

**White Paper**

# Meta-Daten in Microsoft Office und in PDF-Dokumenten

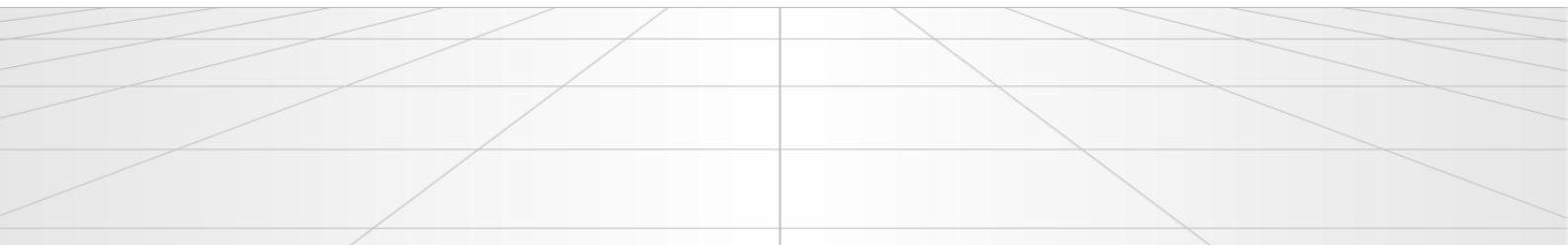
Typen, Übernahme, Anzeige und Löschen

Copyright 2002 - 2009 soft Xpansion GmbH & Co. KG

## Inhaltsverzeichnis

Begriffsdefinition und Typen.....	2
Definitionen.....	2
Einige Typen .....	2
Meta-Daten in Office-Dateien.....	2
Word .....	4
Excel.....	5
PowerPoint .....	5
Meta-Daten in PDF-Dokumenten.....	6
Allgemeine Meta-Daten .....	6
Meta-Daten aus Office-Dokumenten.....	6
Beispiele zur Übernahme.....	7
Darstellung und Anzeige .....	9
Entfernen von Meta-Daten.....	10


**Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind unter Umständen eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.**



## Begriffsdefinition und Typen

### **Definitionen**

Unter Meta-Daten sind nach einer ersten, ganz allgemeinen Definition Daten zu verstehen, die andere Daten beschreiben beziehungsweise Informationen über diese Daten geben. So sind beispielsweise die Buchstaben oder Zeichen in einem Text Daten. Die Anzahl der Buchstaben in dem Text sind Meta-Daten, die auch als Zusatzinformationen über die „eigentlichen“ Daten hinaus bezeichnet werden können.

Wenn man den Begriff etwas enger fasst und im Zusammenhang mit Dateitypen verwendet, sind zum Beispiel Informationen wie der Name und Titel einer Datei, ihres Verfassers, Stichwörter zum Inhalt der Datei oder das Datum der Speicherung Meta-Daten. Dann können Meta-Daten oder Meta-Informationen so charakterisiert werden, dass sie in einer Datei abgespeichert werden, aber in der Regel nicht auf den ersten Blick sichtbar sind, wenn man die Datei in einer Anwendungssoftware öffnet. Auch gehen beim Ausdruck eines Dateiinhalts (zum Beispiel eines Dokuments) auf Papier die meisten Meta-Daten verloren. Da dies bei der Weitergabe von Dateien in der Regel nicht passiert, können Dateien mehr Informationen enthalten, als der Ersteller eigentlich preisgeben will.<sup>1</sup> 

Ein zweifellos positiver Effekt von Meta-Daten in Dateien ist andererseits: Sie können dazu dienen, Datenbestände nach bestimmten, allgemeinen Kriterien leichter zu katalogisieren und gezielter durchsuchen zu können.

### **Einige Typen**

Neben den Meta-Daten in Textdokumenten enthalten zum Beispiel auch Musik- und Videodateien, Fotos oder HTML-Seiten Meta-Informationen. In Musikdateien sind dies unter anderem Informationen über den Interpreten und den Titel oder spezielle Tags. Bei Videos zum Beispiel die Wiedergabedauer, Größe und Bildrate. In Fotos sind Meta-Daten in Form von Bild-Informationen häufig im Exif- oder IPTC-Format abgespeichert. Meta-Daten in HTML-Seiten (Seitentitel, Stichwörter zum Seiteninhalt etc.) werden angelegt, damit die Seiten in den Trefferlisten von Suchmaschinen angezeigt werden.

Die folgenden Ausführungen behandeln ausschließlich die Verwendung von Meta-Daten in Microsoft-Office-Dateien und in PDF-Dokumenten. Dabei werden die drei Office-Anwendungen Word, Excel und PowerPoint betrachtet.

## Meta-Daten in Office-Dateien

### **Office-Dokumente allgemein**

In Office-Dateien wird beim Speichern des Dokuments eine Vielzahl an Meta-Informationen abgelegt. Dies sind zum einen Informationen, die in Word, Excel und PowerPoint als „Dokumenteigenschaften“ bezeichnet werden und bei allen drei Dokumenttypen an derselben Stelle der Anwendung eingesehen (und gegebenenfalls geändert) werden können. Zum anderen sind es Informationen, die nur in der jeweiligen Anwendung vorkommen.

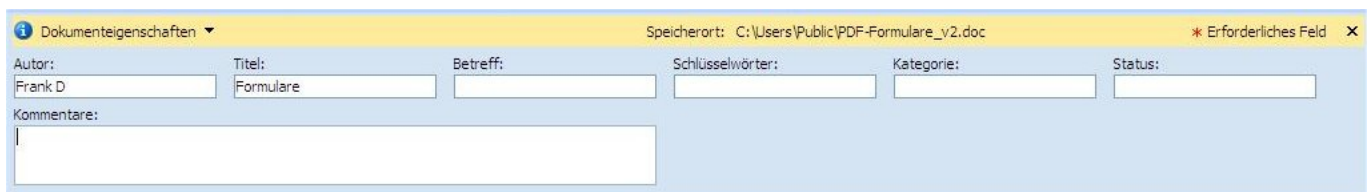
---

<sup>1</sup>Siehe Rost, Martin; Wallisch, Arnold, Was Office-Dateien verraten können, c't Heft 2-2003, Seite 172ff.

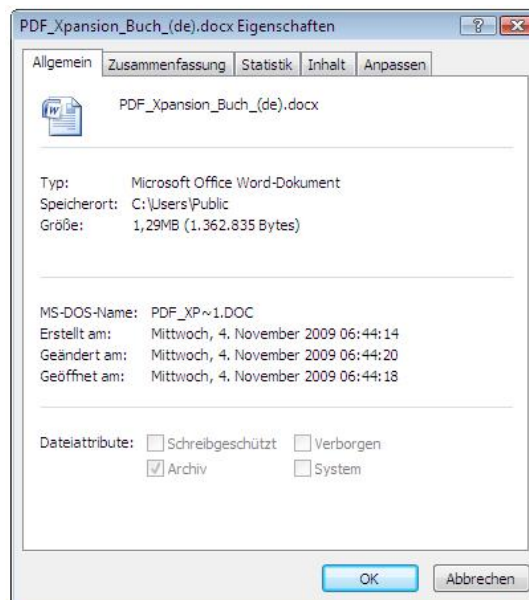
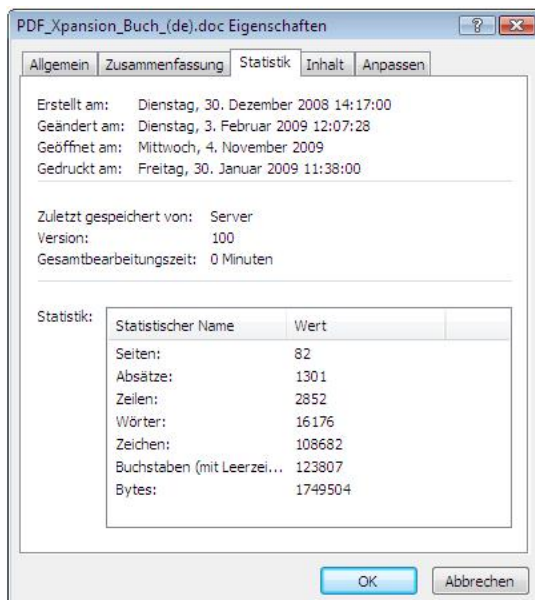
Einen Teil dieser Informationen (die Dokumenteigenschaften) kann man ansehen, wenn in den Versionen bis Office 2003 im Menü „Datei“ auf „Eigenschaften“ geklickt wird. In Word, Excel und PowerPoint 2007 wird zunächst auf die Microsoft Office-Schaltfläche, dann auf „Vorbereiten“ und schließlich auf „Eigenschaften“ geklickt. Alternativ können in Office 2007 die Eigenschaften auch beim Öffnen oder Speichern einer Datei angezeigt werden, und zwar im Dialogfeld „Öffnen“ oder im Dialogfeld „Speichern unter“. Die meisten der Dokumenteigenschaften können für das aktuell geöffnete Dokument nicht nur angesehen, sondern auch geändert werden. Zudem lassen sich neben den standardmäßig vorgegebenen auch benutzerdefinierte Eigenschaften erstellen<sup>2</sup>

Zu den wichtigsten, allgemeinen Dokumenteigenschaften gehören (am Beispiel Word 2007 erläutert):

- ❖ Standardeigenschaften, die vom Anwender im Anzeigedialog geändert werden können. Hierunter fallen Autor (automatisch der Name, den der Anwender bei der Installation angegeben hat), Titel, Betreff, Schlüsselwörter, Kommentare.



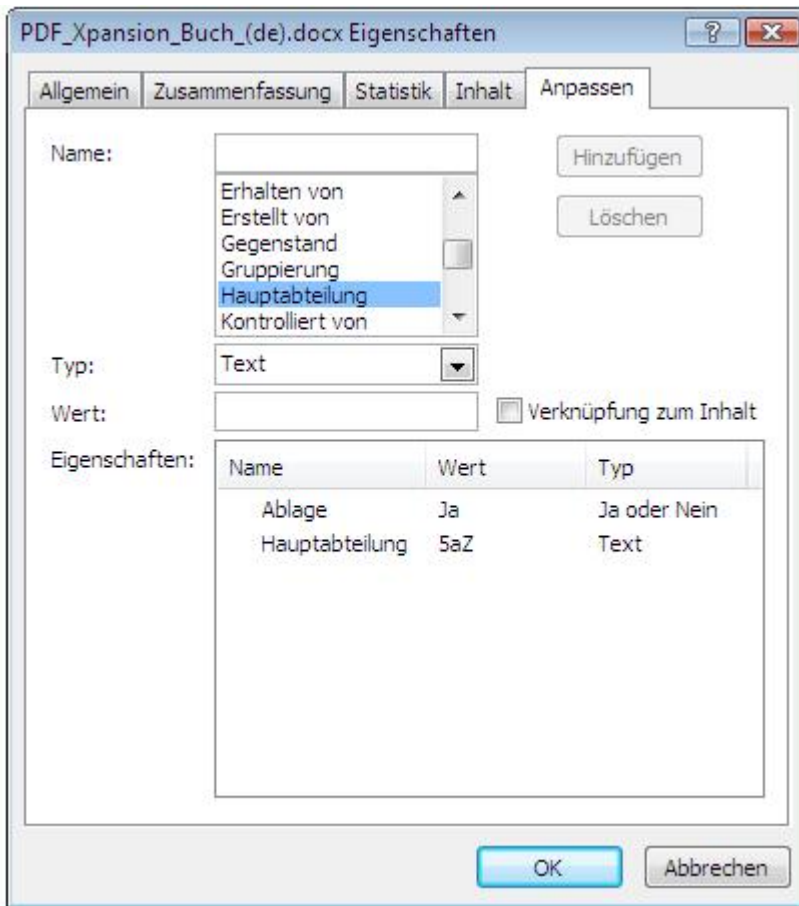
- ❖ Dynamisch aktualisierte oder aktualisierbare Eigenschaften, die vom Anwender im Anzeigedialog nicht geändert werden können. Hierzu zählen Speicher- und Bearbeitungszeiten, Speicherpfad, statistische Daten (Anzahl Seiten, Absätze, Zeilen, Wörter etc.).



<sup>2</sup> Zu Details siehe die Hilfe und Anleitungen auf <http://office.microsoft.com/de-de/help/HA100475241031.aspx>

Zu den dynamischen Metainformationen gehören ganz allgemein auch alle Autotextfelder, also „Speicherorte“ für mehrfach verwendeten Text oder Grafiken, die wieder verwendet werden sollen.

- ❖ Benutzerdefinierte Eigenschaften, die mit Zahlen, Texten, Zeitangaben oder dem Wert „Ja“/ „Nein“ versehen werden können.



Neben den allgemeinen Dokumenteigenschaften sind zudem weitere Meta-Informationen in Office-Dateien gespeichert, die jeweils für Word, Excel und PowerPoint spezifisch sind.

## **Word**

Anwendungsspezifische Meta-Daten in Word-Dateien sind zum Beispiel

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)
- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen den Seiten eines Word-Dokuments und externe Links beziehungsweise Internet-Links
- ❖ Inhaltsverzeichnis (Table of Contents, TOC) mit Sprung-Informationen, welche Seite beim Anklicken einer Überschrift im Inhaltsverzeichnis geöffnet werden soll
- ❖ Informationen über Formularfelder, allgemeine Steuerelemente und Webtools

- ❖ (Ausgeblendete) Anmerkungen und Kommentare
- ❖ Fußnoten, Endnoten
- ❖ Hinweise auf Änderungen
- ❖ (Ausgeblendete) Markup-Hinweise
- ❖ Ausgeblendete Texte und (nicht sichtbare) Seitenhintergründe
- ❖ Das Stichwortverzeichnis (der Index)

## **Excel**

Anwendungsspezifische Meta-Daten in Excel-Arbeitsmappen sind

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)
- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen Blättern einer Excel-Arbeitsmappe und externe Links beziehungsweise Internet-Links
- ❖ Informationen über vorhandene Formularfelder oder allgemeine Steuerelemente
- ❖ Kommentare
- ❖ Gitternetzlinien
- ❖ Zeilen- und Spaltenköpfe

## **PowerPoint**

Anwendungsspezifische Meta-Daten in PowerPoint-Präsentationen sind

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)
- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen den Folien einer Präsentation und externe Links beziehungsweise Internet-Links
- ❖ Informationen über vorhandene Formularfelder oder allgemeine Steuerelemente
- ❖ Notizen/Kommentare
- ❖ Gitternetzlinien
- ❖ Zeilen- und Spaltenköpfe
- ❖ (Ausgeblendete) Folien
- ❖ Folienrahmen
- ❖ Angepasste Graustufen-Darstellung

## Meta-Daten in PDF-Dokumenten

### **Allgemeine Meta-Daten**<sup>3</sup>

PDF-Dateien enthalten unabhängig vom Dateiformat der Ursprungsanwendung üblicherweise allgemeine Meta-Daten wie den Titel und Verfasser des Dokuments sowie Informationen zu seiner Erstellung und zu Änderungen. In PDF-Dateien können Meta-Daten durch den Anwender oder durch Plugin-Erweiterungen geändert oder neu hinzugefügt werden, wobei die Darstellung in PDF-Viewern von Adobe bei den Dokumenteigenschaften auf maximal 255 Bytes begrenzt ist.<sup>4</sup> Mit der Version 1.4 der PDF-Spezifikation können Meta-Daten auch für einzelne Komponenten eines Dokuments definiert werden. In einem PDF-Dokument werden die Meta-Daten entweder in einem Dokumentinformationswörterbuch (Document Information Directory) oder in Meta-Daten-Strömen (Metadata Streams) gespeichert.

### **Meta-Daten aus Office-Dokumenten**

Sofern es sich bei der Ausgangsdatei um eine Office-Datei aus Word, Excel oder PowerPoint handelt, können mit PDF-Convertern, die die Übernahme von Meta-Daten aus Office-Dateien beherrschen, in der Regel darüber hinaus die oben für die drei Programme aufgelisteten anwendungsspezifischen Metainformationen in die PDF-Datei übernommen werden. Solche Converter sind zum Beispiel [Office to PDF Premium](#), [Word to PDF](#) oder [Excel to PDF](#) von soft Xpansion.

Da es sich bei den Office-Dokumenten einerseits und bei den PDF-Dateien andererseits aber um separat voneinander entwickelte Dateitypen handelt, erfolgt die Übernahme vom einen Dateityp in den anderen nicht für alle Arten von Meta-Daten „deckungsgleich“. Aus diesem Grund sind hinsichtlich der Übernahme folgende Gegebenheiten zu beachten:

1. Nicht zwangsläufig alle Meta-Daten aus Office-Dateien können in PDF-Dokumente importiert werden. So fehlen für einige Meta-Daten im Objektmodell des PDF-Formats die technischen Voraussetzungen für eine Übernahme, zum Beispiel bei den Autotextfeldern. Diese Einschränkung gilt allerdings nicht für deren im Dokumenttext angezeigten Feldinhalt.
2. Umgekehrt bietet in manchen Fällen die Übernahme in die PDF-Datei mehr Funktionalität, als in der Office-Datei vorhanden ist. So ist es technisch zum Beispiel möglich, ein Stichwortverzeichnis (Index) aus einer Word-Datei so in eine PDF-Datei zu importieren, dass die einzelnen Einträge im Verzeichnis der PDF-Datei als aktive Verweise angezeigt werden. Auf diese Verweise kann man dann klicken, um zu der Seite zu springen, die im Index angegeben ist.

---

<sup>3</sup> Siehe Adobe Systems Incorporated, PDF Reference, Version 1.7, Seiten 843 - 847, zu den Grundlagen von Meta-Daten in PDF-Dateien

<sup>4</sup> Ebenda, Seite 1125, Anmerkung 160.

Problematisch kann hier sein, dass im PDF-Format, anders als im Word-Dokument, aber keine Möglichkeit besteht, nach dem Hinzufügen oder Löschen von PDF-Seiten den Index in PDF-Datei mit einem „Felder aktualisieren“-Befehl automatisch auf den neuesten Stand zu bringen.

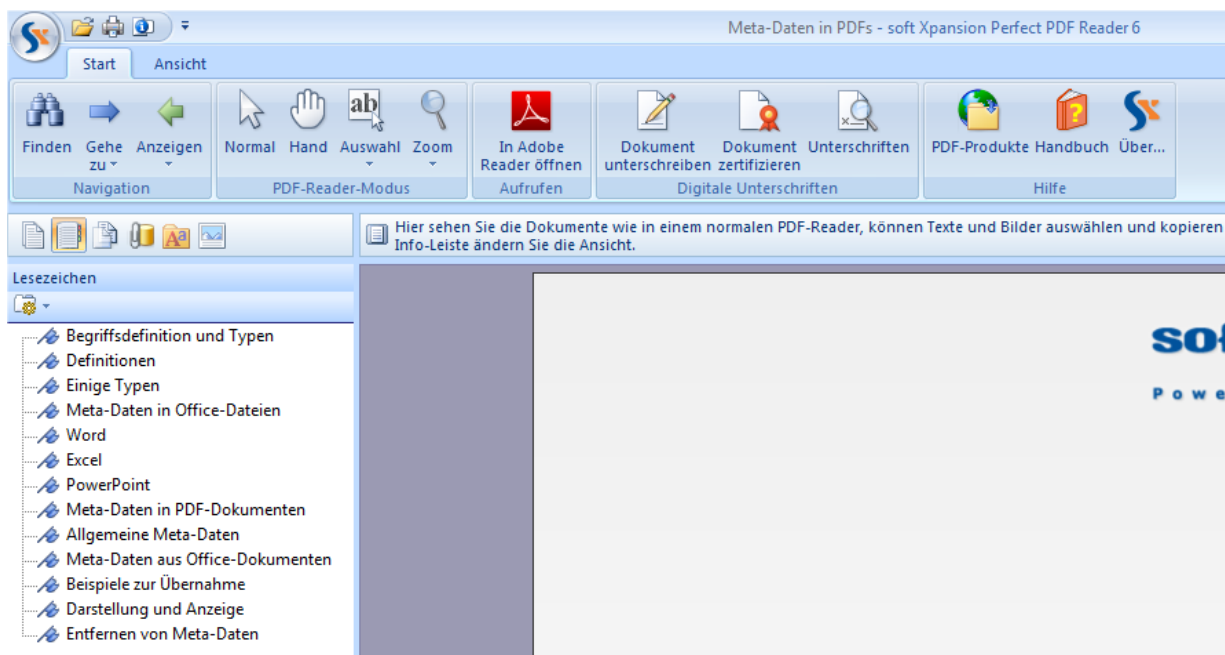
3. Schließlich bestehen bei der Umsetzung der Übernahme für die Entwickler von PDF-Konvertern Freiheitsgrade, wie diese konkret ausgestaltet wird.

## **Beispiele zur Übernahme**

Die folgenden Informationen zu Inhaltsverzeichnis/TOC, Fußnoten/Endnoten und Textüberarbeitungen zeigen exemplarisch, dass die Übernahme und Darstellung von Meta-Daten auf unterschiedlichen Wegen erfolgen kann.

### **1. Inhaltsverzeichnis/TOC**

In Word können Inhaltsverzeichnisse automatisch aus den Überschriften von Kapiteln, Abschnitten und Unterabschnitten erzeugt werden. Ein solches Inhaltsverzeichnis wird am Anfang des Word-Dokuments als Sammlung aktiver Links platziert. Man gelangt von ihm aus mit einem Mausklick auf einen Link an die Stelle im Dokument, an der der jeweilige Textteil beginnt.



**Anzeige des Inhaltsverzeichnisses in einem PDF-Viewer (hier [Perfect PDF Reader](#)) als Lesezeichen**

In der PDF-Datei kann das Inhaltsverzeichnis mit derselben Funktion angezeigt werden wie in der Word-Datei, das heißt, man kann von ihm aus direkt den jeweiligen Verknüpfungen folgen und zu einer bestimmten Textüberschrift springen. Zusätzlich besteht hier aber noch die Möglichkeit, das Verzeichnis als Sammlung von Lesezeichen in Form einer Liste (Verzeichnisbaum) einzubetten, die nicht im

Dokumenttext selbst, sondern getrennt von ihm in einem eigenen Fenster-Bereich des PDF-Viewers angezeigt wird.

## 2. Fußnoten/Endnoten

Fußnoten und Endnoten werden in Word so eingebettet, dass der Fuß- oder Endnotentext oberhalb des Fuß-/Endnotenzeichens in einem Infocfeld angezeigt wird, wenn man den Mauszeiger im Dokument auf das Fuß-/Endnotenzeichen bewegt. Wenn man doppelt auf das Zeichen klickt, springt die Anzeige außerdem direkt zu der zugehörigen Fuß- oder Endnote.

Einen Teil dieser Informationen (die Dokumenteigenschaften) kann man ansehen, wenn in den Versionen bis Office 2003 im Menü „Datei“ auf „Eigenschaften“ geklickt wird. In Word, Excel und PowerPoint 2007 wird zunächst auf die Microsoft Office-Schaltfläche, dann auf „Vorbereiten“ und schließlich auf „Eigenschaften“ geklickt. Alternativ können in Office 2007 die Eigenschaften auch beim Öffnen oder Speichern entweder im Dialogfeld „Öffnen“ oder im Dialogfeld „Speichern“ angesehen, und zwar im Dialogfeld „Öffnen“ oder im Dialogfeld „Speichern“. Dokumenteneigenschaften können für das aktuell geöffnete Dokument angesehen, sondern auch geändert werden. Zudem lassen sich neue vorgegebenen auch benutzerdefinierte Eigenschaften erstellen<sup>2</sup>.

Zu Details siehe die Hilfe und Anleitungen auf <http://office.microsoft.com/de-de/help/HA100475241031.aspx>

In PDF-Dateien werden Fuß- und Endnoten entweder genauso importiert oder aber so, dass nur die eine oder andere Variante gewählt wird. Bei [Office to PDF Premium und Word to PDF](#) von soft Xpansion wird beispielsweise nur das Infocfeld übernommen. Der Doppelklick hat hier eine andere Funktion: Das Feld wird in diesen Programmen um ein Symbol ergänzt und kann per Doppelklick aus- oder dauerhaft eingeblendet werden.

Informationen verloren. Da dies bei der Weitergabe von Dateien in der Regel nicht passiert, können Dateien mehr Informationen enthalten, als der Ersteller eigentlich preisgeben will.<sup>1</sup>

Ein zweifellos positiver Nutzen sind die Datenbestände, die durch die Dienste und gezielter durchsucht werden können.

### Einige Typen

Neben den Meta-Daten von Textdateien, Videodateien, Fotos oder anderen Informationen, zum Beispiel die Wiederherstellung von Bild-Informationen von HTML-Seiten (Seitentitel, Seitennummer) können auch Seiten in den Trefferlisten von Suchmaschinen angezeigt werden.

Die folgenden Ausführungen behandeln ausschließlich die Verwendung von Meta-Daten in Microsoft-Office-Dateien und in PDF-Dokumenten. Dabei werden die drei Office-Anwendungen Word, Excel und PowerPoint betrachtet.

## 3. Textüberarbeitungen (Hinweise auf Änderungen, Markup-Hinweise)

In Word lassen sich bekanntermaßen Textüberarbeitungen nachverfolgen, indem die konkreten Änderungen protokolliert, im Text farbig angezeigt, als Sprechblasen am Rand dargestellt sowie in einem Infocfeld („Gelöscht“, „Eingefügt“) zusätzlich hervorgehoben werden.

3. Schließlich bestehen bei der Umsetzung der Übernahme für die Entwickler von PDF-Convertern Freiheitsgrade, wie diese konkret ausgestaltet wird.

Gelöscht: für

### Beispiele zur Übernahme

Die folgenden Informationen zu Inhaltsverzeichnis/TOC, Fußnoten/Endnoten und Textüberarbeitungen zeigen exemplarisch, dass die Darstellung dieser Meta-Daten in Office-Dokumenten einerseits und in PDF-Dateien andererseits nicht zwangsläufig identisch erfolgen muss.

Die korrespondierende PDF-Datei kann Textüberarbeitungen genauso anzeigen - oder auf andere Weise. In [Office to PDF Premium und Word to PDF](#) von soft Xpansion werden als Hinweise auf Änderungen für Textlöschungen standardmäßig rote Fragezeichen-Symbole und ebenso umrahmte Infofelder verwendet. Und Textergänzungen erkennt man hier an blauen Symbolen und in derselben Farbe umrahmten Infofeldern. Diese enthalten gegenüber der Darstellung in Word mit dem Änderungsdatum, der Änderungszeit und dem Computernamen, auf dem die Änderung durchgeführt wurde, zusätzliche Informationen.

3. Schließlich bestehen bei der Umsetzung der Übernahme für die Entwickler von PDF-Convertern Freiheitsgrade, wie diese konkret ausgestaltet wird.

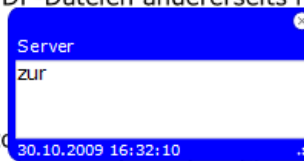
### Beispiele für Übernahme

Die folgenden Informationen zu Inhaltsverzeichnis/TOC, Fußnoten/Endnoten und Textüberarbeitungen zeigen exemplarisch, dass die Darstellung dieser Meta-Daten in Office-Dokumenten einerseits und in PDF-Dateien andererseits nicht zwangsläufig identisch erfolgen muss.

#### 1. Inhaltsverzeichnis/TOC

In Word können Inhaltsverzeichnisse automatisch aus den Titeln von Kapiteln, Abschnitten und Unterabschnitten erzeugt werden. Ein solches Inhaltsverzeichnis wird am Anfang des Word-Dokuments als Sammlung aktiver Links dargestellt. Man kann von ihm aus mit einem Mausklick auf einen Link an die Stelle springen, an der der jeweilige Textteil beginnt.

In PDF-Dateien wird ein Inhaltsverzeichnis mit derselben Funktion angezeigt. In der PDF-Datei, das heißt man kann von ihm aus direkt den jeweiligen Verknüpfungen folgen und zu einem bestimmten Abschnitt springen. Zusätzlich besteht hier aber noch die Möglichkeit, das Verzeichnis als Sammlung



## Darstellung und Anzeige

Meta-Informationen wie Verfasser, Titel, Thema oder Schlüsselwörter werden, sofern eine Übernahme beim Erstellen der Datei bestimmt wurde, in PDF-Dateien üblicherweise in Dialogfenstern angezeigt, die mit „Dokumenteigenschaften“ oder ähnlich bezeichnet sind. Andere Meta-Daten wie zum Beispiel im Originaldokument ausgeblendete Anmerkungen, Notizen und Kommentare, Gitternetzlinien, Folien oder Folienrahmen werden in das PDF-Dokument übernommen und dort auch sichtbar angezeigt, wenn die Übernahme beim Erstellen der PDF-Datei festgelegt beziehungsweise erlaubt wurde.

## Entfernen von Meta-Daten

Bis hier wurde erläutert, welche Meta-Daten aus Office-Dateien in PDFs übernommen werden können und wie dies erfolgen kann. In bestimmten Fällen ist es allerdings sinnvoll, vor der Konvertierung in PDF und Weitergabe einer Datei alle oder bestimmte in ihr enthaltenen Meta-Daten zu entfernen. Nämlich dann, wenn der Verfasser des Dokuments sie nicht bekannt geben möchte. Hier gibt es prominente Fälle, in denen dies aus Sicht der Verfasser versäumt wurde, wie den des Dokuments über die im Irak vermuteten Massenvernichtungswaffen, die den Einstieg Großbritanniens in den Irak-Krieg rechtfertigen sollten<sup>5</sup>. Eine Vielzahl an Meta-Daten aus Office-Dateien komfortabel entfernt werden kann, wird im Folgenden erläutert.

Für das Löschen von Meta-Daten in den Originalanwendungen Word, Excel oder PowerPoint stehen folgende Wege zur Verfügung:

1. Zunächst besteht die Möglichkeit, Programm-Einstellungen innerhalb von Word, Excel und PowerPoint zu nutzen, um als sensibel erachtete Daten aus den Datei-Eigenschaften zu entfernen.<sup>6</sup>
2. Zudem können spezielle von Microsoft zur Verfügung gestellte Tools verwendet werden, um einige Meta-Daten zu löschen:

Bei Office bis Version 2003 erfolgt dies, wie auf der Seite <http://office-2003-xp-add-in-zum-entfernen-von-metadaten.shareware.de/> erläutert.

Bei Office 2007 erfolgt das Entfernen mit Hilfe einer *Dokumentinspektor* genannten Funktion, siehe

<http://office.microsoft.com/de-de/help/HA100375931031.aspx>

Darüber hinaus gibt es Freeware und ein kommerzielles Tool von einem anderen Hersteller, den *Document Trace Remover* und den *Metadata Analyzer*.

---

<sup>5</sup> [http://www.chip.de/artikel/Verraeterisches-Office-2\\_12875097.html](http://www.chip.de/artikel/Verraeterisches-Office-2_12875097.html)

<sup>6</sup> [http://www.chip.de/artikel/Verraeterisches-Office-4\\_12875104.html](http://www.chip.de/artikel/Verraeterisches-Office-4_12875104.html)