ITEXT

PDF/UA:
the inclusive
document format

PDF/UA:
Accessibility is key

Table of contents

PDF/UA:
Accessibility is key

Equal access to information in digital documents

Now that we are rapidly moving towards a world where digital documents are not just an alternative, but the standard, we need to make sure everybody is aboard for this digital revolution and has full access to all information in digital documents. And it is impossible to think about digital documents without including PDF, the most common file type for digital documents.

PDF/UA - the PDF standard for Universal Accessibility - offers a clear framework for accessible PDF documents and aims to assure that all information inside a PDF is equally accessible to everyone. The focus is particularly on including those with disabilities that make use of assistive technology (AT) such as screen readers, magnifiers and adaptive navigation. Considering there are 1 billion people worldwide with disabilities, that is certainly enough to justify the existence of an accessibility standard for PDF, but we'll also see certain aspects of PDF/UA can benefit other implementations reliant on machine interpretation.

In this ebook we will first get to know our audience and the tech they use. Then we will zoom in on the requirements that make up PDF/UA and how to validate their conformance. We'll also learn about the regulations across the world and how PDF/UA can benefit all users, with or without disabilities.

We'll even take it one step further and provide you a tutorial on creating PDF/UA conformant documents with the iText 7 Suite - the leading open-source PDF library/SDK. iText was the first to successfully bring PDF to the backend and has been doing so for over 20 years. As an active member of both the ISO-committee and the PDF Association, iText remains an innovator in PDF technology to this day and its software is deeply intertwined with the evolution of the PDF/UA standard.

1.

Universal Accessibility

Know your audience

If we want to understand PDF/UA, we need to first understand its biggest stakeholder. Let's take a look at how users with disabilities consume information to get a better understanding of who we are making our documents accessible for.

Every user deserves a first-class experience consuming digital documents. The primary goal of PDF/UA is for people with disabilities to be able to experience PDF content with the same successful outcome as those without disabilities. Considering there are over 1 billion people in the world with disabilities[1], not only do they form a vast user group, also the degree in which they are impacted by inaccessible digital documents is continuously growing with more and more information having a digital-only source.

Those users with disabilities should be able to efficiently get to all the information in those documents without help from others. In order to do that a substantial part of them makes use of assistive technologies. When used right, a digital accessible document in combination with assistive technology can offer a lot of advantages to disabled users.

1.1 ASSISTIVE TECHNOLOGIES (AT)

In this subsection we are focusing on the more common impairments and disabilities and their respective AT in order to get a general understanding of how users with disabilities consume information in digital documents. This in no way implies that those are the only users and technologies that benefit from the universal accessibility standard.

---

1   1 Billion people worldwide have disabilities, https://globalaccessibilityawarenessday.org/

1.1.1 Visual impairments

When designing digital documents for users with visual impairments it is important to consider both the output and the input aspect of the end user's interaction when viewing the document.

Screen readers are a great way to translate text to speech, or even Braille on a braille display. A very popular screen reader is NVDA, which was also the world's first PDF/UA conforming screen reader. For those with partial vision a screen magnifier can be useful.
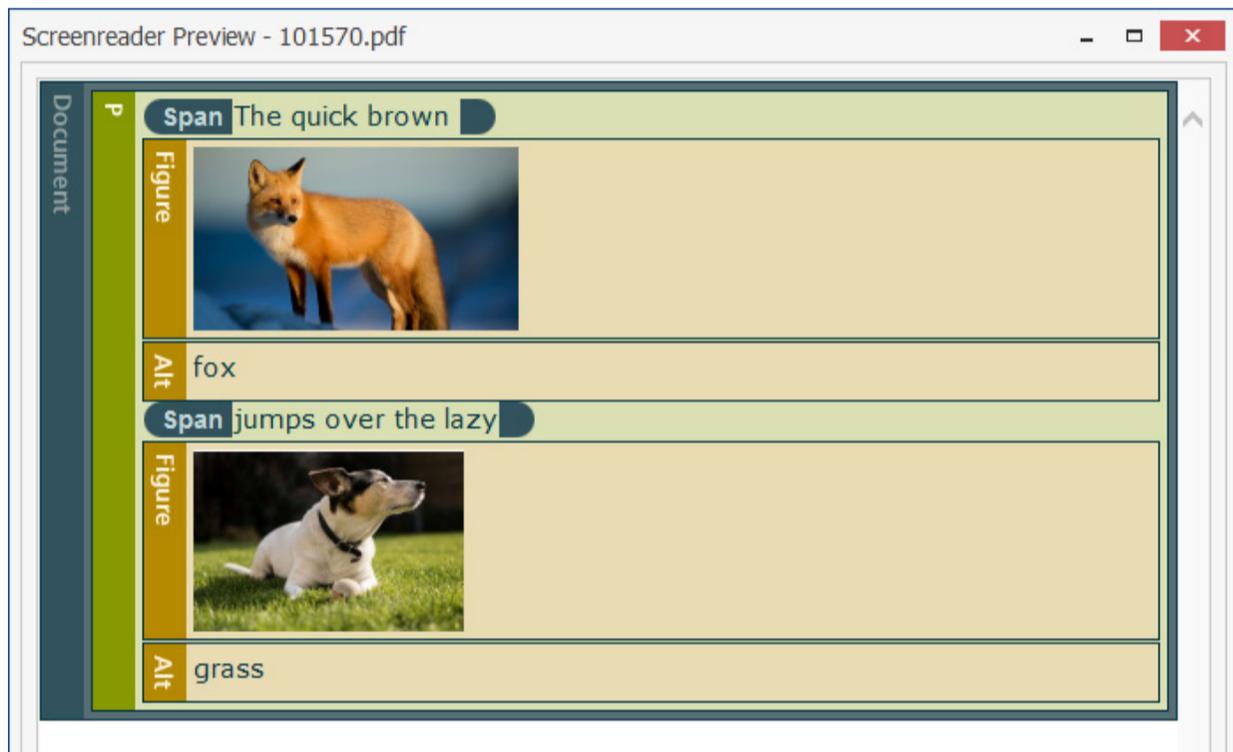
A screenshot from the PAC (PDF Accessibility Checker) Screen Reader Preview, a tool letting you experience how a screen reader interprets the content of your PDF.
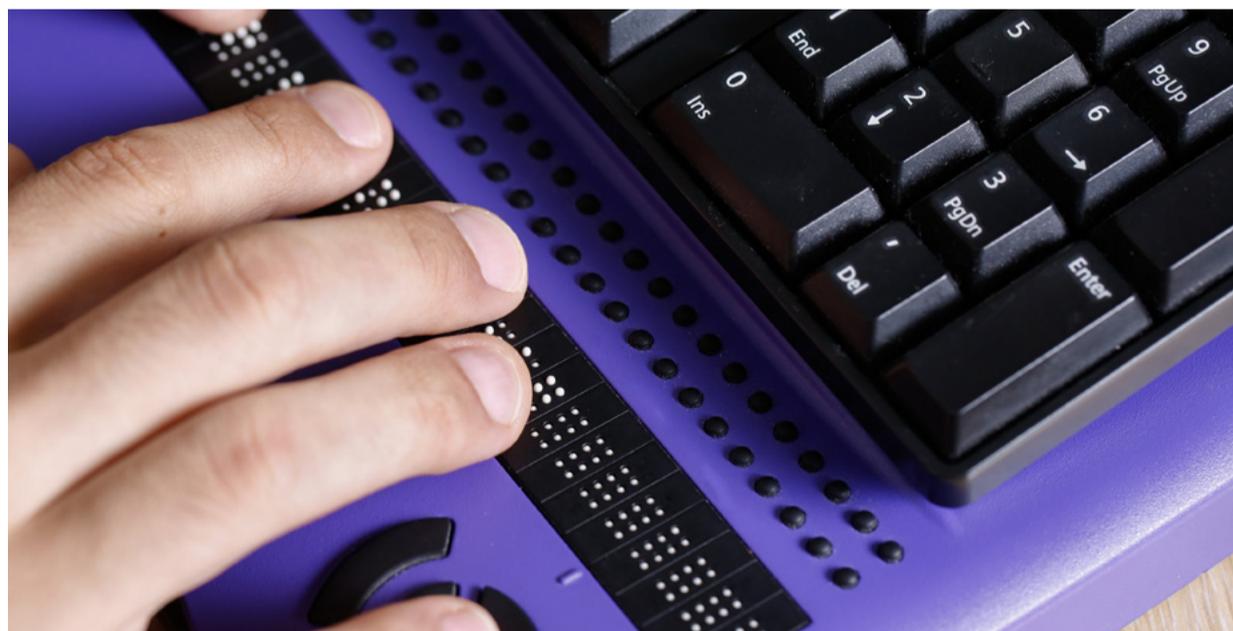
A Braille Display allows the visually impaired user to read the content of the document.

When it comes to user input, navigation is done by means of adaptive keyboards or some of the other assistive technologies we list under motor impairments. Of vital importance it that this type of navigation happens in a sequential way and efficiency requires heavily on an appropriate logical reading order of the document.

1.1.2 Motor impairments

Motor impairments mostly affect the way users navigate PDFs. There are several categories of assistive technologies that let users navigate the document in alternative ways: adaptive keyboards, dictation software, eye-tracking software, sip-and-puff switches, head pointers and so on.

Almost all of these require more effort than the traditional mouse and keyboard setup. This is one reason PDF/UA does not only focus on allowing users to access all information in the document, but also that they can do so efficiently.
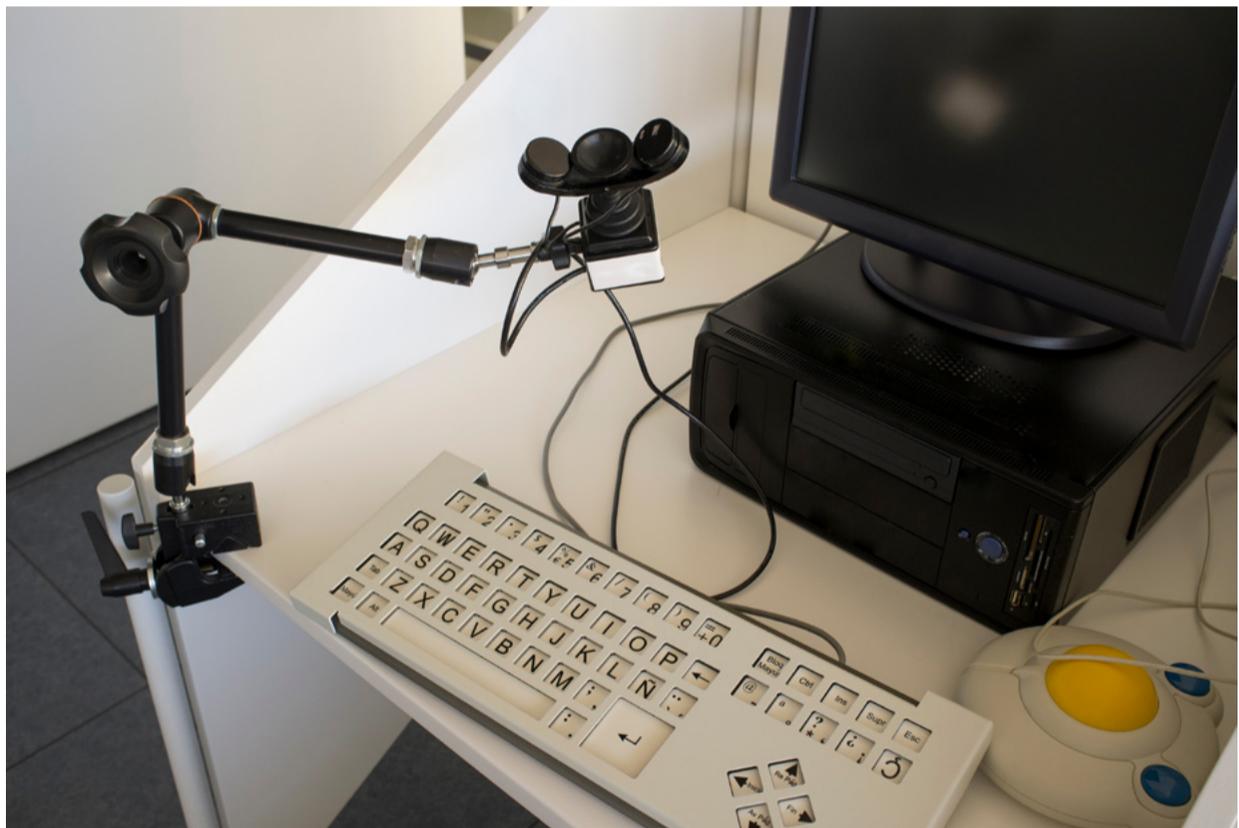
Setup with alternate input devices: chin mouse, adaptive keyboard and trackball.

1.1.3 Cognitive disabilities and impairments

For users with cognitive disabilities it can be helpful to digest the information in a document in an alternate way. For example, a user with dyslexia could be helped by software that highlights and reads individual words out-loud.

Want to generate and manipulate your PDFs with an open source (AGPL) or commercially licensed PDF library and SDK? Get started with iText 7 today!

Text to speech is most helpful when it highlights the words as they are spoken. Dyslexic people say this focuses their attention and helps their understanding of the content.

2.

Understanding PDF/UA

What makes a document PDF/UA compliant?

2.1 TECHNICAL REQUIREMENTS

PDF/UA-1, specified in ISO 14289-1:2014, is based on PDF 1.7, specified in ISO 32000-1, and is consistent with the Web Content Accessibility Guidelines (WCAG), a W3C Recommendation for accessible web content.

To ensure Universal accessibility, PDF/UA does not only address the requirements for the **file format**, but also sets requirements for **applications** and **assistive technologies** tailoring to those with disabilities.

2.1.1 File format

PDF/UA is not a separate file format, but is based on the general PDF specifications and has some additional requirements and constraints.

≡ All meaningful content (also referred to as real content in the PDF/UA specifications) needs to be tagged in a **logical and semantically correct way**. More on tagged PDF in the next section.

≡ All other content such as pagination or repeating table headers, is considered an **artifact** and should be flagged as such. Artifacts describe an element as decorative and unimportant, and those elements will be ignored by AT.

≡ The **structure tree** made up of the tags should represent a logical reading order for a human user.

- The main language of the document and any changes in language throughout the document should be marked.

- A **title** needs to be given and set to display in the window's title bar (rather than the filename).

- **Fonts** must be **embedded** and Unicode mapped.

- **Meaning** should never be conveyed by color, contrast or position alone.

- Images that convey meaning must contain **alternative descriptions** and must be tagged with a Figure tag.

- Flickering, blinking or flashing content is **forbidden**.

- **Encryption settings** must allow access by assistive technologies (which is standard in PDF 2.0 and thus PDF/UA-2).

2.1.2 Applications

PDF applications form the layer between the PDF document and the user interpreting it with assistive technology. In order to ensure full access to all the information in the document, PDF/UA sets out a number of requirements for those PDF applications.

- The application must be able to read **all information** and make it available to the assistive technology. This includes all tags and artifacts.

- The application should be able to recognize the **language** and language changes.

- **Navigation** by page number, bookmarks or structure tree is accommodated

- Media content should only be played **on command**

- **All functionality** must be available, including representation of digital signatures, annotations and optional content.

2.1.3 Assistive technology

Assistive technology, as we have discussed in chapter 1, are key to users with abilities to access content in an efficient way. To make sure these assistive technologies work with the PDF/UA standard, a number of requirements have been set here as well.

They interlace with the other requirements and can in short be described as letting the user read all information and navigate efficiently by using their specific assistive technology.

2.2 ZOOM IN ON TAGGED PDF

A tagged PDF is a PDF that identifies and provides meta-information around the **structural elements** of the content making up a PDF. Each of the tags identifies the type of element and thereby also its place in the hierarchical structure tree of the document. The order of the tags sets the reading order of the PDF.

While the tagging process can be automated, semantically correct tagging in the context of accessibility always requires human inspection and corrections. It is also important to note that only a **standard set of tags** can be used in tagged PDF. The PDF Association published a best practice guide for Tagged PDF where all relevant tags and their meaning can be found. Semantically correct tags are vital to PDF/UA, but as we now know they are by themselves not enough to form an accessible PDF.

Possible ways to create tagged PDF include using authoring software such as the MS Office suite or specialist Design tools such as Adobe InDesign or iText DITO. Using the iText 7 Core Library is also a really easy way to create a tagged PDF programmatically in C# (.NET) or Java.
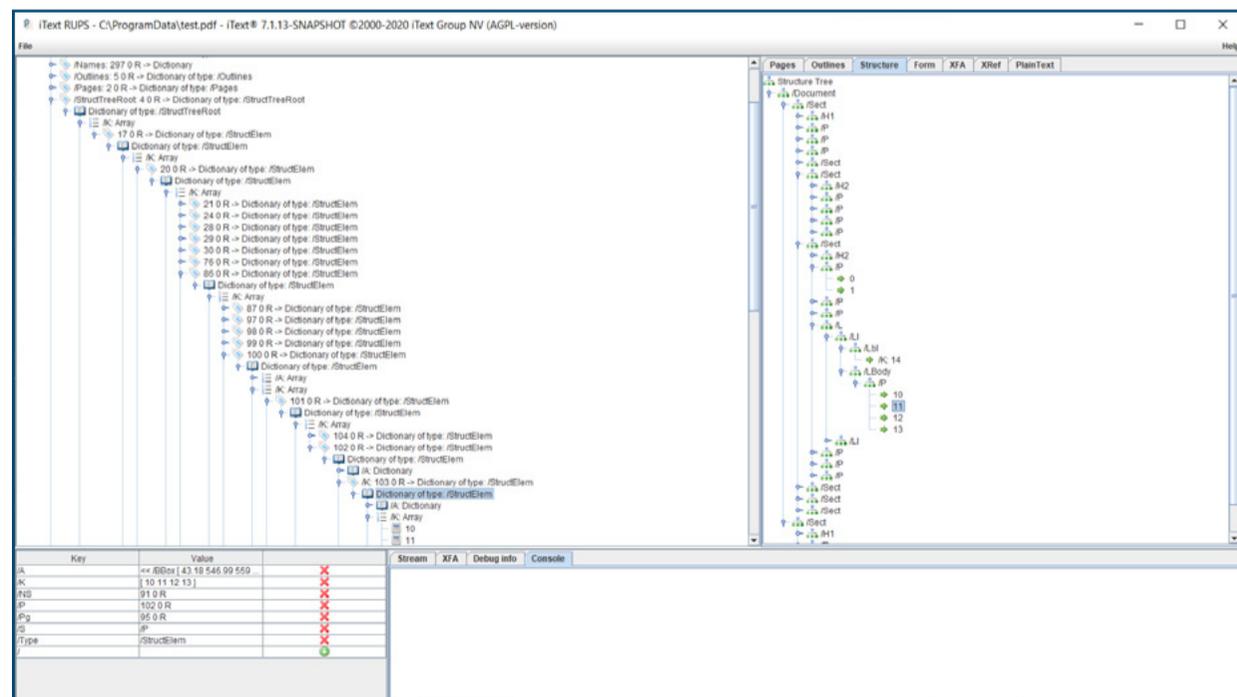

But how do you know if an existing PDF is tagged? The free PDF inspector iText RUPS allows you to inspect the complete structure tree and even all the low-level data.

2.3 PDF/UA-2

PDF/UA-2 (ISO 14289-2) is the **second part of the standard** and is expected to be released at the start of 2024. Through a number of ISO Working Group meetings the basis for PDF/UA-2 is now being finalized. As an active member of the ISO committee, several employees of iText Software are participating in the creation of these specifications.

Being based on PDF 2.0, implies a new set of tags can be used, while some other features that were deprecated in PDF 2.0 will no longer be available.
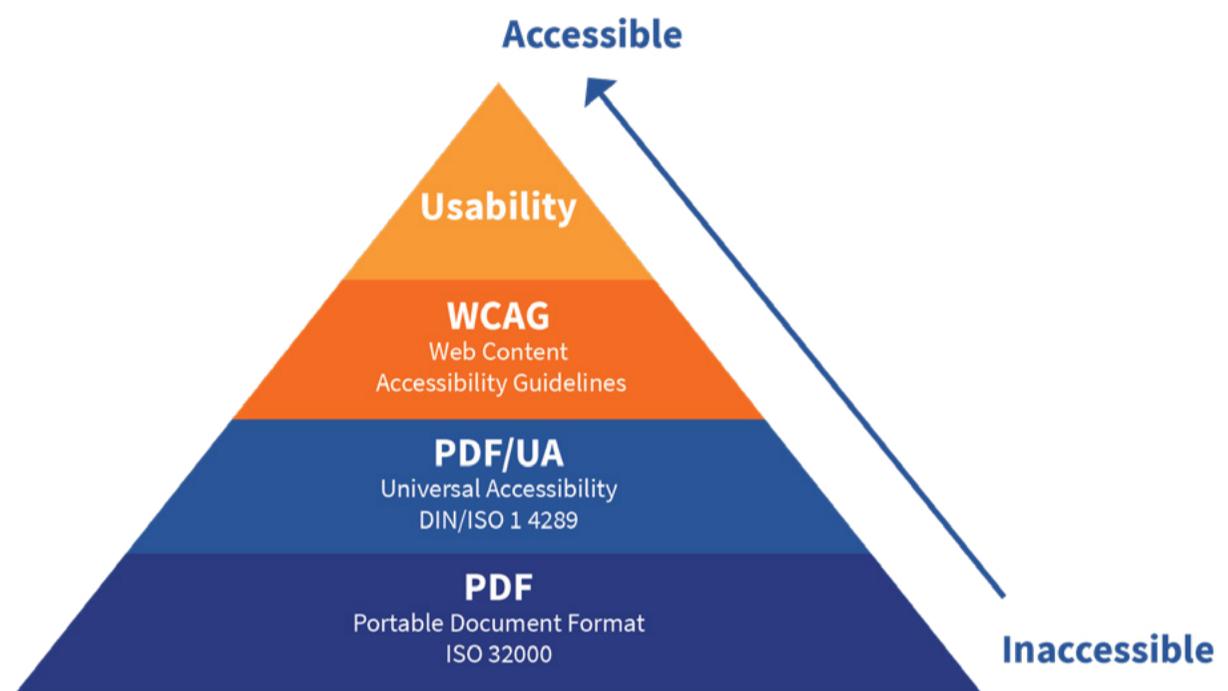
PDF accessibility requirements can be seen as a four-layer pyramid. Every layer builds on the previous layers, as the PDF/UA Foundation beautifully describes. Source: https://pdfua.foundation/en/why-pdf-ua

3.

Validation

How to check if your documents are accessible?

The accessibility of a document can't be verified with a quick glance at the document. Full compliance validation consists of two parts. A **machine test** using an accessibility checker **and a human verification**. The accessibility checker software can only check the purely technical requirements of the PDF/UA standard.

Manual checking is needed to perform a qualitative check of these requirements. For example, while an alternate description might be present, the software cannot check if it is an appropriate and meaningful description. Furthermore, human verification checks certain semantic aspects that cannot be tested by a piece of software, such as the correct reading order and color contrast.

Perhaps in the future **Artificial Intelligence (AI)** will evolve to allow for more of the requirements to be met and checked automatically, but for now it is certainly not adequate. For example, image recognition is not accurate enough to always provide the right alternative text. You can read more on this in our blog **"Can we use machine learning to achieve PDF/UA compliance?"**.

3.1 THE MATTERHORN PROTOCOL

The Matterhorn Protocol is a list of **31 checkpoints** representing all possible PDF/UA violations that one can make. It was created by the PDF Association's PDF/UA Competence Center in an effort to have a more direct guide for software developers and other stakeholders in document accessibility. With 16 pages compared to the 17 pages of the PDF/UA specifications - whose subsections it does refer to - it has a very similar size, but it offers the information in a bite-sized checklist and it is available free of charge.

The 136 failure conditions consist of 87 items that can be checked by software alone, 47 items that most likely require human involvement and two that have no specific test. The conditions are grouped into 31 topical checkpoints ranging from real content tagged to fonts.

It is interesting to note that the Matterhorn Protocol is published as a reference-quality PDF/UA file.

3.2 PDF ACCESSIBILITY CHECKER (PAC)

PAC was the first automated PDF/UA compliance checker on the market. It is developed by the PDF/UA Foundation and is available for free. It is known to be the **most commonly used validation tool** for PDF/UA. The 2021 version provides machine-testable checks for both PDF/UA and WCAG compliance.

In addition, it features a screen reader preview tool and allows you to view the tag tree and its properties.


The PAC Accessibility Checker indicates that we lack Alternate Descriptions and natural language in our PDF document.

3.4 VERAPDF

When it comes to validation of **PDF/A** the most popular and reliable solution so far has been veraPDF. Since version 1.18 released in April 2021 it also supports PDF/UA-1 machine validation following the guidelines set out in the Matterhorn Protocol.

veraPDF is an **open-source** solution that was originally funded by the EU and is supported by the PDF community represented by the Open Preservation Foundation, the PDF Association and the Digital Preservation Coalition. Today veraPDF is maintained by the Open Preservation Foundation and developer Dual Lab.
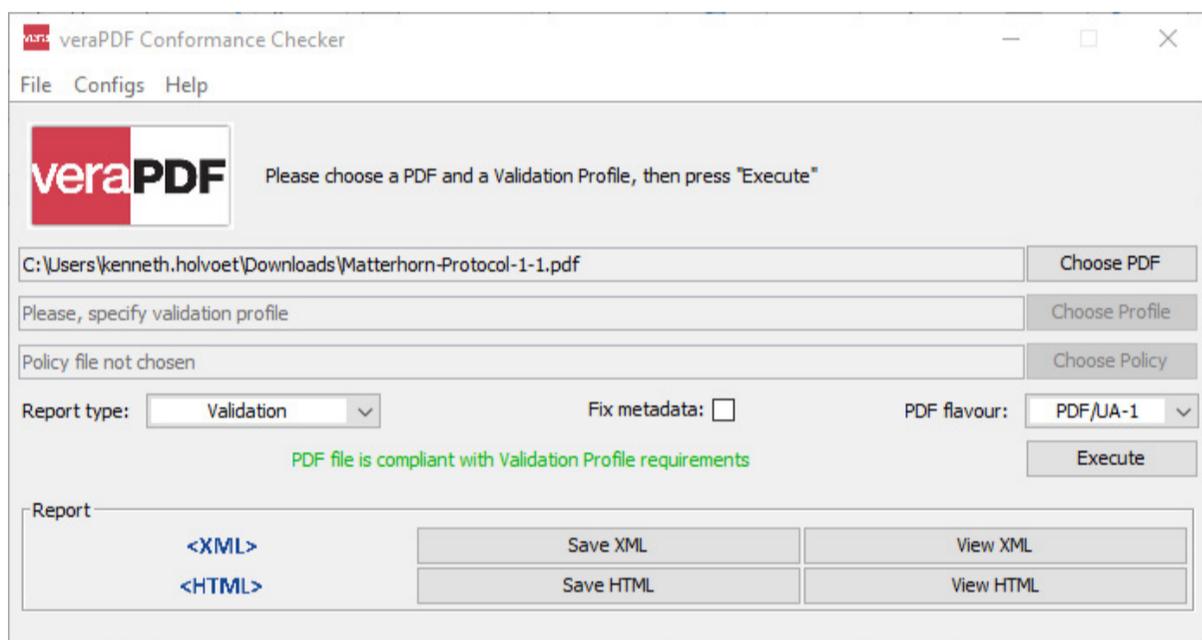
Not surprisingly, the Matterhorn Protocol PDF file passed all (machine identifiable) PDF/UA requirements.

4.

Laws and regulations

PDF/UA is more than just the right thing to do

It is very important for organizations to understand that Universal Accessibility and providing PDF/UA documents is often not just a moral consideration, but one which is mandated by law. One such law addressing the accessibility of all publicly available content, is the Americans with Disabilities Act (ADA). In 2020 alone, approximately 11,000 ADA lawsuits related to accessibility of private businesses were filed of which 3500 relate to digital accessibility. And those numbers are rising every year.

With settlements for individual plaintiffs often adding up to thousands of dollars, it becomes very clear that PDF/UA is not just the right thing to do. Organizations seeking to protect themselves from litigation should know the relevant laws for their area. While many of these legislations refer to WCAG, it is important to note that PDF/UA and WCAG are by no means conflicting standards. Rather, they complement one another: you can think of PDF/UA as a standard mechanism to encode the criteria of WCAG or any content accessibility specification in a way that all PDF/UA processors understand. So, PDF/UA is always the safe choice.

The following list of legislation is by no means a definitive list and does not constitute legal advice.

4.1 USA

Rehabilitation Act Section 508

In 1998 Section 508 was added to the Rehabilitation Act, requiring electronic and information technology to be equally accessible to people with disabilities. In 2018 Section 508 was updated to include new requirements such as the Web Content Accessibility Guidelines (WCAG). Section 508 however does not apply to private organizations.

Americans with Disabilities Act (ADA)

ADA was enacted in 1990 to end discrimination of differing abilities. While there surprisingly is no specific mention of access to digital content, Title III of the ADA requires every owner lesser, or operator of a place of public accommodation to provide equal access. While this can have an open interpretation, legal precedents have shown that this also applies to digital access of private businesses.

4.2 EU

EU Web Accessibility Directive

The EU Web Accessibility Directive applies to public institutions and requires online content to conform to WCAG 2.1 AA standards.

European Accessibility Act

The EAA applies to private companies and was adopted by the EU in June 2019. It is required to be added to members states' national law by June 2025.

While both of these directives are not laws themselves, they do outline what EU member countries should include in national laws.
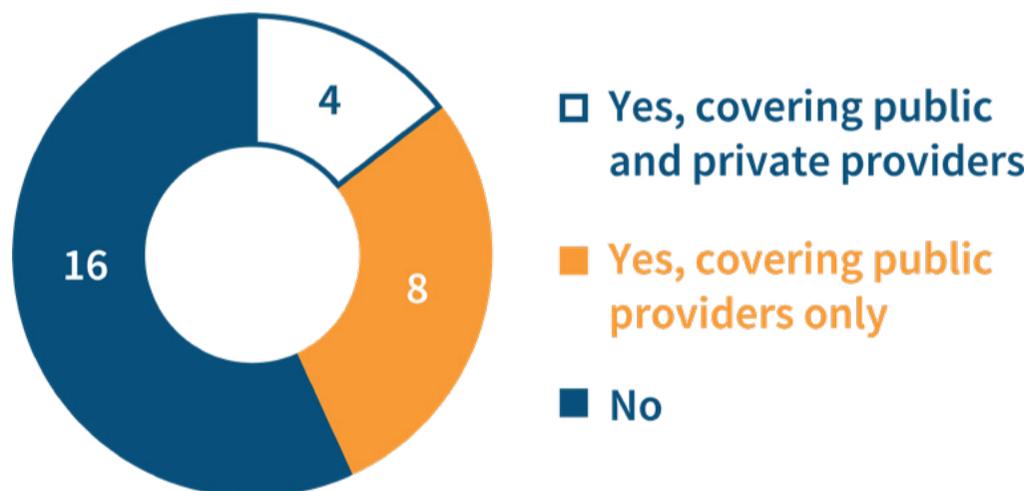
Sixteen EU Member States have no legal accessibility standards in place, eight member states define web accessibility standards for the public sector while only four do so for both public and private sectors. Source: https://fra.europa.eu/en/publication/2014/indicators-right-political-participation-people-disabilities/standards-web

4.3 UK

Disability Discrimination Act

The 2010 act was supplemented with the "Public Sector Bodies (Websites and Mobile Applications) Accessibility Regulations" in 2018 and applies to public institutions. PDFs published before 23 September 2018 are exempt, unless you need them to use a service.

4.4 CANADA

While having multiple accessibility laws on both the federal and provincial level linked to big fines does show Canada's commitment to a more inclusive society, it can be confusing at times. However, it's reassuring to note that PDF/UA meets the compliance requirements of them all.

Accessible Canada Act

The ACA came into effect in July 11, 2019 and aims to make Canada barrier-free and mandate compliance for digital content and technology for public institutions and private organizations under federal jurisdiction such as banks, telecom and transportation companies.

Canada's Standard on Web Accessibility

The standard ensures the uniform application of a high level of web accessibility across Government of Canada websites and web applications since 2011.

Accessibility for Ontarians with Disabilities Act and other Provincial acts

AODA regulates both public, private and non-profit organizations in Ontario. The latter two only have to comply if they have more than 50 employees.

While Ontario was the first to implement an accessibility act with this scope, other provinces such as Nova Scotia and Manitoba followed suit. Quebec has an act that only applies to the public sector.

4.5 INDIA

Rights of Persons with Disabilities Act

This law enacted in 2016 extends the rules for web content of the earlier Guidelines for Indian Government Websites to also apply to private organizations. For documents it requires ePUB or Optical Character Recognition (OCR) enabled PDF. A deadline for compliance was set to June 15, 2019.

4.6 JAPAN

Japanese Industrial Standard X 8341-3

The 2016 issued JIS X 8341-3 provides guidelines for accessibility for older persons and persons with disabilities and includes guidelines web content provided by central and local governments.

5.

Benefits for all users

Improved usability and machine reading

Making your document conform to the PDF/UA standard doesn't just make them accessible to people, the fixed semantics also makes them more **accessible for machine interpretation** and greatly **improves usability** on devices with smaller screens and limited controls. This in turn has advantages for all users, both with and without disabilities, that can improve our everyday experiences.

Thanks to **text-to-speech** it is now possible to let your car or your home assistant (such as Alexa, Siri, Google Assistant and Cortana) read your documents for you, while you can go about your business and have your hands free. PDF/UA documents will already have the right reading order, language detection and alt descriptions for images in place to potentially make it a far better experience.

And what about those devices with limited input controls we use every day? Viewing and navigating complex and inaccessible documents on our **phones and tablets** can be a frustrating experience. Having an accessible PDF/UA document available can already alleviate much of those frustrations by allowing for automatically adjusting of content to readable size. And the Tagged PDF component of PDF/UA even opens up options to evolve to the **next generation of PDF (ngPDF)** making documents fully responsive, so they look good on mobile devices.

Finally, making a PDF document accessible also allows easier web crawling by **search engines** such as Google. This forms an important part of SEO (Search Engine Optimization) and improves the likelihood of your PDF finding its way to your target audience.

5.1 NGPDF

ngPDF stands for the next generation of PDF and focuses on the challenge of making PDFs a first-class citizen of the web, where users consume information through a wide range of devices with diverging screen sizes.

The Tags in PDF and the structure tree are comparable to the HTML tags and the DOM (Document Object Model) they form and this is exactly what ngPDF leverages. It does so by using a derivation algorithm developed by the PDF association that can produce **a reliable HTML presentation of a properly Tagged PDF document**. For example, for a paragraph the PDF structure element *P* translates to a *p* tag in HTML. There is a demo available that serves as a proof of concept and where you can see this algorithm in action. This demo is created by Dual Lab and powered by the iText PDF Library.

You can read more about ngPDF in the whitepaper co-authored by iText and Dual Lab Web-Friendly PDFs with ngPDF.
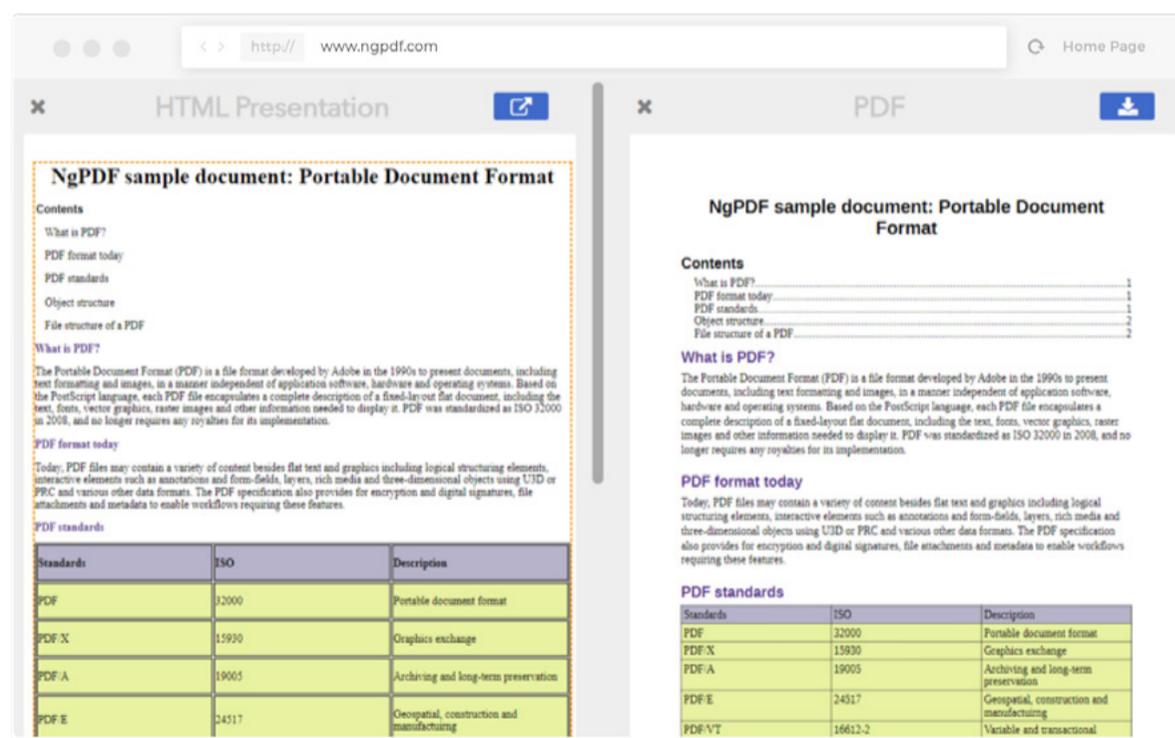
The derivation algorithm that form the core of ngPDF in action in the online demo. By converting Tagged PDF to HTML, we get a responsive solution that is easy to view on all screen sizes.

6.

Creating PDF/UA documents with iText

Using leading PDF software to create accessible PDFs

6.1 ITEXT 7 CORE OPEN-SOURCE SDK/LIBRARY

iText 7 Core offers you the most versatile and well-documented **open-source PDF library**, written for Java and .NET (C#). With two decades of improvements and millions of users, the iText 7 Core PDF library offers you an experience that is both robust and extensive. It can meet all the needs of your document workflow. From creating compliant PDF, to secure digital signing of documents. iText 7 Core does it all, and does it well.

In addition to the ability to create PDFs that are fully compliant with the general PDF standard, iText 7 Core also has the ability to generate PDF/UA (and PDF/A, see our other ebook "PDF/A: digital documents to withstand the sands of time") documents built into its core package.

iText 7 Core provides developers with the necessary tools to automate the creation of accessible PDFs, and with just a few lines of code you can meet the following PDF/UA requirements in your workflows:

- Create Tagged PDF
- Add a language specifier
- Set the title of the document to display in the title bar
- Embed fonts
- Set alt text

Guide:

≡ Java: https://kb.itextpdf.com/home/it7kb/ebooks/itext-7-jump-start-tutorial-for-java/chapter-7-creating-pdf-ua-and-pdf-a-documents

≡ .NET(C#): https://kb.itextpdf.com/home/it7kb/ebooks/itext-7-jump-start-tutorial-for-net/chapter-7-creating-pdf-ua-and-pdf-a-documents-net

Code Example:

https://kb.itextpdf.com/home/it7kb/examples/itext-7-jump-start-tutorial-chapter-7 (use switch to change from Java to .NET)

6.2 ITEXT DITO
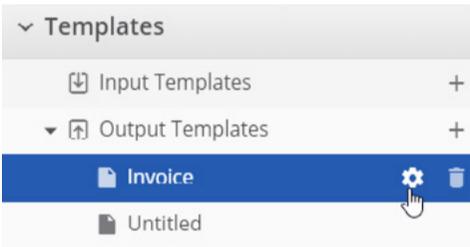
iText DITO is a data-driven, template-based PDF generator that simplifies the process of creating and maintaining PDFs. It is built around an **intuitive visual template designer** and it uses a powerful REST SDK/API for processing and production. You can leverage your data as input and output thousands of documents a day. It also lets you manage all your resources like templates, images and fonts.
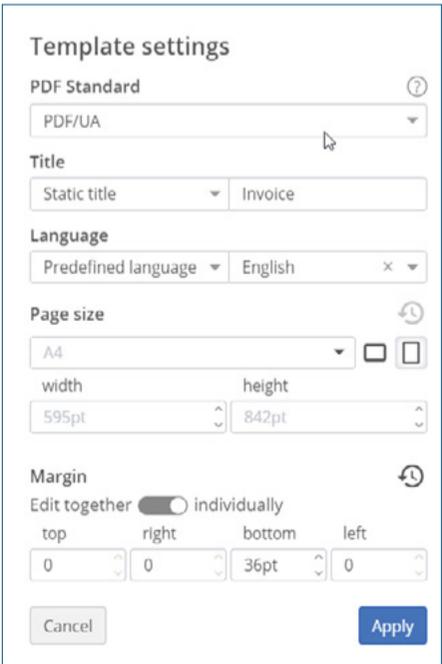
In iText DITO you can indicate in the template settings that you aim to let the resulting documents comply with the PDF/UA standard. This setting activates the PDF/UA compliance assistant, which will help you achieve compliance through a range of property settings and validation checks. To open the template settings panel, click the gear icon of your active template in the Templates panel.

iText DITO will automatically validate your template against PDF/UA requirements where possible and suggest actions to make your template comply.

- Content descriptions for all images (and image-like elements like barcodes) must be specified. The description can be set via the "Alternative description" property.
- Descriptions of all textual hyperlink targets must be specified. The description can be set via the "Link alternative description" property.
- A document title must be specified. The title can be set in the Template Settings.
- A default document language must be specified. It should be the natural language in which the title and the majority of content is written. The language can be set in the Template Settings.
- Semantic row and column headers must be tagged as such in tables. You can do this from the toolbar when a cell/row/column is selected.
- Semantic document headings must be tagged as such. You can do this from the toolbar while editing rich text elements. By default all rich text is treated as a paragraph, but it can be changed to Heading 1-6. You should use heading levels consecutively, e.g., the first heading is always Heading 1, and there are no gaps in a descending heading sequence.

If the template violates one or more of these requirements a warning panel at the bottom of the screen will guide you to address these issues one by one.

As always this should be accompanied by a human check of your template to address the requirements of PDF/UA that can't be machine checked.

Get to know iText

Your boarding pass for your flight. Or an invoice, receipt or form in a PDF format…. Most likely they were generated by iText technology!

iText is a global leader in innovative award-winning PDF software. It is used by millions of users - both open source and commercial - around the world to create digital documents for a variety of purposes: invoices, credit card statements, mobile boarding passes, legal archiving and more.

iText works and works well. Our customers choose iText because of our world-class software quality, and our reliable, mature, and proven technology. We are recognized as a global thought leader and innovator in PDF solutions and functionalities. Our PDF solutions can be embedded into the document workflows of various industries and their applications to enable creation and manipulation of PDFs, and advanced features like secure content redaction, encryption, digital signatures, and ensuring documents are accessible and archivable.

Our diverse customer base includes many of the Fortune 500 companies, as well as small companies and government agencies. We strongly believe in the value of open-source software. Our core library, iText 7, is available under the AGPL license. We also offer commercial licensing for customers that do not wish to comply with AGPL and want to keep their source code private.

VISION

In a world in which speed and efficiency are paramount, we enable companies and people to build the most reliable solutions for document and data exchange, effortlessly.

MISSION

It's our mission to be the most trusted and comprehensive technology provider which perfectly leverages the power of PDF, by offering open-source and enterprise solutions that streamline the generation and consumption of documents and data.

CONTACT

marketing@itextpdf.com
www.itextpdf.com

ITEXT

OUR OFFICES

EUROPE, MIDDLE EAST, AFRICA & CIS

AA Tower
Technologiepark-Zwijnaarde 122
9052 Zwijnaarde
Belgium

sales.isb@itextpdf.com
Tel +32 9 298 02 31
Fax +32 9 270 33 75

AMERICAS

530 Harrison Ave,
Second Floor
Boston, MA 02118
United States

sales.isc@itextpdf.com
Tel +1 617 982 2646
Fax +1 617 982 2647

ASIA & OCEANIA

Republic Plaza
9 Raffles Place, Level 6, Republic Plaza 1
SINGAPORE 048619
Singapore

sales.isa@itextpdf.com
Tel +65 6932 5062

itextpdf.com