

### File structure [\[ edit \]](#)

A PDF file is a 7-bit [ASCII](#) file, except for certain elements that may have binary content. A PDF file starts with a header containing the [magic number](#) and the version of the format such as `%PDF-1.7`. The format is a subset of a COS ("Carousel" Object Structure) format.<sup>[15]</sup> A COS tree file consists primarily of *objects*, of which there are eight types:<sup>[16]</sup>

- [Boolean](#) values, representing *true* or *false*
- Numbers
- [Strings](#), enclosed within parentheses ( `( ...)` ), may contain 8-bit characters.
- Names, starting with a forward slash ( `/` )
- [Arrays](#), ordered collections of objects enclosed within square brackets ( `[ ... ]` )
- [Dictionaries](#), collections of objects indexed by Names enclosed within double pointy brackets ( `<< ... >>` )
- [Streams](#), usually containing large amounts of data, which can be compressed and binary
- 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 *generation number* and defined between the `obj` and `endobj` keywords. An index table, also called the cross-reference table and marked with the `xref` keyword, follows the main body and gives the byte offset of each indirect object from the start of the file.<sup>[17]</sup> This design allows for efficient [random access](#) to the objects