

PM Doc~Flow Report

Nr. 01/ 2006



PDF/A: Das richtige Format für die Langzeitarchivierung?

Das populäre PDF-Format kennt man schon lange – kaum ein PC, auf dem der Acrobat-Reader nicht zu finden ist. Jetzt gibt es eine neue Variante: „PDF/A“ – dabei steht das „A“ für „Archiv“. Was verbirgt sich dahinter?

Im DMS- und Archivierungsumfeld befasst man sich immer wieder mit der Diskussion um das „richtige“ Format für die zu speichernden Dokumente, wobei es das „richtige“ Format gar nicht gibt.

Die Diskussion wird immer dann konkreter, wenn es um die langfristige Archivierung der Dokumente geht. Dies umfasst Zeiträume von 10 Jahren und mehr! Steuerrelevante und betriebswirtschaftliche Unterlagen sind 6 oder 10 Jahre aufzubewahren. Wenn man aber z.B. über rentenrelevante Unterlagen, technische Anlagendokumentationen oder über Dokumente im Zusammenhang mit einer Produkthaftung spricht, dann ist man schnell in Dimensionen von 30 Jahren und mehr. Die technischen Probleme beginnen aber oft schon mit Dokumenten, die unter 10 Jahre alt sind. Die Ursachen sind unter anderem die Weiterentwicklung der Editoren, der Wechsel der Rechnerplattform und ein Tausch von Speichermedien.

Bevor man die Alternativen zur Speicherung diskutiert, sind die Anforderungen die ein Format für die langfristige Archivierung erfüllen soll, zu klären. Primär sind immer folgende Punkte zu beachten.

- **Langfristige Stabilität:** Das Format soll sich nicht ständig ändern. Eine (möglichst internationale) Normung erhöht diese Sicherheit. Auch wenn Konverter zur Umsetzung in eine neue Version zur Verfügung stehen, bedeutet das immer Aufwand und die Gefahr von Informationsverlusten.
- **Wiedergabetreue:** Bei der Wiedergabe der Dokumente muss eine inhaltliche und layoutmäßige Integrität gewährleistet werden, unabhängig davon auf welchem Medium das Dokument ausgegeben wird.
- **Hohe Verbreitung:** Ein Format, das eine hohe Verbreitung hat, ist tendenziell auch langlebig. Hinzu kommt, dass die Dokumente dann auch leichter bei veränderten organisatorischen Verhältnissen (Fusion, Outsourcing, etc.) und den damit verbundenen technischen Veränderungen weiter genutzt werden können. Im schlimmsten Fall, wenn das Format „sterben“ sollte, findet man bei einem gängiges Format leichter einen Konverter für das neue Zielformat, als bei einem Exoten.
- **Unabhängigkeit** von der Ausgangsapplikation: Damit möglichst viele Nutzer die Dokumente einfach und kostengünstig lesen können, sollte es preiswerte oder frei verfügbare Viewer geben, die auf möglichst allen relevanten technischen Plattformen zur Verfügung stehen.

Bisher war das TIFF-Format immer der Top-Favorit für eine langfristige Archivierung. Es hat praktisch alle oben genannten Anforderungen erfüllt. Ursprünglich als Übertragungsformat für Telefax entwickelt, hat es inzwischen zahlreiche Erweiterungen

und Verbesserungen erfahren. Der größte Nachteil ist, dass es als reines Bitmap-Format inhaltlich nicht direkt interpretierbar ist, sondern zu diesem Zweck durch OCR-Verfahren konvertiert werden muss. Dieser Vorgang ist immer mit einer gewissen Fehlerrate behaftet. Dazu kommen weitere Punkte, wie hoher Speicherbedarf, verschiedene Varianten und Erweiterungen.

Andere Formate haben – oft in speziellen Umgebungen – eine gewisse Verbreitung erfahren, konnten sich aber nie wirklich für eine langfristige Archivierung etablieren. Einzig das PDF-Format der Firma Adobe hat sich langsam den Status eines Quasi-Standards erworben, da es praktisch alle obigen Punkte erfüllte und darüber hinaus weitere Vorteile bietet. Ein Hauptthemmnis war bisher aber immer die mangelnde Normung durch ein internationales Gremium und damit die hohe Abhängigkeit von der Firmenpolitik von Adobe.

Nun liegt mit der **Norm ISO 19005-1** das standardisierte Dateiformat **PDF/A** vor. Es basiert auf der (älteren) PDF-Version 1.4, hat aber weniger Funktionalitäten, als die aktuelle PDF-Version von Adobe. Damit will man es in Bezug auf das Ziel der langfristigen Archivierung optimieren. Und so fehlen hier einige Funktionen für Annotation, Watermarks, Signaturen usw.. Aber alle für die Archivierung notwendigen Features sind vorhanden. Es müssen in PDF/A alle verwendeten Schriften im Dokument eingebettet sein, verboten sind Audio- und Video-Daten ebenso wie Java-Scripts. Dadurch werden Elemente verbannt, die die langfristige Nutzbarkeit des Dokuments gefährden könnten. Geregelt werden in der Norm die Graphiken, Schriftarten, Metadaten und die logische Struktur.

Der große Vorteil gegenüber dem TIFF-Format ist, dass eine Volltextrecherche möglich ist. Da es sich um eine Seitenbeschreibungssprache handelt, ist das System sehr flexibel und erweiterbar. Darüber hinaus bietet das PDF-Format weitere wesentliche Vorteile für das Dokumenten-Management, wie die Einbindung von elektronischen Signaturen.

Aus diesen Gründen wird das PDF/A-Format voraussichtlich immer mehr in der Archivierung und dem Dokumenten-Management genutzt werden und kann wahrscheinlich mittelfristig dem TIFF-Format den Rang ablaufen. Auch ist zu erwarten, dass Applikationen, die heute schon eine Archivierung mit dem PDF-Format anbieten (wie SAP) auch bald die PDF/A-Variante anbieten werden.

Nichts ist für die Ewigkeit! Auch heute sollte man sich schon bewusst machen, dass auch PDF-Dokumente irgendwann in ein anderes Format konvertiert werden müssen. Da auch die Datenträger physisch altern und auch deren Format sich ändert, werden die Daten auf den Datenträgern nach einigen Jahren auf neue Datenträger kopiert. Bei dieser Gelegenheit kann auch problemlos eine Konvertierung der Datenformate eingeplant und umgesetzt werden.