


Slide 1

**PDF A11y 102:
PDF/UA & the Matterhorn Protocol**

By Bouton Jones
IT Business Systems Analyst Senior
bouton.jones@gmail.com
My pronouns are he and him



May 13-16, 2024
John Slatin AccessU 2024
St. Edward's University

Presentation Title: PDF A11y 201

(formerly “Going Further with PDF/UA”)

Author: Bouton Jones

Bouton Jones
IT Business Systems Analyst Senior
bouton.jones@gmail.com
My pronouns are he and him

May 13-16, 2024
John Slatin AccessU 2024
St. Edward's University

Many of us look at A11y from the perspective of web development, but ignoring office documents and PDFs would be a huge mistake. When we do apply A11y to PDFs, it's by adapting WCAG principles. That's a great place to start. But there are other relevant guidelines to follow. In this session, we'll compare the WCAG and PDF/UA guidelines and recognize which individual guidelines apply to PDF Accessibility. We'll also look at the Matterhorn Protocol and see how it relates to PDF/UA. We'll also cover the necessity and techniques of embedding fonts in both MSO documents and PDFs.

This will build upon my earlier presentation, but it will not be necessary to have attended the earlier one to understand this next one.

HIGHLIGHTS:

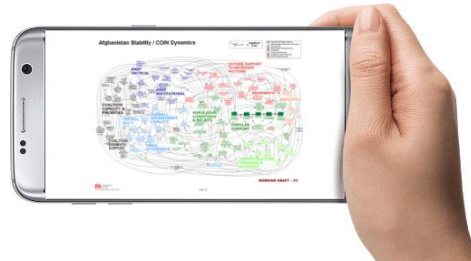
- How do the WCAG and PDF/UA guidelines relate to each other?
- Which specific guidelines apply to PDFs?
- What are their respective limitations and potential pitfalls?
- What is the Matterhorn Protocol and how does it apply?
- What are the checkpoints unique to the PDF/UA and which checkpoints might not be relevant and in what situations?
- Why would you lose headings when exporting Word documents to PDF?
- Why and how do we embed fonts?
- What happens to the document title when it is exported from a Word document to PDF?
- How can OCR be deceptive?
- How should AT address the content in headers and footers and what does that mean to document authors?
- When are "article threads" an issue?
- What is XFA and when is it a problem?
- What are the Base 14 fonts?
- What is the PDF Association's recommendation on the best approach to generating universally-accessible PDF documents?

THREE TAKE-AWAYS:

- Making the most accessible PDFs possible
- Locating the best single checklist for PDF A11y
- Identifying which guidelines are N/A

Slide 2

Life is too short to be spent
photographing slides.




A handout containing all the slides and speaker notes can be found at
<http://bit.ly/2y0p5kJ> . (Someone will post this URL to chat.)

Life is too short to be spent photographing PowerPoint slides.

A handout containing all the slides and speaker notes can be found at <http://bit.ly/2y0p5kJ> (or <http://www.boutonjones.com/pub/a11y.html> .) Someone will post this URL to chat.

Special Thanks

- David Ondich
- Dax Castro



Bouton Jones

3

Acknowledgement

Special Thanks to David Ondich.

David Ondich is a huge reason for why I put this presentation together as well and the manner and the content. David is the ADA Program Manager for the City of Austin and he is a user of assistive technology. Over the past two years David and I have been discussing and sharing information and perspective on this this topic. He's made me aware of situations that are otherwise invisible to me and I've explained to him the technological causes of his accessibility frustrations.

We've co-presented on Accessibility together several times. This session is largely based on that content but it's been expanded an refocused for a more technical perspective.

David Ondich, M.S., is the City of Austin, Americans with Disabilities Act Program Administrator and in 2014 was named Legal Subject Matter Expert on the Americans with Disabilities Act. His experience and education include

- 16 Years as ADA Program Manager or Coordinator for Austin and Fort Worth
- Master's Degree in Rehabilitation Counselling Psychology from UT Southwestern Medical School of Allied Health and 48 Doctoral hours in Rehabilitation Counseling from The University of Texas at Austin

He currently oversees municipal compliance with Titles I and II of the Americans with Disabilities Act, which includes:

- Providing feedback on accessibility to electronic information.
- Reviewing all employee requests for reasonable accommodations;
Consulting with department personnel regarding legal obligations under Titles I and II of the ADA
- Arranging for accommodations to allow the public access to all city facilities, programs and activities
- Providing training to city personnel regarding obligations under the ADA as amended.

“Nothing about us without us.”


Also

PDF Accessibility (A11y) 101 Review


The screenshot shows a Meetup event page. At the top left is the Meetup logo. To its right is a search bar with the placeholder text 'Search for keywords'. Below the search bar, the date 'Thursday, January 14, 2021' is displayed. The event title is 'Should You Remediate Your PDFs?'. Below the title, it says 'Hosted by Bouton J. and Paul A.' with a small profile picture icon. A 'Details' section follows, containing two paragraphs of text. The first paragraph discusses easy ways to ensure Microsoft Office documents and PDF files are fully accessible. The second paragraph states that the session will emulate the experience of using AT to read accessible and inaccessible documents from the perspective of PWD, and will explain how to apply optimal accessibility.

Handout: https://www.boutonjones.com/pub/a11y/a11y_accessu_2021_handout.pdf

In Review (Slide 1 of 5): Do Not Scan to PDF



PITFALL: When you scan a printed document – e.g. a signed contract, police report, corporate policy, or anything requested for discovery – it is no more than an image of the hard copy. It will not be accessible to screen readers in its current form.



“A picture may be worth a thousand words --- but not if it's a picture of a thousand words that you're trying to read with accessible technology.”

- David Ondich
ADA Program Manager for the City of Austin

If you must scan to PDF, then perform OCR and then remediate.

Bouton Jones 5

In Review (Slide 1 of 5): Do Not Scan to PDF

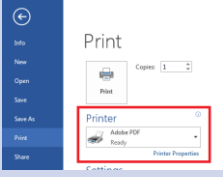
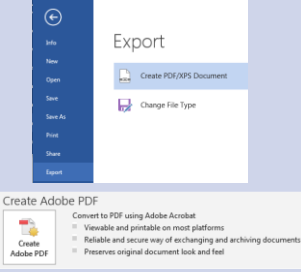
PITFALL: When you scan a printed document – e.g. a signed contract, police report, corporate policy, or anything requested for discovery – it is no more than an image of the hard copy. It will not be accessible to screen readers in its current form.

If you can't “copy and paste” text, it most likely is an image.

As David Ondich (The ADA Program Manager for the City of Austin) has said “A picture may be worth a thousand words --- but not if it's a picture of a thousand words that you're trying to read with accessible technology.”

Slide 6

In Review (Slide 2 of 5): Export to PDF

Never <i>Print</i> to PDF!	Always <i>Export</i> to PDF instead
	
Never print to PDF. Always export to --- or save as --- PDF. That way the semantic formatting (the tags) will be included.	
<i>Export to PDF is a better option than Save as PDF.</i>	

Wednesday, May 8, 2024 Bouton Jones 6
Get the slides and notes at <http://bit.ly/2y0p5kj> 6

In Review (Slide 2 of 5): Save to PDF

TABLE: Two column table comparing Print to PDF with Export to PDF.

Never print to PDF. Always export to --- or save as --- PDF. That way the semantic formatting (the tags) will be included.

Export to PDF is a better option than Save as PDF.

In Review (Slide 3 of 5)

- Only use PDFs when they are the best option.
- It's better to remediate and export an accessible source document to PDF than to remediate an inaccessible PDF.
- Testing
 - Perform both automated *and* manual testing.
 - Check the accessibility of your **source documents** before exporting to PDF.
 - Check the accessibility of your **PDFs** before publishing.

In Review (Slide 3 of 5)

- Don't use PDFs unless they are the best option.
- It's better to remediate and export an accessible source document to PDF than to remediate an inaccessible PDF.
- Testing
 - Perform both automated and manual testing.
 - Check the accessibility of your source documents before exporting to PDF.
 - Check the accessibility of your PDFs before publishing.

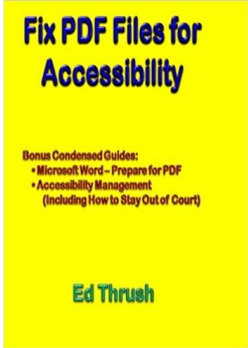
In Review (Slide 4 of 5)

- Apply WCAG Guidelines to PDFs
 - Use Semantic markup
 - Use Alternative text or captions for non-decorative images.
 - Specify the document language
 - Color guidelines (e.g., contrast, color blindness, etc.,)
 - Avoid using tables for layout
- Electronic Forms
 - Make sure form fields are in the correct order
 - Make sure form fields are properly labeled
- Properly redact documents electronically (do not mask.)

In Review (Slide 4 of 5)

- Apply WCAG Guidelines to PDFs
 - Use Semantic markup.
 - Use Alternative text or captions for non-decorative images.
 - Specify the document language
 - Color guidelines (e.g. contrast, color blindness, etc.,)
- Electronic Forms
 - Make sure form fields are in the correct order
 - Make sure form fields are properly labeled
- Properly redact documents electronically (do not mask.)

In Review (Slide 5 of 5)



Fix PDF Files for Accessibility: The Easy and Ultimate Guide to Applying Section 508 and WCAG 2.0 to PDF Files

1st Edition
by Ed Thrush
Paperback : 148 pages
ISBN-10 : 1514608936
ISBN-13 : 978-1514608937

<https://www.amazon.com/Fix-PDF-Files-Accessibility-Ultimate/dp/1514608936/>

Bouton Jones 9

In Review (Slide 5 of 5): A Book about PDF Remediation

Fix PDF Files for Accessibility: The Easy and Ultimate Guide to Applying Section 508 and WCAG 2.0 to PDF Files

1st Edition

by Ed Thrush

Paperback : 148 pages

ISBN-10 : 1514608936

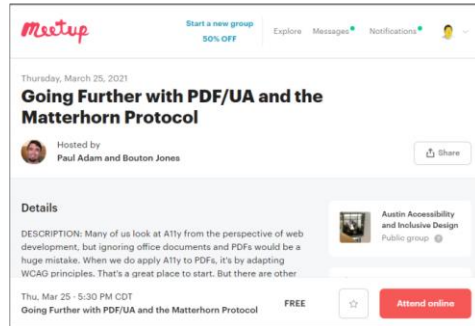
ISBN-13 : 978-1514608937

<https://www.amazon.com/Fix-PDF-Files-Accessibility-Ultimate/dp/1514608936/>

AMAZON.com DESCRIPTION: This book provides the easiest way to get to the required accessibility of a PDF file. It is rare for any PDF to come out perfect for accessibility and as a result almost all PDF files will need some type of work. In addition to the easy nuts and bolts methods to make PDF files work for the disabled, there are also condensed methods for preparation of documents commonly converted to PDF. In addition, there are instructions for managers and even contract language. You can get a legal history or technical lesson elsewhere, this book is for getting to work fixing PDF files for accessibility.

Slide 10

PDF A11y 102



The screenshot shows a Meetup event page. At the top, there's a navigation bar with the Meetup logo, a 'Start a new group' button with a '50% OFF' badge, and links for 'Explore', 'Messages', and 'Notifications'. Below this, the date 'Thursday, March 25, 2021' is displayed. The event title is 'Going Further with PDF/UA and the Matterhorn Protocol'. It is hosted by 'Paul Adam and Bouton Jones'. A 'Details' section follows, with a description: 'DESCRIPTION: Many of us look at A11y from the perspective of web development, but ignoring office documents and PDFs would be a huge mistake. When we do apply A11y to PDFs, it's by adapting WCAG principles. That's a great place to start. But there are other...'. To the right of the description is a profile for 'Austin Accessibility and Inclusive Design', a public group. At the bottom of the event card, it says 'Thu, Mar 25 - 5:30 PM CDT', 'Going Further with PDF/UA and the Matterhorn Protocol', 'FREE', a star icon, and a red 'Attend online' button.

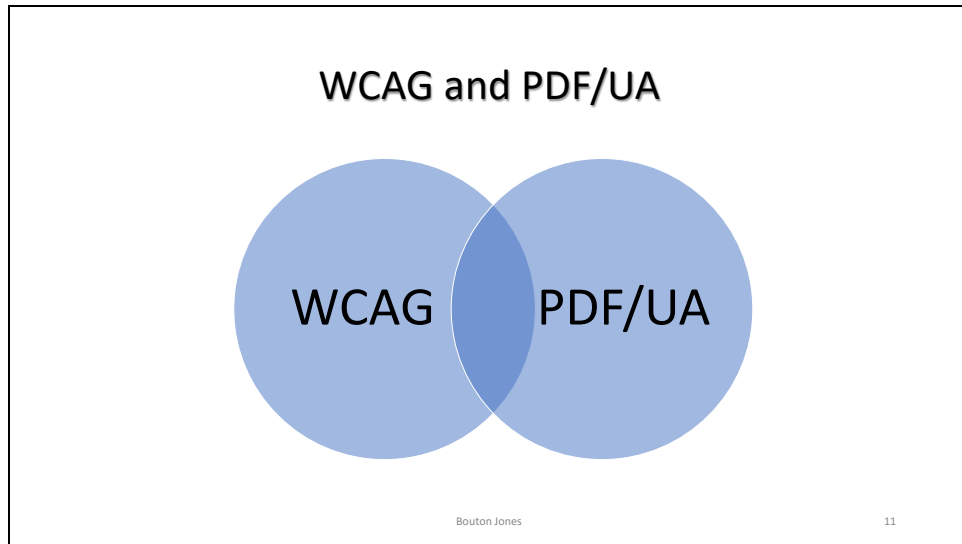
Meetup: <https://www.meetup.com/a11yATX/events/276826282/>

Bouton Jones

10

PDF A11y 201

This presentation will pick up where the last one left off.



Title Slide: WCAG and PDF/UA

IMAGE: Very simple Venn diagram for WCAG and PDF/UA. Two overlapping circles. No data points.

Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C)	PDF Association
<ol style="list-style-type: none">1. Web Content Accessibility Guidelines (WCAG)2. Authoring Tool Accessibility Guidelines (ATAG)3. User Agent Accessibility Guidelines (UAAG)	PDF/Universal Accessibility (PDF/UA) or ISO 14289-1 <ul style="list-style-type: none">• How to produce accessible PDFs• “Define compliant assistive technology and PDF reading software.”

Bouton Jones

12

Two Organizations and Their Guidelines

A table with two columns, a header row, and a data row.

First column

Header: The Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C)

Data:

- Web Content Accessibility Guidelines (WCAG)
- Authoring Tool Accessibility Guidelines (ATAG)
- User Agent Accessibility Guidelines (UAAG)


Second Column

Header: PDF Association



Data:

PDF/Universal Accessibility (PDF/UA) or ISO 14289-1

- How to produce accessible PDFs
- “Define compliant assistive technology and PDF reading software.”



Guidelines published by the International Organization for Standardization (ISO)



- Web Content Accessibility Guidelines (WCAG) 2.0
- World Wide Web Consortium (W3C)
- For websites, HTML-based material, and additional web technologies (*including PDFs*)
- Guidelines for A11y web designers, web developers, and CMS authors.
- 50 Checkpoints/Guidelines

- PDF/Universal Accessibility (PDF/UA) or ISO 14289-1
- PDF Association
- For PDF documents
- Guidelines for producing accessible PDFs but also "define compliant assistive technology and PDF reading software."
- 31 Checkpoints (many of which overlap with the WCAG)

Bouton Jones 13

International Organization for Standardization (ISO) Guidelines

- PDF/UA (Universal Access, or ISO 14289)
- Web Content Accessibility Guidelines (WCAG) 2.0

The International Standard for accessible PDF *technology* is ISO 32000-1 "PDF/UA (PDF/Universal Accessibility)"

On February 18, 2015 the US Access Board announced its Proposed Rule for US federal policy on accessibility, commonly known as Section 508. The proposed rule identifies PDF/UA as equivalent to WCAG 2.0 for "appropriate content".[4]

You can create PDFs to meet common accessibility standards, such as Web Content Accessibility Guidelines (WCAG) 2.0 and PDF/UA (Universal Access, or ISO 14289).

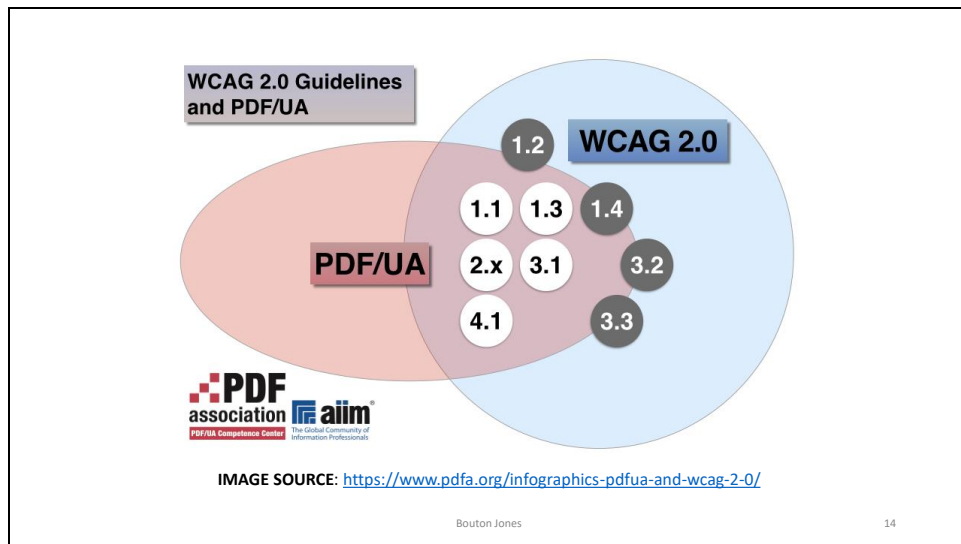


Diagram from the PDF Association

IMAGE: an overly complex Venn Diagram for PDF/UA and WCAG.

IMAGE SOURCE: <https://www.pdfa.org/infographics-pdfua-and-wcag-2-0/>

I didn't find this article helpful.

Guideline Comparison

Standard	Number of Guidelines
WCAG (not PDF applicable)	27
Guidelines in Common	23
PDF/UA Only	8

NOTE: This is subjective. These numbers are rough estimates and subject to interpretation. These guidelines rarely correspond one-to-one. Your findings will be close but might vary from mine.

Bouton Jones 15

Guideline Comparison

There is a large overlap in guidelines between WCAG and PDF/UA but each one also have guidelines that the other does not

DISPLAY: a two column table for 1) the specific standard and 2) number of guidelines

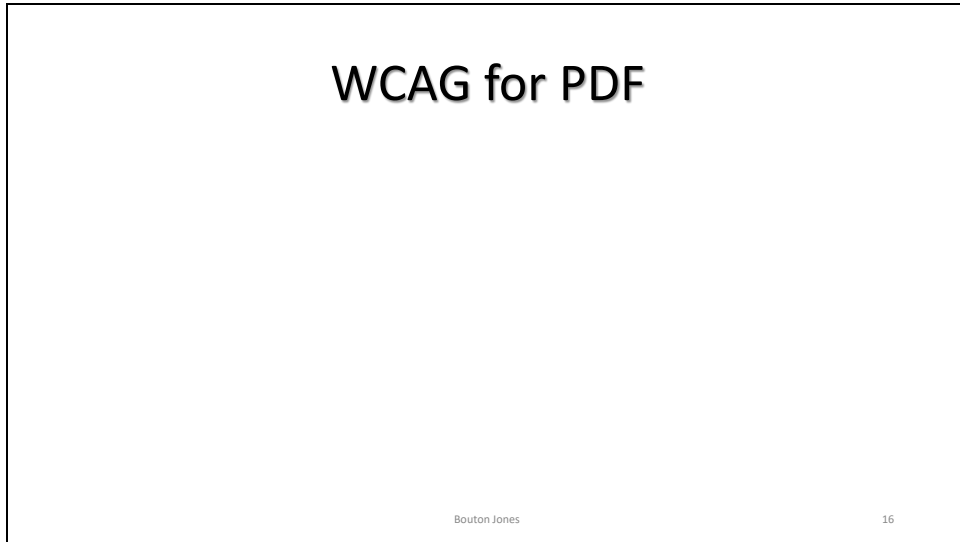
- 27 WCAG guidelines that are not applicable to PDF
- 23 guidelines in common with both WCAG and PDF/UA
- 8 more guidelines for PDF/UA only

NOTE: This is subjective. These numbers are rough estimates and subject to interpretation. These guidelines rarely correspond one-to-one. Your findings will be close to mine but might vary a bit.

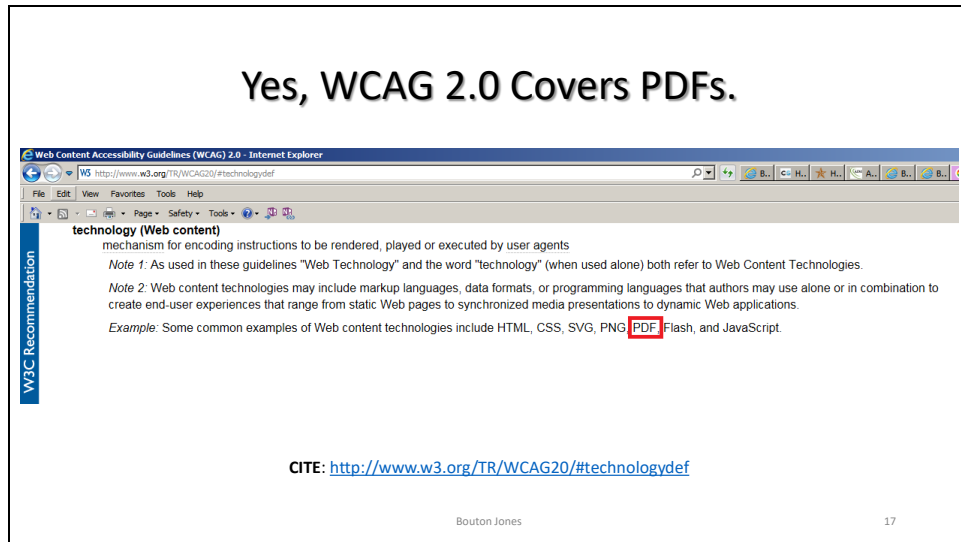
CITE:

- <https://www.w3.org/WAI/GL/WCAG20-TECHS/pdf.html>
- <https://www.pubcom.com/blog/2017-01-03/wcag-pdf-standards.html>

Slide 16



Title slide: WCAG for PDF



Yes, WCAG 2.0 Covers PDFs.

IMAGE: Screen capture of a page on the W3C website.

The W3C includes "PDF" as a web technology in their Web Content Accessibility Guidelines (WCAG) 2.0. Here's a blockquote from their website:

technology (Web content)

mechanism for encoding instructions to be rendered, played or executed by user agents

Note 1: As used in these guidelines "Web Technology" and the word "technology" (when used alone) both refer to Web Content Technologies.

Note 2: Web content technologies may include markup languages, data formats, or programming languages that authors may use alone or in combination to create end-user experiences that range from static Web pages to synchronized media presentations to dynamic Web applications.

Example: Some common examples of Web content technologies include HTML, CSS, SVG, PNG, PDF, Flash, and JavaScript.

Source: Web Content Accessibility Guidelines (WCAG) 2.0, Appendix A: Glossary, <http://www.w3.org/TR/WCAG20/#technologydef>

PDF Guidelines: WCAG Only

“WCAG 2.0 Requirements Beyond PDF/UA” (on the AIIM web site):
https://www.aiim.org/Global/AIIM_Widgets/Community_Widgets/Achieving_WCAG

Bouton Jones 18

PDF Guidelines: WCAG Only

- “WCAG 2.0 Requirements Beyond PDF/UA” (on the AIIM web site):
https://www.aiim.org/Global/AIIM_Widgets/Community_Widgets/Achieving_WCAG

From “WCAG 2.0 Requirements Beyond PDF/UA” (on the AIIM web site)

--- Begin Blockquote ---

WCAG 2.0 Requirements Beyond PDF/UA

WCAG 2.0 provides guidelines for general web content and is thus far broader in scope than PDF/UA. In some cases, therefore, PDF/UA conformance alone cannot assure conformance with WCAG 2.0 and direct reference to the WCAG 2.0 Guidelines is required.

The following list explains terms used in the Achieving WCAG 2.0 with PDF/UA table to help identify cases in which a WCAG 2.0 conformance claim cannot use PDF/UA alone to conform to WCAG 2.0 but requires validation against WCAG 2.0 directly in addition to applicable PDF/UA requirements.

- Time-based media alternatives, test, sensory and CAPTCHAs. PDF/UA does not address these items. Refer to WCAG 2.0.

- Audio and/or Video content. PDF/UA includes syntactical requirements but does not address the actual audio and video content. Refer to WCAG 2.0.
- Video captions. PDF/UA does not address video captions. Refer to WCAG 2.0.
- JavaScript design. While identifying some requirements for accessible scripting with PDF technology, PDF/UA is not a JavaScript accessibility standard. Refer to WCAG 2.0.
- JavaScript or Media content. As above, PDF/UA is not a JavaScript or multimedia accessibility standard. Refer to WCAG 2.0.
- Control device specific. PDF/UA is control device-independent. Refer to WCAG 2.0.
- Design consideration. In general, PDF/UA does not address visual appearance, content or design choices and relies on semantic equivalents. Refer to WCAG 2.0.
- Dynamic XFA-PDF. While prohibited in PDF/UA, dynamic XFA-PDF is not prohibited by WCAG 2.0.

--- End Blockquote ---

PDF Techniques in WCAG 2.0 (from the W3C WAI)

1. Applying text alternatives to images with the Alt entry in PDF documents
2. Creating bookmarks in PDF documents
3. Ensuring correct tab and reading order in PDF documents
4. Hiding decorative images with the Artifact tag in PDF documents
5. Indicating required form controls in PDF forms
6. Using table elements for table markup in PDF Documents
7. Performing OCR on a scanned PDF document to provide actual text
8. Providing definitions for abbreviations via an E entry for a structure element
9. Providing headings by marking content with heading tags in PDF documents
10. Providing labels for interactive form controls in PDF documents
11. Providing links and link text using the Link annotation and the /Link structure element in PDF documents
12. Providing name, role, value information for form fields in PDF documents
13. Providing replacement text using the /Alt entry for links in PDF documents
14. Providing running headers and footers in PDF documents
15. Providing submit buttons with the submit-form action in PDF forms
16. Setting the default language using the /Lang entry in the document catalog of a PDF document
17. Specifying consistent page numbering for PDF documents
18. Specifying the document title using the Title entry in the document information dictionary of a PDF document
19. Specifying the language for a passage or phrase with the Lang entry in PDF documents
20. Using Adobe Acrobat Pro's Table Editor to repair mistagged tables
21. Using List tags for lists in PDF documents
22. Indicating when user input falls outside the required format or values in PDF forms
23. Providing interactive form controls in PDF documents

CITE: <https://www.w3.org/WAI/GL/WCAG20-TECHS/pdf.html>

Bouton Jones

19

PDF Techniques in WCAG 2.0 (from the W3C)

This is a huge web page! Here's an outline of the 23 techniques it covers.

1. Applying text alternatives to images with the Alt entry in PDF documents
2. Creating bookmarks in PDF documents
3. Ensuring correct tab and reading order in PDF documents
4. Hiding decorative images with the Artifact tag in PDF documents
5. Indicating required form controls in PDF forms
6. Using table elements for table markup in PDF Documents
7. Performing OCR on a scanned PDF document to provide actual text
8. Providing definitions for abbreviations via an E entry for a structure element
9. Providing headings by marking content with heading tags in PDF documents
10. Providing labels for interactive form controls in PDF documents
11. Providing links and link text using the Link annotation and the /Link structure element in PDF documents
12. Providing name, role, value information for form fields in PDF documents
13. Providing replacement text using the /Alt entry for links in PDF documents
14. Providing running headers and footers in PDF documents
15. Providing submit buttons with the submit-form action in PDF forms
16. Setting the default language using the /Lang entry in the document catalog of a PDF document
17. Specifying consistent page numbering for PDF documents

18. Specifying the document title using the Title entry in the document information dictionary of a PDF document
19. Specifying the language for a passage or phrase with the Lang entry in PDF documents
20. Using Adobe Acrobat Pro's Table Editor to repair mistagged tables
21. Using List tags for lists in PDF documents
22. Indicating when user input falls outside the required format or values in PDF forms
23. Providing interactive form controls in PDF documents

CITE: <https://www.w3.org/WAI/GL/WCAG20-TECHS/pdf.html>

PDF Guidelines: WCAG *and* PDF/UA

- “PDF Techniques for WCAG 2.0” (W3C WAI):
<https://www.w3.org/WAI/GL/WCAG20-TECHS/pdf.html>
- “Accessibility Support Documentation for PDF” (W3C WAI) This web page on the W3C web site includes "Support Information for WCAG 2.0 Checkpoints": https://www.w3.org/WAI/GL/WCAG20/implementation-report/PDF_accessibility_support_statement
- “WCAG Mapping to PDF/UA” (from Adobe):
<https://helpx.adobe.com/acrobat/using/create-verify-pdf-accessibility.html#WCAGmappingtoPDFUA>

PDF Guidelines: WCAG Only

- “PDF Techniques for WCAG 2.0” (W3C WAI): <https://www.w3.org/WAI/GL/WCAG20-TECHS/pdf.html>
- “Accessibility Support Documentation for PDF” (W3C WAI) This web page on the W3C web site includes "Support Information for WCAG 2.0 Checkpoints":
https://www.w3.org/WAI/GL/WCAG20/implementation-report/PDF_accessibility_support_statement

“WCAG mapping to PDF/UA” (from Adobe)

WCAG mapping to PDF/UA

WCAG 2.0	ISO 14289 -1 (File)	Techniques
1.1.1 Non-text Content (A)	<ul style="list-style-type: none">• 7.3 addresses content requiring text alteration.• 7.18.1 paragraph four addresses control descriptions.• 7.18.6.2 addresses time-based media alternatives. Text, Sensory, and CAPTCHA use-cases are addressed via the technical means used.• 7.1 paragraph 1, sentence 2 addresses decoration.	<ul style="list-style-type: none">• PDF1• PDF4
1.2.1 Audio-only and Video-only (Prerecorded) (A)	<ul style="list-style-type: none">• 7.18.6.2 addresses time-based media alternatives. Design-specific. It's necessary that authors and developers consider this provision and ensure conformance.	<ul style="list-style-type: none">• General Techniques: http://www.w3.org/WAI/WCAG20/quickref/#media-equiv-av-only-alt

URL: <https://helpx.adobe.com/acrobat/using/create-verify-pdf-accessibility.html#WCAGmappingtoPDFUA>

“WCAG Mapping to PDF/UA” (from Adobe)

IMAGE: Screen Capture of a detail of the mapping on a web page on Adobe.com.

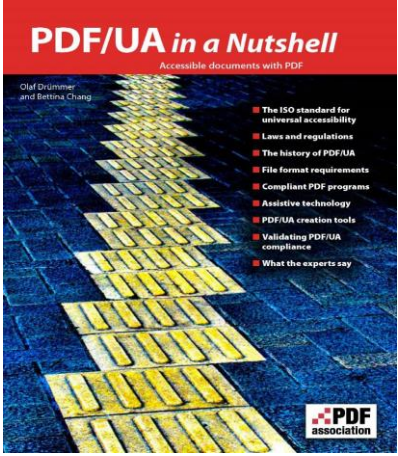
URL: <https://helpx.adobe.com/acrobat/using/create-verify-pdf-accessibility.html#WCAGmappingtoPDFUA>

Slide 22

PDF/UA
and the Matterhorn Protocol

Bouton Jones 22

Title Slide: PDF/UA and the Matterhorn Protocol



The image shows the cover of the brochure 'PDF/UA in a Nutshell'. The title is in a red box at the top. Below it, the authors' names and the subtitle 'Accessible documents with PDF' are listed. A central graphic shows a perspective view of a blue and yellow grid pattern. To the right of the grid is a list of topics covered in the brochure, including the ISO standard, laws, history, file format requirements, compliant programs, assistive technology, creation tools, validation, and expert opinions. The PDF Association logo is at the bottom right of the cover.

PDF/UA in a Nutshell
Accessible documents with PDF
Olaf Drümmer and Bettina Chang

- The ISO standard for universal accessibility
- Laws and regulations
- The history of PDF/UA
- File format requirements
- Compliant PDF programs
- Assistive technology
- PDF/UA creation tools
- Validating PDF/UA compliance
- What the experts say

PDF association

PDF/UA in a Nutshell
By Olaf Drümmer and Bettina Chang
Published by the PDF Association
August 1, 2013
21 pages
Free
<https://www.pdfa.org/resource/pdf-ua-in-a-nutshell/>

“Rather than applying to the PDF file format alone, these clear specifications also define compliant assistive technology and PDF reading software.”

Bouton Jones 23

PDF/UA in a Nutshell

By Olaf Drümmer and Bettina Chang
Published by the PDF Association
August 1, 2013
21 pages
Free

<https://www.pdfa.org/resource/pdfua-in-a-nutshell/>

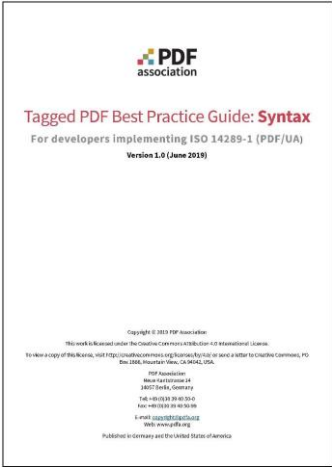
ISO standard 14289-1, known as PDF/UA, was first published in the summer of 2012. It is the first complete definition of a set of requirements for universally accessible PDF documents. Rather than applying to the PDF file format alone, these clear specifications also define compliant assistive technology and PDF reading software.

In signing the UN Convention on the Rights of Disabled Persons, many countries have chosen to work towards building a more inclusive society. They are now required to turn these inclusive principles into a reality on the national level and to remove existing barriers to social inclusion. One key area of activity here is in access to information and communications in the form of documents a subject which PDF/UA now stands to greatly simplify.

PDF/UA in a Nutshell will show you just what PDF/UA is. This brochure provides a clear summary of the key facts, the standard s requirements, the latest developments and the current

legal situation regarding PDF accessibility, all in one place. In addition, it suggests potential applications for the standard and lists the tools that can be used to create and validate PDF/UA-compliant documents.

CITE: <https://www.pdfa.org/resource/pdfua-in-a-nutshell/>



The image shows the cover of the document "Tagged PDF Best Practice Guide: Syntax". At the top, it features the PDF Association logo. The title "Tagged PDF Best Practice Guide: Syntax" is prominently displayed in red, followed by the subtitle "For developers implementing ISO 14289-1 (PDF/UA) Version 1.0 (June 2019)". The cover also includes copyright information for the PDF Association and contact details for the PDF Association, including a website URL, phone number, and email address. The document is published in Germany and the United States of America.

Tagged PDF Best Practice Guide: Syntax

"A publication of the PDF/UA Technical Working Group, this document is intended for developers implementing tagged PDF and PDF/UA. Others (including authors with some technical knowledge of PDF's accessibility mechanisms) may also benefit from this document.

"This guide is also intended to be useful for those performing detailed accessibility testing on PDF documents claiming conformance with PDF/UA, or on PDF documents claiming to be accessible according to some other specification."

URL: <https://www.pdfa.org/resource/tagged-pdf-best-practice-guide-syntax/>

Bouton Jones 24

Tagged PDF Best Practice Guide: Syntax

A publication of the [PDF/UA Technical Working Group](https://www.pdfa.org/resource/tagged-pdf-best-practice-guide-syntax/), this document is intended for developers implementing tagged PDF and PDF/UA. Others (including authors with some technical knowledge of PDF's accessibility mechanisms) may also benefit from this document.

This guide is also intended to be useful for those performing detailed accessibility testing on PDF documents claiming conformance with PDF/UA, or on PDF documents claiming to be accessible according to some other specification.

Keep in mind that as of today most PDF documents do not follow these recommendations.

URL: <https://www.pdfa.org/resource/tagged-pdf-best-practice-guide-syntax/>

Introduction to the Web Content Accessibility Guidelines 2.0

The requirements for PDF/UA are based on the principles of Version 2.0 of the W3C consortium's Web Content Accessibility Guidelines (WCAG), also published in identical form as an ISO standard in October 2012 (ISO/IEC 40500). WCAG 2.0 guides the creation of accessible web content, focusing on the kinds of content and applications usually provided on websites.

Some WCAG 2.0 tips of a purely informative nature suggest how barriers to access can also be avoided in PDF documents, but a systematic explanation or guide is not provided. PDF/UA, on the other hand, exclusively and fully addresses the PDF format, bringing the ideas behind WCAG 2.0 into the world of PDF and setting PDF-specific rules.

CITE: <https://www.pdfa.org/resource/pdfua-in-a-nutshell/>

Bouton Jones 25

The PDF/UA is based on WCAG

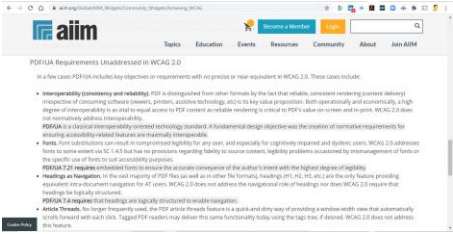
“The requirements for PDF/UA are based on the principles of Version 2.0 of the W3C consortium’s Web Content Accessibility Guidelines (WCAG), also published in identical form as an ISO standard in October 2012 (ISO/IEC 40500). WCAG 2.0 guides the creation of accessible web content, focusing on the kinds of content and applications usually provided on websites.”

“Some WCAG 2.0 tips of a purely informative nature suggest how barriers to access can also be avoided in PDF documents, but a systematic explanation or guide is not provided. PDF/UA, on the other hand, exclusively and fully addresses the PDF format, bringing the ideas behind WCAG 2.0 into the world of PDF and setting PDF-specific rules.”

CITE: <https://www.pdfa.org/resource/pdfua-in-a-nutshell/>

PDF Guidelines: PDF/UA Only

“PDF/UA Requirements Unaddressed in WCAG 2.0” (AIIM):
https://www.aiim.org/Global/AIIM_Widgets/Community_Widgets/Achieving_WCAG



The screenshot shows a web browser displaying the AIIM website. The article title is "PDF/UA Requirements Unaddressed in WCAG 2.0". The content discusses key objectives and requirements for PDF/UA that are not covered by WCAG 2.0. Key points include:

- Interoperability (consistency and reliability):** PDF is distinguished from other formats by the fact that reliable, consistent rendering (content delivery) irrespective of consuming software (viewers, printers, assistive technology, etc.) is its key value proposition. Both operationally and economically, a high degree of interoperability is as vital to equal access to PDF content as reliable rendering is critical to PDF's value on-screen and in-print. WCAG 2.0 does not normatively address interoperability.
- PDF/UA is a classical interoperability-oriented technology standard.** A fundamental design objective was the creation of normative requirements for ensuring accessibility-related features are maximally interoperable.
- Fonts:** Font substitutions can result in compromised legibility for any user, and especially for cognitively impaired and dyslexic users. WCAG 2.0 addresses fonts to some extent via SC (Success Criterion 1.4.12).
- PDF/UA PDF require embedded fonts** to ensure the accurate appearance of the author's intent with the highest degree of legibility.
- Headings as navigation:** In the vast majority of PDF files (as well as in other file formats), headings (H1, H2, H3, etc.) are the only future-proofing mechanism for document navigation for all users. WCAG 2.0 does not address the navigational role of headings nor does WCAG 2.0 require that headings be highly or custom.
- PDF/UA PDF require that headings are highly structured to enable navigation:**
- Article Threads:** An (often) frequently used PDF article threads feature is a quick-and-dirty way of providing a window into views that automatically scroll forward with each click. Tagged PDF headers may deliver the same functionality today using the tags tree. If done, WCAG 2.0 does not address this feature.

Bouton Jones 26

PDF Guidelines: WCAG Only

“PDF/UA Requirements Unaddressed in WCAG 2.0” (AIIM):
https://www.aiim.org/Global/AIIM_Widgets/Community_Widgets/Achieving_WCAG

--- Begin Blockquote ---

In a few cases PDF/UA includes key objectives or requirements with no precise or near-equivalent in WCAG 2.0. These cases include:

- **Interoperability (consistency and reliability).** PDF is distinguished from other formats by the fact that reliable, consistent rendering (content delivery) irrespective of consuming software (viewers, printers, assistive technology, etc.) is its key value proposition. Both operationally and economically, a high degree of interoperability is as vital to equal access to PDF content as reliable rendering is critical to PDF's value on-screen and in-print. WCAG 2.0 does not normatively address interoperability.

PDF/UA is a classical interoperability-oriented technology standard. A fundamental design objective was the creation of normative requirements for ensuring accessibility-related features are maximally interoperable.

- **Fonts.** Font substitutions can result in compromised legibility for any user, and especially for cognitively impaired and dyslexic users. WCAG 2.0 addresses fonts to some extent via SC

1.4.5 but has no provisions regarding fidelity to source content, legibility problems occasioned by mismanagement of fonts or the specific use of fonts to suit accessibility purposes.

PDF/UA 7.21 requires embedded fonts to ensure the accurate conveyance of the author's intent with the highest degree of legibility.

- Headings as Navigation. In the vast majority of PDF files (as well as in other file formats), headings (H1, H2, H3, etc.) are the only feature providing equivalent intra-document navigation for AT users. WCAG 2.0 does not address the navigational role of headings nor does WCAG 2.0 require that headings be logically structured.

PDF/UA 7.4 requires that headings are logically structured to enable navigation.

- Article Threads. No longer frequently used, the PDF article threads feature is a quick-and-dirty way of providing a window-width view that automatically scrolls forward with each click. Tagged PDF readers may deliver this same functionality today using the tags tree, if desired. WCAG 2.0 does not address this feature.

PDF/UA 7.12 requires that if present, article threads shall follow the logical structure (tags) of the document.

--- End Blockquote ---

PDF/UA Technical Requirements

- All real (meaningful) content shall be tagged. Artifacts (non-meaningful content) shall not be tagged.
- The “tag tree” created by the individual tags (representing the document’s content) should be sorted to reflect the document’s logical reading order.
- All structure types in use shall be standard or mapped to standard
- A manual check of the document is required to ensure that there is no information conveyed using visual means alone (e.g., contrast, color etc.), and that there is no flickering, blinking, or flashing content.
- A document title must be given in the metadata. In the Document Properties, the Initial View settings must be set up so that the title (rather than the file name) appears in the window title (under Window Options).
- Any non-text elements (e.g., images, charts, graphs, infographics, logos, etc.) must have corresponding alternative text.

CITE: <https://accessible-docs.com/overview-pdf-ua-standard/>

Bouton Jones

27

PDF/UA Technical Requirements

“WCAG 2.0 compliance

“The W3C’s Web Content Accessibility Guidelines 2.0 (WCAG 2.0) was written with Web content in mind. The application of WCAG 2.0 to other electronic content requires interpretation to determine how to apply the WCAG 2.0 Principles, Guidelines and Success Criteria. There is not always a 1:1 mapping.

“In the case of PDF documents, the international standard (ISO14289-1:2014) defines requirements specifically for accessible PDF files. The requirements for the PDF/UA standard are complementary to the WCAG 2.0 success criteria, and are therefore compliant with its requirements. This means that a PDF file which is compliant to the PDF/UA standard will also be compliant to ADA, Section 508 and AODA requirements.

“PDF/UA technical requirements

“The following are just a few of the technical requirements as set out by the PDF/UA standard:”

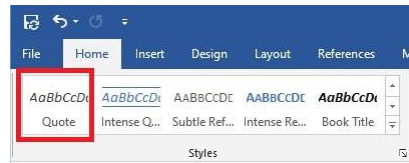
- All real (meaningful) content shall be tagged. Artifacts (non-meaningful content) shall not be tagged.
- The “tag tree” created by the individual tags (representing the document’s content) should be sorted to reflect the document’s logical reading order.

- All structure types in use shall be standard or mapped to standard
- A manual check of the document is required to ensure that there is no information conveyed using visual means alone (e.g. contrast, color etc.), and that there is no flickering, blinking or flashing content.
- A document title must be given in the metadata. In the Document Properties, the Initial View settings must be set up so that the title (rather than the file name) appears in the window title (under Window Options).
- Any non-text elements (images, charts, graphs, infographics, logos, etc.) must have corresponding alternative text.

CITE: <https://accessible-docs.com/overview-pdf-ua-standard/>

PDF/UA *Beyond* the WCAG (Slide 1 of 2)

- Document Title
- Quotes and Blockquotes
- Embed Fonts
- Article Threads
- Captions
- Don't use background images or watermarks
- Don't use flickering, blinking or flashing content
- Don't use Dynamic XFA forms



PDF/UA Beyond the WCAG (Slide 1 of 2)

- Document Title
- Quotes and Blockquotes
- Embedded Fonts
- Article Threads
- Captions
- Don't use background images or watermarks
- Don't use Dynamic XFA forms. (If you don't know what they are don't worry about it.)

PDF/UA *Beyond* the WCAG (slide 2 of 2)

- Use *standard structure types* --- or map to standard structure types with a role map in the structure tree root.
- Tag any information --- that was conveyed by visual means alone (contrast, color, format, or layout) --- to reflect the intended meaning.
- Headings
 - Only use Heading levels 1 through 6
 - You *can* have more than one Heading level 1
 - Do not separate a Heading into two parts

PDF/UA Beyond the WCAG (Slide 2 of 2)

- Use standard structure types or map to standard structure types, with a role map in the structure tree root.
- Any Information conveyed by contrast, color, format or layout must also be tagged to reflect the intended meaning.
- Headings
 - You can have more than 1 Heading level 1
 - Only use heading levels H1 – H6. Unlike HTML, the PDF format doesn't include additional heading levels (e.g. H7.) This is not an issue in Word
 - Do not separate a Heading into two parts

What is the Matterhorn?



Bouton Jones

30

Chat Poll: What is the c?

IMAGE: The Matterhorn

It is a mountain of the Alps with a summit 14,692 feet (or 4,478 metres) high, making it one of the highest summits in Europe.

FROM Wikipedia: “The first ascent of the Matterhorn was in 1865 from Zermatt by a party led by Edward Whymper; it ended when four of its seven members fell to their deaths on the descent. This disaster, later portrayed in several films, marked the end of the golden age of alpinism. It is estimated that over 500 alpinists have died on the Matterhorn, making it one of the deadliest peaks in the world.”

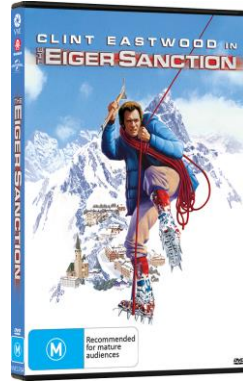
What the word “Matterhorn” means to me:

- Massive
- Monolithic
- Insurmountable
- Dangerous
- Unavoidable
- One of the highest summits in Europe
- “One of the deadliest peaks in the world.” (From Wikipedia)

What the word “Matterhorn” means to me.

- Massive
- Monolith
- Insurmountable
- One of the highest summits in Europe
- “It is estimated that over 500 alpinists have died on the Matterhorn, making it one of the deadliest peaks in the world.” (From Wikipedia)

The Title “Matterhorn Protocol”



Bouton Jones

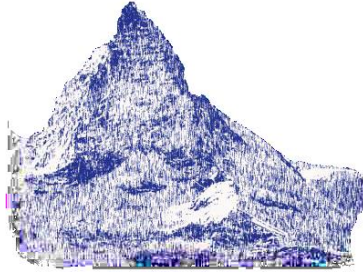
32

The Title “Matterhorn Protocol”

It reminds me of the book and the movie The Eiger Sanction.

What is the Matterhorn Protocol?

Is it another set of **massive** and **insurmountable** A11y guidelines?



Bouton Jones

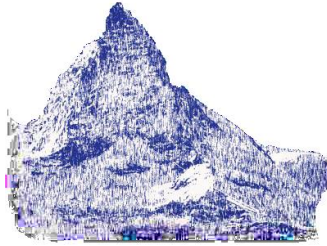
33

What is the Matterhorn Protocol?

Is it another set of massive, insurmountable, deadly A11y guidelines to learn?

Matterhorn Protocol:

A list of all the possible ways to fail PDF/UA produced by the PDF Association's PDF/UA Competence Center. It consists of 31 Checkpoints comprised of 136 Failure Conditions.



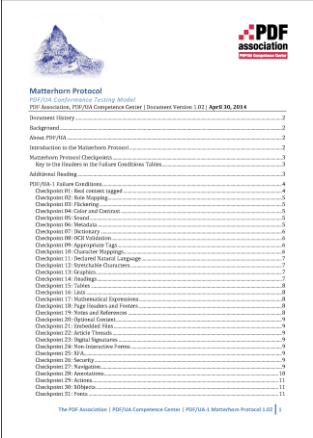
It's based on the PDF/UA.

Bouton Jones

34

Matterhorn Protocol defined

Matterhorn Protocol: a list of all the possible ways to fail PDF/UA produced by the PDF Association's PDF/UA Competence Center. It consists of 31 Checkpoints comprised of 136 Failure Conditions.



The image shows the cover page of the Matterhorn Protocol PDF document. At the top left is a logo of a mountain peak. At the top right is the PDF Association logo. The title 'Matterhorn Protocol' is centered, followed by the subtitle 'PDF User Certification Training Model' and 'PDF Association, PDF/A Compliance Center | Document Version 1.02 | April 30, 2014'. Below this is a detailed table of contents listing sections like 'Background', 'Introduction to the Matterhorn Protocol', and 'Matterhorn Protocol Objectives', along with a list of 32 chapters. At the bottom of the page, it states 'Document Version 1.02 | April 30, 2014 | 13 pages' and provides a URL: https://www.pdfa.org/wp-content/uploads/2014/06/MatterhornProtocol_1-02.pdf. The footer of the slide contains the name 'Bouton Jones' and the page number '35'.

Download the Matterhorn Protocol PDF

Document Version 1.02 | April 30, 2014 | 13 pages

URL: https://www.pdfa.org/wp-content/uploads/2014/06/MatterhornProtocol_1-02.pdf

Matterhorn Protocol: PDF/UA-1 Checkpoints

- | | | |
|------------------------|------------------------------|---------------------------|
| 1. Real content tagged | Language | 21. Embedded Files |
| 2. Role Mapping | 12. Stretchable Characters | 22. Article Threads |
| 3. Flickering | 13. Graphics | 23. Digital Signatures |
| 4. Color and Contrast | 14. Headings | 24. Non-Interactive Forms |
| 5. Sound | 15. Tables | 25. XFA |
| 6. Metadata | 16. Lists | 26. Security |
| 7. Dictionary | 17. Mathematical Expressions | 27. Navigation |
| 8. OCR Validation | 18. Page Headers and Footers | 28. Annotations |
| 9. Appropriate Tags | 19. Notes and References | 29. Actions |
| 10. Character Mappings | 20. Optional Content | 30. XObjects |
| 11. Declared Natural | | 31. Fonts |

Matterhorn Protocol: PDF/UA-1 Checkpoints

- Real content tagged
- Role Mapping
- Flickering
- Color and Contrast
- Sound
- Metadata
- Dictionary
- OCR Validation
- Appropriate Tags
- Character Mappings
- Declared Natural Language
- Stretchable Characters
- Graphics
- Headings
- Tables
- Lists
- Mathematical Expressions
- Page Headers and Footers
- Notes and References
- Optional Content
- Embedded Files

- Article Threads
- Digital Signatures
- Non-Interactive Forms
- XFA
- Security
- Navigation
- Annotations
- Actions
- XObjects
- Fonts

Checkpoint 01: Real content tagged

Use *standard structure types* --- or map to standard structure types with a role map in the structure tree root.

Failure Condition 1-006 : The *structure type* and attributes of a structure element are not semantically appropriate for the structure element.

48 PDF Structure Types: Document, Part, Art, Sect, Div, BlockQuote, Caption, TOC, TOCI, Index, NonStruct, Private, P, H, H1, H2, H3, H4, H5, H6, L, LI, Lbl, LBody, Table, TR, TH, TD, THead, TBody, TFoot, Span, Quote, Note, Reference, BibEntry, Code, Link, Annot, Ruby, Warichu, RB, RT, RP, WT, WP, Figure, Formula, and Form.

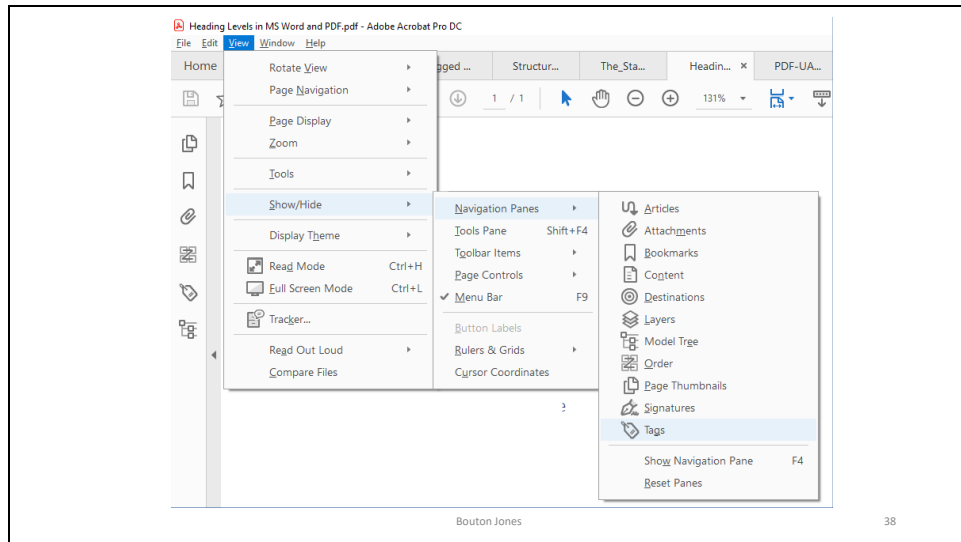
Bouton Jones 37

Checkpoint 01: Real content tagged

Failure Condition 1-006 : The structure type and attributes of a structure element are not semantically appropriate for the structure element.

48 PDF Structure Types: Document, Part, Art, Sect, Div, BlockQuote, Caption, TOC, TOCI, Index, NonStruct, Private, P, H, H1, H2, H3, H4, H5, H6, L, LI, Lbl, LBody, Table, TR, TH, TD, THead, TBody, TFoot, Span, Quote, Note, Reference, BibEntry, Code, Link, Annot, Ruby, Warichu, RB, RT, RP, WT, WP, Figure, Formula, and Form.

Slide 38



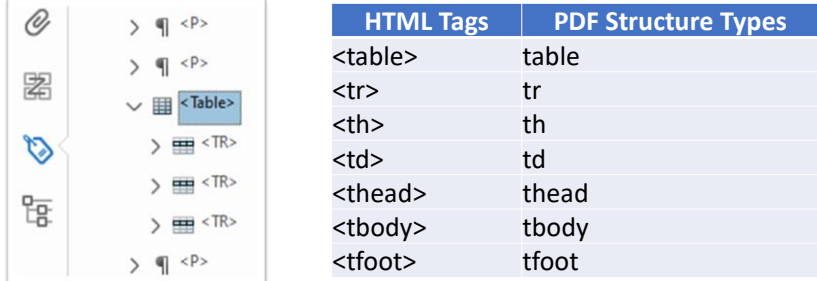
Navigation to the Tag Panel In Acrobat

IMAGE: Screen capture showing the following navigation route.

- Select the View drop down menu
- Select the Show/Hide option
- Select the Navigation Panes option
- Select the Tags option

The Tags pane will appear on the right-hand side of the Acrobat window.

Table Structure Types



The image shows a screenshot of the Acrobat Tag pane on the left, with a list of tags including <P>, <Table>, <TR>, and <TD>. The <Table> tag is selected. To the right is a table with two columns: 'HTML Tags' and 'PDF Structure Types'. The table lists the following pairs: <table> to table, <tr> to tr, <th> to th, <td> to td, <thead> to thead, <tbody> to tbody, and <tfoot> to tfoot.

HTML Tags	PDF Structure Types
<table>	table
<tr>	tr
<th>	th
<td>	td
<thead>	thead
<tbody>	tbody
<tfoot>	tfoot

THEAD, TBODY, and TFOOT help with pagination.
Adobe InDesign provides TFOOT support.

Bouton Jones 39

Structure Types for Tables

IMAGE: Screen capture of a few table tags in Acrobat’s Tag pane.

Two Column Table with a column header for “HTML Tags” and “PDF Structure Types”. There are 7 rows and the contents for both cells in each row are identical. The repeated contents of the rows are: Table, TR, TH, TD, THead, TBody, Tfoot.

So there is a one-to-one correspondence for table tags in HTML and PDF Structure Types for tables.

THEAD, TBODY, and TFOOT help with pagination.

Adobe InDesign provides TFOOT support.

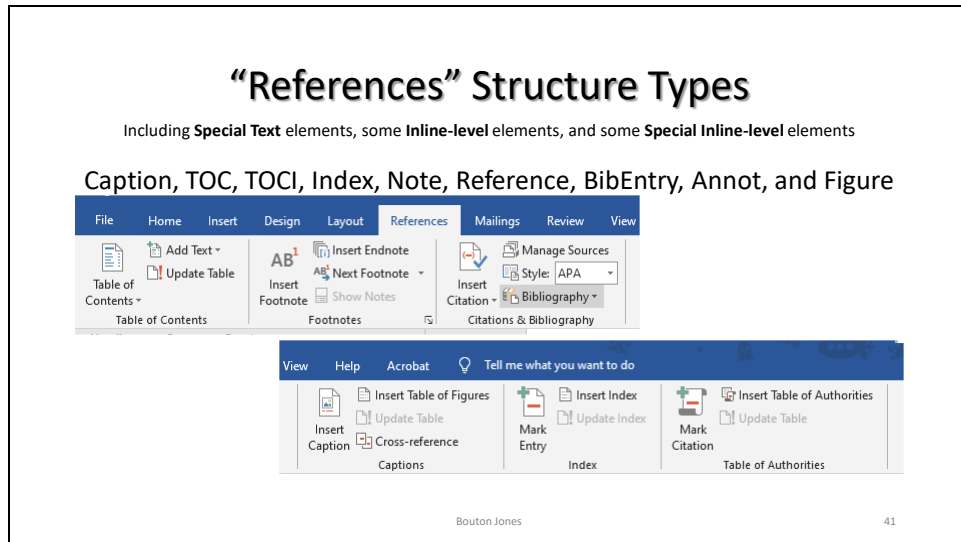
Label and List Structure Types

PDF Structure Types	HTML Tags
L - the List tag, which contains one or more LI tags.	 &
LI - the List Item tag. List item tags can contain Lbl and LBody tags.	
Lbl - the list item label. Contains distinguishing information such as a item number or bullet character.	*
LBody - the list item body. Contains list item content, or in the case of a nested list, it may contain additional List tag trees.	

***NOTE:** HTML's <label> tag is only used in *forms*. It doesn't apply to lists.

Structure Types for Lists

- **L** - the List tag, which contains one or more LI tags.
- **LI** - the List Item tag. List item tags can contain Lbl and LBody tags.
- **Lbl** - the list item label. Contains distinguishing information such as a item number or bullet character.
- **LBody** - the list item body. Contains list item content, or in the case of a nested list, it may contain additional List tag trees.

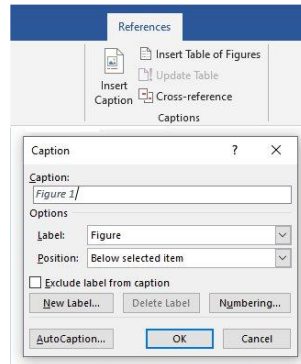


Structure Types for References

- Caption
- TOC
- TOCI
- Index
- Note
- Reference
- BibEntry
- Annot
- Figure

IMAGE: Screen capture of Microsoft Word ribbon with the References tab selected. It shows the Word options that are equivalent to many of the PDF Structure Type (e.g. Table of Contents, Insert Footnote, Insert Citation, Bibliography, Insert Caption, Insert Index, etc.,)

Adding Captions in Word



Bouton Jones

42

Adding Captions in Word

IMAGE: Screen capture of a detail of the References tab in Word. A captions window below

Header and Paragraph Structure Types

- H (heading, strongly structured)*
- H1
- H2
- H3
- H4
- H5
- H6
- P

**"Due to a lack of suitable tools, the H structure element is impractical, and its use is not recommended." (CITE: Tagged PDF Best Practice Guide: Syntax from the PDF Association)*

Remaining Structure Types

- BlockQuote
- Code
- Div
- H (heading, strongly structured)*
- H1
- H2
- H3
- H4
- H5
- H6
- P
- Quote
- Span

*Due to a lack of suitable tools, the H structure element is impractical, and its use is not recommended.

NonStruct

From the PDF Association's *Tagged PDF Best Practice Guide*:

Has no substantive role or meaning; interpretation is out of scope for consumers of tagged PDF.
NonStruct's value is primarily as a utility for role mapping custom structure element types for which no corresponding standard structure element type is suitable.

A document containing custom structure elements for which there is no corresponding standard structure type, but where the content and child elements are intended to be real content. In such a case appropriate semantics require a writer to role map the custom element type to NonStruct.

NonStruct has no semantic significance, but its content and contained structure elements may well have significance. A consuming processor would be expected to pass these children to AT and other processors.

CITE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

NonStruct

From the PDF Association's *Tagged PDF Best Practice Guide*:

Has no substantive role or meaning; interpretation is out of scope for consumers of tagged PDF. NonStruct's value is primarily as a utility for role mapping custom structure element types for which no corresponding standard structure element type is suitable.

A document containing custom structure elements for which there is no corresponding standard structure type, but where the content and child elements are intended to be real content. In such a case appropriate semantics require a writer to role map the custom element type to NonStruct.

NonStruct has no semantic significance, but its content and contained structure elements may well have significance. A consuming processor would be expected to pass these children to AT and other processors.

CITE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Private

From the PDF Association's *Tagged PDF Best Practice Guide*:

Akin to <NonStruct>, this element differs from <NonStruct> insofar as not only the structure element itself, but *the children of the structure element, including structure elements and content, are also ignored*.

CITE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Private

From the PDF Association's *Tagged PDF Best Practice Guide*:

Akin to <NonStruct>, this element differs from <NonStruct> insofar as not only the structure element itself, but the children of the structure element, including structure elements and content, are also ignored.

CITE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Remaining Structure Types

- BlockQuote
- Code
- Div
- P
- Quote
- Span

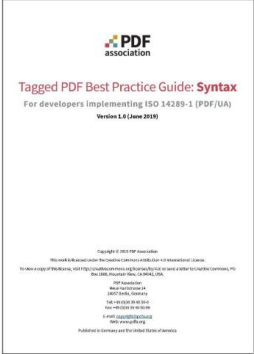
Bouton Jones 47

Remaining Structure Types

- BlockQuote
- Code
- Div
- H (heading, strongly structured)*
- H1
- H2
- H3
- H4
- H5
- H6
- P
- Quote
- Span

*Due to a lack of suitable tools, the H structure element is impractical, and its use is not recommended.

For more about Structure Types:



See the PDF Association's Tagged PDF *Best Practice Guide* at <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Bouton Jones 48

For more about Structure Types:

See the PDF Association's Tagged PDF Best Practice Guide at <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

IMAGE: First Page of the PDF Association's Tagged PDF Best Practice Guide .

Checkpoint 06: Metadata

Index	Failure Condition	Section
06-001	Document does not contain XMP metadata stream	UA1:7.1-8
06-002	The metadata stream in the Catalog dictionary does not include the PDF/UA identifier.	UA1:5
06-003	Metadata stream does not contain dc:title	UA1:7.1-8
06-004	dc:title does not clearly identify the document	UA1:7.1-8

We're going to cover this Checkpoint in two parts:

1. 06-001 and 06-002
2. 06-003 and 06-004

Checkpoint 06: Metadata

Index	Failure Condition	Section
06-001	Document does not contain XMP metadata stream	UA1:7.1-8
06-002	The metadata stream in the Catalog dictionary does not include the PDF/UA identifier.	UA1:5
06-003	Metadata stream does not contain dc:title	UA1:7.1-8
06-004	dc:title does not clearly identify the document	UA1:7.1-8

*Dublin Core (DC): a set of fifteen "core" elements (properties) for describing resources. This fifteen-element Dublin Core has been formally standardized as ISO 15836, ANSI/NISO Z39.85,[2] and IETF RFC 5013. The core properties are part of a larger set of DCMI Metadata Terms. "Dublin Core" is also used as an adjective for Dublin Core metadata, a style of metadata that draws on multiple RDF vocabularies, packaged and constrained in Dublin Core application profiles. (CITE: https://en.wikipedia.org/wiki/Dublin_Core)

Checkpoint 06: 001 & 002 XMP Metadata & PDF/UA Identifier

Index	Failure Condition	Section
06-001	Document does not contain <i>XMP metadata</i> stream	UA1:7.1-8
06-002	The metadata stream in the Catalog dictionary does not include the <i>PDF/UA identifier</i> .	UA1:5
06-003	Metadata stream does not contain dc:title	UA1:7.1-8
06-004	dc:title does not clearly identify the document	UA1:7.1-8

Bouton Jones 50

Checkpoint 06: Metadata

Index	Failure Condition	Section
06-001	Document does not contain XMP metadata stream	UA1:7.1-8
06-002	The metadata stream in the Catalog dictionary does not include the PDF/UA identifier.	UA1:5
06-003	Metadata stream does not contain dc:title	UA1:7.1-8
06-004	dc:title does not clearly identify the document	UA1:7.1-8

Checkpoint 06: 001 XMP



Failure Condition 06-001: Document does not contain *XMP metadata stream*.

“The Extensible Metadata Platform (XMP) provides Adobe applications with a common XML framework that standardizes the creation, processing, and interchange of document metadata across publishing workflows. You can save and import the document metadata XML source code in XMP format, making it easy to share metadata among different documents. You can also save document metadata to a metadata template that you can reuse in Acrobat.”

CITE: <https://helpx.adobe.com/acrobat/using/pdf-properties-metadata.html>

Checkpoint 06: Failure Condition 06-001

Failure Condition 06-001: Document does not contain XMP metadata stream

“The Extensible Metadata Platform (XMP) provides Adobe applications with a common XML framework that standardizes the creation, processing, and interchange of document metadata across publishing workflows. You can save and import the document metadata XML source code in XMP format, making it easy to share metadata among different documents. You can also save document metadata to a metadata template that you can reuse in Acrobat.”

CITE: <https://helpx.adobe.com/acrobat/using/pdf-properties-metadata.html>

Fixes for Missing XMP Data

Instructions to Add Metadata with an XMP File:

- Taggedpdf.com: <https://taggedpdf.com/508-pdf-help-center/pdfua-identifier-missing/>
- Accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

Fixes for Missing XMP Data

Instructions to Add Metadata with an XMP File:

- Taggedpdf.com: <https://taggedpdf.com/508-pdf-help-center/pdfua-identifier-missing/>
- Accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

Checkpoint 06: 002 PDF/UA Identifier

Failure Condition 06-002: The metadata stream in the Catalog dictionary does not include the *PDF/UA identifier*.

This is an issue for PDFs created in:

- Microsoft Word 2013 for Windows
- Any version of Adobe InDesign (formerly Adobe PageMaker)

Test with:

- Adobe Acrobat
- PAC 2021 (free testing tool from -----)
- axesPDF tool from axes4

Checkpoint 06: 002 PDF/UA Identifier

Failure Condition 06-002: The metadata stream in the Catalog dictionary does not include the PDF/UA identifier.

A missing PDF/UA Identifier can be an issue for PDFs created in:

- Microsoft Word 2013 for Windows
- Any version of Adobe InDesign (formerly Adobe PageMaker)

Test with:

- Adobe Acrobat
- PAC 3 (free testing tool from Access for all)

An example of a complete PDF/UA flag:

```
<?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 5.6-c015
91.163280, 2018/06/22-11:31:03 " >
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:pdfuaid="http://www.aiim.org/pdfua/ns/id/"
    <dc:title>
      <rdf:Alt>
        <rdf:li xml:lang="x-default">PDF/UA Document</rdf:li>
        <rdf:li xml:lang="en">PDF/UA Document</rdf:li>
      </rdf:Alt>
    </dc:title>
    <pdfuaid:part>1</pdfuaid:part>
  </rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?>
```

CITE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Bouton Jones 54

An example of a complete PDF/UA flag

IMAGE: XML code for a complete PDF/UA flag

---- BEGIN CODE ---

```
<?xpacket begin="" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 5.6-c015
91.163280, 2018/06/22-11:31:03 " >
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:pdfuaid="http://www.aiim.org/pdfua/ns/id/"
    <dc:title>
      <rdf:Alt>
        <rdf:li xml:lang="x-default">PDF/UA Document</rdf:li>
        <rdf:li xml:lang="en">PDF/UA Document</rdf:li>
      </rdf:Alt>
    </dc:title>
    <pdfuaid:part>1</pdfuaid:part>
  </rdf:Description>
</rdf:RDF>
</x:xmpmeta>
```

<?xpacket end="w"?>

---- END CODE ---

IMAGE and CODE SOURCE: <https://www.pdfa.org/wp-content/uploads/2019/06/TaggedPDFBestPracticeGuideSyntax.pdf>

Why include the PDF/UA Identifier?

- “This enables reader applications and assistive technologies to fully process PDF/UA conforming files to maximize accessibility.” (CITE: <https://accessiblewebsiteservices.com/accessible-pdfs-how-to-add-a-pdf-ua-identifier/>)
- “In order for a PDF document to be technically accessible, the [PDF/UA](#) identifier (metadata) must be set.” (CITE: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>)

So, why is that?

Bouton Jones 55

Why must we include the PDF/UA Identifier?

“This enables reader applications and assistive technologies to fully process PDF/UA conforming files to maximize accessibility.” (CITE: <https://accessiblewebsiteservices.com/accessible-pdfs-how-to-add-a-pdf-ua-identifier/>)

“In order for a PDF document to be technically accessible, the PDF/UA identifier (metadata) must be set.” (CITE: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>)

So, why is that?

The PDF Association on the PDF/UA Identifier

From an email from info@pdfa.org:

The PDF/UA flag is a quality indicator. It prompts AT to expect high-quality tagging and other structures required for accessibility, as well as alt. text, validated disambiguation of real content and artifacts, validated reading-order and so on.

But how does including the PDF/UA identifier improve the experience of a PWD?

Bouton Jones

56

The PDF Association on the PDF/UA Identifier

From an email from info@pdfa.org:

The PDF/UA flag is a quality indicator. It prompts AT to expect high-quality tagging and other structures required for accessibility, as well as alt. text, validated disambiguation of real content and artifacts, validated reading-order and so on.

But, how does including the PDF/UA identifier improve the experience of a PWD?

IMHO

You shouldn't fix the PFD/UA identifier unless/until:

1. your document meets all other PDF/UA compliance criteria
2. your Accessibility policy explicitly requires full PDF/UA compliance.

Solutions for PDF/UA identifier missing (1 of 3)

1. Adobe Acrobat DC

- A. Open the PDF document in Adobe Acrobat DC.
- B. Select More Tools > Print Production > Preflight
- C. In the Preflight dialog, be sure PDF Standards is selected from the dropdown.
- D. Activate the wrench icon above the list of Preflight checks. This should open the Document info and Metadata settings.
- E. Select the Set PDF/UA-1 entry option, then click the Fix button.
- F. In the next dialog box, you will need to re-save your file with a different name.

CITE: <https://accessiblewebsiteservices.com/accessible-pdfs-how-to-add-a-pdf-ua-identifier/>

Solutions for PDF/UA identifier missing (1 of 3)

- Adobe Acrobat DC
 - Open the PDF document in Adobe Acrobat DC.
 - Select More Tools > Print Production > Preflight
 - In the Preflight dialog, be sure PDF Standards is selected from the dropdown.
 - Activate the wrench icon above the list of Preflight checks. This should open the Document info and Metadata settings.
 - Select the Set PDF/UA-1 entry option, then click the Fix button.
 - In the next dialog box, you will need to re-save your file with a different name.

CITE: <https://accessiblewebsiteservices.com/accessible-pdfs-how-to-add-a-pdf-ua-identifier/>

Solutions for PDF/UA identifier missing (2 & 3)

2. Add Metadata with an XMP File

- Instructions at taggedpdf.com: <https://taggedpdf.com/508-pdf-help-center/pdfua-identifier-missing/>
- Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

3. In **InDesign** (using the **MadeToTag** plugin from axaio)

- Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>
- About MadeToTag: <https://www.axaio.com/doku.php/en:products:madetotag>

Solutions for PDF/UA identifier missing (2 & 3)

2. Add Metadata with an XMP File

- Instructions at taggedpdf.com: <https://taggedpdf.com/508-pdf-help-center/pdfua-identifier-missing/>
- Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

3. Automatic approach in InDesign (MadeToTag)

Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

Solutions for PDF/UA identifier missing (4)

4. Use the **axesPDF** tool from axes4 (<https://www.axes4.com/>) to add the identifier to compliant documents.

Solutions for PDF/UA identifier missing (2 & 3)

2. Add Metadata with an XMP File

- Instructions at taggedpdf.com: <https://taggedpdf.com/508-pdf-help-center/pdfua-identifier-missing/>
- Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

3. Automatic approach in InDesign (MadeToTag)

Instructions at accessible-pdf.info: <https://accessible-pdf.info/tutorials/general/pdf-ua-identifier-is-missing>

It's worth repeating ...

A missing PDF/UA Identifier can be an issue for PDFs created in:

- Microsoft Word 2013 for Windows
- Any version of Adobe InDesign (formerly Adobe PageMaker)

So, it might not be relevant to you.

It's worth repeating ...

A missing PDF/UA Identifier can be an issue for PDFs created in:


- Microsoft Word 2013 for Windows
- Any version of Adobe InDesign (formerly Adobe PageMaker)

So, what does that mean for you?

If you don't use these programs or don't maintain legacy documents created with these programs, this is probably not an issue for you.

Checkpoint 06: 003 & 004 DC:Title

Index	Failure Condition	Section
06-001	Document does not contain XMP metadata stream	UA1:7.1-8
06-002	The metadata stream in the Catalog dictionary does not include the PDF/UA identifier.	UA1:5
06-003	Metadata stream does not contain dc:title*	UA1:7.1-8
06-004	dc:title does not clearly identify the document	UA1:7.1-8



***Dublin Core (DC):** a set of fifteen "core" elements (properties) for describing resources. This fifteen-element Dublin Core has been formally standardized as ISO 15836, ANSI/NISO Z39.85,[2] and IETF RFC 5013. The core properties are part of a larger set of DCMI Metadata Terms. (CITE: https://en.wikipedia.org/wiki/Dublin_Core)

Bouton Jones
62

Checkpoint 06: 003 & 004 (Dublin Core) Title

06-003: Metadata stream does not contain dc:title

06-004: dc:title does not clearly identify the document

*Dublin Core (DC): a set of fifteen "core" elements (properties) for describing resources. This fifteen-element Dublin Core has been formally standardized as ISO 15836, ANSI/NISO Z39.85,[2] and IETF RFC 5013. The core properties are part of a larger set of DCMI Metadata Terms. "Dublin Core" is also used as an adjective for Dublin Core metadata, a style of metadata that draws on multiple RDF vocabularies, packaged and constrained in Dublin Core application profiles. (CITE: https://en.wikipedia.org/wiki/Dublin_Core)

In HTML, the Title Tag is in the *HEAD* Section

```
1 <html>
2 <head>
3   <title>Heading Levels in HTML</title>
4   <link rel="stylesheet" type="text/css"
5     href="styles.css" />
6 </head>
7 <body lang="EN-US">
8
9 <p class=MsoTitle>MS Word Title</p>
10 <p class=MsoSubtitle>MS Subtitle</p>
11 <h1>MS Word Heading Level One</h1>
12 <h2>MS Word Heading Level Two</h2>
13 <h3>MS Word Heading Level Three</h3>
14 <h4>MS Word Heading Level Four</h4>
15 <h5>MS Word Heading Level Five</h5>
16 <h6>MS Word Heading Level Six</h6>
17 <p class=MsoHeading7>MS Word Heading Level Seven</p>
18 <p class=MsoHeading8>MS Word Heading Level Eight</p>
19 <p class=MsoHeading9>MS Word Heading Level Nine</p>
20
21 </body>
22 </html>
```

Bouton Jones

63

In HTML, the Title Tag is in the HEAD Section

IMAGE 1 of2: HTML Markup

---- Begin Markup ----

```
<html>
<head>
  <title>Heading Levels in HTML</title>
  <link rel="stylesheet" type="text/css" href="styles.css" />
</head>
<body lang="EN-US">
```

```
<p class=MsoTitle>MS Word Title</p>
<p class=MsoSubtitle>MS Subtitle</p>
<h1>MS Word Heading Level One</h1>
<h2>MS Word Heading Level Two</h2>
<h3>MS Word Heading Level Three</h3>
<h4>MS Word Heading Level Four</h4>
<h5>MS Word Heading Level Five</h5>
<h6>MS Word Heading Level Six</h6>
<p class=MsoHeading7>MS Word Heading Level Seven</p>
<p class=MsoHeading8>MS Word Heading Level Eight</p>
```

```
<p class=MsoHeading9>MS Word Heading Level Nine</p>
```

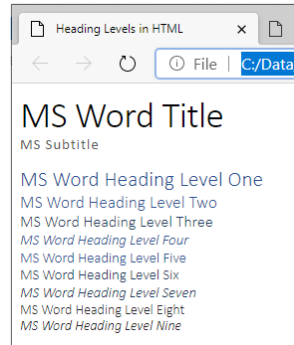
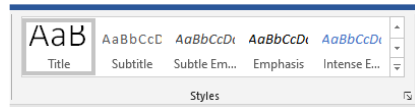
```
</body>
```

```
</html>
```

```
---- End Markup ----
```

IMAGE 2 of 2: Browser Tab displays the page title as being “Heading Levels in HTML”

In Word, users add the Title Tag into the *body* of the document.



In Word, users add the Title Tag into the body of the document.

Question: How Does Word's **Title** Tag Export to PDF?

1. As the PDF's Title *Property* (in the META data?)
2. Or as the top-most heading level?


Q: How Does Word's Title Tag Export to PDF?

Does it export as the PDF's Title Property (not to be confused with the filename?)

Does it export as the top-most heading level?

What the Heck!?

It does *not* export to the PDF title in the Meta Data (as I had assumed until recently.)



The screenshot shows a 'Document Properties' dialog box with a 'Description' tab selected. The 'File' field contains 'unfriendly_file_name_2021_08_11_experiment_num_1A.pdf'. The 'Title' field is empty and highlighted with a red rectangle. The 'Author' field contains 'Jones, Bouton'. The 'Subject' and 'Keywords' fields are empty.

Another Surprise

Nor does Word's Title export to the PDF's top-most heading level. It exports to a *paragraph* tag.

This is the Title

This is Heading Level Two

This is Also Heading Level Two

Navigation

Headings Pages Results

This is Heading Level Two

This is Also Heading Level Two

<P>

This is the Title

<H2>


This is Heading Level Two


Bouton Jones 67

Another Surprise

Nor does Word's Title export to the PDF's top-most heading level. It exports to a *paragraph* tag.

Two Options for the Title Paragraph

1. Change the Paragraph tag into an H1 tag.


The diagram shows two side-by-side boxes. The left box contains a paragraph icon, a dropdown arrow, and the text '<P>' above the text 'This is the Title'. The right box contains an H1 icon, a dropdown arrow, and the text '<H1>' above the text 'This is the Title'.
2. Remove the Paragraph tag and the title text inside it.


The diagram shows two side-by-side boxes. The left box contains a paragraph icon, a dropdown arrow, and the text '<P>' above the text 'This is the Title'. Below this is an H2 icon, a dropdown arrow, and the text '<H2>' above the text 'This is Heading Level Two'. The right box contains an H1 icon, a dropdown arrow, and the text '<H1>' above the text 'This is Heading Level Two'.

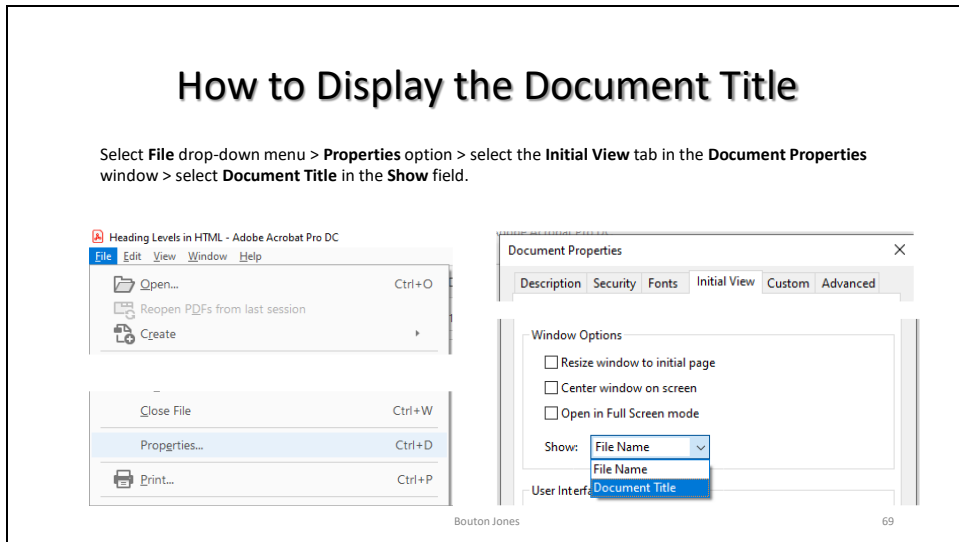
Bouton Jones 68

Two Options for the Title Paragraph

- Change the Paragraph tag into an H1 tag.
- Remove the Paragraph tag and the title text inside it.

How to Display the Document Title

Select **File** drop-down menu > **Properties** option > select the **Initial View** tab in the **Document Properties** window > select **Document Title** in the **Show** field.



How to Display the Document Title

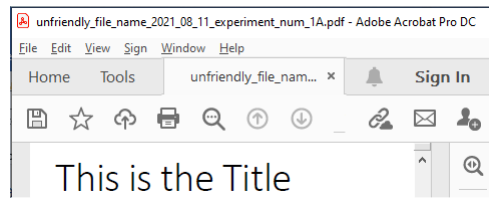
Select **File** drop-down menu > **Properties** option > select the **Initial View** tab in the **Document Properties** window > select **Document Title** in the **Show** field.

Example: Displaying the File Name

By default, the Title is empty and the PDF is set to display the File Name instead of the Title

Title:

Show:



Bouton Jones

70

Example: Displaying the File Name

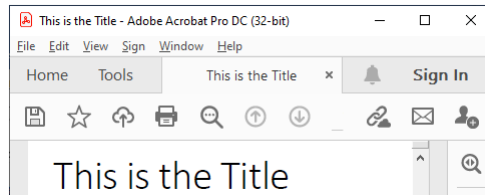
By default, the Title is empty and the PDF is set to display the File Name instead of the Title

Example: Displaying the Title

You must manually enter the Title and set Show to display the Document Title.

Title:

Show:




Example: Displaying the Title

You must manually enter the Title and set Show to display the Document Title.

Checkpoint 08: OCR

Index	Failure Condition	Section
08-001	OCR-generated text contains significant errors.	UA1:7.1-10
08-002	OCR-generated text is not tagged	UA1:7.1-10



Bouton Jones 72

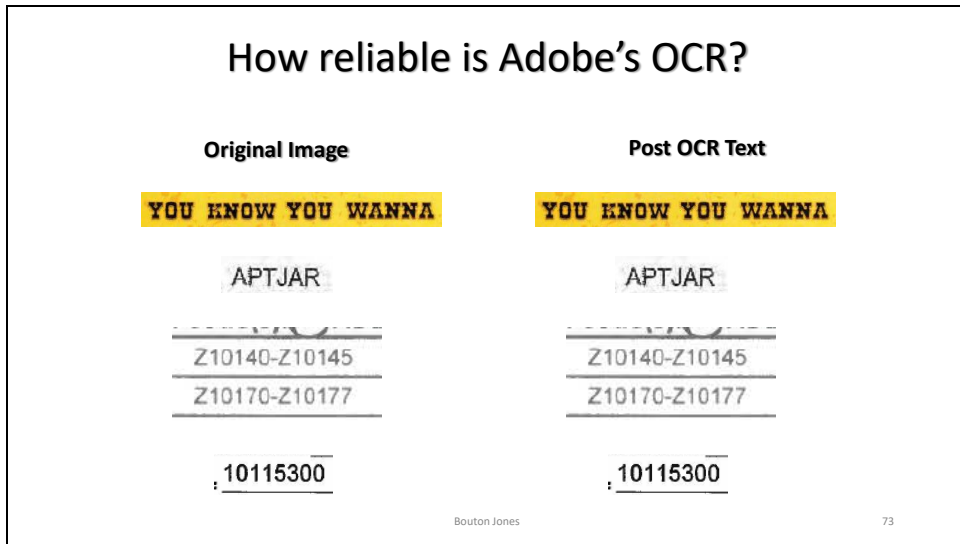
Checkpoint 08: OCR Validation

- OCR-generated text contains significant errors.
- OCR-generated text is not tagged

IMAGE: The acronym O. C. R. being viewed under a magnifying glass.

Optical character recognition or optical character reader is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposed on an image.

CITE: Wikipedia



How Reliable is OCR?

Table with two columns and four rows

Column One: Original Image

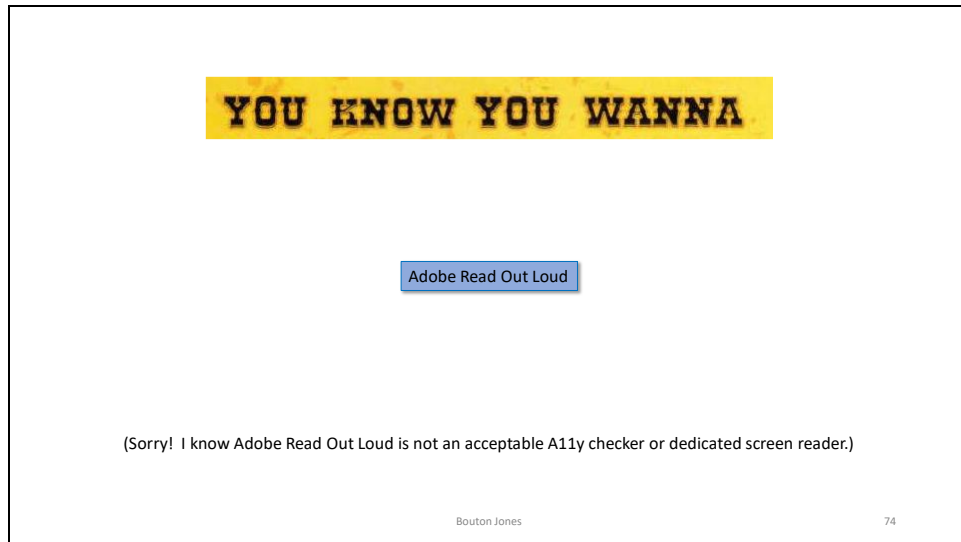
Column Two: Post OCR Text

The screen captures in column one and column two appear identical.

After performing OCR, usually the resulting renderable text looks exactly like the scanned image of text made with a decorative font or Word Art (e.g. "You Know You Wanna.") It's visually indistinguishable to the original to me. But now I can select the new text to copy or edit.

Because electronic signatures were not an option for the last three examples, the documents were filled out online, printed, signed, and scanned.

Slide 74



How would a screen reader identify the following text?

IMAGE: Stylized text in a decorative font – “You Know You Wanna”

How would Adobe’s Read Out Loud identify the Post OCR text above?

As “You Know You Wanna?”

Or something completely incomprehensible?

AUDIO FILE: Adobe Read Out Loud function saying “YDV H!fDW YOU WAH!fA?”

Sorry. I know Adobe Read Out Loud is not acceptable an A11y checker or dedicated screen reader.

How reliable is Adobe's OCR (really?)

Post OCR Text	Post OCR Text Copied and Pasted
YOU KNOW YOU WANNA	YDV H!dW YOU WAHIfA
APTJAR	Af>T JAR
Z10140-Z10145	2 10140- 2 10145
Z10170-Z10177	Z10170-Z10177
:10115300	_1_0_1_1_5_3_0_0

Bouton Jones 75

How Reliable is OCR (really?)

Table with two columns and four rows

Column One: Post OCR Text

Column Two: Post OCR Text

Often the Post OCR text can appear exactly the same as the original text in graphic but in reality, the characters in the new text will not correspond with the original text.

For example: The text in graphic might have been meant to be “You know you wanna” but when converted in OCR to actual text the characters are “YDV H!dW YOU WAHIfA.” And to a sighted person they may appear to be “You know you wanna” but a screen reader will identify it as “YDV H!dW YOU WAHIfA.”

POLL QUESTION: What does the “Recognize Text” (OCR) function do?

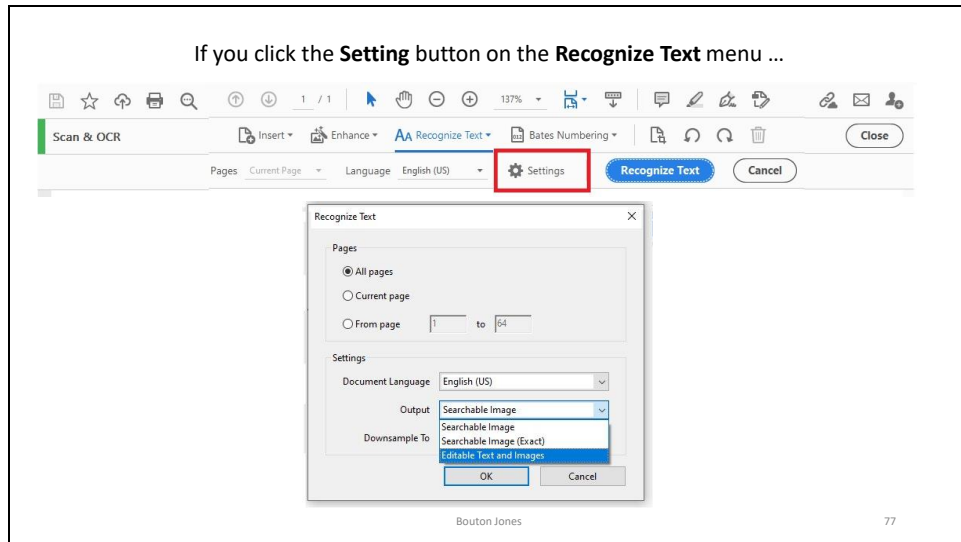
- A. Does it convert the image of text in the image-only PDF into true text that can be edited like a Word document?
- B. Does Acrobat place an invisible layer of "renderable text" --- i.e. vector (computer generated) text --- on top of the image layer?

What does the “Recognize Text” (OCR) function do?

Does it convert the image of text in the image only PDF into true text that can be edited like a Word document?

Or does Acrobat place an invisible layer of "renderable text" --- i.e. vector (computer generated) text --- on top of the image layer?

Slide 77



Output Options for the Recognize Text (OCR) Function in Adobe Acrobat 2017

If you click the Setting Button on the Recognize Text menu, you can choose among three options in the Recognize Text dialog box.

- Searchable Image
- Searchable Image (exact)
- Editable Text and Image

Editable Text and Image	Searchable Image and Searchable Image (Exact)
Converts the graphical text into vector text. The vector text will replace the “text in image.”	Places an invisible “layer” (not <i>really</i> a layer) on top of the original image.
This setting creates and embeds a custom font that looks like the original font in the image.	This setting does not alter the text in the image or embed a custom font.
You can now edit the text.	You can’t easily edit the document, but it can be read by assistive technology (such as JAWS and NVDA.)
<small>Bouton Jones</small>	<small>78</small>

Two Output Options:

1. Editable Text and Image:

Converts the graphical text into vector text. The vector text will replace the “text in image.”

It creates and embeds a custom font that looks like the original font in the image.

You can now edit the text.

The file size is larger

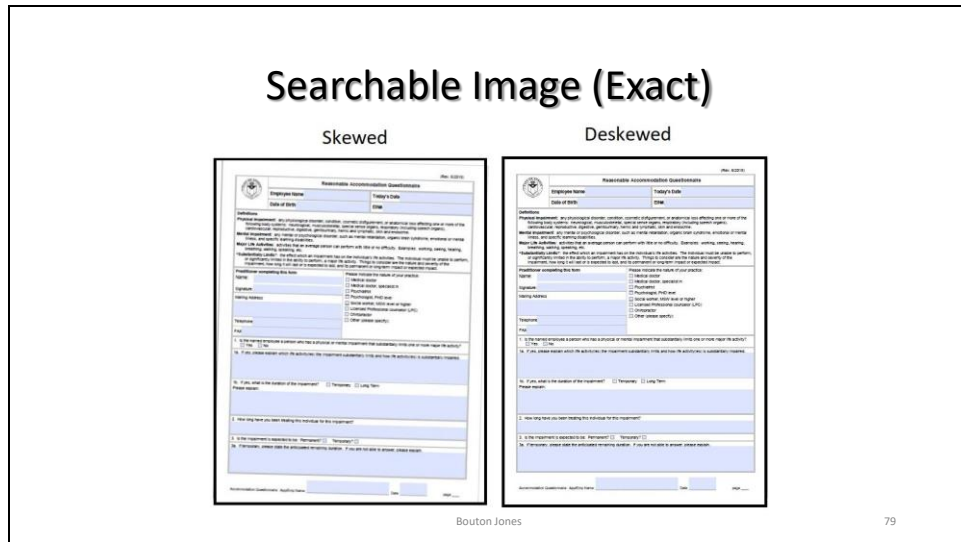
2. Searchable Image

Places an invisible “layer” (not really a layer) on top of the original image.

You can’t easily edit the document but it can be read by assistive technology (such as JAWS and NVDA.)

Additional Resources on the Output Options

- Understanding OCR options in Adobe Acrobat: “Searchable Image”, “Searchable Image (Exact)”, and “Editable Text and Images” (Superuser) - <https://superuser.com/questions/1276594/understanding-ocr-options-in-adobe-acrobat-searchable-image-searchable-imag>
- Searchable Image vs. Searchable Image (Exact) - Quality of OCR (Adobe Community) - <https://community.adobe.com/t5/acrobat/searchable-image-vs-searchable-image-exact-quality-of-ocr/td-p/4734730?page=1>



Searchable Image (Exact)

Deskew: The process of straightening an image that has been scanned or photographed crookedly — that is an image that is slanting too far in one direction, or one that is misaligned. (Vangie Beal on webopedia)

CITE:

https://www.webopedia.com/TERM/I/image_de_skew.html#:~:text=By%20Vangie%20Beal%20The%20process,Also%20called%20deskew.%5C

IMAGES: An original skewed page and a deskewed version side by side

The Searchable Image setting will deskew the image before OCR.

The Searchable Image (Exact) setting will NOT deskew the image before OCR.

Often deskewing the image will lead to a smaller file size.

But there may be a reason that you don't want to alter the document. (Ask a lawyer)

RESEARCH: Will the editable text setting deskew the image?

Checkpoint 09: Appropriate Tags

Index	Failure Condition	Section
09-001	Tags are not in logical reading order.	UA1:7.2-1
09-002	Structure elements are nested in a semantically inappropriate manner. (e.g., a table inside a heading).	UA1:7.2-1
09-003	The structure type (after applying any role-mapping as necessary) of a structure element is not semantically appropriate.	UA1:7.2-1
09-004	A table-related structure element is used in a way that does not conform to the syntax defined in ISO 32000-1, Table 337.	UA1:7.2-1
09-005	A list-related structure element is used in a way that does not conform to Table 336 in ISO 32000-1.	UA1:7.2-1
09-006	A TOC-related structure element is used in a way that does not conform to Table 333 in ISO 32000-1.	UA1:7.2-1
09-007	A Ruby-related structure element is used in a way that does not conform to Table 338 in ISO 32000-1.	UA1:7.2-1
09-008	A Warichu-related structure element is used in a way that does not conform to Table 338 in ISO 32000-1.	UA1:7.2-1

Bouton Jones

80

Checkpoint 09: Appropriate Tags

09-001: Tags are not in logical reading order.

09-002: Structure elements are nested in a semantically inappropriate manner. (e.g., a table inside a heading).

09-003: The structure type (after applying any role-mapping as necessary) of a structure element is not semantically appropriate.

09-004: A table-related structure element is used in a way that does not conform to the syntax defined in ISO 32000-1, Table 337.

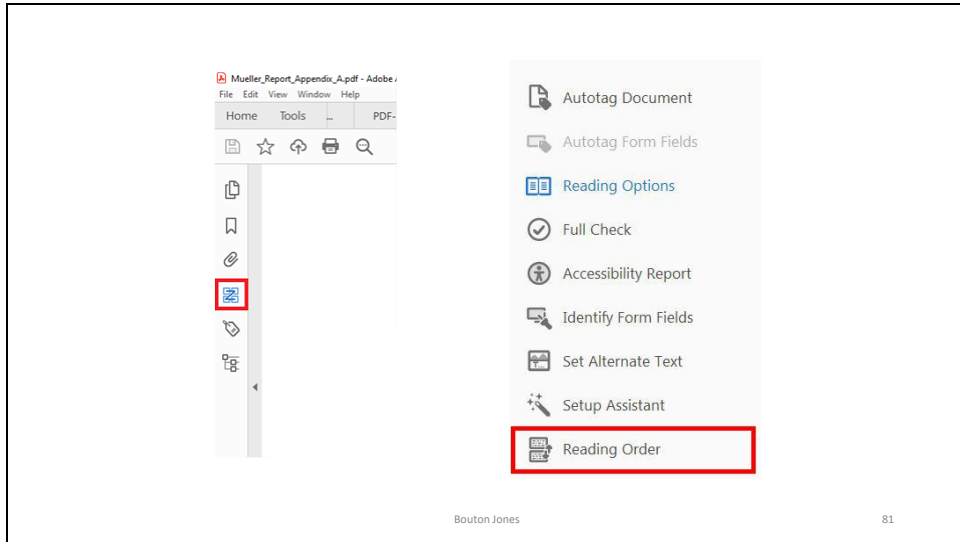
09-005: A list-related structure element is used in a way that does not conform to Table 336 in ISO 32000-1.

09-006: A TOC-related structure element is used in a way that does not conform to Table 333 in ISO 32000-1.

09-007: A Ruby-related structure element is used in a way that does not conform to Table 338 in ISO 32000-1.

09-008: A Warichu-related structure element is used in a way that does not conform to Table 338 in ISO 32000-1.

Slide 81



Reading Order

Checkpoint 14: Headings

Index	Failure Condition	Section
14-001	Headings are not tagged.	UA1:7.4-1
14-002	Does use numbered headings, but the first heading tag is not H1.	UA1:7.4.2-1
14-003	Numbered heading levels in descending sequence are skipped (Example: H3 follows directly after H1).	UA1:7.4-1
14-004	Numbered heading tags do not use Arabic numerals.	UA1:7.4.3-1
14-005	Content representing a 7 th level (or higher) heading does not use an “H7” (or higher) tag.	UA1:7.4.3-1
14-006	A node contains more than one H tag.	UA1:7.4.4-1
14-007	Document uses both H and H# tags.	UA1:7.4.4-3
<p>NOTE: In weakly-structured documents, headings always take the form “Hn” (e.g., H1, H2, Hn) without intervening whitespace or numerical separators.</p>		

Checkpoint 14: Headings

Index

Failure Condition

Section

14-001

Headings are not tagged.

UA1:7.4-1

14-002

Does use numbered headings, but the first heading tag is not H1.

UA1:7.4.2-1

14-003

Numbered heading levels in descending sequence are skipped (Example: H3 follows directly after H1).

UA1:7.4-1

14-004

Numbered heading tags do not use Arabic numerals.

UA1:7.4.3-1

14-005

Content representing a 7th level (or higher) heading does not use an “H7” (or higher) tag.

UA1:7.4.3-1

14-006

A node contains more than one H tag.

UA1:7.4.4-1

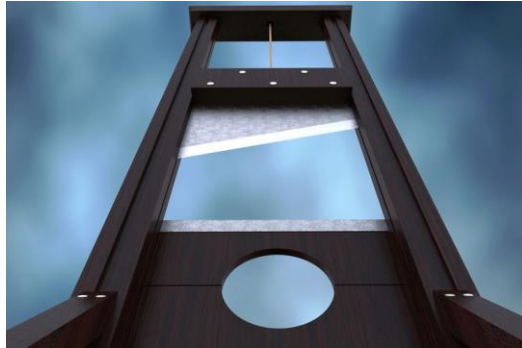
14-007

Document uses both H and H# tags.

NOTE: In weakly-structured documents, headings always take the form “Hn” (e.g., H1, H2, Hn) without intervening whitespace or numerical separators.

UA1:7.4.4-3

Why would you lose headings when exporting from Word to PDF?



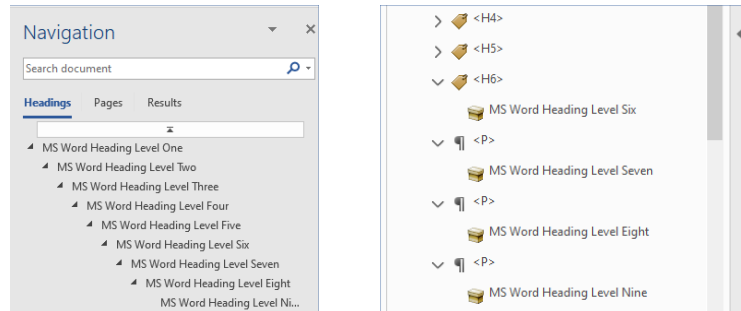
Bouton Jones

83

Why would you lose headings when exporting from Word to PDF?

IMAGE: guillotine. an apparatus designed for efficiently carrying out executions by beheading. The device consists of a tall, upright frame with a weighted and angled blade suspended at the top. The condemned person is secured with stocks at the bottom of the frame, positioning the neck directly below the blade.

Exporting Heading Levels from Word to PDF



Try to stop at 6.

Exporting Heading Levels from Word to PDF

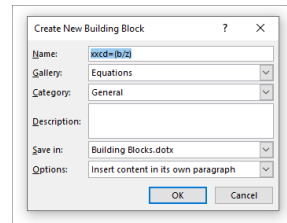
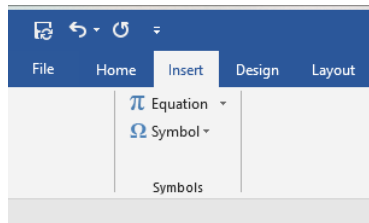
First screen capture: MS Word document showing 9 levels of headings

Second screen capture: PDF tag panel showing the tag panel for a PDF that was exported from the previously mentioned MS Word document

The headings at level 7, 8, and 9 have exported as paragraphs in a PDF.

Checkpoint 17: Mathematical Expressions

Index	Failure Condition	Section
17-001	Content is a mathematical expression but is not tagged with a Formula tag.	UA1:7.7-1
17-002	Formula tag is missing an Alt attribute.	UA1:7.7-1
17-003	Unicode mapping requirements are not met.	UA1:7.7-2



Bouton Jones

85

Checkpoint 17: Mathematical Expressions

17-001: Content is a mathematical expression but is not tagged with a Formula tag.

17-002: Formula tag is missing an Alt attribute.

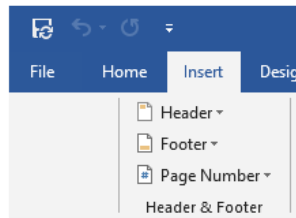
17-003: Unicode mapping requirements are not met.

IMAGES: Two Screen captures showing

- the Insert tab selected in the Word ribbon. The Symbols section shows the Equation and Symbol functions
- The Create New Building Block window

Checkpoint 18: Page Headers and Footers

Index	Failure Condition	Section
18-001	Headers and footers are not marked as pagination artifacts.	UA1:7.8-1
18-002	Header or footer artifacts are not classified as Header or Footer subtypes.	UA1:7.8-1



*These three semantic tags are listed under the **Insert** tab in Word.*

Bouton Jones

86

Checkpoint 18: Page Headers and Footers

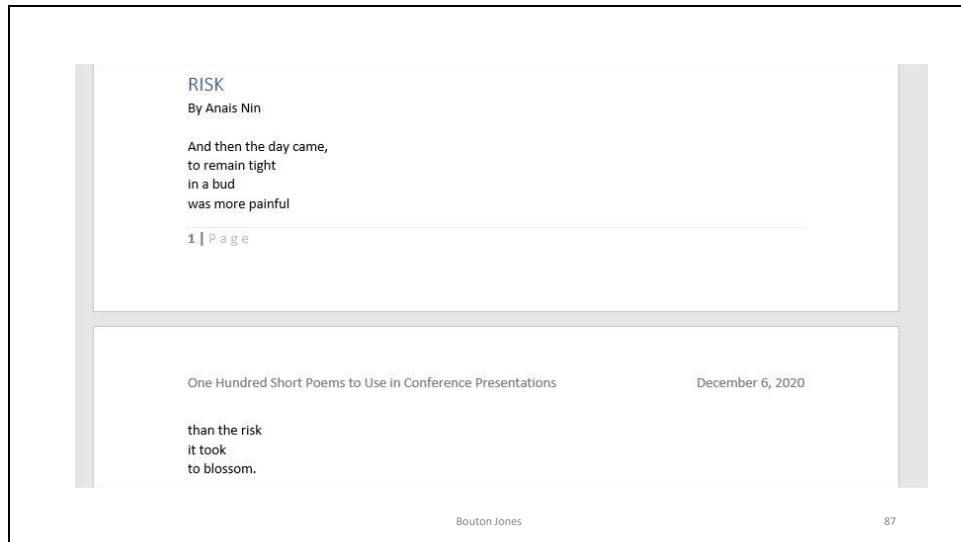
18-001: Headers and footers are not marked as pagination artifacts.

18-002: Header or footer artifacts are not classified as Header or Footer subtypes.

IMAGE: These three semantic tags (Header, Footer, and Page Number) are listed under the Insert tab in Word.

This is specific to Word. It doesn't apply to HTML.

Slide 87



Will AT read the headers and footers?

IMAGE: Screen capture of PDF showing the header and footer of a document

If the PDF was exported from a properly tagged Word document, screen readers should not read the headers and footers. But you should confirm this before and after properly exporting the source document to PDF.



So, what if the information in the header and footer is important?

If the information in the header and footer is important to the understanding of the content, you must repeat that information in the body of the document to ensure an equitable experience for the screen reader users.

SCREEN CAPTURE: First page of a document repeating information from the header --- i.e. the document title and the published date --- in the body of the document (specifically on the first page.)

Checkpoint 22: Article Threads

22-001: Article threads do not reflect logical reading order.

In PDFs, articles are optional electronic threads that the PDF author defines within that PDF. Articles lead readers through the PDF content, jumping over pages or areas of the page that are not included in the article. It is the same way that you skim through a traditional newspaper or magazine, following one specific story and ignoring the rest. When you read an article, the page view zooms in or out so that the current part of the article fills the screen.

CITE: https://helpx.adobe.com/acrobat/using/navigating-pdf-pages.html#article_threads

Checkpoint 22: Article Threads

22-001: Article threads do not reflect logical reading order.

"In PDFs, articles are optional electronic threads that the PDF author defines within that PDF. Articles lead readers through the PDF content, jumping over pages or areas of the page that are not included in the article. It is the same way that you skim through a traditional newspaper or magazine, following one specific story and ignoring the rest. When you read an article, the page view zooms in or out so that the current part of the article fills the screen." (CITE: https://helpx.adobe.com/acrobat/using/navigating-pdf-pages.html#article_threads)

Defining Article Threads

"You create an article by defining a series of boxes around the content in the order in which you want the content read. The navigational path you define for an article is known as the article thread. You create a thread connecting the various boxes, unifying them into a continuous text flow.

....

"Most desktop publishing programs allow you to generate article threads automatically as you convert the files to Adobe PDF. If the file you're viewing has articles, you can show the names of the articles on a tab and navigate easily through them."

CITE: <https://helpx.adobe.com/acrobat/using/pdf-articles.html>

Defining Article Threads

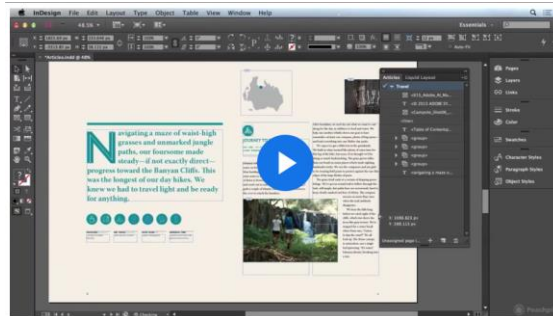
"You create an article by defining a series of boxes around the content in the order in which you want the content read. The navigational path you define for an article is known as the article thread. You create a thread connecting the various boxes, unifying them into a continuous text flow.

"Note:

"Most desktop publishing programs allow you to generate article threads automatically as you convert the files to Adobe PDF. If the file you're viewing has articles, you can show the names of the articles on a tab and navigate easily through them."

CITE: <https://helpx.adobe.com/acrobat/using/pdf-articles.html>

Adobe InDesign Video: "The Articles panel"



URL: <https://helpx.adobe.com/indesign/how-to/indesign-articles-panel.html>



Bouton Jones

91

Adobe InDesign Video: "The Articles panel"

URL: <https://helpx.adobe.com/indesign/how-to/indesign-articles-panel.html>

Checkpoint 23: Digital Signatures

<h3 style="text-align: center;">Electronic Signature</h3> <p>An electronic symbol attached to a contract or other record, used by a person with an intent to sign.</p> <p>It can be a scanned image of a signature. Or a text field or text fields.</p> 	<h3 style="text-align: center;">Digital Signature</h3> <p>A cryptographically secure electronic signature that guarantees an electronic document is authentic. They use a certificate-based digital ID issued by an accredited Certificate Authority (CA) or Trust Service Provider (TSP.)</p> 
---	---

Digital Signatures are a type of Electronic Signature. All Digital Signatures are Electronic Signatures but not all Electronic Signatures are digital.

Bouton Jones 92

Checkpoint 23: Digital Signatures

Types of eSignatures

Electronic Signatures: An electronic symbol attached to a contract or other record, used by a person with an intent to sign.

Digital Signature: A cryptographically secure electronic signature that guarantees an electronic document is authentic. They use a certificate-based digital ID issued by an accredited Certificate Authority (CA) or Trust Service Provider (TSP.)

Digital Signatures are a type of Electronic Signature. All Digital Signatures are Electronic Signatures but not all Electronic Signatures are digital.

Testing your Digital Signature Fields

"No test specific to digital signatures is required, however other provisions apply (form fields)."

Digital signatures are form fields with a lot of functionality.

I suggest testing digital signatures "manually" with screen readers and keyboards (not mice.)

Bouton Jones 93

Testing your Digital Signature Fields

"No test specific to digital signatures is required, however other provisions apply (form fields)."

Digital signatures are form fields with a lot of functionality.

I suggest testing digital signatures "manually" with screen readers and keyboards (but not mice.) Can you read the document (including the digital signature field?) Can you apply the digital signature using your screen reader and keyboard (but not your mouse?) Can you confirm that you successfully sign the document using a screen reader and a mouse?

Checkpoint 25: XFA

"XFA (also known as XFA forms) stands for XML Forms Architecture, a family of proprietary XML specifications ... to enhance the processing of web forms. The XML Forms Architecture ... has been **deprecated in PDF 2.0**."

"XFA forms are saved internally in PDF files or as XDP (XML Data Package) files which can be opened in **Adobe's LiveCycle Designer** software. An XDP can package a PDF file, along with XML form and template data. XDP provides a mechanism for packaging form components within a surrounding XML container."

CITE: <https://en.wikipedia.org/wiki/XFA>

Checkpoint 25: XFA

So What are XFA Forms?

"XFA (also known as XFA forms) stands for XML Forms Architecture, a family of proprietary XML specifications that was suggested and developed by JetForm to enhance the processing of web forms. It can be also used in PDF files starting with the PDF 1.5 specification. The XFA specification is referenced as an external specification necessary for full application of the ISO 32000-1 specification (PDF 1.7). The XML Forms Architecture was not standardized as an ISO standard,[3] and has been deprecated in PDF 2.0." <https://en.wikipedia.org/wiki/XFA>

So, if you do *not* use Adobe LiveCycle ...



Bouton Jones

So, if you don't use Adobe LiveCycle ...

XFA forms shouldn't be a problem.

IMAGE: Man bent backward to dodge bullets fired at him.

Checkpoint 30: XObjects

Index	Failure Condition	Section
30-001	A reference XObject is present.	UA1:7.20-1
30-002	Form XObject contains MCIDs and is referenced more than once.	UA1:7.20-2

Form XObjects: a way of describing objects (text, images, vector elements,...) within a PDF file. It is meant for repetitive objects that only get stored once in a PDF but referenced multiple times. In a way, Form XObjects are similar to a mini-PDF embedded inside the main PDF document ... (CITE: <https://www.prepressure.com/pdf/basics/form-xobjects>)

Reference XObject: a form XObject which contains an optional Ref sub-dictionary. The sub-dictionary specifies a page in some target PDF file. (CITE: <https://www.adobe.com/devnet-docs/acrobatetk/tools/PrefRef/Windows/AVDisplay.html>)

SEE: <https://www.prepressure.com/pdf/basics/form-xobjects>

Checkpoint 30: XObjects

- 30-001** A reference XObject is present. UA1:7.20-1
30-002 Form XObject contains MCIDs and is referenced more than once.
 UA1:7.20-2

Form XObjects: a way of describing objects (text, images, vector elements,...) within a PDF file. It is meant for repetitive objects that only get stored once in a PDF but referenced multiple times. In a way, Form XObjects are similar to a mini-PDF embedded inside the main PDF document ... (CITE: <https://www.prepressure.com/pdf/basics/form-xobjects>)

Reference XObject: a form XObject which contains an optional Ref sub-dictionary. The sub-dictionary specifies a page in some target PDF file. (CITE: <https://www.adobe.com/devnet-docs/acrobatetk/tools/PrefRef/Windows/AVDisplay.html>)

SEE: <https://www.prepressure.com/pdf/basics/form-xobjects>

Checkpoint 31: Fonts

```
38 HTML,  
39 BODY {  
40     font-family: Calibri, "Trebuchet MS", Arial, Georgia, "Times New Roman", Helvetica, Serif;  
41     font-size: 100%;  
42     line-height: 1.125em;  
43     margin-top: 0px;  
44     margin-bottom: 0px;  
45 }
```

Businesses continue to be at great risk of a legal action regarding the accessibility of digital technologies, including websites, mobile applications, and kiosks. The number of demand letters and lawsuits filed on this basis has continued unabated for quite a few

Bouton Jones 97

Checkpoint 31: Fonts

In HTML and CSS, you can choose alternative fonts

--- Begin CSS ---

```
HTML,  
BODY {  
    font-family: Calibri, "Trebuchet MS", Arial, Georgia, "Times New Roman", Helvetica, Serif;  
    font-size: 100%;  
    line-height: 1.125em;  
    margin-top: 0px;  
    margin-bottom: 0px;  
}
```

--- End CSS ---

If the Calibri is not installed on the user's computer, the browser will use Trebuchet MS. If Trebuchet MS is not installed, the browser will use Arial. And so on and so forth.

But in Word you can only select one font to use at a time.

If You are Missing an Embedded Font:



When You Are Missing an Embedded Font

Adobe Acrobat's Base 14 Fonts

<p>Monospaced Fonts</p> <ul style="list-style-type: none">• Courier New• Courier New Bold• <i>Courier New Oblique</i>• <i>Courier New Bold Oblique</i>	<p>Proportional Fonts (Sans Serif)</p> <ul style="list-style-type: none">• Helvetica• Helvetica Bold• Helvetica Oblique• Helvetica Bold Oblique
<p>Proportional Fonts (Serif)</p> <ul style="list-style-type: none">• Times New Roman• Times New Roman Bold• <i>Times New Roman Italic</i>• <i>Times New Roman Bold Italic</i>	<p>Remaining Fonts</p> <ul style="list-style-type: none">• Symbol (e.g., Σμμβολ.)• <u>ZapfDingbats</u>

Bouton Jones 99

Adobe Acrobat's Base 14 Fonts

The Base 14 Fonts are defined in ISO 32000-1:2008(E) in Section 9.6.2.2.

They are also referred to as the "Standard 14 Fonts," the "Standard Type 1 Fonts", and the "Standard Fonts."

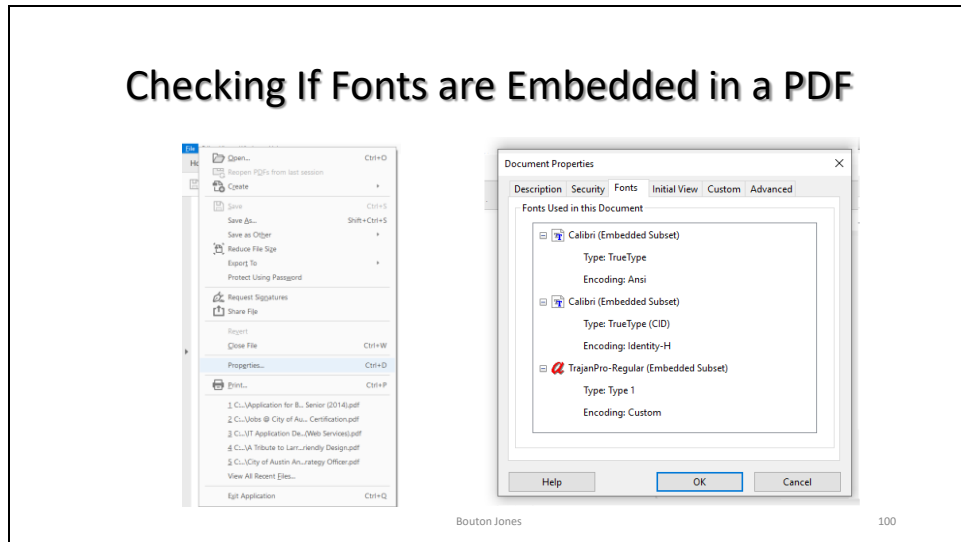
Monospaced Fonts: Courier, Courier Bold, Courier Oblique, Courier Bold-Oblique

Proportional Fonts (Sans Serif): Helvetica, Helvetica Bold, Helvetica Oblique, Helvetica Bold-Oblique

Proportional Fonts (Serif): Times Roman, Times Bold, Times Italic, Times Bold-Italic

Symbol

Zapf Dingbats



Checking If Fonts are Embedded in a PDF.

To check whether the fonts are all embedded in your PDF file or not:

- Open your PDF file
- Click File> Document Properties
- Click on the Fonts Tab to display the list of all fonts
- All fonts are either Type 1 or TrueType fonts
- All fonts should show as “Embedded Subset”

CITE: <https://kbpdfstudio.qoppa.com/how-to-embed-fonts-in-a-pdf-document/>

Embedding Fonts in *Word Docs*

1. In the **Files** tab select the **More** option.
2. Choose the **Options** option.
3. In the **Words** pane, select the **Save** option.
4. Select the “Embed fonts in the file” checkbox
5. Leave the “Embed only the characters used in the document (best for reducing file size)” checkbox.
6. Select the “Do not embed common systems fonts” checkbox.
7. Click the OK button to close this window.
8. You will need to use the “Save As PDF” or “Export to PDF”

The UI varies among releases of Microsoft Word. Microsoft is likely to change it for future versions of Word

Embedding Fonts in Word Documents

- In the **Files** tab select the **More** option.
- Choose the **Options** option.
- In the **Words** pane, select the **Save** option.
- Select the “Embed fonts in the file” checkbox
- Leave the “Embed only the characters used in the document (best for reducing file size)” checkbox.
- Select the “Do not embed common systems fonts” checkbox.
- Click the OK button to close this window.
- You will need to use the “Save As PDF” or “Export to PDF”

The UI varies among releases of Microsoft Word. Microsoft is likely to change it for future versions of Word

Embedding Fonts in *PDFs*

1. Click File > Print
2. Click Adobe PDF
3. Click the Properties button to the right of the Printer Name text box
4. Select the tab Adobe PDF Settings
5. Edit the Default Settings
6. Click Fonts
7. For Subset embedded fonts when percent of characters used is less than: Set the percentage to 100%
8. Select the Embed all Fonts option
 - A. For Embedding, select the folder with the fonts you want to embed from the drop-down list
 - B. Make sure the fonts you need to embed are in the Always Embed box and not in the Never Embed box

Bouton Jones

102

Embedding Fonts in PDFs

Pro tip: a font can only be embedded if it contains a setting by the font vendor that permits it to be embedded.

- * Click File > Print
- * Click Adobe PDF
- * Click the Properties button to the right of the Printer Name text box
- * Select the tab Adobe PDF Settings
- * Edit the Default Settings
- * Click Fonts
- * For Subset embedded fonts when percent of characters used is less than: Set the percentage to 100%
- * Select the Embed all Fonts option
 - * For Embedding, select the folder with the fonts you want to embed from the drop-down list
 - * Make sure the fonts you need to embed are in the Always Embed box and not in the Never Embed box

CITE: <https://www.printivity.com/insights/2020/09/13/how-to-embed-fonts-in-pdfs/>

See also: VIDEO "How to Embed Fonts in a PDF using Acrobat pro-2017" (1:21)
<https://www.youtube.com/watch?v=Su50hSBGcW8>

How much does embedding fonts increase the file size?

NOTE: This was a quick and dirty check --- not a rigorous scientific study. Your results might vary.

File	Size
PDF with no embedded fonts	239 KB
PDF with one embedded font	248 KB
PDF with two embedded fonts	250 KB
PDF with three embedded fonts	250 KB
PDF with four embedded fonts	260 KB

NOTE: A common typography best practice is to limit the number of fonts you use on web pages and documents to 2 – 3.

Bouton Jones

103

How much does embedding fonts increase the file size?

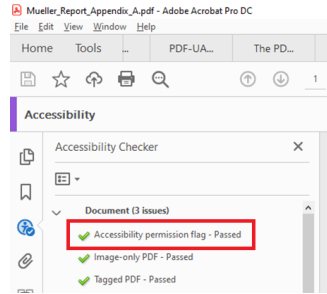
File	Size
PDF with no embedded fonts	239 KB
PDF with one embedded font	248 KB
PDF with two embedded fonts	250 KB
PDF with three embedded fonts	250 KB
PDF with four embedded fonts	260 KB

NOTE: This was a quick and dirty check --- not a rigorous scientific study. Your results might vary.

NOTE: A best practice in web design is to limit the number of fonts you use on web sites. It's probably a good idea to limit the number of fonts you use in a document whether embedded or not.

Accessibility Permission Flag

If "Accessibility permissions flag" fails, then content copying for assistive technology is not allowed and the document fails Baseline #23



Bouton Jones

104

Accessibility Permission Flag

If "Accessibility permissions flag" fails, then content copying for assistive technology is not allowed and the document fails Baseline #23

Accessibility permission flag is set

From "Using the Acrobat Pro DC Accessibility Checker":

Accessibility permission flag: A document author can specify that no part of an accessible PDF is to be copied, printed, extracted, commented on, or edited. This setting can interfere with the user of a screen reader's ability to read the documents screen readers must be able to copy or extract the document's text to convert it to speech.

This flag reports whether it's necessary to turn on the security settings that allow accessibility.

To fix the rule automatically, select Accessibility Permission Flag on the Accessibility Checker panel. Then, choose Fix from the Options menu.

(CITE: <https://www.adobe.com/accessibility/products/acrobat/using-acrobat-pro-accessibility-checker.html>)

Accessibility permission flag is set

"A document author can specify that no part of an accessible PDF is to be copied, printed, extracted, commented on, or edited. This setting can interfere with the user of a screen reader's ability to read the documents screen readers must be able to copy or extract the document's text to convert it to speech." (CITE: <https://www.adobe.com/accessibility/products/acrobat/using-acrobat-pro-accessibility-checker.html>)

One Last Point: the PDF Association's Recommendation

PDF/UA: Creation tools

In principle, universally-accessible PDF documents can either be generated directly during document creation, using a suitable program, or indirectly generated by adapting an existing PDF document. The indirect approach tends to require a great deal of work, as all tags and numerous other settings will need to be provided manually.

This work can also become void as soon as a new version of the PDF document replaces the old one, if any changes to the document's content need to be made. The adaptation process must then be started again from scratch.

The direct approach is almost always the better option; post-creation PDF edits should be avoided or at least kept to a minimum. Either way, it is essential that the document creation program can perform the functions required and that the document creator can make use of them. Although only a few programs currently support PDF/UA in its entirety, a wide range of options still exist in the form of the programs listed alphabetical below:

CITE: <https://www.pdfa.org/resource/pdfua-in-a-nutshell/>

Bouton Jones

106

Last Point: The PDF Association's Recommendation

“In principle, universally-accessible PDF documents can either be generated directly during document creation, using a suitable program, or indirectly generated by adapting an existing PDF document. The indirect approach tends to require a great deal of work, as all tags and numerous other settings will need to be provided manually.”

“This work can also become void as soon as a new version of the PDF document replaces the old one, if any changes to the document's content need to be made. The adaptation process must then be started again from scratch.”

“The direct approach is almost always the better option; post-creation PDF edits should be avoided or at least kept to a minimum. .”

CITE: <https://www.pdfa.org/resource/pdfua-in-a-nutshell/>

Additional Resources

- The PDF Association: <https://www.pdfa.org/>
- The PDF Accessibility group on Facebook: <https://www.facebook.com/groups/PDFAccessibility/>
- Tagged PDF: <https://taggedpdf.com/>
- Austin Accessibility and Inclusive Design (Meetup Group): <https://www.meetup.com/a11yATX>
- AccessU (A11y conference in May): <https://knowbility.org/programs/accessu/2021/>

Additional Resources

The PDF Association: <https://www.pdfa.org/>

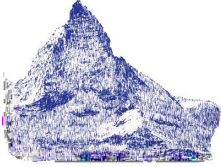


The PDF Accessibility group on Facebook: <https://www.facebook.com/groups/PDFAccessibility/>

Tagged PDF: <https://taggedpdf.com/>

Austin Accessibility and Inclusive Design (Meetup Group): <https://www.meetup.com/a11yATX>

AccessU (A11y conference in May): <https://knowbility.org/programs/accessu/2021/>

Closing Summary



Bouton Jones 108

Closing Summary

Follow both sets of guidelines --- WCAG and PDF/UA --- and use the Matterhorn Protocol as a checklist for the PDF/UA.



Any Questions?

Slide 110



Thank You. This concludes this session.

Hacking Redacted PDFs

[Open the Slides for "Hacking Redacted PDFs"](#)

Slide 112

December 10, 2020

Mueller Report: Appendix A

Evidence clearly indicated that ██████████ of Austin, TX is a double agent employed variously by the Special Executive for Counter-intelligence, Terrorism, Revenge and Extortion (SPECTRE) and Hydra.

He has also been implicated in the theft of yogurt from the 6th floor break room of the office building where he works. He strongly denies that accusation.

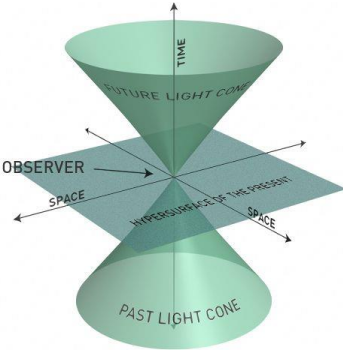
CONFIDENTIAL

Bouton Jones 112

Correcting Searchable Image PDFs

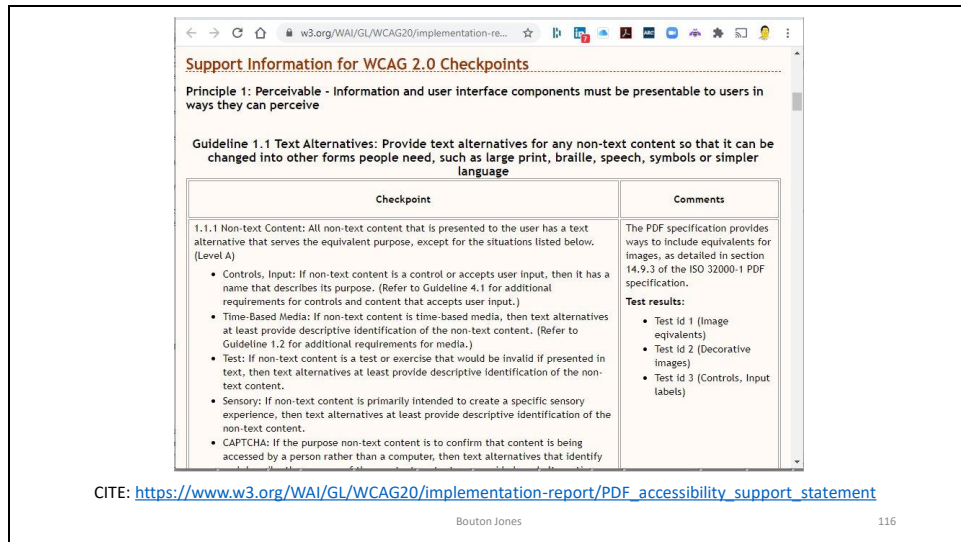
Correcting Searchable Image PDFs

Spaces Existent



Which Guidelines Do PDF A11y Checkers Use?

Slide 116



The screenshot shows a web browser window with the address bar displaying "w3.org/WAI/GL/WCAG20/implementation-re...". The page title is "Support Information for WCAG 2.0 Checkpoints". Below the title, it states "Principle 1: Perceivable - Information and user interface components must be presentable to users in ways they can perceive". Underneath, it says "Guideline 1.1 Text Alternatives: Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language".

Checkpoint	Comments
<p>1.1.1 Non-text Content: All non-text content that is presented to the user has a text alternative that serves the equivalent purpose, except for the situations listed below. (Level A)</p> <ul style="list-style-type: none">• Controls, Input: If non-text content is a control or accepts user input, then it has a name that describes its purpose. (Refer to Guideline 4.1 for additional requirements for controls and content that accepts user input.)• Time-Based Media: If non-text content is time-based media, then text alternatives at least provide descriptive identification of the non-text content. (Refer to Guideline 1.2 for additional requirements for media.)• Text: If non-text content is a test or exercise that would be invalid if presented in text, then text alternatives at least provide descriptive identification of the non-text content.• Sensory: If non-text content is primarily intended to create a specific sensory experience, then text alternatives at least provide descriptive identification of the non-text content.• CAPTCHA: If the purpose non-text content is to confirm that content is being accessed by a person rather than a computer, then text alternatives that identify	<p>The PDF specification provides ways to include equivalents for images, as detailed in section 14.9.3 of the ISO 32000-1 PDF specification.</p> <p>Test results:</p> <ul style="list-style-type: none">• Test id 1 (Image equivalents)• Test id 2 (Decorative images)• Test id 3 (Controls, Input labels)

CITE: https://www.w3.org/WAI/GL/WCAG20/implementation-report/PDF_accessibility_support_statement

Bouton Jones 116

Support Information for WCAG 2.0 Checkpoints

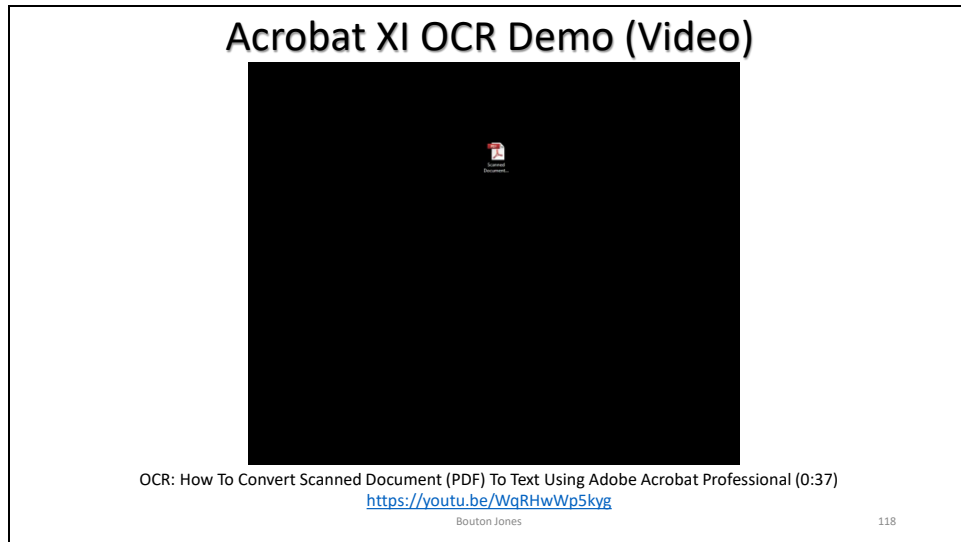
The Accessibility Support Documentation for PDF page on the W3C web site includes this section

https://www.w3.org/WAI/GL/WCAG20/implementation-report/PDF_accessibility_support_statement



OCR Steps in Adobe Acrobat 2017

- Select the Tools tab
- Select the Enhance Scans icon
- Select the Recognize Text menu
- Select the In This File Option
- Select the Recognize Text button



OCR Demo (Video)

OCR: How To Convert Scanned Document (PDF) To Text Using Adobe Acrobat Professional (:37)

This video has no narration, transcript, or close captioning.

In this video a PDF is displayed inside Adobe Acrobat. The user tries to select the text. But the PDF is image only so the text cannot be selected. The user runs OCR on the entire PDF. The pages are de-skewed (i.e. straightened) The text inside the images are converted into real text. They appear to retain the original fancy font face.

CITE: <https://youtu.be/WqRHwWp5kyg>

Three Ways to Check the OCR Accuracy

1. Compare the visible text to what a screen reader finds.
2. Copy and paste the newly editable text into a text document.
3. Change the font face --- of the *editable text* --- throughout the revised document. (Later you can change it back to the fonts closest to the original document's fonts.)

Three Ways to Check the OCR Accuracy

- Compare the visible text to what a screen reader finds.
- Copy and paste the newly editable text into a text document.
- Change the font face --- of the editable text --- throughout the revised document. (Later you can change it back to the fonts closest to the original document's fonts.)

How do you add the <tfoot> structure type?

- HTML (not?!)
- Adobe InDesign

How do you add the <tfoot> structure type?

How to *not* include Headers and Footers on the 1st page.

Side Note: Fonts for Dyslexia

- Readers with dyslexia are likely to read words by shape. They can confuse different letters with similar outlines: M-W, p-q, d-b, 1-l, O-0, etc.,
- Readily available fonts include: Arial, Comic Sans, Century Gothic, Verdana, Trebuchet, and Calibri.
- Specially designed fonts include Dyslexie, Open Dyslexic, Sylexiad, and Read Regular
- Dyslexie font, the revolutionary font for people with dyslexia - <https://www.dyslexiefont.com/en/typeface/>

Bouton Jones

122

Side Notes: Fonts for Dyslexia

Readers with dyslexia are likely to read words by shape. Confusing different letters that look similar: M-W, p-q, d-b, 1-l, O-0, etc.,

Readily available fonts include: Arial, Comic Sans, Century Gothic, Verdana, Trebuchet, and Calibri.

Specially designed fonts include Dyslexie, Open Dyslexic, Sylexiad, Dyslexie, and Read Regular

Dyslexie font, the revolutionary font for people with dyslexia
<https://www.dyslexiefont.com/en/typeface/>

OpenDyslexic CSS
<https://gist.github.com/mvilroxx/4536584>

Q&A Section: XMP Data

“The XMP specification includes more than a dozen predefined schemas with hundreds of properties for common document and image characteristics. The most widely used predefined XMP schema is called the Dublin Core, or dc. It includes general properties such as Title, Creator, Subject, and Description. In addition to predefined schemas, custom schemas can be defined to cover company- or industry-specific metadata requirements.

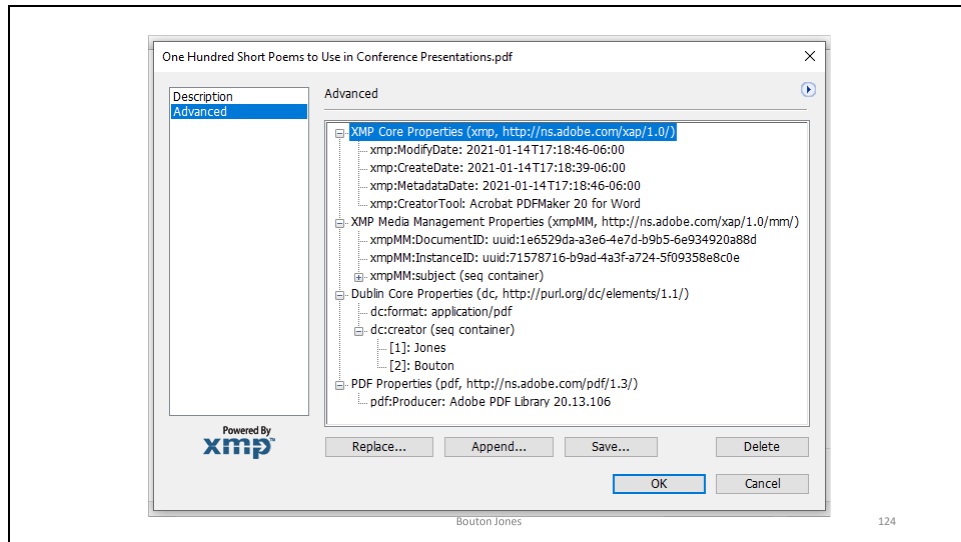
“The Dublin Core has been standardized as ISO 15836 (published in 2003, revised in 2009)”

XMP Resources

- Adobe's main XMP page: <http://www.adobe.com/products/xmp.html>
- XMP 2012 specification and Adobe's XMP developer's page with XMP Toolkit: <http://www.adobe.com/devnet/xmp.html>

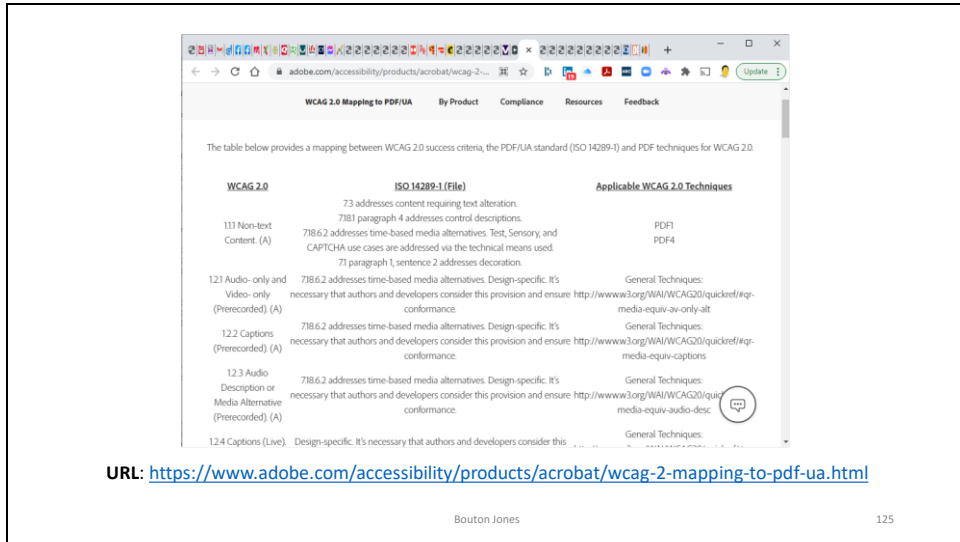
CITE: <https://www.pdflib.com/pdf-knowledge-base/xmp/xmp-overview/>

Slide 124



Dublin Core Properties in PDFs

Slide 125



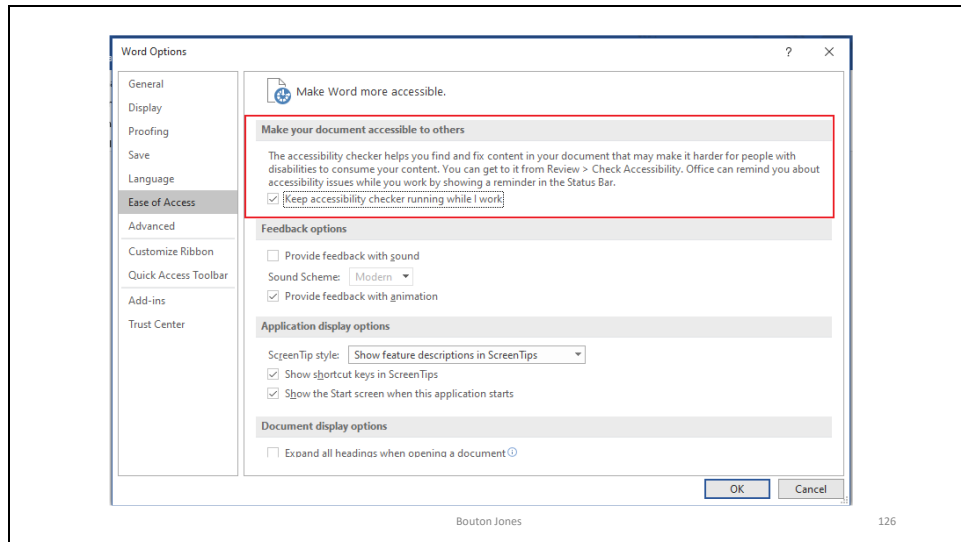
WCAG 2.0 Mapping to PDF/UA (Adobe)

Image: Screen capture of a Adobe.com web page with a table comparing WCAG and PDF/UA.

I couldn't figure out this table. I was confused by the display. I can't tell where the rows are. I'm not sure who this is for.

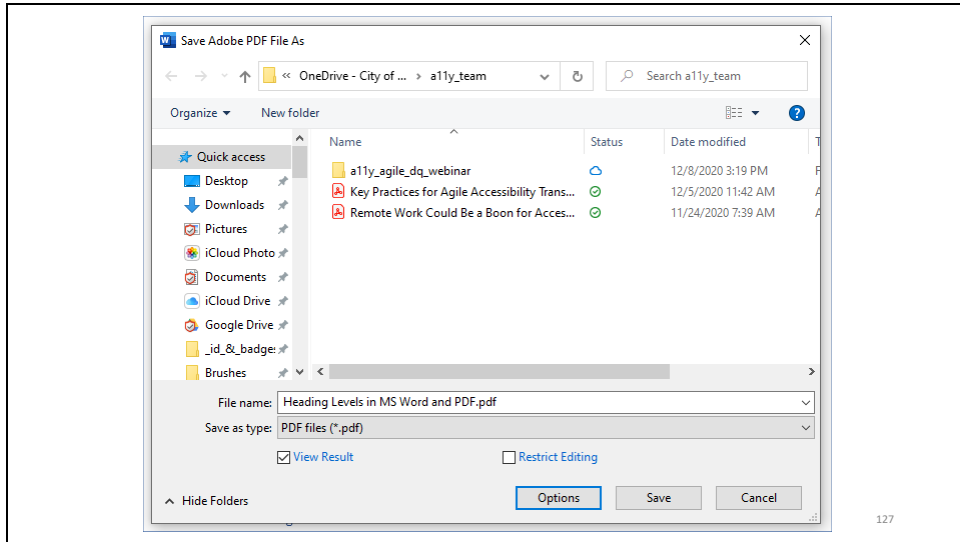
URL: <https://www.adobe.com/accessibility/products/acrobat/wcag-2-mapping-to-pdf-ua.html>

Slide 126

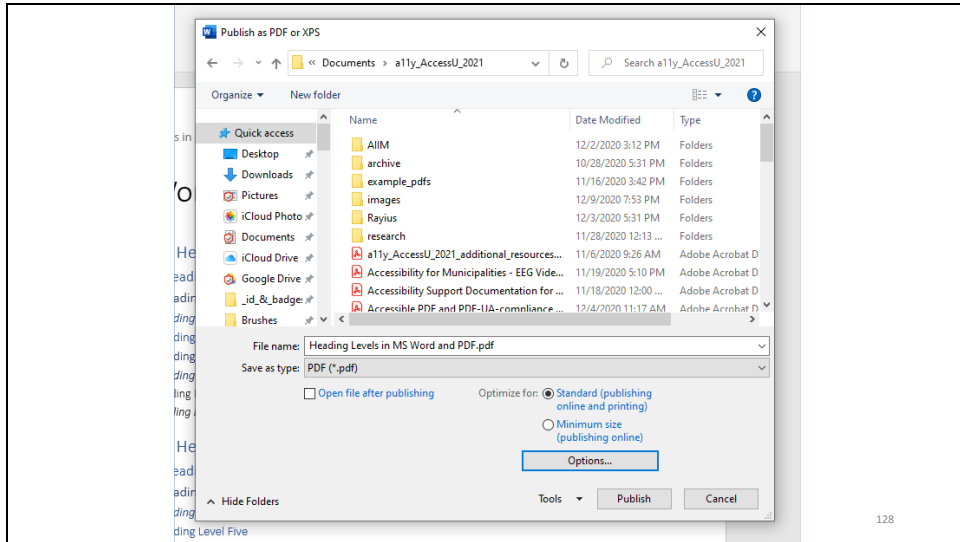


Keep the Word Accessibility Checker Running while You Work.

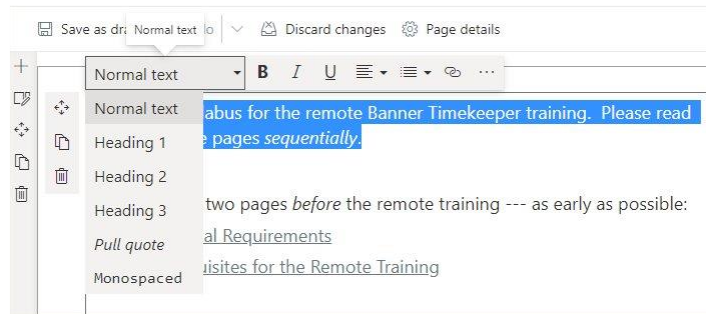
Slide 127

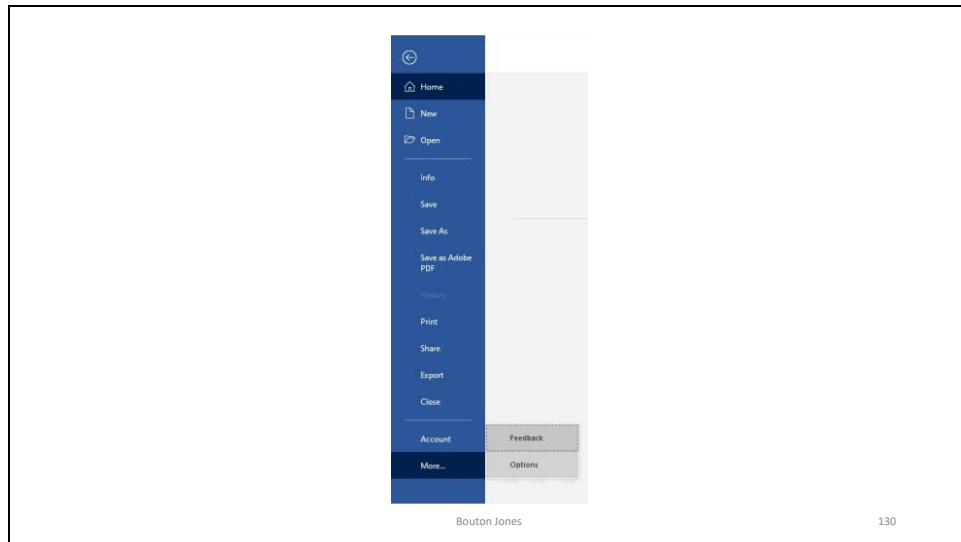


Slide 128



Semantic Tagging in SharePoint

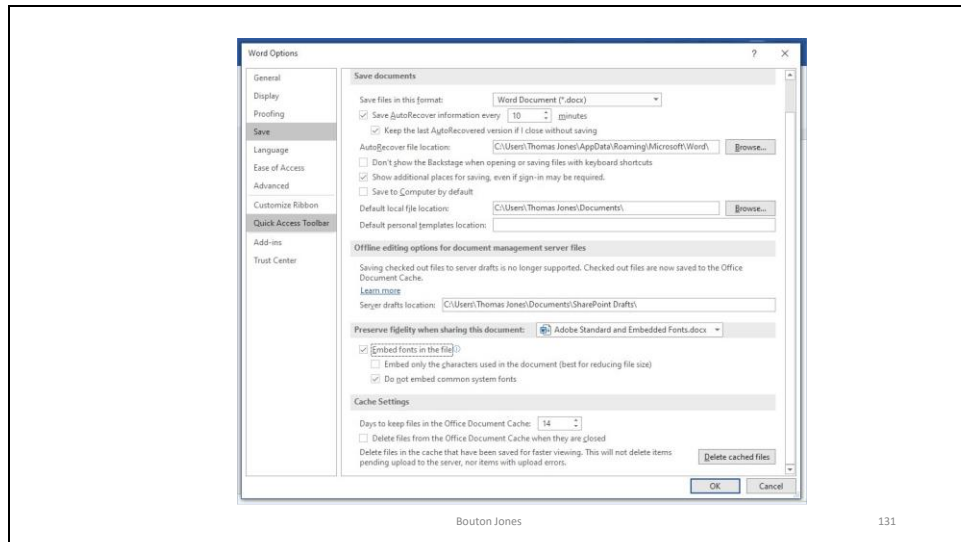




Embedding Fonts in Word Documents (Slide 2 of 3)

- In the **Files** tab select the **More** option.
- Choose the **Options** option.
- In the **Words** pane, select the **Save** option.
- Select the “Embed fonts in the file” checkbox
- Leave the “Embed only the characters used in the document (best for reducing file size)” checkbox.
- Select the “Do not embed common systems fonts” checkbox.
- Click the OK button to close this window.
- You will need to use the “Save As PDF” or “Export to PDF”

The UI varies among releases of Microsoft Word. Microsoft is likely to change it for future versions of Word



Bouton Jones

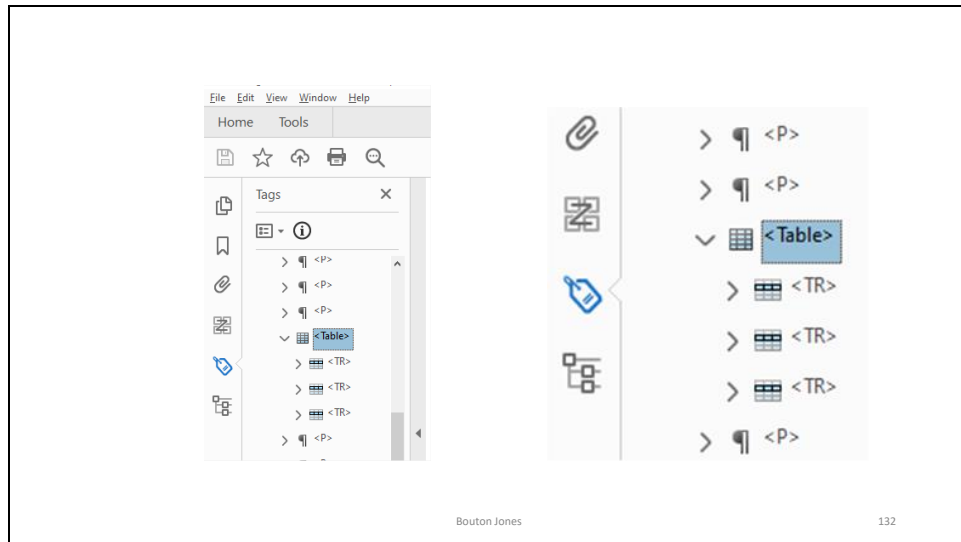
131

Embedding Fonts in Word Documents (Slide 3 of 3)

- In the **Files** tab select the **More** option.
- Choose the **Options** option.
- In the **Words** pane, select the **Save** option.
- Select the “Embed fonts in the file” checkbox
- Leave the “Embed only the characters used in the document (best for reducing file size)” checkbox.
- Select the “Do not embed common systems fonts” checkbox.
- Click the OK button to close this window.
- You will need to use the “Save As PDF” or “Export to PDF”

The UI varies among releases of Microsoft Word. Microsoft is likely to change it for future versions of Word

Slide 132



Bouton Jones

132

Viewing and Editing Table Tags in Acrobat

Two Screen Captures: A few table tags in Acrobat's Tag pane.

Dublin Core Metadata Element Set

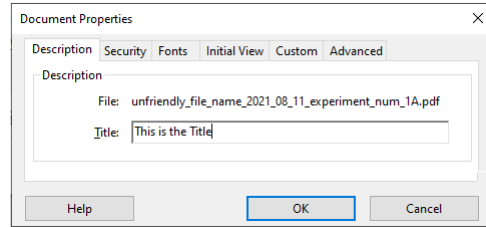
Dublin Core Element	Use
Title	A name given to the resource.
Subject	The topic of the resource.
Description	An account of the resource.
Creator	An entity primarily responsible for making the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Type	The nature or genre of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A related resource from which the described resource is derived.
Language	A language of the resource.
Relation	A related resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Rights	Information about rights held in and over the resource.

Dublin Core in HTML and PDF

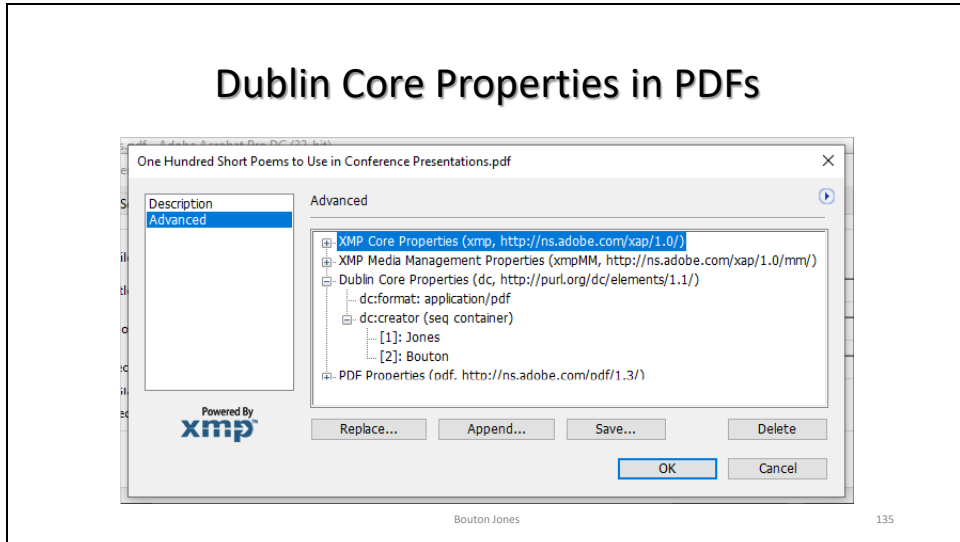
HTML

```
<HTML>
<HEAD>
  <TITLE>This is the Title</TITLE>
  <META NAME="DC.Title" CONTENT="Title
of the Document">
</HEAD>
<BODY>
  ...
</HTML>
```

PDF

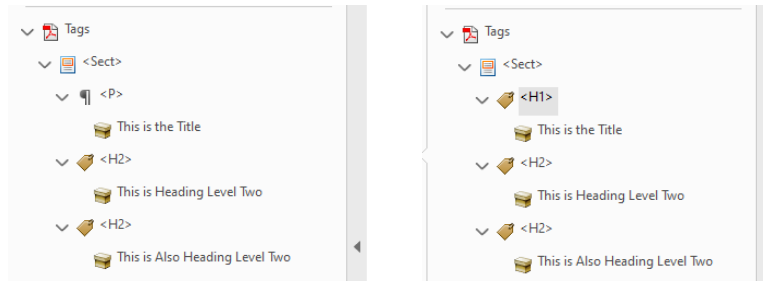


Dublin Core Properties in PDFs



Dublin Core Properties in PDFs

You can change the (Title) paragraph into H1.

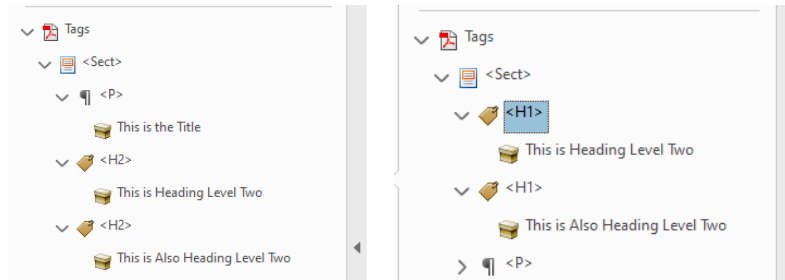


You can change the (Title) paragraph in H1.

This will not change the title meta data. Or the appearance (formatting) of the text in th PDF.

IMAGES: Two screen captures showing the Tag panel in Adobe Acrobat. In the first, the text “This is the title” is nested inside the P tag. In the second, the same text is nested inside the H1 tag.

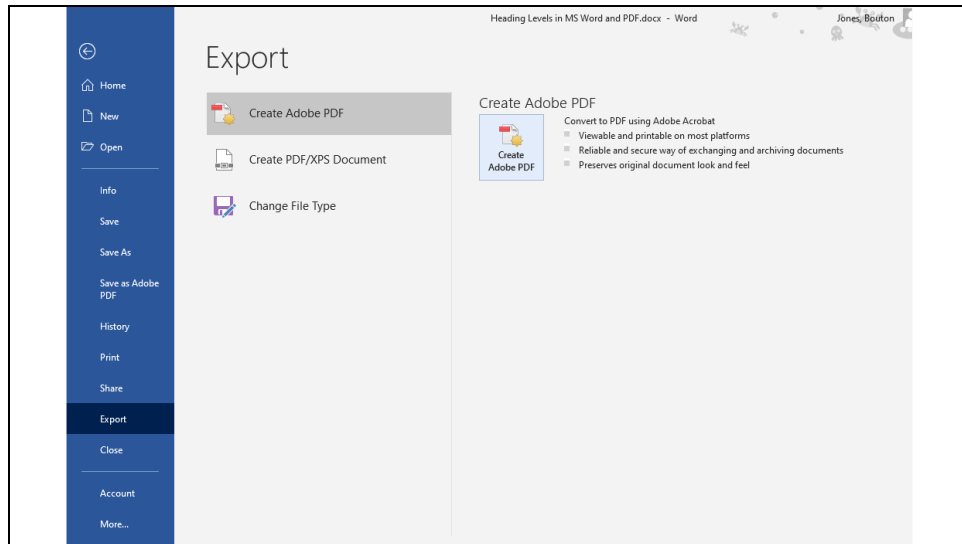
Or you can remove the paragraph entirely ---
and leave the Title property in the metadata.



Or you can remove the paragraph entirely and leave the property in the metadata.

IMAGES: Two screen captures showing the Tag panel in Adobe Acrobat. In the first, the text “This is the title” is nested inside the P tag. In the second, the P tag and the same text have been removed.

Slide 138



Heading levels above H6 are *not* standard structure types.

Table of Available Headers Levels by Medium

HTML 5	Word	PDF
6	9	6

Heading Levels by Medium

Heading levels above H6 are not standard structure types

Table of Available Headers Levels by Medium

- HTML 5 has 6 heading level tags: H1 through H6.
- Microsoft Word has 9 heading levels.
- PDF has 6 heading levels.