PDF Accessibility Report

Authored by
Kim Andersen, Helmut Hoss & Catherine Bridge

1st ed. July 2020

www.homemods.info

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification Information Clearinghouse, UNSW Sydney

Abstract

This report aims to highlight the accessibility of the documents found on four different organisations' websites. The purpose of this is to highlight the accessibility issues found in PDFs hosted on the internet and what these issues mean for users with disabilities. This report is useful for both document creators and managers who wish to better understand the components of an accessible PDF and how to assess their own document creation processes.

The analysis in this report was conducted using the freely available PDF Accessibility Checker. Throughout the analysis it was clear that many PDFs had fundamental flaws with respect to accessibility, even though visually they looked fine. Some issues can be traced back to the PDF generator used and the document creation process, while others are a result of incorrect document set up. Of the documents assessed 36% were deemed to be wholly inaccessible due to missing internal data in the PDFs. While some organisations' websites performed better than others, all have significant room for improvement in their provision of accessible PDFs.

Keywords

PDF Accessibility; Document Creation; Standards; Microsoft Word; Adobe

Contribution of Authors

This is the first edition of the Occasional Paper: PDF Accessibility Report for the Home Modification Information Clearinghouse, UNSW Sydney.

Kim Andersen undertook the audit and wrote the report.

Helmut Hoss undertook the audit and aided in writing the report, including checking for technical accuracy.

Catherine Bridge supervised the project and edited the report.

Acknowledgements

HMinfo have a policy of undertaking a review process prior to the publication of research documents. The reviews are performed by Specialist Review Panels in accordance with the HMinfo Specialist Review Panel: Terms of Reference, available on the HMinfo website: http://www.homemods.info.

The following Specialist Review Panel members provided their expertise and feedback for this document:

Name	Organisation	Panel
Jackie Leach Scully	Director, Disability Innovation Institute	Guest Reviewer

Liability Statement

The Home Modification Information Clearinghouse team gives no warranty that the information or data supplied contain no errors. However, all care and diligence has been used in processing, analysing and extracting the information. The Home Modification Information Clearinghouse will not be liable for any loss or damage suffered upon the direct or indirect use of the information supplied in this document.

Reproduction of Material

This material has been published by the Home Modification Information Clearinghouse within the faculty of Built Environment, UNSW Sydney (University of New South Wales).

Any table or material published in this Occasional Paper may be reproduced and published without further license, provided that due acknowledgement is made of this source. The preferred acknowledgment style is:

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification
Information Clearinghouse, UNSW Sydney

Contents

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification
Information Clearinghouse, UNSW Sydney

Figures

Tables

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification Information Clearinghouse, UNSW Sydney

1.   Introduction

For people with disabilities, whether a place or thing is accessible may be a daily consideration. In the ever-changing and expanding digital world this is increasingly important. A large number of services and documentation are now only accessible online meaning that content must be available to all users, no matter what assistive devices they may be using.

With the rise of the Portable Document Format (PDF) as a means to share documents among users with potentially incompatible software it has become an unofficial standard for fixed-format documents. For documents being read on a computer, an appropriately set up PDF is accessible when combined with the most popular screen readers. By contrast, the Human Rights Commission takes the position that PDFs are not accessible on mobile devices due to the limitations of screen readers available for those devices[1]. PDF accessibility is governed by an international standard[2] that explains the proper use of the PDF specification to enhance accessibility.

While in recent years there has been an increasing focus on web accessibility there is comparatively less attention paid to the accessibility of PDFs. In legal terms, under the Disability

---

[1] https://www.humanrights.gov.au/our-work/disability-rights/world-wide-web-access-disability-discrimination-act-advisory-notes-ver#pdf

[2] ISO 14289-1:2014: Document management applications – electronic document file format enhancement for accessibility – Part 1: Use of ISO 32000-1 (PDF/UA-1)

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification Information Clearinghouse, UNSW Sydney

Discrimination Act (1992)[3] content must be accessible to all users and not discriminate against any particular group, either directly or indirectly. Inaccessible content, including PDFs, is considered to be discriminatory under section 5 of the Disability Discrimination Act (1992) by treating people with disabilities less favourably than those without. Refusal to make a reasonable adjustment to remedy the issues is considered discrimination. In practice, if a website or document is deemed to be inaccessible a complaint may be filed against the host organisation. All cited legislation and standards in this report are current at the time of writing and may change and evolve with time.

In an effort to educate users and organisations on best practice and help them better understand the problems surrounding PDF accessibility, this report will detail the results of an accessibility check run on a selection of articles posted online for download in PDF format.

1.1.   PDF Components

When assessing a PDF for accessibility a number of criteria are assessed. These criteria vary in importance to accessibility and with their ability to be directly influenced by the user. The following list will detail each criterion, except embedded files, in order of importance. Embedded files are omitted from this list due to their comparatively low usage.

1.1.1.   PDF Syntax

Incorrect syntax in a PDF would likely result in an unusable PDF file and as such presents the greatest barrier to accessibility, for

---

[3] https://www.legislation.gov.au/Details/C2018C00125

anyone. Due to the severity of the issue however it is not expected that a document that has incorrect syntax would be uploaded as it would be obviously flawed. As such this metric will not be discussed for the remainder of the report.

1.1.2. Structure Elements

The structure elements criterion verifies the correct implementation of structure elements (also known as structure tags) for the item in question (heading, figure, etc.), correct use of heading levels and even table regularity. When these tags are not present it creates what is known as an "untagged PDF". Untagged PDFs are unable to achieve even partial accessibility as it directly affects other elements of the PDF internal structure. These will appear visually identical to a tagged PDF since the visual layout is defined separately. A screen reader may attempt to read an untagged PDF but as it lacks the tags to indicate structure there is no guarantee this will be successful. It is very difficult to repair an untagged PDF and as such this should be avoided.

A potential cause of this is that document generators may provide the option to exclude structure elements when saving PDFs. For example, Microsoft Word 2013/2016 allows the user to save the PDF as "optimised for minimum size (publishing online)", which by default excludes structure elements. It is also possible to remove all structure elements from a PDF after it has been generated using specialised PDF editors.

Part of the structure elements accessibility stems from decisions made in the original documents (e.g. table regularity and heading levels) and are best fixed at that level before it is converted to a PDF. An example of irregular table rows is shown in Figure 1. The rows in this example are irregular since the header row contains 1 merged cell above a row with 5 columns in the same space. The parameters to indicate that these rows occupy the same amount of table space have not been specified. Heading levels should progress

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification Information Clearinghouse, UNSW Sydney

without skipping levels, for example, heading 1 should not be followed by heading 3. A common mistake is to use the heading style that looks as desired regardless of level. This can cause confusion to a user of a screen reader, who may question where the missing heading is located or whether they have missed it themselves while moving through the document.
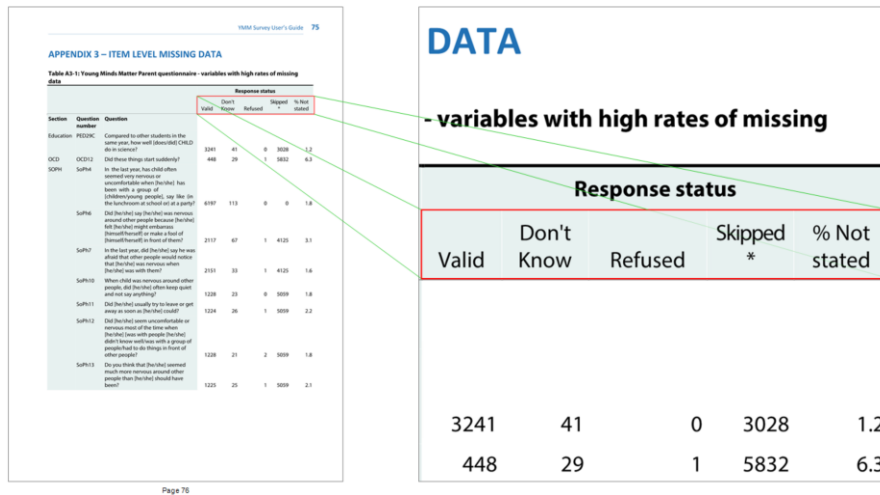

Figure 1: Example of an irregular table row from a sample document.

1.1.3.   Structure Tree

The structure tree provides flow and order to the tags in the document. These must be in a particular order to adequately convey meaning to a user. Another way to consider the structure tree is that it gives context to the document tags. If this context is lacking then a screen reader will be unable to accurately read the PDF. The ease of addressing issues in the structure tree varies with the number of errors for this metric. As fixing this requires removal and replacement of existing elements in the document structure the issue is more of quantity that difficulty of the fix itself.

1.1.4.   Role Mapping

There are 49 standard PDF structure elements used for tagging content, such as figure, note and table. Most standard structure elements are general and thus have a broad scope of application. Roles allow content to be defined more specifically, which is

beneficial in cases where tagging content using a standard structure element would not convey sufficient meaning.

In order for a role to be defined, it must ultimately be mapped to a standard structure element that it intends to expand upon. A role mapping is non-compliant when the role is semantically outside the scope of the standard structure element. Examples of compliant role mappings are a chart mapped to figure or a footnote mapped to note. Examples of non-compliant role mappings are a caption mapped to table or a hyperlink mapped to quote.

Improperly mapped roles can confuse screen readers where one type of content is expected but another is present. As the roles are at a higher level than the actual elements these are simple to fix with minimal effort. There is some ambiguity caused by the creation of custom tags for roles which may result in failures when checking accessibility that are false negatives. Judgment needs to be used in these cases as to whether keeping the role under the custom tags or changing it to something that passes is the best option.

1.1.5.   Natural Language

The natural language metric refers to the language of different text types in the document and the inclusion of the appropriate language identifiers. These include text objects, alternative text, and annotations, among others. Natural language refers to a human language as opposed to a computer language and is useful for the interpretation of text by the display or reading device. For example, a screen reader may load appropriate pronunciation guides for the specified natural language in which the text was written. Incorrect specification of the natural language may confuse the user or alter meaning as the screen reader mispronounces words. This is especially important for words that have different meanings depending on pronunciation (heteronyms).

In general, this is a simple fix with appropriate software however is dependent on the method used to specify natural language. As the natural language can be specified in a parent/child relationship, where the natural language of the parent element is passed to the child, it may be defined at a high level and propagate through or it may be defined paragraph by paragraph. Both are correct but the former has fewer areas to check and fix than the latter.

1.1.6. Fonts

When a PDF is created the fonts used in the parent document should be embedded in the resulting document. This enables the reading device to load the document as it is intended to look. If these fonts are not embedded it can result in characters not loading correctly and random symbols appearing. In some cases, it may default to the system font which may be difficult to read.

It is difficult to fix this post-PDF creation as the missing font needs to be available to place in the document. In general, the PDF generator used is relied upon to embed the fonts correctly and a better generator being used is the best option.

1.1.7. Content

The content metric refers to a number of different tag types directly related to content; the larger the document the more tags will be present. This criterion fails or passes based on whether objects are tagged (e.g. paths and text), characters are correctly mapped to Unicode[4], among other things. Success or failure on this metric is

---

[4] Unicode is a computing standard for the consistent usage and display of text expressed in most of the world's writing systems. It contains characters, symbols and even emojis.

strongly linked to correct implementation of structure elements, hence the relatively low ranking in this list. Non-compliance increases user confusion by making navigation more difficult while using a screen reader. The ease of addressing this issue varies with the number and severity of the specific issues hence it is challenging to determine the exact difficulty of fixing any issues that arise.

1.1.8. Alternative Descriptions

When including graphical media in a document they should be accompanied by alternative description, typically referred to as alternative text. The inclusion of these descriptors allows a wider range of users to access and utilise the document and avoids the exclusion of a subset of the community. Descriptions should be short and meaningful to the reader as well as contributing content to the document, while avoiding repetition. For example, an image that is described by a caption with an alternative description matching that caption would read both to the user. This would be frustrating and break the flow of the document. Alternatively, the description may focus on elements of the picture not described in the caption. As images also includes page elements, such as lines and coloured header blocks, it is not feasible to provide alternative descriptions to these elements. Depending on the version of Microsoft Word or document creator used, these can be marked as decorative to allow a screen reader to ignore them. To pass this criterion the description needs to be defined or set to decorative; it is not appropriate to leave the description blank as this may read out the file name in lieu of a description.

Correction of this issue is relatively simple, requiring the user to enter the missing alternative text as required. A caveat to that is that complex images are difficult to describe in a concise manner and would be better described in the body of the text. If this is not done prior to PDF creation then it may be difficult to include afterwards. Such images will require judgment as to whether enough value is

added with longer alternative text to justify the imposition on the user and whether it can be fixed in the original document.

1.1.9. Document Settings

This criterion is related to the settings in the document itself. This includes if the "DisplayDocTitle" entry is set, if the document is marked as tagged and tab order entry. There are other document settings that may be included depending on the document and PDF creators used, but these are the most common document settings. The DisplayDocTitle is not output in Microsoft Word 2019 (Office365) and would have been manually added for documents made in this program, if present. The tab order entry is important for navigation of content and must reflect the logical order of the PDF content as would be apparent to sighted users. This tab order is determined by tab order properties configured per page. An example of where this would be useful is tables with multiple columns; the flow of the table may be visually clear as being read column by column, but a screen reader may read the data row by row if the underlying page is not appropriately configured.

Non-compliance with this measure is of relatively low impact and in general the PDF will work for those using screen readers. The tab order is easily fixed at the PDF level and specifying the DisplayDocTitle is a straightforward process in remediation software.

1.1.10. Metadata

The final metric has three components that are checked; whether the UA identifier is present, the XMP metadata is included, and the title is included in the XMP. The UA identifier is used to indicate that a PDF is accessible and compliant with the ISO 14289-1:2014 standard mentioned previously. XMP refers to the Extensible Metadata Platform and its purpose is to allow for the creation, processing and interchange of document metadata for digital documents and data sets. It is used for tracking metadata throughout the document creation process, allowing each software package to add its own

metadata. The format of XMP is governed by ISO 16684-1:2019. XMP
metadata is not required for defining critical document metadata and
absence of XMP metadata in documents will not limit screen readers
from correctly interpreting them, therefore missing XMP metadata
has minimal impact on most users. However, XMP metadata is
technically required to achieve full PDF/UA compliance due to the
UA identifier being in XMP format. It is very difficult to add XMP
metadata if it is not included, requiring a manual rebuild or
specialised tools. This would also be highly time intensive.

2. Methodology

2.1.　　Accessibility Tool

The accessibility check was completed using PDF Accessibility
Checker 3 (PAC3), downloadable for free from https://www.access-
for-all.ch/en/pdf-lab/pdf-accessibility-checker-pac.html. The tool is
created by the Swiss Access for all Foundation, a non-profit group
dedicated to promoting the use of technology for people with
disability.

Figure 2: PDF Accessibility checker 3 main window showing results from a sample document.

The checker assesses PDFs against 14 metrics related to the PDF specification. The meaning of each metric will be described with the results of the analysis. Results from each metric were collated and normalised to the overall number of results for that metric and that document. For example, a document's PDF syntax results would be normalised to the sum of the passes, warnings and failures for PDF syntax for that document. This normalisation was completed due to the disparate sizes of documents chosen causing dramatic increases in values for some documents versus smaller ones.

A summary report for each document can be created which includes the information from the main screen of PAC3 shown in Figure 2. Detailed results, including a graphical representation of the error in

the PDF, can be accessed in the PAC3 program. These results take the form of an expandable tree and use graphical tags to indicate areas with errors or warnings, as shown in Figure 3. Results reported on the main window and PDF summary report for passes, warnings and failures are doubled, however the correct values are shown in the detailed results. The cause of this duplication is unknown however as the error is systematic it does not affect the overall results. Where results are quoted below they are the correct values from the detailed report.
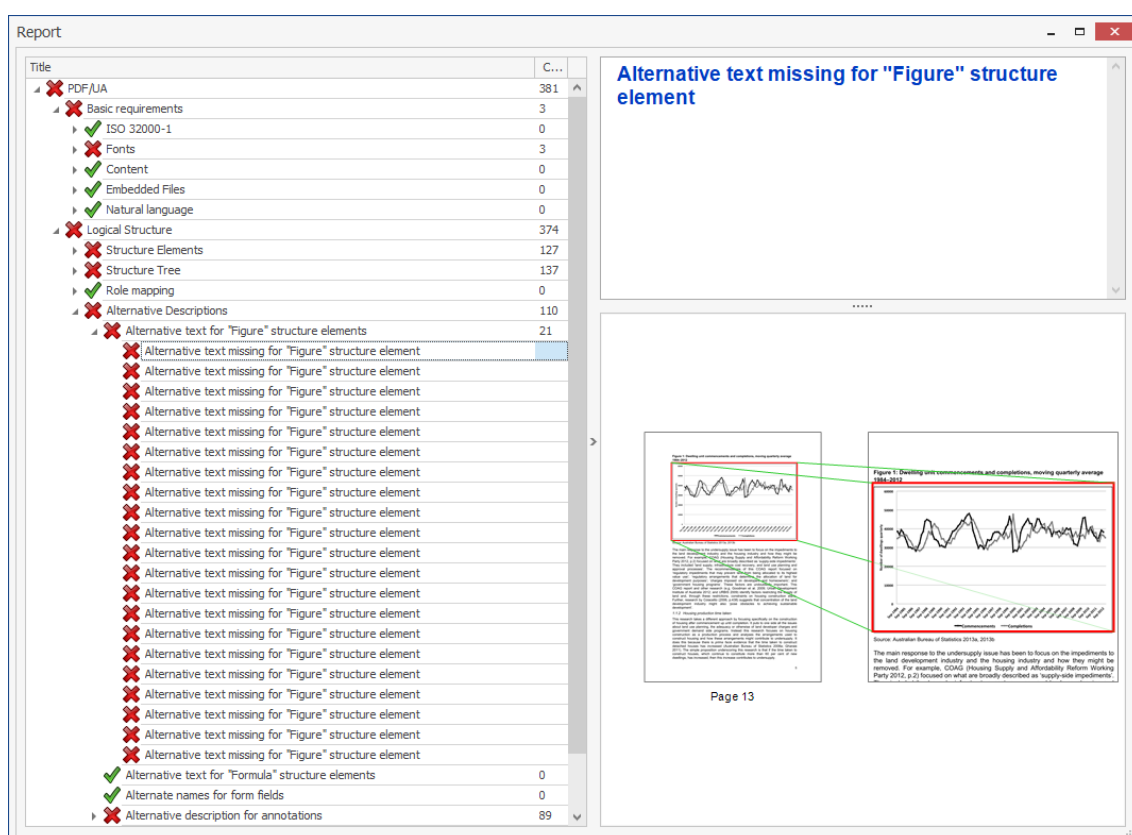

Figure 3: Detailed report example from a sample document.

2.2. PDF Selection

The PDFs assessed in this analysis were selected from four websites. These organisations are repositories for data and grey literature repository and will be referred to as Org1, Org2, Org3 and Org4. From the two websites that function as a document repository (Org1 and Org2) PDFs were selected ranging from 2013 to 2019 with emphasis on those years where possible. These years were selected

due to differences in versions of Microsoft Word available between those times and hence different tools available for users. For the remaining two websites (Org4 and Org3) PDF were selected from what was available. This includes a PDF generated on the Org3 website via a third-party. A total of 28 documents were selected for assessment, as listed in Table 1. It is important to note that not all documents assessed were published by the host organisation and may come from a third-party, as are marked in the table.

Table 1: Documents included in the accessibility analysis. Documents marked with a * are not published by the host organisation.

#	Org	Year	Subject Area	# of pages
1	Org1*	2019	Digital Communication	10
2	Org1*	2019	Telehealth	11
3	Org1	2019	Digital Literacy	4
4	Org1	2019	Healthcare Policy	17
5	Org1*	2013	Social Media	18
6	Org1	2013	Car Manufacturing	3
7	Org1	2013	Australian Elections	11
8	Org1*	2012	Water Management	28
9	Org2	2019	Bathroom Modification	4
10	Org2	2019	Home Modification	45
11	Org2	2018	Lighting	62
12	Org2	2016	Colour	49
13	Org2	2016	Water Management	30
14	Org2	2015	Fire Safety	53
15	Org2	2014	Landscape Modification	2
16	Org2	2014	Home Modification	5
17	Org2	2014	Landscape Modification	3
18	Org2	2013	Liveable Cities	58

#	Org	Year	Subject Area	# of pages
19	Org3*	2020	Documentation	4
20	Org3	2019	Documentation	11
21	Org3	2019	Water Management	2
22	Org3*	2018	Documentation	12
23	Org4*	2019	Documentation	30
24	Org4	2018	Documentation	2
25	Org4*	2016	Survey	24
26	Org4	2015	Documentation	3
27	Org4*	2015	Documentation	111
28	Org4	2013	Documentation	11

3.Results and Analysis

3.1. Interpreting the Graphs

The results for each metric assessed will be presented as stacked column graphs. Due to the large variation in numbers of tags/fonts/etc. between the documents these results are normalised to the total number of results for that metric and that document. The graph shows the passes, warnings and failures as a percentage of total results. For example, document #3 had 156 passes and 4 failures resulting in a graph showing 97.5% pass and 2.5% fail for that document and metric (Figure 5). Passes are indicated in green, warnings in blue (patterned) and failures in yellow. The document number and the organisation who hosted it are listed on the bottom axes and documents are ordered from newest to eldest within organisation.

3.2. Overall Performance

Due to the disparate nature of the documents included in the analysis, differing by host organisation, authors, content type and creation method, among other things, it is difficult to directly

compare performance. That being said, an average score for each organisation on each metric was generated to serve as an indication of the accessibility of their content overall, as shown in Table 2 and Figure 4. The bubble size corresponds to the percent failure of that measure, separated by organisation, with the worst result for each metric shown on the appropriate bubble.

Table 2: Organisation level failure rates for each metric

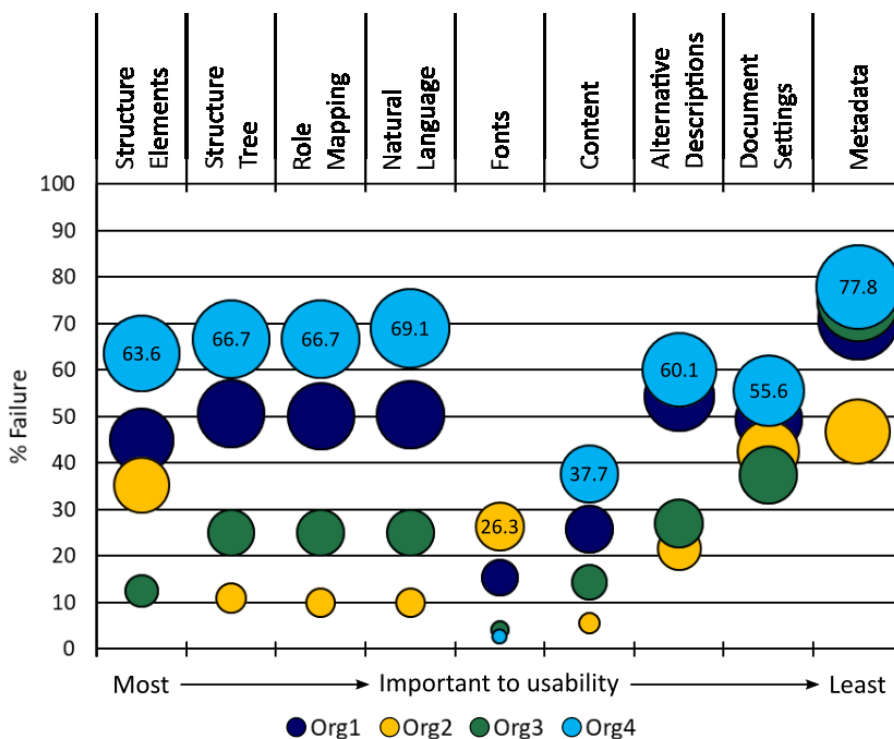| Organisation	Structure Elements	Structure Tree	Role Mapping	Natural Language	Fonts	Content	Alternative Descriptions	Document Settings	Metadata
Org1	44.8	50.7	50.0	50.3	15.2	25.7	54.3	49.3	70.8
Org2	35.2	10.9	10.0	10.0	26.3	5.5	21.6	42.5	46.7
Org3	12.5	25.0	25.0	25.0	4.2	14.4	26.9	37.5	75.0
Org4	63.6	66.7	66.7	69.1	2.8	37.7	60.1	55.6	77.8

Figure 4: Overall results for each organisation ranked from most important to least important.

The results show that Org2 performed the best on nearly all measures and Org4 the worst. It is important to note that 50% of the Org4 documents were from a third-party and hence they would have limited control on the document's accessibility, while Org2 authored all their documents. Furthermore, Org2 has had accessibility measures built into its document creation protocols and as such is expected to perform well. Removing the documents from third-parties had mixed results; Org4 performed worse, Org3 better and Org1 comparably to the results below. Notably, on the first four metrics Org3 had no errors, while for the same metrics Org4 had a 100% failure rate. This highlights the need for each organisation to review their protocols for document creation and the hosting of third-party documents on their website to create the best outcome for all users.

3.3. PDF Tags

This check is a binary yes or no to whether a title and the language of the document are specified. Title tags were included on 36% (N = 10) of documents. The inclusion of this tag varied across the organisations assessed and was not affected by whether the documents were authored by the host organisation. In Microsoft Word, the title can be set in the document properties. When saving to PDF, the PDF/UA Compliant box must be checked to output the title to the created PDF.

It should be noted that it does not assess whether these are appropriate or correct. For example, document #9 by Org2, Consumer Factsheet: Bidet douche seats, flexible hoses and bidets, lists the title as "Consumer Factsheet:" which is only a part of the title and in this case refers to the type of document. The same occurs on documents #15 and #16, both from Org2 suggesting the title beginning is included in the template but has not been updated on the document itself. Another example is the Org1 document #3 and Org2 document #11 which have empty titles. In total, eight documents passed this measure but did not include an appropriate

title on further investigation. This makes the effective pass rate for the title tag 7% (N = 2).

The language tags were specified on 54% (N = 15) of documents with nearly full compliance on Org2 documents and half of Org3 documents. As with the title tag, the inclusion of the language tag was not affected by whether the author was the host organisation. Language tags should transfer when creating a PDF from a Microsoft Word document.

3.4.   Structure Elements

One quarter of documents (N = 7, #'s 5, 6, 7, 18, 23, 24 and 26) indicated the presence of no structure elements, pass or otherwise and as such are untagged. This is a severe breach of accessibility guidelines and means that these documents cannot achieve any modicum of accessibility. Org3 was the only document to host no untagged PDFs in the sample selected.
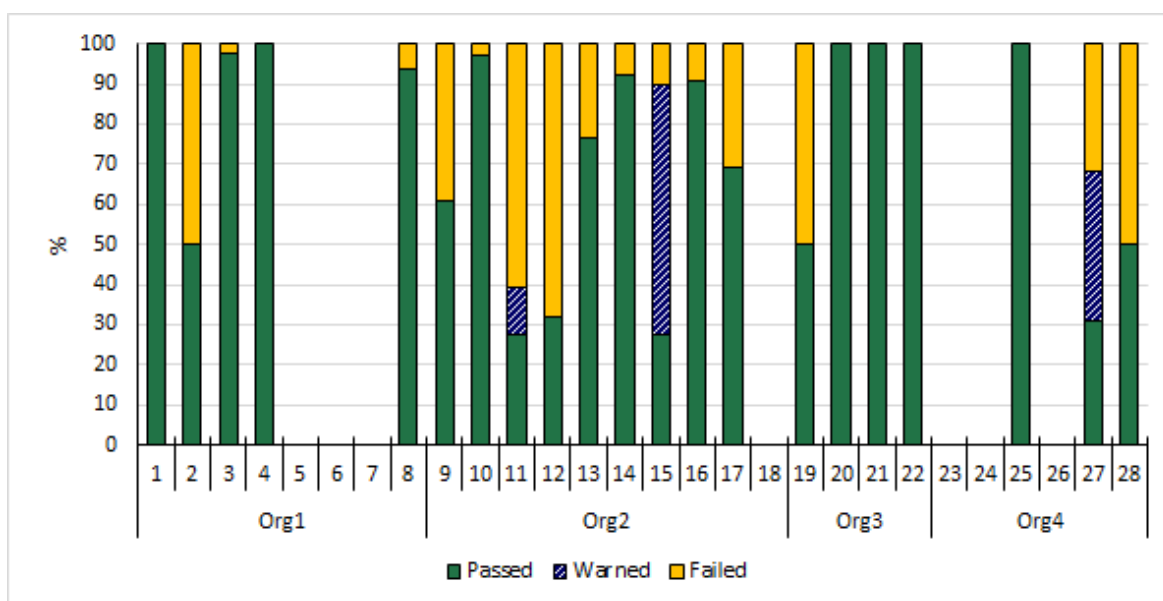

Figure 5: Structure element metric output.

Failed criteria tended to be related to headings skipping levels, correct use of link annotations, missing figure bounding boxes and table header cell assignments. Warnings were all related to irregular table rows as shown in Figure 1. These errors generate a warning as

the software cannot conclusively determine compliance on its own. As a warning is generated for each individual row that may be non-compliant, usually every row in the offending table, the large number of warnings is directly related to the size of the tables in question. Only one document showing warnings (#15) did not also fail to correctly define the table headers.

3.5.   Structure Tree

In the absence of structure elements it is not possible to define a structure tree. Additionally, document #'s 2, 19 and 28 do not include a structure tree representing 36% (N = 10) of the documents. Org1 and Org4 performed the worst on this measure and results were split between those documents authored by the host organisation and not, however all organisations had at least one document with no tree. This could be a result of a third-party PDF generator that is not appropriately developed or tested being used to generate the file. These generators may also be the cause of the lack of a language tag. Each document without a structure tree also does not specify the language tag, accounting for all except two of those documents.

Otherwise, compliance with the standard on this measure was good, with minimal or no errors or warnings for the assessable PDFs. The most common warnings related to inappropriate use of sect (section) and figure elements, while others related to span, note and P (paragraph) elements.
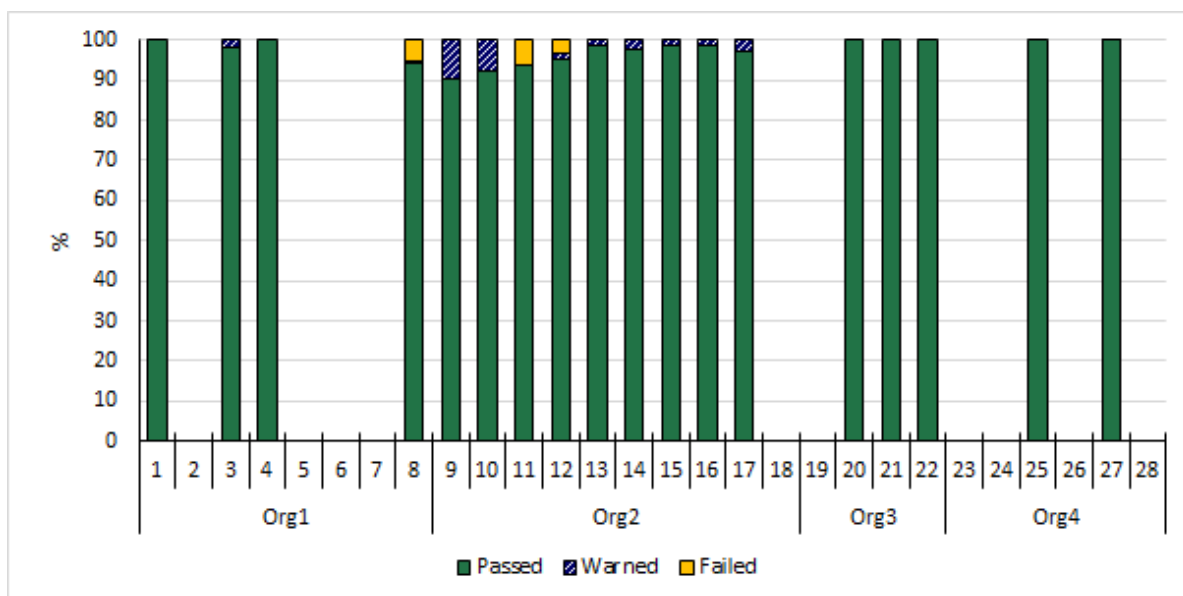
Figure 6: Structure tree metric output.

3.6.   Role Mapping

Role mapping compliance cannot be determined for PDFs missing
structure elements as there are no elements present for roles to map
to. Similarly, if no structure tree is present the roles cannot be
adequately determined. Hence, the documents with no results for
this measure also correspond to the documents that returned no
results for the structure tree. All documents that could be assessed
for this measure returned a 100% pass with the exception of
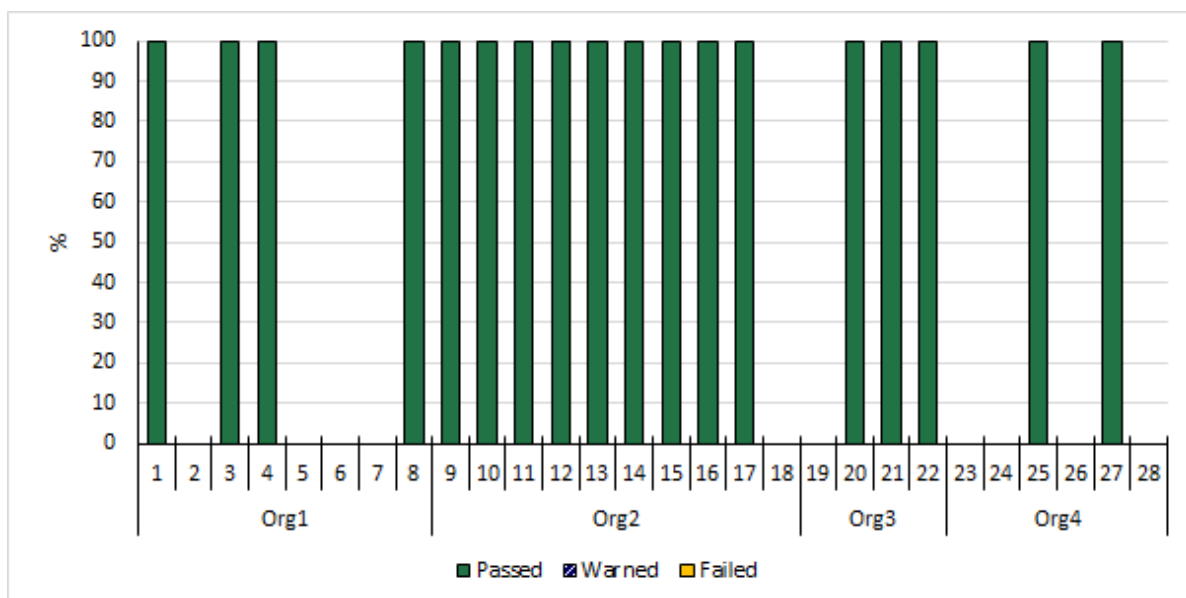document #12 which returned one failure out of 1626 roles.

Figure 7: Role mapping metric output.

3.7.   Natural language

Compliance with this measure mostly corresponds to the specification of language for the document earlier discussed, as expected. When language is not provided at the file level, typically all text objects and outline items (bookmarks) will be non-compliant. However, it is possible for some structure elements to be compliant if their parent element has been tagged with a language. For example, documents #1 and #25 are the only results that are not a strict pass/fail on this measure and document #21 passes this metric (100%) however does not have a defined language tag. The file language tag is not defined but the document (root structure element) has been assigned a language, meaning the root/parent element passes the language to the child structure elements. Documents without a structure tree will fail to comply as they cannot inherit the language from the parent element as it does not exist, they can only inherit from the file tag, which in this case is undefined.
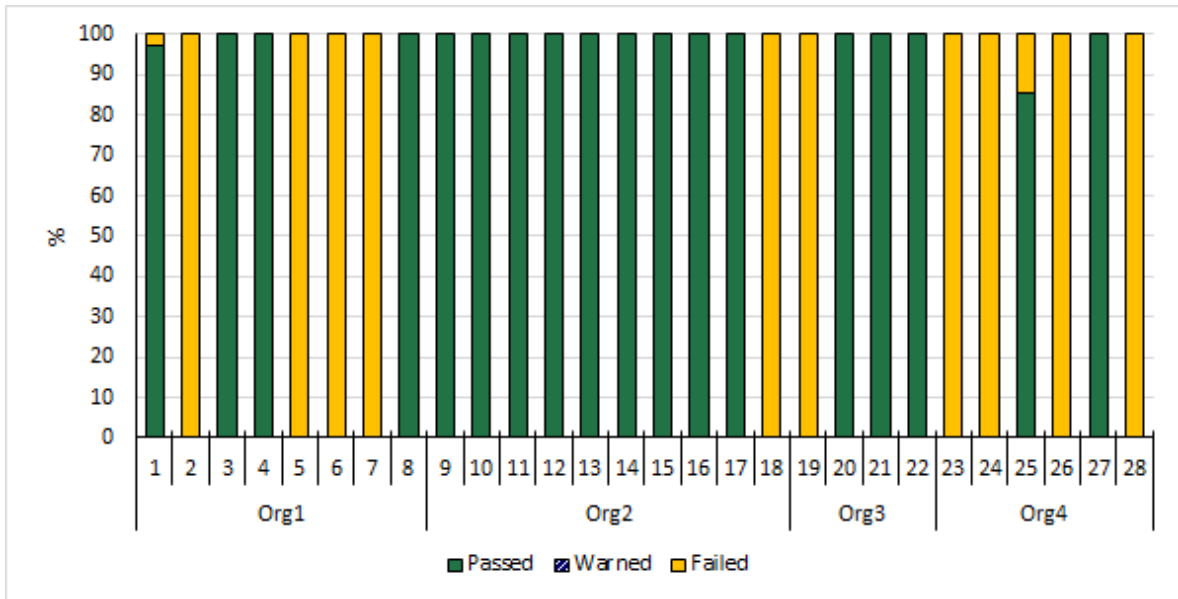
Figure 8: Natural language metric output.

3.8. Fonts and Embedded Files

In general, fonts were embedded across the documents. The number of fonts included in the documents varied from 3 fonts (#15, 100% non-compliant) to 63 fonts (#5, 100% compliant). Only one Org1 document and those from Org2 with errors on this metric were authored by the host organisation. With the difficulties in addressing this metric after document creation it is reasonable to assume that these third-party errors would be unfixable by the host organisation.

Only 1 document (#17) embedded files (N = 4) in the PDF documents. These files could be other PDF files, images and even multimedia. These were fully compliant with the specifications.
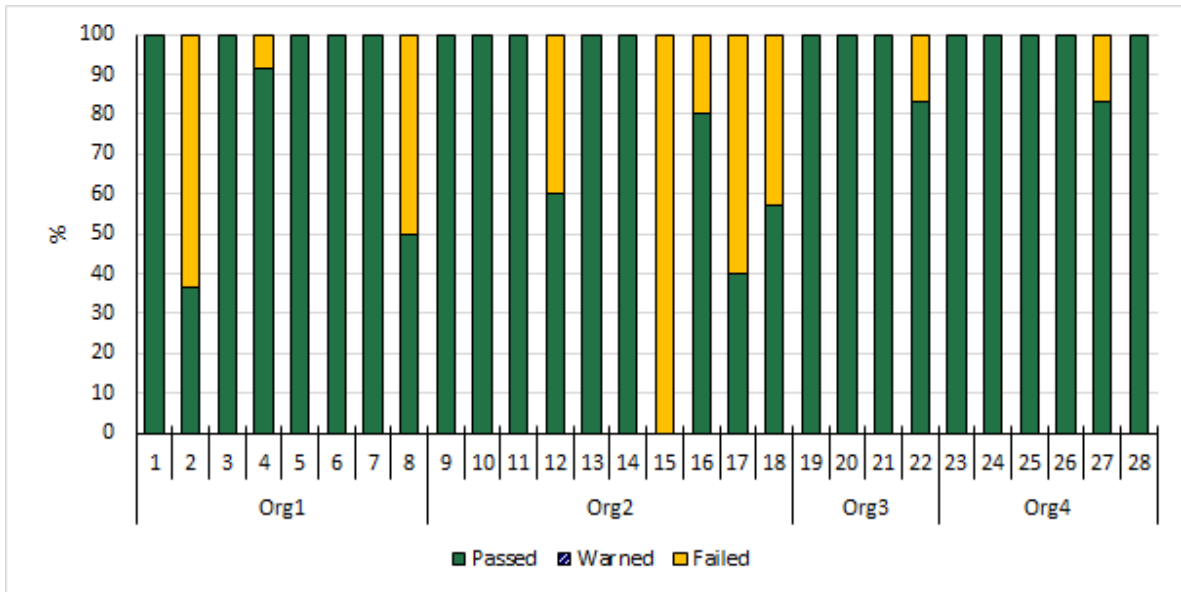
Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification
Information Clearinghouse, UNSW Sydney

Figure 9: Font metric results.

3.9. Content

Most commonly, errors occurred with correctly tagging text and paths in the document. Additionally, some symbols did not convert correctly to Unicode characters, which can cause issues on the display device, including screen readers. After removing the documents without a structure tree from the analysis, most documents were fully compliant or nearly so on this measure. The average pass rate was 98% among the remaining documents.
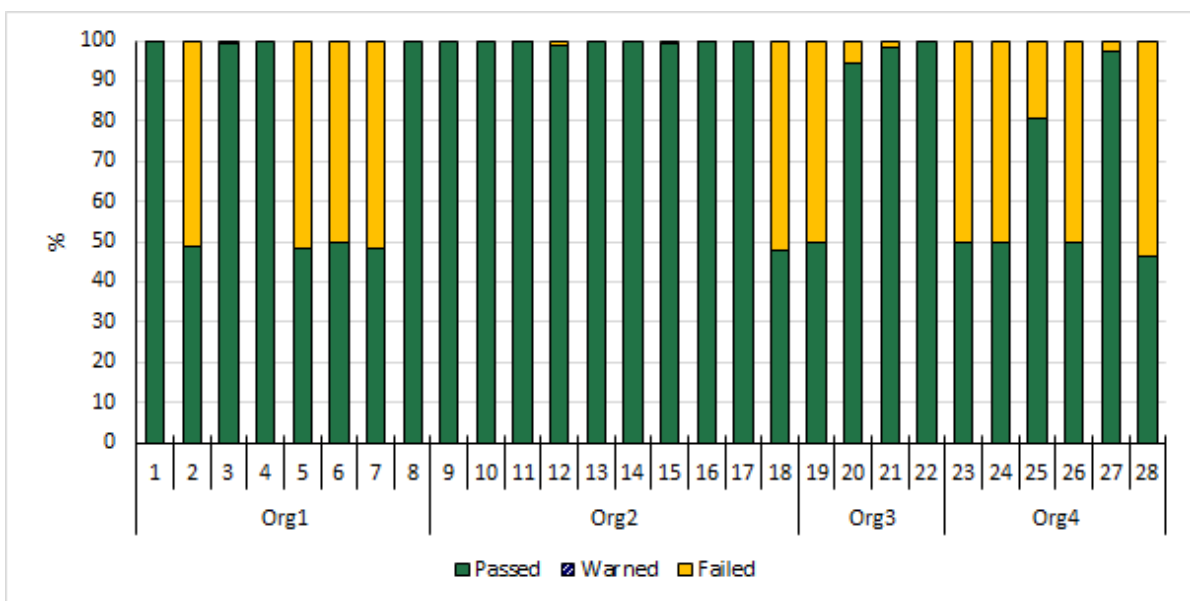

Figure 10: Content metric results.

3.10. Alternative Descriptions

The documents showing no results here correspond to those missing structure elements, i.e. untagged PDFs, because there is no element that can be described by alternative text and as such compliance cannot be assessed. Those documents lacking structure trees can however still be assessed as the element still exists in those documents. Nearly 20% of the assessable documents included alternative descriptions on 50% or less of the applicable items in those documents. No document returned 100% compliance for this measure, despite appearances in Figure 11. Not all alternative descriptions are placed on images, they are also placed on annotations, form fields and formula. Annotations are often links to external sources such as websites, though they could also be to other documents and were the most common error that wasn't related to an image. Form fields and formula elements relate to editable PDFs, none of which were included in the sample.
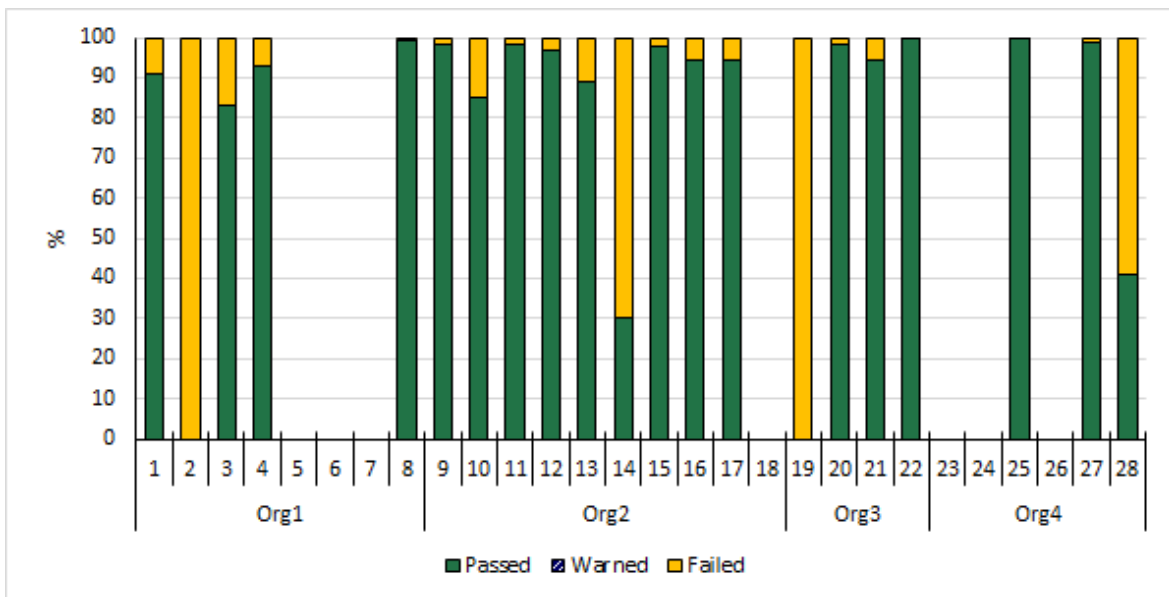

Figure 11: Alternative descriptions metric output.

3.11. Document settings

Compliance with this measure varied between documents, with some doing well and others very poorly. This was inconsistent within organisation and shows no particular trend with author organisation.

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification
Information Clearinghouse, UNSW Sydney

The most common error across documents was the lack of
DisplayDocTitle settings containing valid values (N = 22).
Interestingly, the two documents that had 100% compliance on this
measure (#4 and 22) and hence do specify a valid DisplayDocTitle
value, did not include a title tag for the document itself. Correctly,
the documents without a structure tree were not marked as tagged
and no other documents lacked this identifier. The tab order entry
was not compliant for 32% of documents (N = 9), with 44% (N = 4) of
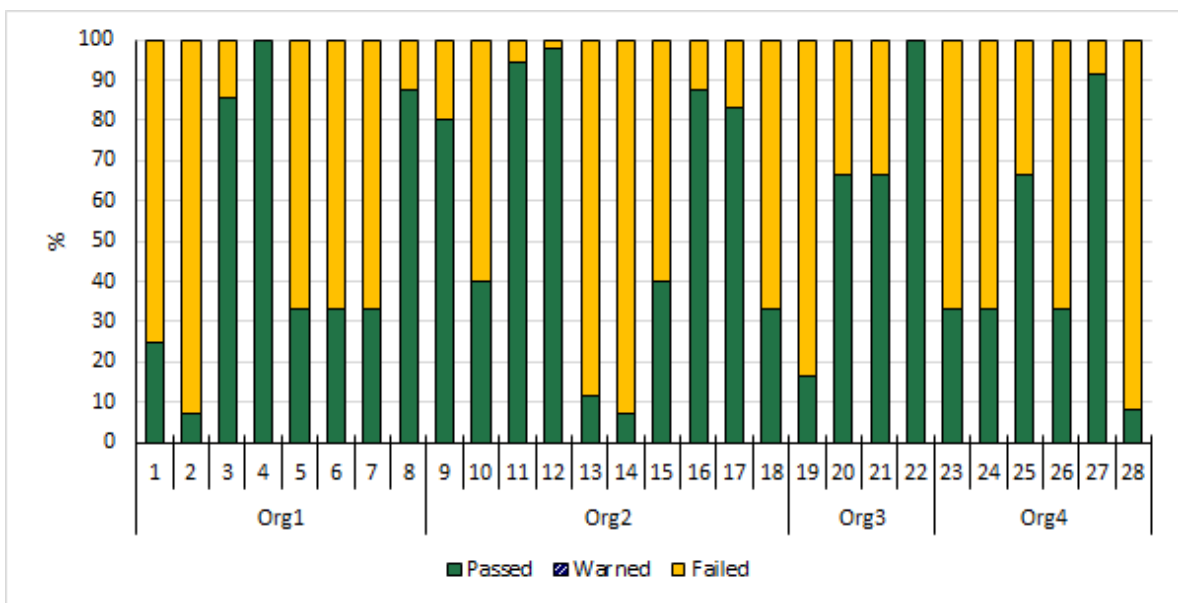these returning errors on every page in the document.

Figure 12: Document settings metric output.

3.12. Metadata

No document included the PDF/UA identifier to indicate an
accessible PDF. Given the results discussed previously this is
correct as no PDF assessed was compliant with the standard for
accessible PDFs. Had the identifier been present it would have been
in error and would have misled users. This means that while
compliance on this metric is overall very poor, in this one instance a
negative result is not an inherently bad result.

In comparison, the XMP metadata was included in the majority of
documents (N = 20). As mentioned in the description for this metric,

the exclusion of XMP metadata will not limit screen readers from correctly interpreting documents.

The XMP title is only included on documents with the XMP metadata defined, as expected. Comparing the documents with no XMP title (N = 10) with those with no title tag (N = 18) or poor titles (N = 8) showed that no document lacking the XMP title had a title tag specified. The documents with poor titles all had XMP titles defined, however given the title quality and the inability to check the XMP title with the software provided, it can only be assumed the XMP title is a duplication of the tag and therefore poor as well.
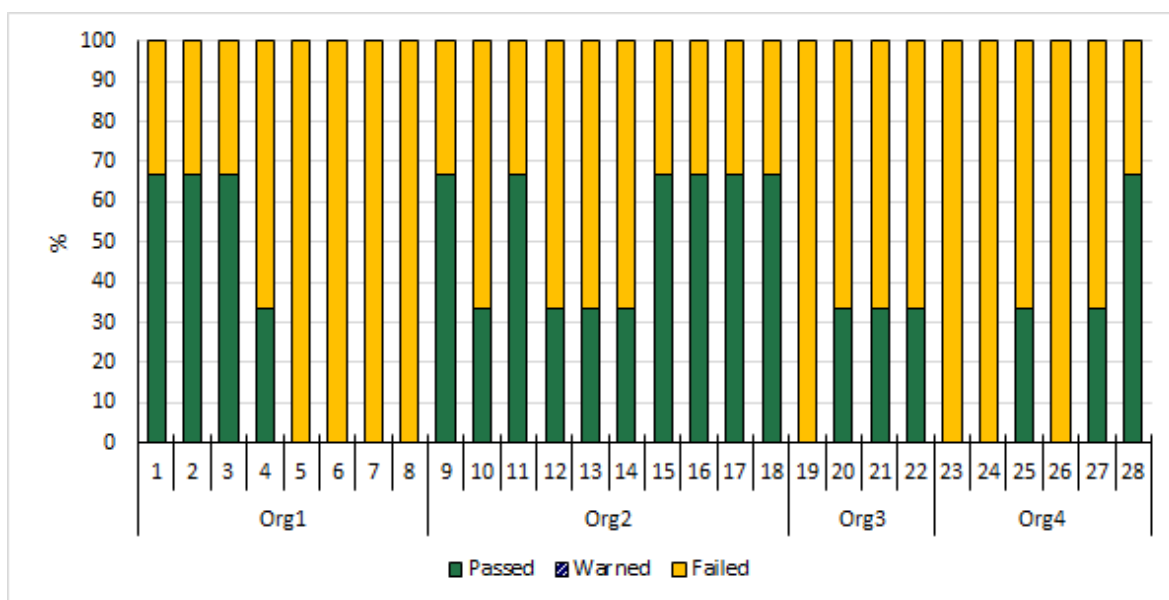

Figure 13: Metadata metric output.

3.13. Third-Party Documents

As mentioned previously, not all documents assessed were authored by the host organisation. Examining those documents separately shows mixed results on most measures. As with the other documents, a lack of a structure tree and structure elements makes evaluating role mapping and alternative descriptions impossible for those documents. Four Org1 documents selected were journal articles and while all show errors, documents #1 and #8 return better results on most measures than the other documents. Conversely,

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification
Information Clearinghouse, UNSW Sydney

document #5 lacks a structure tree and is untagged making it completely inaccessible to a non-sighted user.

Document #19 was created using an online generator called PrintFriendly located on a page on the Org3 website. This generator fails to specify a title, the language, metadata, the majority of the document settings and is 50% untagged. Some of this may relate back to the webpage itself. For example, if an image on the page lacks an alternative description then the PDF generator cannot include one unless it creates its own. The alternative description error in this case related to annotations and these were not present on the webpage. In general, this generator does create a printable PDF, it does not however create an accessible one.

The extent to which the PDF authored by third-parties need to be accessible on the website is for each to determine. It is however suggested that if the documents are not accessible that they are accompanied by a disclaimer. It is possible to fix some errors to make the PDF usable by assistive technology using specialised software. The benefit of doing this does need to be weighed with the time required, cost of the software and expertise needed. Documents that are untagged would result in a significant time investment to fix, while others such as document #'s 1, 7, 22, 25 and 27 which returned comparatively few errors on most measures might be repaired relatively quickly.
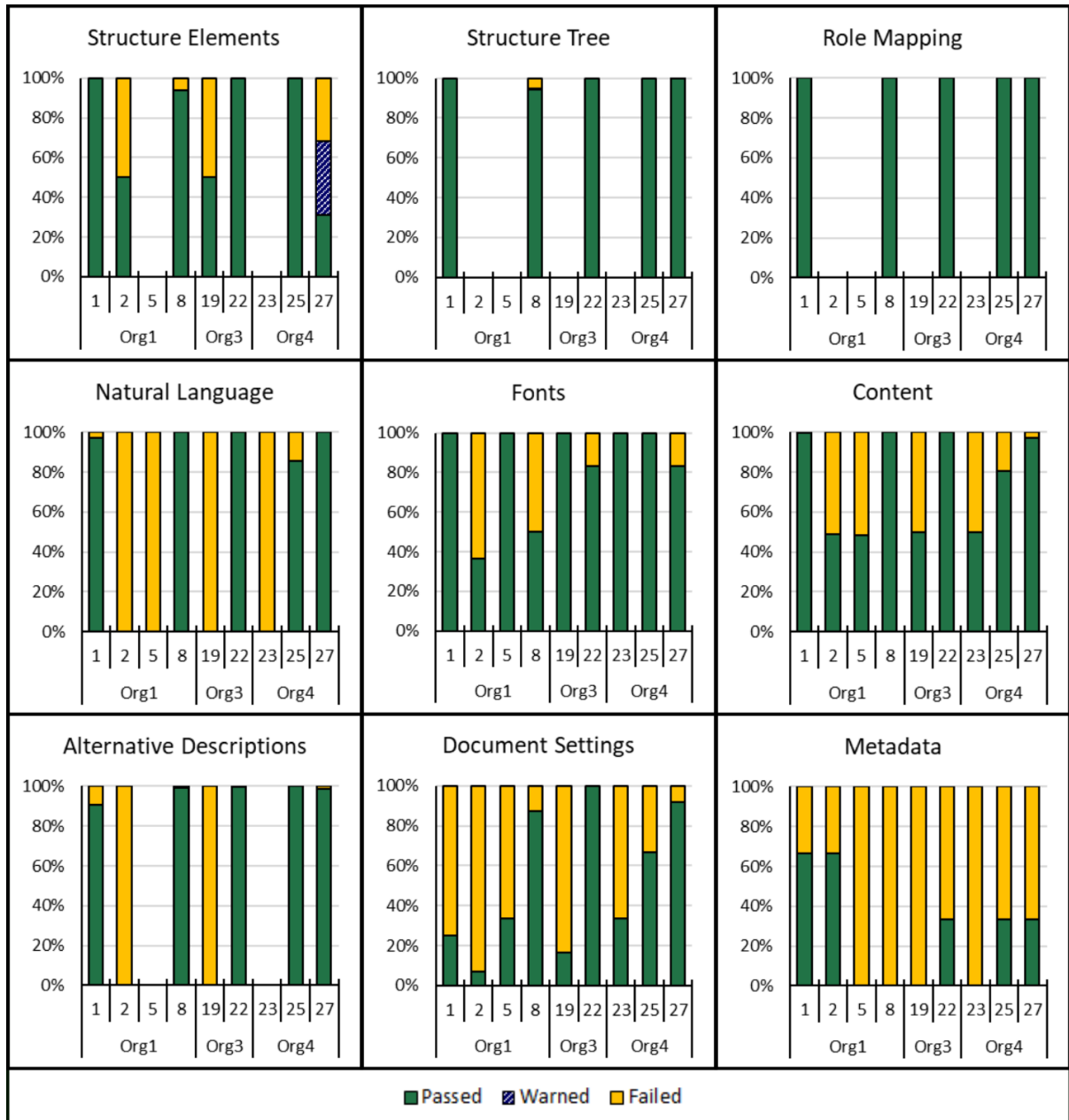
Figure 14: Third-party document results for all measures.

4. Assessing a PDF Generator

In many cases in the documents assessed the errors stemmed from the generator or processes used to create the PDF rather than specific errors on the user's part. While an appropriately set up document will remedy a number of errors it is important to delineate between errors from a user and errors related to the process. A guide for how to correctly create a PDF from a Microsoft Word

document has been developed[5], while a sample document to assess
the PDF generation process is included in Appendix – PDF Test
Document.

Using the example document described, a PDF was generated using
Microsoft Word's Print to PDF function and Adobe PDF's save option.
This creates two PDF that are visually identical but different in size
(Adobe: 396kb, Microsoft: 725kb) and creator. The results vary
significantly considering the common start point, as shown in Figure
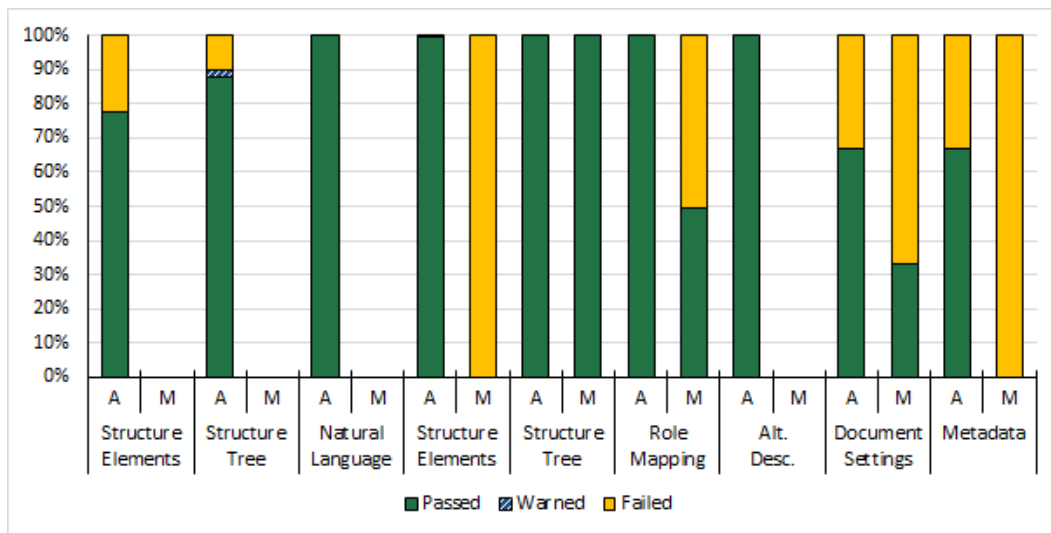15.

Figure 15: Generator comparison results for all metrics.

The Adobe PDF performed better than the Microsoft PDF on all
assessable measures. Errors with the Microsoft PDF were primarily
related to the lack of tags. One reason for this is an additional
option available when saving to an Adobe PDF that allows the PDF

---

[5] Andersen, K., Hoss, H. & Bridge, C. (2020) *Writing Accessible PDFs
Using Word*, UNSW, Sydney, Accessed at
www.homemods.info/resources/hminfo-research-publications/occasional

to be output with tags for accessibility. No such option could be found when printing to PDF with the inbuilt creator. The DisplayDocTitle setting value is not set correctly despite the title being specified in the document settings manually and a title style being used in document.

Errors in the Adobe PDF related to the natural language, structure elements and tree, metadata and document settings. The natural language was not defined for alternative text and structure element and tree errors are all related to the table. Namely, the header rows are not appropriately associated with the cells below them and paragraph structure elements are incorrectly assigned in the table cells. Finally, it lacked a correctly set DisplayDocTitle setting value, even though the title tag was defined.

A similar analysis can be completed for any PDF generator and word processor combination. While not all possible elements for a large document are included in the sample document the basics are assessed. These results can be extrapolated, with care, to consider the adequacy of the process selected. Based on the above results, it is recommended that Adobe generators are used when creating documents in Microsoft Word rather than the inbuilt creator. The caveat to this recommendation is that this analysis was completed using one computer and other settings may play a role in the PDF output that have not been considered here. For best results, it is recommended that a new analysis be conducted, where possible, to confirm findings based on individual processes, settings and available software.

5. Conclusion

Accessibility is vitally important to ensuring that all members of the community can access, contribute and participate in all society. Beyond the social benefits of enabling access for all, it is also a legal requirement under the Disability Discrimination Act (1992). This report has assessed the accessibility of 28 PDF documents

hosted on the four LIEF partners websites: Org1 (8), Org2 (10), Org3 (4) and Org4 (6). Of these the majority were authored by the host organisation, however 9 were from third-parties. No publication assessed was 100% accessible and 7 documents were completely inaccessible due to a lack of structure element tags. The results were poor overall with a large spread of results across each metric and host organisation. The importance of PDF generator choice, in addition to appropriate document set up has been demonstrated throughout the report. Combined with the poor results, it emphasises the point that document creation processes and protocols are essential to producing accessible PDF content. These processes need to be updated with changes to the PDF specifications and software to ensure that over time they remain current and applicable. There is significant work to be done at each organisation to transition to an accessible PDF creation process and develop guidelines or protocols regarding the publishing of third-party documents on their websites. Once completed, however, each website will be able to engage with potentially new groups of users.

6. Appendix – PDF Test Document

The information included below may be copied into a new document and recreated in the word processor of choice. To aid in this, the example document styles and formatting will be listed in order below. Normal and body text may be used interchangeably. All text is left justified and at default sizes. The title should be specified in the document properties if using Microsoft Word.

1. Title
2. Normal
3. Heading 1
4. Normal
5. Image, wrapped in-line with text
6. Caption
7. Page break
8. Heading 1
9. Normal
10. Heading 2
11. Normal
12. Table, header row selected and banded rows

6.1.   Example Document

The Definitive Guide to Australian Crows

Introduction

Foreigners have often asked me about the cultural significance of crows in Australia and I thought that it was about time that I documented my lengthy explanations to avoid repetition.

Indeed, the next time I am queried about crows I shall direct the inquisitive soul to this book, which by the way, I hope you have paid for and not downloaded from one of those notorious pirate websites.

Australian Crow

Authored by Kim Andersen, Helmut Hoss & Catherine Bridge for the Home Modification Information Clearinghouse, UNSW Sydney

A Somber Call

A typical feature of any Australian morning in New South Wales is the loud crowing of seemingly unhappy crows. Do not be mistaken, these rascals are not at all upset and are probably cleverer than your toddler.

Differences

The astute observer would have noticed that crows in other countries sound quite different to Australian crows, while nonetheless maintaining a gloomy tone.

For instance, crows in Japanese cities sound somewhat higher pitched and the length of their crow is shorter.

To better illustrate the wide variety of birds within the same genus as Australian crows, I have included a short table below.

Countries/Regions	Name
United States, southern Canada, northern Mexico	American crow
Eastern Africa	Somali crow
Jamaica	Jamaican crow
Cuba, Haiti, Dominican Republic	Palm crow