

Dokumentautomation mit XML am Beispiel einer Banddiskografie

Hochschulinformationstag in Merseburg – 05. April 2014

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 (Informationsdesign und Medienmanagement)
- Lehr- und Arbeitsgebiete:
 - Auszeichnungs- und Skriptsprachen
 - Online-Dokumentation und Web-Entwicklung
 - XML-Technologien



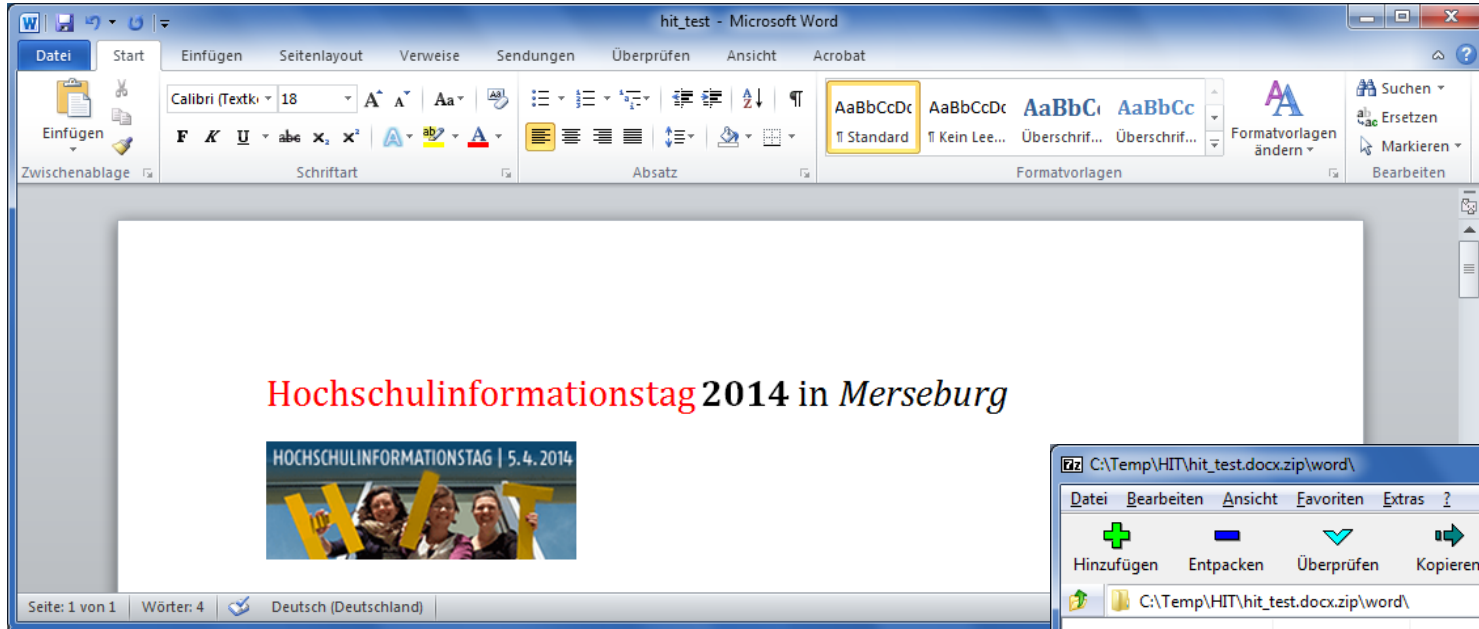
Technische Redaktion in Merseburg

- Diplomstudiengang 1995 bis 2008
Kommunikation und Technische Dokumentation
- Masterstudiengang seit 2006
Technische Redaktion und Wissenskommunikation
ab 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.

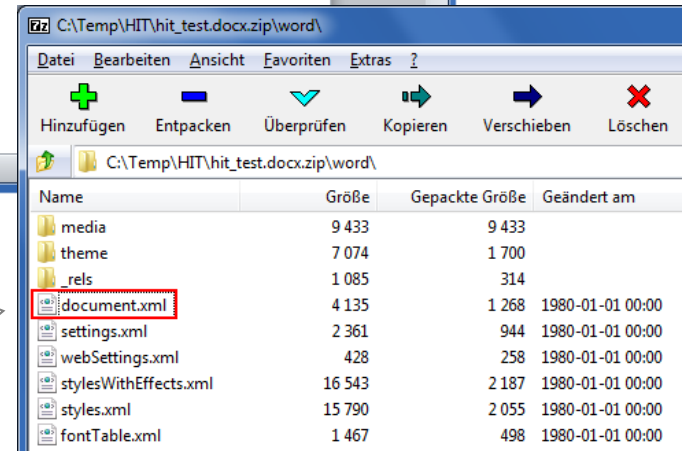


Experiment: »In Word steckt XML drin!« (1)

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



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



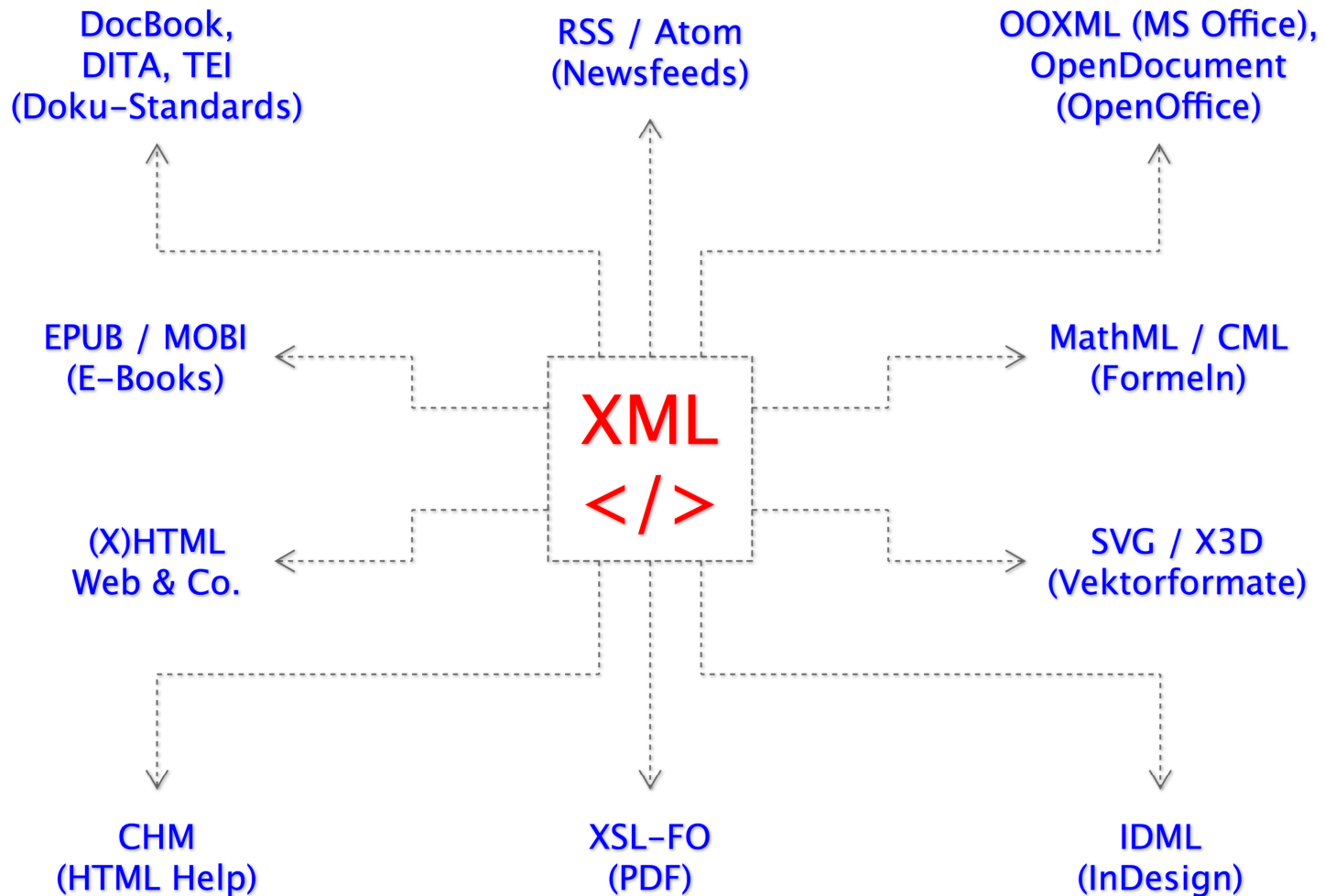
Experiment: »In Word steckt XML drin!« (2)

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



```
29 <w:r w:rsidRPr="00C51FF7">
30 <w:rPr>
31 <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
32 <w:color w:val="FF0000"/>
33 <w:sz w:val="36"/>
34 <w:szCs w:val="36"/>
35 </w:rPr>
36 <w:t>Hochschulinformationstag</w:t>
37 </w:r>
38 <w:r w:rsidRPr="00C51FF7">
39 <w:rPr>
40 <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
41 <w:sz w:val="36"/>
42 <w:szCs w:val="36"/>
43 </w:rPr>
44 <w:t xml:space="preserve"> </w:t>
45 </w:r>
46 <w:r w:rsidRPr="00C51FF7">
47 <w:rPr>
48 <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
49 <w:b/>
50 <w:sz w:val="36"/>
51 <w:szCs w:val="36"/>
52 </w:rPr>
53 <w:t>2014</w:t>
54 </w:r>
55 <w:r w:rsidRPr="00C51FF7">
56 <w:rPr>
57 <w:rFonts w:asciiTheme="majorHAnsi" w:hAnsiTheme="majorHAnsi"/>
58 <w:sz w:val="36"/>
59 <w:szCs w:val="36"/>
```

XML-Anwendungen (→ TR-Studium)



XML ...

- ... kurz für **Extensible Markup Language**.
- ... 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>
```

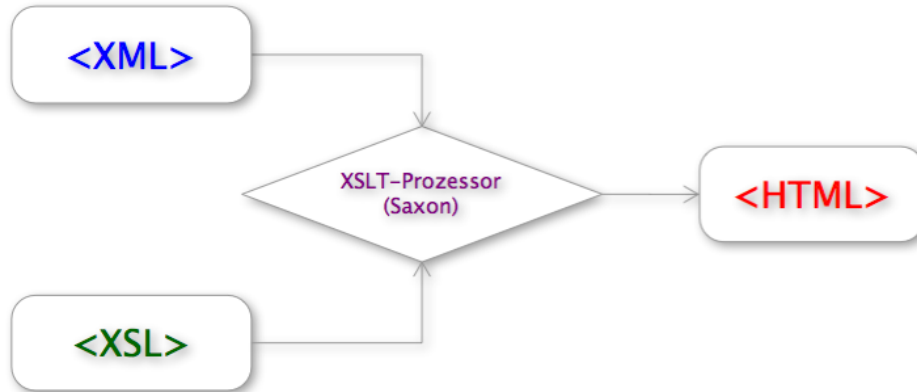
Zusätzliche Informationen vermitteln **Attribute**:

```
name="wert"
```

Spezielle Datenmodelle helfen beim Bearbeiten und Prüfen.

XML-Verarbeitung (1)

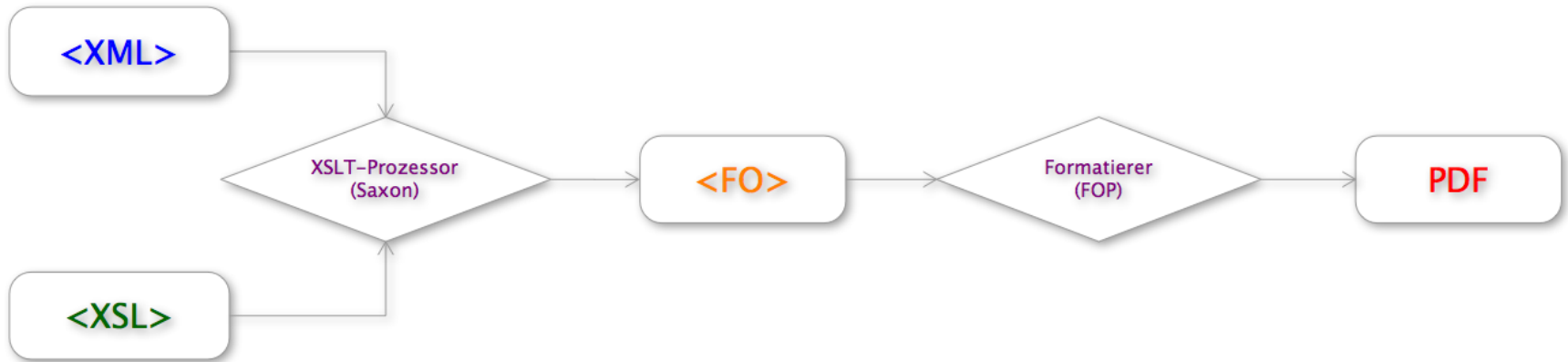
→ Ablaufschema zur **HTML**-Generierung:



- 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)

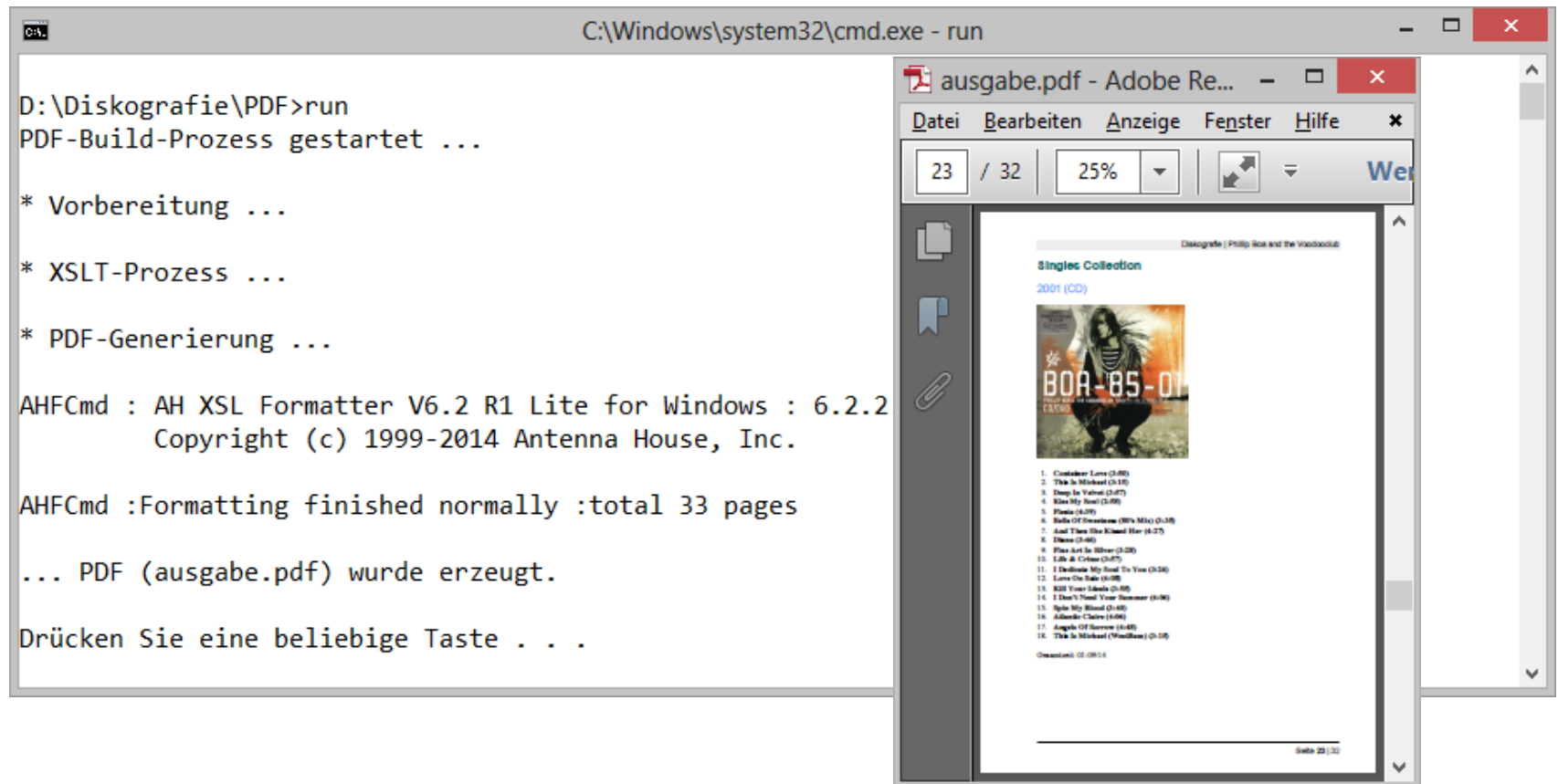
→ Prozesskette der automatisierten **PDF**-Produktion:



- 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 screenshot shows a Windows command prompt window titled "C:\Windows\system32\cmd.exe - run" and an Adobe Reader window titled "ausgabe.pdf - Adobe Re...".

The command prompt shows the following output:

```
D:\Diskografie\PDF>run
PDF-Build-Prozess gestartet ...

* Vorbereitung ...

* XSLT-Prozess ...

* PDF-Generierung ...

AHFCmd : AH XSL Formatter V6.2 R1 Lite for Windows : 6.2.2
        Copyright (c) 1999-2014 Antenna House, Inc.

AHFCmd :Formatting finished normally :total 33 pages

... PDF (ausgabe.pdf) wurde erzeugt.

Drücken Sie eine beliebige Taste . . .
```

The Adobe Reader window shows the PDF content, which is a discography for Philip Sosa and the VooDooBats. The title is "Single Collection" and the year is "2001 (CD)". The cover image shows a person in a dark, hooded outfit. The track list is as follows:

1. Cradle of Love (3:05)
2. This Is Myland (3:15)
3. Deep In Your (3:07)
4. Kiss My Soul (3:05)
5. Plans (3:07)
6. Back Off Sweetness (With Mike) (3:15)
7. And You Are Kissing Her (3:27)
8. Heaven (3:05)
9. She Ain't In Love (3:05)
10. Life Is Crime (3:07)
11. I Believe My Soul Is Free (3:24)
12. Love On the Side (3:05)
13. Kiss Your Soul (3:05)
14. I Don't Need Your Love (3:05)
15. Split My Soul (3:05)
16. Adrenaline Choke (3:05)
17. Angles Of Heaven (3:05)
18. This Is Myland (Remix) (3:15)

The footer of the PDF page reads "Discography 01-08-01" and "Seite 23 | 33".

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 Skripten gesteuert (Demos).

{Praktische Demonstration ...}



Fertige Diskografie-Ausgaben

1 = CHM (Help Viewer)

2 = IDML (InDesign)

3 = EPUB (iBooks)

4 = DOCX (Word)

5 = PDF (Adobe Reader)

Web-Anwendung

→ Entwickelt mit Saxon-CE,
XSLT 2.0 und JavaScript.

Firefox

Diskografie (CEdition) +

Diskografie (CEdition)

Katalogeinträge: [Einstürzende Neubauten](#) [Nick Cave & The Bad Seeds](#) [Phillip Boa and the Voodooclub](#)

Phillip Boa and the Voodooclub

Bandinfo: Phillip Boa (* 18. Januar 1963 in Dortmund) ist ein deutscher Independent-Musiker, Songwriter, Sänger, Gitarrist und Komponist. Sein bürgerlicher Name lautet Ernst Ulrich Figgen. Er lebt in Dortmund, Hamburg und auf Malta. 2012 schrieb John Robb, ein renommierter englischer Musikkritiker, im Louder Than War Magazine: »Boa ist einer der bedeutendsten deutschen Künstler der letzten 30 Jahre.« (Wikipedia)

Referenzen: Eigene Musiksammlung | [Offizielle Band-Website](#) | [Wikipedia \(de\)](#) | [voodoophenia](#) | [indiepedia](#)


Album:

- ▷ **Most Boring World**
- ▷ Philister
- ▷ Aristocracie
- ▷ Philistrines
- ▷ Copperfield
- ▷ 30 Years Of Blank Expression
- ▷ Hair
- ▷ Hispanola
- ▷ LIVE! Exile On Valletta Street
- ▷ Helios
- ▷ Boaphenia
- ▷ Voodocult – Jesus Killing Machine
- ▷ God
- ▷ Voodoocult
- ▷ Fine Art On Silver
- ▷ She
- ▷ Masterseries
- ▷ Lord Garbage
- ▷ My Private War
- ▷ The Red
- ▷ Singles Collection
- ▷ C90
- ▷ BOA Best Singles
- ▷ Decadence & Isolation
- ▷ iTunes Essentials
- ▷ Faking To Blend In
- ▷ Diamonds Fall
- ▷ LIVE! Exile On Strait Street
- ▷ Loyalty
- ▷ REDUCED! A more or less acoustic performance

Details:

Most Boring World

1984 (LP)



Titel:

1. All I Hate Is You (2:58)
2. Dishwashers (2:40)
3. Heroes (3:45)
4. Dance The Candican (3:23)
5. Lovers (4:32)
6. Constricted (1:49)

Gesamtzeit: 00:19:07

Kommentar:
Mini-Album, nur als Vinyl, limitiert auf 500 Stück erschienen.

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).