

# Portable Document Format

aus Wikipedia, der freien Enzyklopädie

Das **Portable Document Format (PDF)**; deutsch: *(trans)portables Dokumentenformat*) ist ein plattformunabhängiges Dateiformat für Dokumente, das vom Unternehmen Adobe Systems entwickelt und 1993 veröffentlicht wurde.

Ziel war es, ein Dateiformat für elektronische Dokumente zu schaffen, das diese unabhängig vom ursprünglichen Anwendungsprogramm, vom Betriebssystem oder von der Hardwareplattform originalgetreu weitergeben kann. Ein Leser einer PDF-Datei soll das Dokument immer in der Form betrachten und ausdrucken können, die der Autor festgelegt hat. Die typischen Konvertierungsprobleme (wie veränderter Seitenumbruch oder falsche Schriftarten) beim Austausch eines Dokuments zwischen verschiedenen Programmen entfallen dadurch.

Neben Text, Bildern und Grafik kann eine PDF-Datei auch Hilfen enthalten, die die Navigation innerhalb des Dokumentes erleichtern. Dazu gehören zum Beispiel anklickbare Inhaltsverzeichnisse und miniaturisierte Seitenvorschauen.

PDF ist mittlerweile weit verbreitet und wird von vielen elektronischen Zeitschriften (E-Journals) genutzt. Es gibt auf dem Markt zahlreiche Softwareprodukte, die PDF-Dateien erzeugen können.

## Portable Document Format (PDF)



Dateiendung: .pdf

MIME-Type: application/pdf

Magische Zahl: %PDF

Entwickelt von: Adobe Systems

Erstveröffentlichung: 1993

Aktuelle Version: 1.7 (Stand: Juli 2008)

Standard(s): Adobe: PDF Reference  
([http://www.adobe.com/devnet/pdf/pdf\\_reference.html](http://www.adobe.com/devnet/pdf/pdf_reference.html))

## Inhaltsverzeichnis

- 1 Übersicht
  - 1.1 Grundlagen und Software
  - 1.2 Erzeugung und Umwandlung
- 2 Verwendung und Eigenschaften
  - 2.1 Umfang von Dokumenten
  - 2.2 Speicherung von Informationen im Dokument
  - 2.3 Sicherheit von Dokumenten
  - 2.4 Bearbeiten von Dokumenten
  - 2.5 Weitere Eigenschaften
- 3 Die Eigenschaften von PDF-Dateien
- 4 Sicherheitslücken durch menschliches Versagen
  - 4.1 Prominente Fallbeispiele
  - 4.2 Vermeidung
- 5 Standard-Schriftarten
- 6 PDF in Betriebssystemen
- 7 Angaben der Seitengeometrie
  - 7.1 MediaBox
  - 7.2 CropBox
  - 7.3 BleedBox
  - 7.4 TrimBox
  - 7.5 Art Box
- 8 Formularprozesse
- 9 Konvertierung anderer Formate zu PDF
  - 9.1 XML
  - 9.2 HTML
  - 9.3 TeX
  - 9.4 DVI
  - 9.5 3D-CAD-Daten
  - 9.6 Videoformate
  - 9.7 Dateien aus Büroanwendungen

- 9.8 Diverse Formate
- 10 Normen und Standards
- 11 Siehe auch
- 12 Literatur
- 13 Weblinks
- 14 Einzelnachweise

# Übersicht

## Grundlagen und Software

Aus Texten, Bildern und Grafiken – gemischt oder einzeln – lassen sich mit entsprechenden Programmen PDF-Dokumente und -Dateien erzeugen (z. B. mit freien Programmen wie dem PDFCreator und Büropaketen LibreOffice/OpenOffice.org, kostenpflichtigen wie Adobe Acrobat oder einfach über den Druckdialog) und mit entsprechenden Leseprogrammen darstellen (z. B. Evince, Ghostscript, Okular, Adobe Reader, Foxit, Vorschau). Der Ersteller einer PDF-Datei kann diese in vielfältiger Weise vor ungewolltem Gebrauch schützen, indem er die Sicherungsmechanismen von PDF aktiviert. Durch Verschlüsselung soll der Zugriff Unbefugter verhindert werden. Je nach Bedarf kann bereits das Öffnen der Datei ein Passwort erfordern, oder das Kopieren von Inhalten aus der Datei oder das Ausdrucken wird nicht zugelassen. Die dafür in PDF implementierten Schutzmechanismen sind jedoch nicht zuverlässig; insbesondere sind einfachere Formen der Verschlüsselung leicht zu überwinden.

In der Startphase war der Adobe Reader kostenpflichtig. Erst die kostenfreie Weitergabe der Software ermöglichte die Verbreitung im heutigen Ausmaß. PDF war lange Zeit ein kommerzielles (proprietäres), aber offengelegtes Dateiformat, das im PDF Reference Manual von Adobe dokumentiert ist. Anfang 2007 hatte Adobe es in den Standardisierungsprozess der ISO eingebracht und mit der Veröffentlichung am 1. Juli 2008 ist PDF in Version 1.7 als ISO 32000-1:2008 ein Offener Standard.

Bestimmte Methoden beim Umgang mit PDF wurden zur Erleichterung des Datenaustausches in der Druckvorstufe (als PDF/X) schon vorher und zur Langzeitarchivierung von PDF-Dateien (als PDF/A-1 in ISO 19005-1:2005) von der ISO genormt.

## Erzeugung und Umwandlung

PDF ist eine vektorbasierte Seitenbeschreibungssprache, die die freie Skalierbarkeit der Darstellung erlaubt. PDF-Dateien beschreiben das mit dem Erstellungsprogramm erzeugte Layout in einer vom Drucker und von Voreinstellungen unabhängigen Form weitgehend originalgetreu. Das ist einer der wesentlichen Unterschiede zwischen PDF und fortgeschrittenen Beschreibungs- und Auszeichnungssprachen wie SGML oder HTML, wenn es um die Forderung nach unbedingter Layouttreue geht.

Um eine Darstellung auch auf Ausgabegeräten mit kleiner Anzeigefläche – wie beispielsweise PDAs oder Mobiltelefonen – zu optimieren, können in einem PDF Auszeichnungen (ähnlich HTML-Tags) eingelagert werden, die ein Umbrechen der Seiteninhalte – dann zwangsläufig unter Einschränkung der Layouttreue – ermöglichen. Solche Auszeichnungen ermöglichen es zudem einem Vorleseprogramm, sehbehinderten Nutzern das Dokument vorzulesen und erleichtern das Konvertieren des Inhalts in andere Formate.

Ein häufig verwendetes Programm zur Erzeugung von PDF-Dateien ist *Adobe Acrobat Distiller*, das aus PostScript-Dateien PDFs erstellt. Acrobat Distiller ist als Desktopprodukt für Windows und Mac OS verfügbar. Serverversionen sowie der kostenlose Adobe Reader existieren auch für andere Plattformen. Mit Hilfe der freien Software Wine läuft Acrobat Distiller auch unter Linux. Diverse Office- und DTP-Programme anderer Hersteller bieten einen direkten PDF-Export an und sind auf einer Reihe von Plattformen verfügbar. Des Weiteren ist es mit pdfTeX möglich, aus LaTeX direkt eine PDF-Datei zu erzeugen. Mit inzwischen zahlreichen Werkzeugen und Programmierbibliotheken lassen sich auf unterschiedlichste Weise mit unterschiedlicher Spezialisierung PDF-Dateien erstellen, die Erzeugung von PDF-Dateien ist nahezu auf jeder Plattform möglich.

Adobe räumt Entwicklern teilweise das Recht ein, eigene Anwendungen zur Generierung und Bearbeitung von PDF-Dokumenten zu entwickeln, behält sich aber das Copyright über die Spezifikationen vor.<sup>[1]</sup> PDF als Seitenbeschreibungssprache kann als Weiterentwicklung des Grafikmodells von PostScript angesehen werden, das ebenfalls offengelegt ist. Ende Februar 2007 hat Adobe angekündigt, die PDF-Spezifikation in der Version 1.7 in die ISO-Normierung einzubringen und arbeitet dafür mit dem amerikanischen Branchenverband AIIM zusammen, der das Sekretariat des ISO-Komitee TC 171 innehat. Adobe drohte Microsoft bei Integration des offenen Standards PDF mit kartellrechtlichen Klagen.

## Verwendung und Eigenschaften

Eine PDF-Datei kann Dokumente eines Ursprungsprogramms, einschließlich aller Farben, Raster- und Vektorgrafiken, sehr präzise wiedergeben. Das gilt grundsätzlich auch für Schriften.

## Umfang von Dokumenten

PDF-Dokumente können einen Umfang von einer bis zu mehreren hunderttausend Seiten haben. Dabei ist die Seitengröße durch das Format selbst nicht begrenzt. In Adobe Acrobat gibt es jedoch durch die Implementierung bedingte Grenzen (bis Version 3 auf  $45 \times 45$  Zoll [etwa 1,14 m], bis Version 6 auf  $200 \times 200$  Zoll [5,08 m], und seit Version 7 auf das 75000-fache, das sind  $15.000.000 \times 15.000.000$  Zoll [381 km]).

## Speicherung von Informationen im Dokument

In PDF-Dateien werden alle Informationen als nummerierte Objekte abgespeichert. Objekte sind beispielsweise Schriftinformationen, Zeichenbreiten, verwendete Zeichen-Encodings (Mac/PC ...), Seitenbeschreibung, Parameter für Decoder, Crop-Boxes, einzelne Lesezeichen, Farbdefinitionen, Seitenreihenfolgen, Bitmaps, Formulare, Sprungmarken und alles andere, was in PDF-Dateien gespeichert werden kann. Eine hundertseitige PDF-Datei kann ohne Weiteres 10.000 Objekte enthalten.



Größenvergleich Europa und ein darin befindliches PDF mit einer Seitenlänge von je 381 km.

PDF basiert auf dem gleichen Grafikmodell wie die Seitenbeschreibungssprache PostScript, erlaubt jedoch gegenüber dieser einige zusätzliche Funktionen – insbesondere interaktive Elemente wie Lesezeichen, Kommentare, Formularfelder und deren Programmierung mit JavaScript sind möglich. Auch das Grafikmodell für die Seiteninhalte wurde gegenüber PostScript um Funktionen wie Transparenz oder optionalen oder schaltbaren Inhalt (in der Acrobat-Bedienungsoberfläche als Ebenen bezeichnet) oder die Unterstützung von ICC-Profilen und OpenType-Schriften erweitert.

Schriften (mit Ausnahme von Pixelschriften) und Vektorgrafiken können beliebig ohne Qualitätsverlust vergrößert werden. Große Netzwerkpläne und Datenmodelle lassen sich unter diesen Voraussetzungen verlustfrei auf einer PDF-Seite unterbringen.

Aus PDF-Dokumenten lassen sich Textpassagen, Tabellen und Grafiken (auch Ausschnitte davon) leicht in anderen Anwendungsprogrammen durch Kopieren und Einfügen der jeweiligen Elemente weiterverarbeiten, *falls* der Ersteller des Dokumentes es zugelassen hat. Text kann nicht nur zur Weiterverarbeitung in anderen Anwendungen, sondern auch zum Durchsuchen oder zur Verwendung mit anderen Ausgabemedien, wie Screenreadern, extrahiert werden. Durch die Textsuche im einzelnen Dokument oder die Volltextrecherche innerhalb einer PDF-Dokumentensammlung lassen sich Detailinhalte sehr einfach auffinden. Das funktioniert selbst dann, wenn der Text grafisch verzerrt, etwa in Kreis- oder Kurvenform, dargestellt ist.

## Sicherheit von Dokumenten

Eine Besonderheit von PDF ist der optionale Dokumentenschutz mit 40- oder 128-Bit-Verschlüsselung. Durch Vergabe eines Benutzerpasswortes ist es so möglich, das Dokument nur einem beschränkten Personenkreis zugänglich zu machen. Weiterhin kann der Autor mit einem separaten Besitzerpasswort gezielt die Rechtevergabe des betreffenden Dokuments festlegen. So kann verhindert werden, dass Benutzer das Dokument abändern, ausdrucken oder Teilinhalte kopieren können. Auch ohne Kenntnis des Besitzerpasswortes lassen sich diese Rechteeinschränkungen jedoch mithilfe diverser Werkzeuge insbesondere dann leicht entfernen, wenn zwar ein Besitzerpasswort verwendet wird, das Öffnen des Dokuments jedoch kein Passwort erfordert (das bedeutet, dass kein Benutzerpasswort gesetzt ist). Die mit der Verschlüsselung verbundenen Nutzungseinschränkungen lassen sich durch Speichern von Screenshots als Bitmaps und darauf folgende optische Texterkennung leicht umgehen.

Durch entsprechende Werkzeuge lassen sich auch Rechte vergeben, die es ermöglichen, PDF-Dokumente mit Notizen, Kommentaren und Dateianhängen zu versehen oder Formulareinträge abzuspeichern. Ursprünglich konnten diese Merkmale nur mit Adobe Acrobat genutzt werden, seit Version 7 ist es jedoch auch mit dem kostenlosen Adobe Reader möglich, Notizen und Kommentare hinzuzufügen und interaktive Formularfelder auszufüllen, sofern das entsprechende Dokument vom Verfasser mit den notwendigen Berechtigungen versehen wurde.

Mittlerweile existieren auch DRM-geschützte PDF-Dateien. Diese sind unter anderem mit Adobe Digital Editions lesbar.

## Bearbeiten von Dokumenten

PDF wurde als ein Austauschformat für fertiggestellte Dokumente konzipiert. Inzwischen gibt es eine Reihe von Programmen und Erweiterungen für Adobe Acrobat, mit denen sich PDF-Dateien bearbeiten lassen. Das Format ist jedoch nicht mit Dateiformaten von Textverarbeitungsprogrammen oder Grafikprogrammen vergleichbar und eignet sich, abgesehen von der Notiz- und Kommentarfunktion, nur begrenzt zur Weiterverarbeitung von Dokumenten. Es ist allerdings innerhalb gewisser Grenzen möglich, beispielsweise Tippfehler zu entfernen. Vorteile im Desktop-Publishing sind für Grafiker und Designer die Einbindung aller Elemente für die Druckerstellung.

PDF-Dokumente können abhängig vom Einzelfall sowohl größer als auch kleiner als die Dateien der Ursprungsanwendung sein. Die Größe eines Dokuments hängt von der Art der enthaltenen Daten, von der Effizienz des Erstellungsprogramms und davon ab, ob Schriften eingebettet wurden. Schriften können entweder vollständig, als Untermenge der tatsächlich im Dokument verwendeten Zeichen oder aber überhaupt nicht eingebettet werden. Soll ein Dokument zuverlässig darstellbar sein, unabhängig davon, ob auf der Zielpattform die verwendeten Schriften installiert sind, müssen mindestens die tatsächlich verwendeten Zeichen eingebettet werden.

## Weitere Eigenschaften