

Tagged PDF 2.0

NORMEX

Roman Toda,
Normex

What PDF 2.0 tagging brings in the context of format interchange and repurposing?

Roman Toda, Normex

Yulian Gaponenko, Dual Lab



Yulian Gaponenko,
Dual Lab



New in ISO 32000-2

- Major rework
- Removed tags (Standard Structure Types)
 - Grouping elements: Art, BlockQuote, Index, NonStruct, Private, Sect, TOC, TOCI
 - Inline elements: BibEntry, Code, Note (FENote), Quote, Reference
- New tags
 - Artifact, Aside, DocumentFragment, Em, FEnote - replacement for Note
 - H7..Hn, Strong, Sub, Title

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Artifact sample

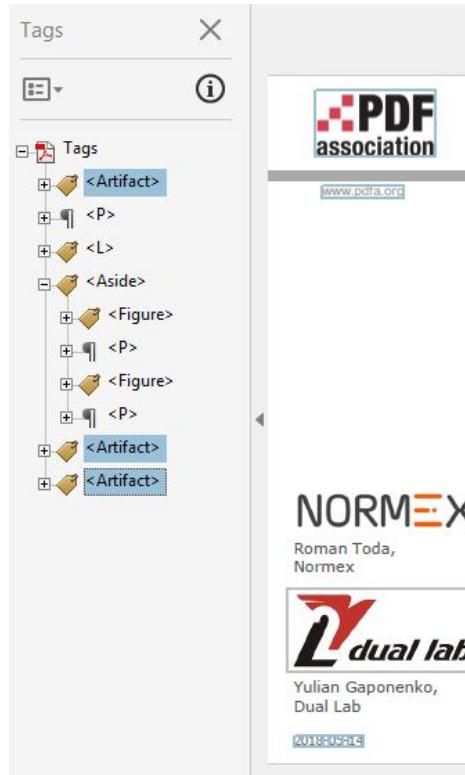
NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

2018-05-14

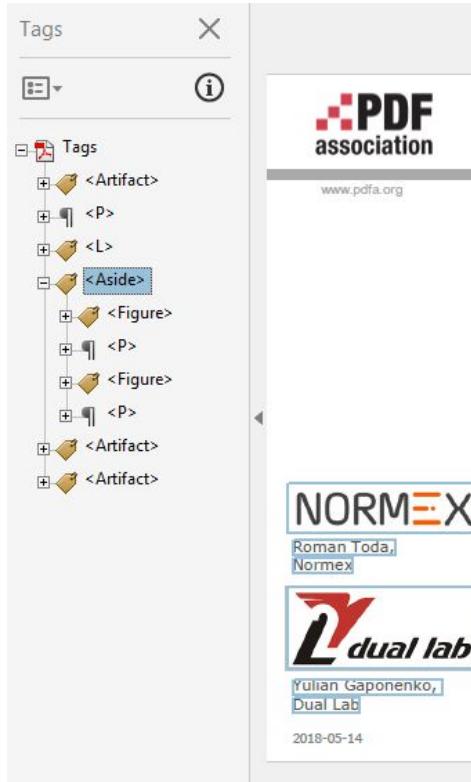


The screenshot shows a PDF editor interface. On the left, a 'Tags' panel lists various document structures: Tags, <Artifact>, <P>, <L>, <Aside>, <Figure>, <P>, <Figure>, <P>, <Artifact>, and <Artifact>. The <Artifact> item is highlighted with a blue selection bar. The main pane displays a PDF page with the 'PDF association' logo at the top, followed by the text 'www.pdfa.org'. Below this is the 'NORMEX' logo with 'Roman Toda, Normex' underneath, and the 'dual lab' logo with 'Yulian Gaponenko, Dual Lab' underneath. At the bottom of the page, there is some small text including '2018-05-14' and a copyright notice.

New in ISO 32000-2

- Major rework
- Removed tags (Standard Structure Types)
 - Grouping elements: Art, BlockQuote, Index, NonStruct, Private, Sect, TOC, TOCI
 - Inline elements: BibEntry, Code, Note (FENote), Quote, Reference
- New tags
 - Artifact, Aside, DocumentFragment, Em, FENote - replacement for Note
 - H7..Hn, Strong, Sub, Title

Aside sample



The screenshot shows a PDF editor interface. On the left, a 'Tags' panel lists the document's structure. It includes a 'Tags' section with icons for 'Artifact', 'P', 'L', and 'Aside'. The 'Aside' section is expanded, showing a nested structure: 'Figure', 'P', 'Figure', 'P', 'Artifact', and 'Artifact'. The main pane displays the PDF content, which consists of three logos: 'PDF association' at the top, 'NORMEX' in the middle, and 'dual lab' at the bottom. Below the logos, there is some text: 'Roman Toda, Normex' under NORMEX, and 'Yulian Gaponenko, Dual Lab' under dual lab. At the bottom of the main pane, the date '2018-05-14' and the copyright notice 'A PDF Association Presentation · © 2018 by PDF Association · www.pdfa.org' are visible.

New in ISO 32000-2

- Major rework
- Removed tags (Standard Structure Types)
 - Grouping elements: Art, BlockQuote, Index, NonStruct, Private, Sect, TOC, TOCI
 - Inline elements: BibEntry, Code, Note (FENote), Quote, Reference
- New tags
 - Artifact, Aside, DocumentFragment, Em, FENote - replacement for Note
 - H7..Hn, Strong, Sub, Title

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

New attribute owners

- Artifact
- HTML-5.00
- CSS-3.00
- RDFa-1.10
- ARIA-1.1
- NSO
 - use /NS to define namespace

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

2018-05-14



New attribute - ListNumbering

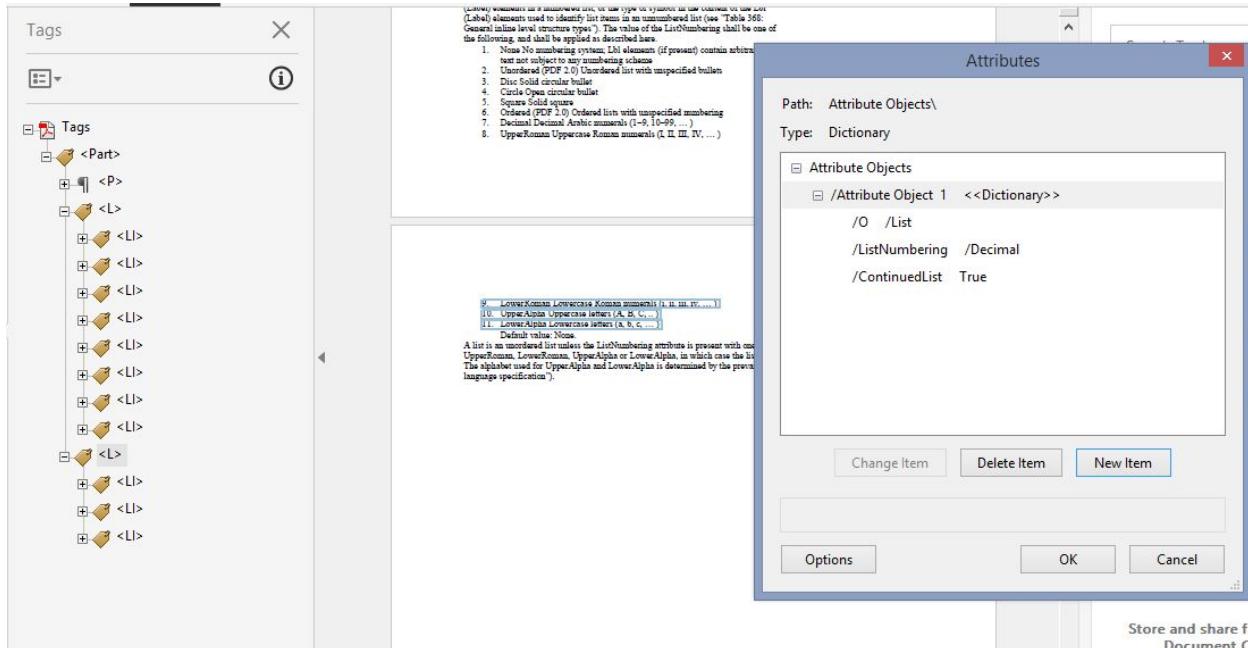
NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

2018-05-14



The screenshot shows a software interface for managing PDF structure. On the left, the 'Tags' panel displays a hierarchical tree of XML tags, primarily `<L>` elements under a `<Part>` container. In the center, a detailed description of the `/ListNumbering` attribute is provided, stating it is used to identify list items in an ordered list. It lists nine bullet types: None, Unordered, Ordered, Circle, Square, Decimal, Roman, Arabic, and Upper/Roman. Below this, a list of values for the `/ListNumbering` attribute is shown: Lower-Comma, LowerAlpha, LowerKoma, LowerAlphaKoma, Upper-Period, UpperAlpha, and LowerAlpha. A note specifies that if no value is present, the default is None. On the right, the 'Attributes' dialog box is open, showing the path `/Attribute Objects\` and type `Dictionary`. It contains one item: `/Attribute Object 1 <<Dictionary>>`, which defines `/List` with `/ListNumbering /Decimal` and `/ContinuedList True`. Buttons for `Change Item`, `Delete Item`, `New Item`, `Options`, `OK`, and `Cancel` are visible at the bottom.

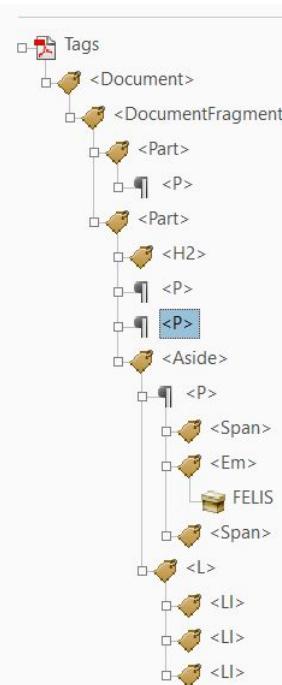


NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



adipiscing elit, sed do eiusmod tempor
incididunt ut labore et dolore magna aliqua.
Ut enim ad minim veniam, quis nostrud
exercitation ullamco laboris nisi ut aliquip
ex ea commodo consequat. Duis aute irure
dolor in reprehenderit in voluptate velit esse
cillum dolore eu fugiat nulla pariatur.
Excepteur sint occaecat cupidatat non
proident, sunt in culpa qui officia deserunt
mollit anim id est laborum.

Ac orci phasellus egestas

Lorem ipsum dolor sit amet, consectetur
adipiscing elit, sed do eiusmod tempor
incididunt ut labore et dolore magna aliqua.
Scelerisque in dictum non consectetur a erat
nam at. Imperdiet massa tincidunt nunc
pulvinar sapien et ligula. Amet nisl purus in
mollis nunc sed. Duis ut diam quam nulla
porttitor massa id neque. Fames ac turpis
egestas integer. Mauris a diam maeccenas
sed enim ut sem viverra aliquet. Nunc non
blandit massa enim nec dui nunc mattis. At
risus viverra adipiscing at in tellus integer.
Ridiculus mus mauris vitae ultricies. Nibh
sit amet commodo nulla facilisi nullam
vehicula ipsum a. Mauris rhoncus aenean
vel elit scelerisque mauris pellentesque. Non
odio euismod lacinia at quis risus sed. Elit
ut aliquam purus sit amet luctus venenatis
lectus magna. Fringilla phasellus faucibus
scelerisque eleifend. A iaculis at erat
pellentesque adipiscing commodo elit at.

Quam ut venenatis tenus in metus vapitate.
Pellentesque elit ullamcorper dignissim cras
tincidunt lobortis. Tellus at urna
condimentum mattis pellentesque id. Urna
cursus eget nunc scelerisque viverra mauris
in. Euismod nisi porta lorem mollis aliquam
ut porttitor leo. Mauris pellentesque
pulvinar pellentesque habitant morbi
Pretium nibh ipsum consequat nisl.
Dignissim sodales ut eu sem integer vitae.
Lorem ipsum dolor sit amet consectetur
adipiscing elit ut aliquam. Id cursus metus
aliquam eleifend mi. Penatibus et magnis
dis parturient montes nascetur. Vitae aliquet
nec ullamcorper sit. Semper risus in
hendrerit gravida rutrum quisque non tellus.
Penatibus et magnis dis parturient montes
nascetur. Dignissim cras tincidunt lobortis
feugiat vivamus at augue eget arcu.
Tristique senectus et netus et. Risus at
ultrices mi tempus imperdiet nulla
malesuada. Cursus turpis massa tincidunt
dui. Lobortis elementum nibh tellus
molestie. Eget mi proin sed libero enim.

SED FELIS EGEST VELIT

- Aliquet sagittis id consectetur
elementum
- Purus ut faucibus pulvinar
- Integer enim neque volutpat ac
tincidunt vitae

What's inside?

9 0 obj <<

/Type /StructElem

/NS 48 0 R

/S /P

/P 8 0 R

/K [...]

>>

Namespace
dictionary

4 0 obj <<

/Type /StructTreeRoot

47 0 R

/Namespaces [48 0 R]

/K [8 0 R 10 0 R]

...

>>

Associated
files

10 0 obj <<

/Type /StructElem

/NS 48 0 R

/S /P

/P 4 0 R

/K [...]

/Ref [...]

>>

Reference
structure
elements

8 0 obj <<

/Type /StructElem

/NS 48 0 R

/S /Div

/P 4 0 R

/K [...]

/AF 101 0 R

>>

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

Namespaces crash course

- every structure type belongs to a namespace;
- several namespaces may be specified in document;
- every namespace can have own role mappings;
- namespaces are defined by the *Namespace dictionary*.

 NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Namespace dictionary

48 0 obj <<

/Type /Namespace

/NS (http://iso.org/pdf2/ssn)

/RoleMapNS <<

 /CUSTOMROLE [/Div 48 0 R] % ind ref to current NS dict

>>

/Schema 67 0 R % ind ref to a Filespec dict of the schema file

>>

endobj

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Standard structure namespaces

- Standard structure namespace for PDF 2.0
"<http://iso.org/pdf2/ssn>" - ISO 32000-2 Clause 14.8, "Tagged PDF"
- Standard structure namespace for PDF 1.7
"<http://iso.org/pdf/ssn>" - ISO 32000-1 Clause 14.8, "Tagged PDF"

NORMEX

Roman Toda,
Normex

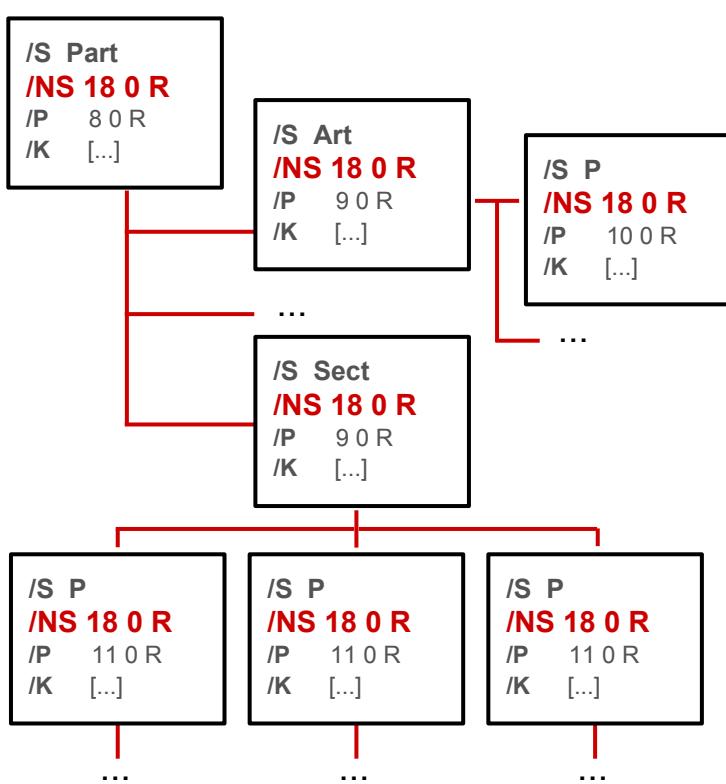
...the default standard structure namespace shall be "<http://iso.org/pdf/ssn>".



Yulian Gaponenko,
Dual Lab



Standard structure namespace for PDF 1.7



18 0 obj <<

/Type /Namespace

/NS (<http://iso.org/pdf/ssn>)

>>

endobj

NORMEX

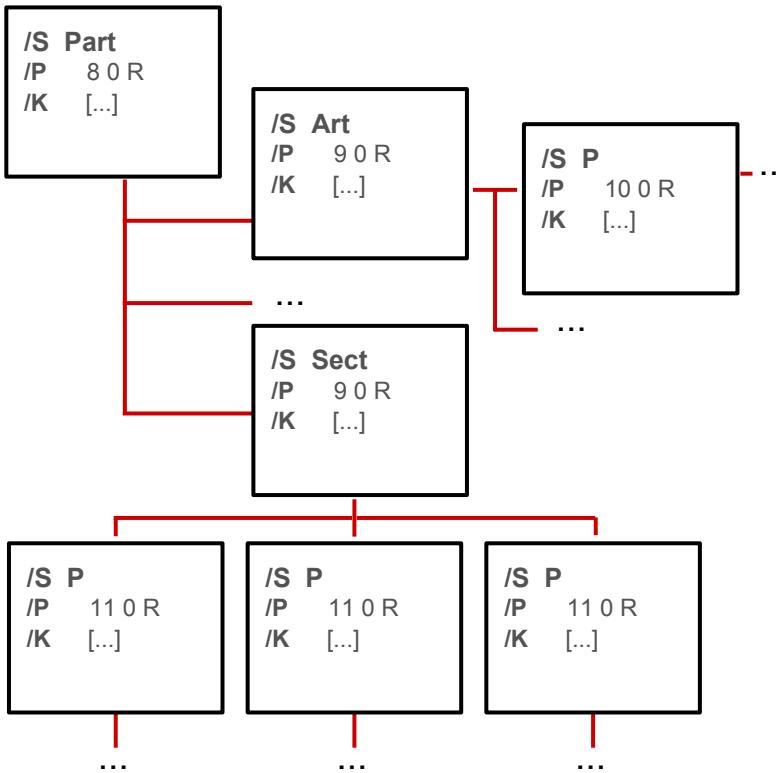
Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Standard structure namespace for PDF 1.7



NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



What possibilities do namespaces provide?

- standardized way of reusing domain specific languages;
- more data preserved;
- facilitates accessibility and extractability.

NORMEX

Roman Toda,
Normex



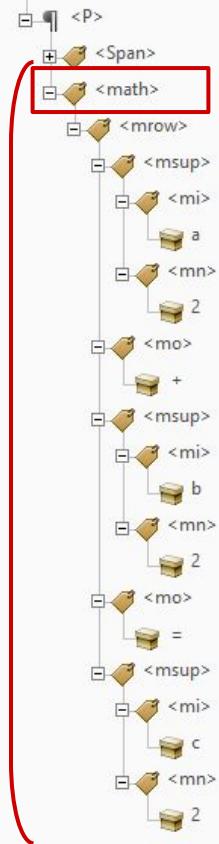
Yulian Gaponenko,
Dual Lab



Domain specific namespaces. MathML

In a right angled triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides. i.e, If a and b are the legs and c is the hypotenuse then

$$a^2 + b^2 = c^2$$



```
11 0 obj <<
/Type /StructElem
/S /math
/NS 28 0 R
/P 8 0 R
/K 12 0 R
>>
```

```
28 0 obj <<
/Type /Namespace
/NS (http://www.w3.org/1998/Math/MathML)
>>
```

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Structure destination

- Destination defines a place in a document (same or remote)
 - 1.7 Explicit (page)
 - 1.7 Named Destination
- 2.0 Structure Destination - associated with Structure Element
 - through object reference
 - Through ID in SE (when GotoR is used)

NORMEX

Roman Toda,
Normex



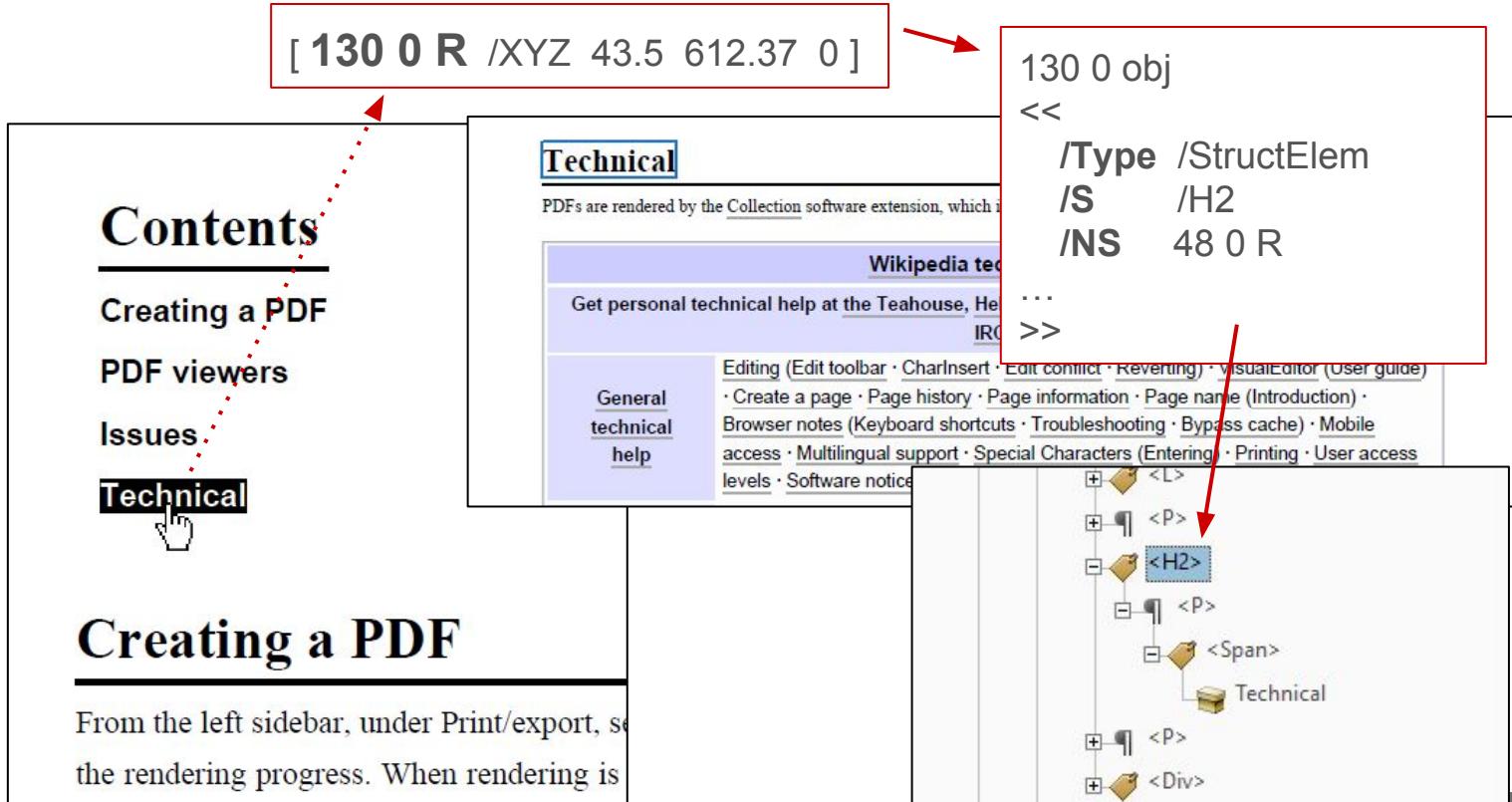
Yulian Gaponenko,
Dual Lab

2018-05-14



Structure destination - sample

[130 0 R /XYZ 43.5 612.37 0]



The diagram illustrates the mapping between a PDF structure and its XML representation. On the left, a sidebar shows a 'Contents' menu with 'Technical' selected. The main area displays a 'Creating a PDF' page. A red box highlights the coordinate [130 0 R /XYZ 43.5 612.37 0]. An arrow points from this coordinate to a detailed view of the PDF structure on the right. This view shows a 'Technical' section with various links and sub-sections. A red box highlights the object reference '130 0 obj'. Another arrow points from this reference to the corresponding XML structure, which is a 'StructElem' node with type '/H2'. The XML structure also includes attributes like '/S' and '/NS'.

```

130 0 obj
<<
/Type /StructElem
/S /H2
/NS 48 0 R
...
>>

```

Technical

PDFs are rendered by the [Collection](#) software extension, which i

[Wikipedia tec](#)

Get personal technical help at the [Teahouse](#), [Help](#) or [IRC](#).

[General](#) [technical](#) [help](#)

Editing ([Edit toolbar](#) · [CharInsert](#) · [Edit commit](#) · [Reverting](#)) · [visualEditor](#) ([User guide](#)) · [Create a page](#) · [Page history](#) · [Page information](#) · [Page name](#) ([Introduction](#)) · [Browser notes](#) ([Keyboard shortcuts](#) · [Troubleshooting](#) · [Bypass cache](#)) · [Mobile access](#) · [Multilingual support](#) · [Special Characters](#) ([Entering](#) · [Printing](#) · [User access levels](#) · [Software notices](#))

<L>

<P>

<H2>

<P>

Technical

<P>

<Div>

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

Associated files in 2.0

- /AF key could be used “anywhere”
 - Catalog, Page, Xobject, Annotation, Graphics Object (through marked content)
 - Structure Element
 - Points to File Specification dictionary
 - /AFRelationship **shall** be used in FS
 - Doesn’t mandate anything when it comes to processing
 - **Source, Data, Alternative, Supplement, EncryptedPayload, FormData, Schema, Unspecified**

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



The screenshot shows a software interface with a sidebar on the left containing various icons and a tree view of HTML tags. The tree view is expanded to show a table structure with rows for <Table>, <TR>, and <TD>. To the right of the sidebar is a table with four columns: Segment, Employed 2000, Employed 2013, and Change (%). The data in the table is as follows:

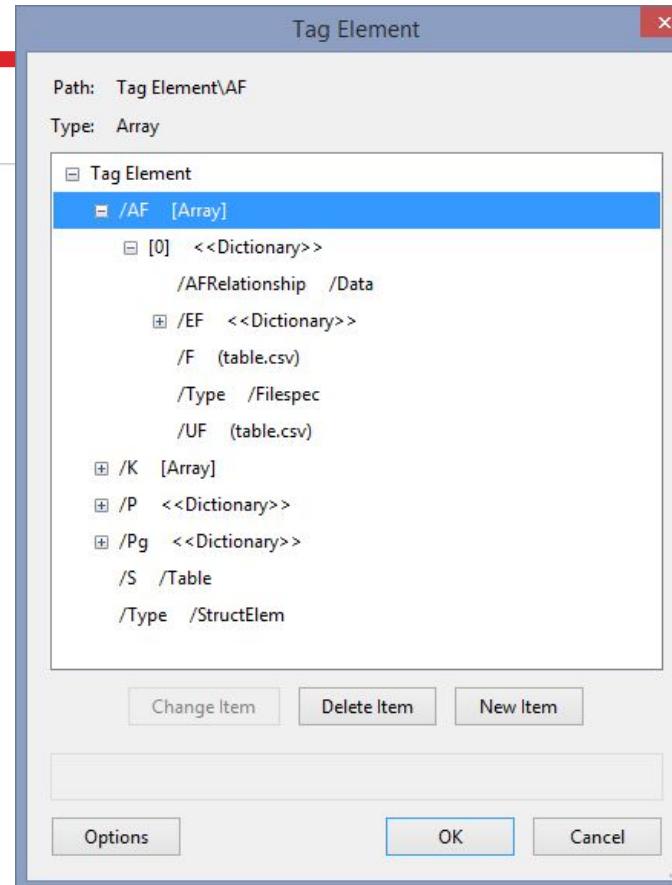
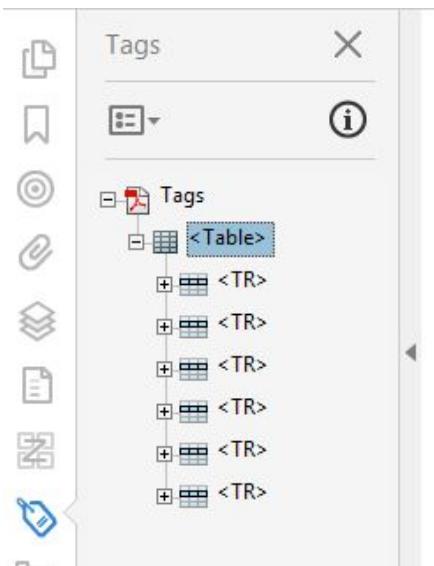
Segment	Employed 2000	Employed 2013	Change (%)
Overall	894	950	+6.4%
No degree	88	70	-21.3%
HS, no college	280	250	-10.0%
Some college	247	260	+5.4%
Bachelor's +	278	370	+33.0%

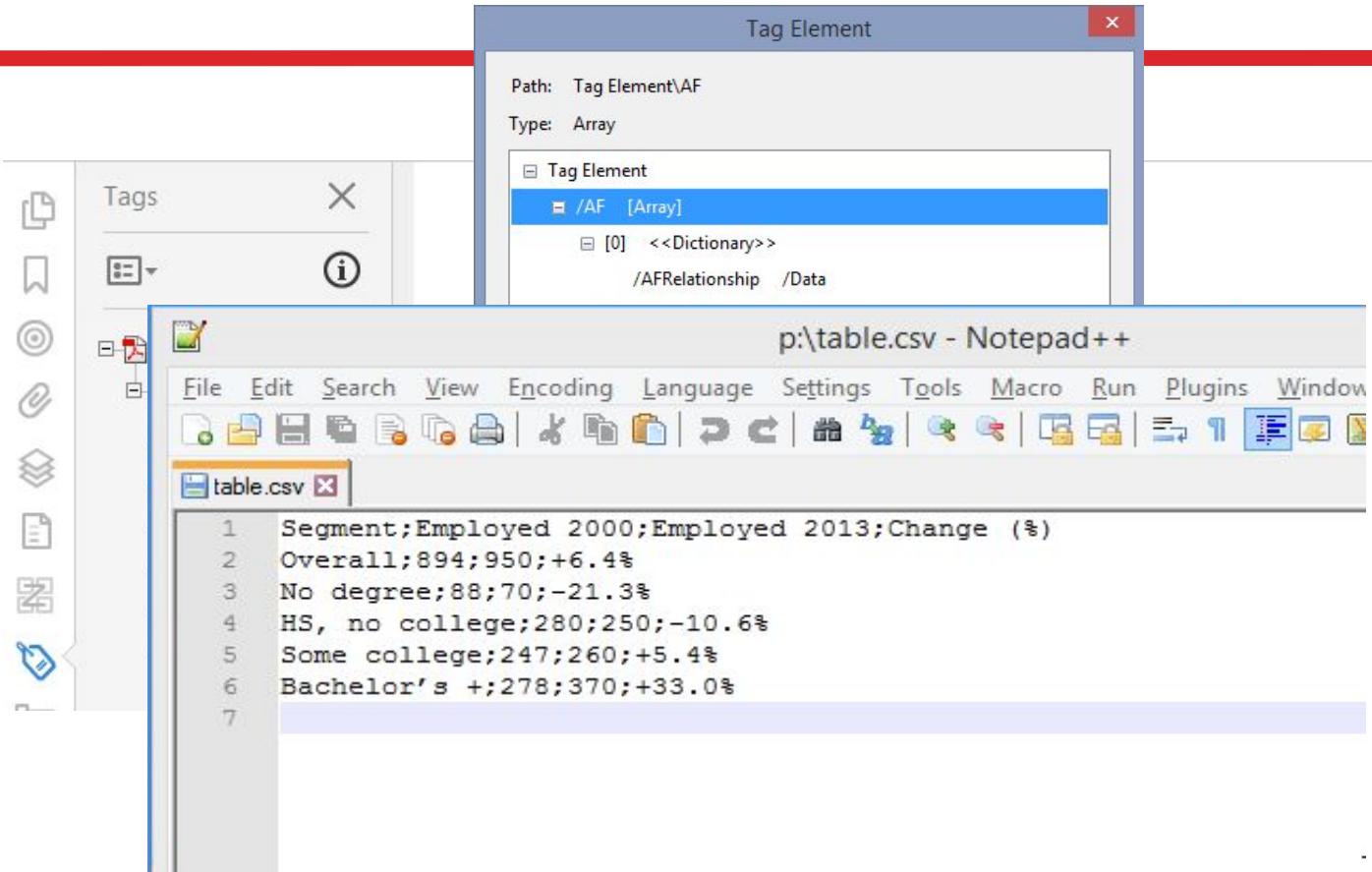
NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

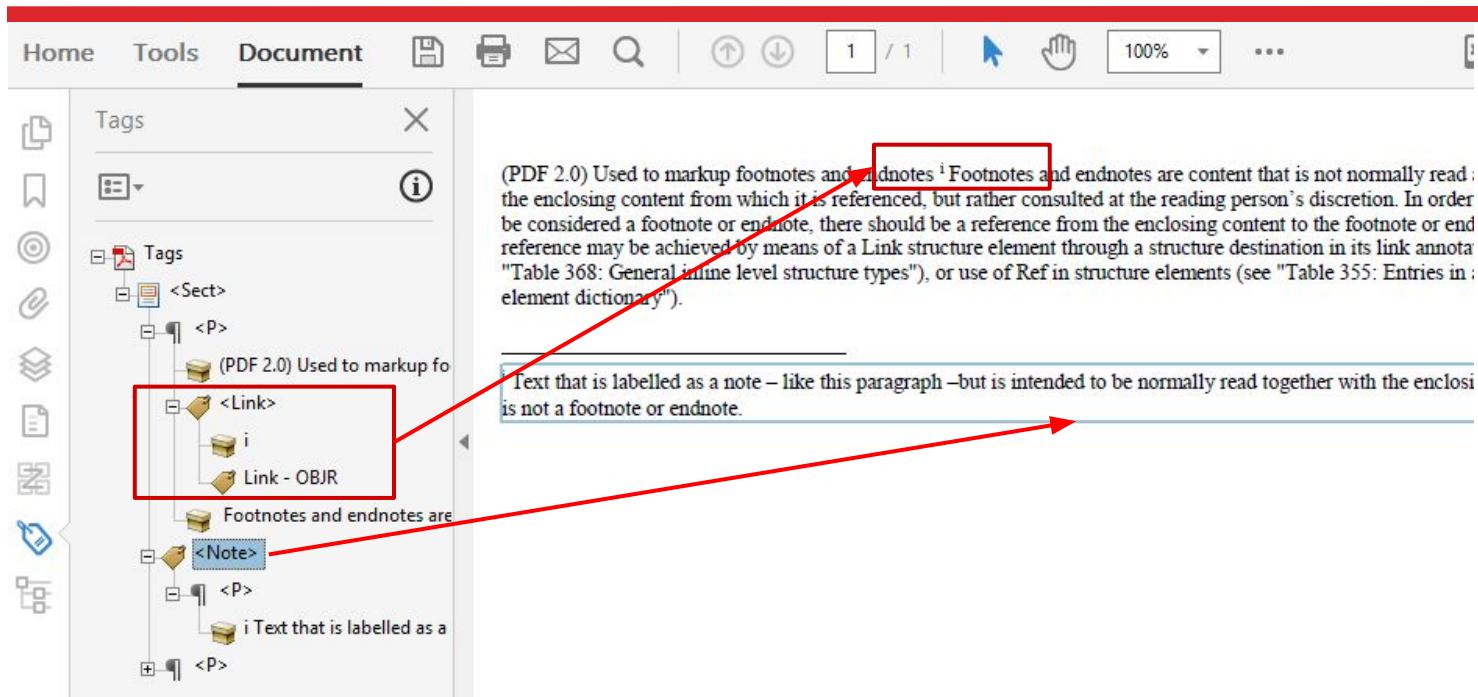




The screenshot shows a PDF editor interface with a 'Tag Element' dialog open. The dialog displays the path 'Tag Element\AF' and the type 'Array'. It contains a tree view with nodes: '/AF [Array]' and '[0] <<Dictionary>>'. Below the tree, the path '/AFRelationship /Data' is shown. In the foreground, a 'Notepad++' window is open, displaying a CSV file named 'table.csv'. The content of the CSV file is:

Segment	Employed 2000	Employed 2013	Change (%)
Overall	894	950	+6.4%
No degree	88	70	-21.3%
HS, no college	280	250	-10.6%
Some college	247	260	+5.4%
Bachelor's	278	370	+33.0%

Tagging Note in 1.7 using Link



The screenshot shows the Adobe Acrobat Pro DC interface with the 'Tags' panel open on the left. The 'Document' tab is selected in the top menu bar. The main workspace displays a note with the following text:

(PDF 2.0) Used to markup footnotes and endnotesⁱ Footnotes and endnotes are content that is not normally read from the enclosing content from which it is referenced, but rather consulted at the reading person's discretion. In order to be considered a footnote or endnote, there should be a reference from the enclosing content to the footnote or end note. This may be achieved by means of a Link structure element through a structure destination in its link annotation ("Table 368: General inline level structure types"), or use of Ref in structure elements (see "Table 355: Entries in element dictionary").

Text that is labelled as a note – like this paragraph – but is intended to be normally read together with the enclosing content is not a footnote or endnote.

Annotations in the screenshot include:

- A red box highlights the text "(PDF 2.0) Used to markup fo".
- A red box highlights the text "Footnotes and endnotes are".
- A red arrow points from the first red box to the second red box, indicating a relationship between the two highlighted sections.
- A red arrow points from the bottom text "Text that is labelled as a note – like this paragraph – but is intended to be normally read together with the enclosing content is not a footnote or endnote." to the "Footnotes and endnotes are" section in the Tags panel.

NORMEX

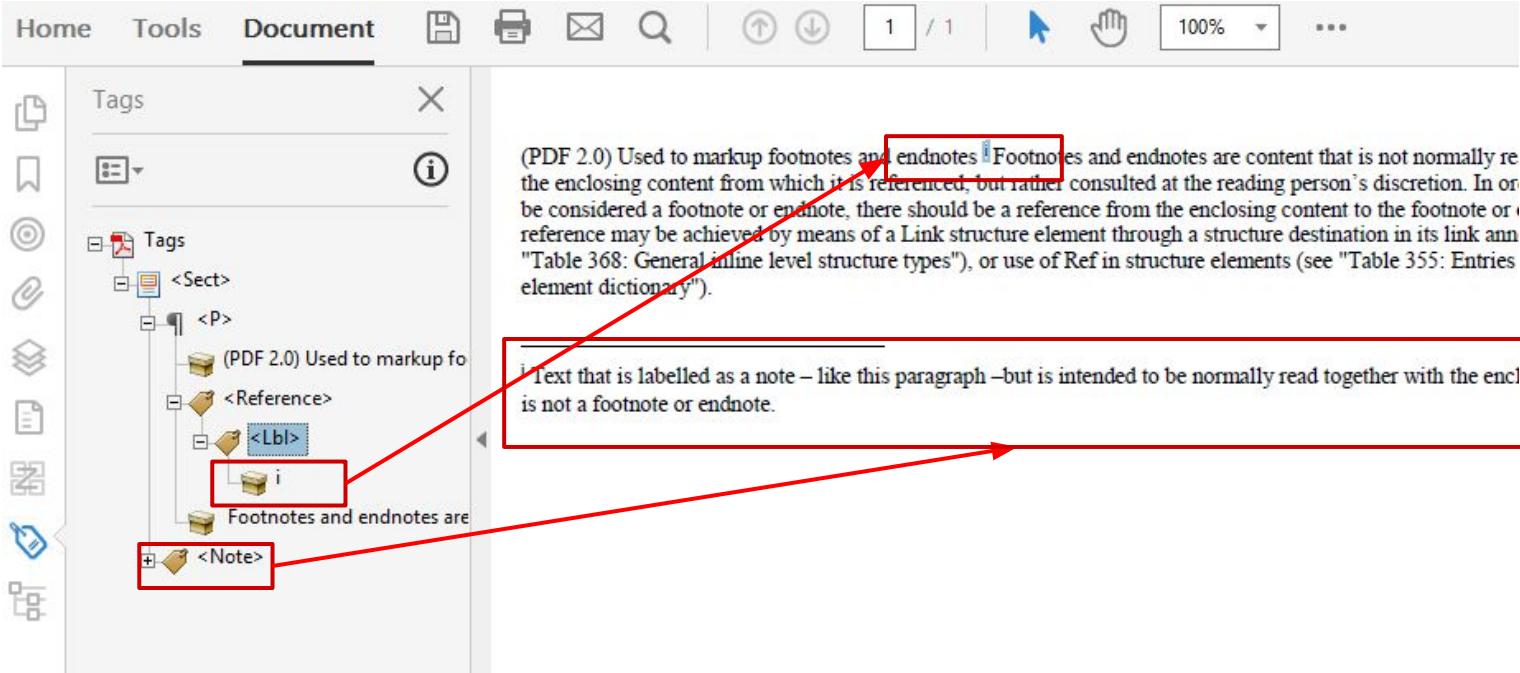
Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab



Tagging Note in 1.7 using Reference



(PDF 2.0) Used to markup footnotes and endnotes. Footnotes and endnotes are content that is not normally read together with the enclosing content from which it is referenced, but rather consulted at the reading person's discretion. In order to be considered a footnote or endnote, there should be a reference from the enclosing content to the footnote or endnote. This reference may be achieved by means of a Link structure element through a structure destination in its link annotation ("Table 368: General inline level structure types"), or use of Ref in structure elements (see "Table 355: Entries element dictionary").

Text that is labelled as a note – like this paragraph – but is intended to be normally read together with the enclosing content from which it is referenced, but rather consulted at the reading person's discretion.

NORMEX

Roman Toda,
Normex



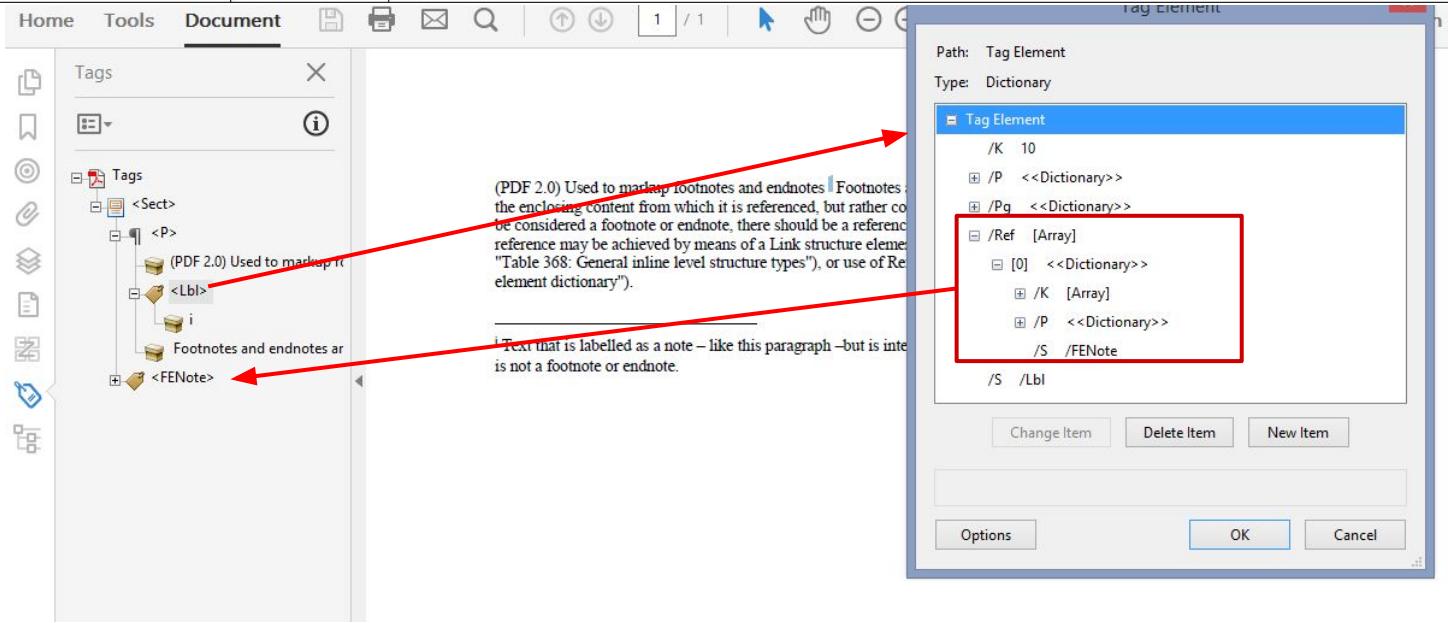
Yulian Gaponenko,
Dual Lab



/Ref - new language in 2.0

www.pdfa.org

Ref	array	(Optional; PDF 2.0) An array containing zero, one or more indirect references to structure elements.
A Ref identifies the structure element or elements to which the item of content, contained within this structure element, refers (e.g. footnotes, endnotes, sidebars, etc.).		



The screenshot shows a PDF editing interface with a 'Tags' panel on the left and a 'Tag Element' dialog box on the right.

Tags Panel:

- Shows a tree structure under the 'Tags' tab.
- Root node: <Sect>
- Child nodes: <P>, <Lb>, <FENote>.
- Details for <FENote> node:
 - (PDF 2.0) Used to markup footnotes and endnotes. Footnotes; the enclosing content from which it is referenced, but rather considered a footnote or endnote, there should be a reference may be achieved by means of a Link structure element ("Table 368: General inline level structure types"), or use of Ref element dictionary".
 - Text that is labelled as a note – like this paragraph – but is intended is not a footnote or endnote.

Tag Element Dialog:

- Path: Tag Element
- Type: Dictionary
- Structure:
 - <Tag Element>
 - /K 10
 - /P <<Dictionary>>
 - /Pg <<Dictionary>>
 - /Ref [Array]** (highlighted with a red box)
 - [0] <<Dictionary>>
 - /K [Array]
 - /P <<Dictionary>>
 - /S /FENote
 - /S /Lb

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

Annex L -Defines relationship between tags

Matrix representation of ISO 32000-2, Annex L, "Parent-child relationships between standard structure elements"

This table is the source data for the table in Annex L



NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

Annex L -Defines relationship between tags

This table is the source data for the table in Annex L.

→ parent	StructTreeRoot	Document	DocumentFragment	Part	Div	Aside	Title	Sub	P	Hn	H	Ibl	Em	Strong	Span	Link	Annot	Form	Ruby	RB	RT	RP	Wrt	WP	FENote	L				
↓ child																														
Document	0..1	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
DocumentFragment	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
Part	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
Div	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n				
Aside	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
P	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
Hn	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
H	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
Title	0	0..n	0..n	0..n	0..n	0..n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0..n				
Sub	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n				
Ibl	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n				
Em	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Strong	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Span	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Link	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Annot	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Form	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n			
Ruby	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n		
RB	0	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n								
RT	0	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n								
RP	0	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n								
Wrtchu	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	
WT	0	0	0	0	0	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
FENote	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n

FENote

L	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
Table	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
TR	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
TH	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
TD	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
THead	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
TBody	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
TFoot	0	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
Caption	0	0..n	0..1..1	0..1..1	0..1..1	0..1..1	0..n																								
Figure	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
Formula	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
Artifact	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
actual content (no child elements)	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n
OBJR content	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n	0..n



Annex L -Defines relationship between tags

This table is the source data for the table in Annex L

NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

HTML conversion to PDF

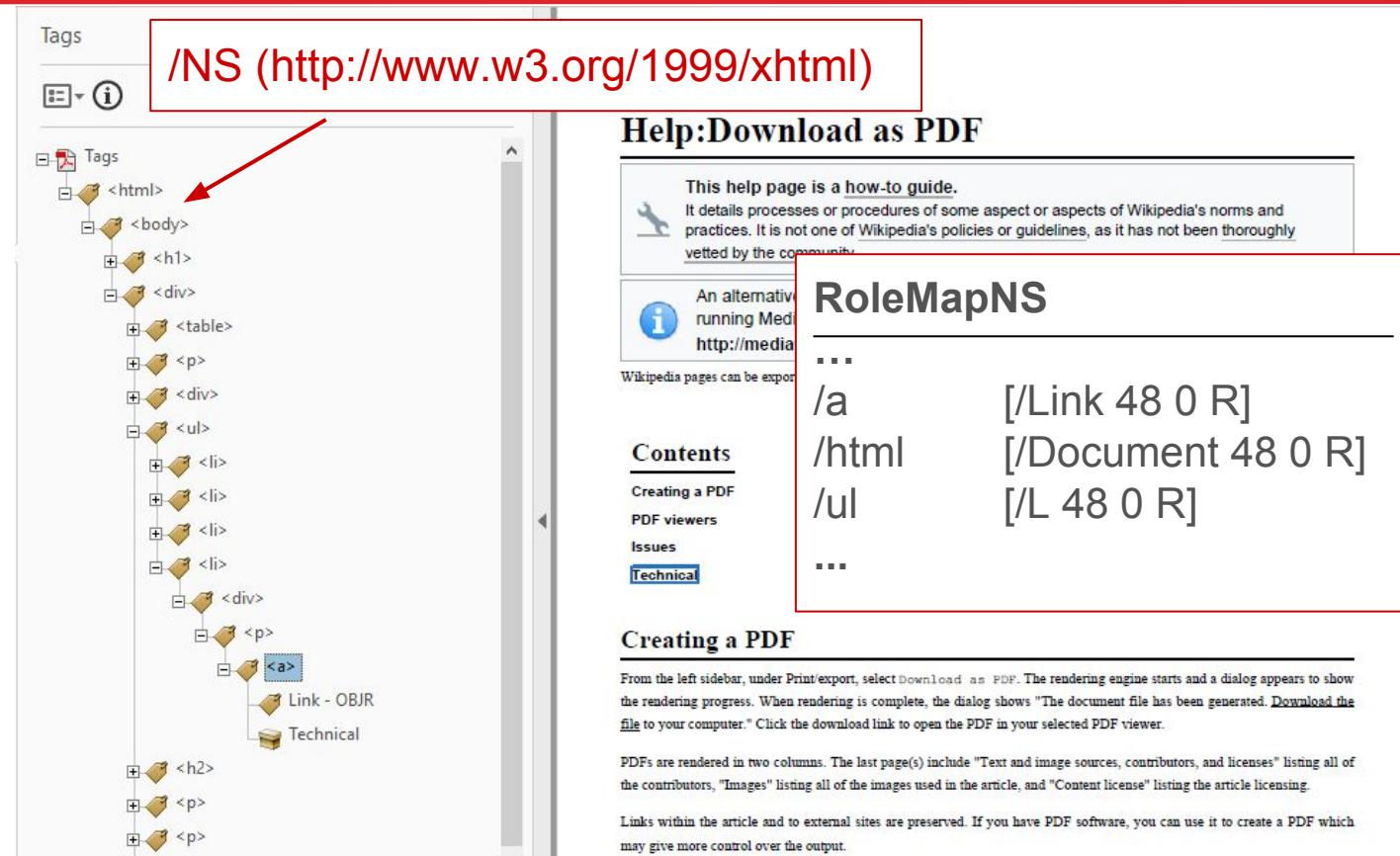
NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

2018-05-14



The screenshot shows a software interface for PDF conversion. On the left, a tree view labeled "Tags" displays the hierarchical structure of an HTML document. A red arrow points from the text box above to the "<html>" node in the tree. The text box contains the URL [/NS \(http://www.w3.org/1999/xhtml\)](http://www.w3.org/1999/xhtml). To the right, a help page titled "Help:Download as PDF" is displayed. The page includes a sidebar with links like "Creating a PDF", "PDF viewers", "Issues", and "Technical". A red box highlights the "RoleMapNS" section, which lists several PDF objects corresponding to HTML elements: "/a [Link 48 0 R]", "/html [Document 48 0 R]", and "/ul [/L 48 0 R]."

This help page is a how-to guide.
It details processes or procedures of some aspect or aspects of Wikipedia's norms and practices. It is not one of Wikipedia's policies or guidelines, as it has not been thoroughly vetted by the community.

An alternative running Mediawiki
<http://media>

Wikipedia pages can be exported as PDFs.

RoleMapNS

...

/a	[Link 48 0 R]
/html	[Document 48 0 R]
/ul	[/L 48 0 R]

...

Creating a PDF

From the left sidebar, under Print/export, select Download as PDF. The rendering engine starts and a dialog appears to show the rendering progress. When rendering is complete, the dialog shows "The document file has been generated. Download the file to your computer." Click the download link to open the PDF in your selected PDF viewer.

PDFs are rendered in two columns. The last page(s) include "Text and image sources, contributors, and licenses" listing all of the contributors, "Images" listing all of the images used in the article, and "Content license" listing the article licensing.

Links within the article and to external sites are preserved. If you have PDF software, you can use it to create a PDF which may give more control over the output.

A PDF Association Presentation · © 2018 by PDF Association · www.pdfa.org

HTML conversion to PDF

www.pdfa.org

/O /CSS-3.00
/list-style-type /none
/overflow /hidden

/O /HTML-5.00
/href (#Technical)
/hreflang (en-us)

Tags

- Tags
- <html>
-
 -
 -
 -
 -
 -
 - <div>
 - <p>
 - <a>
- <h2>
- <p>
- <p>
- <p>

practices. It is not one of Wikipedia's policies currently vetted by the community.

An alternative option is to run MediaWiki on a local server at <http://mediawiki2018.pdfa.org>. Wikipedia pages can be exported and imported into PDF files.

Contents

- Creating a PDF**
- PDF viewers**
- Issues**
- Technical**

/AF

[<<
/Type Filespec
/EF 415 0 R
/F (styles.css)
/UF (styles.css)
/AFRelationship /Supplement
>>]

...
blockquote {
 overflow: hidden;
 margin: 1em 0;
 padding: 0 40px
}
strong.selflink {
 font-weight: 700
}
...
the rendering progress. When rendering is complete, the dialog will

A PDF Association Presentation · © 2018 by PDF Association · www.pdfa.org

NORMEX

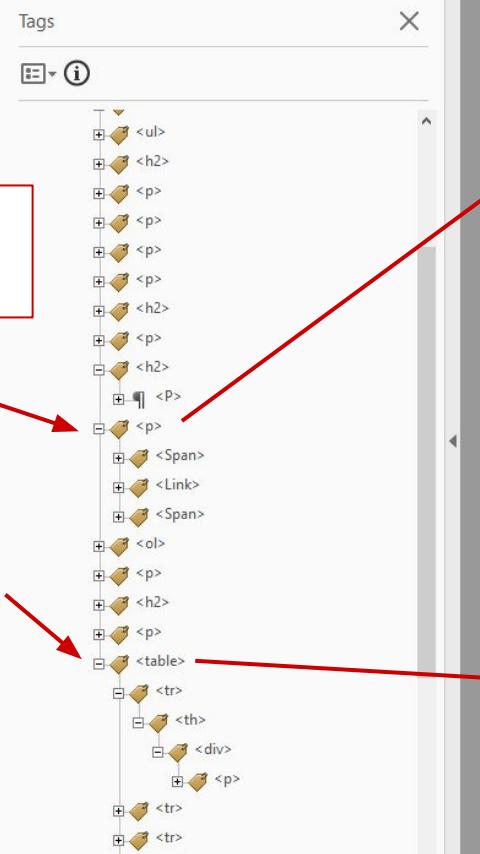
Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

**/O /ARIA-1.1
/role /note**

**/O /ARIA-1.1
/role /navigation
/aria-labelledby ("Wikipedia_technical_help")**



For browser support, see [Comparison of web browsers](#) § [Image format support](#). For PDF app software.

Issues

For more information, see the [list of bugs on Phabricator](#)

1. PDFs are rendered in two columns; this cannot be changed
2. The table of contents is not rendered
3. Tables, including infoboxes, are not rendered; see [T85777](#), [T73296](#), [T73311](#)
4. Columns in the article are rendered as a single column within the PDF columns
5. Grouped references do not render; see [T41095](#)
6. Many CSS properties are ignored, causing template issues; see [T32702](#)
7. `<ruby>` elements do not render; see [T23504](#)
8. Long equations are overflowing

See [T134205](#) for a method to generate PDF files proposed in May 2016.

Technical

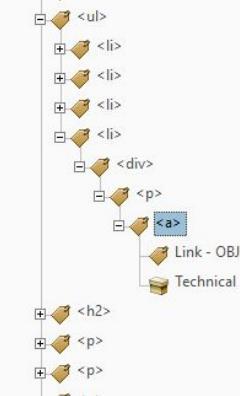
PDFs are rendered by the Collection software extension, which in turn uses the PDF Writer extensi

Wikipedia technical help	
Get personal technical help at the Teahouse, Help desk, Village pump (technical), IRC.	
General technical help	Editing (Edit toolbar · CharInsert · Edit conflict · Reverting) · Vis Create a page · Page history · Page information · Page name Browser notes (Keyboard shortcuts · Troubleshooting · Bypass access · Multilingual support · Special Characters (Entering) · Page levels · Software notices (Editnotice) · IRC (Tutorial))



HTML conversion to PDF

```
...  
<a href="#Technical">Technical</a>  
...
```



NORMEX

Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

2018-05-14

```
...  
<h2 id="Technical">Technical</h2>  
...
```



Contents

Creating a PDF

PDF viewers

Issues

Technical

Creating a PDF

From the left sidebar, under Print/export

1. {{chem}}; see [T42042](#)

7. <ruby> elements do not render; see [T2350](#)

8. Long equations are overflowing

See [T134205](#) for a method to generate PDF files pro

Technical

PDFs are rendered by the [Collection](#) software exten

Wiki

Get personal technical help at the Teal

Editing (Edit toolbar · C

General · Create a page · Page

More to discover?

- **/PhoneticAlphabet** and **/Phoneme**, **/PronunciationLexicon**
- **NSO** attribute owner and **/NS** entry in structure attributes

Other formats?

- XSL-FO, ODF, SVG
- GraphML, CellML, railML, etc.



Roman Toda,
Normex



Yulian Gaponenko,
Dual Lab

Next steps for developers?

- invent own practices
- save the source information



Thank you!

NORMEX

Roman Toda,
Normex

Any questions?

Get in touch: toda@digitaldocuments.org

yulian.gaponenko@duallab.com

Yulian Gaponenko,
Dual Lab

