

# Dokumentautomation mit XML am Beispiel einer Banddiskografie

Hochschulinformationstag in Merseburg – 18. April 2015

**Dr. Thomas Meinike**

Hochschule Merseburg | FB Informatik und Kommunikationssysteme

<http://www.iks.hs-merseburg.de/~meinike/>

[thomas.meinike@hs-merseburg.de](mailto: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 Informationsdesign und Medienmanagement
- Lehr- und Arbeitsgebiete:
  - Auszeichnungs- und Skriptsprachen
  - Online-Dokumentation und Web-Entwicklung
  - XML-Technologien



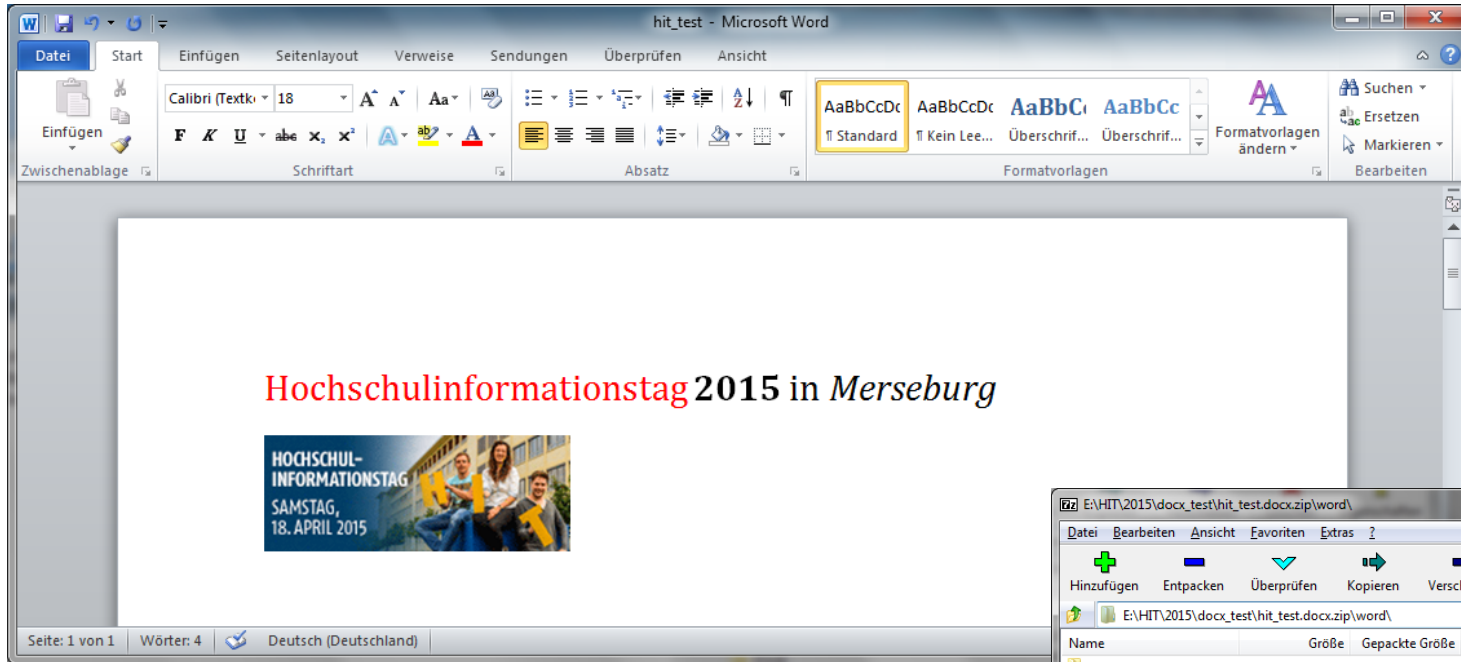
# Technische Redaktion in Merseburg

- Diplomstudiengang 1995 bis 2008  
Kommunikation und Technische Dokumentation
- Masterstudiengang seit 2006  
Technische Redaktion und Wissenskommunikation  
seit 2014: Informationsdesign und Medienmanagement (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.

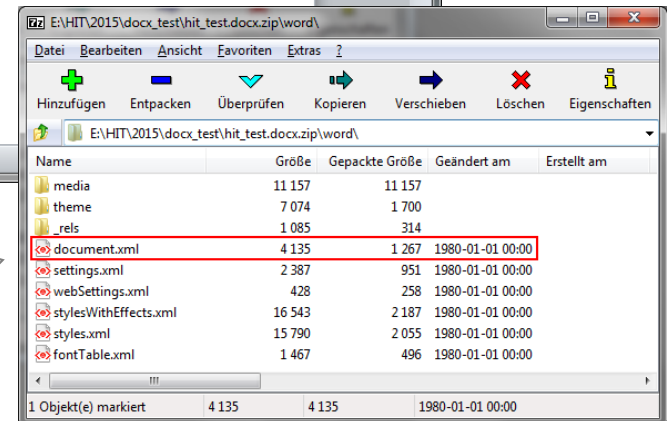


# Experiment: »In Word steckt XML drin!« (1)

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

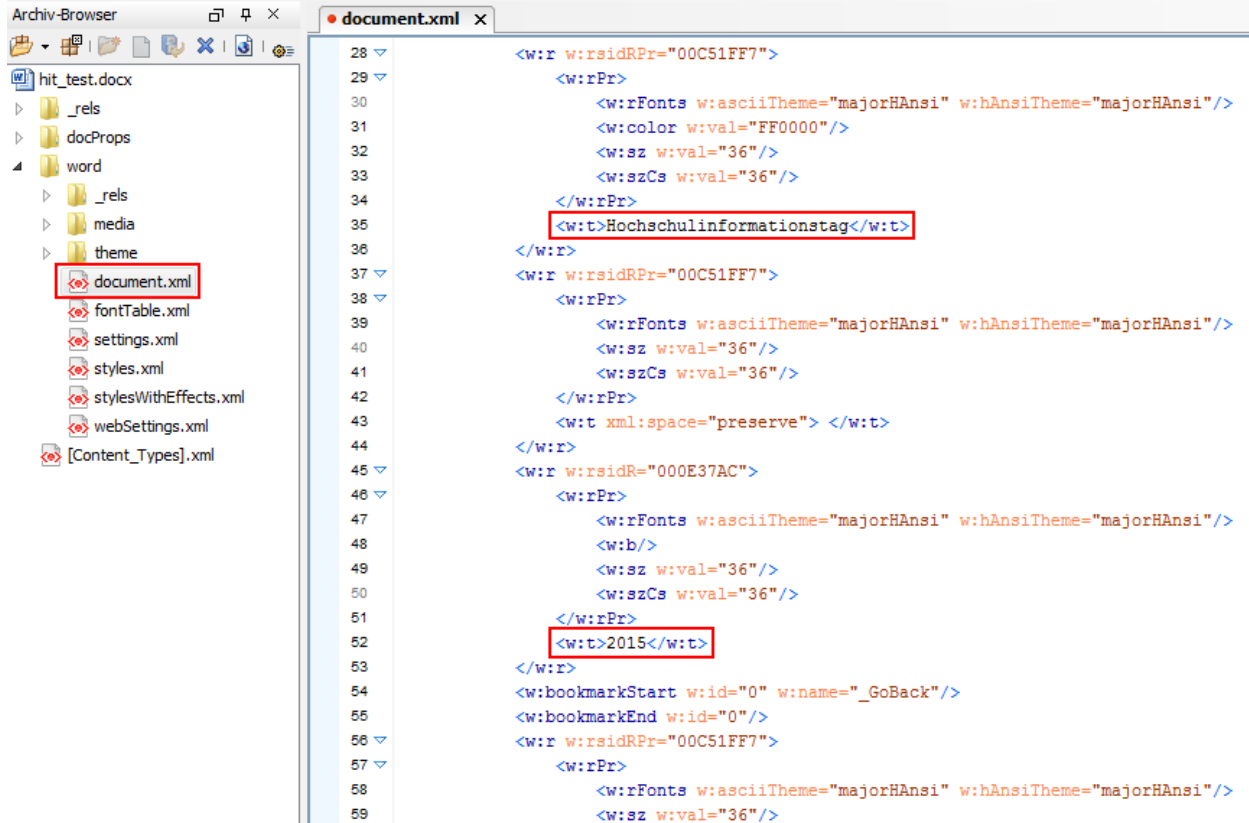


➔ Dateiendung *.docx* in *.zip* umbenennen und in 7-Zip öffnen (*document.xml* enthält den obigen Text):



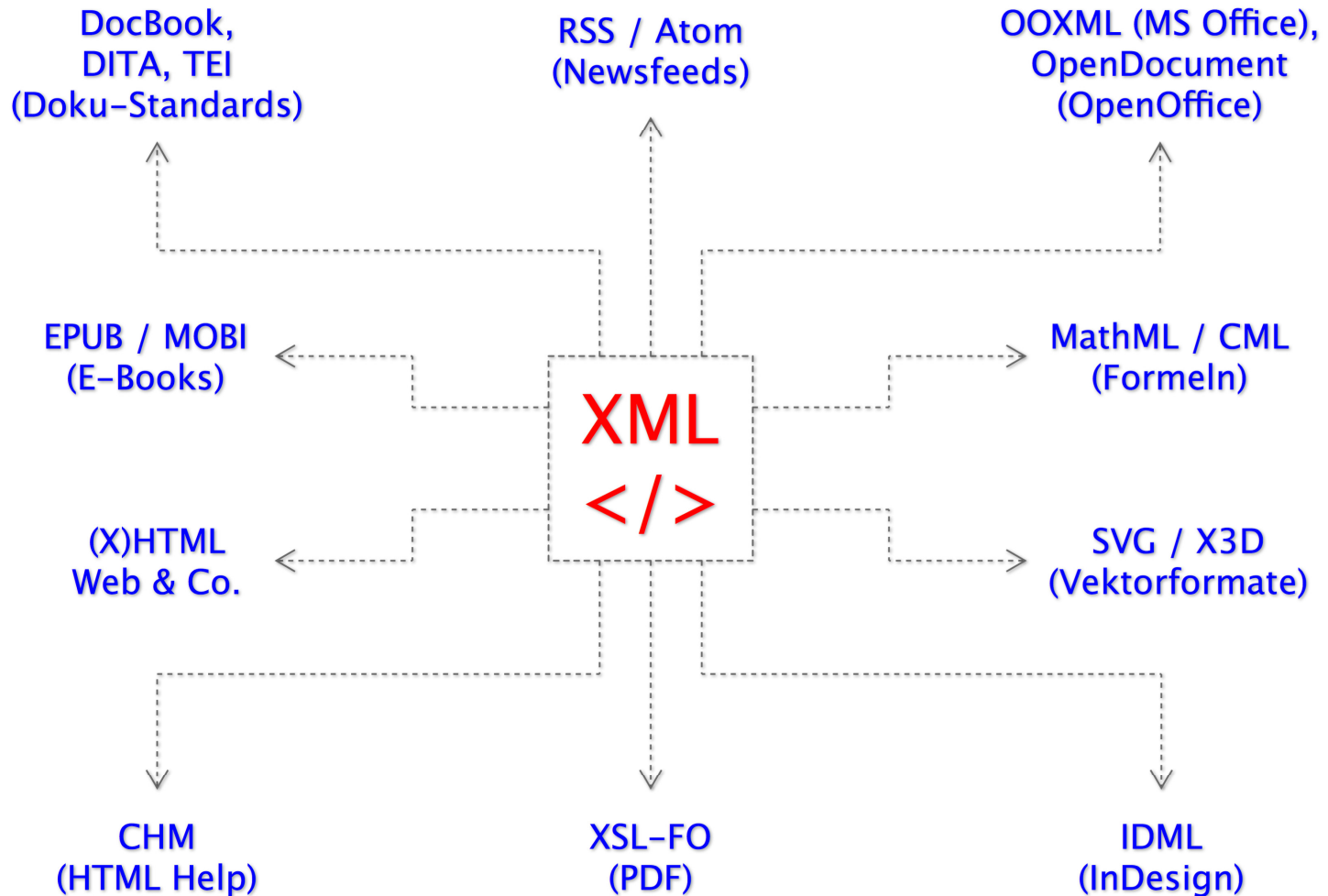
# Experiment: »In Word steckt XML drin!« (2)

➔ Mit Detailkenntnissen lässt sich ein Word-Dokument aus den Bestandteilen im Archiv auch **ohne Word** erzeugen:



```
28 <w:r w:rsidRPr="00C51FF7">
29   <w:rPr>
30     <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
31     <w:color w:val="FF0000"/>
32     <w:sz w:val="36"/>
33     <w:szCs w:val="36"/>
34   </w:rPr>
35   <w:t>Hochschulinformationstag</w:t>
36 </w:r>
37 <w:r w:rsidRPr="00C51FF7">
38   <w:rPr>
39     <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
40     <w:sz w:val="36"/>
41     <w:szCs w:val="36"/>
42   </w:rPr>
43   <w:t xml:space="preserve"> </w:t>
44 </w:r>
45 <w:r w:rsidR="000E37AC">
46   <w:rPr>
47     <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
48     <w:b/>
49     <w:sz w:val="36"/>
50     <w:szCs w:val="36"/>
51   </w:rPr>
52   <w:t>2015</w:t>
53 </w:r>
54 <w:bookmarkStart w:id="0" w:name="_GoBack"/>
55 <w:bookmarkEnd w:id="0"/>
56 <w:r w:rsidRPr="00C51FF7">
57   <w:rPr>
58     <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
59     <w:sz w:val="36"/>
```

# XML-Anwendungen (→ TR-Studium)



# XML ...

- ... kurz für **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 Regeln zur Auszeichnung von Dokumenten.
- ... ermöglicht die (automatisierte) Erstellung von Webinhalten, Onlinehilfen und E-Books sowie Print-Erzeugnissen (als PDF).
- ... zur Verarbeitung dienen spezielle Standards und Technologien.

# XML-Codebeispiel

→ 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 **Elemente**, welche mittels Tags `<...>` markiert werden:

```
<name>Inhalt</name>
```

Zusätzliche Informationen vermitteln **Attribute**:

```
name="wert"
```

Spezielle Datenmodelle helfen beim Bearbeiten und Prüfen.

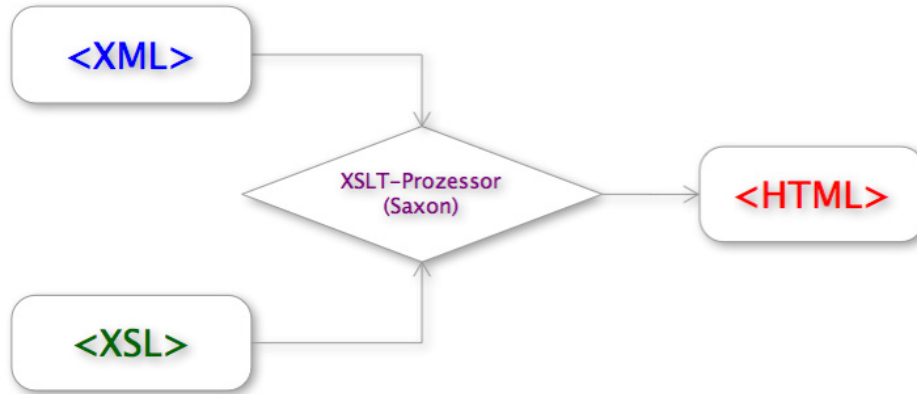
# XML-Codebeispiel

→ Grundaufbau der Banddiskografie (ein werk-Element):

```
<werk jahr="2013" typ="CD" bewertung="10">  
  <werkname>Push The Sky Away</werkname>  
  <kommentar>Großartige Platte für die Insel. [...]</kommentar>  
  <coverbild breite="300" hoehe="300">pushtheskyaway.jpg</coverbild>  
  <titelliste>  
    <titel spielzeit="4:04">We No Who U R</titel>  
    <titel spielzeit="3:40">Wide Lovely Eyes</titel>  
    <titel spielzeit="3:49">Water's Edge</titel>  
    <!-- drei weitere titel-Elemente ... -->  
    <titel spielzeit="7:50">Higgs Boson Blues</titel>  
    <titel spielzeit="4:07">Push The Sky Away</titel>  
  </titelliste>  
</werk>
```

# XML-Verarbeitung (1)

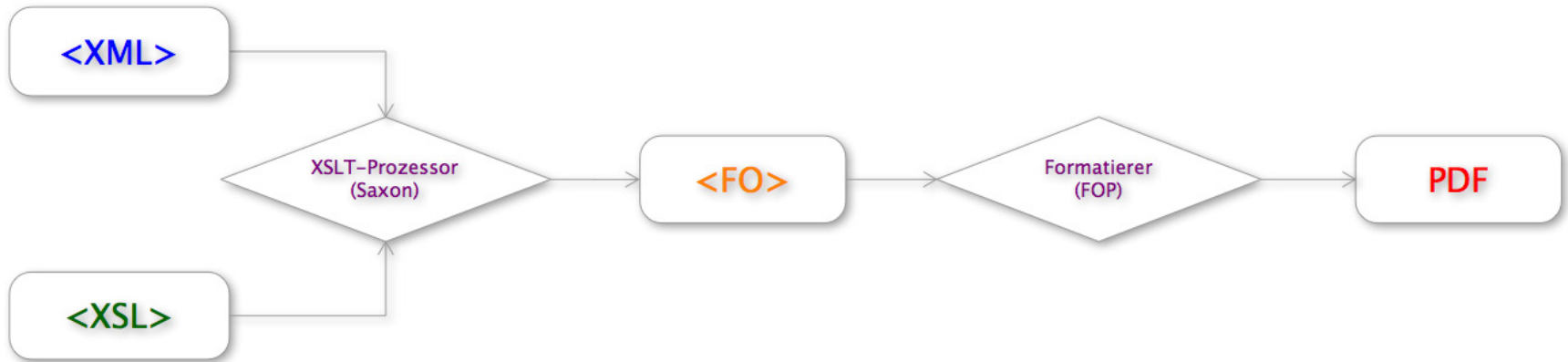
→ Ablaufschema zur **HTML**-Generierung:



- Die XML-Struktur wird mit einem XSLT-Stylesheet (Programmcode) in das Ausgabeformat **HTML** transformiert.
- Medienprodukte wie E-Books und Onlinehilfen enthalten **HTML**.

# XML-Verarbeitung (2)

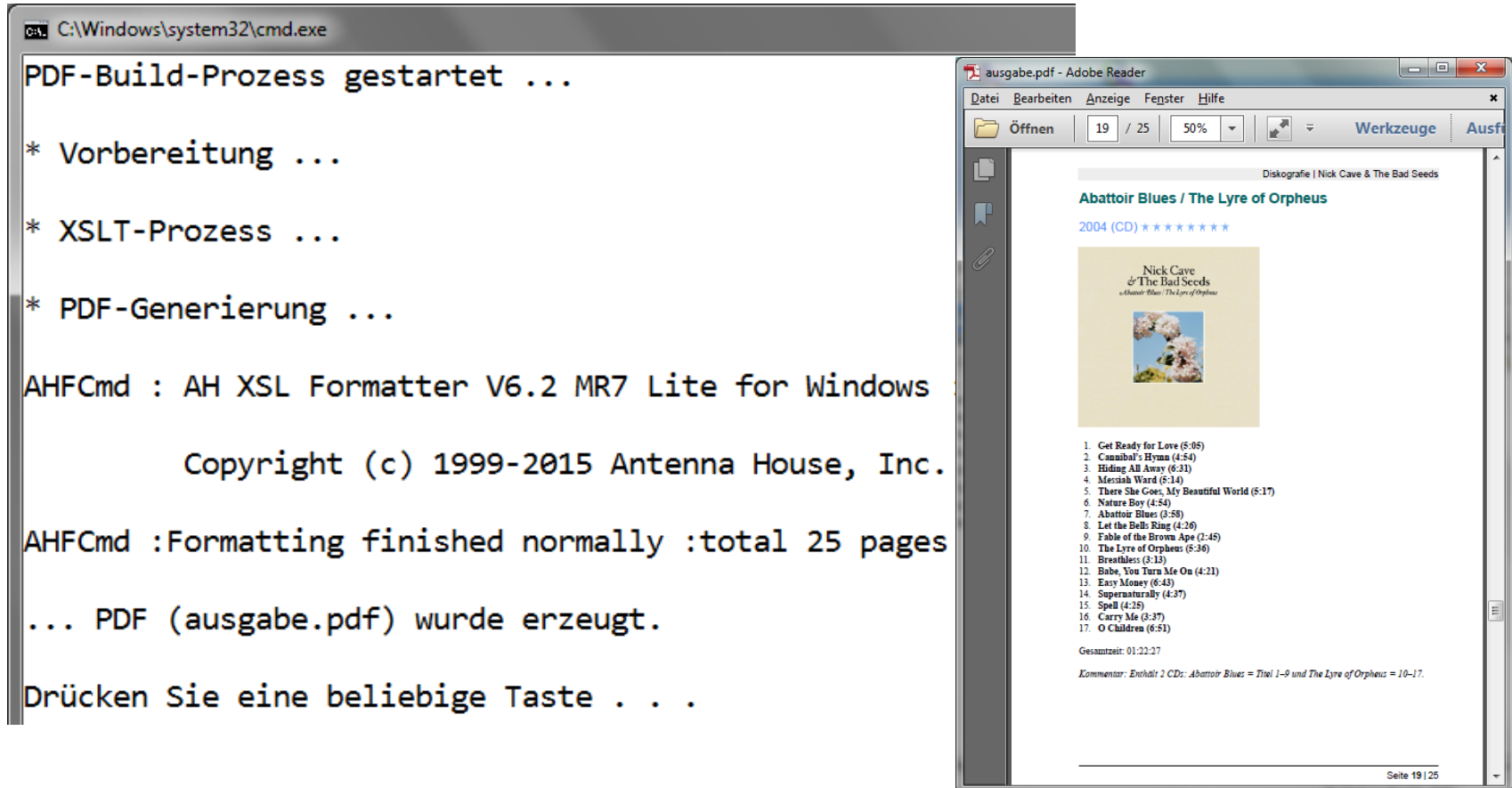
→ Prozesskette der automatisierten **PDF**-Produktion:



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

# XML-Verarbeitung (3)

→ Ablauf eines Konsolenskriptes zur **PDF**-Produktion:



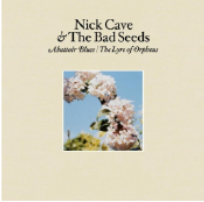
The image shows two overlapping windows. The left window is a command prompt titled "C:\Windows\system32\cmd.exe" with the following text:

```
PDF-Build-Prozess gestartet ...  
  
* Vorbereitung ...  
  
* XSLT-Prozess ...  
  
* PDF-Generierung ...  
  
AHFCmd : AH XSL Formatter V6.2 MR7 Lite for Windows  
  
Copyright (c) 1999-2015 Antenna House, Inc.  
  
AHFCmd :Formatting finished normally :total 25 pages  
  
... PDF (ausgabe.pdf) wurde erzeugt.  
  
Drücken Sie eine beliebige Taste . . .
```

The right window is Adobe Reader titled "ausgabe.pdf - Adobe Reader" showing a discography page for "Abattoir Blues / The Lyre of Orpheus" by Nick Cave & The Bad Seeds. The page includes a tracklist and a comment.

ausgabe.pdf - Adobe Reader  
Datei Bearbeiten Anzeige Fenster Hilfe  
Öffnen 19 / 25 50% Werkzeuge Ausf

Diskografie | Nick Cave & The Bad Seeds  
**Abattoir Blues / The Lyre of Orpheus**  
2004 (CD) ★★★★★★



1. Get Ready for Love (5:05)  
2. Cannibal's Hymn (4:54)  
3. Hiding All Away (6:31)  
4. Messiah Ward (5:14)  
5. There She Goes, My Beautiful World (5:17)  
6. Nature Boy (4:54)  
7. Abattoir Blues (3:58)  
8. Let the Bells Ring (4:26)  
9. Fable of the Brown Ape (7:45)  
10. The Lyre of Orpheus (5:36)  
11. Breathless (3:13)  
12. Babe, You Turn Me On (4:21)  
13. Easy Money (6:43)  
14. Supernaturally (4:37)  
15. Spell (4:25)  
16. Carry Me (3:37)  
17. O Children (6:51)

Gesamtzeit: 01:22:27  
Kommentar: Enthält 2 CDs: Abattoir Blues = Titel 1-9 und The Lyre of Orpheus = 10-17.

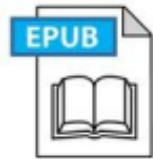
Seite 19 | 25

# XML-Verarbeitung (4)

→ 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 mit Batch-Skripten gesteuert.

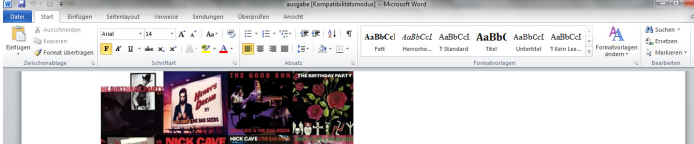
# {Ausgaben erzeugen ...}



# Fertige Diskografie-Ausgaben

**[7] Henry's Dream**  
1992 (LP) ★★★★★★

1. Papa Won't You Leave Henry (5:43)
2. I Had A Dream, Joe (3:42)
3. Straight To You (4:35)



## Einführung

The Bad Seeds ist eine australische Rockband, die 1983 vom Sänger Nick Cave, Multinstrumentalist Mick Harvey und Gitarrist Blixa Bargeld nach Auflösung von The Birthday Party in West-Berlin gegründet wurde. Der Name basiert auf dem Roman The Bad Seed (deutsch: „Böse Saat“) von William March. (Wikipedia)

## Referenzen

- Eigene Musiksammlung
- Offizielle Band-Website
- Wikipedia (de)
- Wikipedia (en)
- indiepanda

4

**Kicking Against The Pricks**  
1985 (LP) ★★★★★★

1. Muddy Water (5:15)
2. I'm Gonna Kill That Woman (3:44)
3. Sleeping Amulet (3:14)
4. Long Black Veil (3:47)
5. Hey Joe (3:56)
6. The Singer (3:09)
7. All Tomorrow's Parties (5:52)
8. By The Time I Get To Phoenix (3:39)
9. The Hammer Song (3:00)
10. Something's Gotten Hold Of My Heart (3:44)
11. Jesus Met The Woman At The Well (2:00)
12. The Carrion & Crow (3:16)

Kommentar: LP-Version ohne Titel 7 und 8.

1 = CHM (HTML Help Viewer)

2 = IDML (InDesign)

3 = EPUB (iBooks / iPad)

4 = DOCX (Word)

5 = PDF (Adobe Reader)

**[15] Nocturama**  
2003 (CD) ★★★★★★

1. Wonderful Life (6:49)
2. He Wants You (3:31)
3. Right Out of Your Hand (5:15)
4. Bring It On (5:23)
5. Dead Man in My Bed (4:41)
6. Still in Love (4:44)

3

**Tender Prey**  
1988 (LP) ★★★★★★

1. The Mercy Seat (7:17)
2. Up Jumped The Devil (5:16)
3. Deanna (3:46)
4. Watching Alice (4:01)
5. Mercy (6:22)
6. City Of Refuge (4:48)
7. Slowly Goes The Night (5:23)
8. Sunday's Slave (3:40)
9. Sugar Sugar Sugar (5:01)
10. New Morning (3:46)
11. The Mercy Seat (Video Mix) (5:05)

5

# 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).