



Extensible Metadata Platform: mehr als nur Metadaten

Jeremias Märki <info@jeremias-maerki.ch>



Über mich

- Selbstständiger Software-Entwickler & Berater aus Luzern (Schweiz)
- Hauptfokus auf Dokumentenproduktion und -handling
- Starker Fokus auf Open Source Software und offene Standards
- Mitglied der Apache Software Foundation
- Mitarbeit an Apache FOP (XSL-FO Formatierer) und anderen Open Source Projekten

Website & Blog: <http://www.jeremias-maerki.ch/>

Inhalt

- Das Prinzip von Metadaten
- Wieso Metadaten?
- Was will XMP?
- Verhältnis zu anderen Standards wie RDF, Dublin Core etc.
- Beispiele aus der Praxis

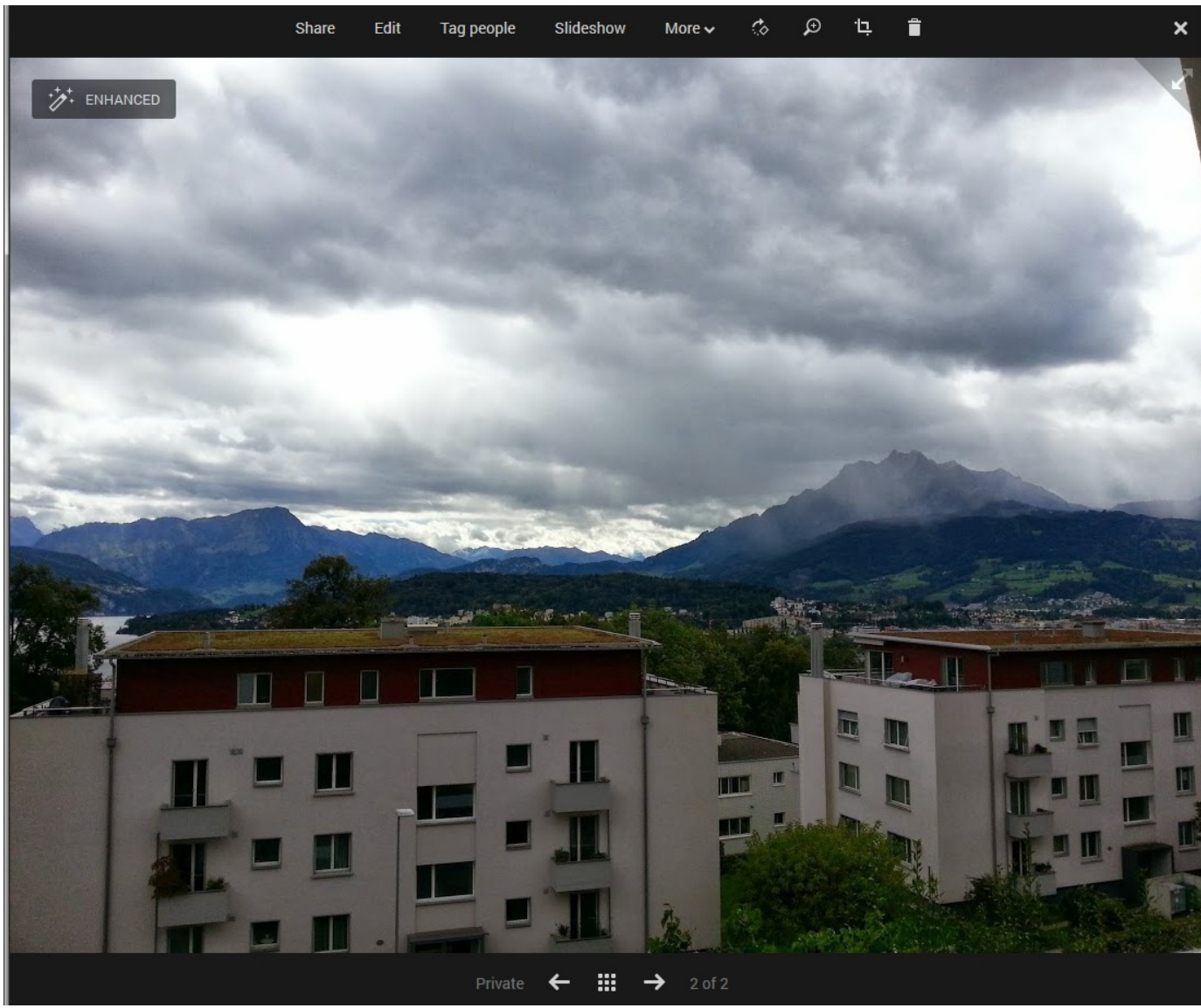
Was sind Metadaten?


Daten über Daten

Naja, das ist eine Vereinfachung...

Was für Daten?


- Mittel zur Datenerstellung
- Zweck der Daten
- Datum und Uhrzeit der Erstellung
- Erzeuger oder Autor der Daten
- Ort der Erstellung
- etc. etc.



 **Jeremias Märki**
14:21

Pilatus am 18.9.2013

Photo details

 [Edit](#) [Remove](#)

Share location data for this album with viewers

Location	47.06, 8.332
Date taken	18/09/2013 14:21
Dimensions	2048 x 1536
File name	20130918_142104.jpg
File size	285.07K
Camera	GT-N7100
Focal Length	3.7 mm
Exposure	1/2028
F Number	f/2.6
ISO	80
Camera make	SAMSUNG
Flash	Not used
Exposure Bias	-
Views	-

[+1](#) [Share](#)

Add a comment...

Google Picasa: Metadaten eines Fotos

Dateigröße 1.7 MB	Blitz Nicht verwendet	JPEG-Qualität 96 (422)	Belichtungszeit 11
Dimensionen 3264 x 2448 Pixel	Brennweite 3.7 mm	GPS-Breite 47 3\35" N	Helligkeit 9.21
Kameramarke SAMSUNG	Belichtungszeit 0.000493097 s (1/2028)	GPS-Länge 8 19\55" E	Belichtungsausrichtung 0
Kameramodell GT-N7100	Blende 2.76	GPS-Höhe 539	Maximale Blende 2.76
Kameradatum 2013:09:18 14:21:03	F-Zahl F/2.6	Eindeutige ID 2e134571c7972af800000000000000000	Nutzerkommentar {4096 Bytes}
Digitalisiertes Datum 2013:09:18 14:21:03	ISO 80	Auflösung X 72	Flashpix-Version {4 Bytes}
Änderungsdatum 2013:09:18 14:21:03	Weißabgleich Automatisch	Y-Auflösung 72	Belichtungsmodus Automatisch
Dateidatum 18.09.2013 14:21:03	Messmodus Mittenbetonte Belichtungsmessung	Auflösungseinheit Zoll	Szenenaufnahmetyp Standard
Ausrichtung Normal	Belichtungsprogramm Blendenpriorität	Software N7100XXDME6	GPS-Versions-ID 2.2.0.0
	Farbraum sRGB	YCbCr-Positionierung Zentriert	GPS-Höhenreferenz 0
		Komponentenkonfiguration {4 Bytes}	GPS-Zeitstempel 12, 20, 49
			GPS-Verarbeitungsmethode {8 Bytes}
			GPS-Datumsstempel 2013:09:18

Metadaten über was?

Foto Person
Rechnung Konferenz
Telefonanruf
Smartphone Auto
Buch Fahrschein

Inhalt vs. Metadaten: der Unterschied?



upc cablecom

upc cablecom GmbH
Industriestrasse 19
8112 Otelfingen

Rechnungsnummer
103052836

Kundennummer
5398756-1

Rechnungsdatum
04.09.2013

Zahlungstermin
26.09.2013

P.P. 8021 Zürich

5398756-1/103052836/04092013/B/M1A/0001

Herr Jeremias Märki
Lützelmatthstrasse 14
6006 Luzern

Rechnung September 2013

	Betrag	Gesamtbetrag
Saldo der letzten Rechnung	62.00	
Zahlungen berücksichtigt bis zum 3. September 2013	-62.00	
Ihr Guthaben		0.00
Abonnementspreise und Zusatzdienste	62.00	
Gesamtbetrag aktuelle Rechnungsperiode		62.00
Rechnungstotal gerundet (zahlbar bis 26.09.2013)		62.00

Zahlungseingänge sind bis zum 3. September 2013 berücksichtigt.

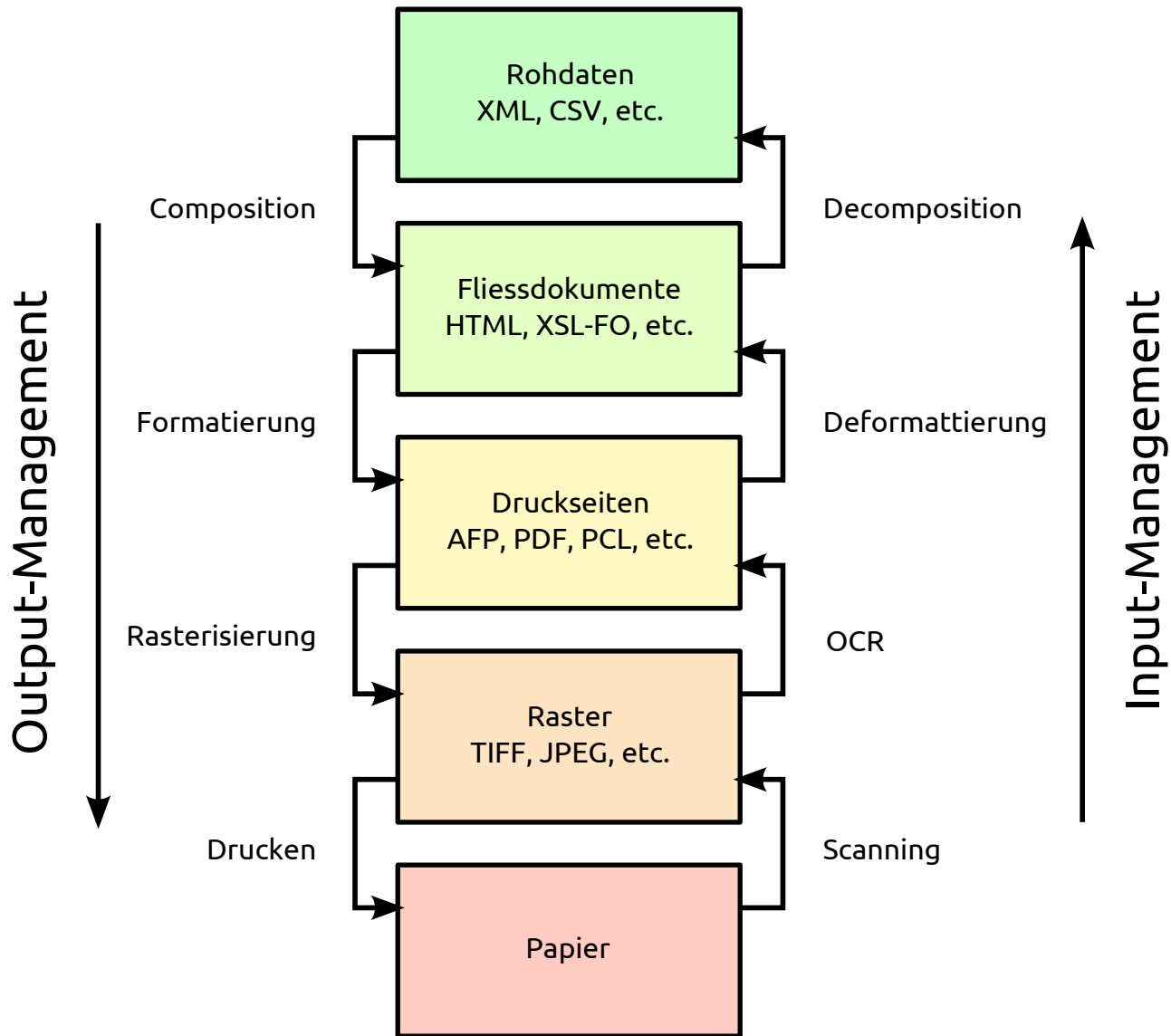
Wieso Metadaten?

Kategorisierung
Information
Zuordnung Zugriffskontrolle
Dinge finden

Aber das steht alles auf dem Papier, oder?

- Mehrheitlich
- Nur für Menschen zugänglich
- Der Computer kennt die **Semantik** der Daten nicht
- Der Computer hat vielleicht nicht einmal den vollen Text (Seiten als Rastergrafik)
- OCR nur um den Volltext zu bekommen

Es geht um Informationsverlust!



Also... was ist mit



Über XMP

- Ursprünglich entwickelt von Adobe Systems Inc.
- Neuerdings ein ISO Standard: ISO 16684-1:2012
- Standardisiertes Datenmodell
- Serialisierungsformat (eine Teilmenge von RDF/XML)
- Spezifiziert einen Satz von Kern-Eigenschaften
- Richtlinien für die Einbettung von XMP in verschiedene Dateiformate:
 - PDF, PostScript, JPEG, PNG, TIFF, HTML, XML uvm.

XMP und sein Verhältnis zu anderen Standards

- Eine Teilmenge von RDF/XML als Serialisierungsformat
- Wiederverwendung von Dritt-Ontologien
 - Dublin Core (ISO 15386:2009)
 - IPTC
 - Exif

RDF: Resource Description Format

- W3C Spezifikation
- Erstmals publiziert im 1999
- Ursprünglich als "Metadaten Datenmodell" gestaltet
- Mehrere Serialisierungsformate
 - Das wichtigste: RDF/XML
- Ressourcen-Identifikation über URIs
 - Beispiel: urn:isbn:0-486-27557-4
- Erlaubt Aussagen wie:
Der "Author" von "Romeo und Julia" ist "William Shakespeare".

Dublin Core

- IETF RFC 5013, ISO 15386:2009
- Definiert im 1995 an einem OCLC/NCSA Metadaten Workshop in Dublin, Ohio, USA
- Definiert Metadaten-Elemente wie:
 - Titel
 - Erzeuger
 - Thema
 - Beschreibung
 - Datum
 - Identifikator
 - Sprache
 - und mehr

XMP in der Praxis

- RDF ist relativ komplex, XMP vereinfacht das ein bisschen
- Es gibt Werkzeuge, aber die sind nicht immer gut integriert
- Man kann eigene Eigenschaften definieren. Beispiele:
 - Kundennummer
 - Policennummer
 - Ablaufdatum eines Dokuments
 - Fälligkeitsdatum einer Rechnung
 - Name und Version einer Dokument-Vorlage

Beispiel: PDF/A und XMP

- PDF/A hat einen starken Fokus auf Metadaten.
- PDFs müssen ein XMP Paket enthalten. Damit werden Metadaten gefördert.
- Empfiehlt einen eindeutigen Dokumentidentifikator zu benutzen.
- Empfiehlt die Herkunftsinformationen durch die Dokumentverarbeitung zu bewahren.
- Eigene Eigenschaften müssen im XMP Paket beschrieben sein (über ein eingebettetes Schema)
 - Leider scheint das viele von deren Benutzung abzuhalten.

XMP in PDF

```
<?xpacket begin="ï»¿" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description xmlns:dc="http://purl.org/dc/elements/1.1/" rdf:about="">
      <dc:language>de</dc:language>
      <dc:title>Extensible Metadata Platform: mehr als nur Metadaten</dc:title>
      <dc:creator>Jeremias Märki</dc:creator>
      <dc:type>Präsentation</dc:type>
      <dc:subject>
        <rdf:Bag>
          <rdf:li>XMP</rdf:li>
          <rdf:li>Metadaten</rdf:li>
        </rdf:Bag>
      </dc:subject>
      <dc:date>2016-11-05T15:50:15+01:00</dc:date>
    </rdf:Description>
    <rdf:Description xmlns:pdf="http://ns.adobe.com/pdf/1.3/" rdf:about="">
      <pdf:PDFVersion>1.4</pdf:PDFVersion>
      <pdf:Producer>Apache FOP Version 1.0</pdf:Producer>
    </rdf:Description>
    <rdf:Description xmlns:xmp="http://ns.adobe.com/xap/1.0/" rdf:about="">
      <xmp:CreateDate>2016-11-05T15:50:15+01:00</xmp:CreateDate>
      <xmp:MetadataDate>2016-11-05T15:50:15+01:00</xmp:MetadataDate>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end="r"?>
```

Dublin Core schema

The Dublin Core schema provides a set of commonly used properties.

- The schema namespace URI is <http://purl.org/dc/elements/1.1/>
- The preferred schema namespace prefix is `dc`

Property	Value type	Category	Description
<code>dc:contributor</code>	bag ProperName	External	Contributors to the resource (other than the authors).
<code>dc:coverage</code>	Text	External	The extent or scope of the resource.
<code>dc:creator</code>	seq ProperName	External	The authors of the resource (listed in order of precedence, if significant).
<code>dc:date</code>	seq Date	External	Date(s) that something interesting happened to the resource.
<code>dc:description</code>	Lang Alt	External	A textual description of the content of the resource. Multiple values may be present for different languages.
<code>dc:format</code>	MIMETYPE	Internal	The file format used when saving the resource. Tools and applications should set this property to the save format of the data. It may include appropriate qualifiers.
<code>dc:identifier</code>	Text	External	Unique identifier of the resource.
<code>dc:language</code>	bag Locale	Internal	An unordered array specifying the languages used in the resource.
<code>dc:publisher</code>	bag ProperName	External	Publishers.

XMP für PDF/A

The Identification schema defined in Table 6 uses the namespace URI `<http://www.aiim.org/pdfa/ns/id>`. The required schema namespace prefix is **pdfaid**.

Table 6 — PDF/A identification schema

Property	Value type	Category	Description
<code>pdfaid:part</code>	<i>Open Choice of Integer</i>	Internal	PDF/A version identifier
<code>pdfaid:amd</code>	<i>Open Choice of Text</i>	Internal	Optional PDF/A amendment identifier
<code>pdfaid:conformance</code>	<i>Closed Choice of Text</i>	Internal	PDF/A conformance level: A or B

```
<x:xmpmeta xmlns:x="adobe:ns:meta/">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    [..]
    <rdf:Description rdf:about=""
      xmlns:pdfaid="http://www.aiim.org/pdfa/ns/id/">
      <pdfaid:part>1</pdfaid:part>
      <pdfaid:conformance>B</pdfaid:conformance>
    </rdf:Description>
    [..]
  </rdf:RDF>
</x:xmpmeta>
```

XMP für ZugFeRD

- ZugFeRD, das elektronische Rechnungsformat für Deutschland

```
<x:xmpmeta xmlns:x="adobe:ns:meta/">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    [...]
    <rdf:Description rdf:about=""
      xmlns:zf="http://www.awv-net.de/ns/zugferd/1/">
      <zf:DocumentID>471102</zf:DocumentID>
      <zf:DocumentDate>2013-03-05</zf:DocumentDate>
      <zf:DocumentType>INVOICE</zf:DocumentType>
      <zf:DocumentFileName>ZUGFeRD_Example_1.xml</zf:DocumentFileName>
      <zf:Version>0.11</zf:Version>
      <zf:ConformanceLevel>BASIC</zf:ConformanceLevel>
    </rdf:Description>
    [...]
  </rdf:RDF>
</x:xmpmeta>
```

Ein Beispiel aus der Praxis



upc cablecom

upc cablecom GmbH
Industriestrasse 19
8112 Otelfingen

Rechnungsnummer
103052836

Kundennummer
5398756-1

Rechnungsdatum
04.09.2013

Zahlungstermin
26.09.2013

P.P. 8021 Zürich

5398756-1/103052836/04092013/B/M1A/0001

Herr Jeremias Märki
Lützelmattestrasse 14
6006 Luzern

Rechnung September 2013

	Betrag	Gesamtbetrag
Saldo der letzten Rechnung	62.00	
Zahlungen berücksichtigt bis zum 3. September 2013	-62.00	
Ihr Guthaben		0.00
Abonnementspreise und Zusatzdienste	62.00	
Gesamtbetrag aktuelle Rechnungsperiode		62.00
Rechnungstotal gerundet (zahlbar bis 26.09.2013)		62.00

Zahlungseingänge sind bis zum 3. September 2013 berücksichtigt.

Bessere Metadaten

- [-] Dublin Core Properties (dc, <http://purl.org/dc/elements/1.1/>)
 - ... dc:format: application/pdf
 - [-] dc:language (bag container)
 - ... [1]: de
 - [-] dc:title (alt container)
 - ... [x-default]: UPC Cablecom Rechnung 103052836
 - [-] dc:creator (seq container)
 - ... [1]: upc cablecom GmbH
 - [-] dc:type (bag container)
 - ... [1]: Invoice
 - [-] dc:date (seq container)
 - ... [1]: 2013-10-07T17:01:53+02:00
 - [-] dc:subject (bag container)
 - ... [1]: Rechnung
- [-] <http://www.aiim.org/pdfa/ns/id/>
 - ... pdfaid:conformance: B
 - ... pdfaid:part: 1
- [+] PDF Properties (pdf, <http://ns.adobe.com/pdf/1.3/>)
- [+] XMP Core Properties (xmp, <http://ns.adobe.com/xap/1.0/>)
- [-] <http://upc-cablecom.ch/ns/metadata>
 - ... cab:payment-due: 2013-09-26
 - ... cab:period: 2013-09
 - ... cab:client: 5398756-1
 - ... cab:invoice-nr: 103052836
 - ... cab:invoice-date: 2013-09-04

Und jetzt?

- Schauen Sie sich um nach Metadaten in Ihrer Umgebung!
- Gehe hin und bewahre Metadaten!
- Reichern Sie Ihr Dokumente mit mehr Metadaten an!
- Bieten Sie Metadaten für den Erzeuger **und** für den Empfänger an!

Vielen Dank!!! Fragen?

Jeremias Märki
Software-Entwicklung und Beratung
Lützel mattstrasse 14
CH-6006 Luzern
Tel. +41 76 302 49 27
E-Mail: info@jeremias-maerki.ch
Website: <http://www.jeremias-maerki.ch>