MULTILINGUAL INFORMATION ONLINE

VICTORIAN GOVERNMENT GUIDELINES ON POLICY AND PROCEDURES

CONTENTS

FOREWORD 3

INTRODUCTION 5

VICTORIAN GOVERNMENT POLICIES AND STANDARDS 6
  Victorian Government digital standards and design principles 6
  Using credentialed translators 6
  Machine automated interpreting and translating tools 6
  Text to speech tools 7

PREPARING CONTENT FOR WEB TRANSLATIONS 8
  Who will access the website? 8
  Web access models 8
  Considerations for different languages and audiences 8
  Using audiovisual content 9
  Culturally appropriate content and design 9
  Additional material for translation 10
  Quality control for translated content 10
  Technical considerations 10
  Website navigation 11
  Language selection features 11
  Logo indicating translated material 11
  Interpreter symbol 11
  Metadata 12

ENSURING INFORMATION QUALITY 13
  Final checks before going live 13
  Reviewing multilingual content 13
  Promoting translated material 14

TECHNICAL NOTES 15
  Adding translated content to a website 15
  Content Management Systems 16
  Encoding 16
  Specifying page encoding 17
  What to do when you are not using Unicode 17
  Indicating languages 18
  Indicating change of language 19
  Text direction 20

APPENDIX 1: LANGUAGE TAGS 22

APPENDIX 2: INTERNATIONALISATION AND ACCESSIBILITY 23

APPENDIX 3: WEBSITE LINKS 25

This document was developed as part of the Multicultural Government Information Online project undertaken by the State Library Victoria for the Department of Premier and Cabinet. It was subsequently updated by Enabling Languages.

FOREWORD

More than one million Victorians speak a language other than English at home. Many speak several languages. Language skills act as a bridge between people and between the cultures that make up our community. Our linguistic diversity reflects the multicultural and cosmopolitan nature of our State.

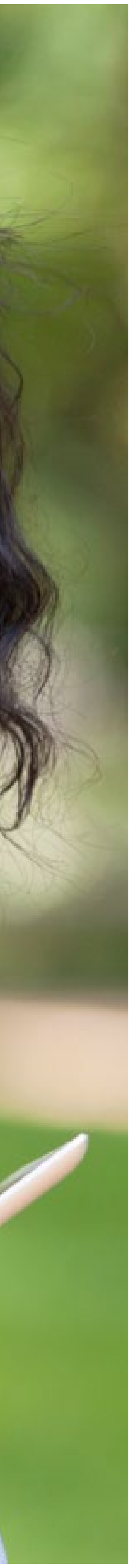
The Victorian Government aims to ensure that high quality interpreting and translation services are available for all Victorians who require language assistance when accessing government services.

In a multicultural society, such as Victoria, websites play an increasingly important role in providing information about government services. Victorians who prefer information in a language other than English should also enjoy the benefits online delivery offers.

Many government departments and agencies already provide information on their websites in languages other than English. These Guidelines will help all departments to provide online multilingual information effectively by improving the navigation and accessibility of online information in other languages.

I trust that all government departments and agencies will find these Guidelines useful in delivering high quality and accessible services to culturally and linguistically diverse Victorians.

ROBIN SCOTT MP
Minister for Multicultural Affairs



INTRODUCTION

These Guidelines aim to assist Victorian Government departments and agencies to improve the availability of multilingual information on their websites and other digital mediums. They are designed for people developing online content for translation, web teams deploying multilingual online content, and professional translators working on website content.

The Guidelines focus on preparing and deploying multilingual information online, and making it more accessible. The Guidelines should be read in conjunction with the companion publication, *Effective Translations: Victorian Government Guidelines on Policy and Procedures*.

Over one million Victorians speak a language other than English at home and over 200,000 Victorians have limited English proficiency. Language services are critical for many Victorians to access government services and information.

While internet access and usage varies between communities, digital platforms are increasingly important in making government information available in other languages. The ABS census showed that internet use among Victorians, originating from countries where people are less likely to speak English, increased by 53 percent between 2006 and 2011.

Online delivery complements traditional ways of providing multilingual information. It also offers a number of unique features compared to hard copy translated information. It can more easily reach wide and dispersed audiences. Costs can be lower compared to hardcopy distribution, and online information is easier to keep up to date. Another advantage is that web-based multilingual audiovisual information can also be used to complement the written word.

Victorian Government departments and agencies provide a range of translated materials on their websites. However, navigating websites to find translated information can often be difficult without a knowledge of English. One reason is that translated information is often displayed in file formats, such as PDF, which may not contain searchable text. Making translated information more 'discoverable' is facilitated by having the content in HTML or by providing optimised MS Word or PDF files and improving search tools.

Improving website navigation will make translated information easier to find. As web technology changes and improves, new solutions are becoming available to enable better online accessibility of multilingual information.

The following companion publications are also available:

- **Using Interpreting Services**: Victorian Government Guidelines on Policy and Procedures, provides information on arranging interpreting services
- **Effective Translations**: Victorian Government Guidelines on Policy and Procedures, provides guidance on arranging translations of written information into languages other than English.

All three guidelines can be downloaded from **www.multicultural.vic.gov.au**

VICTORIAN GOVERNMENT POLICIES AND STANDARDS

The *Multicultural Victoria Act 2011* (the Act) states that all individuals in Victoria are equally entitled to access opportunities and participate in and contribute to the social, cultural, economic and political life of the state. Availability of information online translated into languages other than English is important to ensuring this is achieved.

Government departments and agencies have a responsibility to ensure people with limited English, and people who are Deaf or hard of hearing, are given information in their own language to participate in decisions that affect their lives.

The Act also requires all Government departments to report annually on their use of interpreting and translation services. This includes reporting on the accessibility of information on government services in languages other the English.

Further detail about relevant Victorian Government legislation and policies is available in *Effective Translations: Victorian Government Guidelines on Policy and Procedures*.

VICTORIAN GOVERNMENT DIGITAL STANDARDS AND DESIGN PRINCIPLES

The Victorian Government digital standards and digital design principles articulate the principles of government website management and are available at: **vic.gov.au/digitalstandards**

The guidance in the standards on ensuring that website material is discoverable and usable is particularly important for deploying translated content in community languages.

These requirements should be taken into account when planning and implementing translated government information.

USING CREDENTIALED TRANSLATORS

Victorian Government policy is that interpreters and translators should be appropriately credentialed by the National Accreditation Authority for Translators and Interpreters (NAATI). This is important to ensure the quality of online multilingual information.

It is advisable to avoid using translators based overseas as they may not be NAATI-credentialed. Also, overseas translators may not have a good understanding of the local community or issues and may not be familiar with Australian English.

MACHINE AUTOMATED INTERPRETING AND TRANSLATING TOOLS

Machine automated interpreting and translating tools undertake translating or interpreting with no human involvement and can, for example, automatically translate information on a website from one language to another.

Victorian Government policy strongly recommends engaging NAATI credentialed interpreters and translators and currently advises against the use of automated interpreting and translating tools, which cannot at present be guaranteed to be accurate. While some machine tools are improving, they still have a reasonably high chance of incorrectly translating information.

Accessibility *refers to the features of a website, and other digital channels, that enable all people, regardless of linguistic or other needs, to access its information.*

Discoverability *refers to how easily information can be found. For a translated webpage to be useful, people need to be able to find it through a search engine or a link from another website.*

Machine automated interpreting and translating tools may be unable to take into account:

- variations in dialect and language
- linguistic preferences of communities
- actual meaning (i.e. word for word translation does not consider overall comprehension)
- specific cultural references
- other nuances such as politeness level.

There may be risks of legal action due to distorted translations. It is unlikely that a disclaimer about the content in an automatic translation would relieve an organisation of the responsibility for the information provided.

Written content that has been translated by a machine should always be checked for accuracy by a NAATI credentialed translator.

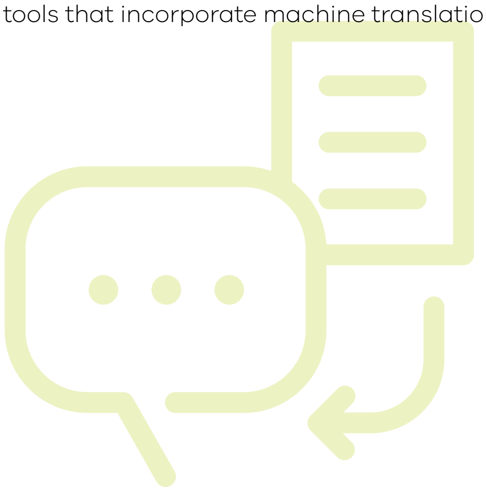
Also, machine translations may not support all languages that may be required.

TEXT TO SPEECH TOOLS

Text to speech tools can be integrated into websites. These tools can improve accessibility, and can be appropriate for users who may have limited ability to read English but are able to understand spoken English.

These tools often support a number of different languages, and some tools also integrate machine translation. Not all community languages needed may be available.

Considerations that apply to machine translation tools also apply to text to speech tools that incorporate machine translation.



PREPARING CONTENT FOR WEB TRANSLATIONS

WHO WILL ACCESS THE WEBSITE?

It is important to determine how the translated information is to be accessed. The website may be intended for people to directly read or listen to information in their own language, or for service providers to find information on a client's behalf.

ONLINE ACCESS MODELS

Determining who will access the website will help to decide which online access model is most appropriate.

Direct access

- Navigation from the homepage to translated documents is available in languages other than English
- Enables people to find the translated information for themselves.

Mediated access

- Navigation is in English only
- Service providers or other English-speakers access the translated information on behalf of people who require it.

Dual access

- Navigation is both in English and in the languages of translation
- Labels for links and documents are in English and in the translated languages to enable both direct and mediated access to translations.

CONSIDERATIONS FOR DIFFERENT LANGUAGES AND AUDIENCES

Some languages can present particular challenges to achieving online accessibility and discoverability. Factors to consider when translating and deploying information in these languages include:

- Linguistic diversity within languages. For example, some languages contain a large number of dialects which use different terminology
- Literacy levels within a community that speaks a particular language
- Lexical gaps. For example, there may not always be equivalent concepts or words in another language
- Lack of style guides and information on typesetting and typography for certain scripts.

USING AUDIOVISUAL CONTENT

Alternatives to written translations are available to cater for varying literacy needs and language requirements. It is important to understand the communication preferences of the target audience. For example, some people are unable to read the language they speak. Also, some languages are rarely displayed in their written form and are largely oral. In these instances audiovisual content may be more effective.

To meet accessibility requirements, audiovisual content, including any English language transcripts, will also require translation by a NAATI credentialed translator.

Translated content may be either written subtitles or spoken (over-dubbing).

Audiovisual material can be expensive to produce so it is important to identify which communities would most benefit from this type of delivery format.

CULTURALLY APPROPRIATE CONTENT AND DESIGN

Check any images and content associated with multilingual information to ensure these are appropriate. If in doubt consult relevant community organisations for advice.

Consider that some symbols and expressions used in Australia may not be familiar to new migrants and refugees. For example, images of parking signage such as ‘no standing’ and ‘clearway’ zones could require additional explanation.

Backspace
Enter

ADDITIONAL MATERIAL FOR TRANSLATION

In addition to the main content, other material to be translated may include:

- Introductory text
- Title of documents
- Alternative text for images
- Words and phrases needed for navigation
- Document metadata
- Accessibility, copyright and privacy statements
- Contact information
- Audio transcripts and video scripts
- Video closed captioning or subtitling.

All material to be translated needs to be thoroughly identified. The additional material will form part of the brief to the translator.

QUALITY CONTROL FOR TRANSLATED CONTENT

All content to be translated needs to be carefully checked before it is submitted to a translator. It needs to be clear, concise, appropriate and accurate.

When briefing a language services provider or translator ensure the following:

- Specify that translations will be used on a website and will need to be in Unicode
- Ask the language services provider to perform a final check of the translations after these are loaded onto the website.

TECHNICAL CONSIDERATIONS

When preparing multilingual content for the web, consider:

- The translated information may take up more or less space than the English text. Text expansion and reduction should be taken into account when creating the design template for the publication. Consult with both the language service provider and the digital team for advice on space requirements
- Translations may involve languages that do not use spaces to delineate words. Web browsers are inconsistent with line breaking for such languages. It may be necessary to use Cascading Style Sheets (CSS) and Javascript to improve line-breaking for some languages
- Translations may entail bi-directional scripts. Bi-directional text (known as bidi) contains information that runs both left-to-right, and right-to-left. It generally involves text containing different types of alphabets. Some content management systems, or the templates they use, need to be adapted to enable such scripts to display correctly
- The format in which translations should be provided (HTML or MS Word files)
- Whether both HTML and MS Word files (or the less accessible PDF files) should be used to enable printing content from the site
- Formats for multimedia content.

WEBSITE NAVIGATION

Websites should provide clear navigation from the home page to the translated content. To ensure that content is accessible and user-friendly:

- Multilingual content should be in HTML rather than, or in addition to, MS Word or PDF format. Using HTML allows search engines to locate the information in a language other than English
- Ensure both the language and publication title is included in English at the beginning of the translation for easy identification and to assist with distribution of printed versions
- Include navigation to both the English version and the non-English translation on the same page.

The English language sitemap should provide an index of translations by language.

LANGUAGE SELECTION FEATURES

A language selector should be a prominent design element on the website. If the language selector is not included on the initial viewport, a navigation link to the language selector (such as the 'in your language' logo) should be available on the site's masthead across the site.

To search for target languages easily when navigation is in English only, link labels to translated documents should be made bilingual i.e. in the target language and English.

The site should also use user friendly URLs (in English), with the language name included in the URL. For example: **http://www.multicultural.vic.gov.au/italian**

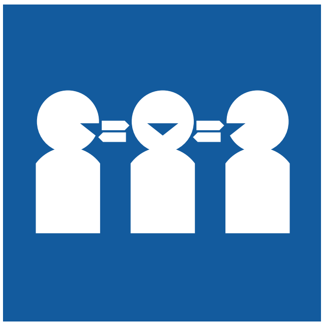
LOGO INDICATING TRANSLATED MATERIAL

In your language

The above logo can be used to indicate the availability of translated information online.

This logo can be accessed from: **http://www.multicultural.vic.gov.au**

INTERPRETER SYMBOL



The Interpreter symbol was designed to show where someone can ask for language assistance. It provides a simple way to help people with low English proficiency access government services. The symbol indicates that a person with low English proficiency can ask for help to communicate in their own language.

This symbol can be used on a website to link to information about accessing or using an interpreter phone service, or other advice about communicating with a department or agency in a language other than English.

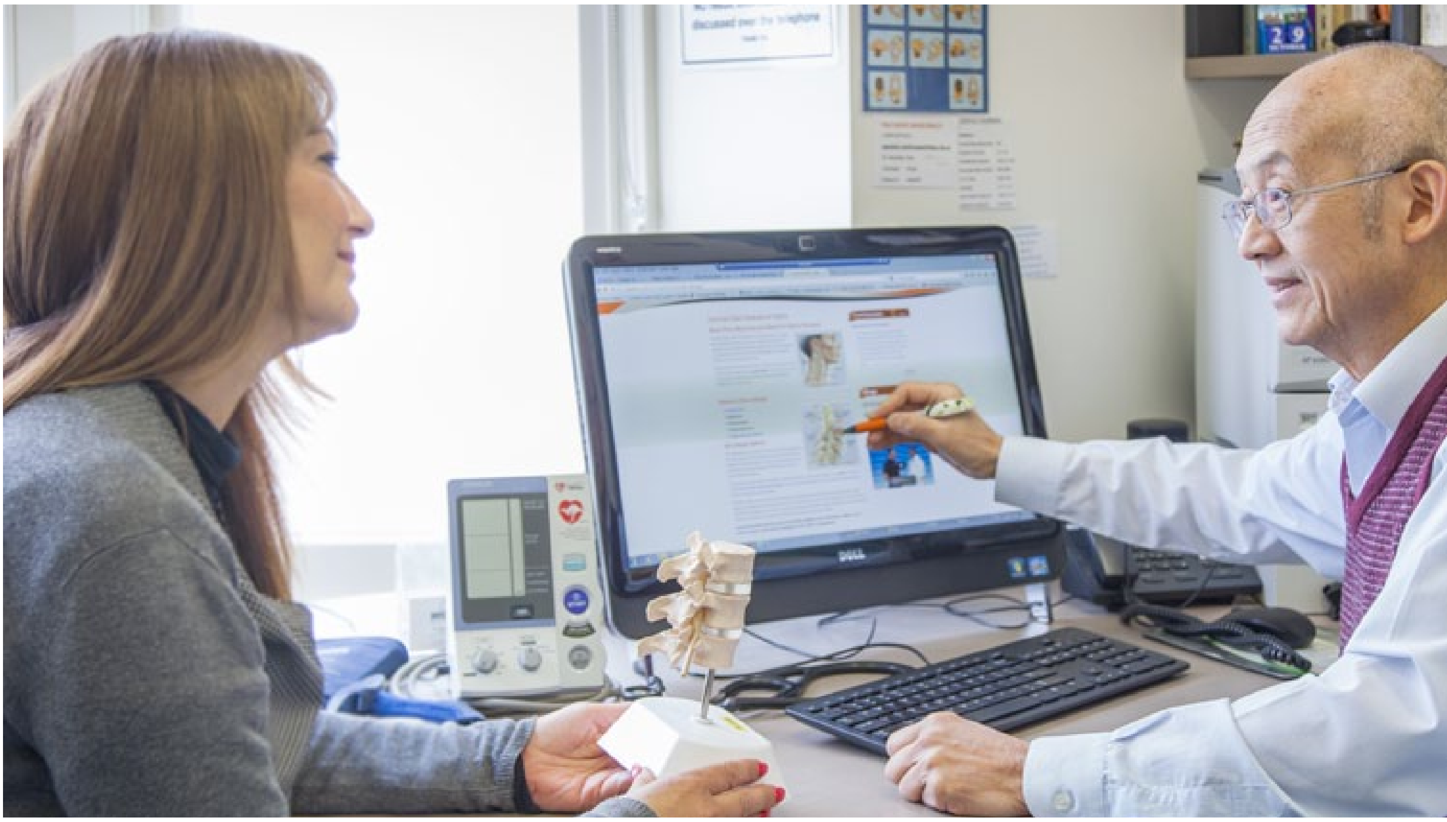
METADATA

Metadata summarises information about a webpage, MS Word or PDF file.

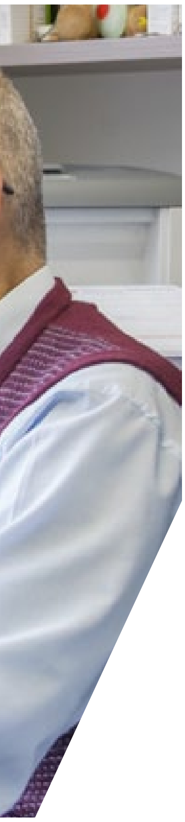
The web access model should determine the language that relevant metadata should be in. For example:

- For websites based on the direct access model, metadata should be translated;
- For the mediated access model, metadata should be in English; and
- For a dual access or a bilingual page, metadata should be provided in both English and the other language.

Further information on improving website access and navigation is in the Technical Notes section of these guidelines.



ENSURING INFORMATION QUALITY



FINAL CHECKS BEFORE GOING LIVE

Translated content should go through final checking before it is made publicly available. Some steps will require checking by a NAATI credentialed translator while others can be done by the digital team.

Consider the following:

- Is the text rendering correctly?
- Is a suitable font being used?
- Did the text become corrupted when it was added to the website?
- Are lines wrapping or breaking in acceptable places?
- Are languages that are written from right-to-left, such as Arabic and Persian, displaying correctly? Text alignment, positioning of bullets, punctuation and phone numbers should be checked
- Final checking of the translated webpage from the language services provider should be scheduled before the webpage goes live.

REVIEWING MULTILINGUAL CONTENT

Translated material on the web should be reviewed periodically to determine whether the information is still relevant and up to date.

- Update translated material on a website whenever the original English version changes
- Assess the effectiveness of the translated publication in conveying the intended information. This might include specifically requesting feedback or conducting surveys of the target audience and relevant service providers
- Review the languages the translated content has been translated into. Other languages may need to be added from time to time, to reflect Victoria's changing migration and resettlement patterns
- Monitor the distribution of the translated material by collecting website data on visits to translated pages, choice of language and the referral traffic. This data can improve understanding of who accesses the website
- Keep original English versions of translations. This is helpful when making corrections or updates, or repurposing content to make a brochure, printed publication or new webpage. Because most translations are costed on a per word basis, making minor updates to existing documents is cheaper than translating a new document.

PROMOTING TRANSLATED MATERIAL

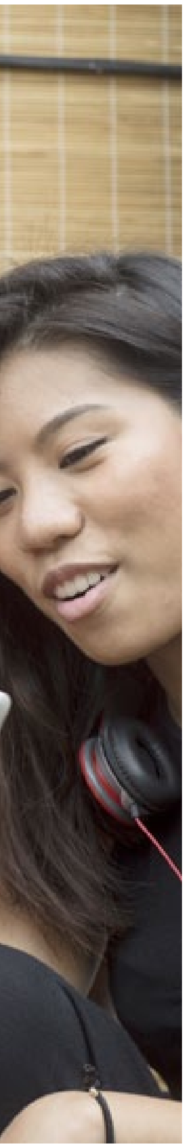
Promoting translated material on websites can be done by sending information and links to organisations with strong connections to Victoria's culturally and linguistically diverse communities.

The following links provide a starting point for the promotion of translated materials:

- The Victorian Multicultural Commission – **multicultural.vic.gov.au**
- The Ethnic Communities' Council of Victoria – **eccv.org.au**
- Health Translations Directory – **healthtranslations.vic.gov.au**
- The Centre for Culture, Ethnicity and Health – **ceh.org.au**
- Action on Disability within Ethnic Communities – **adec.org.au**
- The Federation of Ethnic Communities' Councils of Australia – **fecca.org.au**
- The Refugee Council of Australia – **refugeecouncil.org.au**

Be sure to specify the languages included on your website as this will assist directing information to relevant communities.

TECHNICAL NOTES



ADDING TRANSLATED CONTENT TO A WEBSITE

To maximise accessibility and discoverability HTML should be used.

Print-friendly MS Word versions are preferable to PDF and can be provided alongside HTML content. While PDFs are widely used for translated content their format is often not suitable as they may not contain searchable text. As such, they may not appear in search results and can be very difficult to find in some languages.

If PDF files are still required in addition to MS Word, PDF/UA should be used. PDF accessibility requirements are documented in PDF techniques for WCAG 2.0 and ISO 14289-1:2014. Some community languages have additional requirements. Appendix 2 documents some aspects of the accessibility of HTML and PDF files in relation to community languages.

Key points for displaying translated content:

- Use characters rather than escaped characters. An escaped character is an alternative way of representing a character, used in some programming languages
- Indicate the language of each document and any change in language, using the lang attribute on relevant HTML elements
- Use style sheets for consistent page presentation;
- Use appropriate encoding on forms and servers that support Australian formats for names, addresses, dates and time
- Keep text separate from graphics. The space taken up by a translation will often differ from the space taken up by the English version
- Include a clearly visible navigation system to localised content on each page, using the target language (see section on logo indicating translated material)
- For writing systems that are rendered from right-to-left, such as Arabic, clearly indicate the base text direction (right-to-left) of the document and indicate changes in text direction when the language of the content changes
- Check and validate work before publishing it.

CONTENT MANAGEMENT SYSTEMS

The themes and templates for a website may need to be updated to support community languages appropriately.

Thought should also be given to how the editing interfaces can be optimised to support editing and markup of community language content. The editing interface should be able to handle all the languages being translated.

The following features should also be available:

- Ability to control the overall directionality of content in the editing interface
- Add mark-up to control directionality of block level and inline elements
- Marking-up change in language on block level and inline elements
- Display of translations in fonts appropriate to the language within the editing interface.

Not all Content Management Systems in use across Victorian Government websites support Unicode. This may present challenges at the editing interface.

ENCODING

Character encoding refers to the way a character (such as a letter or number) is represented in binary data by a computer. ASCII and Unicode are the most common systems of character encoding, and Unicode is best for multilingual content as it supports a larger set of characters from different alphabets and scripts.

All translated content should be provided in Unicode. HTML content should use the UTF-8 character encoding.

Key resources include:

- Introducing Character Sets and Encodings
- Character encodings for beginners
- Character encodings: Essential concepts
- Choosing & applying a character encoding
- Declaring character encodings in HTML
- Declaring character encodings in CSS

SPECIFYING PAGE ENCODING

It is essential to declare the encoding of the documents.

- The character encoding of a document can be specified in the web server's HTTP Response Header, or the information can be included in the actual web page
- If the character encoding is declared in the HTTP Response Header, it should also be included within the web page as well
- The value in the HTTP Response Header must match the value declared in the web page.

Depending on the document type, there are different ways to declare the encoding. The table below indicates the declarations required for UTF-8 encoded HTML4, HTML5 and XML documents.

Document type	**Document type**	**Notes**
HTML 4	<meta http-equiv="Content-Type" content="text/html; charset=utf-8">	Declared in a meta element within the head element
HTML 5	<meta charset="utf-8">	Declared in a meta element within the head element
XML	<?xml version="1.0" encoding="utf-8"?>	Declared before XML root element

WHAT TO DO WHEN YOU ARE NOT USING UNICODE

When your CMS is using a legacy encoding, it is possible to convert Unicode content into a format that can be used in a non-Unicode CMS.

It is possible to convert the characters in the HTML content into Numerical Character References (NCR). These are HTML entities that identify a particular character (by decimal or hexadecimal numbers). Browsers will substitute the correct character or letter. For instance the lowercase Greek letter alpha (U+03b1) can be represented as a decimal character reference, for example &#945; or it can be represented in hexadecimal notation, for example: &#x3b1;.

INDICATING LANGUAGES

It is essential to indicate the language of a web page to: enhance accessibility; enable language specific searching within search engines, and; for browsers to select the appropriate fonts.

There is a distinction between the primary language of a document and the text processing language. The text processing language is the language in which the text of the document is written, processed, displayed or read by a screen reader. The lang and xml:lang attributes are used to indicate the text processing language.

It is necessary to declare the default text processing language for the whole document. Declaring a text processing language in the HTML element will specify the default language for the whole document. Do not declare the language of a document in the body element.

If the document has multiple main languages, it will be necessary to decide whether one of the languages is declared as a text processing language in the HTML element, or leave the default text processing language undefined.

For Victorian Government websites, the language of the page is best set to "en" (English) or "en-AU" (Australian English), even when the unique content is not in English.

Document type	**Language declaration**	**Notes**
HTML 4 and HTML5	<html lang="am">	Declared primary language of document in a lang attribute in html element
XML	<html xml:lang="am" xmlns="http://www.w3.org/1999/xhtml">	Declare primary document language in xml:lang attribute of root element

INDICATING CHANGE OF LANGUAGE

It is necessary to declare any language changes within a document. Use the lang or xml:lang attributes around any changes in language within a document. If there is no appropriate element to add the language declaration to, use the div element for a block change and use a span element for an inline change. For example:

```
<p>The Chinese title is <span lang="zh-Hant">哮喘病簡介</span></p>
```

The specification of a text processing language not only applies to the content of the element but also to the content of attributes used by the same element. If the text attribute values and the element content is in different languages, consider using a nested approach. For example:

Use nested tags as follows:

```
<li lang="en-AU" title="Emergency Relief and Recovery – help is available">
   <a lang="din" href="">Akuny wëi kë cï tuöl ku bën-pïïr – kuony aluthïn</a>
</li>
```

Instead of the following code:

```
<li>
   <a lang="din" title="Emergency Relief and Recovery – help is available"
href="">Akuny wëi kë cï tuöl ku bën-pïïr – kuony aluthïn</a>
</li>
```

If there are multiple main languages within the document, the web developer should divide the document into blocks at the highest possible level. The appropriate text processing language should be declared for each of these blocks.

When using Unicode it is important to declare the language of text written in Chinese and Japanese. These languages share Unicode characters, but the glyphs may differ between traditional Chinese, simplified Chinese and Japanese.

If the languages are declared in the mark-up, web browsers can use appropriate default fonts for each language/writing script.

For most government sites deploying translated content, the overall language of the site templates will be English. Therefore, it is good practice to wrap the translated content in a div element, or other block level element, with the appropriate lang attribute:

```
<div id="translationContent" lang="hi"></div>
```

Key resources:

- Choosing a Language Tag
- Declaring language in HTML
- HTTP headers, meta elements and language information
- IANA Language Subtag Registry search tool
- Language on the Web
- Language tags in HTML and XML
- Why use the language attribute?
- Working with language in HTML

Appendix 1 contains a list of languages used on Victorian Government websites and the preferred language tag for each.

TEXT DIRECTION

Bi-directional text (known as bidi) contains information that runs both left-to-right, and right-to-left. It generally involves text containing different types of alphabets, i.e. scripts that are read right-to-left and left-to-right.

The design of templates or themes needs to accommodate both RTL (right-to-left) and LTR (left-to-right) languages. It is important to handle bidirectional text with care. In HTML Unicode documents, it is possible to add the dir attribute to a HTML entity to indicate the directionality of text within that element.

For a web page written in a right-to-left script, the overall document direction should be indicated in the html element. For example:

```
<html lang="ar" dir="rtl">
```

Do not add dir="rtl" to the body element. The default direction of a web page is LTR.

For web pages written in languages using LTR scripts, it is not necessary to indicate the primary direction of a web page.

Government website templates will be in English, so a more practical approach is to wrap the translated content in an appropriate block level element and apply lang and dir attributes to that block level element.

```
<div id="translationContent" lang="prs" dir="rtl"></div>
```

Key resources:

- Bidi space loss
- Creating HTML Pages in Arabic, Hebrew and Other Right-to-left Scripts
- CSS vs. markup for bidi support
- How to use Unicode controls for bidi text
- Inline markup and bidirectional text in HTML
- Structural markup and right-to-left text in HTML
- Unicode controls vs. markup for bidi support
- Unicode Bidirectional Algorithm basics

The authoring techniques for handling bi-directional text recommend that web developers:

- Do not use Cascading Style Sheets (CSS) to control directionality. Mark-up should be used instead;
- Only add bi-directional mark-up to a document when it is needed. The Unicode bi-directional algorithm should be sufficient in most cases; and
- To change the direction of a block level element, add the dir attribute to that element. The content of all nested block elements will inherit directionality.

It is important to take care with bidirectional nesting. It is common in translations to leave some text in English or include the common English equivalent when the term is translated into the target language. Examples include government department names. Care should be taken to ensure that nested English content within a language written in a Right-to-Left (RTL) script renders correctly.

- Double check all punctuation is located correctly, especially mirrored punctuation like brackets and parentheses;
- Phone numbers should be treated explicitly as Left-to-Right (LTR) text; and
- Background images, and images for list markers should be checked to ensure appropriate placement and orientation within RTL text.

APPENDIX 1:
LANGUAGE TAGS

List of language tags for some of the languages used by Victorian Government departments and agencies, using valid BCP-47 language codes.

For written Chinese content it is best to use a language code based on the writing system used, either "zh-Hans" or "zh-Hant". For Audiovisual material, it is best to use a language tag that identifies the spoken language or dialect used.

Language	Tag
Albanian	sq
Amharic	am
Arabic	ar
Arabic, Juba	pga
Arabic, Sudanese	apd
Armenian	hy
Assyrian	aii
Bari	bfa
Bengali (Bangla)	bn
Bosnian	bs
Burmese	my
Cantonese	yue
Chinese, Simplified Zh-Hans	
Chinese, Traditional	Zh-Hant
Chin, Hakha	cnh
Croatian	hr
Czech	cs
Dari	prs
Dinka	din
Dutch	nl
Ewe	ee
Fanti (Akan)	fat
Fijian	fj
Filipino	fil
French	fr
German	de
Greek	el
Hakka (Kejia)	hak
Hazaragi	haz
Hindi	hi
Hmong Daw	mww
Hungarian	hu
Igbo	ig
Indonesian	id
Italian	it
Japanese	ja
Karen, S'gaw	ksw
Karen, Western Pwo	pwo
Khmer (Cambodian)	km
Kirundi (Rundi)	rn
Korean	ko
Kurdish (Arabic script)	ku-Arab
Kurdish (Latin script)	ku-Latn
Kurdish, Kermashani	sdh
Kurdish, Kurmanji	kmr
Kurdish, Sorani	ckb
Lao (Laotian)	lo
Macedonian	mk
Malay	ms
Maltese	mt
Mandarin	zh (or cmn)
Nepali	ne
Nuer	nus
Oromo	om
Pashto	ps
Persian (Farsi)	fa
Polish	pl
Portuguese	pt
Punjabi	pa
Rohingya	rhg
Romanian	ro
Russian	ru
Samoan	sm
Serbian	sr
Shilluk (Chollo)	shk
Sinhala (Sinhalese)	si
Slovak	sk
Slovene (Slovenian)	sl
Somali	so
Spanish	es
Swahili	sw
Tagalog	tl
Tamil	ta
Tetum	tet
Thai	th
Tigrinya	ti
Tongan	to
Turkish	tr
Turkmen	tk
Twi (Akan)	twi
Ukrainian	uk
Urdu	ur
Vietnamese	vi

APPENDIX 2: INTERNATIONALISATION AND ACCESSIBILITY

Victorian Government websites must meet WCAG 2.0 (Level AA) requirements. When adding content in community languages it is also necessary to meet accessibility requirements. The obvious accessibility requirements relate to identifying the language of content and change in languages, but there are a number of stumbling blocks in providing accessible content in community languages.

Other core internationalisation best practice, such as the need to correctly select and identify the character encoding used by text, or applying appropriate bidirectional markup and control characters, that affect the readability and comprehension of the text, are assumed but unarticulated in WCAG 2.0.

Legacy and pseudo-Unicode encodings (HTML, MS Word and PDF)

WCAG 2.0 makes an important distinction between text and non-text content. Text is a string of characters in a human language that can be programmatically determined.

For accessible community language content, it is necessary to select and correctly identify the character encoding used within a document. For HTML, it must be an encoding supported by web browsers. The HTML5 Encoding specification identifies which encodings a user agent can support.

If the character encodings are unsupported, or misidentified, the content should be treated as non-text content when assessing the accessibility of web resources.

What this means in practical terms is that translated content, regardless of file formats, should be sourced from language service providers as Unicode text. HTML documents must be in the UTF-8 character encoding.

It is common to receive translated content in certain languages in a non-Unicode character encoding.

For instance, Burmese content is often supplied in the Zawgyi pseudo-Unicode encoding, while Sgaw Karen is often supplied in an unsupported eight bit legacy encoding.

Using non-Unicode content (either legacy or pseudo-Unicode encodings) will often require additional steps to make the content accessible.

MS Word specific considerations?

Care needs to be taken with language identification in Microsoft Word documents as some community languages are not supported by Microsoft Office. It may not be possible to correctly tag all translations, thus impacting on the accessibility of the document.

You can use the document properties dialog to set a metadata value identifying the document's language.

MS Word will automatically assign the default editing language as the document language. If the document is opened on another computer where the MS Word default editing language setting is different, the document language will be changed when the file is saved.

When English content is included in a translation, it is necessary to change the proofing language appropriately. For translations written in scripts that are read from the right to the left of a page, it is necessary to set the direction, not just for paragraphs, but also for sections, columns, tables and text boxes. It is not sufficient to only use text alignment.

PDF specific considerations

ISO 14289-1:2014 and PDF techniques for WCAG 2.0 document requirements and techniques for creating accessible PDF files.

For a PDF file to accessible text, the textual content of the PDF must resolve to Unicode. Software that accesses or displays PDF files, uses the file's ToUnicode mappings for each font to resolve glyphs to Unicode codepoints. The ability to correctly resolve text in a PDF to a valid sequence of Unicode characters is dependent on the font, its internal mapping of glyphs to codepoints, and also on the nature of the writing system (script) the language is written in. Fonts designed for complex scripts may reorder glyphs and use alternative glyphs in ways that cannot be adequately represented in the ToUnicode mappings.

When the text in the PDF cannot be resolved to a meaningful Unicode sequence the user can understand, first try alternative fonts to see if they provide a better result. Otherwise, it is necessary to treat the content as non-text content and add ActualText attributes to each of the relevant tags.

Website search tools need to use the content of the ActualText attributes for indexing and searching these PDF files, in order to make the content discoverable.

Summary

I18n	HTML5	WCAG 2.0	Recommendation
Declare character encoding	charset attribute on meta element.	refer to the definition of text vs non-text content	Use Unicode for all text. For HTML documents use the UTF-8 encoding. For PDF files, use ActualText attributes of tags for languages that require it.
Declare language of document	lang attribute on root element	3.1.1 Language of page	Use a valid and correct BCP-47 language tag to identify the primary language of a document. For MS Word documents ensure the default editing language is set correctly.
Declare change of language	lang attribute of relevant element-version="1.0" encoding= "utf-8"?>	3.1.2 Language of parts	Use valid and correct BCP-47 language tags to identify change of language within a document. For MS Word documents select the appropriate proofing languages for content.
Bidirectional support	dir attribute on relevant HTML elements	–	For HTML5, use markup rather than CSS to handle bidirectional text. Use control characters as required. For other file formats use appropriate techniques available when editing content.

APPENDIX 3: WEBSITE LINKS

The website URLs that appear in these Guidelines are listed below:

Guidelines and standards

Effective translations – http://www.multicultural.vic.gov.au/projects-and-initiatives/improving-language-services/standards-and-guidelines

ISO 14289-1:2014 – Document management applications – Electronic document file format enhancement for accessibility – Part 1: Use of ISO 32000-1 (PDF/UA-1) – http://www.iso.org/iso/catalogue_detail.htm?csnumber=64599

PDF techniques for WCAG 2.0 – http://www.w3.org/TR/2014/NOTE-WCAG20-TECHS-20140408/pdf.html

Victorian Government digital standards and digital design principles articulate the principles of government website management and are available at: vic.gov.au/digitalstandards

Iconography

National interpreter symbol – http://multicultural.vic.gov.au/projects-and-initiatives/improving-language-services/interpreter-card-and-symbol

Web internationalisation

Getting Started with the W3C I18n site – http://www.w3.org/International/getting-started/

Internationalization techniques: authoring HTML & CSS – http://www.w3.org/International/techniques/authoring-html-dynamic

W3C internationalization checker – https://validator.w3.org/i18n-checker/

Character encodings

Introducing Character Sets and Encodings – https://www.w3.org/International/getting-started/characters

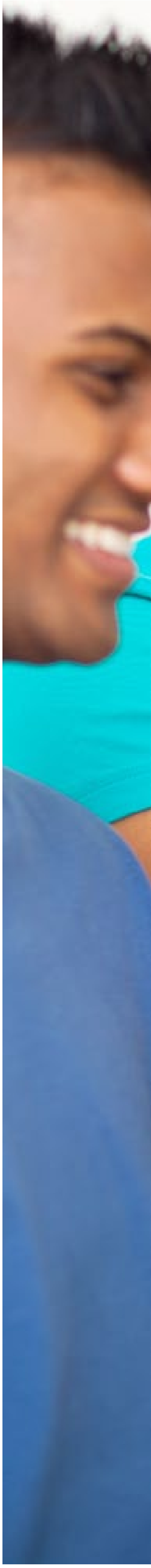
Character encodings for beginners – https://www.w3.org/International/questions/qa-what-is-encoding

Character encodings: Essential concepts – http://www.w3.org/International/articles/definitions-characters/

Choosing & applying a character encoding – https://www.w3.org/International/questions/qa-choosing-encodings

Declaring character encodings in HTML – https://www.w3.org/International/questions/qa-html-encoding-declarations

Declaring character encodings in CSS – https://www.w3.org/International/questions/qa-css-charset



Language tagging

Choosing a Language Tag – https://www.w3.org/International/questions/qa-choosing-language-tags

Declaring language in HTML – https://www.w3.org/International/questions/qa-html-language-declarations

HTTP headers, meta elements and language information – https://www.w3.org/International/questions/qa-http-and-lang

IANA Language Subtag Registry search tool – http://r12a.github.io/app-subtags/

Language on the Web – https://www.w3.org/International/getting-started/language

Language tags in HTML and XML – http://www.w3.org/International/articles/language-tags/Overview

Why use the language attribute? – https://www.w3.org/International/questions/qa-lang-why

Working with language in HTML – http://www.w3.org/International/tutorials/language-decl/

Text direction

Bidi space loss – https://www.w3.org/International/questions/qa-bidi-space

Creating HTML Pages in Arabic, Hebrew and other right-to-left scripts – https://www.w3.org/International/tutorials/bidi-xhtml/

CSS vs. markup for bidi support – https://www.w3.org/International/questions/qa-bidi-css-markup

How to use Unicode controls for bidi text – https://www.w3.org/International/questions/qa-bidi-unicode-controls

Inline markup and bidirectional text in HTML – https://www.w3.org/International/articles/inline-bidi-markup/

Structural markup and right-to-left text in HTML – https://www.w3.org/International/questions/qa-html-dir

Unicode Bidirectional Algorithm basics – https://www.w3.org/International/articles/inline-bidi-markup/uba-basics

Unicode controls vs. markup for bidi support – https://www.w3.org/International/questions/qa-bidi-controls

Authorised and published by the Victorian Government,
1 Spring Street, Melbourne
Printed by Doculink Australia, Port Melbourne

ACCESSIBILITY

If you would like to receive this publication in an accessible format, such as large print or audio, contact the Department of Premier and Cabinet on email:
language.services@dpc.vic.gov.au

This document is also available in PDF and Word format on the internet at
www.multicultural.vic.gov.au

Images by John Palermo Photography