

# Verarbeitung von PDF-Dateien

## Szenarien für die Erfassung, Aufbereitung und Archivierung von PDF Dateien für das ZDA der Finanz Informatik

### Das allgegenwärtige PDF

PDF-Dateien sind aus dem Dokumentenmanagement nicht mehr wegzudenken.

Auch im Umfeld der Sparkassen fallen sowohl im Tagesbetrieb als auch aus Altbeständen oder Konvertierungen PDF-Dokumente an, die systemkonform in das Zentrale Dokumentenarchiv ZDA der Finanz Informatik überführt werden müssen.

Für die Verarbeitung sind insbesondere folgende Aspekte relevant:

- Aus welchen Quellen stammen die PDF-Dokumente?
- Sind die PDF-Dokumente im passenden Archiv-Format der Finanz Informatik?
- Sind die Kunden-/Kontoinformationen für die Indizierung aus den PDF-Dokumenten ersichtlich?



### Korrektes Format

Um PDF-Dokumente im ZDA anliefern und archivieren zu können, müssen diese im Format **PDF/A-1b** vorliegen. Dieses Format gewährleistet die spätere visuelle Reproduzierbarkeit der Dokumente für Recherchen. Aus diesem Grund müssen für PDF/A-1b-Konformität alle verwendeten Schriften und Farben im Dokument eingebettet sein.

Sparkassen können z.B. über den DocuPrinter, den die Finanz Informatik anbietet, Dokumente manuell(!) „drucken“ (Drucken -> DocuPrinter) und somit in das PDF/A-1b-Format konvertieren und anschließend archivieren. Auch über das Scannen von gedruckten Dokumenten kann das passende Archivformat erzeugt werden.

### Automatische Archivierung in WANDO

Für die effiziente und automatische Massenverarbeitung kommt die Wissensbasierte Dokumentverarbeitung **WANDO** zum Einsatz. Die Verarbeitung in **WANDO** beinhaltet die Komponenten Interface und Imageverarbeitung sowie alternativ den **WebUpload** und die **Archivierung direkt per E-Mail-Anhang**.

Der Ablauf der Archivierung ist davon abhängig, in welcher Form die Ursprungsdaten vorliegen.

Voraussetzung ist immer, dass die Dokumente auf einem Speichermedium/Festplattenbereich bereitliegen.

Folgende Varianten werden unterschieden:

- A:** Keine Indexdaten vorhanden
- B:** Indexdaten in Dateinamen enthalten
- C:** Indexdaten als Indexdatei vorhanden

#### **A: Keine Indexdaten vorhanden**

In dieser Variante nehmen die Dokumente den normalen Weg über die Indizierung, wie andere gescannte Dokumente auch. Der Ablauf in **WANDO** gestaltet sich wie folgt:

- Konfiguration des automatischen Imports des Plattenbereiches des Mandanten
- Zyklische Abholung der PDF-Dateien (z.B. stündlich) Anlegen als „Scan“-Stapel in **WANDO**
- Aufbereitung per Imageverarbeitung für die Archivierung in PDF/A-1b (dabei auch Erzeugung von Bildern der einzelnen PDF-Seiten, somit Volltexterkennung möglich)
- Indizierung der Dokumente in der Korrektur

Liegen zu den PDF-Dateien bereits Indexkriterien wie Dokumenttyp, Kontonummer, Personennummer und Datum vor, kann die Indizierung vollautomatisiert wie bei Barcodedokumenten laufen.

Die Indexdaten können entweder aus dem Dateinamen der PDFs entnommen werden oder sie liegen als separate Indexdatei vor.

### **B: Indexdaten in Dateinamen enthalten**

Beispiel:

Der Dateiname *0000004711041130915.pdf* enthält die folgenden Angaben:

- Kontonummer: 4711
- Dokumenttyp : 041
- Datum: 13.09.2015

Der Verarbeitungsprozess in **WANDO** ist dann sehr ähnlich zu Variante A.

Zusätzlich wird beim Import der PDFs Konto/ Person, Dokumenttyp und Datum aus dem Dateinamen ermittelt und diese Daten werden für die automatische Indizierung herangezogen.

### **C: Indexdaten als Indexdatei vorhanden**

Liegt zu den PDF-Dateien jeweils eine weitere Datei vor, die die Indexkriterien enthält, können diese Dokumente ebenfalls automatisch verarbeitet werden.

Dabei sind folgende Fälle möglich:

- Zu jeder PDF-Datei gibt es eine Indexdatei gleichen Namens.
- In der Indexdatei ist der Verweis, sprich der Dateipfad auf die PDF-Datei enthalten.

Die Indexdateien können beispielsweise als CSV-, XML- oder INI-Format vorliegen.

**WANDO** ordnet die Indexdaten automatisch den Bildern zu.

Die Verarbeitung erfolgt somit ähnlich zu in Variante B.

## **WANDO WebUpload**

Der K7 WANDO WebUpload steht z.B. für den Fall zur Verfügung, dass der Kunde eine PDF-Datei als E-Mail-Anhang an seinen Sparkassenberater schickt.

Der Ablauf ist dann wie folgt:

- Der Berater speichert die Mail auf seinem Arbeitsplatz und öffnet den Browser.

- Anschließend wird die PDF-Datei per **WANDO WebUpload** an **WANDO** gesandt.
- Schon beim Hochladen der Datei hat der Berater bereits die volle Unterstützung der Bestandszugriffe wie in der Korrektur und kann das Dokument bereits komplett mit Dokumenttyp, Datum und Konto-/Kundendaten versehen und somit indizieren.
- Anschließend erfolgt dann im **WANDO**-System die vollautomatisierte Verarbeitung wie bei den oben erwähnten Varianten A und B.

### **Vorteile des WANDO WebUploads**

- Es können Dateien bis zu einer Größe von 30 MegaByte verarbeitet werden, da die Dokumente über den normalen Archivweg wie gescannte Dokumente archiviert werden (der Web-Upload der Finanz Informatik ist auf max. 5 MB große Dateien beschränkt).
- PDF Dateien, die ursprünglich nicht in PDF/A-1b vorliegen, werden entsprechend umgewandelt und formatkonform im ZDA angeliefert.

### **Per Mail direkt ins ZDA**

Eine weitere sehr effiziente Variante der PDF-Archivierung stellt **WANDO** für den Mailempfang bereit. Nach Vorlage gestaltete E-Mails, die über den Mailserver ankommen, enthalten im Mailanhang eines oder mehrere zu archivierende Dokumente im PDF-Format und zusätzlich im Mailtext die notwendigen Indexdaten.

Diese prüft **WANDO** auf Vollständigkeit und Plausibilität und schickt die Mailanhänge im Idealfall direkt PDF/A-konform ins Archiv. Alternativ kann die Freigabe auch per **WANDO**-Korrektur über einen Sparkassenmitarbeiter erfolgen.

Dedizierte E-Mailadressen wie z.B. [aktiv@sparkasse-musterstadt.de](mailto:aktiv@sparkasse-musterstadt.de) ermöglichen zusätzlich eine gezielte Steuerung.

In **WANDO** werden wichtige Parameter voreingestellt (z.B. Übertragung S/W oder Farbe) und der Workflow kontrolliert.

Der Mailtext lässt sich an das zu archivierende PDF anhängen und auch kennwortgeschützte PDFs werden verarbeitet.