

## PDF/A für gescannte Dokumente

### Papier wird digital

**Carsten Heiermann, LuraTech, Geschäftsführer**  
**Armin Ortmann, LuraTech, Entwicklungsleiter**



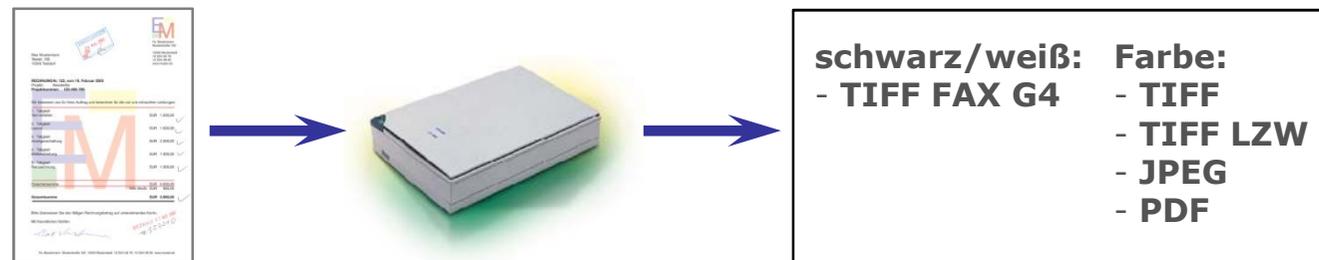
**LuraTech**

Carsten Heiermann  
Geschäftsführer



## Bisherige Lösungen für gescannte Dokumente

- ▶ **schwarz/weiß: TIFF G4**
- ▶ **Farbe: JPEG. Selten verwendet auch PNG, BMP und andere Rastergrafik-Formate**
- ▶ **Spezialversion wie „JPEG in TIFF“**
- ▶ **Nachteile:**
  - ▶ **Schon für gescannte Dokumente mehrere Formate**
  - ▶ **Unzählige Formate für „Character encoded“ (Office)**
  - ▶ **Informationsverlust, z.B. durch TIFF G4**
  - ▶ **Schlechte Bildqualität und große Dateien, z.B. JPEG**
  - ▶ **Keine einheitlichen Metadaten für alle Formate**
  - ▶ **Keine Durchsuchbarkeit (OCR) gescannter Dateien**



## Bisherige Lösungen für gescannte Dokumente

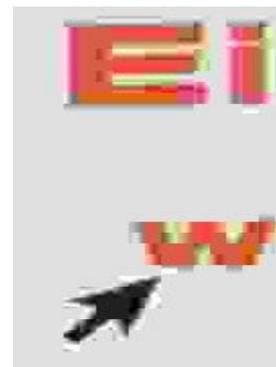
### ❖ Schlechte Bildqualität vs. Dateigröße

**TIFF/BMP**



**23,8 MB**

**JPEG**



**180 kB**

**TIFF G4**



**60 kB**



**LuraTech**  
Carsten Heiermann  
Geschäftsführer

## Alternative Lösung: PDF

- ❖ **Schon sehr weit verbreitet:**
  - ❖ Entweder zur Harmonisierung von Archivformaten
    - ❖ Image2PDF
    - ❖ Aber auch "Office" und andere Formate
  - ❖ Oder um Volltext/Durchsuchbarkeit zu ermöglichen
  - ❖ Oder um bessere Kompression nutzen zu können (z.B. JPEG2000)
  - ❖ Oder um einheitliche Metadaten zu ermöglichen
  
- ❖ **Zusammengefasst: PDF hat nicht die Nachteile der alten/anderen Formate**
  
- ❖ **"Wenn die Entscheidung für PDF gefallen ist, warum dann nicht im sicheren PDF/A Format archivieren – mit allen PDF Vorteilen?"**



**LuraTech**

Carsten Heiermann  
Geschäftsführer



# PDF/A

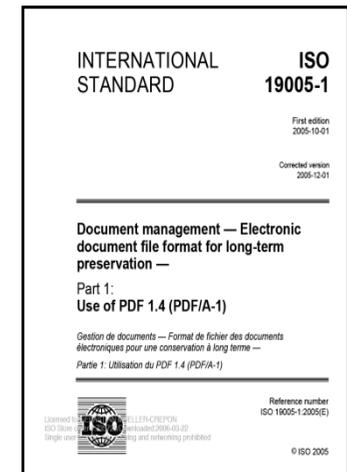
## Was ist PDF/A?

- **ISO 19005-1, Document Management**
- **Electronic document file format for long-term preservation**



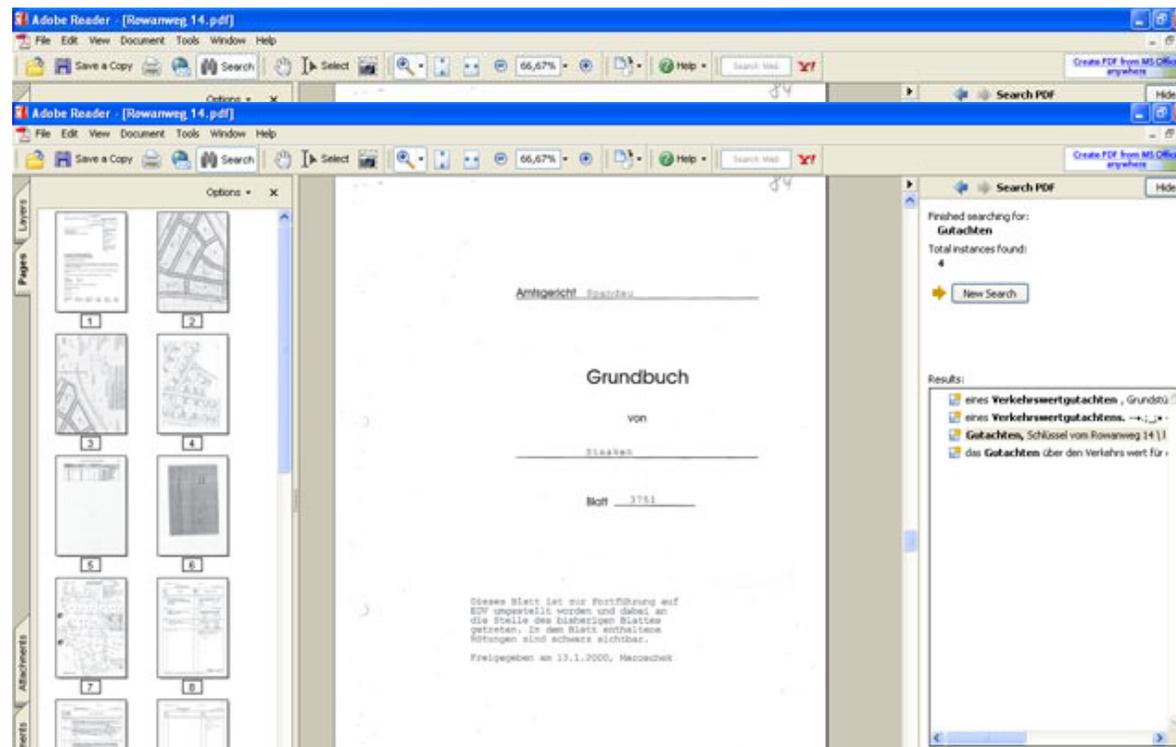
## Ziele von PDF/A

- **Erhalt der Darstellung statischer Inhalte**
- **Einheitliche Behandlung von Metadaten**
- **Optional auch Erhalt der Struktur und semantischen Bedeutung des Inhalts**
- **Transparenter Aufbau erlaubt den dauerhaften Zugang**
- **Beschränkung auf PDF Features, die nicht kritisch für die langfristige Archivierung sind**



## PDF/A – Durchsuchbarkeit im Volltext (OCR)

- ❖ **Vorteil: Durchsuchbarkeit auf Dateiebene**
  - ❖ z.B. digitale Bibliothek, „nach dem Download“
  - ❖ z.B. große oder mehrseitige Handbücher und Bauakten
  - ❖ z.B. Dokumente außerhalb des DMS, Versand an Kunden, Lieferanten, Partner,...



## PDF/A – Verbesserte Kompression

- ❖ Für schwarz/weiße Dokumente
  - ❖ JBIG2 - ISO/IEC 14492
    - ❖ Als Alternative zu TIFF G4
    - ❖ Vollständig oder visuell verlustfreie Kompression
    - ❖ Eingebettet in PDF/A, verfügbar im Acrobat Reader

FAX G4

60 kB

JBIG2/lossless

46 kB

JBIG2/lossy

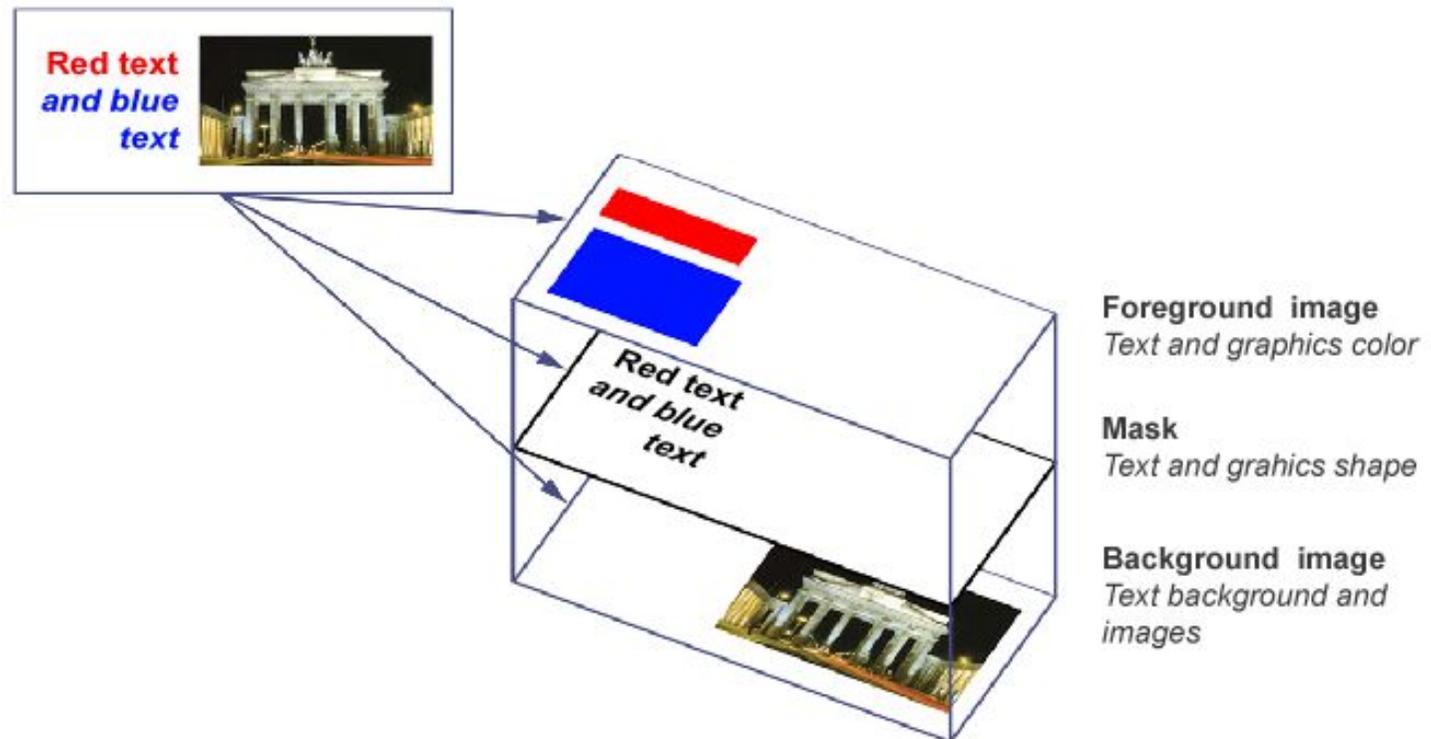
29 kB



LuraTech  
Carsten Heiermann  
Geschäftsführer

## PDF/A – Verbesserte Kompression

- ❖ Für farbige Dokumente
  - ❖ MRC-Kompression, bekannt aus JPEG2000 (JPM)
  - ❖ Segmentierung in drei Layer, die unabhängig optimal komprimiert und dann in PDF/A verpackt werden



## PDF/A – Verbesserte Kompression

- ❖ Für farbige Dokumente
  - ❖ Extreme Kompression, volle Lesbarkeit
  - ❖ Erhält Farbe und visuelle Qualität

TIFF



23,8 MB

TIFF G4



60 kB

JPEG



180 kB

PDF/A



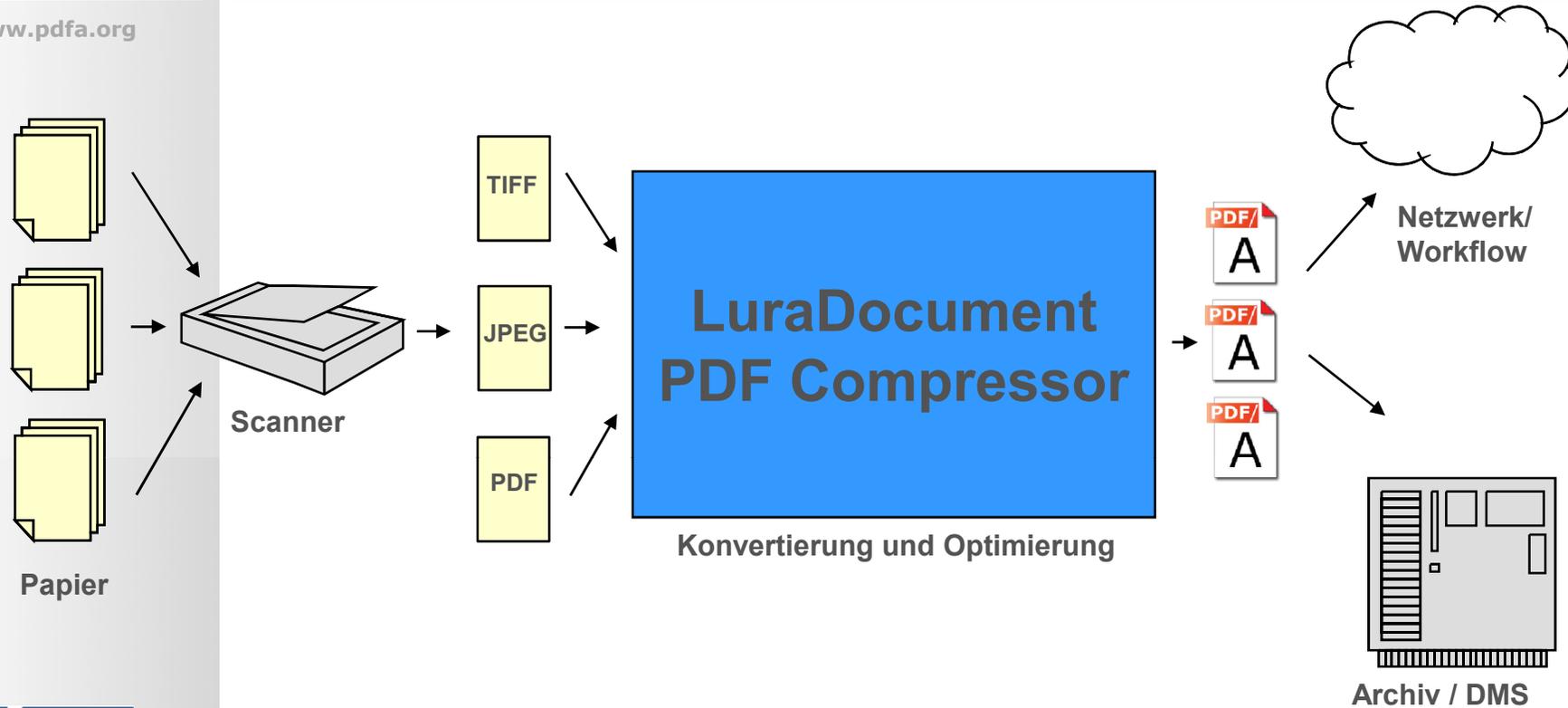
65 kB



LuraTech

Carsten Heiermann  
Geschäftsführer

# Konvertierung in PDF/A, gescannte Dokumente



## ➤ Konvertierung gescannter Dokumente

- Stapelverarbeitung möglich
- „Dunkelverarbeitung“ ohne Benutzerinteraktion
- Volle Automatisierbarkeit



## Demo

---

❖ **Armin Ortmann, praktische Vorführung!**



**LuraTech**  
Carsten Heiermann  
Geschäftsführer



## Frage:

---

- ❖ **PDF/A: Hype oder Archivformat der Zukunft?**



## PDF/A – Beispiel Kreditakten

- ❖ **Posteingang für Kreditakten und Schecks**
- ❖ **Beispiel: Landesbank HeLaBa**
  - ❖ **168 Mrd. Euro Bilanzsumme**
  - ❖ **5.700 Mitarbeiter**
- ❖ **Projekt**
  - ❖ **20 Mio. Seiten Papierarchiv in PDF/A konvertieren**
  - ❖ **Konvertierung aller Tagespost in PDF/A**
  - ❖ **Erzeugung einer komplett elektronischen Kreditakte**
  - ❖ **Werkzeuge: LuraTech PDF Compressor, Kofax Ascent, EMC Centera, Wincor Nixdorf archive:net**
- ❖ **Ergebnisse**
  - ❖ **Vollfarbige Belege im Archiv verfügbar**
  - ❖ **Hoch komprimierte PDF/A-Dateien**
  - ❖ **Kreditakten sind im Volltext durchsuchbar**
  - ❖ **Langfristige Lesbarkeit der Kreditakten gesichert**
  - ❖ **Erster Schritt zum einheitlichen Archivformat**



## PDF/A – Beispiel eGovernment

- ❖ **Meldeakten und Bauakten**
- ❖ **Beispiel: Langzeitarchivierung Stadt Erlangen**
  - ❖ 103.000 Einwohner, über 70% mit Internetzugang
  - ❖ eGovernment-Center Initiative
- ❖ **Projekt**
  - ❖ Erstens: Meldeakten in PDF/A konvertieren
  - ❖ Zweitens: Bauakten in PDF/A konvertieren
  - ❖ Drittens (Plan): PDF/A als einheitliches Aktenformat
- ❖ **Ergebnisse**
  - ❖ PDF/A ist einsetzbar für kleinere Massenbelege und auch für große Dokumente wie Baupläne (mit einigen 100 MB Rohdaten)
  - ❖ Hoch komprimierte PDF/A-Dateien sparen Speicherkosten und Bandbreite im Netzwerk
  - ❖ Langfristige Lesbarkeit garantiert
  - ❖ Akten sind für schnelle tägliche Recherche verfügbar



## Mehrere Milliarden Seiten, die zwischen 10 und mehr als 100 Jahren aufzubewahren sind

- ❖ **APO-Bank (D)**
- ❖ **West LB (D)**
- ❖ **Bank Julius Baer (CH)**
- ❖ **Diverse Sparkassen (D)**
- ❖ **Swiss RE (CH)**
- ❖ **DAK Versicherung (D)**
- ❖ **Universa Versicherungen (D)**
- ❖ **AOK (D)**
- ❖ **OCE (NL/D)**
- ❖ **Southern Nuclear (USA)**
- ❖ **RWE Energie (D)**
- ❖ **Siemens (D)**
- ❖ **Vattenfall (D)**
- ❖ **Airbus (D)**
- ❖ **Blohm & Voss (D)**
- ❖ **Bosch Rexroth (D)**
- ❖ **International Labor Organization (CH)**
- ❖ **British Library (UK)**
- ❖ **Harvard Library (USA)**
- ❖ **Het Utrechts Archief (NL)**
- ❖ **Internet Archive (USA)**
- ❖ **Das sind nur LuraTech Referenzkunden...**



## PDF/A für gescannte Dokumente

---

**Vielen Dank für Ihre Aufmerksamkeit!**  
**Bitte füllen Sie unseren Fragebogen aus.**

**Demo-Software oder weitere Infos?**

**[c.heiermann@luratech.com](mailto:c.heiermann@luratech.com)**



**LuraTech**  
Carsten Heiermann  
Geschäftsführer

