# Making documents accessible

## Introduction

This document is about making documents and more specifically their content, accessible to people with disabilities or impairments, who may engage with the document's content using the help of assistive technology, for example a screen-reader.

Making an accessible document does not mean making a different type of document. It just means being systematically correct in how it is put together. So, it is about best practice.

Beginning this document with a simple statement of what it is about is a deliberate illustration of the use of clear language to explain purpose and convey information in documents, which is always helpful in terms of their usability and accessibility.

- This document outlines the principles of making a document accessible.
- Whilst it references the practicalities of making common document types accessible using typical authoring tools, it is an introductory document and does not go into definitive detail.
- It should be a sufficient guide for general communications and non-specialist document types, if used together with the further resources on making documents accessible using commonly used software that are suggested and linked to at the end of this document.
- Some of the more advanced possibilities are briefly remarked on, because it is relevant to know that they exist.

## Overview

Documents are most accessible if they are correctly structured and have their presentational values separated from what they are actually about. Web technologies do this best because they semantically describe the document structure and allow for complete separation of content and design (when made properly). So, starting from that position—with well-made accessible content, for example in XML, XHTML or HTML5; and then converting to other document formats such as Word or PDF; usually produces derivatives that are also very accessible.

Unfortunately at present most documents are not made that way and begin their life in word processing, or graphic design software that was originally directed at print solutions, so structure and presentation get jumbled up in varying degrees, sometimes quite badly.

However, if Microsoft Office, or the Open Office suite, is used in much the same way as web document mark-up would be applied, they will save and convert documents to reasonably accessible output. In Microsoft Office 2010 and later there is an accessibility checker that will point out significant errors. If these are addressed and the document is well structured to maximise its accessibility in Word, then a save as PDF from Word should also be reasonably accessible.

What makes a PDF accessible is the presence of tagging, which is almost identical to the mark-up used in web languages. In Open Office, there is a box to tick when outputting to PDF that carries the document structure across as PDF tags.

# The essential principles

The principles of making a document accessible are relatively straightforward:

1. Structure the document well, so that its organisation reflects its content:
    a. This means the hierarchical use of headings
    b. The proper use of paragraphs
    c. The correct use of different types of list
    d. Nesting these together in a way that relates items of connected content
    e. All of this could be described as making sure the document has a logical flow
        i. It could for example look like this:

# This is a Heading 1

Some content in a paragraph, or several paragraphs, establishing what the document is about.

## This is a Heading 2

Some more content in a paragraph, or several paragraphs, that represents a subsection, or discrete area, of the subject that the document covers.

- A list of bullet points that follow on from what is being said
- And that relate to this subsection of content

### This is a Heading 3

Further content representing an internal subsection of the subsection it is in.

## Another Heading 2

This is the next logical discrete topic area. In the hierarchy of the document it is below the overview presented under Heading 1 and has equal status to the content in the previous Heading 2 subsection, however, in the logical flow of the content delivery—the sensible order of giving or explaining the information, it comes after it.

*Note 1: There should only ever be one Heading 1 in a document. This one actually contains two, the second being included to illustrate a good document structure in the example given here. This note identifies itself as such, so would be understood as a note without knowing that it is italicised.*

*Note 2: The correct use of paragraphs seems simple enough in principle though isn't always done. Most word processors and similar tools end a paragraph and create a new one using the return key. Some users add spaces with the space bar, or put in line breaks, instead to achieve the layout appearance they want, or add empty paragraphs to adjust the spacing. This is likely to make a document less accessible.*

*Note 3: More complex content often requires the nesting of lists, which means a structure where another list forms part of the content of a list item*

*in a parent list. This is done in this document. Most word processors and similar tools either have a multi-level list option to do this, or will start a nested list by indenting the next list item with the increase indent tool.*
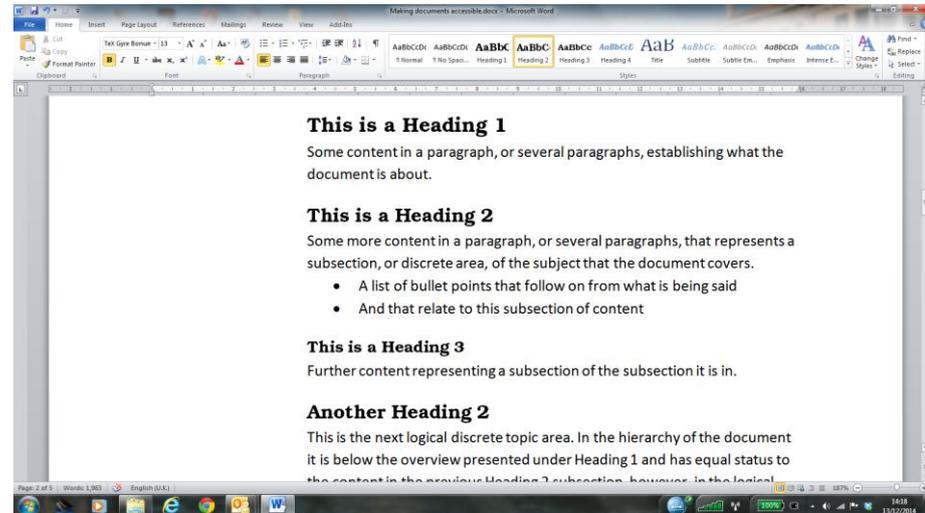
## Illustrations



**Figure 1: The Word 2010 interface with this document loaded, showing the Styles options used for Headings and other structured content across the top right and options for lists to the left of centre, also at the top. These are all in the Home tab of Word. (This information also appears in the 'Alt Text' description of the image, which is found by right clicking on it and selecting the 'Format Picture' dialogue.)**
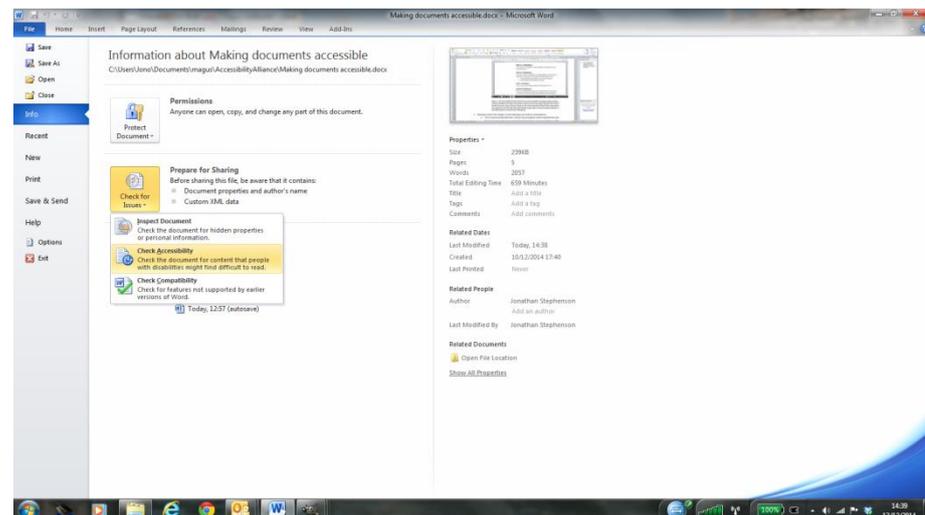


**Figure 2: The Word 2010 interface, showing the File tab, with check accessibility selected under Check for Issues, which appears as a dropdown menu item under Prepare for Sharing. This document reports no accessibility problems.**

## Content and design

2.  Separate content from design, so that meaning is not reliant on presentation:
    a.  This is interconnected with item 1 above. Structuring the content identifies the role of each element in the document and its relationship to what comes before or after it. This is independent of what it looks like, so makes sense regardless of how the document is accessed.

b. Present the document using styling. This means applying a consistent style to each item in the document hierarchy telling it what to look like.

   i. Here is the difference:

## This is a Heading 2

This a real level 2 heading that identifies itself as such to assistive technology and looks as it does because the styling of the document tells all Heading 2 headings to look like this.

## This is not a Heading 2

This is just text that has been made bigger and bolder by the document's creator, so that it may be interpreted as a significant heading by the reader that can see it; however, if the user cannot see it then the presentation is meaningless and it is not identifying itself to assistive technology as being a heading. Neither does it have an exact position or hierarchical relationship in terms of the document's structure.

### Alternative routes, flow and context

3. Provide alternative routes to the information:

a. Alternative descriptions are important where content is potentially inaccessible to some users, or their devices. The most obvious example is an image, which if it conveys important information also needs that to be explained in other ways. Web languages use an 'alt' attribute, or a separate description tag to do this, which is available to assistive technology though otherwise hidden. Some word processing software allows something very similar to be done by adding a title and alternative description when an image is added.

   i. It generally makes sense to add a caption to an image in a document as well as giving it an alternative description, since that will also be informative for all users about what the purpose of the image is; and it may be that what it is illustrating, rather than what it looks like, is the more important factor.

b. Information in tables may be difficult to interpret if the table cannot be seen, or is large and complex. So, only use tables where they are necessary to display tabular data and always make sure that there is a 'Header Row' enabled, as that defines the columns in a more accessible way.

c. Different disabilities and impairments have different obstacles to overcome and it is difficult to predict all scenarios, or what assistive software any given user may have. A way of dealing with this is to offer the accessible document in different formats, for example as web content, in a cross-compatible word processing format and as a PDF version. If the parent document is accessible then derivatives of it in alternative formats should be too.

4. Documents should flow. The content of a document should follow a logical sequence from beginning to middle to end in a linear fashion. This is sometimes referred to as content being linearized, simply meaning that each item of content follows another in an ordered sequence. Content constructed like this behaves like a liquid and 'flows' into whatever it is

put into, taking on its shape, just like water in a glass. So, the end-user's screen size, or the capabilities of their device, or the media channel they are using and many other obstacles to access become irrelevant.

    a. Changing the font size, or the background colour, or the text being read out by a screen reader, should make no difference to flowing content.

    b. Well-made digital content has flowing content underneath, which may be presented visually as though it is not, by means of a linked style-sheet. This means the same content presents itself to different users in different ways. That makes it inherently accessible to the disabled and impaired.

    c. Content designed for print is fixed in concept and by the page layout. As a result it is often less accessible or even inaccessible. The document structure is governed by what it looks like, which may result in content out of sequence, a lack of formal structure underpinning it and 'knocked over' content, adjusted for visual effect.

5. Where the format permits be aware of context and semantics:

    a. Context: This mostly applies to web technologies and related technologies like eBooks where it is possible to do—however, that is a very substantial content area that is continuing to expand exponentially and that is the true digital environment. As it is possible to derive all other formats from this starting point it will inevitably take over. Context also applies to the use of hyperlinks to online or other digital content when they are included in any type of document.

        i. Context here is usually a reference to what a thing does rather than what it looks like. So, for example the most useful alternative description of the logo on a website is typically going to be 'return to home page', not a description of the image.

        ii. Italic and bold describe what certain items of text are presented like in print and are simply descriptions of the font style. Web technologies allow for tags meaning 'emphasise' and 'strong' in their place, which have contextual meaning focusing on their significance when a document is being interpreted with the help of assistive technology. The Styles in Word also permit *Emphasise* and **Strong** to be applied, as they have been here. As a rule of thumb avoid any formatting that is not done with styles.

        iii. Perhaps the most universal consideration is this: When a hyperlink to other content is included in any kind of document it is important to title or describe it contextually, so that the end-user understands what it points at. A web address is often not particularly meaningful to the human reader. The links to further resources at the end of this document are used to provide examples of this.

    b. Semantics: This also only applies to web technologies and related technologies like eBooks where it is possible; and will become more so going forward, as the Semantic Web evolves. As examples:

        i. HTML5 includes elements of mark-up that identify the purpose of content within a document. These are not visible to the end-user when the content is presented to them; however, significantly extend the structural organisation that is possible. This means elements of content may indicate

their role and relevance in relation to the document as a whole to assistive technology (and therefore the user) in far greater detail.

ii. The eBook specifications for ePUB3 (which is for interactive and media enhanced eBooks) and for EDUPUB (which is specifically for eTextbooks) have quite extensive possibilities for semantic mark-up. This has the potential to be game-changing for accessibility.

## Further resources

1. More detailed information on using Word for accessible documents
2. Further information on using the Microsoft Office suite for accessible documents
3. An article on using Open Office to create accessible documents
4. An article on using Microsoft Word to create accessible documents
5. An article on converting documents to accessible PDF

Whereas this document outlines the general principles of making any document accessible and picks out some key points to consider, these further resources (above) give more detailed instruction on how to actually achieve that with particular software programs. Accessibility when using Excel files for example is also included in the content that second link points to.

Note that the hyperlinks given here are attached to meaningful descriptions of what they point at and not simply the web address. A web address describes where something is, rather than what it is—and some of them, particularly if system generated, are extremely long.

Also consider testing with a screen reader: This is a valuable exercise both as a means of flushing out potential problems and from the perspective of understanding the challenges faced by some end-users. The NVDA (Non Visual Desktop Access) screen reader is free. A small donation is suggested, though is not obligatory. Download the NVDA screen reader here.

*Note 4: A final point to note is that the levels of accessibility support in many current document authoring and design tools is a retro-fix development that makes them more capable of delivering an accessible outcome. They were not originally conceived to do this though and it is questionable if they will ever be perfect at it. This is particularly evident when web output is generated from them, which is rarely of a comparable standard to web and digital mark-up produced from a code editor, done directly in the mark-up language and intended to be accessible.*

*Attempting to deliver accessible web content by converting documents directly from word processing or design software cannot be recommended. The same applies to eBooks, which use web technologies and are best authored as such.*

*This highlights a major stumbling block to accessibility, which is that the majority of documents and their content are being created by the wrong skillsets, in outdated workflows, using the wrong technology. The digitally enabled publishing processes that began in the 1980s and 1990s are focused on print production and have very little to do with the present day digital environment or the technologies that should be being used in the future of digital publishing.*

# Disclaimer

This document on Making Documents Accessible is for general information and guidance purposes only and is intended simply as an introduction to and overview of the subject area—and of surrounding and interrelated influences.

Its commentary represents well-reasoned professional opinion based on the author's knowledge and expertise and an informed interpretation of the available resources at the time of writing, some of which may be derived from third-party content from trusted sources that has not been independently verified.

Readers of this document should exercise their own judgement in evaluating its worth for their own purposes. The author shall not be responsible for any interpretation put on it, or the consequences of relying or acting on it in any way.

It is presented 'as is' with no warranties expressed or implied

**Jonathan Stephenson**

**Magus Digital**

[www.magusdigital.com](www.magusdigital.com)

[jps@magusdigital.com](jps@magusdigital.com)