PDF/A Competence Center

www.pdfa.org

DMS EXPO in Köln - 2007

Der PDF/A-Standard:

Einführung, Einsatzgebiete und Wege zu PDF/A

Hans Klimmer, Produkt-Manager
DETEC Decision Technology Software GmbH, Rüsselsheim

DETEC

Hans Klimmer
Produkt-Manager

© 2007 by PDF/A Competence Center, www.pdfa.org

Definition Standard / Norm

- **Ein Standard ist eine vergleichsweise einheitliche, weithin anerkannte und meist auch angewandte Art und Weise, etwas herzustellen oder durchzuführen, die sich gegenüber anderen Arten und Weisen durchgesetzt hat.**
- **Etwas Geläufiges, Etabliertes, weithin Verbreitetes.**
- **Eine verbindliche Richtlinie, die sicherstellen soll, dass Produkte verschiedener Hersteller miteinander kompatibel bzw. interoperabel sind.**
- **Eine Norm ist ein Dokument, das durch eine breite Beteiligung aller interessierten Kreise im Konsens erarbeitet wird.**

PETEC

Hans Klimmer
Produkt-Manager

Ziele der Normung

- Förderung des Austausches von Waren, Dienstleistungen und Dokumenten
- technische Handelshemmnisse verhindern
- Rationalisierung
- Kompatibilität
- Austauschbarkeit
- Gebrauchstauglichkeit
- Sicherheit
- Qualitätsverbesserung und –sicherung

PETEC
Hans Klimmer
Produkt-Manager

Wer normiert die Welt?

International:

ISO
International Organization for Standardization

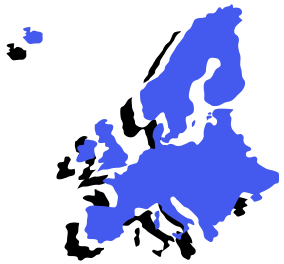
IEC
International Electrotechnical Commission

ITU
International Telecommunication Union

Europa:

CEN
Europäische Komitee für Normung

CENELEC
Europäisches Komitee für elektrotechnische Normung



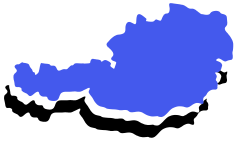
ETSI
Europäische Institut für Telekommunikationsnormen

Deutschland:

DIN
Deutsches Institut für Normung e. V.

Österreich:

ON
Österreichisches Normungsinstitut



Schweiz:

SNV
Schweizerische Normen-Vereinigung

PETEC
Hans Klimmer
Produkt-Manager

Formate im digitalen Archiv – bisher

„Das Gute an den Standards ist, daß es so viele gibt, aus denen man wählen kann."

- **TIFF** (i.d.R. G3 oder G4)
- **PDF** (in allen erdenklichen Ausprägungen)
- **Druckdaten** (z.B. AFP, PCL)
- **Native Applikationsformate**
  (Hersteller-Standard, z.B. DOC, XLS, DWG, DXF)

PETEC
Hans Klimmer
Produkt-Manager

PDF – Standard für digitale Dokumente?

- PDF ist ein universell einsetzbares offenes Format zur Darstellung elektronischer Dokumente
- 10% aller Web-Inhalte liegen im PDF-Format vor
- Als Format für den elektronischen Dokumentenversand de facto Standard
- Eine PDF-Datei gibt die Dokumente des Ursprungsprogramms einschließlich aller Farben, Raster- und Vektorgrafiken präzise wieder (?)
- PDF ist proprietär – aber mit veröffentlichter und frei zugänglicher Spezifikation
- PDF ist äußerst mächtig und reich an Features, die mit jeder Version erweitert werden

PETEC

Hans Klimmer
Produkt-Manager

PDF – Problematisch für die Archivierung?

PDF kann aufgrund seiner Mächtigkeit Strukturen enthalten, die für die Langzeitarchivierung nicht geeignet sind:

- **WYSIWYG nicht unbedingt garantiert: Fonts müssen nicht zwangsläufig eingebettet sein und sind u.U. auf dem darstellenden Gerät nicht vorhanden.**
- **Referenzen in Dokumenten können über einen längeren Zeitraum ihre Gültigkeit verlieren.**
- **Interaktive Dokumentenbestandteile bzw. die Möglichkeit, Daten in einem Dokument zu erfassen, verändern den ursprünglichen Inhalt des Dokuments.**
- **Passwortgeschützte Dokumente sind nicht mehr zu reproduzieren, falls das Passwort verloren geht.**
- **und, und, und …**

PETEC

Hans Klimmer
Produkt-Manager

PDF/A – Anforderungen für die Langzeitarchivierung

- **Geräteunabhängig**
  - Darf keine bestimmten Geräte (Drucker oder Bildschirme) voraussetzen, um in seiner ursprünglichen Form reproduziert werden zu können.
- **Dokument enthält alle benötigten Bestandteile zur Reproduktion**
  - Alle zur Darstellung benötigten Objekte sind eindeutig im Dokument enthalten.
- **Selbst dokumentierend**
  - Zusätzliche Metadaten über Herkunft und Inhalt eines Dokuments sind in einer definierten Struktur anzugeben.

PETEC
Hans Klimmer
Produkt-Manager

PDF/A – Anforderungen für die Langzeitarchivierung

- **Zugänglichkeit**

  Das Dokument selbst darf nicht verschlüsselt sein oder bestimmte Funktionen unterbinden (z.B. Drucken)

- **Öffentlich verfügbar**

  Es muss sich um ein Dateiformat handeln, das öffentlich verfügbar und frei von Patentrechten oder Lizenzabgaben ist. Die Implementierung PDF/A-kompatibler Produkte muss für jedermann ohne Einschränkung möglich sein.

- **Breite Akzeptanz**

  Das zugrunde liegende Format muss von einer breiten Mehrheit akzeptiert sein. Nur so ist eine umfangreiche und nachhaltige Unterstützung des Formats durch Softwarehersteller gewährleistet.

PETEC

Hans Klimmer
Produkt-Manager

PDF/A – Der Standard

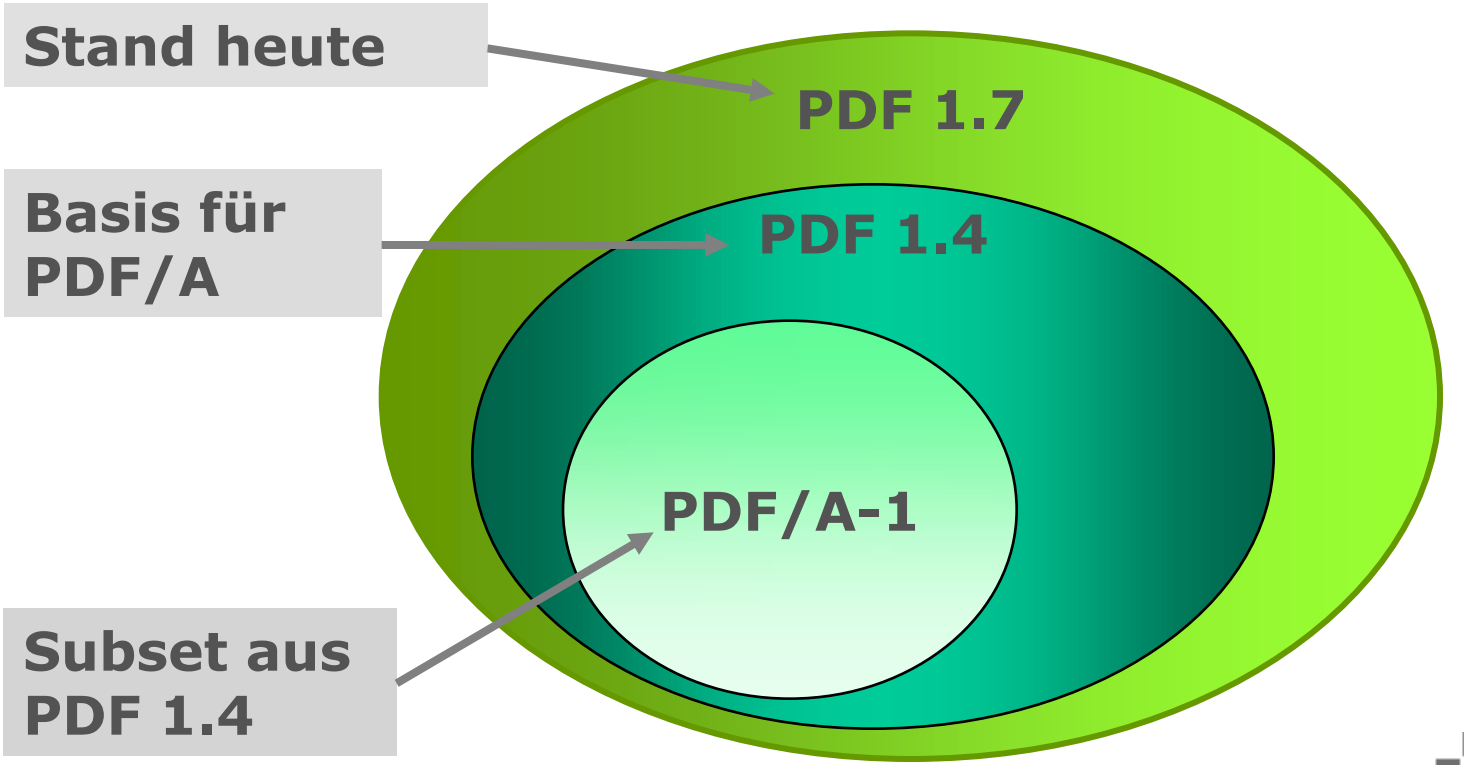
- **ISO 19005-1: Nicht nur Standard, sondern Norm!**
- **Initiiert und betrieben von NPES (National Printing Equipment Association), AIIM (Association for Information and Image Management) und Administrative Office of the U.S. Courts**
- **Viele Hersteller und Anwender sind vertreten: Xerox, Harvard University, Library of Congress, NARA (National Archives and Records Administration), EDS, EMC, Adobe, Merck, Glaxo Smith Kline,...**
- **PDF/A konzentriert sich auf die Bedürfnisse bezüglich der Archivierung in Unternehmen, Bibliotheken und der öffentlichen Hand**

PETEC

Hans Klimmer
Produkt-Manager

PDF/A – ISO 19005-1 Überblick

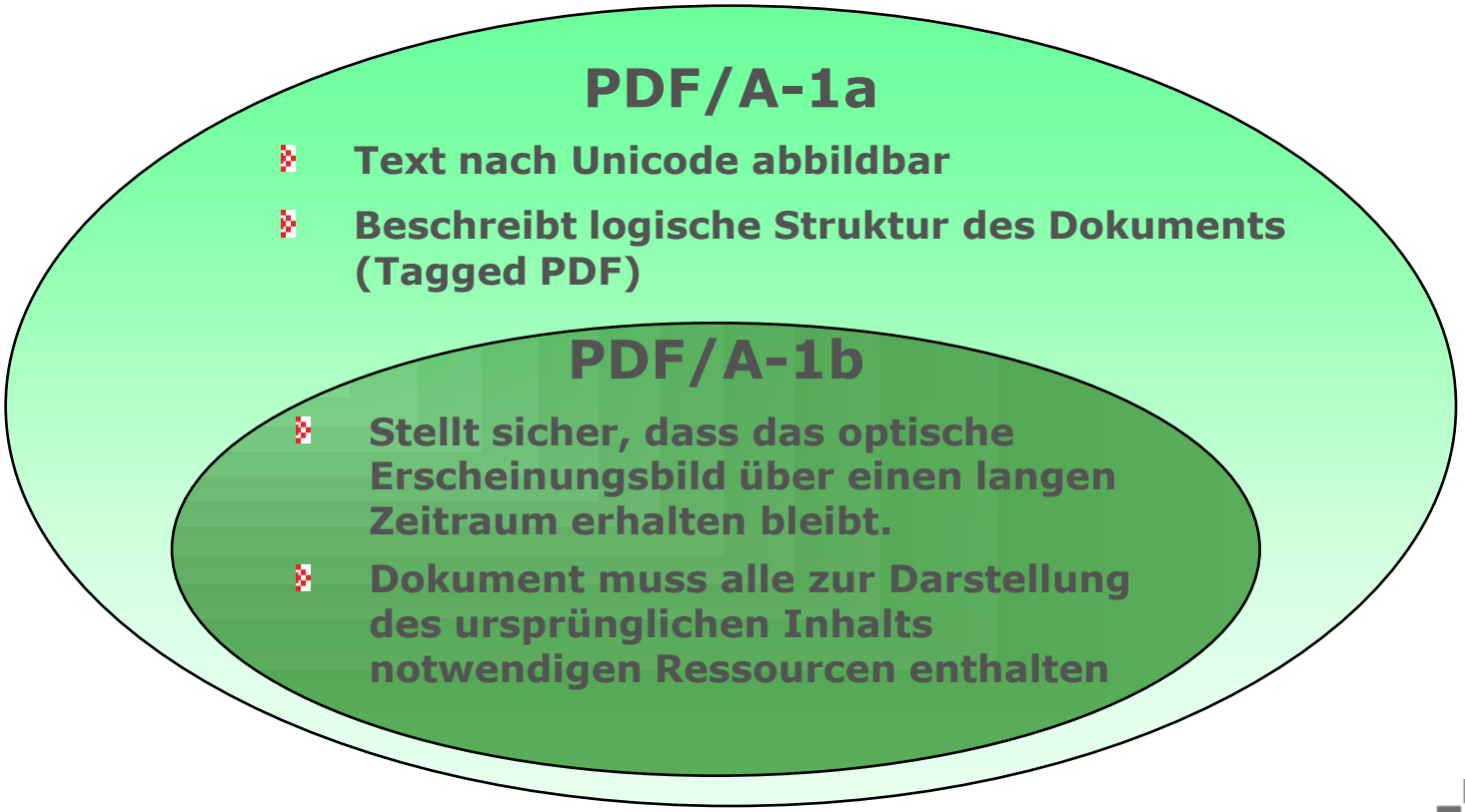
- Die ISO 19005-1 Norm wurde im September 2005 verabschiedet.
- Basiert auf der PDF 1.4 Referenz und definiert ein Subset der darin enthaltenen Funktionalität.



PETEC
Hans Klimmer
Produkt-Manager

PDF/A – ISO 19005-1 Überblick

- Definiert 2 Levels für PDF/A Konformität:
  - PDF/A-1a: konform in vollem Umfang mit dem PDF/A Standard.
  - PDF/A-1b: konform mit den Mindestanforderungen des PDF/A Standard.



PETEC
Hans Klimmer
Produkt-Manager

PDF/A – ISO 19005-1 Überblick

Einige wesentliche Eigenschaften PDF/A konformer Dokumente:

- Alle verwendeten Schriften einbetten.
- Geräteunabhängigen Farbraum verwenden.
- Metadaten im XMP-Format

- Passwort!
- Patentrechtlich geschützte Verfahren!
- Dateien einbetten!
- Referenzen auf externe Objekte!
- Multimedia-Inhalte!

PETEC
Hans Klimmer
Produkt-Manager

PDF/A – Einsatzgebiete

Verwaltung und Behörden

- Als Ablageformat für alle online zur Verfügung stehende Formulare (eGovernment). Dadurch identisches Erscheinungsbild - wo auch immer die Formulare angezeigt bzw. gedruckt werden.
- Als Format für alle elektronischen Akten
- Anwendungsfall Bundesarchiv:
  - Plant, zukünftig alle Dokumente in elektronischen Akten nach PDF/A zu konvertieren.
  - Ziel ist, dass die angeschlossenen Ministerien bereits PDF/A konforme Dokumente liefern.
  - Argument für PDF/A ist der offene und jederzeit nachvollziehbare Standard.

PETEC

Hans Klimmer
Produkt-Manager

PDF/A – Einsatzgebiete

Bildungseinrichtungen

- 90% der bei der Deutschen Bibliothek archivierten Online-Dissertationen liegen in PDF Format vor.
- PDF/A als Langzeitarchivformat für wissenschaftliche Arbeiten und Dissertationen
- Anwendungsfall Universität Potsdam:
  - Bachelor- und Masterarbeiten*) werden auf dem Publikationsserver der Universität einheitlich in PDF/A zur Verfügung gestellt.

*) nur mit „sehr gut" bewertete Arbeiten

Hans Klimmer
Produkt-Manager

PDF/A – Einsatzgebiete

Verlags- und Medienbereich, Bibliotheken

- PDF/A als Format für alle über das Web-Portal (Intra- oder Internet) verfügbaren Publikationen.
- Vorteil: Darstellung des ursprünglichen Inhalts auch beim Konsumenten der Online-Inhalte gewährleistet.

PETEC
Hans Klimmer
Produkt-Manager

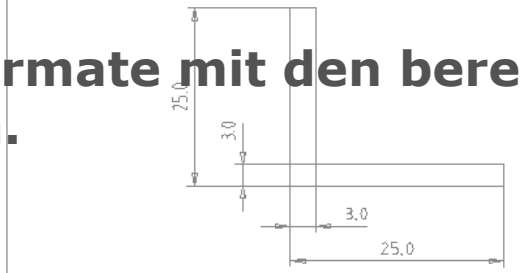
PDF/A – Einsatzgebiete

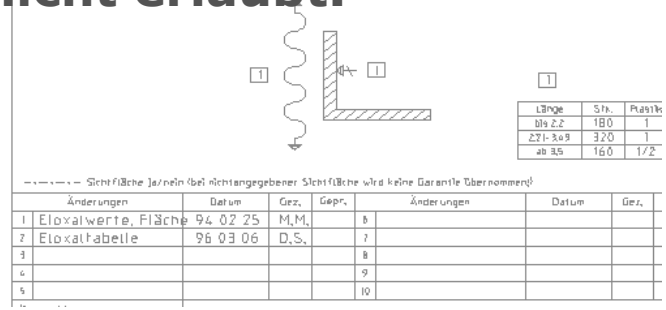
Engineering

- PDF/A als Datenformat für das Zeichnungsarchiv und die technische Dokumentation.
- Keine proprietären Formate mit den bereits genannten Nachteilen.

Aber:

- Vor allem die im CAD-Bereich wichtige Funktion „Layer" ist in PDF/A nicht erlaubt!





Hans Klimmer
Produkt-Manager

PDF/A – Einsatzgebiete

eBilling

- Sonderfall des elektronischen Dokumentenversands: Verfahrensweise durch das SigG gesetzlich geregelt.
- (Qualifizierte) digitale Signatur ist in PDF/A zulässig.
- Vorteile für den Sender und Empfänger:
  - PDF/A Kompatibilität bereits auf Seite des Rechnungsversenders um so wichtiger, da digital signierte PDF-Dokumente nicht nachträglich nach PDF/A konvertiert werden können.
  - Kostenreduktion durch eBilling für Sender und Empfänger durch Wegfall manueller Prozesse (Rechnungsproduktion, Versandkosten, Posteingangsverarbeitung usw.)

PETEC

Hans Klimmer
Produkt-Manager

PDF/A – Einsatzgebiete

Posteingang

- Papierdokumente werden eingescannt – und bisher häufig als TIFF archiviert.
- Eine Konvertierung nach PDF/A in einem erweiterten Workflow bringt zusätzliche Vorteile:
  - OCR – automatische Texterkennung ermöglicht die Suche nach Texten in den Dokumenten.
  - Kompression – die Imagedaten werden mit neueren Kompressionsalgorithmen wie z.B. JBIG2 wesentlich kleiner.
  - Sicherheit – Ausgabe als PDF/A konforme Dokumente
  - Nur ein Viewer für Archivdokumente notwendig

PETEC
Hans Klimmer
Produkt-Manager

PDF/A – Weitere Entwicklung

Roadmap PDF/A

- Ergänzung mit Fehlerkorrekturen zu PDF/A-1.
- Aktuell wird an der Spezifikation zu ISO 19005-1, Part 2 (PDF/A-2) gearbeitet. Freigabe frühestens 2009.
- PDF/A-2 basiert auf PDF Referenz 1.7 (Acrobat 8)
- PDF/A-2 nimmt neuere Elemente aus PDF 1.5 und 1.6 mit auf wie z.B. JPEG2000 Kompression für Farbimages.
- PDF/A-3 soll nicht mehr nur die visuelle Darstellung (statisch) bewahren, sondern zusätzlich auch das Verhalten (dynamisch).
- Adobe hat Ende Januar die komplette PDF 1.7 Referenz gegenüber dem AIIM zur Standardisierung durch die ISO offengelegt.

PETEC
Hans Klimmer
Produkt-Manager

Wege zu PDF/A

DETEC GmbH ist seit über 20 Jahren im Bereich Dokumentendesign und Produktion, Datenkonvertierung und OMS tätig und bietet folgende Produkte zum Thema PDF/A:

- **myPDFconvert:**
  - **Software zur Konvertierung vorhandener Dokumente in den Formaten PDF, TIFF, JPEG, PNG und BMP nach PDF/A-1b**
  - **optional mit Kompression und OCR**
  - **Verfügbar als GUI, Commandline-Version und API**
- **DoXite:**
  - **Softwaresystem zur Konvertierung von Rohdaten, Steuerung von Workflows und Dokumentendesign.**
  - **Produziert Dokumente aus unterschiedlichen Eingabeformaten diverser ERP- und CRM-Systeme.**
  - **Erzeugt sowohl Druckoutput als auch digitale Dokumente für Fax, Mail und Archiv.**
  - **Steuert die benötigten Workflows**
  - **PDF-Output natürlich ebenfalls PDF/A konform.**

DETEC

Hans Klimmer
Produkt-Manager

Ihre Fragen?

Vielen Dank für Ihre Aufmerksamkeit!

Hans Klimmer, hklimmer@detec.de

DETEC

Hans Klimmer
Produkt-Manager