welsh. We need massive Welsh language data sets and relevant corpora to develop speech-to-text capability, new translation systems, and machine learning. To do that, we need to ensure that academic research and development in the field of language technology is supported in the long term. This infrastructure needs to be maintained and updated for future generations.

What resources are available now

Sometimes corpora projects are planned in advance in order to meet specific

needs, and sometimes they're created as a by-product of other activities, for example, as a result of developing dictionaries or terminology databases. There are a number of Welsh or bilingual corpora, which are the result of different research projects over the years. Indeed, any large collection of words, terms or phrases that have been tagged is a corpus. Behind any dictionary or terminology, there will be a corpus, but these aren't always available for everyone to use because of intellectual property constraints.

The main corpus resources available to people who use Welsh are listed below (not all of them are necessarily available to download through open licence).

The National Corpus of Contemporary Welsh (CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes

The CorCenCC project involves Cardiff, Swansea, Lancaster and Bangor Universities. It includes samples of contemporary Welsh which happens naturally in real-life communication, in the form of oral, written and electronic data (for example, Welsh used on social media), presented in the form of an online, searchable and tagged text corpus. It is accompanied by an online teaching and learning toolkit, which draws on data from the corpus to offer resources for learning Welsh to people of all ages and abilities. Cardiff University also has a portal which enables users to search CorCenCC as well as other associated online tools such as WordNet Cymru, Thesawrws, the FreeTxt analysis tool, and the Tiwtiadur for learning Welsh.

Welsh National Corpora Portal

A collection of online written Welsh and bilingual corpora, maintained by the Terminology and Language Technology Unit at Bangor University. It provides an interface which can be used to search for words, terms and phrases, which are displayed in their context. Also, if a translation of the text is available, it can show it side by side with the original. The interface includes software called a lemmatizer, which can find Welsh words in any form (for example, if they've

been mutated, or are shown in plural form). The Corpora Portal also includes links to other corpora that could be of interest to researchers. Some of these were produced by Bangor University, and some by other organisations.

Objectives

Ensuring that extensive and high quality corpus resources are available for everyone to use under an appropriate permissive licence. It's possible that this will happen organically as a by-product of activities which are already taking place, but coordinating the field and maintaining dialogue with the main providers will help ensure that appropriate corpora are available for as many developments as possible in relation to the Welsh language.

Keeping an eye on the gaps that arise in the kinds of corpora that exist, and commission, where appropriate.

How we will do it

- The linguistic infrastructure unit (described in area 6) will have an overview of the field, facilitating work to create and develop corpora where needed.
- Explore how to safeguard corpora that already exist, for instance by providing a facility to keep back-ups of them, to reduce the risk that resources developed with public funds are lost.
- Maintain links to corpus resources (see area 5) in order to help people find corpora that are appropriate for their needs and explaining how they can be useful.
- Help expand the corpora that already exist, for instance by facilitating the availability of public documents to enlarge these resources, where appropriate.
- Facilitate work to enrich corpora by coordinating the relationship between different public providers, and ensure that opportunities to do so are not lost.

- On a central website, maintain links to authoritative corpora, including, for example: lexicons, Welsh and English sentences that have been checked and aligned, place names and street names, lists of named entities, audio corpora that have been transcribed, in order to allow dictionary and terminology database projects to continue to develop based on evidence.
- In order to support new Welsh speakers, it would be beneficial to have a corpus of common forms and mistakes that could be used by organisations such as the National Centre for Learning Welsh to enrich their learning (in the case of common forms, this is already being trialed through a Cardiff University project called Geirfan).

Area 4: Standardisation

Several different kinds of standardisation take place in relation to the Welsh language, and in general they all aim for consistency in the way Welsh is used, by ensuring guidelines exist where they're needed.

The different kinds of standardisation include:

- standardising orthography (for example, how we spell words)
- standardising grammar (for example, the gender of nouns, whether nouns are masculine or feminine; the plural forms of nouns; verb endings; comparative, feminine and plural forms of adjectives)
- standardising place names (with regard to the names of cities, towns, villages and some other geographical features in Wales, the Welsh Language Commissioner leads in this area)
- standardising terms (we address terminology under area 2)

Standardising orthography

Byrger or byrgyr (for 'burger')? Cydbwyllgor or Cyd-bwyllgor (for 'joint-

committee')? Microffon or meicroffon (for 'microphone')? Which one should we use?

At present, the spellings for some Welsh words can vary, and this can lead to inconsistency between different sources, as well as uncertainty for the user. Making sure that the language's orthography (the way Welsh words are spelled) is continuously standardised is important if we want to see more consistency, for instance between different dictionaries and terminologies, and to give more certainty to people who want to use the language.

To get to grips with this, we've set up a Welsh Language Standardisation Panel. The Panel includes academics, terminologists, lexicographers, translators and linguists who convene to standardise Welsh orthography to make it fit for purpose for the modern user. The Standardisation Panel have already met several times since first coming together in January 2021.

The unit in the Welsh Government has a key role in making sure that the members of the Standardisation Panel (for instance, those who maintain dictionaries and terminology databases) recognise the Panel's decisions concerning orthography, and encourage other resources to follow their example.

Standardising grammar and style

After the initial work of standardising the most pressing variations in orthography, the Welsh Language Standardisation Panel will consider the next set of priorities. These may include standardising grammar. This could lead to further advice and guidance being available on using the language, thereby giving clarity to anyone who wants to use Welsh.

With time, this standardisation work will facilitate the development of easy-to-use language guidance. This could include guidance, advice and answers concerning, for instance, conjugating verbs, mutations, appropriate use of

register in certain contexts, and so on, on a central website (see area 5). The unit (described in area 6) will be able to develop this aspect of the work in the long term.

Standardising place names

The Welsh Language Commissioner leads on standardising names of settlements in Wales (for example, names of cities, towns and villages), by convening a Place-names Standardisation Panel which focusses on the spellings of villages, towns and cities in Wales.

The Panel includes specialists on place names and Welsh orthography, and follows specific guidance in order to recommend standard forms for place-names. One part of the work is to maintain a [List of Standardised Welsh Place-names](#), which is an online resource that can be searched or downloaded.

Place names are included here as they're part of the wider standardisation picture. But we believe that the work which we propose in relation to orthography, grammar and style, in tandem with work on place names, will raise the status of standardisation in general. In addition, in collaboration with the Welsh Language Commissioner, it makes sense that the unit in Area 6 would have a role in emphasising the status of standardised place names, and further raising awareness of them.

Objectives

What we have done

The Welsh Language Standardisation Panel has begun to meet the demand for an authoritative voice on orthographical matters in the Welsh language. This

work will continue.

What we will do

Consider whether there is demand to expand the Welsh Language Standardisation Panel's remit to include standardising grammar and style.

Raise the status of what's being standardised and coordinate where necessary the different kinds of standardisation in Wales.

Encourage that variations in spelling between different resources are ironed out over time as providers adhere to the Welsh Language Standardisation Panel's decisions, to ensure more consistency between resources and less confusion for users.

How we will do it

- The linguistic infrastructure unit (see area 6) will coordinate and have oversight of standardisation needs, as appropriate.
- Continue to convene the Welsh Language Standardisation Panel, and providing secretariat support for it, to tackle linguistic issues, for example, standardising orthography. Ensure that the Panel's membership includes a variety of experts who can resolve issues which cause confusion.
- Ensure that the organisations who are members of the Welsh Language Standardisation Panel recognise the Panel's decisions in terms of orthography, and encourage other resources to follow their example.
- Maintain an overview of the field and take steps if it becomes obvious that there is dire need for further standardisation work on specific linguistic aspects.

Area 5: A central website

One of the main issues we're trying to resolve is that so many resources are located in different places, whether they be dictionaries, terminology databases, or corpora. The everyday user can't be expected to know what status these resources have, nor which is most likely to suggest an answer appropriate to their needs. This can also mean that many people can't find relevant resources, or that they don't know about their existence in the first place.

What will we do

Whether trying to find a word or a term, it makes sense that being able to find all the relevant information quickly and easily in one place would make it easier to find answers in relation to Welsh. Therefore, we will create a website where the user can find information about the different Welsh linguistic infrastructure resources in one place.

In the short term, this website will help you easily find the resources most suitable to your needs, be they dictionaries, terminology databases or corpora on-line. It will describe what the different projects have to offer, and have links to take you to them.

Each dictionary or terminology referred to on the new website will remain independent projects with their own websites. This will mean that Welsh Government will be promoting these resources and give them more presence on the internet.

Starting immediately, but coming to fruition in the long term, we will explore whether we could develop this website further to allow users to search easily through several online resources at the same time. This ambition is quite complex, and to achieve it we would need the help and support of those who

provide the resources themselves.

The first step in this regard will be to continue with the work to standardise orthography (see area 4) in order to ensure consistent spellings, and ensure that resources commit to following these standardised forms. There will also be a need to consider how best to tag words and terms from different projects in a consistent way that would allow them to be searched on a central interface. Having done that, we would like to be in a position where you can note what kind of user you are (for example, someone using Welsh at work, a professional translator, school pupil, parent - whether able to speak Welsh or not, teacher, university student, and so forth), or what area you'd like to search, and that the results would be arranged according to relevance. This would give you more certainty that you've found the appropriate answer.

A multipurpose website

This central website will also be a springboard for all kinds of activities in relation to Welsh linguistic infrastructure.

When it's up and running, we will develop the website to help all kinds of Welsh speakers or users. For instance, by drawing attention to resources that show how words mutate or are conjugated in other ways. This could be complemented by the inclusion of various style-guides, as mentioned in area 4. Materials such as these would be especially useful to new speakers and those using the language every day at work.

In order to identify gaps in the provision with regard to words or terms, we will include a form on the website, so that individuals or organisations can enter a specific high-profile term or word, that hasn't been included in the resources as they are. The officials working for the unit (see area 6) will use the information gathered through this process to help to prioritise what kind of standardising terminology work will be done in the future.

The website will also be an ideal place to publish decisions of the Welsh Language Standardisation Panel in relation to orthography, grammar and style (see area 4).

In the case of corpora, it would be relatively simple for the central website to include links to all the appropriate corpora on one interface, in order to help people find corpora that are appropriate for them. The aim would be to facilitate the sharing of different corpora so that people can develop new resources. It would be good practice, and ensure value for public money, to work towards a situation where the most up-to-date versions of resources could be downloaded from the website under open licence. Indeed, we propose that the principle of providing resources through open licence is relevant to as many of the resources mentioned in this document as possible.

What resources are available now

Bangor University Language Technology Unit's website

Provides access to the Bangor Welsh-English Dictionary, as well as the University's Terminology Portal and Corpora Portal. This includes an extensive and searchable collection of terminology and corpus resources across a number of fields. It doesn't include BydTermCymru, the other main terminology database for the Welsh language, and doesn't enable the user to search all the main online dictionaries, for example, Geiriadur Prifysgol Cymru.

The Gweiadur

Provides a suite of searchable linguistic resources, which include Welsh-English and English-Welsh dictionaries, a Welsh Thesaurus, an English-Welsh Thesaurus, and the Verb Conjugator. The resources themselves, as well as the website, were developed by D. Geraint Lewis and Nudd Lewis. The website also includes links that facilitates searches in other resources, like Geiriadur Prifysgol

Cymru, TermCymru, Y Porth Termau, Wikipedia, the Record of the Proceedings of the Senedd, and CorGenCC, by opening new tabs.

European Dictionary Portal

One lesser-known example that acts as a window to a large number of resources is the European Dictionary Portal, which is curated by experts from the European Network of e-Lexicography. It offers a way of searching a number of sources in one place, but rather than providing all the results side by side, the user must go to each individual website to see the results. This website also doesn't provide the user with advice about which resource could be most appropriate for them.

Objectives

In the short term, create a website where users can get information about Welsh language infrastructure resources, as well as links to them, to help users know where to find the best resource for them.

In the long term, consider how to work towards a situation where users can search the main Welsh dictionaries and terminologies simultaneously through one interface.

Develop the website to be a source for all kinds of work in relation to Welsh linguistic infrastructure.

Use the website as a way of marketing and promote the excellent resources we have for the Welsh language, so that more people are aware of them and use them.

How we will do it

In the short term

- Create and launch a website where users can go for advice about which resources are best for them. The website will also promote and market the resources that exist for the Welsh language.
- Establish a process for receiving and responding to enquiries submitted via the website.
- Establish a link between the enquiries that are submitted through the website and the field of standardising terminology outlined in area 4.
- Explore the website's potential from the perspective of corpora, standardisation, sharing best practice, tracking how words and terms are used, providing brief guidelines on use, and so on.

In the medium-long term

- Start discussing immediately with different projects in order to work towards a situation where search results from different databases are included through one central interface, as well as on their own websites.
- Take steps to pave the way towards being able to search through one interface, and continue to ensure consistency in how words are spelt, so that they can appear through the same searches.

Area 6: A new unit to coordinate Welsh linguistic infrastructure

When describing the different resources above, it's obvious that work of an extremely high, professional standard is taking place across all areas of

linguistic infrastructure in Wales. What we also see is that there are gaps that could be filled by maintaining more oversight over the field, and by coordinating more effectively what already exists.

In the draft policy that was subject to consultation, we said that there wasn't one body or team responsible for coordinating Welsh linguistic infrastructure provision in its entirety. It noted that Welsh Government provides funding for many different projects, but that there wasn't any formal coordination, with the individual projects operating independently from each other.

For years, decades even, there has been talk about the need to better coordinate Welsh linguistic infrastructure. A number of experts in the field have also agreed with this.

In other countries with languages in similar situations to Welsh, governments and government agencies play a prominent role in planning and coordinating provision in terms of dictionaries, terminologies, corpora, and helping to standardise the language.

That is why we have established a central unit to coordinate dictionaries, terminology, corpora, standardisation work, and to set up a website to provide resources together in one place. The unit will also work closely with a variety of stakeholders in fields such as translation and information technology.

What we will do

In order to maintain an overview of this field, and help to plan it strategically, we have set up a new unit. The unit will coordinate different aspects of Welsh linguistic infrastructure, and ensure that the provision makes it easier for users to find definitive answers. By establishing such a unit, we envisage that it will lend prestige to the field and promote the status of words and terms in Wales, so that there is more consistency of use across different organisations and the media. In

turn, this will help people to use their Welsh skills more often in more situations. The unit will undertake activities such as those described in areas 1 to 5 in this document.

We have located this unit in the Welsh Government. The advantage of doing this is that it can benefit from holistic planning, thereby linking in with other relevant policy areas, for instance our work as a Government on language technology, which complements what's in this document. It also means the unit is well placed to be aware of developments in different policy areas (for example, education, health, the economy) in order to scan the horizon for terminology needs.

On a practical level, locating it in the Welsh Government also keeps the costs and overheads associated with setting up such a unit lower than establishing it outside the Government. This helps to ensure that any investment is focused on the end-product and specialist staff, rather than overheads, Human Resource services and offices.

The unit's structure

In order to ensure that the unit is well-located to respond to policy developments, to scan the horizon for terminology needs, and respond to enquiries of a linguistic nature, we have started appointing staff with a background in linguistic infrastructure.

It's important for the unit to have a strong role in terms of marketing and engagement, to ensure that target audiences are aware of resources, understand what they're for, and use them. That being so, ensuring that the unit has status, and fostering awareness of what it does, would help lead to more consistency in the use of words and terms, and therefore encourage more people to use them.

As one of its objectives is to develop the long term robustness of Welsh linguistic infrastructure, we would also aim to take advantage of apprenticeship schemes to ensure that trainees are given opportunities to learn about this area of work.

Although the unit is based in the Welsh Government, we will work to develop its own identity and brand, so that it's visible and accessible to the public.

Funding

Our aim is to maintain the funding which at present is distributed to various linguistic infrastructure projects. Any funding provided for the unit's work, and for further developing the field more generally (for example, developing the offer in relation to dictionaries, creating a website) will be new funding for this area.

Objectives

What we have done

We have already set up a unit in the Welsh Government to start work on the objectives in this document. The unit's aim is to coordinate Welsh linguistic infrastructure, and provide strategic leadership so that the different components work together better.

What we will do

The unit will be responsible for the objectives outlined in areas 1 to 5 in this document, and it will prioritise:

- Fostering a strong identity and ensuring that the unit and resources connected with it are marketed effectively.

- Maintaining and developing the relationship with the main stakeholders.
- Developing a website where users can go to get advice about which resources are best for them, and work towards a situation where the user can search through various resources on the same website.
- Encouraging and facilitating the development of skills and expertise through apprenticeships and raising awareness of the field as a career.
- Beginning to develop packages of work under the proposals described in areas 1 to 5.

Welsh place-names

In our Programme for Government, we commit to working to protect Welsh place-names, and this has been reinforced by the Cooperation Agreement between the Welsh Government and Plaid Cymru. The linguistic infrastructure unit will also be responsible for Welsh Government policy on Welsh place-names.

Our initial steps with regards to this have been set out in our [Welsh Language Communities Housing Plan](#), published in October 2022. The aim of that plan is to tackle the challenges facing Welsh-speaking communities with a high concentration of second homes. The linguistic infrastructure unit will be responsible for those steps, which aim to “protect the Welsh language inheritance of our built environment and topography”.

Objective

Act to ensure that Welsh language place names in the built and natural environments are safeguarded and promoted for future generations (The Cooperation Agreement).

What we have done

- Commission research to learn more about how, why and where Welsh place-names are changing, to enable us to develop targeted policy interventions to safeguard names in every part of Wales.
- Establish a Local Authority Place Names Forum, which aims to:
 - share information
 - identify gaps in the way Welsh place-names are dealt with, and opportunities to collaborate
 - support each other to try to find practical solutions to prevent Welsh place-names from being displaced
 - share good practice, so that examples of what works in one organisation can be adopted, where appropriate, in another organisation

What we will do

- In terms of geographic names, as well as the aforementioned research to better understand the processes behind changes, we will work with partners and businesses, as well as prominent cartographers, to better safeguard geographic names in the natural environment.
- We will explore how local authorities discharge their role in this area of work to ensure that all possible steps to safeguard place names are being taken.
- We will consider the recent use of covenants to protect house and field names and explore how these might be used more widely in the future.
- We will explore the use of conveyancing packages as a way of raising awareness of the value of a particular property's name.
- We will explore new ways of raising awareness of and promoting the List of Historic Place Names of Wales.
- We will keep abreast of internal developments regarding the names of electoral wards and of the new Senedd constituencies for the 2026 election, and advise where appropriate.

- We will continue to consider international developments in the field of place names in order to keep an eye on best practice. We will also continue to make use of forums such as the British Irish Council (the Welsh Government chairs the indigenous, minority and lesser used languages group, and are hopeful that safeguarding place names will be included in the Council's next three-year programme of work) and Network to Promote Linguistic Diversity (NPLD).

We believe that this will

- Enable us to have a better understanding of the roles of different agencies and authorities in safeguarding Welsh place and house names.
- Enable us to formulate appropriate interventions to better safeguard Welsh names, which may or may not lead us to consider developing: legislation, the planning system, or any other interventions.
- Increase awareness of Welsh and historic place names of all types, and a greater appreciation of their value.
- Safeguard place names for future generations, both as part of our historic heritage and as visual manifestations of our living, breathing communities.

Epilogue

There's no doubt that the way Welsh linguistic infrastructure currently works needs to be improved. The steps outlined in this paper aim to coordinate the different elements, commission resources such as terminology as needed, make sure that the fruits of this labour are easily available to all, and market the provision effectively. We believe this will make it easier for everyone to use the language confidently, be they new speakers, the parents of school children, or people using Welsh at work in a professional capacity.

The [responses to the consultation](#) have helped us weigh up what exactly the new way of doing things will entail, and what responsibilities the new unit should have.

A lot of work has already been done in this field. For decades, we've recognised that different elements of Welsh language infrastructure are inter-connected. Many components already exist (and we will coordinate them so that they work in a more coherent and effective way), whilst adding the proactive element of scanning the horizon for developments and planning strategically.

We want this Linguistic Infrastructure Policy to play a central part in our response to improve the way this field works, and help to ensure that the provision continues to develop so that people can use the Welsh language easily.

This document may not be fully accessible.

For more information refer to our [accessibility statement](#).

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/welsh-linguistic-infrastructure-policy-html> for the latest version.

Get [information on copyright](#).