



## PDF: Container mit Struktur

**?** Was genau ist PDF eigentlich?

**!** PDF steht für Portable Document Format. Es ist ein Containerformat, das Text, Bilder, Vektorgrafiken, aber auch dynamische Elemente wie Formularfelder, Videos und Animationen enthalten kann. Der Grafiksoftware-Hersteller Adobe hat PDF in den 90er-Jahren auf Basis seiner Seitenbeschreibungssprache PostScript entwickelt, um Dokumente jedweder Herkunft – von Office- bis hin zu Grafikprogrammen – layoutgetreu unabhängig von ihrer Ursprungsanwendung weitergeben und ausdrucken zu können.

Wegen seines ursprünglichen Zwecks – reine Anzeige und Weitergabe an einen Druckdienstleister – unterstützte PDF anfangs vor allem statische Objekte. Man erzeugte es in der Regel über einen virtuellen Drucker, der aus jeder Anwendung heraus

# Portable Document Format

**Das Portable Document Format ist das Austauschformat schlechthin – vom einfachen Bericht bis hin zum großformatigen Gebäudeplan. Weil es so komplex ist, sollte man aber ein paar Fallstricke und Eigenheiten kennen.**

Von Andrea Trinkwalder

den Druckstrom in PDF-Code übersetzte und als Datei speicherte. Später wurde das Format immer dynamischer und interaktiver: mit digitalen Signaturen, JavaScript, Formularfeldern mit Berechnungen und sogar 3D-Objekten. Außerdem kann ein PDF beliebige Dateien (Attachments) und Ebenen enthalten. Von OCR-Software erstellte PDFs etwa bestehen aus mindestens zwei Schichten: dem Scan sowie einer Textebene für die erkannten Zeichen.

Mittlerweile bieten zahlreiche Programme – auch Office-Programme – einen sehr guten PDF-Export, sodass man in der Regel keinen universellen PDF-Drucker mehr braucht. Mit verantwortlich für den Qualitätssprung sind die besseren Grafikbibliotheken der Betriebssysteme: Die Grafikbibliothek Quartz von macOS basiert auf PDF, womit Apple-Rechner nativen PDF-Export und -Anzeige bieten. Der Windows-10-Druckpfad setzt auf den von Microsoft entwickelten PDF-Konkurren-

ten XPS – dessen Objekte sich auch leicht in PDF-Code konvertieren lassen. Nativer Export bietet gegenüber der PDF-Erstellung über das Druckmenü auch den Vorteil, dass interaktive Elemente wie Inhaltsverzeichnis, Querverweise und Formularfelder erhalten bleiben.

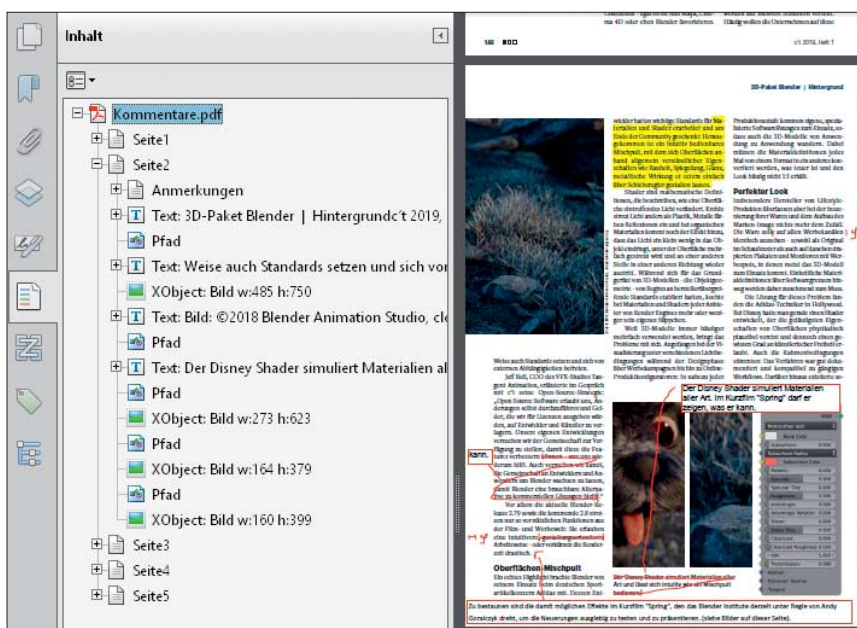
## Zweckdienliche Standards

**?** Was bedeuten PDF/X, PDF/A, PDF/UA und PDF/E?

**!** Seit einigen Jahren wird das Format nicht mehr allein von Adobe weiterentwickelt, sondern von einer ISO-Arbeitsgruppe, die aus mehreren Firmen besteht. Mit dem ISO-Standards PDF 2.0 hat das Konsortium im Jahr 2017 eine Spezifikation geschaffen, die proprietäre Technik verbietet. Vorherige PDF-Versionen erlaubten noch herstellerspezifische Erweiterungen des Formats (Extensions), die mit Betrachtern der Konkurrenz nicht funktionierten.

Außer PDF 2.0 existieren diverse Unterstandards wie PDF/X für die Druckvorstufe, PDF/A für die Langzeitarchivierung, PDF/UA für Barrierefreiheit sowie PDF/E für technische Zeichnungen aus CAD-Anwendungen. Diese Unterstandards verbieten Inhalte oder PDF-Optionen, die beim jeweiligen Einsatzzweck Probleme bereiten können.

PDF/A etwa soll garantieren, dass das Dokument auch nach zig Jahren noch darstellbar ist – und zwar originalgetreu. Daher müssen im Dokument alle Informationen eingebettet sein, die zur Anzeige benötigt werden, zum Beispiel sämtliche Schriftarten. Auch viele dynamische Elemente wie Skripte und Videos gelten als nicht archivierungssicher. Ähnliche Einschränkungen finden sich bei PDF/X; außerdem müssen druckspezifische Angaben in den Metadaten hinterlegt werden, etwa zu Farbräumen, Überfüllung sowie zur Seitengeometrie.



Das Portable Document Format ist ein Container, der außer den sichtbaren Text-, Grafik- und Markup-Objekten auch Skripte und ganze Dateien transportieren kann.

Damit Anwender nicht mühsam in den Spezifikationen nachlesen müssen, was erlaubt ist und was nicht, bringt gute PDF-Software Profile mit, die für die Einhaltung der wichtigsten Richtlinien sorgen. Layout-Programme wie Quark XPress, InDesign oder Scribus betten bei der Auswahl des PDF/X-Profiles automatisch die Schriften ein und befüllen die für den Druckprozess notwendigen Metadaten. Office-Programme wie Word und Libre/OpenOffice hingegen fertigen dank PDF/A- und PDF/UA-Profilen rasch langbelegte und barrierearme Dokumente.

Achtung: Bei der Bearbeitung von PDF/X oder PDF/A in einem Editor können die Eigenschaften der jeweiligen Standards verloren gehen. Wichtig ist daher, die Dokumente wieder als PDF/X et cetera abzuspeichern oder – falls der Editor dies nicht anbietet – mit einem geeigneten PDF-Konverter erneut auszudrucken.

## PDF-Bearbeitung

**?** Womit kann ich PDF am besten bearbeiten, wenn ich Text oder Bilder ändern möchte?

**!** Da PDF eigentlich nicht dafür geschaffen wurde, umfangreich geändert oder gar neu layoutet zu werden, lassen sich die Inhalte auch mit teureren PDF-Editoren nicht grundlegend umbauen. Man kann Bilder austauschen, einzelne Wörter oder Buchstaben korrigieren, im besten Fall auch mehrere Zeilen innerhalb eines Absatzes.

Beim Bearbeiten von Text treten Probleme auf, wenn Schriften weder im System installiert noch im Dokument eingebettet sind. Dann wählt der PDF-Editor eine Ersatzschrift, die unter Umständen nicht hundertprozentig passt. Meist ist nur eine Untergruppe der Schrift im Dokument eingebettet, sodass unter Umständen für die Korrektur benötigte Zeichen fehlen.

Am besten nehmen Sie Änderungen also in der Originaldatei vor und speichern das Ergebnis erneut als PDF. Falls das Original nicht verfügbar ist, behelfen Sie sich mit einem PDF-Editor wie dem kostenlosen LibreOffice Draw oder den kommerziellen Acrobat, Foxit Phantom, PDF-XChange Pro und FlexiPDF. Außerdem gibt es Konverter, die den Inhalt in bearbeitbare Formate exportieren. Das

Ergebnis ist leider oft alles andere als perfekt und muss mitunter aufwendig nachbearbeitet werden. Bei simplen Layouts ohne Spalten und umflossene Bilder gelingen Bearbeitung und Export am besten.

Praktisch: Libre/OpenOffice-Nutzer erstellen auf Wunsch ein Hybrid-PDF, das Original und PDF immer beisammen hält. Dabei wird die Original-ODT-Datei als Attachment ins PDF eingebettet. Beim Öffnen des PDFs startet automatisch der Libre/OpenOffice-Writer, beim Speichern werden sowohl das PDF als auch das Original aktualisiert. Allerdings sollte man bedenken, dass das Original mehr vertrauliche Informationen enthalten kann als das PDF. Vor Veröffentlichung und Weitergabe sollten Sie also überprüfen, ob eine ODT-Datei eingebettet ist oder nicht.

## Sicherer Betrachter

**?** PDF scheint viele Schwachstellen zu haben, man hört ständig von Sicherheitslücken beim Acrobat Reader. Mit welchem Betrachter bin ich auf der sicheren Seite?

**!** Grundsätzlich dürfte jeder Betrachter Schwachstellen haben, die ein entsprechend präpariertes PDF ausnutzen kann, um Schadsoftware einzuschleusen. Durch geschickte Wahl des Viewers – insbesondere für das Öffnen von PDFs aus dem Internet – kann man aber die Wahrscheinlichkeit verringern, dass Angriffe gelingen.

Empfehlenswert sind Betrachter, die potenziell kritische PDF-Eigenschaften nicht unterstützen: Die beispielsweise in Firefox und Chrome eingebauten PDF-Vorschauen sowie Sumatra PDF führen kein JavaScript aus. Das allein stellt schon einen erheblichen Sicherheitsgewinn dar, denn die meisten Exploits missbrauchen die JavaScript-Engine für ihre Zwecke. Außerdem gilt: Je verbreiteter ein Betrachter, umso wahrscheinlicher wird er Ziel eines Angriffs. Ganz oben auf der Angriffsliste stehen demzufolge Acrobat Reader und die populäre Alternative Foxit Reader. Beide bieten immerhin einen eingeschränkten Sandbox-Modus, der die Sicherheit erhöht – aber ebenfalls bereits Ziel von Exploits war.

Wer sich für einen der spartanischen Betrachter entscheidet, verzichtet allerdings auch auf Komfort und blockiert mit-

unter nützliche Eigenschaften des Dokuments. So funktionieren interaktive Formulare mit Berechnungen und Plausibilitätsprüfung ohne JavaScript nicht. Ein Kompromiss: erste Sichtung von PDFs (im Browser) mit einem funktionsreduzierten Viewer, Öffnen vertrauenswürdiger Dokumente im voll ausgebauten Betrachter oder Editor.

## Inhalte dauerhaft löschen

**?** Wenn ich weiße Seitenränder mit dem Crop-Werkzeug abschneide, bleibt die Dateigröße gleich. Woran liegt das und wie kann ich abgeschnittene Inhalte dauerhaft entfernen?

**!** PDF-Editoren verhalten sich in vielerlei Hinsicht anders, als man es von normalen Editoren gewohnt ist. Unter anderem gibt es Bearbeitungswerkzeuge, die gar nicht den Inhalt selbst verändern, sondern lediglich die Art und Weise, wie er angezeigt wird. So verhält sich etwa das Werkzeug mit dem irreführenden Namen „Crop“: Anders als die von Bildbearbeitungsprogrammen bekannte Funktion blendet PDF-Crop den abgeschnittenen Inhalt lediglich aus, löscht ihn aber nicht dauerhaft. Die einfachste Art, den ausgeblendeten Inhalt loszuwerden, ist der erneute Ausdruck des beschnittenen Dokuments in ein PDF.

Auch Inhalte, die man mithilfe der Kommentarwerkzeuge (Markup) ergänzt hat – Textkästen, Markierungen, Handschrift-Unterschriften et cetera – schweben über dem Dokument und können von jedermann nachträglich gelöscht oder verändert werden. Sensible Informationen dürfen also niemals mit schwarzen Kästen aus dem Markup-Sortiment unkenntlich gemacht werden, da der Text dahinter les- und durchsuchbar bleibt. Kommerzielle PDF-Editoren bieten dafür spezielle Schwärzungstools.

Wer die Handschrift-Signatur, ausgefüllte Formularfelder oder andere Markups vor der Weitergabe mit dem Dokument verschmelzen möchte, benutzt die Flatten-Funktion (etwa im kostenlosen Betrachter Xodo). Darüber hinaus lässt sich das Bearbeiten des Dokuments über die Sicherheitseinstellungen unterbinden.

(atr@ct.de)

**Quellen und Dokumentationen:**  
[ct.de/y5bx](http://ct.de/y5bx)