



PDF/UA examples, Accessible HTML

FallIMT 2022 PDF/UA and HTML – Accessible

by Ross Moore

July 2023

FallMT-2022: Management Track Assessments Fall 2022

- PDF/UA: a pre-publication version of FallMT-2022: [PDF version](#);
- HTML: a derivation to [Accessible HTML](#).

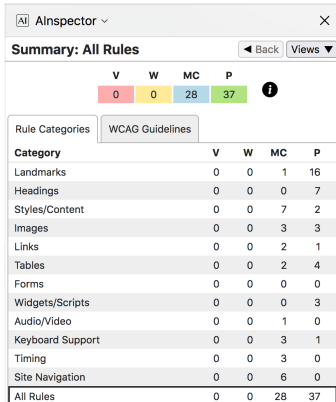
Some features of these PDF and HTML documents will be discussed at the TUG 2023 meeting, in Ross Moore’s presentation video.

The method of production follows these steps.

1. First \LaTeX source is processed to build a Tagged PDF valid for the PDF/UA-1 format.
2. The resulting PDF is derived into an HTML version using [ngPDF](#).
3. The HTML page, and any images, is then downloaded and copied to this location.
4. Some post-processing with sed commands is done on the HTML to remove some incorrect structures, invalid attributes, and the ‘-’ remnants of hyphenation within the PDF.

Accessibility: WCAG 2.1 and ARIA Success Criteria

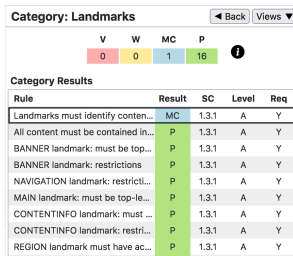
The following images and explanatory notes indicate how WCAG and ARIA recommendations and requirements are handled. We use the ‘AInspector’ software which operates as a plugin to the Firefox browser, to algorithmically check many aspects of these recommendations.



Category	V	W	MC	P
Landmarks	0	0	1	16
Headings	0	0	0	7
Styles/Content	0	0	7	2
Images	0	0	3	3
Links	0	0	2	1
Tables	0	0	2	4
Forms	0	0	0	0
Widgets/Scripts	0	0	0	3
Audio/Video	0	0	1	0
Keyboard Support	0	0	3	1
Timing	0	0	3	0
Site Navigation	0	0	6	0
All Rules	0	0	28	37

The table headings are V: violation, W: warning, MC: manual check, and P: pass.

Note that there are no violations and no warnings, only 28 possible issues requiring manual checking, which will be dealt with below. As well as the 37 rules which are algorithmically verified as Pass, there are 57 further rules/recommendations deemed to be not applicable, making 122 tests in all. For instance, since there are no interactive Form fields in the document, then all 17 checks under the ‘Forms’ category are deemed inapplicable; similarly for 15 under ‘Widgets/Scripts’ and 12 under ‘Audio/Video’ as well as a few in other categories.



Rule	Result	SC	Level	Req
Landmarks must identify conten...	MC	1.3.1	A	Y
All content must be contained in...	P	1.3.1	A	Y
BANNER landmark: must be top...	P	1.3.1	A	Y
BANNER landmark: restricti...	P	1.3.1	A	Y
NAVIGATION landmark: restricti...	P	1.3.1	A	Y
MAIN landmark: must be top-le...	P	1.3.1	A	Y
CONTENTINFO landmark: must ...	P	1.3.1	A	Y
CONTENTINFO landmark: restri...	P	1.3.1	A	Y
REGION landmark must have ac...	P	1.3.1	A	Y

The ARIA concept of ‘Landscape’ has no direct analog in PDF/UA; but corresponds to the highest level nodes in the Tagged PDF ‘structure Tree’. In this document the correspondence is appropriate, as follows; BANNER ↔ Cover page(s), COMPLEMENTARY ↔ \frontmatter, MAIN ↔ \mainmatter, CONTENTINFO ↔ Back cover pages. Furthermore, REGION ↔ sectioning with \section or \subsection, etc. All content is contained in such regions appearing structurally under landscapes.

Category: Headings

V W MC P

0 0 0 7

Category Results

Rule	Result	SC	Level	Req
Headings must be properly nest...	P	1.3.1	A	Y
Headings should have text cont...	P	1.3.1	A	
First landmark heading H2	P	1.3.1	A	
Headings nested in landmarks	P	1.3.1	A	Y
Page should have H1 element	P	2.4.1	A	
H1 should be in MAIN or BANNE...	P	2.4.6	AA	
Sibling headings should be unliq...	P	2.4.6	AA	

Category: Styles/Content

V W MC P

0 0 7 2

Category Results

Rule	Result	SC	Level	Req
Use semantic markup for lists	MC	1.3.1	A	Y
Reading order: CSS positioning	MC	1.3.2	A	Y
Not only shape, size and location	MC	1.3.3	A	Y
Use of color	MC	1.4.1	A	Y
Resize text content	MC	1.4.4	AA	Y
Provide list labels when appropr...	MC	2.4.6	AA	
Identify language changes	MC	3.1.2	AA	Y
Page must use language code	P	3.1.1	A	Y
Text must exceed CCR threshold	P	1.4.3	AA	Y

Category: Images

V W MC P

0 0 3 3

Category Results

Rule	Result	SC	Level	Req
All text must summarize purpose	MC	1.1.1	A	Y
Long description for complex im...	MC	1.1.1	A	Y
Use MathML for mathematical e...	MC	1.1.1	A	Y
Images must have alt text	P	1.1.1	A	Y
All text must not include filename	P	1.1.1	A	Y
Alt text no more than 100 chara...	P	1.1.1	A	
Verify image is decorative	N/A	1.1.1	A	Y

Category: Links

V W MC P

0 0 2 1

Category Results

Rule	Result	SC	Level	Req
Link text must describe the link ...	MC	2.4.4	A	Y
Target focus must be in content...	MC	3.2.1	A	Y
Link text should be unique	P	2.4.4	A	

Category: Tables

V W MC P

0 0 2 4

Category Results

Rule	Result	SC	Level	Req
Data tables should have descrip...	MC	1.3.1	A	
Data cells must use HEADERS a...	MC	1.3.1	A	Y
Data tables must have unique n...	P	1.3.1	A	Y
Identify table markup as data or...	P	1.3.1	A	Y
Header cells must be TH eleme...	P	1.3.1	A	Y
Data tables should have name	P	2.4.6	AA	
Data cells must have row/column...	N/A	1.3.1	A	Y
Name must be different than de...	N/A	1.3.1	A	Y
Layout tables must have meanin...	N/A	1.3.2	A	Y

Category: Forms

V W MC P

0 0 0 0

Category Results

Rule	Result	SC	Level	Req
SELECT must not change context	N/A	3.2.2	A	Y
Forms submitted using buttons	N/A	3.2.2	A	Y
Information on invalid values	N/A	3.3.1	A	Y
Form controls must have labels	N/A	3.3.2	A	Y
Image button must have alt. con...	N/A	3.3.2	A	Y
Radio buttons must have groupi...	N/A	3.3.2	A	Y
BUTTONs must have content	N/A	3.3.2	A	Y
LABEL must reference control	N/A	3.3.2	A	Y
LABEL must have content	N/A	3.3.2	A	Y

The HTML site is just a single (very long) page, with the PDF /Title occurring as H1 on the front cover, which is the ‘BANNER’ landmark. All top-level (H2 or H3) headings in REGION sections have unique textual titles. Lower-level headings may include the fish name, to ensure uniqueness; e.g., ‘Terms of Reference: Pollock’. Even where not, such as ‘References’, it is unique to each fish-stock section; that is, branch of the structure tree.

Most lists are tagged as such; apart from some, such as authors or panel-lists. These have semantic markup conveyed through user-defined environments and commands. Changes in text sizing occur with headings or captions, all of which have meaningful semantic markup. Document divisions without header text are marked using an aria-label attribute; e.g., ‘Coverpage’, ‘Front-matter’, ‘Series Information’, etc. Generally reading order follows the order of appearance in the L^AT_EX source; when perhaps different, then CSS rules are defined and associated through structure tag attributes. The only language changes are for latin names of fish species, being clearly tagged as anchor text for links to a Glossary page.

All the images, whether graphical plots, photographs or line-art, have alternative text. Some have appropriate long descriptions provided as structural attribute string-values. There is no mathematical content presented as an image.

Each internal hyperlink has an aria-label attribute which is unique to the link’s target. This has been constructed from the string keys used within the L^AT_EX processing of regular structures. These are phrases such as: ‘starts on page v’, ‘section 1 starts on page 1’, ‘Table 2 on page 5’, ‘Figure 13 on page 46’, ‘Glossary:NOAA’, ‘1st AOP on page 4’, etc., according to context.

As there is only a single HTML page, focus remains on that page after following an internal hyperlink.

All the tabular material has a caption describing the purpose and a summary description describing the shape; i.e., number of rows and columns as well as how many have header- or data-cells. All (non-blank) data cells have a Headers array specifying the appropriate row- and column-header. Some tables have hierarchical headers.

There are no interactive Forms within this document; neither in PDF nor HTML. So none of the rules/recommendations are applicable.

Category: Widgets/Scripts

V W MC P
0 0 0 3

Category Results

Rule	Result	SC	Level	Req
ROLE must be valid	P	4.1.2	A	Y
ARIA values must be valid	P	4.1.2	A	Y
Attributes that start with ARIA- ...	P	4.1.2	A	Y
Closed shadow DOM requires m...	N/A	2.1.1	A	Y
Remove deprecated ARIA attrib...	N/A	4.1.1	A	Y
Role is not allowed	N/A	4.1.2	A	Y
Widgets must have label	N/A	4.1.2	A	Y
ONCLICK event handlers must h...	N/A	4.1.2	A	Y
Widgets must have properties	N/A	4.1.2	A	Y

There are no interactive scripts. ARIA roles are supplied, where appropriate. Where custom-roles or aria-labels may have been produced by the translation software, these have been removed from the HTML by post-processing scripts.

There are no other kinds of active scripts embedded within the PDF nor HTML versions of the document.

Category: Audio/Video

V W MC P
0 0 1 0

Category Results

Rule	Result	SC	Level	Req
Pause, stop or mute audio	MC	1.4.2	A	Y
AUDIO must have alternative	N/A	1.2.1	A	Y
OBJECT for audio must have alt...	N/A	1.2.1	A	Y
EMBED for audio must have alte...	N/A	1.2.1	A	Y
VIDEO for video only must have ...	N/A	1.2.1	A	Y
OBJECT for video only must hav...	N/A	1.2.1	A	Y
EMBED for video only must hav...	N/A	1.2.1	A	Y
VIDEO must have caption	N/A	1.2.2	A	Y
OBJECT for video must have ca...	N/A	1.2.2	A	Y

None of the images in this document contain embedded audio nor video; so most of the rules do not apply.

Category: Keyboard Support

V W MC P
0 0 3 1

Category Results

Rule	Result	SC	Level	Req
Skip to main content link	MC	2.4.1	A	Y
Focus order must be meaningful	MC	2.4.3	A	Y
Focus must be visible	MC	2.4.7	AA	Y
Interactive functionality must be...	P	2.1.1	A	Y
Widget roles require keyboard s...	N/A	2.1.1	A	Y
No keyboard trap	N/A	2.1.2	A	Y

A 'skip to Main' link is anchored on the logo at the top of the (1st) page, both in PDF and HTML. This anchor is first in the keyboard focus order; indeed it has 'autofocus' in the derived HTML. The focus order includes the ARIA landscapes, as well as (in some browsers) all active hyperlinks.

Category: Timing

V W MC P
0 0 3 0

Category Results

Rule	Result	SC	Level	Req
Control time limits	MC	2.2.1	A	Y
Control moving, blinking or auto...	MC	2.2.2	A	Y
Flashing limits	MC	2.3.1	A	Y

There are no images or other content that 'blinks', 'flashes' or otherwise changes with time. With all content known to be static, manual checks are unnecessary.

Category: Site Navigation

V W MC P
0 0 6 0

Category Results

Rule	Result	SC	Level	Req
TITLE must identify website and...	MC	2.4.2	A	Y
At least two ways of finding con...	MC	2.4.5	AA	Y
Consistent ordering of landmarks	MC	3.2.3	AA	Y
Consistent ordering of H1 and H...	MC	3.2.3	AA	Y
Consistent labeling of landmarks	MC	3.2.4	AA	Y
Consistent H1 and H2 page sect...	MC	3.2.4	AA	Y

The visible document 'Title' is located within the initial 'BANNER' landmark. This string also is used for the HTML <title> content in the <head> section. With the Bookmarks, Table of Contents, List of Tables, List of Figures, Glossaries and Tags tree, there are multiple ways to navigate to particular parts of the document. Keyboard navigation is supported by providing taborder attribute values for primary structural components. Since there is just a single HTML page, the 'consistency' requirements of SC 3.2.3 and 3.2.4 do not apply. Nevertheless, due to the production using L^AT_EX macro methods, documents built from similar sources and processed using the same modules will have a consistent 'look and feel'.