

# What even is a PDF?

January 2024

---

# Contents

---

- What is a PDF?
- History
- Tools
- File structure
- Document structure
- Look at the files
- Common problems

# What is a PDF?

---

- Portable: independent of application software, hardware and operating system.
- Document: complete description of fixed-layout flat document.
- File: everything needed to present the document can be stored within a single file.

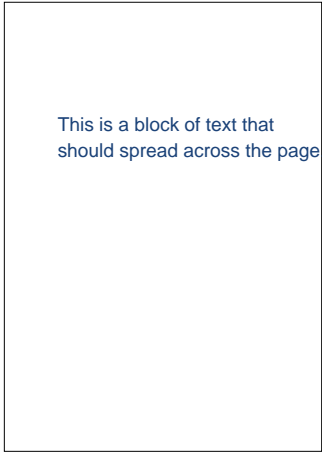
# History of PDF

---

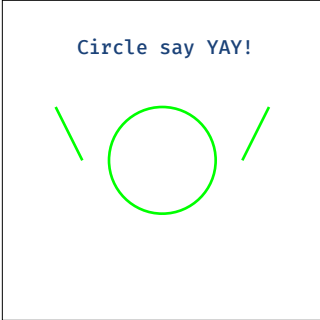
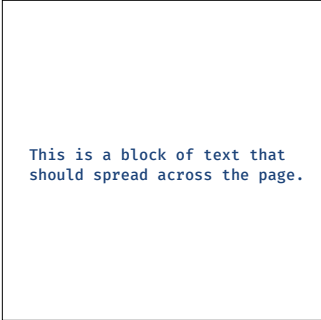
- 1990 The Camelot Project launched
- 1993 Portable Document Format 1.0 launched
- 1994 PDF 1.1 passwords and better encryption
- 2001 PDF 1.4 accessibility and transparency
- 2006 PDF 1.7 stabilisation
- 2008 PDF 1.7 ISO standardisation
- 2017 PDF 2.0 elimination of prioritary elements

# Three documents

## mini.pdf



## maxi.pdf



## web.pdf

PDF - Wikipedia

https://en.wikipedia.org/wiki/PDF

WIKIPEDIA

The Free Encyclopedia

PDF

**Portable Document Format (PDF)**, standardized as **ISO 32000**, is a file format developed by Adobe in 1992 to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems.<sup>[a]</sup> Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991.<sup>[a]</sup> PDF was standardized as ISO 32000 in 2008.<sup>[a]</sup> The last edition as ISO 32000-2:2020 was published in December 2020.

PDF files may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and form-fields, layers, rich media (including video content), three-dimensional objects using U3D or PRC, and various other data formats. The PDF specification also provides for encryption and digital signatures, file attachments, and metadata to enable workflows requiring these features.


**History**

Adobe Systems made the PDF specification available free of charge in 1993. In the early years PDF was popular mainly in desktop publishing workflows, and competed with several other formats, including DjVu, Encaps, Common Ground Digital Paper, Farallon Replica and even Adobe's own PostScript format.

PDF was a proprietary format controlled by Adobe until it was released as an open standard on July 1, 2008, and published by the International Organization for Standardization as ISO 32000-1:2008,<sup>[a]</sup> at which time control of the specification passed to an ISO Committee of volunteer industry experts. In 2008, Adobe published a Public Patent License to ISO 32000-1 granting royalty-free rights for all patents owned by Adobe necessary to make, use, sell, and distribute PDF-compliant implementations.<sup>[a]</sup>

PDF 1.7, the sixth edition of the PDF specification that became ISO 32000-1, includes some

Portable Document Format



Adobe PDF icon

<b>Filename extension</b>	.pdf
<b>Internet media type</b>	application/pdf <sup>[1]</sup>
	application/x-pdf
	application/x-bzpdf
	application/x-gzpdf
<b>Type code</b>	PDF <sup>[1]</sup> (including a single trailing space)
<b>Uniform Type Identifier (UTI)</b>	com.adobe.pdf
<b>Magic number</b>	%PDF
<b>Developed by</b>	Adobe Inc. (1991–2008) ISO (2008–)
<b>Initial release</b>	June 15, 1993
<b>Latest release</b>	2.0
<b>Extended to</b>	PDF/A, PDF/E, PDF/A-2, PDF/VT, PDF/X
<b>Standard</b>	ISO 32000-2
<b>Open format?</b>	Yes
<b>Website</b>	iso.org/standard/75839.html (http://s/iso.org/standard/75839.html)

1 of 22

08/05/2024, 13:39

# PDF tools

---

- **qpdf:** useful for exploring PDF files on the command line.
- **mutool:** useful for extracting fonts and images from PDF files.
- **allsorts:** useful for exploring and subsetting fonts.

# File structure - min.pdf

---

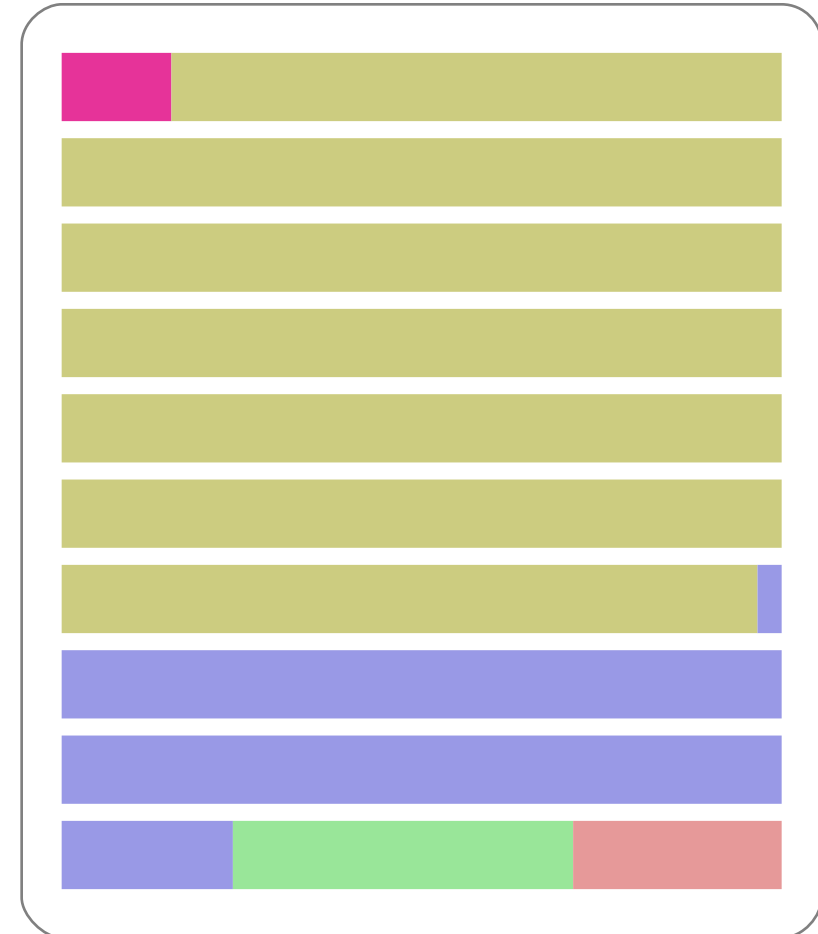
Header

Body (68%)

Cross-reference table

Trailer

Startxref



# File structure - max.pdf

---

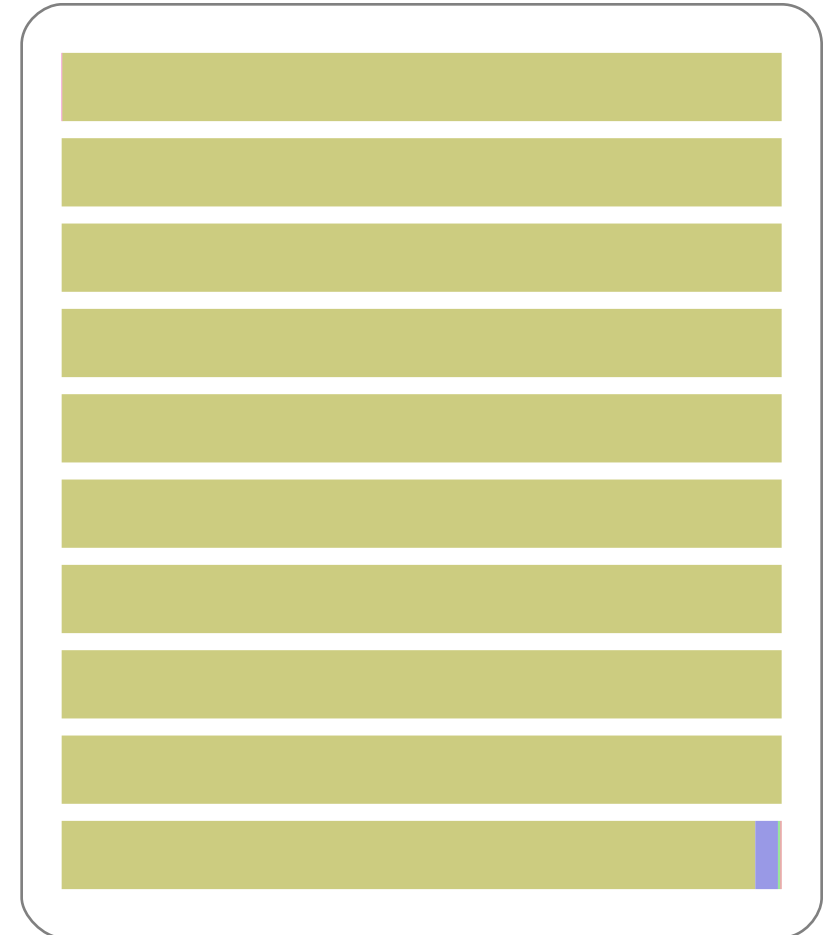
Header

Body (99.6%)

Cross-reference table

Trailer

Startxref





# File structure - appended.pdf

---

Header

Body

Cross-reference table

Trailer

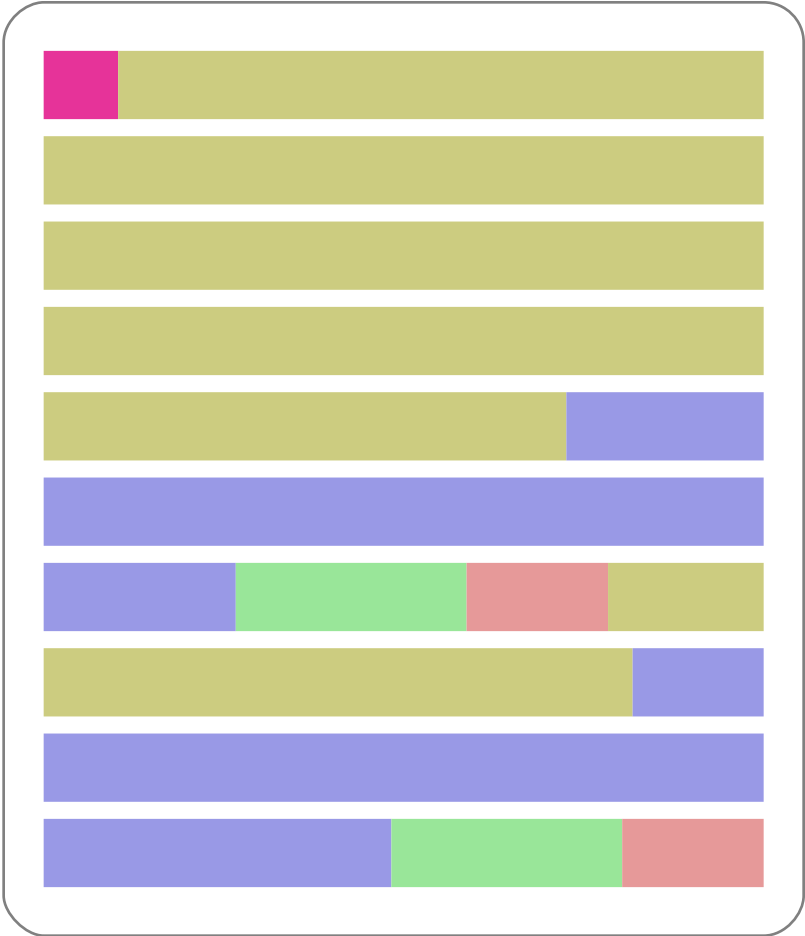
Startxref

Body

Cross-reference table

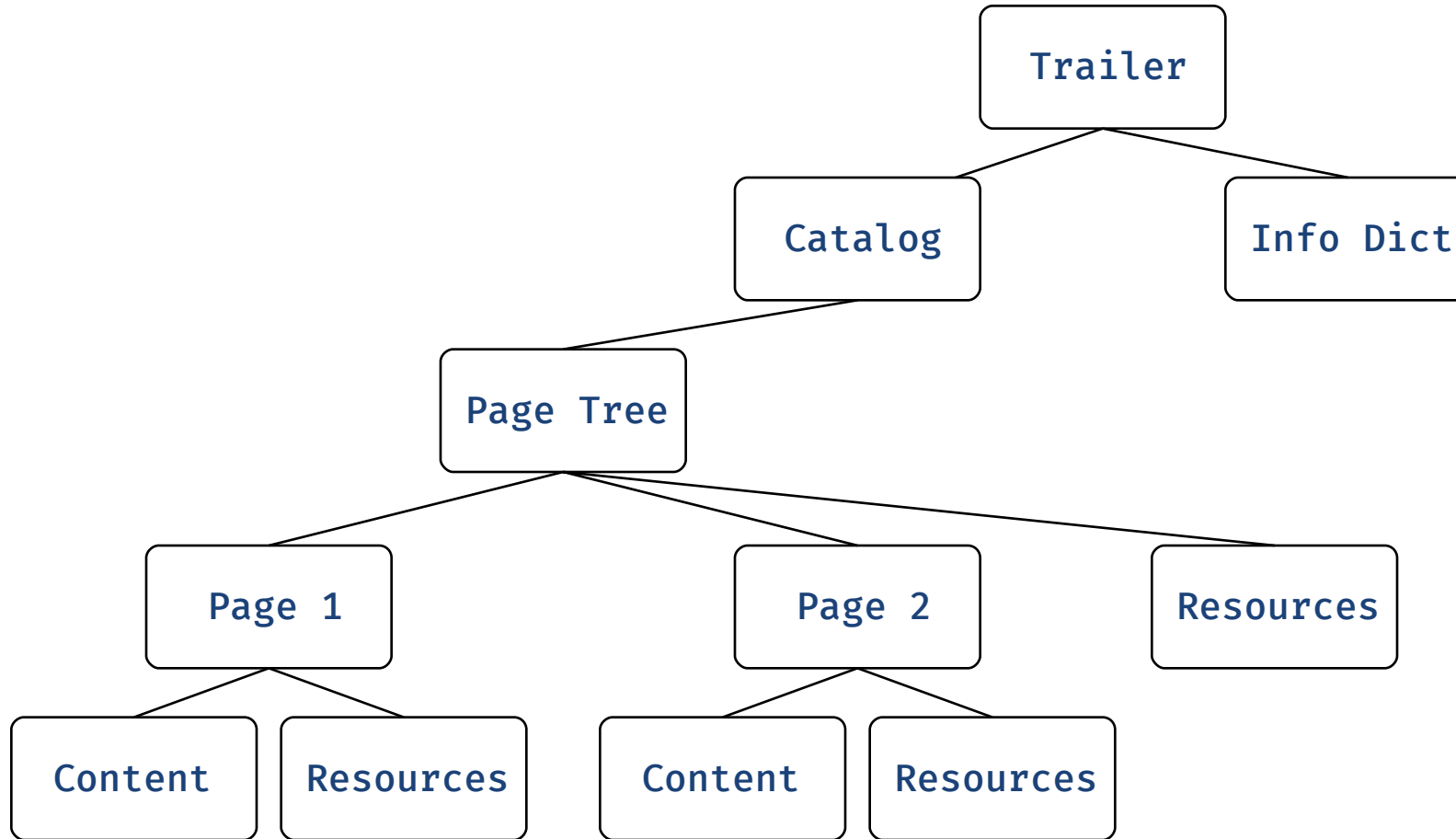
Trailer

Startxref



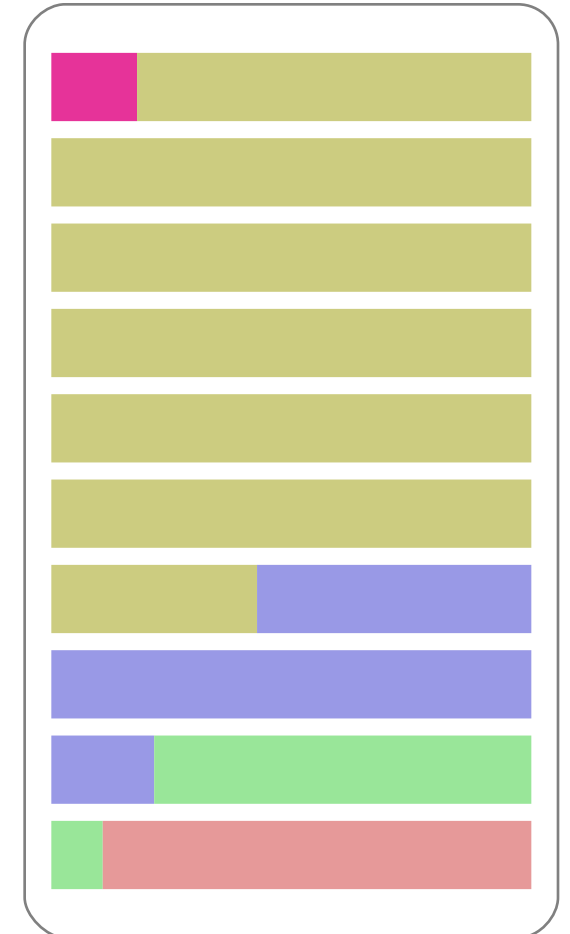
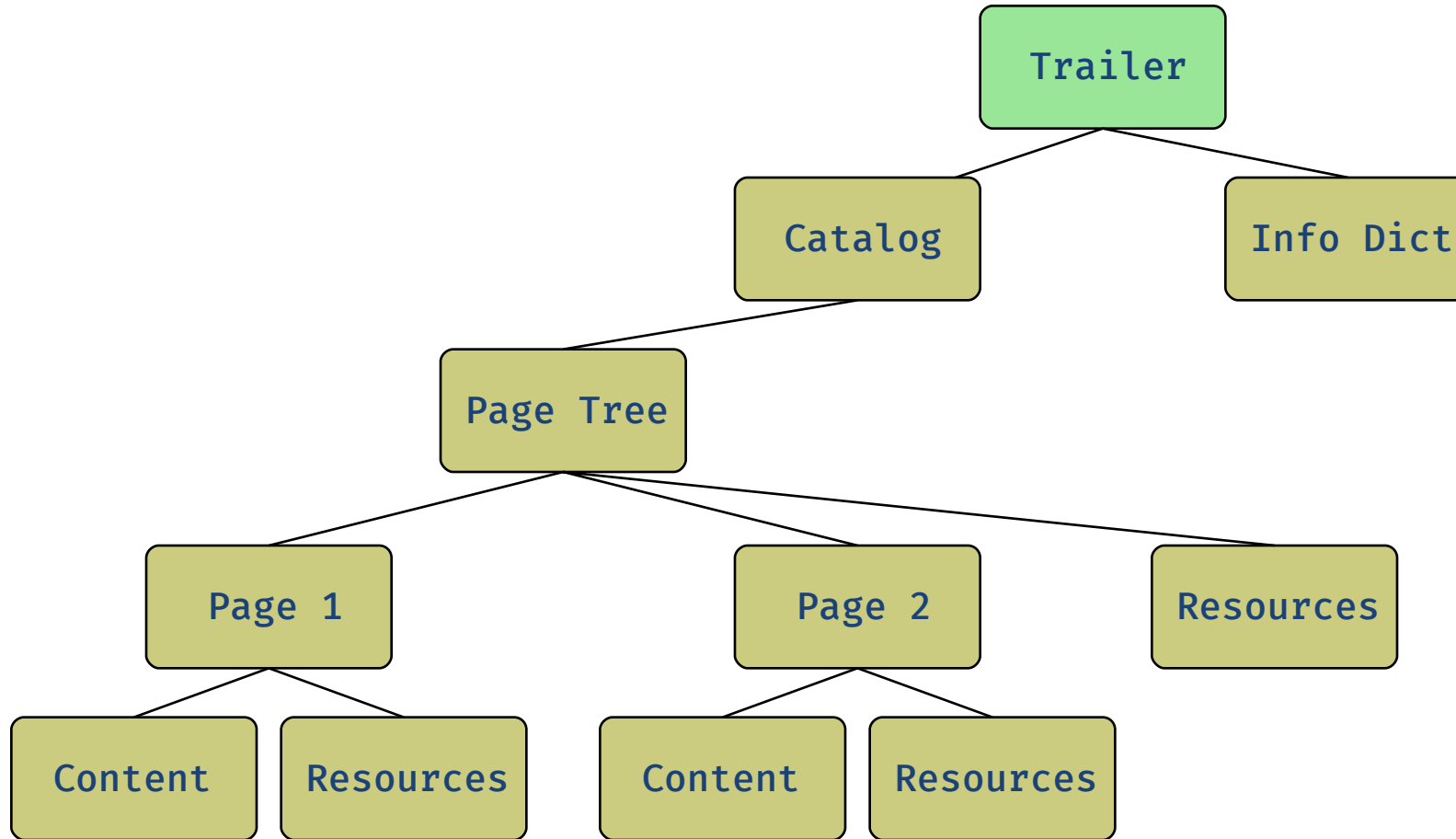
# Document structure

---



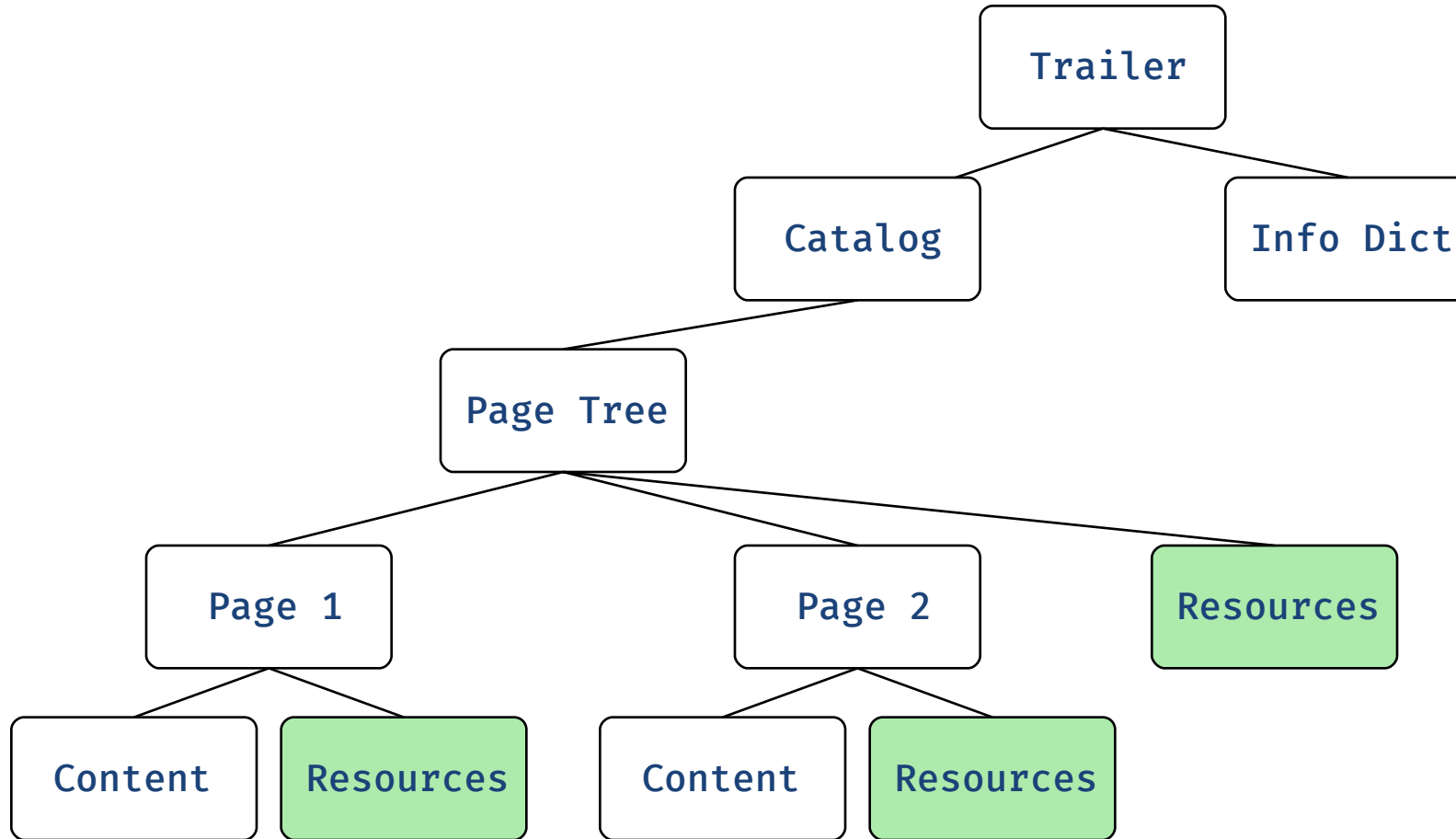
# Document structure - file structure

---



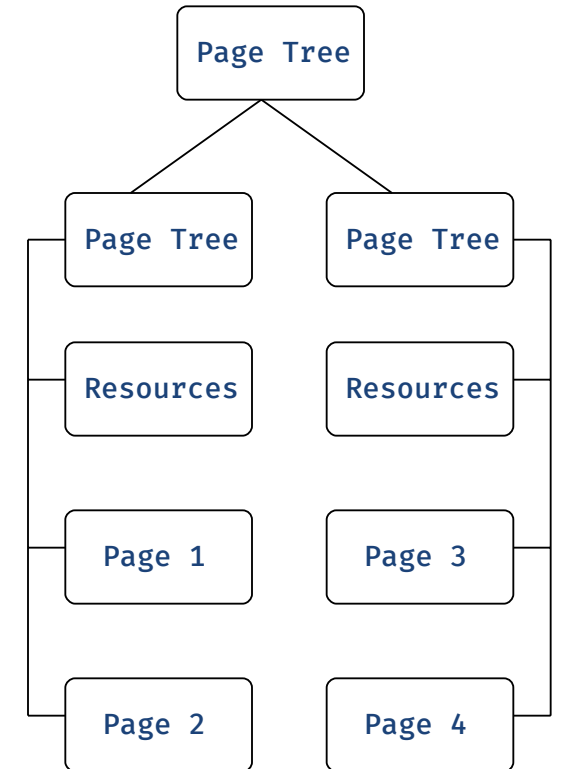
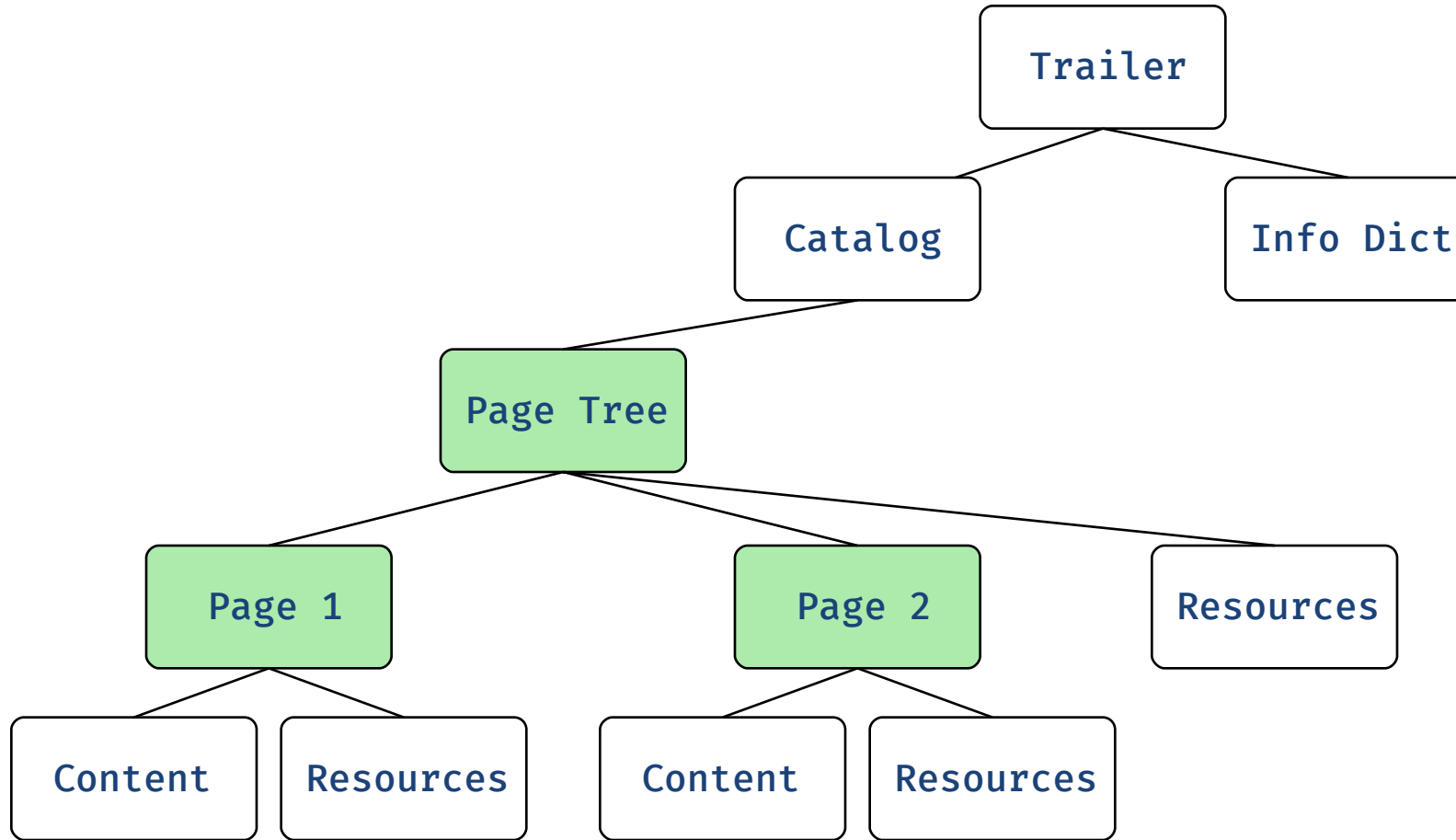
# Document structure - resources

---



# Document structure - page tree

---



Let's take a look

---

Into the files...