Dokumentautomation mit XML am Beispiel einer Banddiskografie

Hochschulinformationstag in Merseburg – 13. April 2013

Dr. Thomas Meinike

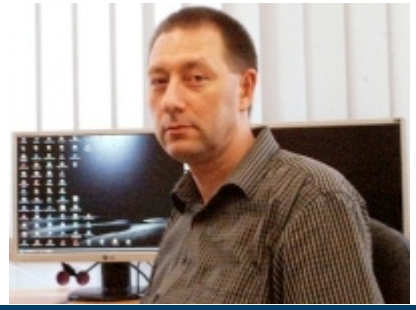
Hochschule Merseburg | FB Informatik und Kommunikationssysteme
http://www.iks.hs-merseburg.de/~meinike/
thomas.meinike@hs-merseburg.de

Zur Person

- Lehrkraft für besondere Aufgaben seit 1997

- Tätig in den Studiengängen im Fachbereich IKS:
  - Bachelor Technische Redaktion und E-Learning-Systeme
  - Master Technische Redaktion und Wissenskommunikation

- Lehr- und Arbeitsgebiete:
  - Auszeichnungs- und Skriptsprachen
  - Online-Dokumentation und Web-Entwicklung
  - **XML-Technologien**



Technische Redaktion in Merseburg

- Diplomstudiengang 1995 bis 2008
  Kommunikation und Technische Dokumentation

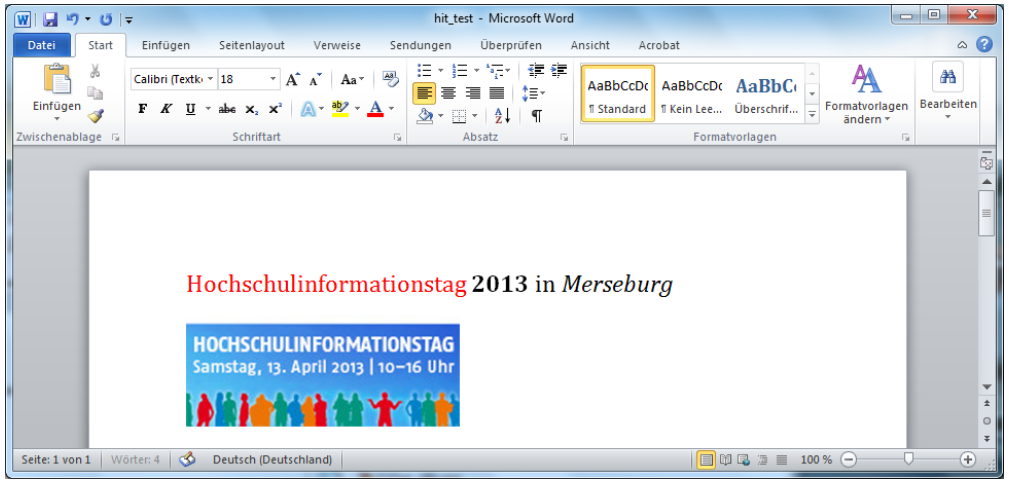
- Masterstudiengang seit 2006
  Technische Redaktion und Wissenskommunikation (4 Sem. / M. A.)

- Bachelorstudiengang seit 2010
  Technische Redaktion und
  E-Learning-Systeme
  (6 Sem. / B. Eng.)

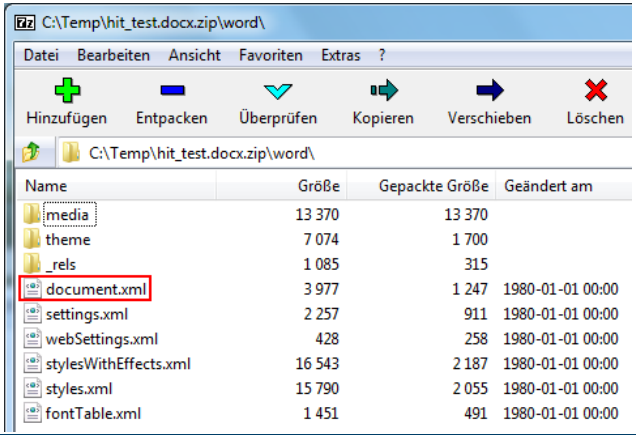
- Weitere Details zu den
  Inhalten gern im Anschluss.

Praktischer Einstieg (»In Word steckt XML drin!«)

➔ Word-Dokument mit Text, Formatierungen und Bild erstellen:



➔ Dateiendung *.docx* in *.zip* umbenennen und in 7-Zip öffnen (*document.xml* enthält den obigen Text):



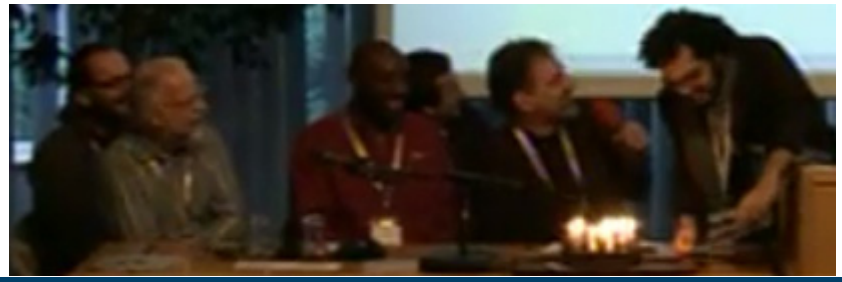
Praktischer Einstieg (»In Word steckt XML drin!«)

➔ Lässt sich ein Word-Dokument aus den Bestandteilen im Archiv auch **ohne Word** erzeugen?

Ja!

XML …

- **Ex**tensible **M**arkup **L**anguage
- … wurde 1998 vom World Wide Web Consortium (W3C) eingeführt
- … ist keine eigenständige Sprache, sondern ein Konzept zur Definition von konkreten Auszeichnungssprachen
- … definiert (relativ wenige) Regeln, welche beim Schreiben und der anschließenden Verarbeitung exakt einzuhalten sind.



15 Jahre XML (Prag, 10.02.13)

XML ...

➔ Grundaufbau der Banddiskografie:

```
<?xml version="1.0" encoding="UTF-8"?>
<diskografie bandname="...>
  <bandinfo>Text</bandinfo>
  <referenzen> <referenz url="http://...">Linktext</referenz> </referenzen>
  <werk jahr="..." typ="...">
    <werkname>Text</werkname>
    <kommentar>Text</kommentar>
    <coverbild breite="..." hoehe="...">bildname.jpg</coverbild>
    <titelliste>
      <titel spielzeit="...">Text</titel>
    </titelliste>
  </werk>
</diskografie>
```

In Kürze:

Die Dokument-Struktur entsteht durch (ver-schachtelte) Elemente, welche mittels Tags markiert werden. Dazwischen steht der eigentliche Inhalt.

<elname>...</elname>

Attribute in der Form atname="atwert" enthalten zusätzliche Informationen.

Spezielle Datenmodelle helfen beim Bearbeiten und Prüfen.

XML …

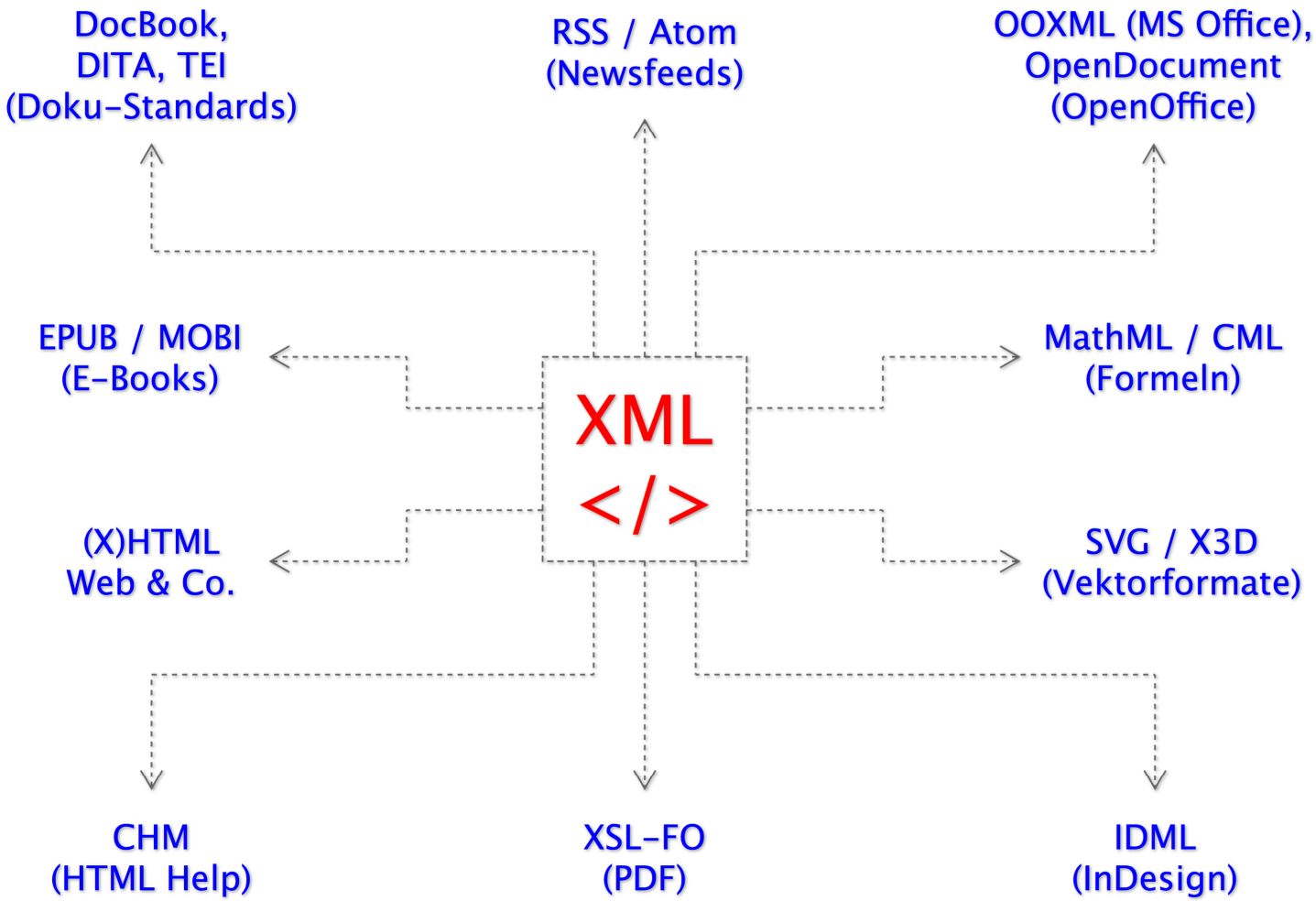
- … bildet mit weiteren Konzepten wie XSLT, XSL-FO und XPath eine Technologiefamilie

- … wird mittlerweile für vielfältige Anwendungen, speziell in der Technischen Dokumentation, eingesetzt (Single-Source-Publishing)

- … unterstützt eigene Entwicklungen neben der Anwendung »gebrauchsfertiger« Sprachen (wie HTML)

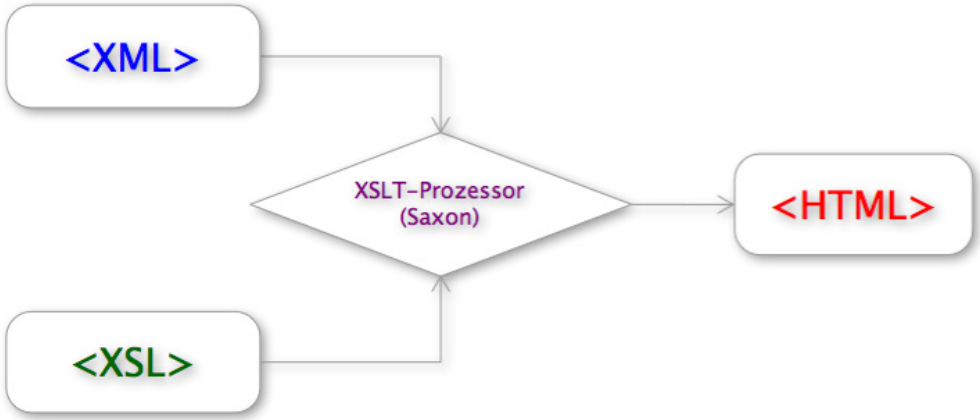
- … ermöglicht in Bereichen wie Technische Redaktion und Verlagswesen automatisierte Umsetzungen von Webinhalten, Onlinehilfen, PDF-Dateien und E-Books.

XML-Anwendungen (→ TR-Studium)



XML-Verarbeitung

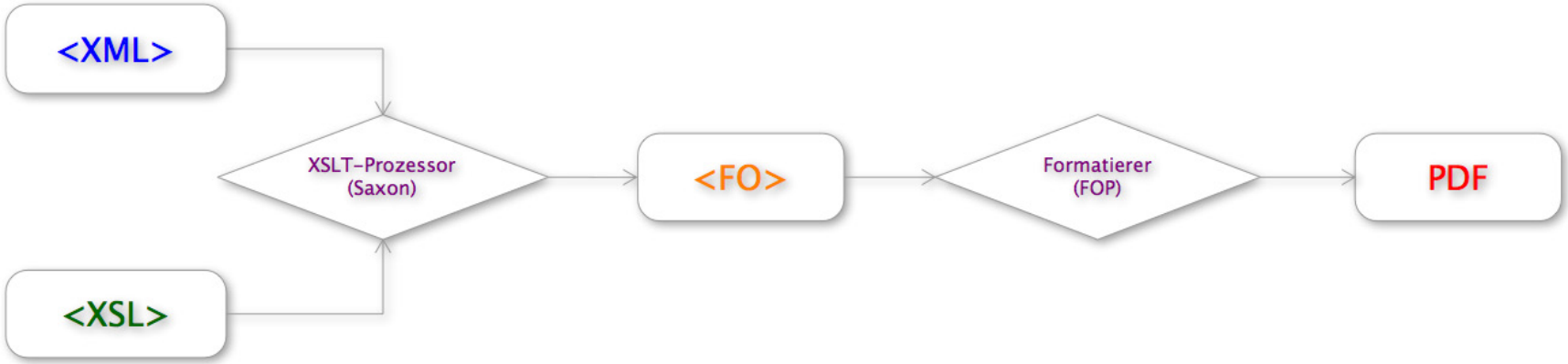
➔ **HTML**-Generierung nach diesem Ablaufschema:



- Die XML-Struktur wird mit einem XSL(T)-Stylesheet (= Programm) in das Ausgabeformat HTML transformiert.
- HTML ist in Medienprodukten wie EPUB und CHM enthalten.

XML-Verarbeitung

➔ Prozesskette der automatisierten **PDF**-Produktion:



– Hier wird zunächst ein Zwischenformat (XSL-FO) erzeugt und dieses im zweiten Schritt zum fertigen PDF verarbeitet.

XML-Verarbeitung

➔ Produktion von Archivformaten wie **DOCX**, **EPUB**, **IDML**, **CHM**:

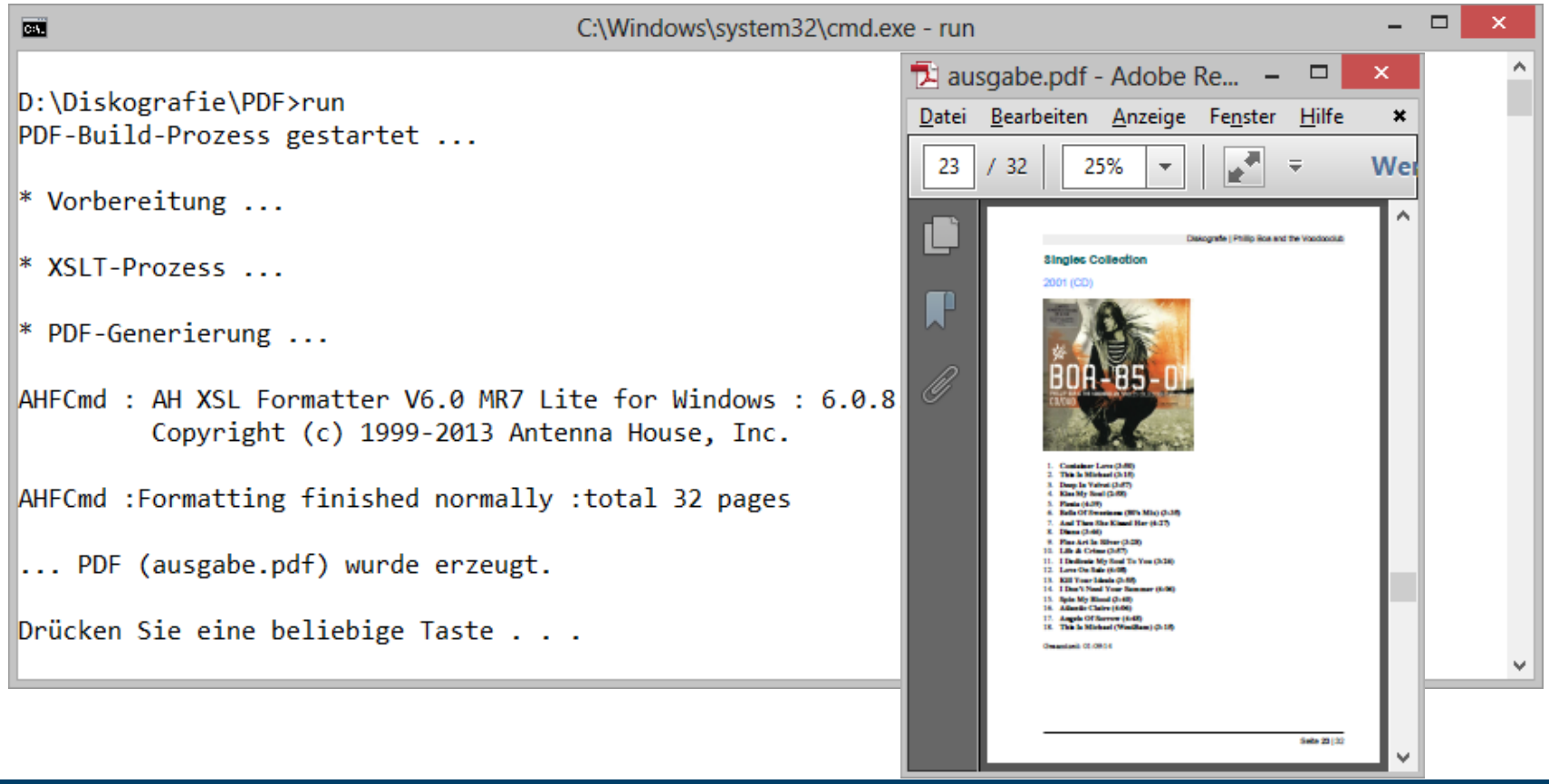
- Detaillierte Analyse der Dokumentstrukturen
- Anlegen der Zielstruktur
- Generierung der Inhalts- und Zusatzdateien (XSLT)
- Hinzufügen statischer Dateien, z. B. Bilder und Fonts
- Kompression mit ZIP-Werkzeug
- Dateiendung .zip umbenennen (→ .docx, .epub, .idml)

- Für das binäre Hilfeformat CHM (HTML Help) wird zusätzlich ein externer Hilfecompiler benötigt (→ .chm).

- Der automatisierte Ablauf wird mittels Batch- bzw. Shellskript gesteuert (siehe Demonstrationen).

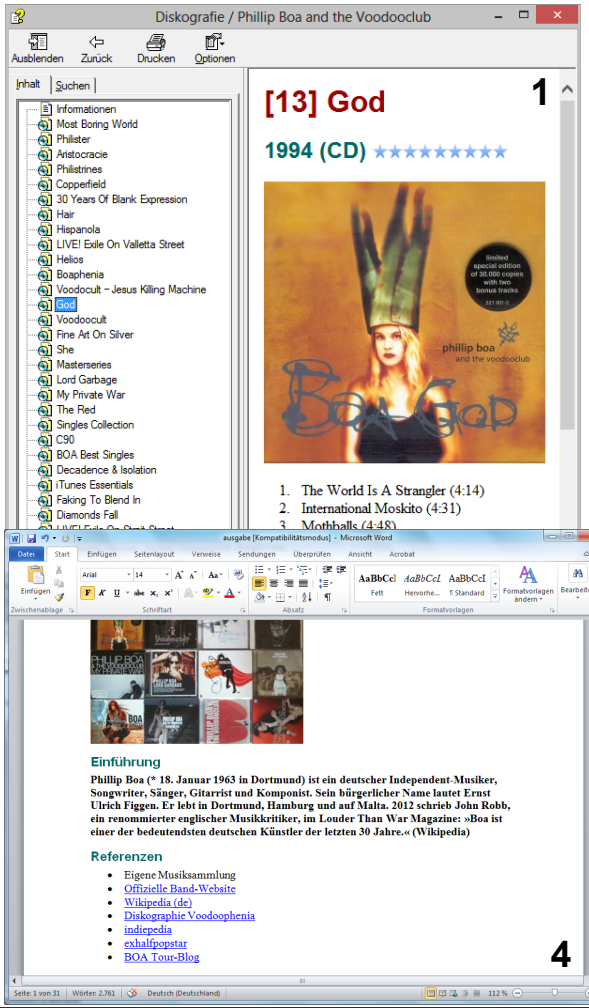
XML-Verarbeitung

➔ Beispiel PDF-Produktion:



{Praktische Demonstration …}

Diskografie-Ausgaben



1 = CHM (Help Viewer)
2 = IDML (InDesign)
3 = EPUB (iBooks)
4 = DOCX (Word)
5 = PDF (Adobe Reader)

Zusammenfassung und Ausblick

- XML und zugehörige Konzepte haben sich in der Informations- und Wissensgesellschaft fest etabliert.
- Das Anwendungsspektrum reicht von der strukturierten Ablage von Informationen, über im Alltag genutzte Dateiformate bis zu komplexen, automatisierten Publikationsszenarien.
- XML-Technologien lassen sich bereits mit einfachen Mitteln studieren und ausprobieren.
- Um letztlich attraktive Medienprodukte zu produzieren, ist neben den theoretischen Grundlagen viel Entwicklerpraxis nötig (z. B. durch Bearbeitung entsprechender Bachelor- und Masterthemen).