

# PDF Remediation Basics

CommonLook PDF



# Concepts covered

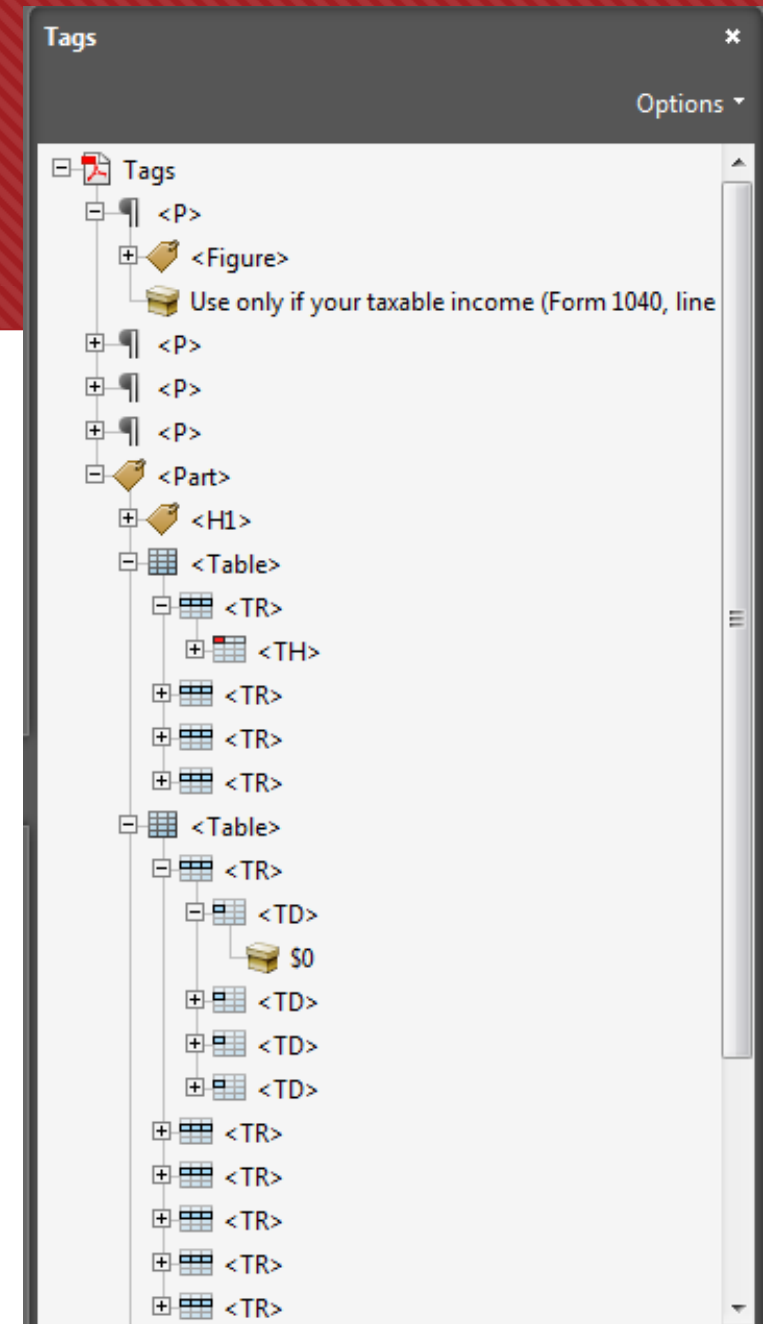
- Intro to PDF and Accessibility
- Intro to the Standards
- Document Analysis/ Creating a Workflow
- Fixing the Reading Order
- Images and Figures – Alt Text and Artifacts
- Working with Simple Tables
- Creating Lists
- Tagging Links
- Simple Forms
- Using Find and Replace
- Verifying Document Compliance
- Metadata

# What is a PDF

- Portable Document Format - Adobe®
- Documents portable across platforms and devices.
- A tagged document consists of three separate views:
  - ▶ Physical View
  - ▶ Content Order View: Used for reflow, by cell phone, tablet, web, etc.
  - ▶ Tag View: Used by Screen Readers, AT (assistive technology)
- If untagged, use Acrobat to add tags to document
- \*A common misconception – the “Order” menu in Acrobat

# Tags Explained

- Tags
  - ▶ Point to Containers from the Content Order and/or other Tags,
  - ▶ Show the Structure of the document (Paragraph, List, Table, etc.),
  - ▶ Are similar to simple HTML (P, TH, TD, etc.).
  - ▶ Changing the order of the tags does not change the physical view.
  - ▶ Deleting tags does not change the physical view.



# The Four Principles of Accessibility

PDF Web Content\* must be:

- **Perceivable** – Information cannot be invisible to all senses.
- **Operable** - User interface components and navigation cannot require interaction that the user cannot perform.
- **Understandable** - Information and the operation of user interface cannot be beyond the level of comprehension of the user.
- **Robust** – 1) PDFs have to render properly on all technologies, 2) As technology changes, content must still be accessible.

If any of these are not true, users with disabilities will not be able to use the Web.

(from <http://www.w3.org/TR/UNDERSTANDING-WCAG20/intro.html#introduction-fourprincs-head>)

\* These principles are from WCAG for web pages, but can be extended to other electronic content like PDF web content

# WCAG and Success Criteria

## WCAG – Web Content Accessibility Guidelines

Checkpoint success criteria levels are based on the checkpoint's impact on accessibility.

### [A]

- If level A is not satisfied, one or more groups will find it impossible to access information in the document. (Example – Alt Text on images.)

### [AA]

- If level AA is not satisfied, one or more groups will find it difficult to access information in the document. (Example – Heading Levels)

### [AAA]

- If level AAA is not satisfied, one or more groups will find it somewhat difficult to access information in the document. (Example – Expansion Text)

# Creating a Workflow

- In Acrobat (if needed)
  - ▶ Run OCR (scanned docs)
  - ▶ Add interactive elements (form fields, links)
  - ▶ Add script work to form if needed
  - ▶ Embed fonts (PDF/UA)
  - ▶ Add Tags/ Check for corruption
  - ▶ Save AS!
- Optional Tips
  - ▶ Print out the Table of Contents for Headings
  - ▶ Check for presence of tables
  - ▶ Look for images that convey information
- In CommonLook
  - ▶ Check/Fix reading order
  - ▶ Check/Fix correctness of tags
    - Clean up empty tags occasionally – esp. at end.
  - ▶ Verify Color use
  - ▶ Check/Fix other checkpoints
  - ▶ Run Validation

**Let's Try it!**

**Hands on training!**



# A Settings Adjustment

- In the Settings Tab
  - ▶ In the Business Group
    - Choose User Preferences
    - Check the checkbox for “Empty Spaces...”
  - ▶ Choose Alternate Highlight Color?

# Fixing the Reading Order and Tags

- Navigate through the tag tree to make sure content is in the correct order.
- Move content/tags to the correct place as needed.
  - ▶ Click and Drag
  - ▶ Cut and Paste
  - ▶ Use “Change Tag Order” (Ctrl+Shift+ Up/Down Arrow)
- Make sure the correct Tag is being used (Heading levels, etc.).
- If not...
  - ▶ Select content (highlight in the Physical view, select in Tags, etc.)
  - ▶ Tag it correctly
    - Merge Tags
    - Create new tag
      - Make sure “Level” is correct
    - Convert existing tags

# Images and Figures

- Alternative Text – needed for images that convey information
  - ▶ Create/ Edit in the Properties Panel
  - ▶ Check/ Fix with Validation later
- Artifacts – (untagged and unread)
  - ▶ Images that are just decoration
  - ▶ Anything else that should not be read (page numbers, footers, headers, etc.)
    - Use the Pagination option for running headers, footers, etc.
  - ▶ Delete key does not delete content – it untags (artifacts) it

# Verification of Compliance

## In the Standards Panel

- Structural Check – ISO 32000-1
  - ▶ Makes sure tags are built correctly (Tables/Lists contain what they need, etc.)
  - ▶ Detects empty tags, empty tags with Alt text, broken links, etc.
- Accessibility – (508, WCAG, PDF/UA, HHS)
  - ▶ Open the menu
  - ▶ Choose your standard(s)
  - ▶ Run from the “Root” of the standard or choose a specific checkpoint
- Validation
  - ▶ Full – checks the whole document every time
  - ▶ Incremental – only checks things in the document that have changed since the last validation
  - ▶ Filter results – Failed, User Verify, Warning, etc.
  - ▶ Use the “Fix Wizard” and “Status” to correct any remaining issues.

# Metadata

- Add “Document” Tag – right click/ open the context menu on the Tags root, choose “Add Document Tag”
- Checkbox for “Display doc title” should be checked
- Keywords – separate by semicolon
- Subject is usually same as doc title
- Author is the organization (not your name!)
- Title – should be easy to identify the document
- Set the Language
- Checkbox for PDF/UA (and other standards) compliance IF it passes (run the check first!)

# Presentation (or Layout) Tables

If a table is used for presentation and is not a data table, it should not be marked as a table in the structure.

Linearize the table into paragraphs.

- Options for Linearizing Presentation Tables
  - ▶ Horizontally – cell by cell, row by row, top to bottom
  - ▶ Vertically – cell by cell down the columns from left to right

# Properly Tagged Data Tables

- Data Tables should be in Table tags
- Table tags contain Table Rows (TR) tags
- TR tags contain Table Data cells (TD) and/or Table Header cells (TH).
- The column and row span of each cell should reflect the physical view.
- The scope of header cells (TH) should also be indicated (column or row).

# Tagging Data Tables

- Highlight the table in the physical view or select content in the tags
- Choose “Table” from the “Generate Tag from Selection” group
- “Level Up” and “Clean-up” as needed
- Right Click (open the context menu) on the <Table> to open the Table Editor
- Check the columns/rows for correct layout
- Add, delete, merge cells, adjust span to fix as needed
- Make sure headers are marked TH and given correct scope.



# Tagging Lists

- Highlight the list text in the Physical View (including bullets, numbers, etc.) or select from within the tags.
- In the “Generate Tag from Selection” Group – Choose List
- “Level Up” and “Clean-up” as needed
- Make sure the list structure is correct (LI, Lbl, and Lbody tags, nesting, etc.) – Fix as needed
- Verify the ListNumbering Attribute value is correct (Properties panel)

Note: Nested Lists go inside the appropriate <LBody>

# Tagging Links

Note: the Link Annotation is the “interactive part” of the link.

- Select the text and the annotation in the physical view (or in the Tags tree)
- In the Insert Tag tab choose <Link>
- Link tags should be inside another structural tag (<P> for example) not on the root.
- In the Annotation Properties, check and edit the link target as needed.
- (WCAG) – Link tags need Alt text (may use Fix Wizard)
- (PDF/UA) – Link Annotation need Contents (like a tooltip on a form). (may use Fix Wizard)
- Sometimes possible to tag the <P> and the Link is automatically tagged.
- Force the tabbing order to follow document structure (next slide).

# Tabbing Order (Links and Forms)

- Force the tabbing order to follow the Document structure
  - ▶ Open the Untagged Content panel
  - ▶ Select the page (or multi-select the pages)
  - ▶ In the Properties Panel, make sure the “S” radio button is selected under Tabs.

Note: This can also be done automatically with the Fix Wizard after Validation

# Forms

Note: an Annotation is the “interactive part” of the form (text field, checkbox, etc.)

Important: Every form annotation needs to be placed in its own <Form> tag.

- When the document contains a form (question), include the form annotation with its text label in a <Form> tag:
  - ▶ Select the text and the annotation (widget) in the physical view or in the Tags tree
  - ▶ In the Insert Tag Tab choose <Form>
  - ▶ Form tags should be inside another structural tag (<P> for example) not on the root.
  - ▶ All Form Annotations need tooltips. (In the annotation’s Properties panel.)
    - (May be assigned “en masse” with the Fix Wizard)
  - ▶ If the annotation does not have a text label, place the annotation in a <Form> tag and provide the descriptive text in the tooltip.
  - ▶ The order of the form elements in the tag tree should follow the logical reading order in the physical view.
  - ▶ Set the document’s tabbing order to follow the document’s structure (previous slide)

# Find and Replace

## Find and Replace Text, Symbols, etc.

A quick way to add Alt Text to “other” text/ characters

- Replacing Symbols
  - ▶ Select the Symbol (text, tag, etc. as needed)
  - ▶ Press Ctrl + H to copy the symbol, text, tag, etc. into the “Find” box in the “Replace” tab
  - ▶ Copy and paste (or type) the replacement text as needed
  - ▶ Use “Replace” or “Replace all” as needed

# Additional Verification Checkpoints Explained

- **Multimedia:** If the document contains multimedia objects, verify that the alternatives are synchronized with the presentation. Videos need captions, audio needs transcripts, etc. That's beyond the scope of this course.
- **Color:** Make sure that all information conveyed with color is also available without color. Use symbols, patterns, etc. This comes into play more at the document creation phase. PDFs that fail this checkpoint often have to be edited at the source.
- **Contrast:** Make sure that there's sufficient contrast between foreground and background. Check ratios for "regular" and "large" text. Documents that fail the contrast check need to be edited at the source.

# Additional Verification Checkpoints Explained (cont'd.)

- Server-side image maps (even though very rare) maybe implemented in PDF. Verify that every active region of the server-side image map has an equivalent link in the document.
- Flicker: If the document contains any objects that may cause flickering (animated gifs, video files, flash components ... etc.) make sure that they do not flicker at a rate between 2 and 55 Hz. (more than 3x per second)

# Additional Verification Checkpoints

(cont'd.)

- **Timed Response:** When a timed response is required, the user 1) shall be alerted and 2) given sufficient time to indicate that they need more time. (Example, online banking time-out notifications.)
- **Text-only Alternative:** (Last resort – not allowed by all standards) When compliance cannot be accomplished in any other way, a text-only page, with equivalent information or functionality, shall be provided to make a web site comply with the provisions of these standards. The content of the text-only page shall be updated whenever the primary page changes. This is generally not applicable to PDF because PDFs can be made accessible.



# Thanks for Coming

And happy remediating!

Paul Rayius

[prayius@commonlook.com](mailto:prayius@commonlook.com)