

*The future: re-imagining the best
possible definition of PDF*



PDF Days Europe 2022 | Berlin

*Specifications, standards,
agility, correctness and
the reliability imperative*

Why?



Addison-Wesley Publishing Company

POSTSCRIPT®

Language

Reference Manual

Adobe Systems
Incorporated

Includes Detailed Programming
Information on the
LaserWriter™

Addison-Wesley Publishing Company, Inc.

POSTSCRIPT

Language

Program Design

Adobe Systems
Incorporated

Addison-Wesley Publishing Company

POSTSCRIPT™

Language

Tutorial and Cookbook

Adobe Systems
Incorporated

Copyright © 1986

PDF Reference

SECOND EDITION
version 1.3

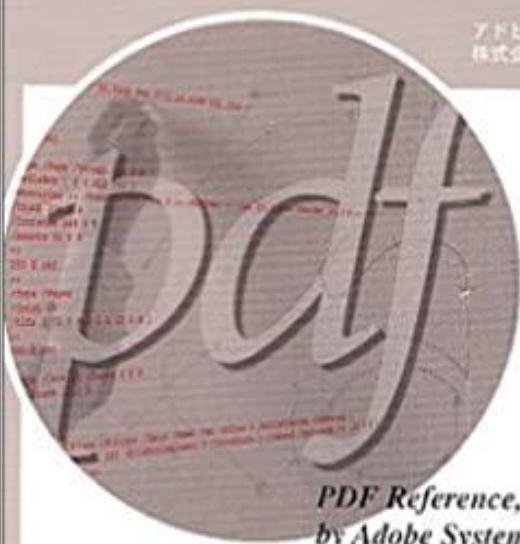


Adobe

PDFリファレンス 第2版

Adobe Portable Document Format Version 1.3

アドビシステムズ＝著
株式会社ドキュメントシステム＝訳



PDF Reference, Second Edition
by Adobe Systems Incorporated



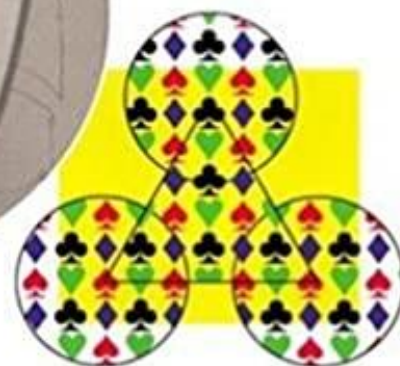
ピアソン・エデュケーション

PDF Reference

THIRD EDITION
version 1.4



- New transparent imaging model
- Accessibility support for disabled users
- Format used by Adobe® Acrobat® 5



Adobe Systems Incorporated

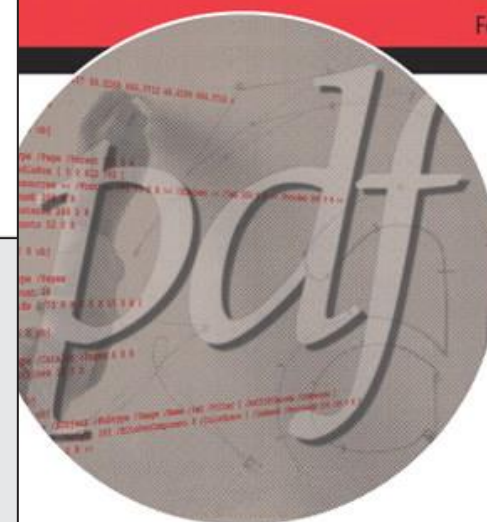


Adobe

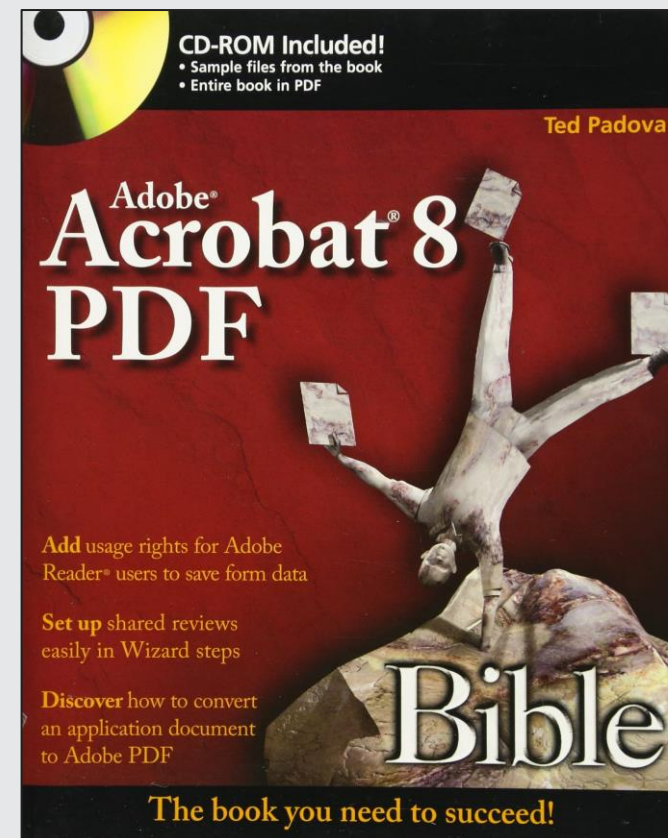
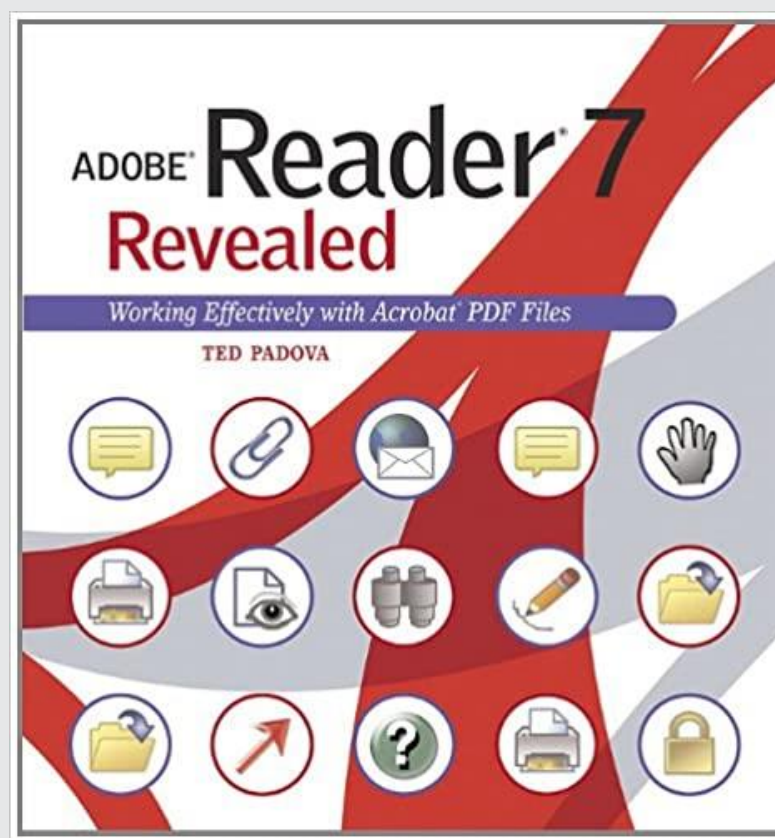
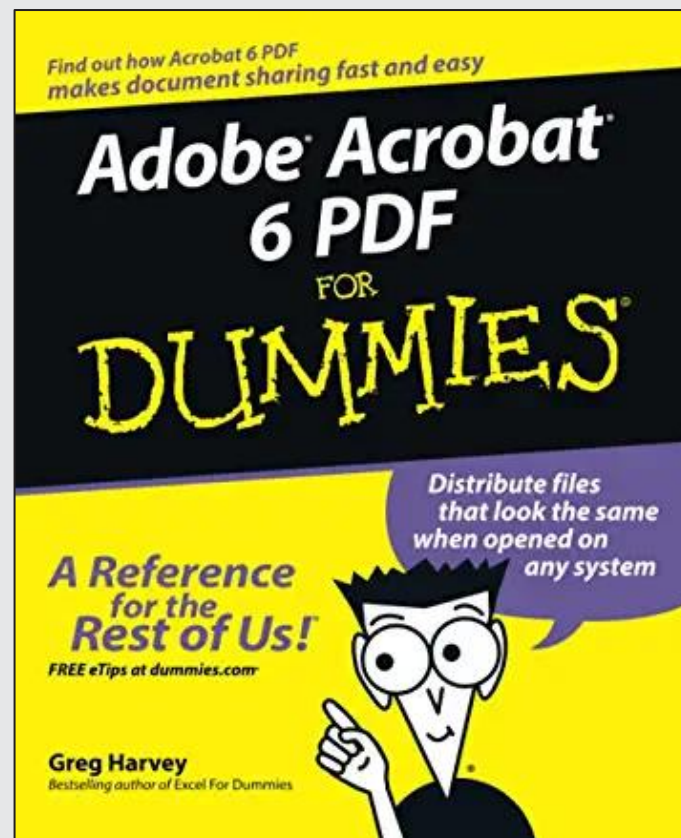
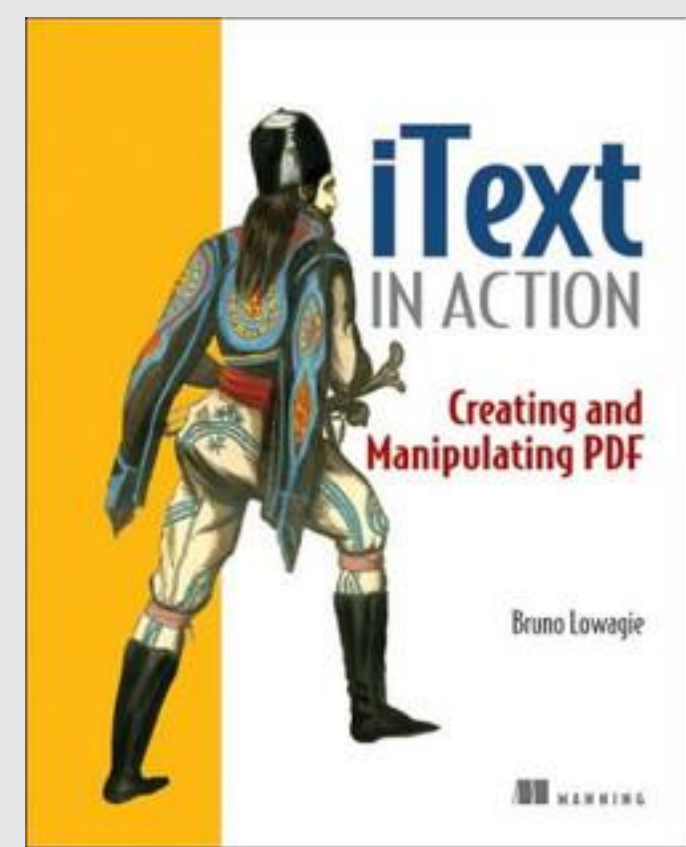
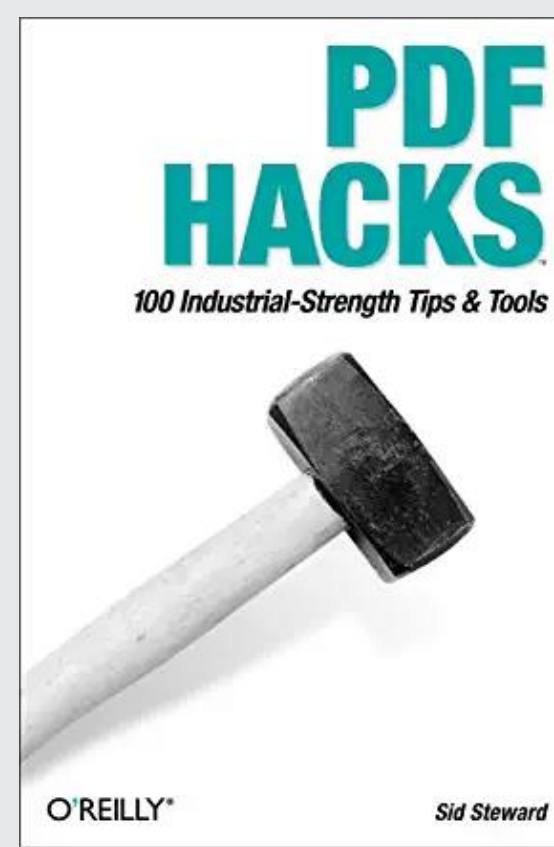
PDF Reference

FIFTH EDITION
version 1.6

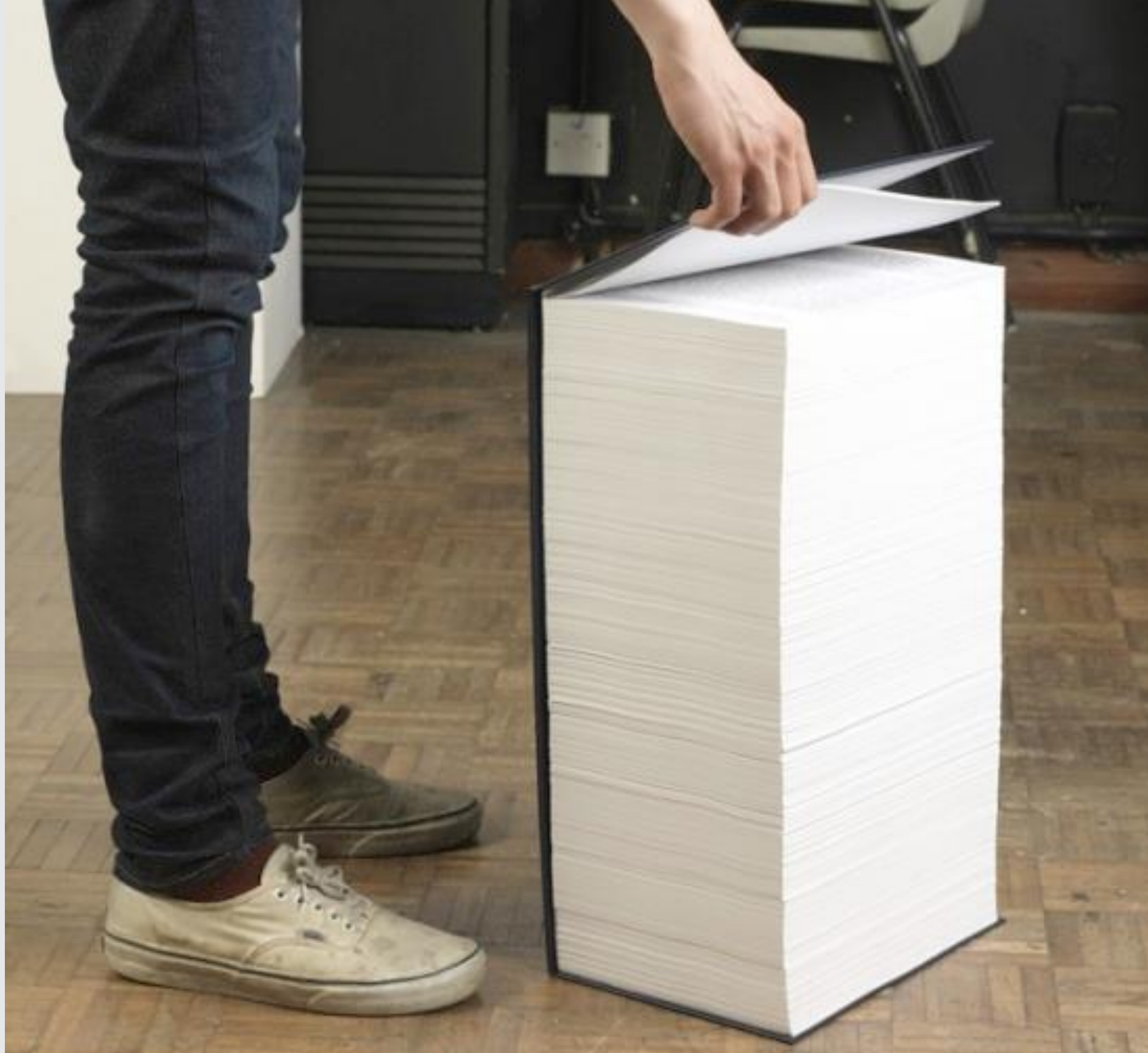
Format used by Adobe® Acrobat® 7



Adobe







The word cloud visualizes the 15 dimensions of the CLEAR PDF framework. The words are arranged in a circular pattern, with 'Clear' and 'PDF' being the largest and most central. Other prominent words include 'Error-free', 'Accurate', 'Unambiguous', 'Usable', 'Correct', 'Precise', 'Accessible', 'Relevant', 'Informative', 'Machine-readable', 'Referenceable', 'Exact', 'Indexable', 'Educational', 'Meaningful', 'Explanatory', 'Searchable', 'Maintained', 'Illustrative', 'Formalized', 'Unusable', 'Inaccurate', 'Ambiguous', 'Unusable', 'Incorrect', 'Imprecise', 'Inaccessible', 'Irrelevant', 'Uninformative', 'Unmachine-readable', 'Unreferenceable', 'Inexact', 'Unindexed', 'Uneducational', 'Unmeaningful', 'Unexplanatory', 'Unsearchable', 'Unmaintained', 'Unillustrative', 'Unformalized', 'Unusable', 'Inaccurate', 'Ambiguous', 'Unusable', 'Incorrect', 'Imprecise', 'Inaccessible', 'Irrelevant', 'Uninformative', 'Unmachine-readable', 'Unreferenceable', 'Inexact', 'Unindexed', 'Uneducational', 'Unmeaningful', 'Unexplanatory', 'Unsearchable', 'Unmaintained', 'Unillustrative', 'Unformalized'.

Key	Type	Value
Dests	dictionary	(Optional; PDF 1.1; shall be an indirect reference) ...
Outlines	dictionary	(Optional; shall be an indirect reference) ...
Threads	array	(Optional; PDF 1.1; shall be an indirect reference) ...
Lang	text string	<p>(Optional; PDF 1.4) A language identifier that shall specify the natural language of all text in the document except where overridden by language specifications in structure elements or marked-content (see 14.9.2, "Natural language specification"). If this entry is absent <u>or invalid (see 14.9.2, "Natural language specification")</u>, the language shall be considered unknown.</p> <p><u>NOTE All text in a document includes PDF text strings (see 7.9.2.2 "Text string type") as well as textual content.</u></p>

5. Otherwise, the body element is null, but there's a document element. Append the new value to the document element.

ISO/IEC 14492:2019, Information technology – Lossy/lossless coding of bi-level images. <https://www.iso.org/standard/75396.html>

NOTE The ISO/IEC 14492 dated reference for JBIG2 was updated in the 2020 dated revision of PDF 2.0.

ISO/IEC 14496-22:2019, Information technology – Coding of audio-visual objects — Part 22: Open Font Format. <https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>

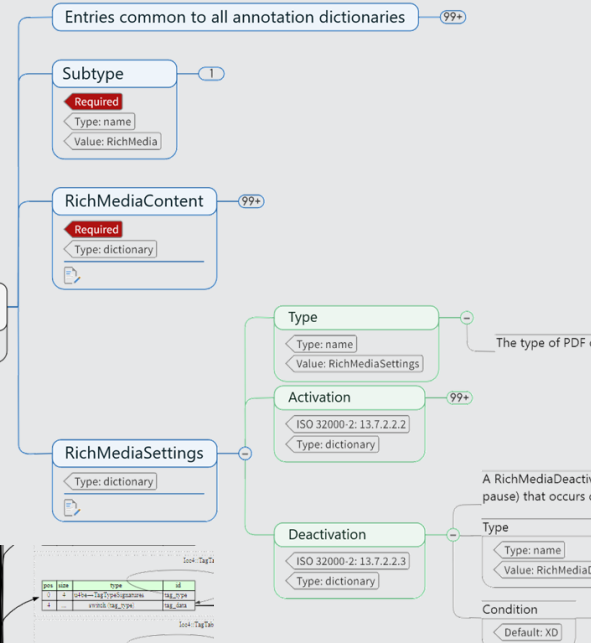
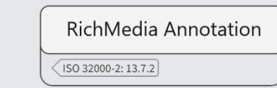
NOTE The ISO/IEC 14496-22 dated reference was updated in the 2020 dated revision of PDF 2.0.

ISO 14739-1, Document management – 3D use of Product Representation Compact (PRC) format — Part 1: PRC 10001. <https://www.iso.org/standard/54948.html> (This is an undated reference in PDF 2.0, but the URL here is to the ISO 2014 published edition. ISO TC 171 WG 5 is currently working on a PRC dated revision to fix errata so the URL will updated when PRC achieves DIS/FDIS status)

ISO 15076-1, Image technology colour management – Architecture, profile format and data structure — Part 1: Based on ICC.1:2010. <https://www.iso.org/standard/54754.html>

NOTE **Errata #181** changed ISO 15076-1 to an undated reference for all ICC profiles that contain an ICC header profile version of 4.x. This supports the dated revision of ISO 15076 that is in progress with ISO TC 130. An equivalent specification is already available at <https://www.color.org/specification/ICC.1-2022-05.pdf>.

ISO/IEC 15444-1:2019, Information technology — JPEG 2000 image coding system: Core coding system. <https://www.iso.org/standard/78221.html>



5. Otherwise, [the body element](#) is null, but there's a [document element](#). [Append](#) the new value to the [document element](#).

Note
The value returned by the `body` getter is not always the one passed to the setter.

Example

In this example, the setter successfully inserts a `body` element (though this is non-conforming since SVG does not allow a `body` as child of `SVG svg`). However the getter will return null because the document element is not `html`.

```
<svg xmlns="http://www.w3.org/2000/svg">
<script>
document.body = document.createElementNS("http://www.w3.org/1999/xhtml", "body");
console.assert(document.body === null);
</script>
</svg>
```

For web developers (non-normative)

```
document.images
    Returns an HTMLCollection of the img elements in the Document.

document.embeds
document.plugins
    Returns an HTMLCollection of the embed elements in the Document.

document.links
    Returns an HTMLCollection of the a and area elements in the Document that have href attributes.
```

- PostScript is an **interpreted programming language** with an implicit global state, so instructions accompanying the description of a document must be processed to determine the correct appearance of a given page, whereas each page in a PDF document is usually processed **independently** before being able to display the destination page.

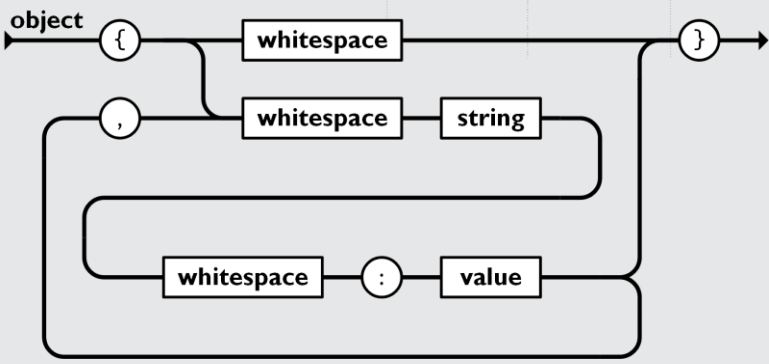
In computer science, an **array type** is a data type that contains a collection of elements, each identified by an index or key. In the most general type of array, the elements are all of the same type, but in more restricted types of array (such as **strings** and **structs**) the index determines the data type. 3D drawings can be embedded using **U3D** or **PRC** a

In computer science, an **array type** is a data type that represents a collection of *elements*, each selected by one or more indices that can be computed at run time during program execution. Such a collection is usually called an **array variable**, **array value**, or simply **array**. By analogy with the mathematical

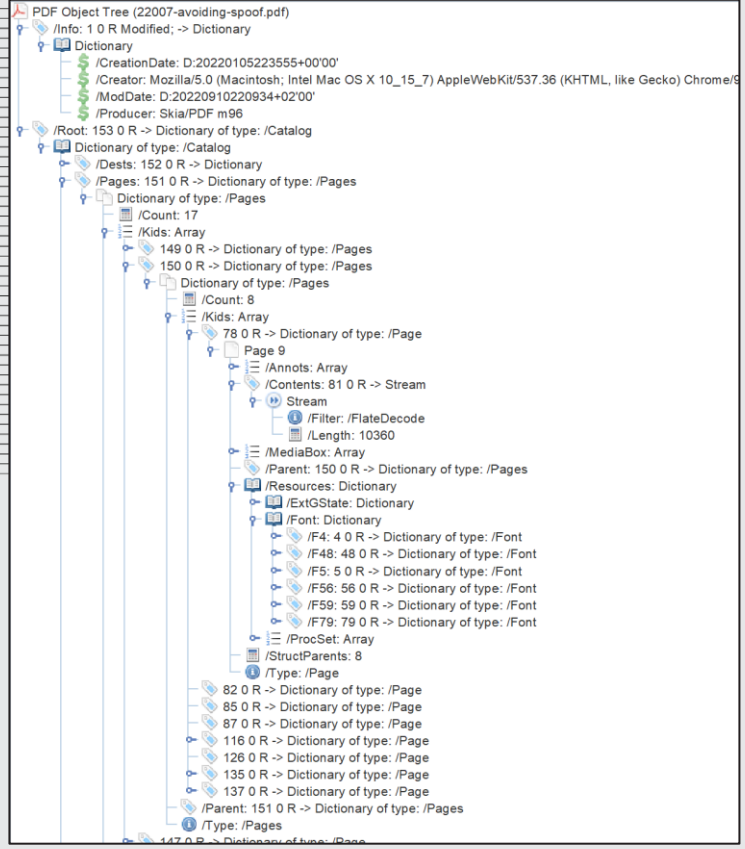
- **Arrays**, ordered collections of objects enclosed within square brackets (`[...]`)
- **Dictionaries**, collections of objects indexed by names enclosed within double angle brackets (`<<...>>`)
- **Streams**, usually containing large amounts of optionally compressed binary data, preceded by a dictionary and enclosed between `<<...>>`
- The **null** object

Furthermore, there may be comments, introduced with the percent sign (%). Comments may contain 8-bit characters.

Objects may be either *direct* (embedded in another object) or *indirect*. Indirect objects are numbered with an *object number* and a *group number*.

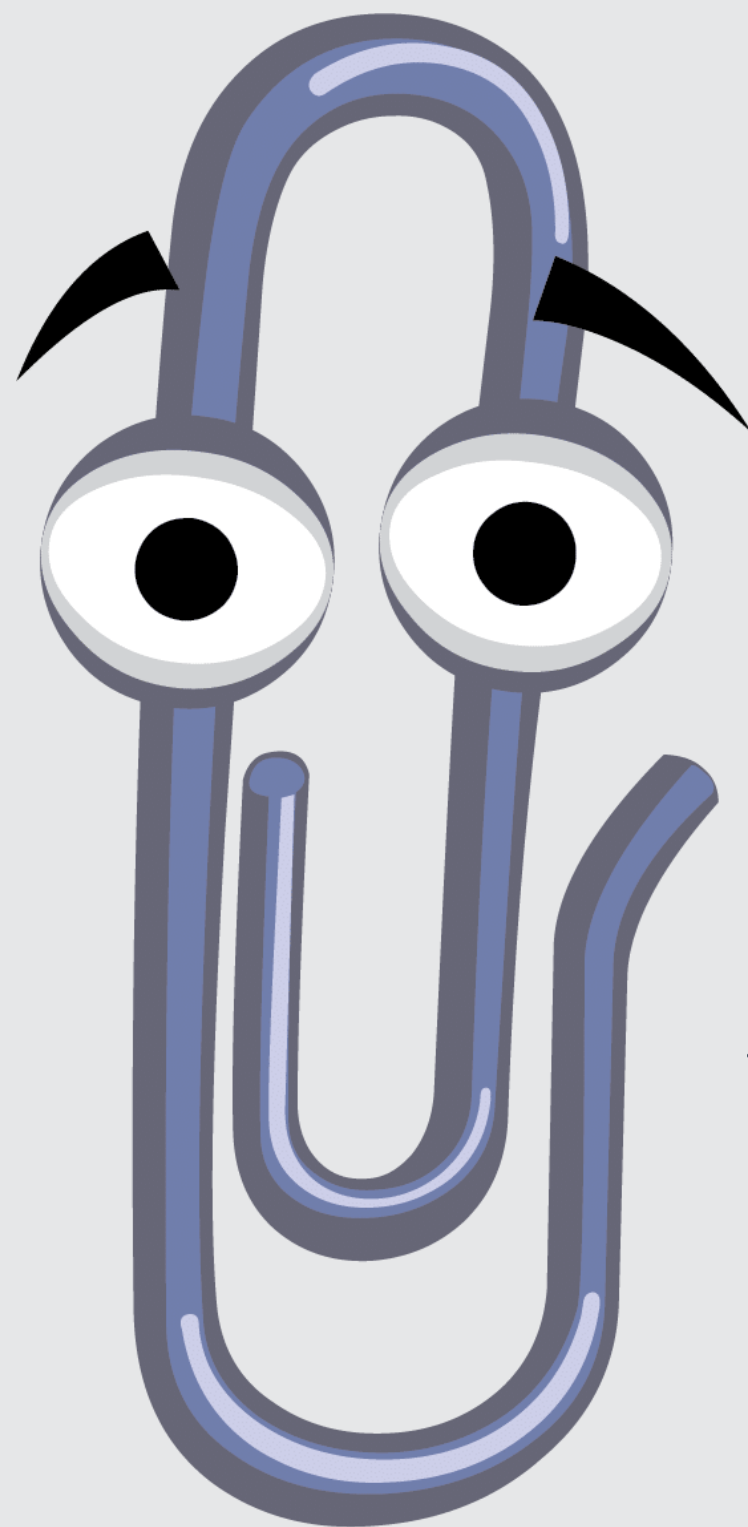


7.2266083 0 Td <01EC> Tj
ET
BT
/F4 17.33 Tf
1 0 0 -1 183.68973 113.454666 Tm
<072A> Tj
ET
0 0 0 RG 0 0 0 rg
BT
/F5 16 Tf
1 0 0 -1 96 142.731995 Tm
<072A> Tj
0 0 Td <06F7> Tj
5.29599 0 Td <0062> Tj
10 719986 0 Td <000E> Tj









It looks like there are errors and inconsistencies in the PDF specification. Would you like help with that?



Why now?



Thank you



PDF Days Europe 2022 | Berlin

Peter Wyatt
CTO, PDF Association
peter.wyatt@pdfa.org