Guidelines for Accessible E-text

November 2009

Round Table on Information Access for People with Print Disabilities

Guidelines for Accessible E-text

About these guidelines

These guidelines are published by the Round Table on Information Access for People with Print Disabilities Inc. The Round Table is an umbrella organisation which brings together producers, distributors and consumers of information in alternative formats to print; blindness agencies, tertiary institutions and government departments in Australia and New Zealand.

These guidelines are available from Round Table in accessible formats.

Acknowledgements

Contents

Introduction to these guidelines

What is e-text?

E-text is structured electronic text which is accessible to people with a print disability, that is, to people who can't access information from regular print.

E-text may be an alternative version of a print document, that provides a print-disabled person with equivalent access to the information. The term "e-text" can also refer to a document that is originally created in digital format, and is accessible to all readers.

People with print disabilities use a range of technologies to access e-text. Typical reading methods include:

Synthetic speech - using screen reader software like JAWS or Window Eyes, or speech built into software like Adobe Acrobat.

Refreshable braille - using a portable braille display attached to a desktop computer or laptop, or a stand-alone braille notetaker.

Viewing on screen - using screen enlarging software, modified colour combinations, tools that visually highlight words while they are read with synthetic speech, magnifiers and zoom functions built into different software or operating systems, or not using any special software.

Typical formats for accessible electronic text include:

- **plain text**.
  This may include textual "markup" to indicate structure (e.g. "Round Table" markup, Markdown)
- **word processor documents**.
  Common formats are Microsoft Word, RTF or OpenDocument
- **HTML** and **XHTML**.
- **DAISY books**.
  Can contain the full structured electronic text of a book, and may contain audio.

About these guidelines

These guidelines have been produced to provide document creators with an understanding of accessibility principles, and some best practice accessibility methods across a variety of electronic formats in common use.

In a print-based society, people are significantly disadvantaged if they are cannot access information in print format. The trend towards mainstream electronic communication provides the opportunity for equity of information access, but the reality is that in 2009, many electronic documents are not designed to be fully accessible to readers with a print disability. Electronic documents need to be created with a standards-based approach to ensure accessibility.

The Round Table's goal has been to make the guidelines clear and easy to read in the hope they will be widely adopted.

Background

Version 1.0 of **Guidelines for preparation of text materials on computer disk for people with print disabilities** was produced by a sub-committee of the Round Table in July 1995.

In 2007 the Round Table Committee set up a working group to review the guidelines, inviting representation from people with experience in developing accessible formats. The group's purpose was to bring the guidelines up-to-date with newer technologies and to reflect the increase in end reading formats and editing processes as well as the increasing practice of converting electronic documents between different formats.

The new version reflects input received on earlier versions and takes into consideration the Australian Government style guide, as well as e-text guidelines produced by other organisations around the world.

In compiling the guidelines the working group sought input from people with experience in preparing different e-text formats and consulted widely with professionals and people with a print disability to determine the methods currently considered best practice.

Related standards, guidelines and production manuals were reviewed, to ensure that the draft guidelines were consistent where applicable. These included Round Table's **Guidelines for Conveying Visual Information**, the World Wide Web Consortium's **Web Content Accessibility Guidelines**, the DAISY standard, and production manuals from Vision Australia and RNZFB.

Workshops on the draft guidelines were presented at the 2007 and 2008 Round Table Conferences to stimulate discussion on the draft and feedback from participants was incorporated into the document.

Scope

The guidelines describe basic principles for preparing accessible e-text documents which apply regardless of which file format is used. These principles will mean people with a print disability are able to use and understand the content and have equivalent access to the information contained in a document.

The guidelines then elaborate on how these principles apply to text in particular file formats. The new guidelines focus on e-text as an end reading format to be used by people with a print disability rather than as a base file that can be converted to other formats, but many of the principles of good document formatting will mean that an e-text document will convert more easily to braille or large print.

Who these guidelines are for

Anyone preparing electronic documents will find the guidelines useful. They have been written for professional transcribers as well as a general audience. Possible uses include:

- Specialist accessible format production agencies and educators may use the guidelines when converting printed text into accessible electronic versions
- Government agencies may reference the guidelines when preparing public documents
- Organisations may adopt the guidelines as part of their internal procedures for publication of reports and documents
- Conference organisers may specify that abstracts and papers should comply with the guidelines
- People with a print disability may find the guidelines useful in explaining their accessibility requirements

General principles

Summary checklist

1. Equivalent to print

- Include all meaningful elements of the print document.
- Ensure accuracy.
- Mark changes to the print with producers' notes.
- Include metadata.

2. Accessible

- Arrange text in a linear reading order.
- Include structural markup.
- Verbalise images and visual elements.
- Express special characters and languages unambiguously.

3. Clear visual style

- Use a clear visual design.
- Allow users to control visual appearance.

4. Standards, Guidelines and Best Practice

- Follow applicable standards
- Adjust for individual need and preference
- Ensure consistency

1. Equivalent to print

An accessible e-text version of a print document should provide the same information as the original print. Differences from the print, or producer's interpretations of visual material, should be marked with a producer's note.

1.1 Include all meaningful elements of the print document

For an e-text document to be an accessible equivalent of print, all important content in the print document must be included.

When producing a full book, information from the cover and preliminary pages such as the synopsis, review quotes, and notes about the author should be included as part of the e-text document. Common practise is to include this at the beginning of the document.

Print page numbers are important as they allow an e-text reader to navigate to a particular page number, use the index and contents list, or reference which page a particular quote or concept is from.

A note on formats: DAISY has a page number structure which differentiates between normal numbered pages, preliminary pages and pages in a special sequence, like a series of images inserted between two numbered pages. Readers can use DAISY software to navigate directly to a page. In other formats, include the print page number on a separate line, marked by words (Page, Print Page), letters (p, pp) or Round Table-style codes (<pp> #, <p #>).

1.2 Ensure accuracy

The e-text version must provide equivalent information to the print - this requires careful checking of spelling.

Make sure correct characters are used, for example, don't confuse O and 0, 1 and l. **Appendix 5: Tips for using Scanners and OCR** contains a list of errors that are commonly introduced in the OCR process.

Remove hyphens that appear at the end of print lines (unless the word would otherwise be hyphenated).

1.3 Mark changes to the print with producers' notes

A producer's note should be used to comment about any changes made to the book when converting it to e-text, such as formatting changes or omissions. For example, if a

line of text is unreadable or omitted from the print, insert a producer's note such as "The following line was not legible in the print and has been omitted from this e-text version."

Producers' notes can appear at the beginning or end of the file, or within the text where the note relates to a particular point.

Make it clear where text is a producer's note rather than part of the original text. DAISY has its own structure for this, and Round Table markup uses <tn> and </tn>. The words "producer's note" and "end of producer's note" can work in other formats.

1.4 Include metadata

Publishing information related to the original print document should be included either within the e-text file itself or in a separate "readme" file. This information should include the title, author, publisher, copyright owner, special notes about the e-text version (e.g. omissions), transcription date, any identifier of the text like ISBN or ISSN and the identity and copyright permissions of the producing agency.

2. Accessible

An accessible e-text document must be fully usable by a person with a print disability, in the sense of perceivable, operable, understandable and robust (http://www.webaim.org/articles/pour/)

2.1 Arrange text in a linear reading order

Text should be arranged so that it can be read in a linear way, that is, pieces of information follow each other in logical sequential order. For example, "floating" textboxes should be incorporated into the flow of the text.

Footnotes

- refer to the position of the footnote in the text.
- avoid putting footnote text where it appears visually on the print page (i.e. at the end of the page) as this can interfere with reading and comprehension.
- common practice is to move all footnotes to the end of the file.
- alternatively, footnotes can be presented inline with the original text, that is, the footnote text can appear in the original position of the footnote reference, or at the end of the relevant sentence or paragraph. However, depending on their length, inline footnotes can interfere with reading and comprehension.

A note on formats: DAISY has its own footnote structure which provides an ideal way of representing footnotes. Readers can choose to read a document with footnotes

switched on or off, and can navigate directly to a footnote. Word footnotes are not reliably accessible with the range of adaptive technology currently in use. In HTML, a common practice is to make the footnote reference a link to the footnote text, which appears at the bottom of the document. The footnote text then includes a link back to the related reference.

Text boxes and margin notes

If a print document contains text boxes or margin notes these should be inserted in the most logical place in the text, within the linear flow of the document. Avoid using features like "Text Boxes" in Microsoft Word which separate content from the main reading order.

The beginning and end of each text box should be marked in some way so that it can be perceived by someone using a screen reader.

A note on formats: Text boxes could be represented in Rich Text Formats by marking the beginning and end of the box with distinctive styles, marking the contents of the box with a distinctive style, or putting the box inside a 1-column, 1-row table. In DAISY, use the "sidebar" structure. In other formats, the words "box" and "end of box" may be useful.

Headers and footers

Print documents often include title or pagination information in the header and footer sections of the page. An e-text version should only include this information where it is informative and not repeated elsewhere in the text. For example, page numbers should be included in an e-text version, but a running chapter heading that appears on each page of a chapter should not be included, as the heading will have been included at the beginning of the chapter.

2.2 Include structural markup

Structural elements of the text, such as headings, paragraphs, lists and tables, should be "marked up" so that adaptive technology can interpret their significance, meaning and context. Structural markup methods include using "Styles" in word processing software, Round Table codes in plain text documents, or semantically-correct HTML elements.

For example, mark the main headings in a Word document with the Heading 1 style, then modify the visual appearance of that Style, rather than selecting each heading individually and applying visual markup (bold, bigger text size). In HTML, use structural

elements like <h1> and control the visual appearance with stylesheets, rather than <font size=+2>.

The resulting document might **look** exactly the same but the significance of the visual structuring can be interpreted for different reading methods. For example, a screen reader can extract a list of headings in a document and use them for navigation.

Structural elements that should be marked up include headings, paragraphs, lists, tables, and emphasised words.

Headings styles are usually hierarchical. Most texts can be divided into main sections such as chapters, then into smaller sections, then into sub-sections, and so forth. Headings are used to label each piece of the text. The major headings through the document should be assigned Heading 1. Heading 2 applies to subheadings, and so on. HTML allows a maximum of 6 levels of heading.

Common structures should be treated consistently, for example all level 2 headings should be structurally marked in the same way, even if they don't look the same visually.

2.3 Verbalise images and visual elements

Where an image or visual element adds meaning to the text, that meaning should be expressed in words, or verbalised, within the accessible e-text file.

Text explanations can be included as a producer's note in the text, as "alt" attributes attached to an image file or as captions below an image.

For some readers, it may be more appropriate to provide images in an alternative format such as a tactile or enlarged graphic. If an image has been produced as a physical enlarged or tactile image, this should be explained with a producer's note.

The key to conveying visual information in e-text format is to interpret what information the illustration is adding to the text, then describing this information in a simple, structured and straightforward manner. A good guideline is to imagine reading the document to another person, and think about what would be said in place of the image.

Some general guidelines for verbalising images are:

- Identify the key concepts conveyed in the image and explain these.
- Describe the overall image first, then provide further detail as necessary.
- Organise the details of descriptions in a linear order, for example from left to right or following a process.

- Separate information into easily readable pieces by using bullet points or line breaks.
- Include all titles, captions and labels provided in the original text.
- Avoid making judgements, assumptions or interpretations of the images as much as possible. If a judgement is required, state within a producer's note that it is the producer's interpretation.

Refer to the Round Table's **Guidelines for Conveying Visual Information** for more detailed examples of verbalisation.

2.4 Express special characters and languages unambiguously

Special characters

Ideally, use a Unicode character set and represent all special characters in the document as their correct Unicode equivalent. Unicode is a widely used cross platform standard for representing characters in any language that is backwards compatible with plain ASCII. Representing a character correctly in Unicode is less ambiguous than replacing the character with words.

At the time of writing these guidelines, assistive technologies do not reliably interpret some of the Unicode characters, for example the numero sign № is not reliably read by current screen readers. If a document is being prepared for a particular person whose technology cannot interpret Unicode, replace characters with words (for example, replace ° with "degrees", £ with "pounds").

Avoid using proprietary Windows characters like "smart quotes" (change to " and ') and ellipses (change to …). Change em and en dash characters to plain hyphens, and ligature characters to single letters.

Languages

In word processing formats, HTML and DAISY, it is possible to indicate which language a document is written in. In these formats, ensure that the language is set correctly and that changes in language (for example, French words in an English document) are marked consistently. This allows synthetic speech tools to read content more accurately.

Mathematics

Mathematics can be conveyed unambiguously using MathML or LaTeX. The usefulness of these formats depends on the user's reading technology and familiarity.

3. Clear visual style

An accessible e-text document should have a clear, legible visual style that considers the needs of people with a range of print disabilities.

3.1 Use a clear visual design

Clear visual design of an electronic text document for viewing on screen requires similar design decisions to creating clear print documents. General principles are:

- use a simple and clear typeface.

- use an adequately large default text size.

- use good spacing between lines, paragraphs and columns.

- use adequate colour contrast between text and background. Black text on a white background generally provides the best contrast. Avoid placing text over background images.

- ensure all text is left-aligned and set horizontally.

- use a consistent and logical layout. Unusual spacing or layout can make it difficult to locate information on the page.

- avoid using italics, underline and blocks of capital letters to emphasise words.

- do not convey information solely through images or colours. Provide information in text as well.

- give tables a bold, visible border and ensure adequate space between the border and contents of a cell.

- ensure that structural elements like headings and lists stand out and have a consistent appearance.

3.2 Allow users to control visual appearance

Use stylesheets to control text and background colour so these can be adjusted by the user. In formats where it is possible, use relative rather than absolute text sizing so that the reader can easily resize text.

4. Standards, Guidelines and Best Practice

An accessible e-text document should follow relevant standards and best practice where they exist, only deviating from these where they are needed for particular reasons such as limitations in adaptive technology. A document that is produced according to standards is more likely to work with a wider range of adaptive technology, and is easier to adapt - whether to other e-text formats including webpages or to other formats like braille, large print or synthetic audio.

4.1 Follow applicable standards

Standards that may be applicable include:

- a semantic approach to HTML: http://en.wikipedia.org/wiki/HTML#Semantic_HTML

- WCAG (Web Content Accessibility Guidelines): http://www.w3.org/WAI/

- DAISY: http://daisy.org/

- Unicode: http://unicode.org/

4.2 Adjust for individual need and preference

When a file is being prepared with a particular person in mind, individual need should take precedence over standards. If a file might be used by multiple people, this can be achieved by producing the file in a standards-based way and customising it for individual need. An example is using Unicode to represent special characters, then replacing these characters for people whose technology can't interpret them. Some examples of where there would be an individual need or preference include:

- adaptive technology limitations e.g. a screen reader that is unable to read Unicode.
- software availability e.g. a person might not necessarily have Microsoft Word.
- varying levels of computer literacy.
- reading preferences.

4.3 Ensure consistency

These guidelines are not a prescriptive formatting guide. Sometimes a producer or document author will need to make choices about document-specific formatting, such as heading styles. An important guiding principle is consistency - within a document or across a set of documents. For organisations creating digital documents regularly, it may be helpful to develop an organisational style guide to help maintain consistency.

Particular formats

HTML

HTML is the language in which the content of most webpages is written. It is simple to create, highly navigable and supported by most adaptive technology.

HTML elements wrap pieces of text content within short codes that provide information about the structure and meaning of content. For example:

<h1>This is a top level heading</h1>

This is a list item

HTML is designed to describe the structure of a document, rather than its appearance. For example, the HTML table element implies a data table and the blockquote element indicates that content inside this element is a direct quote. Adaptive technology can interpret information from structural elements and change behavior accordingly - for example, most screen readers can extract a list of headings from a document and use them like a table of contents.

Once the content has been properly marked-up, all the HTML elements can be styled using CSS, and in this way can be easily customised to achieve a desired presentational effect, such as formatting in columns or indenting blocks of text, without affecting the meaning of the markup. Separating the content structure (HTML) from the layout/appearance (CSS) means that visual style can be adjusted easily for different users.

Check documents carefully when converting to HTML from another format, as many built-in conversion tools also embed visual formatting information within the document (for example, font colour and size).

When creating HTML documents:

- Use a semantic approach to writing HTML:
  http://en.wikipedia.org/wiki/HTML#Semantic_HTML
- Ensure documents are valid to the HTML standard: http://validator.w3.org/
- Use valid style sheets to control layout and presentation:
  http://jigsaw.w3.org/css-validator/
- Conform with the latest version of the W3C Web Content Accessibility Guidelines:
  http://www.w3.org/WAI/intro/wcag.php

Rich Text Format (RTF)

Word processing formats, such as RTF, Microsoft Word and OpenDocument are widely supported by current adaptive technology. It is preferable to use RTF as this is supported by a wider range of software, however MS Word or OpenDocument can be used if requested by individuals. RTF can be created in a variety of word processing programs, and files in Word and OpenDocument can be saved as RTF.

RTF can be useful for providing downloadable content on the web, so that a whole document is available as one file.

When creating documents in these formats, use "Styles" to add semantics and ensure consistent presentation.

General

- Use Styles to indicate structures like headings and lists, rather than formatting text using purely visual tools. Some adaptive technology can access Style information, so the reader can locate structures like headings and lists.
- Use bold text for emphasis.
- Arrange content so that it can be accessed in a linear order using the keyboard. Do not use floating "Text Boxes" in Word, as these cannot be accessed with the keyboard. Avoid using columns, as keyboard navigation can be difficult.
- If informative content is included in headers or footers, ensure this is repeated elsewhere in the document.

Automatic contents lists

- In a document originally created in digital format, an automatic contents list can be a useful way of adding navigation information.
- In documents that are converted from print, don't use automatic contents pages. A document converted from print contains two sets of page numbers (original print pages and on-screen pages in the word processor document), so automatic contents lists can be confusing.

Tables

- Use the table tools available in Word processing software.
- Ensure tables have a clear visual style (see General Principle 3.1).
- Avoid merging cells, as this makes keyboard navigation difficult.

Plain text

Plain text is the most universally accessible file format. It is useful when producers are unsure of what adaptive technology will be used to access the file.

Different conventions exist for formatting plain text documents.

- The Round Table's 1995 **Guidelines for preparation of text materials on computer disk for people with print disabilities** recommended formatting electronic text with a system of HTML-like "codes" (Round Table markup). Codes were a way of added information such as print page numbers, heading levels and word emphasis to a plain text document.
- **Markdown** is a human-readable markup system based on plain text email conventions: http://daringfireball.net/projects/markdown/
- **Project Gutenberg** has developed formatting guidelines based on internet conventions (http://www.gutenberg.org/wiki/Gutenberg:Volunteers%27_FAQ).

A note regarding Round Table markup

Round Table markup was designed to help a reader understand the structure of a document clearly. Codes consisted of characters enclosed in angle brackets which marked structures like headings, page numbers and diagrams.

Round Table markup is a way of adding additional information and structure to a document that is useful for people who are already familiar with these conventions. Codes add extra verbosity so it is important to use judgement as to the level of markup.

How codes should be applied

Codes should be separated from the surrounding text by a space or a new line. Where codes mark a block of text, such as a passage in italics, use a "start code" (e.g. <it>) and an "end code" (</it>).

Include a list of codes in a separate readme file, or at the beginning or end of the file.

Refer to **Appendix 3: Round Table markup** for a list of codes.

Document setup

- The first line of document should be the title (omit any blank lines before the title).
- If you have used Round Table or other text markup, provide a list of the codes at the beginning of the file, or in a separate readme file.

Line breaks

- Insert a blank line between paragraphs
- Leave two blank lines before all headings and one blank line between consecutive headings. Leave a blank line after all top level headings.
- Indicate the beginning of a print page on a new line. Leave a blank line before a page indicator.
- Blank lines can be added between list items or rows of a table if needed for clarity.

Emphasis

Different types of emphasis include block capitals, bold, italic and underline. These can often be omitted unless they are needed to understand the meaning of the document. A common internet convention uses asterisks and underscores for *bold* and _italics_.

Lists

- Retain numbering, letters and roman numerals to indicate list points. Insert a fullstop character after the number or letter to clearly differentiate it from the following text.
- Indicate bulleted lists with asterisks. For multi-level lists, use - for different levels.

Tables

Indicate the beginning and end of a table with the words Table and End of Table or the Round Table codes <table> and </table>. If the table has a title, use Table: Title

Use a producer's note to explain how the table has been presented.

Separate cells with a distinctive character (| is commonly used). Indicate blank cells with a hyphen (-) or a word like "nil" or "blank".

A basic way of presenting tables

Table

Producer's note: The following table has three columns with the headings: "A", "B", "C". The columns are separated by the | character. End of Producer's note.

A | B | C

End of Table

DAISY text

DAISY is a standard for structured digital books that has been developed by the DAISY Consortium, which is made up of print disability organisations around the world. A DAISY book can include audio, text and images. The treatment of text in a DAISY book is based on XML, and allows a closer representation of the print than any other electronic text format. For example:

- Footnotes can be marked as a separate structure that the reader can choose to navigate to directly, skip over or access inline with the text.
- A reader can navigate directly to a particular page number.

Full text DAISY books consist of marked-up text that can be accessed and navigated using DAISY software (http://daisy.org/about_us/dtbooks.shtml)

The DAISY website contains guidelines for producing DAISY books: http://www.daisy.org/

Appendix 1: A note about PDF

Portable Document Format (PDF) is a file format developed by Adobe that makes it possible to reliably create, combine, and control text and graphical documents with the original formatting intact.

Although there has been some recent improvements in the software used to create PDFs, this file format is not regarded as sufficiently accessible for documents designed for general distribution by the Human Rights and Equal Opportunity Commission (HREOC) or by other accessibility experts world wide including Royal National Institute of Blind People (RNIB) in the UK.

HREOC's position is: "Despite considerable work done by Adobe, PDF remains a relatively inaccessible format to people who are blind or vision-impaired. Software exists to provide some access to the text of some PDF documents, but for a PDF document to be accessible to this software, it must be prepared in accordance with the guidelines that Adobe have developed. Even when these guidelines are followed, the resulting document will only be accessible to those people who have the required software and the skills to use it. The Commission's view is that organisations who distribute content only in PDF format, and who do not also make this content available in another format such as RTF, HTML, or plain text, are liable for complaints under the Disability Discrimination Act (DDA). Where an alternative file format is provided, care should be taken to ensure that it is the same version of the content as the PDF version, and that it is downloadable by the user as a single document, just as the PDF version is downloaded as a single file." Source: Human Rights and Equal Opportunity Commission (August 2002). World Wide Web Access: Disability Discrimination Act Advisory Notes [Electronic Version 3.2]. Retrieved 29 August 2007 from http://www.hreoc.gov.au/disability_rights/standards/www_3/www_3.html#s2_3

Where several alternative formats of a document are provided, the best practice approach is:

- ensure the accessible format contains the same information as other formats.
- place the most accessible document first.
- include the download file format and file size as part of the link label e.g. document title (rtf 750 kb).

Appendix 2: Glossary

Accessible: Usable by a person with a print disability, in the sense of perceivable, operable, understandable and robust. http://www.webaim.org/articles/pour/

Alternative Format: Any text that has been reproduced in either large print, electronic text (e-text), braille, or audio format.

Adaptive Technology: Software programs that enable individuals with a print disability to view printed materials

Caption: Written description that normally accompanies a picture, chart or diagram

Character Set: A defined list of characters recognised by computer software. Each character is represented by a number. The ASCII character set, for example, uses the numbers 0 through 127 to represent all English characters as well as special control characters. European ISO character sets are similar to ASCII, but they contain additional characters for European languages. Unicode is a widely used character set for representing characters in any language that is backwards compatible with ASCII.

Colour Contrast: The contrast between the brightness and hue of text and background colours.

Cascading Style Sheets (CSS): A style sheet language used to control the presentation of a document written in a markup language. Its most common application is to style documents written in HTML and XHTML. CSS is used to define colours, fonts, layout, and other aspects of document presentation. It is designed primarily to enable the separation of document content (written in HTML or a similar markup language) from document presentation (written in CSS).

DAISY: An open standard for accessible digital books. http://www.daisy.org

Endnote: A note placed at the end of an article, chapter, or book that comments on or cites a reference for a designated part of the text

Footnote: A note placed at the bottom of a page of a book that comments on or cites a reference for a designated part of the text.

Large Print: Print that is enlarged and reformatted for clarity, designed to be optimally legible for people with low vision.

LaTeX: A document markup language and document preparation system for the TeX typesetting program, which can be used to lay out complex mathematical equations.

Markdown: A human-readable markup system based on plain text email conventions: http://daringfireball.net/projects/markdown/

MathML: Mathematical Markup Language (MathML) is an application of XML developed by the W3C for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulae into text documents. http://www.w3.org/Math/

Metadata: information **about** a document that is used to facilitate the understanding, use and management of the document.

Optical Character Recognition (OCR): The electronic identification and digital encoding of printed or handwritten characters by means of an optical scanner and specialised software.

Print Disability: An individual with a print disability is unable to access the information disseminated in a regular print format because they 1. are blind or vision impaired; 2. have physical disabilities which limit their ability to hold or manipulate information in a printed form; 3. have perceptual or other disabilities which limit their ability to follow a line of print or which affect their concentration; or 4. cannot comprehend information in a print format due to insufficient literacy or language skills

Producer: Someone responsible for the production of alternate format material

Producer's Note: A note included by the producer to indicate differences from the print or the producer's interpretation of visual material

Project Gutenberg: Project Gutenberg is the first and largest single online collection of free electronic books, or eBooks. http://www.gutenberg.org

Semantic Markup: using the most meaningful tag to describe the type of content.

Styles: A style is a set of formatting characteristics that can be applied to text in a document to quickly change its appearance.

Subscripts and Superscripts: a number, figure, symbol, or indicator that appears smaller than the normal line of type and is set slightly below or above it - subscripts appear below the baseline, while superscripts are above. Subscripts and superscripts are typically used in formulas, mathematical expressions, and descriptions of chemical compounds or isotopes, but have many other uses as well.

Tactile Graphics: Images that use raised surfaces so that a visually impaired person can feel them. They are used to convey non-textual information such as maps, paintings, graphs and diagrams

W3C: World Wide Web Consortium. The main international standards organisation for the World Wide Web. http://www.w3.org/

WCAG: Web Content Accessibility Guidelines (WCAG) are published by the W3C's Web Accessibility Initiative. They give guidance on making content accessible, primarily for disabled users, but also for all user agents, including highly limited devices, such as mobile phones. http://www.w3.org/WAI/

XML: Stands for **Extensible Markup Language**. A general-purpose markup language used to encode structured information. A document written in XML contains content (text) and structure (elements or "tags" which describe the content's purpose), but does not describe how the content should be displayed. This separation of content and presentation allows an XML document to be easily converted between different presentations such as a webpage, print document, braille or synthetic speech.

Appendix 3: Round Table markup

Conventions for Round Table markup were originally described in **Guidelines for preparation of text materials on computer disk for people with print disabilities, 1998**. In practice, actual codes used have differed between organisations, with new codes developed as they were needed. The following list is based on the original Guidelines, as well as production manuals by Vision Australia and RNZFB.

If using Round Table codes within an e-text document, always include a list of codes at the beginning or end of the file, or in a separate readme file.

Page numbers: conventions include:
<pp> #
<p #>
<opp> # (for original print page)
<hp #> (for handout pages)

Headings: <h#> where # is the level of heading (up to <h6>)

Emphasis:
<b> for bold
<it> or <i> for italics
 or <u> for underlining
<other> for other types of emphasis e.g. small caps.

Footnotes:
<fn> or <fn#> for footnotes
<mn> for margin notes

Tables:
<table>
<row # col #> where each cell needs separate markup

Superscript and subscript:
<sub>
<super> or <sup>

Images:
<figure>

<caption>
<diagram>, <graph>, <photograph>, <picture>, <map>, <cartoon>

Boxes and indented text:
<box>
<box#>
<indent>

Special print symbols
<sym copyright>
<sym registered trademark>

Producer's note:
<transcriber's note>
<tn>

Line numbers
<l#>

Appendix 4: Tips for working in Word

Viewing hidden characters

Hidden characters include line breaks and spaces. It can be useful to work with hidden characters visible. Click **Show/Hide ¶** on the **Standard** toolbar

Turning off Smart Quotes

On the **Tools** menu, click **AutoCorrect Options**,

- then click the **AutoFormat As You Type** tab, tab down to **Replace as you type**, arrow down to **"Straight quotes" with "smart quotes"** and unselect
- then click the **AutoFormat** tab, tab down to **Replace**, arrow down to **"Straight quotes" with "smart quotes"** and unselect

Word Short Cut Keys

File Shortcuts

- **New (Ctrl+N)** - Opens a New Document
- **Close (Ctrl+W)** - Closes the document
- **Open (Ctrl+O)** - Opens a Saved Document
- **Save (Ctrl+S)** - Saves a Document
- **Cycle Documents (Ctrl+F6)** - Cycles through open documents

Editing Shortcuts

- **Undo (Ctrl+Z)** - Undoes your last action
- **Repeat (Ctrl+Y)** - Repeats your last action
- **Cut (Ctrl+X)** - Cuts selected text/object (so it can be moved)
- **Copy (Ctrl+C)** - Copies text/object to clipboard
- **Paste (Ctrl+V)** - Pastes whatever is on clipboard into document
- **Select All (Ctrl+A)** - Selects everything in the document
- **Find (Ctrl+F)** - Finds a word or phrase
- **Replace (Ctrl+H)** - Use this to replace one word with another word
- **Spellcheck (F7)** - Checks the spelling in a document

Formatting Shortcuts

- **Heading 1 (Ctrl+Alt+1)** - Applies Heading 1 style to text
- **Heading 2 (Ctrl+Alt+2)** - Applies Heading 2 style to text
- **Heading 3 (Ctrl+Alt+3)** - Applies Heading 1 style to text
- **Normal (Ctrl+Shift+N)** - Converts text to Normal style
- **Bold (Ctrl+B)** - Bolds selected text
- **Italics (Ctrl+I)** - Italicises selected text
- **Underline (Ctrl+U)** - Underlines selected text
- **Double Underline (Ctrl+Shift+D)** - Double Underlines selected text
- **Superscript (Ctrl+=)** - Superscripts selected characters
- **Subscript (Ctrl+Shift+=)** - Subscripts selected characters
- **Small Caps (Ctrl+Shift+K)** - Converts text to small capitals

Moving and Selecting

- Start of word (Ctrl+Right Arrow)
- End of word (Ctrl+Left Arrow)
- Start of line (Home)
- End of line (End)
- Start of Document (Ctrl+Home)
- End of Document (Ctrl+End)
- To Select content between your current cursor location and another point, hold down Shift along with the above commands (e.g. to select everything from your current location to the end of the document, use Ctrl+Shift+End)

Find-and-replace tips

The ^p character finds a paragraph mark, or hard line break.

To add an additional line between paragraphs:
Find ^p and replace with ^p^p.

Where a document has retained line breaks within paragraphs, with an additional line break between paragraphs:
1. Find ^p^p and replace with *****
2. Find ^p and replace with a space character
3. Find ***** and replace with ^p (or ^p^p if an additional line break between paragraphs is needed)

Appendix 5: Tips for using Scanners and OCR

The advent of scanners and Optical Character Recognition (OCR) software has enabled the easy transfer of printed text into an electronic format. However, production of a well formatted and correct document will usually require additional proofreading and formatting.

There are a number of OCR software packages available and the following suggested tips are generic in nature and not specific to any particular software.

General tips

- Bound books cannot always be laid perfectly flat for scanning. The curve where the spine occurs can create distortion. If it is feasible to unbind or remove the binding of a book do so or else you may need to use manual pressure to ensure that each page is as flat against the scanner glass as possible.

Formatting

- Literal layout vs linear formatting: When using OCR to create a document for an electronic format, it is important to first have a clear understanding of these Guidelines for Accessible E-text. Always keep in mind that the e-text format is a representation of the document rather than a literal recreation of the actual layout. e.g. text in single column, rather than 2 or 3 columns, no text boxes, etc.
- Reading Order: Not all printed material is linear in its presentation. Ensure that a logical reading order is maintained. Some software will allow you to select blocked areas of a page in the order that OCR is to occur.
- Columns: Where there are columns on the printed page, it is important that the OCR software recognises the separate columns and does not "read across the page."
- Tables: ensure that rows and columns are correct in complex tables. eg where a single cell of a table in print goes over more than one line. Some OCR software will allow the marking of rows and columns of tables to ensure that each cell is correctly identified.

Proofreading

There are a number of common errors which can occur during OCR, not all of which will be picked up by a spell checker. Careful proofreading is still required for an accurate document. Listed below are some common errors which are worth looking for.

- If your OCR has a spell and/or grammar checker, utilise these as well as those in your Word Processor. Different software will pick up different errors.
- Misrepresented letters: These are not always picked up in spell checkers. For some documents, search and replace can be a useful tool.
  - 1 for I (numeral one and capital i)
  - O and 0 (capital o and numeral zero)
  - b and h (eg he instead of be and vice versa) - grammar checkers will sometimes identify these
  - m for rn (eg modem instead of modern)
- Punctuation:
  - Spacing: eg additional spaces before periods, question marks, etc
  - Quotation marks: Ensure that open and close quotes are balanced i.e. all open quotes have corresponding close quotes. Open quotes can often be missed at the beginning of paragraphs by the OCR software.

Foreign languages

Make sure correct language settings are used so that accents on characters are recognised correctly.