

# PDF/A und Metadaten

## XML, RDF und XMP: Was, wie und woher?



# Agenda

---

- ❖ **Metadaten**
- ❖ **RDF und XMP**
- ❖ **XMP, PDF und PDF/A**
- ❖ **XMP Schema**
- ❖ **XMP Erweiterungen**

# Was sind Metadaten?

- ❖ **Metadaten sind Informationen über Daten**
- ❖ **Von Einfach:**
  - ❖ Wer hat die Daten erstellt?
  - ❖ Wovon handeln die Daten?
  - ❖ Wann wurden die Daten geändert?
- ❖ **Über spezifisch:**
  - ❖ Welche ID haben diese Daten?
  - ❖ Welches Gerät hat die Daten produziert?
- ❖ **Bis hin zu komplex:**
  - ❖ Über was handeln die Daten?
  - ❖ Welchen Prozessstatus haben die Daten?

## Struktur von Metadaten?

- ❖ **Oft Fehlanzeige**
- ❖ **z.B. Word: Dokumenteigenschaften**
- ❖ **z.B. PDF: Dokumentinfo**
- ❖ **z.B. Bilder: EXIF, TIFF Tags**
- ❖ **...etc...**



## Eine Lösung: RDF

- ❖ **RDF bietet eine strukturierte Abbildung**
  - ❖ **Resource Description Framework**
  - ❖ **Initiative des W3C**
  
- ❖ **Ressourcen werden abgebildet als**
  - ❖ **Objekt -> Prädikat -> Subjekt**
  - ❖ **z.B. Dokument -> erstellt -> Author**
  - ❖ **z.B. Prozessnummer -> verschoben -> Archiv**

# XMP (Extensible Metadata Platform)

- ❖ **XMP von Adobe**
- ❖ **Basiert auf RDF**
- ❖ **Beschreibt primitive Datentypen (z.B. Boolean, Integer, Datum etc...)**
- ❖ **Beschreibt die Ablage in verschiedenen Formaten**
  - ❖ **PDF**
  - ❖ **JPEG / TIFF / PNG / GIF**
  - ❖ **PS**
- ❖ **Vordefinierte Schemas für Metadaten**
- ❖ **Erweiterbar durch eigene Schemas**

# XMP Schema

- ❖ **Nicht identisch mit XML Schema (xsd)**
- ❖ **Fasst Properties zusammen**
  - ❖ z.B. Autor, Thema, Titel in DC Schema
  - ❖ z.B. Besitzer, Zertifikat, Rechte in Rights Management Schema
- ❖ **Eindeutige Namespace-URI pro Schema**
  - ❖ z.B. DC Schema = <http://purl.org/dc/elements/1.1/>
  - ❖ Präfix: dc

# XMP im PDF

- ❖ **Seit PDF 1.4 integriert**
  
- ❖ **Kann an vielen Stellen im PDF hängen**
  - ❖ **Dokumentebene**
  - ❖ **Einzelne Bilder**
  - ❖ **Fonts**
  - ❖ **Farbräume**
  - ❖ **etc...**
  
- ❖ **Aber: DokumentInfo <> Dokument-XMP**



## PDF/A und XMP

- ❖ **Verpflichtender Bestandteil: PDF/A Konformität**
- ❖ **Basiert auf XMP 2004**
- ❖ **Erlaubt alle vordefinierten Schemas aus XMP 2004**
- ❖ **Erfordert gespiegelte Einträge zum Dokumentinfo**
- ❖ **Sehr strenge Regeln für XMP Schema**

## Vordefinierte Schemas in PDF/A

---

- ❖ **Dublin Core: Standard-Eigenschaften**
- ❖ **XMP Basic: Beschreibende Eigenschaften**
- ❖ **XMP Rights Management**
- ❖ **XMP Media Management: z.B. Scanning, History**
- ❖ **Adobe PDF**



## Der komplexe Teil: Erweiterungen

- ❖ **Zweck von XMP Erweiterungen**
  - ❖ **Abbildung spezifischer Anforderungen**
  - ❖ **z.B. Patientenakten**
  - ❖ **z.B. Workflow Informationen**
  - ❖ **z.B. bibliothekarische Informationen**
- ❖ **XMP Erweiterungen sind Schemas**
- ❖ **Müssen in PDF/A beschrieben werden**
- ❖ **„Extension Schema Container Schema“ !**



# XMP Erweiterungen

- ❖ **Extension Schema Container Schema**
- ❖ **Definiert neue Eigenschaften**
  - ❖ **Name und Beschreibung**
  - ❖ **Datentyp**
- ❖ **Kann neue, komplexe Datentypen definieren**



## Warum XMP Erweiterungen

---

- ❖ **Eigene Metadaten können im starren PDF/A abgebildet werden**
- ❖ **Geeignete Software kann Metadaten interpretieren**
- ❖ **Plausibilitätstest durch Datentypen**
- ❖ **Strukturierte Durchsuchbarkeit**
- ❖ **Indexierung für Archive**



## XMP Erweiterungen: Wie?

- ❖ **Formalisieren**
  - ❖ **Eigenschaften**
  - ❖ **Datentypen**
- ❖ **Gibt es schon vordefinierte Eigenschaften / Datentypen?**
  - ❖ **PDF/A CC TechNote 0008**
- ❖ **Definition des Erweiterungsschema**
  - ❖ **PDF/A CC TechNote 0009**
- ❖ **Einbringen des XMP in die PDF Dokumente**
  - ❖ **Bei Erstellung**
  - ❖ **Nachbearbeitung**
  - ❖ **...**



## Vielen Dank für Ihre Aufmerksamkeit

### Gerne beantworte ich Ihre Fragen



Karsten Krieg  
Telefon: +49 721 38479 – 28  
E-Mail: [KKrieg@intarsys.de](mailto:KKrieg@intarsys.de)

intarsys consulting GmbH  
Bahnhofplatz 8, D-76137 Karlsruhe

Geschäftsführer:  
Dr. Bernd Wild, Karl Kagermeier, Michael Traut  
Sitz: Karlsruhe, Amtsgericht Mannheim - HRB 107535  
Internet: [www.intarsys.de](http://www.intarsys.de)