

# Next-generation PDF

Meeting today's needs  
Staying true to core values



# PDF 2.0





INTERNATIONAL  
STANDARD

ISO  
32000-2

---

**Document management —  
Portable document format —**

**Part 2:  
PDF 2.0**

*Gestion de documents — Format de document portable —  
Partie 2: PDF 2.0*

**Coming really soon!**



Reference number  
ISO 32000-2:2017(E)

© ISO 2017



- Improved encryption
- State-of-the-art digital signature technology
- Unencrypted wrapper document
- Document parts
- Associated files
- Support for PRC
- Support for geospatial data
- ...and much more



- PDF documents that suit the user's device and viewing preference (text size, font, etc.)
- PDF content that may be easily and reliably reused
- PDF content that may include additional resources (e.g., data for a chart, a metadata schema, an alternative representation) within the document
- Requirements for next-generation PDF processors
- PDF viewers may provide additional styling or dynamic resources to enhance the user experience
- A PDF Association-mediated industry collaboration



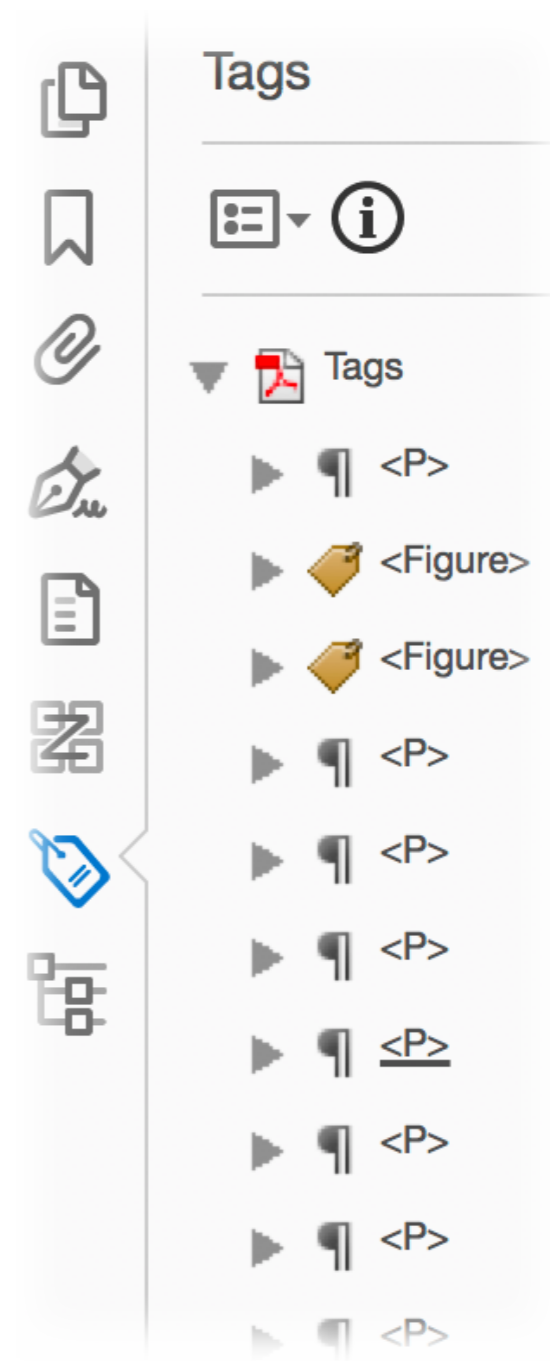
- A new file format  
Next-generation PDF is based on PDF 2.0
- A "website in a can"  
Next-generation PDF is as document-centric as PDF itself
- An Adobe product or technology  
Next-generation PDF is under development by interested members of the PDF Association in a collaborative community setting



- Based on the latest PDF - ISO 32000-2:2017 (PDF 2.0)
  - Tagged PDF revised, enhanced and formalized
  - Namespaces for tagged PDF
  - Associated files
- Processor roles
  - “Derivation” from Tagged PDF: a standardized model for representing PDF content as HTML
  - Support Open Web Platform (OWP) technology in derived HTML content



- Next-generation PDF files are “Tagged PDF”
- The tags and their respective attributes meets high standards for quality:
  - Logical reading-order
  - Semantic appropriateness
  - Distinction from artifacts





- Standard structure elements (“tags”) rationalized and simplified
- 13 structure elements removed, 8 added
- Unambiguous parent/child/sibling rules
- Revised description (function) for most tags
- Revised structure element attributes
- Pronunciation hints
- Namespaces

## Not defined in PDF 2.0

~~Sect~~  
~~Art~~  
~~BlockQuote~~  
~~TOC & TOCI~~  
~~Index~~  
~~NonStruct~~  
~~Private~~  
~~Quote~~  
~~Note~~  
~~Reference~~  
~~BibEntry~~  
~~Code~~

## New in PDF 2.0

DocumentFragment  
 Aside  
 Hn (where n > 6)  
 Title  
 FENote  
 Sub  
 Em  
 Strong  
 Artifact



- Mechanism for identifying and exchanging custom structure types
- Familiar from the XML world
- Many PDF documents are already authored by software with its own rich structures and content
- Namespaces make it possible to readily include MathML or other non-PDF structure elements in the PDF tagging model
- Processors that understand these tags can express the content to the user accordingly



- A means of associating embedded or external files with PDF file objects, and identifying the semantic relationship between them
- Originally introduced with PDF/A-3 in 2012, now part of PDF 2.0
- *Associations* may be created for:
  - The PDF as a whole
  - A PDF page
  - A graphics object (e.g., an image)
  - A structure element in a tagged PDF (e.g., a table)
  - An annotation (e.g., a comment, highlight markup, etc.)



# Possible associations

---

- *Source* (e.g., the Word file used to create the PDF)
- *Data* (e.g., information used to derive a graph)
- *Alternative* (e.g., alternative representation of content)
- *Supplement* (e.g., a MathML version of an equation)
- *Schema* (e.g., an XML schema associated with metadata)
- *Form data* (e.g., data associated with form-fields)
- *Encrypted payload* (see unencrypted wrapper document)



# Next-generation PDF processors



- HTML is “derived” from tagged PDF according to a standardized model (the “derivation algorithm”)

A standardized derivation algorithm allows derived HTML to be created in a predictable and interoperable manner, based on the logical structure in a tagged PDF

- Styling is derived from the PDF’s fonts and layout and supplemented by the viewer with the user’s preferred CSS
- Dynamic features are provided by the viewer in response to data-structures found in the tags, or to associated files. For example, a viewer could provide users with the ability to sort table data, or display supplementary information for the content of a given structure element



- Developers of next-generation PDF processors can leverage next-generation PDF files with web-technology features, including, but not limited to:
  - CSS
  - JavaScript
  - Media Query
  - RDFa
  - ARIA
  - And more...



- Browsers
- Mobile viewers
- Desktop viewers
- Search-engines
- Content and Document Management systems
- Assistive technology (AT) for disabled users





# Use cases



- **The biggest complaint** about PDF in mobile (phone-sized) viewers is the need to pinch, zoom and pan. More than just swiping, this fact negatively impacts:
  - **Navigation** (no help there; on an iPhone PDF bookmarks are a blunt instrument)
  - **Appearance** (printable pages aren't mobile-friendly or user-adjustable)
  - **Search** (inconsistent and missed hits)
- **Interactivity** (PDF page content is “dead”)



# Next-generation PDF on mobile devices

---

- Next-generation PDF makes it possible to easily derive (and thus reflow) PDF content into HTML for diverse screen-sizes – **no more pinching and panning!**
- Next-generation PDF viewers can derive structure (headings, tables, etc.) from the PDF and provide **device-optimized navigation**
- HTML derived from next-generation PDF may be styled via CSS provided by the viewer to provide **rich user preferences in a device-optimized context**
- **Search operates flawlessly**, including across pages
- Viewers can provide **JavaScript for interactivity**, including forms, sorting tables, and much more



# PDF accessibility and the government

---

- Governments in the US, EU, Canada, Australia and elsewhere are requiring accessible content from their agencies and contractors... and are increasingly requiring accessible B2C communications as well
- Today, most Assistive Technology (AT) does not (properly) support PDF... but does support HTML
- Government agencies don't want to give up on the reliability and portability of PDF
- Regulations tend to identify “technology-neutral” WCAG 2.0 Level AA as their performance benchmark, but WCAG 2.0 is most easily applied to HTML, not PDF



- Next-generation PDF may be consumed (including by AT) as HTML. For accessibility purposes, every next-generation PDF file is an HTML page
- Governments need not give up on PDF's existing and much-appreciated value proposition in order to achieve the highest degree of accessibility
- As an HTML representation of the PDF, next-generation PDF may be readily contrasted with the “actual” PDF page and assessed for WCAG 2.0 conformance



# Side-benefit: searching PDF documents

---

- Today, searching PDF documents is often a poor experience:
  - Page content that's unaligned with logical reading order
  - Page-breaks interrupting word or phrase searches
  - No way to distinguish author-intended content from publication artifacts such as running headers
- If search software used Tagged PDF it would find:
  - Logical reading order unambiguously defined and page-breaks managed appropriately (i.e., ignored)
  - Real content distinguished from artifacts such as page running headers, improving search quality
  - Heading (H1..Hn), paragraph, table and other tags that may be leveraged to provide richer search results



- Next-generation PDF means:
  - Mobile-friendly representations of PDF documents
  - Improved navigation (richer Tables of Contents, collapsible heading, etc.) and more accurate, richer searching
  - Interactive "smart" objects
  - Leverage embedded Associated Files



- You can participate in PDF and ISO standards development too!
- Via the PDF Association's "category A" ISO liaison
- Directly, via your own country's ISO Member Body (DIN in Germany, ANSI in the US, Standards Australia in Australia, etc.)
- Join us as we continue to push PDF forwards!

