



# XML-Die Lösung? -- Konvertierungstools DissOnline Aktivitäten

INETBIB-Tagung 2002 Göttingen

Thomas Pieruschka<sup>†</sup>; Dr. Nikola Korb<sup>‡</sup>

<sup>†</sup>Universitätsbibliothek der Universität der Bundeswehr  
München

<sup>‡</sup> Koordinierungsstelle *DissOnline*; Kathrin Schroeder DDB  
Susanne Dobratz, Matthias Schulz Humboldt-Univ. Berlin



# Inhalt

Online-Dissertationen  
Persistent Identifier

XML/SGML  
Dokument-Struktur

Konvertierung  
Word, Latex

XML-Editoren



# DissOnline-Aktivitäten

## Warum Online-Dissertationen?

- gezielte Recherche in Dissertationen auf verteilten Servern,
- sofortige Verfügbarkeit des Dokuments und damit,
- rasche Integration neuer Ergebnisse in die eigene Forschung,
- sofortige Überprüfbarkeit von Verweisen und Zitaten,
- kostengünstige Veröffentlichungen mit vielfältigen Darstellungsmöglichkeiten,
- Beschleunigung bibliothekarischer Arbeit sowie platzsparende Archivierung.



## Zugriff auf Online-Dissertationen

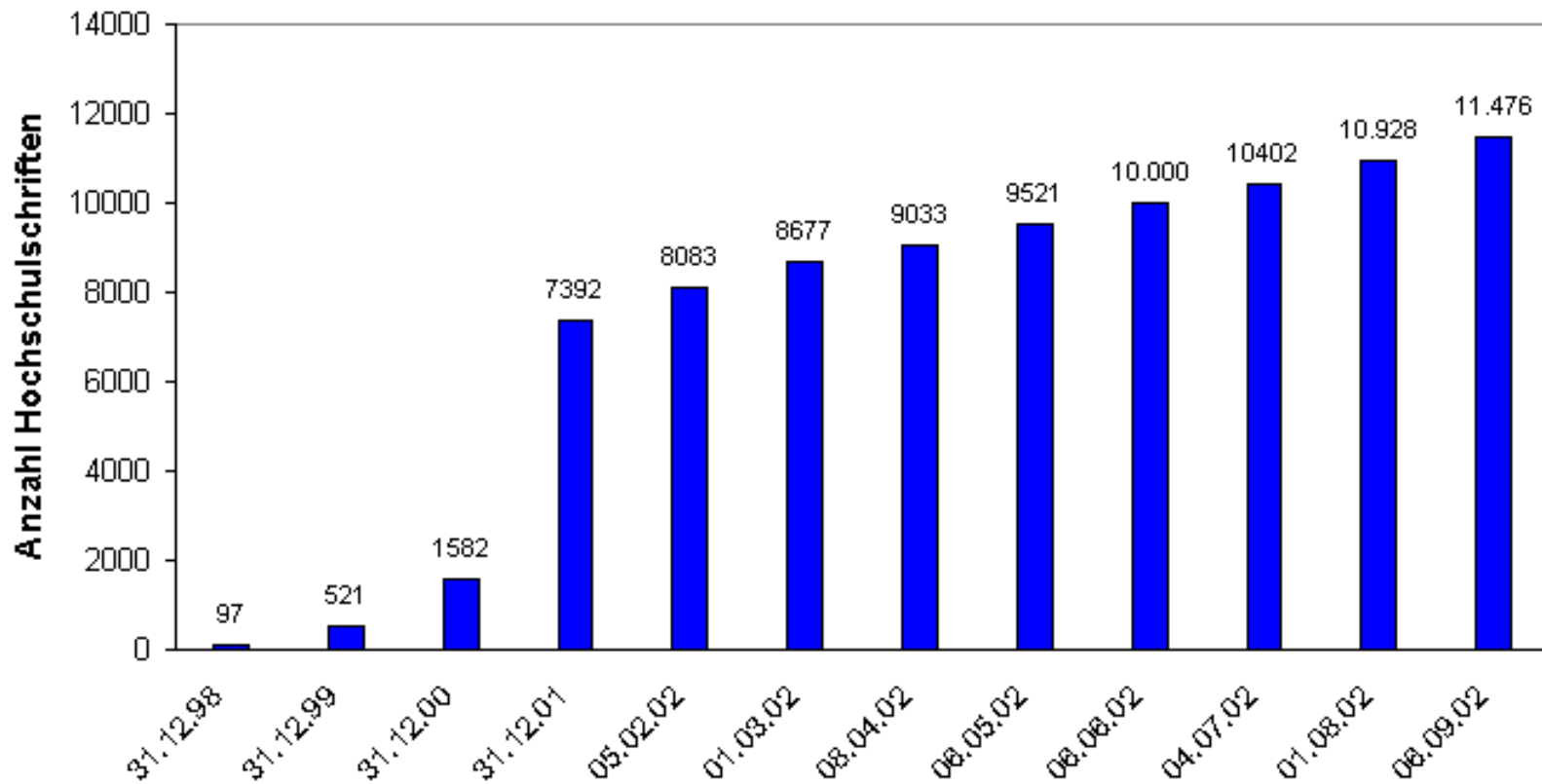
- <http://dbf-opac.ddb.de/>
- <http://www.iwi-iuk.org/dienste/TheO>
- <http://elib.uni-stuttgart.de/opus/>
- <http://www.dini.de/oaisuche/>
- <http://scholar.lib.vt.edu/theses/>



# Aufgaben der Koordinierungsstelle

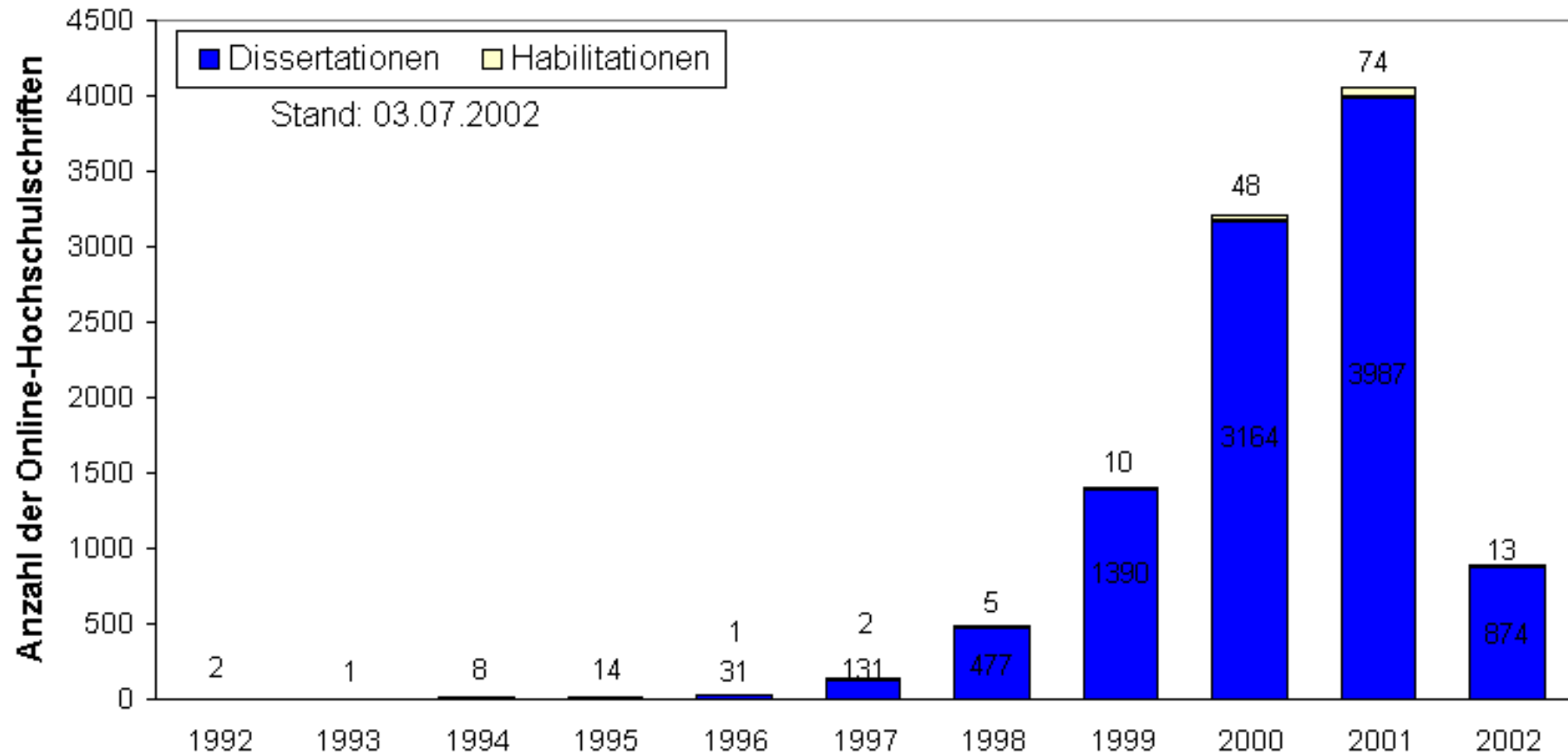
- Ansprechpartner und Kontaktstelle  
<http://www.dissonline.de>
- Anregung zur Kooperation der beteiligten Einrichtungen und Know-How-Transfer
- Koordinierung von Weiterentwicklungen und Anpassungen
- Öffentlichkeitsarbeit / Statistik:  
[http://deposit.ddb.de/netzpub/web\\_online-hochschulschriften\\_stat.htm](http://deposit.ddb.de/netzpub/web_online-hochschulschriften_stat.htm)
- Zusammenarbeit mit nationalen und internationalen Organisationen  
<http://www.hu-berlin.de/etd2003/>

## Anzahl der archivierten Online-Hochschulschriften in DDB





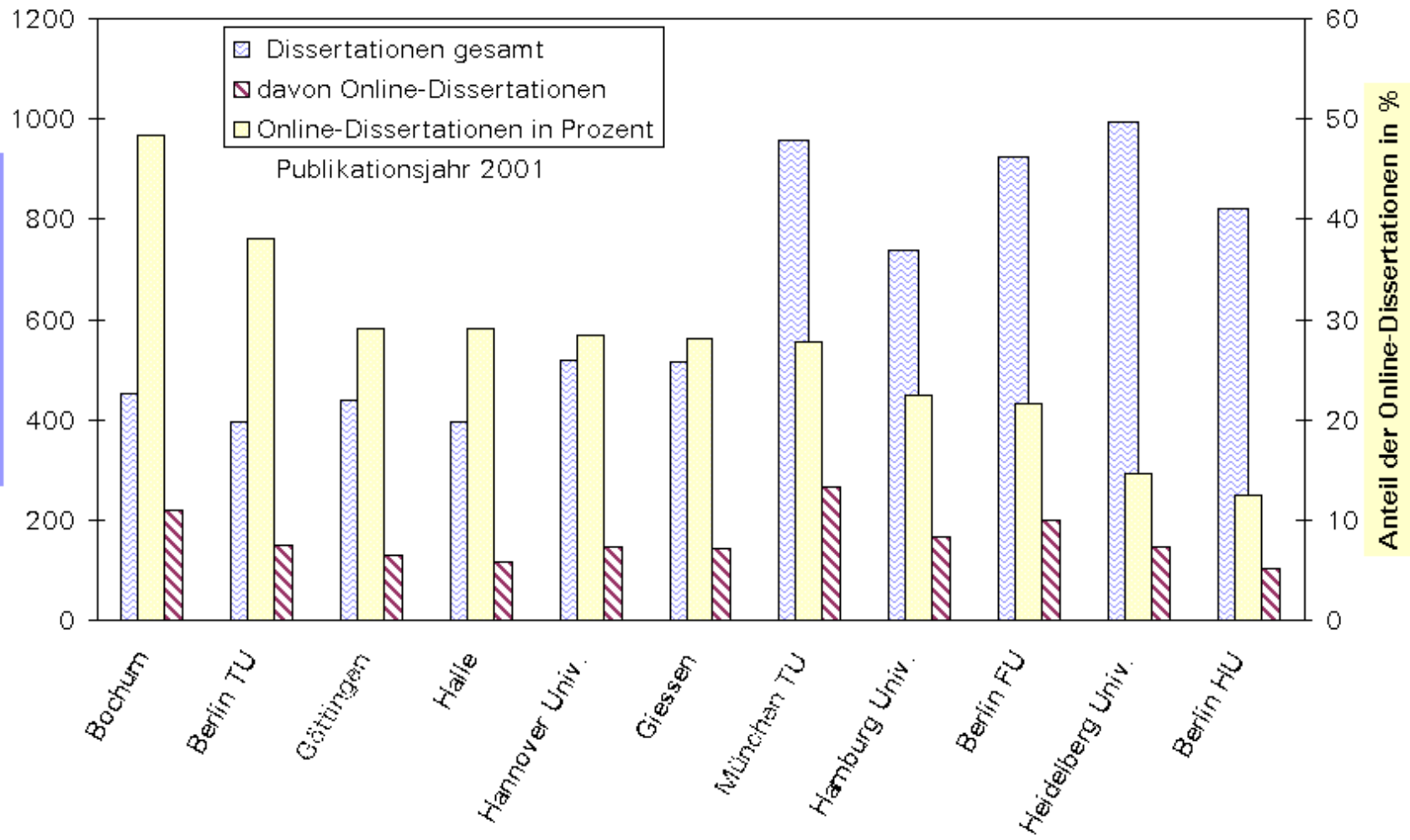
## Erscheinungsjahr der Online-Hochschulschriften



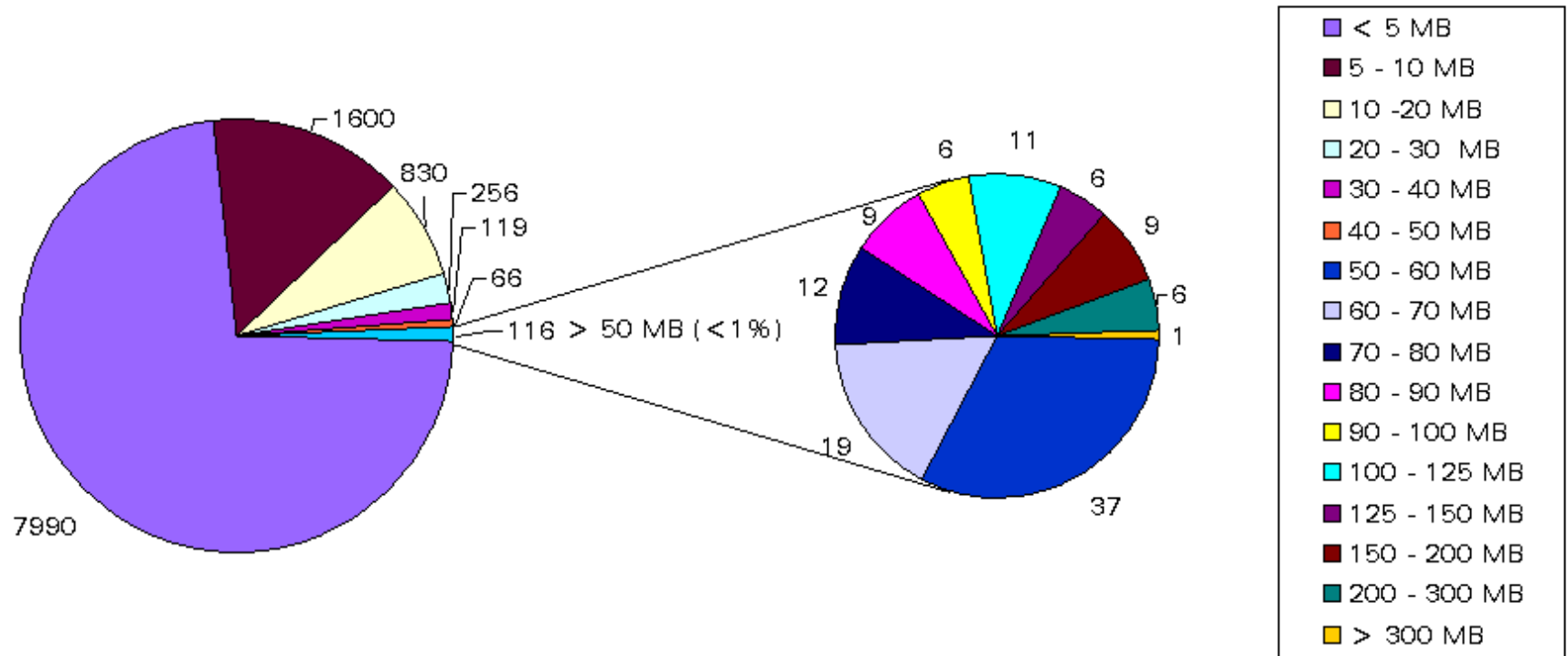




### Dissertationen je Hochschule

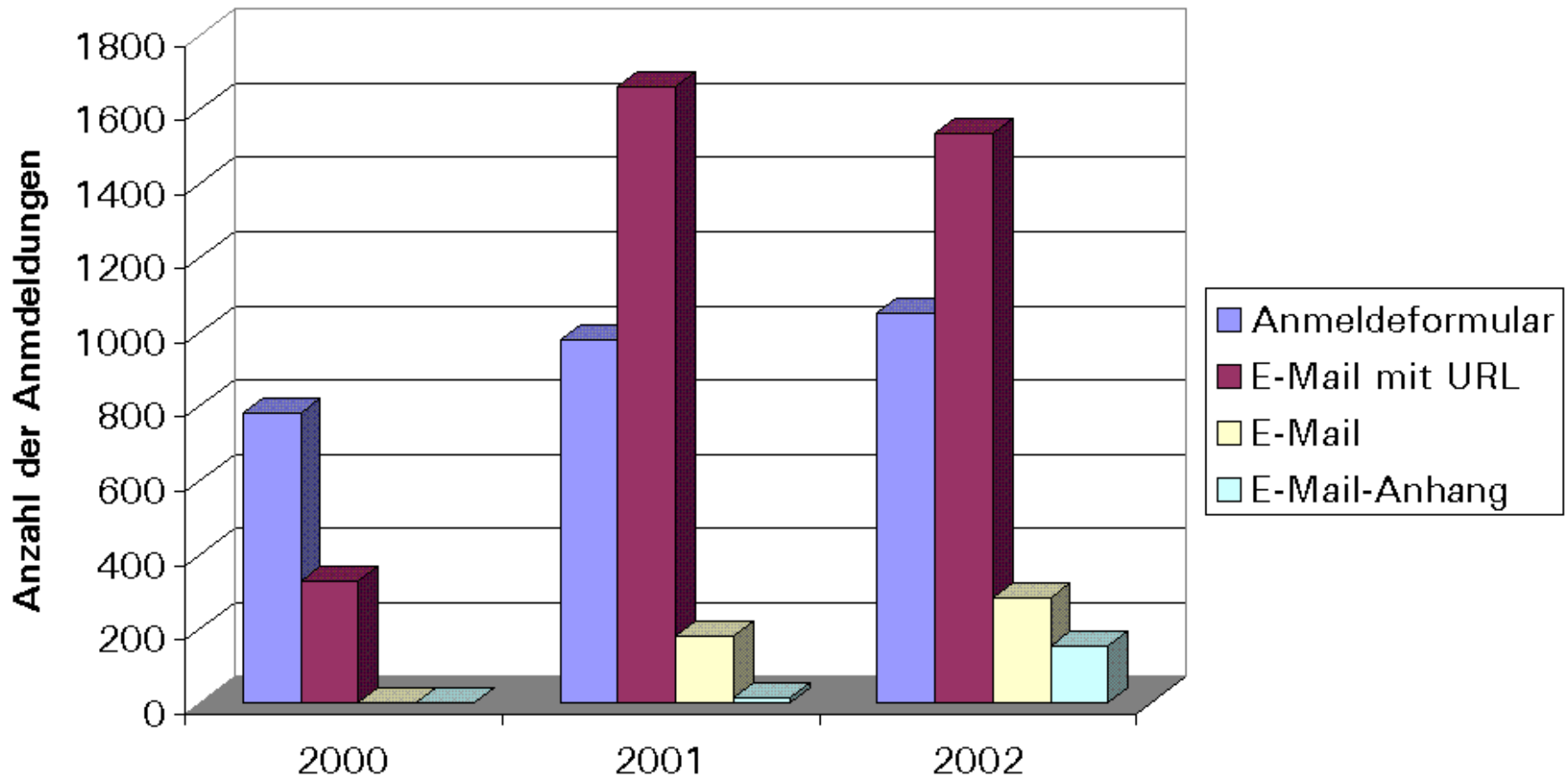


Anzahl der Online-Dissertationen in verschiedenen Dateigrößen 2002  
 gesamt ca. 11000 Dissertationen mit 59,1 GB





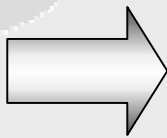
## Art der Anmeldung von Online-Dissertationen





## Persistent Identifier

- Steigende Zahl von elektronischen Publikationen
- Probleme mit dem „Uniform Resource Locator“
  - Veränderung des Standortes
  - Zitierfähigkeit



Verwendung von **dauerhaften und eindeutigen Adressierungsverfahren** für Online-Publikationen



# Persistent-Identifizier URN

- Kontrolle durch „**Internet Engineering Task Force**“ (IETF)
- basiert auf **hierarchischem** Aufbau von Teilbereichen

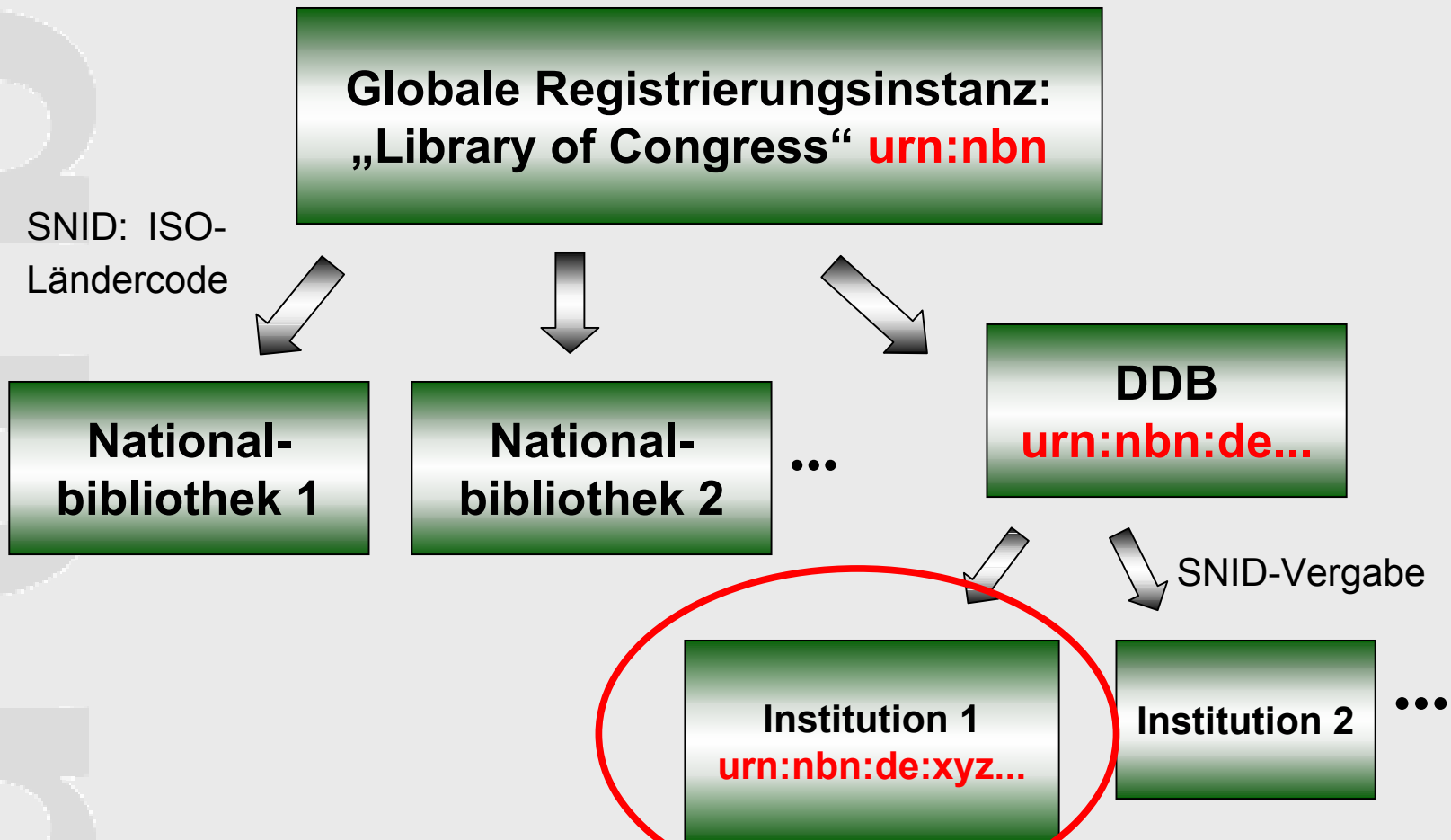
## URN:NID:SNID-NISS

**NID:** Namensraum (Namespace Identifier)  
Beispiele: ISBN, ISSN und **NBN**

**SNID:** Teilbereich aus dem Namensraum  
(Subnamespace oder Sub-Prefix)

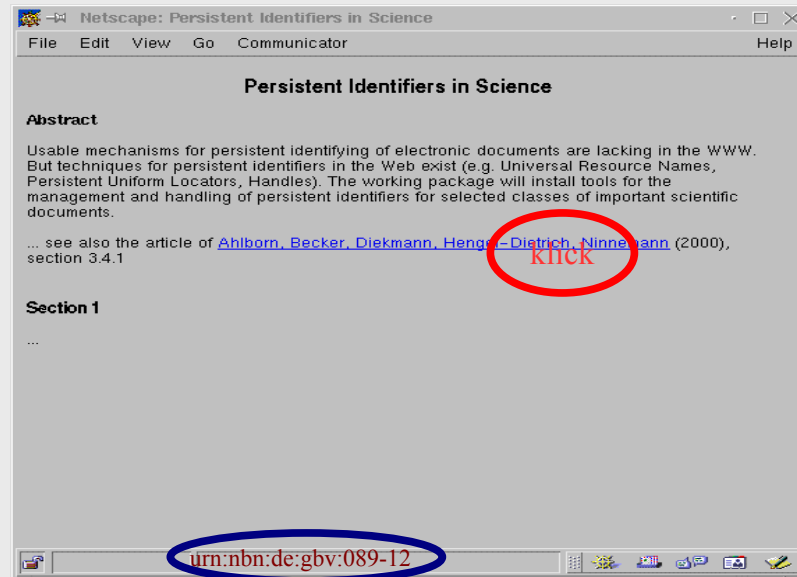
**NISS:** Namespace Specific String

# URN-Namespace „National Bibliography Number“ (NBN)





# Darstellung eines URN-Requests



Darstellung als URN

urn:nbn:de:gbv:089-12

URL-Darstellung

http://nbn-resolving.de/urn/resolver.pl?  
urn=urn:nbn:de:gbv:089-12

... ist universell



# URN-Resolvingdienst von DDB

[English version](#)



Die Deutsche Bibliothek



## Persistent-URL-Resolving (PURL) Der Deutschen Bibliothek im Rahmen des Projektes [CARMEN-AP4](#)

Der "PURL" (Persistent Uniform Resource Locator)-Resolving-Mechanismus leitet ein Persistent-Identifizier-Request an den zugehörigen URL in Form eines HTTP-Redirects weiter. Der Vorteil dieses Resolving-Mechanismus ist seine Browser- und Plattformunabhängigkeit. Die Verwendung des HTTP-Protokolls ermöglicht dem Benutzer, diese Dienstleistung unmittelbar anzuwenden. Weitere Informationen zu dieser Thematik finden Sie [hier](#).

Bitte geben Sie den fehlenden Teil der URN ein:

URN:NBN:DE  URLs ausgeben als:  Liste  Dokument

### Hinweis:

Für die Ausgabe der Ergebnisse der formulargestützten Persistent-Identifizier-Anfrage werden die folgenden zwei Möglichkeiten angeboten.

- Dokument
- Liste

Beim Anklicken des Buttons **Liste** werden alle vorhandenen URLs, die zu einem Persistent Identifier gehören, angezeigt. Wird der Button **Dokument** gewählt, wird anhand eines Rankings eine URL ermittelt. Der Nutzer wird direkt zum entsprechenden Online-Dokument geführt.

Wenn Sie weitergehende Fragen bzw. Informationen benötigen, können Sie uns unter der folgenden E-Mail kontaktieren: [nbn-request@ddb.de](mailto:nbn-request@ddb.de)

[Kathrin Schroeder](#)

Copyright: Die Deutsche Bibliothek.  
Kathrin Schroeder / 11.07.2001

Web-Formulars:

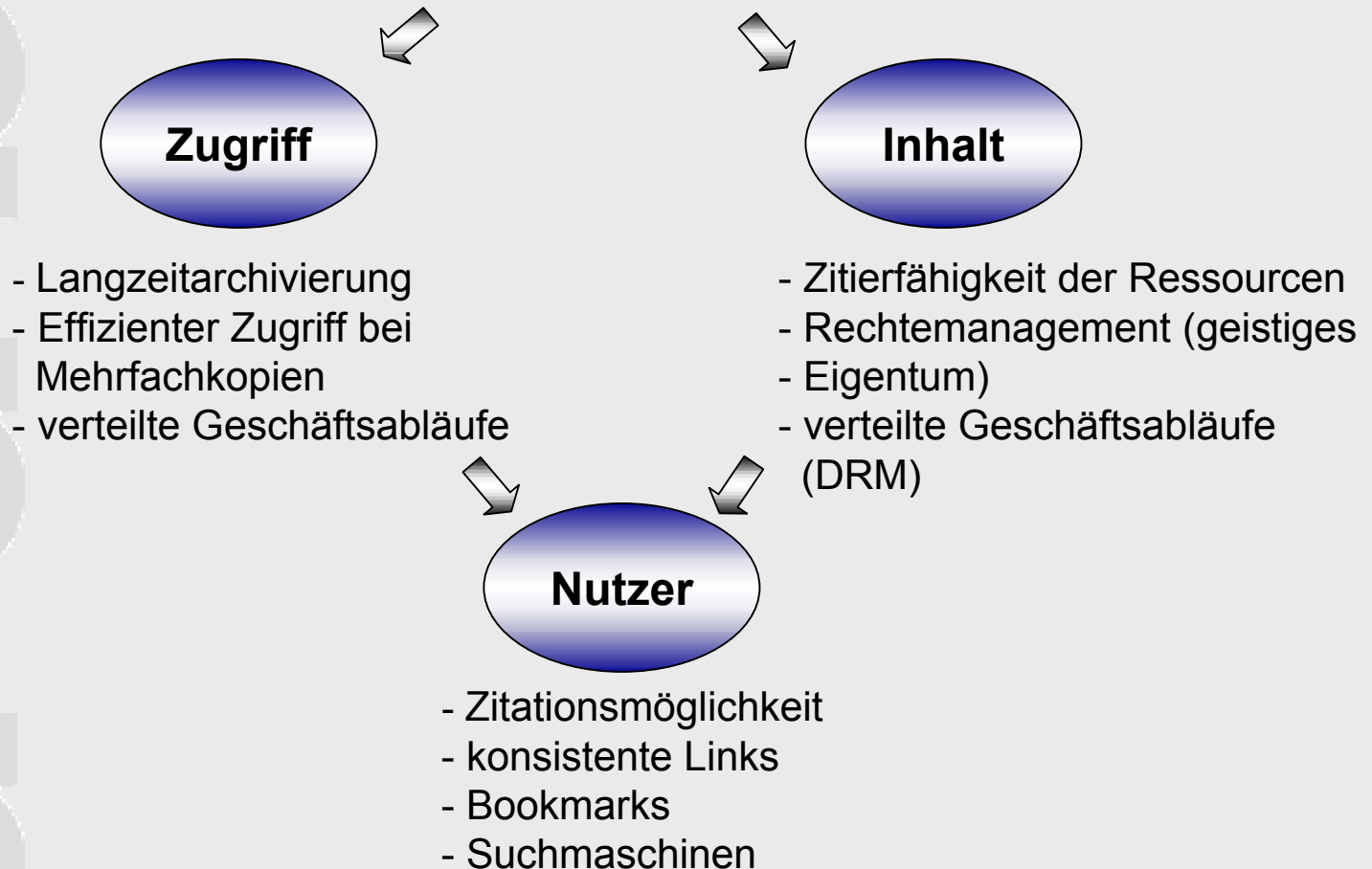
<http://nbn-resolving.de/Resolving.html>

Direkte URN-Eingabe:

<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bsz:93:opus-59>



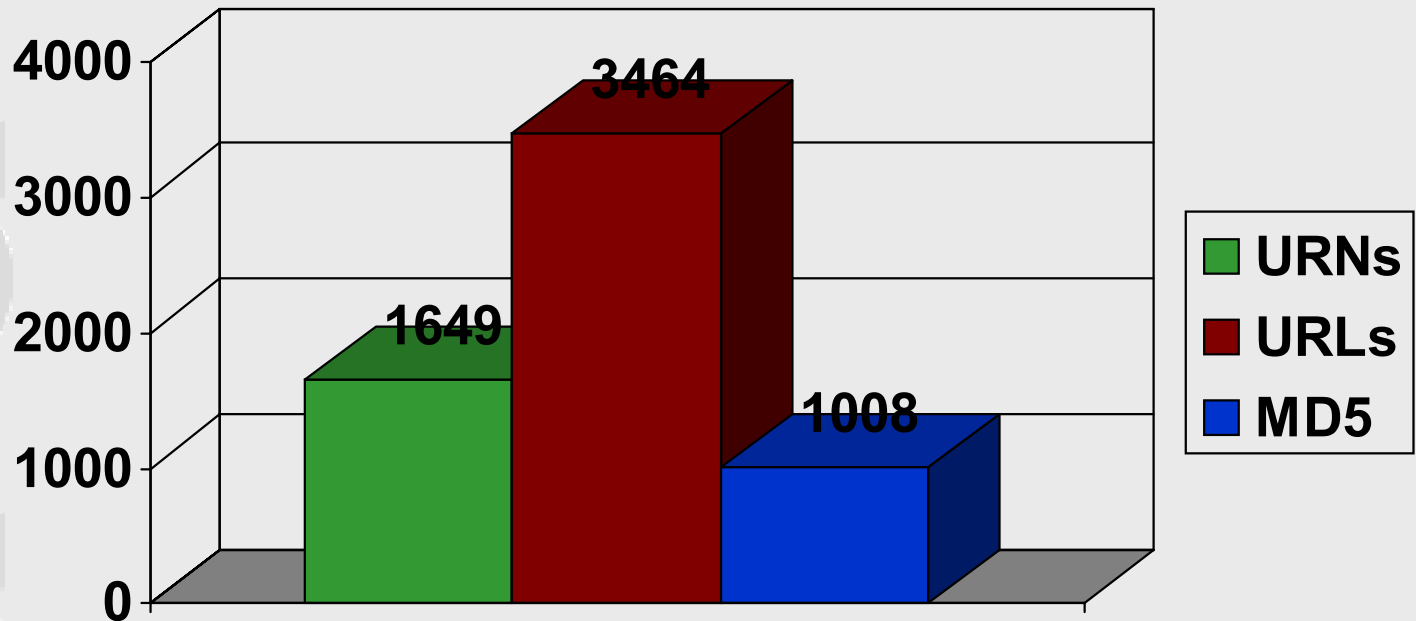
# Persistent-Identifizier – Produkte/Dienstleistungen





# Gesamt-Statistik

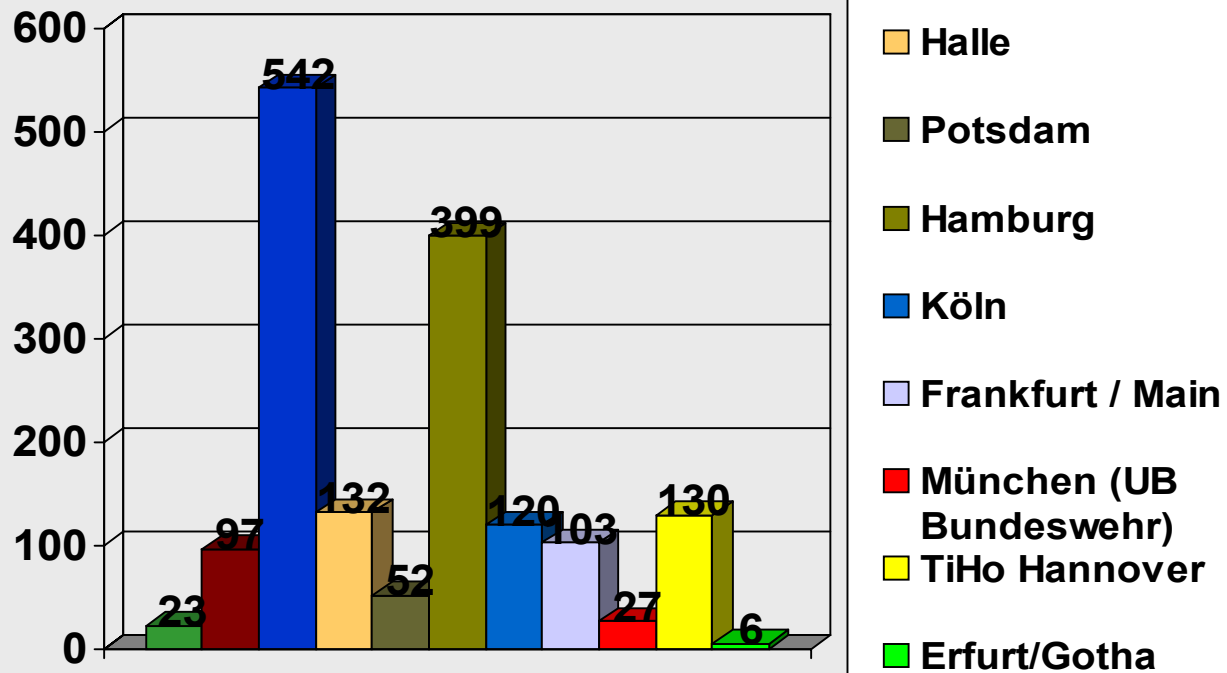
Anzahl URNs, URLs, MD5  
(10.09.2002)



# Aktiv meldende Universitätsbibliotheken

## URNs pro Universitätsbibliothek

(10.09.2002)





# XML-Die Lösung?



# Was ist SGML?

- SGML: Standard Generalized Markup Language – Standard zur Beschreibung von Dokumenten  
ISO 8879 (1986)
- Ziel: logische und strukturelle Auszeichnung
- Meta-Sprache zur Beschreibung von Auszeichnungs- oder Markup-Sprachen für Dokumente
- Hard- und Softwareunabhängigkeit
- Publikation der Daten auf unterschiedlichen Medien

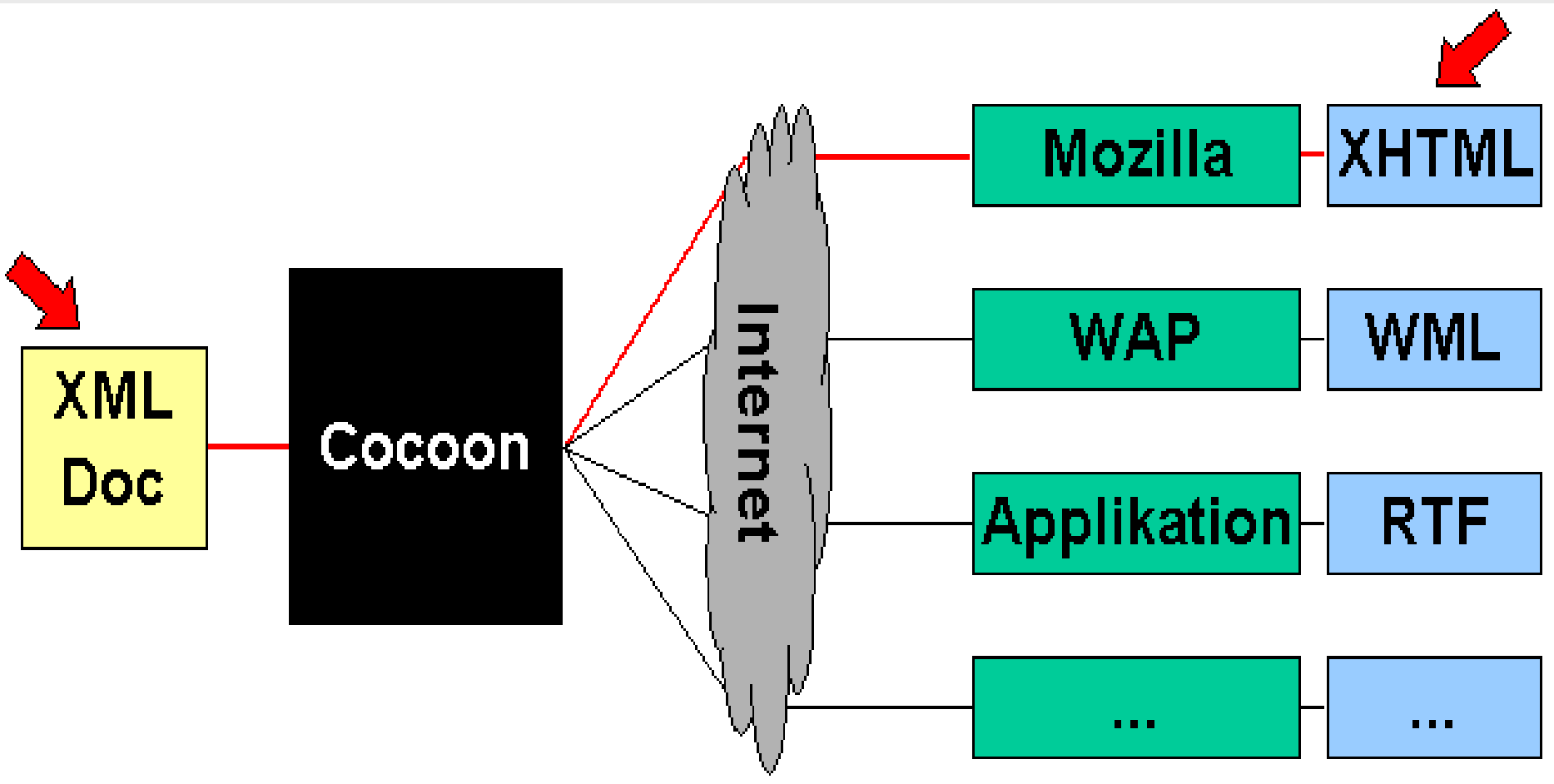


## Was ist XML?

- XML: Extensible Markup Language ist ein Anwendungsprofil bzw. eine eingeschränkte Form von SGML, aber ebenfalls eine Meta-Sprache
- einfache Form von SGML, die über das Web geliefert werden kann
- ersetzt SGML nicht, sondern erweitert (WWW)

## Vorteile:

- Auszeichnung der Inhalte (Struktur), nicht der Erscheinungsform (Layout)
- Unabhängigkeit der Daten von Formaten
- maschinell verarbeitbare Daten (Geschwindigkeit)
- Maschinell prüfbare Daten (Qualität)
- Datenaustausch (SOAP, OAI, EDI)
- unterschiedliche Anwendungen aus einer Quelle







## DTD

- DTD: Document Type Definition
- Festlegung aller Elemente einer Dokumentstruktur und deren Eigenschaften (Attribute) sowie die Beziehungen zwischen den Elementen
- DTD's beschreiben die Struktur eines Dokuments das in SGML verfasst wurde
- XML ist eine Untermenge von SGML; eine DTD kann also die Struktur eines XML-Dokuments definieren



# SGML/XML als Archivformat

- Langzeitarchivierung, da mensch- und maschinenlesbar
- Verfügbarkeit auf verschiedenen Hardwareplattformen und Betriebssystemen
- Internationale Normierung durch ISO8879
- Einfache Transformation in
  - neue Präsentationsformen oder
  - Druckformate
- Integration von Multimediaobjekten
- Automatische Migration möglich



# SGML/XML als Rechercheformat

- Vereinheitlichung von Dokumentstrukturen (DTDs)
- Semantische Kodierung
- Zielgenaue Recherchen
- Automatisierte Katalogisierung möglich



# Historie 1

1987

- erste Dissertations-DTD (ETD) von Juri Rubinsky

1997

- HU-Berlin: SGML als Archivierungsformat
- Anpassung ETD an deutsche Verhältnisse durch Matthias Schulz
- "DIML,, (DissertationMarkupLanguage)
- Entwicklung einer Word-Dokumentvorlage
- Konvertierung Word to SGML mit SGML Author for Word



## Historie 2

1999

- Einbringen der HU-Erfahrungen einschl. Autorenschulungen in das Projekt  
Dissertationen Online

Seit 2000

- Verbreitung von XML/SGML zur Publikation auch im wiss. Bereich



# Strukturiertes Dokument

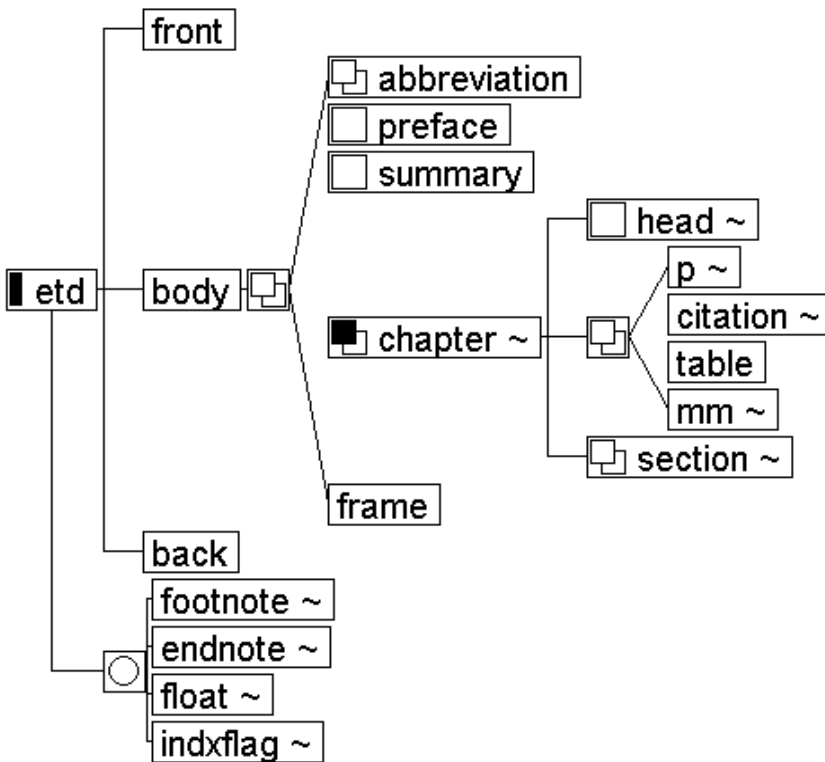
# Was ist ein Dokument?

- **Struktur** DTD: DIML, MATHML, CML, CALS  
– Aufteilung und Abfolge der Information
- **Inhalt** Volltext - XML Instanz  
– zu vermittelnde Information in Schrift, Bild, Ton
- **Layout** Stylesheets: XSL:FO, XSLT, CSS, DSSSL  
– Visualisierung des Inhaltes und der Struktur des Dokumentes
- **Metadaten** Dublin Core, RDF  
– Beschreibung des Dokuments



## DiML-Aufbau

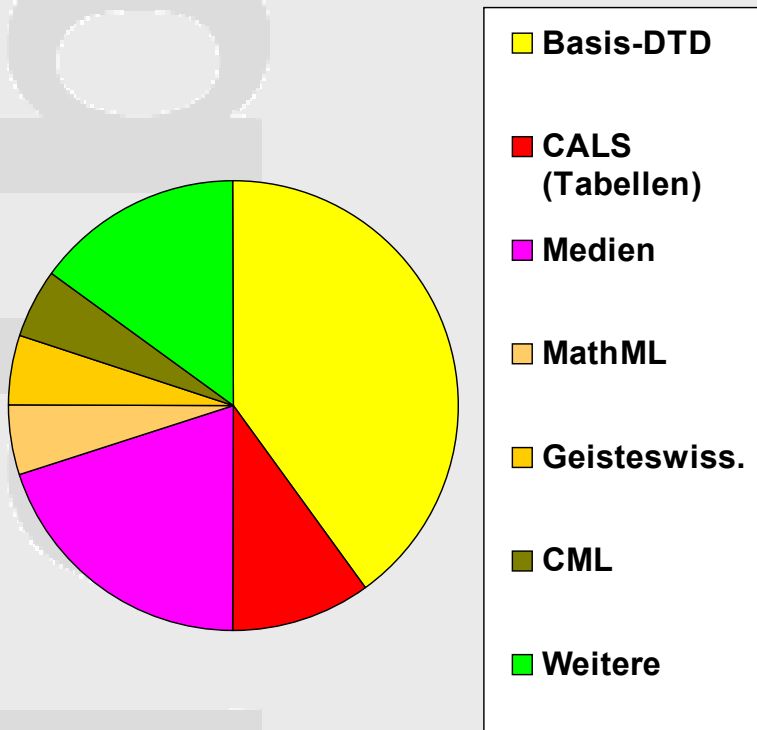
## Die Diml-DTD



- eine SGML-DTD
- DiML-Instanzen sind mit der XML-Spezifikation kompatibel, da nur diejenigen Eigenschaften von SGML verwendet wurden, die auch in XML enthalten sind



# xDiML3.0 "Pizzamodell"



- Basis-DTD
  - Front, Body, Back
  - Titelblatt
  - Bibliographie
- Medien
  - SVG, SMIL, weitere
  - Bitmapformate
- Geisteswiss.
  - Gedichte, Rezension
- Weitere
  - Offene definierte Schnittstelle
  - z.B. TEI-Module, Dublin Core

<b>Partner</b>	<b>HU Berlin</b>	<b>Virginia Tech</b>	<b>Univ. of Iowa</b>	<b>Univ. of Montreal/ Lyon</b>	<b>Tech. Univ. Helsinki</b>	<b>Univ. of Michigan</b>	<b>Univ. Oslo</b>
<b>DTD</b>	<b>DiML</b>	<b>ETD</b>	<b>TDM</b>	<b>TEI-Light</b>	<b>HutPubl</b>	<b>TEI- Light</b>	<b>ISO- Book</b>
<b>Konv. nach SGML/ XML</b>	<b>SGML- Author for Word</b>	<b>SGML- Author</b>	<b>Majix</b>	<b>Omnimark rtf2sgml</b>	<b>Frame- Maker+ SGML</b>	<b>Omni- mark rtf2sgml</b>	<b>Balise</b>
<b>Von SGML nach HTML, PDF</b>	<b>Perl- script, DSSSL</b>	<b>Perl- script</b>	<b>CSS</b>	<b>XSL</b>	<b>DSSSL, Frame- Maker</b>	<b>-</b>	<b>-</b>

siehe <http://dohost.rz.hu-berlin.de/epdiss/dtd-workshop>



# Konvertierung



# XML als Archivformat für Dissertationen

- ... ABER darin schreibt kein Autor

MS Word	75%
LaTeX	22%
WordPerfect	2,6%
FrameMaker und Andere	1,4%
SGML/XML	keine

von 300 eingereichten Arbeiten an der Humboldt-Univ. im Zeitraum von Jan 1998 - Juni 2001



# Voraussetzung zur Konvertierung

## –Anforderungen an die Autoren in der Textverarbeitung Word, LaTeX

- Dokumentvorlage
- Formeln mit Formeleditor
- Literatur mit Reference Manager o.ä.
- Behandlung von Abbildungen, etc. als externe Dateien

## –Schulungsinhalte

## –Konvertierungswerkzeuge

- Problem Qualität der Umsetzung
- Ausgabe in XML muss nach xDiML und nicht nach einer Anbieter-DTD erfolgen



# Word und Formatvorlage

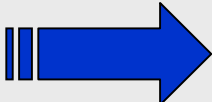
- Text erstellen mit Formatvorlage

[http://edoc.hu-berlin.de/e\\_autoren/vorlage.php?nav=diss](http://edoc.hu-berlin.de/e_autoren/vorlage.php?nav=diss)

- Konvertierung nach XML mit upcast-Konverter



## Probleme: Word to XML

- Ständige Anpassung der Formatvorlage an aktuelle Word-Versionen
- Anpassung der Konvertierung
  - neue Werkzeuge
  - manuelle Nachbearbeitung
- Formeln nur als Image darstellbar  
 Keine Suchmöglichkeit



# Konvertieren von LaTeX nach XML

**Hintergrund**  
**Konvertierungsprogramme**  
**Funktionsweise von TeX4ht**  
**Fazit**

Thomas Pieruschka  
Bibliothek der Universität der Bundeswehr  
München





# Hintergrund

LaTeX ist ein Textsatz-Programm:

- einfach erlernbar
- professionelles Layout
- unübertroffen bei Eingabe mathematischer Formeln
- in Naturwissenschaften/Technik weit verbreitet



# Hintergrund

## Funktionsweise von LaTeX

- TeX-Datei, z.B. test.tex

```
\documentclass{article}
\usepackage{german}
\title{LaTeX to XML}
\author{Mark Tech}
\begin{document}
\section{Hintergrund}
In diesem Teil ....
...
\section{Werkzeuge}
...
\end{document}
```



# Hintergrund

## Funktionsweise von LaTeX

- Aufruf des TeX-Kompilers:

```
latex test.tex
```

- Erzeugen der DVI-Datei, z.B. test.dvi
- Umwandeln der DVI-Datei ins Druckformat, z.B.

Postscript: `dvips test`

PDF: `pdftex test`



# Konvertierungsprogramme

3 Möglichkeiten LaTeX nach XML zu konvertieren:

- Analysieren der TeX-Datei mit Programmen, die in konventioneller Sprache geschrieben sind

**TtH:** C-Programm

(<http://hutchinson.belmont.ma.us/tth>)

**LaTeX2HTML:** Perl-Programm

(<http://cbl.leeds.ac.uk/nikos/tex2html/doc/latex2html/latex2html.html>)



# Konvertierungsprogramme

- Neuprogrammieren des TeX-Kompilers, so dass anstelle einer DVI-Datei XML erzeugt wird

**Omega:** Unterstützung von Unicode  
MathML als Ausgabe möglich

Release 3: mehrere Ausgabeformate  
vorgesehen, u. a. XML, MathML, PDF...

(<http://omega.cse.unsw.edu.au:8080/index.html>)



# Konvertierungsprogramme

- Systeme, die den TeX-Kompiler benutzen um zusätzliche Informationen in die DVI-Datei zu schreiben. Anschließend wird die DVI-Datei von einem weiteren Programm analysiert.

**TeX4ht:** viele Möglichkeiten der Konfiguration, momentan das geeignetste Tool

aber: schwer erlernbar

(<http://www.cis.ohio-state.edu/~gurari/TeX4ht/mn.html>)



# Konvertierungsprogramme

- { 4. Möglichkeit, beruht nicht mehr auf LaTeX:
- Entwicklung einer neuen Markup-Sprache, die von Anfang an zur Erzeugung von XML konzipiert ist

**GELLMU:** Generalized Extensible LaTeX-Like Markup  
(<http://math.albany.edu:8000/math/pers/hammond/igl.html>)

}



# Funktionsweise von TeX4ht

TeX4ht führt seine Umwandlung in 3 Arbeitsschritten durch: Kompilieren der TeX-Datei, Aufruf von tex4ht, Aufruf von t4ht

```
latex test.tex
```

```
latex test.tex
```

```
latex test.tex
```

```
tex4ht test.tex
```

```
t4ht test.tex
```





# Funktionsweise von TeX4ht

## 1. Kompilieren des Dokuments:

Aufgabe: Erzeugen einer DVI-Datei, die die zur Konversion nötigen Informationen enthält

- dreimaliges Ausführen von `latex`: nötig um Verweise korrekt zu setzen
- beim Kompilieren wird Erweiterungspaket `tex4ht.sty` **geladen**

```
\documentclass{article}
\usepackage[xhtml,dim1]{tex4ht}
...
\begin{document}
...
\end{document}
```



# Funktionsweise von TeX4ht

`tex4ht.sty` fügt die Informationen, die zur Konvertierung nach XML benötigt werden, mit Hilfe des `\special{}`-Befehls in die TeX-Datei ein:

`\special{...}`: Argument wird vom TeX-Kompiler nicht verarbeitet, sondern direkt in die DVI-Datei geschrieben



# Funktionsweise von TeX4ht

Um zu bestimmen, wo die `\special{}`-Befehle eingefügt werden, stellt `tex4ht.sty` spezielle Ankerelemente, sog. Hooks, bereit.

Es gibt Hooks für nahezu alle LaTeX-Befehle

Bsp.: Hook für `\section{}`:

Dieser Hook fügt `\special{}`-Befehle zu Beginn und am Ende der `section` ein.

Das Ende der `section` wird durch das Auftreten eines neuen `\section{}`-Befehls oder eines `\end{document}` o.ä. ermittelt.



# Funktionsweise von TeX4ht

Der konkrete Code, den die Hooks einfügen, wird in nicht von `tex4ht.sty` festgelegt, sondern in externen Moduldateien (Dateiendung `.ht`)

Per Default wird immer die Datei `tex4ht.ht` eingelesen, aber auch weitere Module können geladen werden, z.B.

```
\usepackage[xhtml,diml]{tex4ht}
```

Jedes Ausgabeformat (z.B. DocBook, TEI, DiML) sollte eigene `.ht`-Datei besitzen: `diml.ht`



# Funktionsweise von TeX4ht

Die einzelnen Hooks werden in den `.ht`-Dateien über `\Configure{}`-Befehle konfiguriert, z. B.

```
\Configure{section}
  {\Hcode{<section>}}  {\Hcode{</section>}}
  {\Hcode{<head>}}    {\Hcode{</head>}}
```

Der Befehl `\Hcode{ }` fügt HTML-Code in die TeX-Datei ein.



# Funktionsweise von TeX4ht

Mit dieser Konfiguration erzeugt TeX4ht aus folgendem TeX-Code

```
...  
\section{LaTeX nach XML konvertieren}  
Dieser Abschnitt ...  
...  
\section{XML bearbeiten}
```

die XML-Ausgabe:

```
<section>  
  <head>LaTeX nach XML konvertieren</head>  
  Dieser Abschnitt ...  
</section>  
...
```



# Funktionsweise von TeX4ht

## 2. Aufruf von tex4ht

Der 1. Schritt liefert eine DVI-Datei mit `\special{}`-Befehlen. Diese Befehle wertet das Programm `tex4ht` aus, um an den richtigen Stellen die angegebenen Tags zu schreiben und die richtigen Ausgabedateien zu erzeugen.

Ergebnis: Dokument im gewünschten Ausgabeformat und eine verkleinerte DVI-Datei (Endung `.idv`), die den Code enthält, der in Bilder konvertiert werden muss



# Funktionsweise von TeX4ht

## 3. Aufruf von t4ht

Die Erzeugung der Bilder aus der `.idv`-Datei übernimmt `t4ht`.

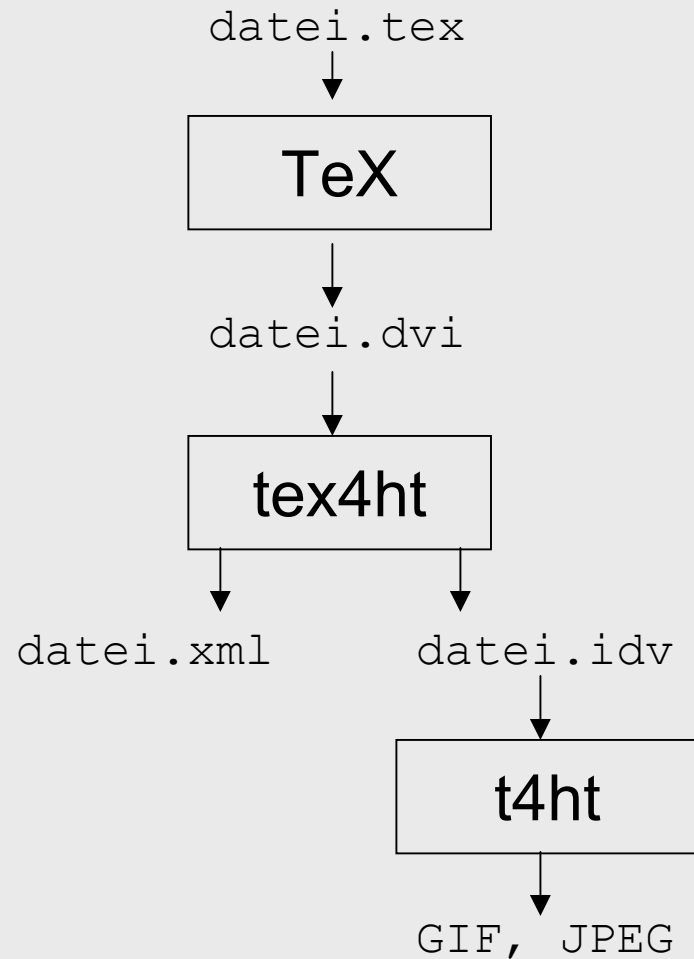
Zu den Bildern gehören neben den echten im TeX-Code enthaltenen auch diejenigen, die als Ersatz für nicht darstellbare Zeichen (z.B. mathematische Symbole) gedacht sind.

`t4ht` konvertiert die Bilder in das richtige Format. Dazu verwendet es Programme wie ImageMagick und Ghostscript.





# Funktionsweise von TeX4ht





## Fazit

Mehrere Möglichkeiten LaTeX nach TeX zu konvertieren, z.B. TtH, Omega, TeX4ht

Zur Zeit TeX4ht das geeignetste Werkzeug, da am flexibelsten konfigurierbar

An UniBwM Diplomarbeit, die TeX4ht benutzt, um LaTeX-Dokumente in die DiML zu konvertieren.



# Probleme der Konvertierung von LaTeX nach XML

**Prinzipielle Probleme**

**Mathematische Formeln**



# Prinzipielle Probleme

- Für LaTeX gibt es eine große Anzahl an Erweiterungspaketen
  - 1 Konvertierungsprogramm an Pakete anpassen? Beschränken welche Pakete die Autoren verwendet dürfen?
- Autoren können TeX-Makros selbst definieren
  - 1 Setzt das Konvertierungsprogramm die Makros richtig um?



# Prinzipielle Probleme

- LaTeX für Layout gedacht, nicht für Markup
  - 1 Nur die Strukturen, für die es LaTeX-Befehle bzw. Umgebungen gibt, sind nach XML konvertierbar

Bsp: In DiML gibt es Tags für Vor- und Zuname des Autors, in LaTeX nicht

```
<author>  
  <given>Mark</given>  
  <surname>Tech</surname>  
</author>
```

3

```
\author{Mark Tech}
```



# Mathematische Formeln

DTD für mathematische Formeln:

**MathML** (<http://www.w3.org/Math/>)

Zwei Arten von MathML:

**Presentation Markup (P.M.):** Auszeichnung der Formel erfolgt Layout-orientiert

**Content Markup (C.M.):** Semantik einer Formel wird beschrieben, z.B. ob es sich beim Ausdruck um eine Variable, Operator oder Zahl handelt



# Mathematische Formeln

Bsp.:  $(a + b)^2$

## P.M.:

```
<mrow>
  <msup>
    <mfenced>
      <mrow>
        <mi>a</mi>
        <mo>+</mo>
        <mi>b</mi>
      </mrow>
    </mfenced>
    <mn>2</mn>
  </msup>
</mrow>
```

## C.M.:

```
<mrow>
  <apply>
    <power/>
    <apply>
      <plus/>
      <ci>a</ci>
      <ci>b</ci>
    </apply>
    <cn>2</cn>
  </apply>
</mrow>
```



# Mathematische Formeln

Da LaTeX Formeln nicht semantisch auszeichnet, kann bei Konvertierung nur P.M. erzeugt werden.

Probleme:

- Exponenten, Basen: wo Beginn, Ende?
- aufgespaltene Formeln
- Schachtelung von Klammern und Begrenzern
- implizite Operatoren, z.B. unsichtbare Multiplikation
- Argumente von Funktionen





## Fazit

- Automatische Konvertierung setzt Verwendung von Formatvorlagen bzw. Standard-Paketen/Makros voraus
- Bisher keine Konvertierung vollständig ohne manuelle Nachbearbeitung möglich
- **Alternative:** Schreiben in XML



# Dokumenterstellung in XML

**XEmacs**

**Epic**

**XMLmetal**



# XML-Editoren

Prinzipiell kann man XML-Dateien mit jedem Editor erzeugen. Manche Editoren bieten jedoch eine spezielle XML-Unterstützung an.

Von den vielen erhältlichen Systemen (<http://www.xmlsoftware.com/editors.html>) werden kurz vorgestellt:

- **XEmacs mit PSGML**
- **Epic Editor**
- **XMetaL**



# XEmacs mit PSGML

XEmacs bzw. Emacs (<http://www.xemacs.org/>) ist „der“ Editor auf Linux-Rechnern. Inzwischen gibt es ihn auch für Windows.

XEmacs ist modular in LISP programmiert und läßt sich deshalb beliebig mit Modulen erweitern.

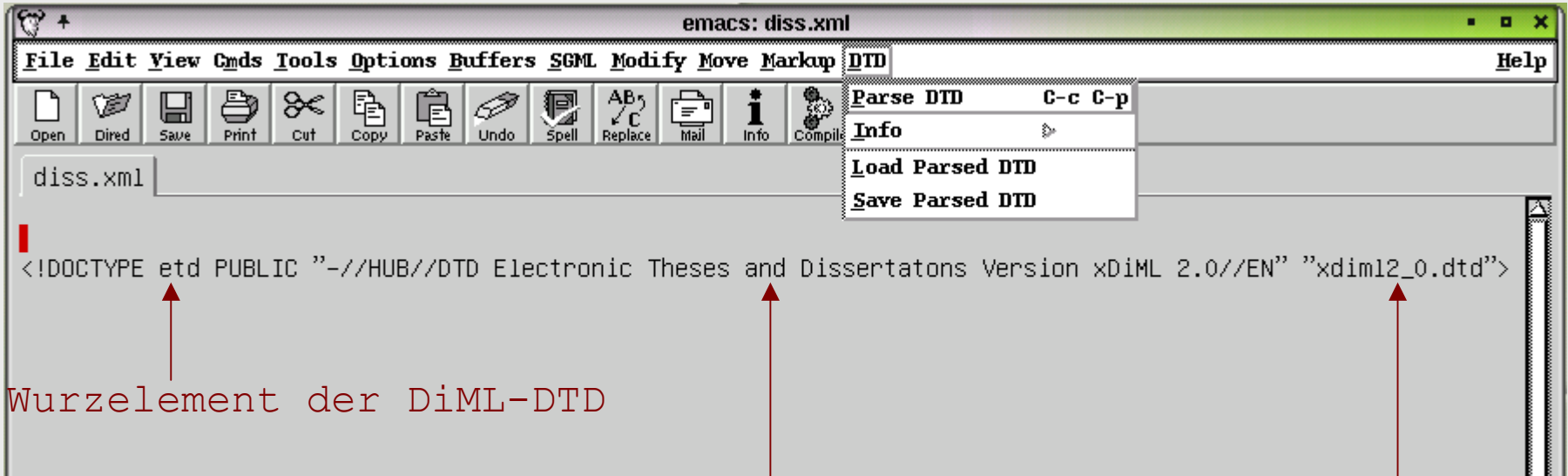
Ein komfortables Modul für die Eingabe von XML ist PSGML.

Bsp: Editieren einer Dissertation `diss.xml` mit der DiML-DTD



# XEmacs mit PSGML

Parsen der DTD:



↑  
Wurzelement der DiML-DTD

↑  
Public Identifier

↑  
Name der DTD-Datei



# XEmacs mit PSGML

Die geparste DTD sollte man zunächst speichern, dann kann mit ihr gearbeitet werden.

The screenshot shows the XEmacs editor window titled 'emacs: diss.xml'. The menu bar includes 'File', 'Edit', 'View', 'Cmds', 'Tools', 'Options', 'Buffers', 'SGML', 'Modify', 'Move', 'Markup', 'DTD', and 'Help'. The toolbar contains icons for Open, Dired, Save, Print, Cut, Copy, Paste, Undo, Spell, Replace, Mail, Info, Compile, Debug, and News. The main text area displays the following XML code:

```
diss.xml
<?xml version="1.0" ?>
<!DOCTYPE etd PUBLIC "-//HUB//DTD Electronic Theses and Dissertations Version x2
DiML 2.0//EN" "xdiml2_0.dtd">
<etd>
  <front>
    <title>Konvertierung von LaTeX nach XML</title>
    <author>
      given
    </author>
  </front>
  <body>
  </body>
  <back>
  </back>
</etd>
```

A popup menu titled 'Element' is open over the 'given' text, listing the following elements: 'given', 'organization', 'suffix', and 'surname'. A red arrow points from the text 'Drücken der rechten Maustaste zeigt, welche Elemente hier eingefügt werden können.' to the popup menu.



# XEmacs mit PSGML

XEmacs bietet auch die Möglichkeit XML-Dokumente zu validieren.

Standardmäßig ist hierzu das Tool `nsgml` (<http://www.jclark.com/sp/index.htm>) vorgesehen. Da dieses für SGML entwickelt wurde, bringt es beim Überprüfen der DTD viele Fehlermeldungen.

Hier wurde daher Xerces-Parser (<http://xml.apache.org/xerces2-j/index.html>) verwendet, der in Java programmiert ist. Er wird durch das Skript `xmltest` aufgerufen.



# XEmacs mit PSGML

## Links ein fehlerhaftes Dokument

```
emacs: diss.xml
File Edit View Cmds Tools Options Buffers SGML Modify Move Markup DTD Help
Open Dired Save Print Cut Copy Paste Undo Spell Replace Mail Info Compile Debug News
*sgml validation*
<!DOCTYPE etd PUBLIC "-//HUB//DTD Electronic Theses and Dissertations Version x
DiMl 2.0//EN" "xdiml2_0.dtd">
<etd>
  <front>
    <title>Konvertierung von LaTeX nach XML</title>
    <author>
      <given>Mark
      <surname>Tech</surname>
    </author>
  </front>
  <body>
  </body>
  <back>
</hack>
IS08--**XEmacs: diss.xml (XML [etd])---7%---
cd /home/thomas/XML/
xmltest diss.xml
Parser-Fehler:
The element type "given" must be terminated by the matching end-tag "</given>".
Zeile : 9 Spalte : 13
SGML validation exited abnormally with code 2 at Fri Sep 13 09:54:32
```

```
emacs: diss.xml
File Edit View Cmds Tools Options Buffers SGML Modify Move Markup DTD Help
Open Dired Save Print Cut Copy Paste Undo Spell Replace Mail Info Compile Debug News
*sgml validation*
<?xml version="1.0" ?>
<!DOCTYPE etd PUBLIC "-//HUB//DTD Electronic Theses and Dissertations Version x
DiMl 2.0//EN" "xdiml2_0.dtd">
<etd>
  <front>
    <title>Konvertierung von LaTeX nach XML</title>
    <author>
      <given>Mark</given>
      <surname>Tech</surname>
    </author>
  </front>
  <body>
  </body>
  <back>
</hack>
IS08----XEmacs: diss.xml (XML [etd])---Top---
cd /home/thomas/XML/
xmltest diss.xml
diss.xml: kein Fehler gefunden!
SGML validation finished at Fri Sep 13 09:51:35
```





# XEmacs mit PSGML

- Open Source
- Unix und Windows
- Unterstützung von DTDs
- Validierung: Nutzung externer Parser (nsgmls oder Xerces)
- Zahlreiche weitere Tools, z.B.  
**TDTD** zur Erstellung von DTDs  
(<http://www.menteith.com/tDTD/>)  
**xslide** für XSLT-Stylesheets  
(<http://sourceforge.net/projects/xslide>)

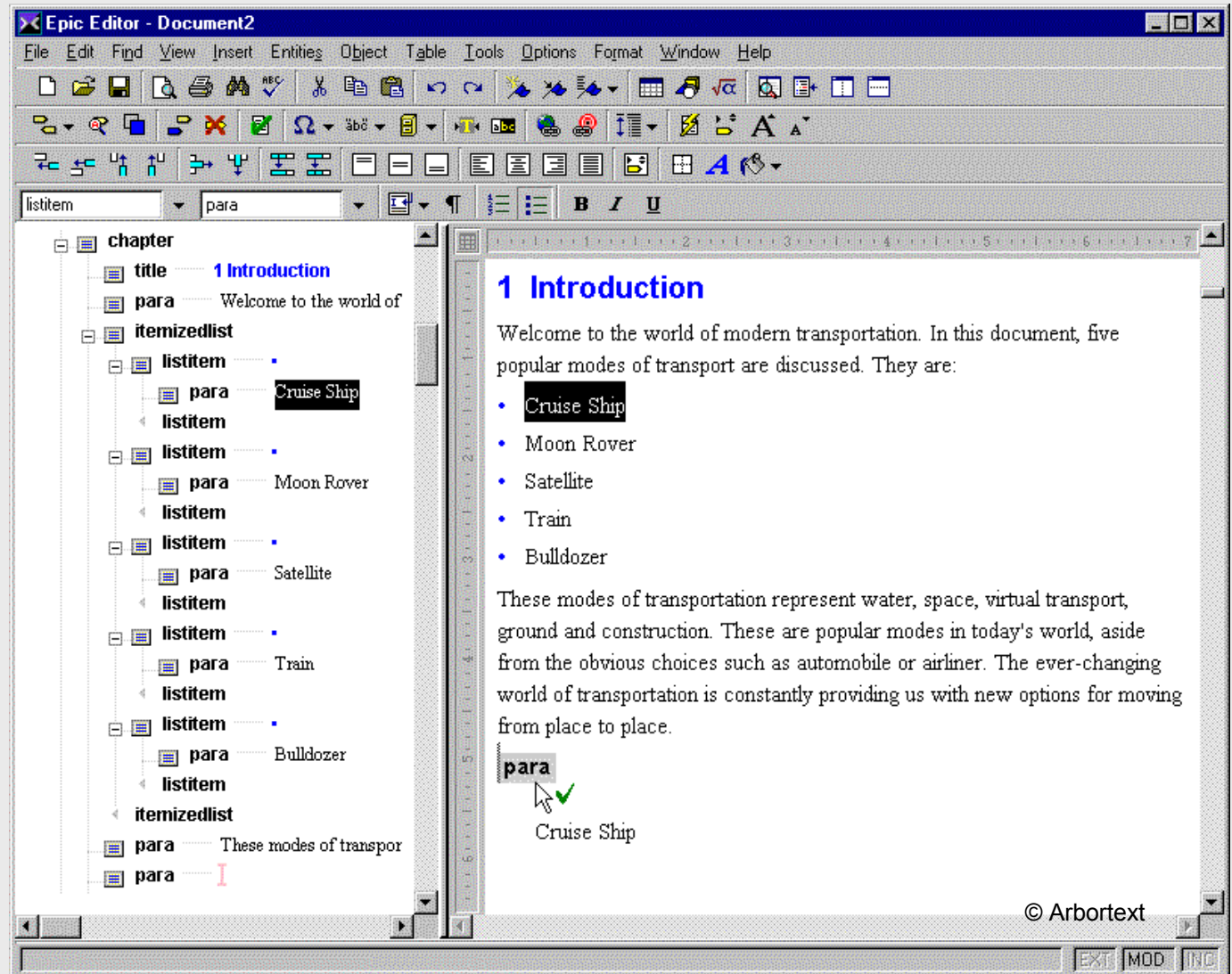


# Epic Editor

- Editor von Arbortext: 10 Jahre SGML-Erfahrung (<http://www.arbortext.com>)
- nur Windows
- Unterstützung von DTDs
- Validiert bei Texterfassung
- WYSIWYG-Ansicht per Stylesheets (XSLT, XSL-FO, FOSI)
- Schnittstellen zu COM (Windows), Java
- Einzellizenz 695\$



# Epic Editor





# XMetaL

- Editor von Softquad (<http://www.sq.com>)
- nur Windows
- Unterstützung von DTDs und als einziger Editor von XML Schema
- Validiert bei Texterfassung
- WYSIWYG per Stylesheets (CSS, XSL, XSLT)
- Schnittstellen zu COM, Java
- Einzellizenz 495\$



# XMetaL

XMetaL - [Zarathustra.xml]

File Edit View Insert Format Tools Table Window Section Track Changes Annotations Help

Title

Article

- Title -> Also sprach Zarathustra
- Sect1
  - Title -> Zarathustras Vorrede
  - Para
    - InlineGraphic
  - TABLE
    - TBODY
      - TR
        - TH -> SwitchboardID
        - TH -> ItemNumber
        - TH -> ItemText
        - TH -> Command
        - TH -> Argument
      - TR
        - TD -> 1
        - TD -> 0
        - TD -> Hauptübersich
        - TD -> 0
        - TD -> Standard
  - Para -> Als Zarathustra dreißig Jahre alt war, verließ er seine Heimat und den See seiner heimat und ging in das

Frederich Nietzsche

SwitchboardID	ItemNumber	ItemText	Command	Argumen
1	0	Hauptübersicht	0	I Standard

Als Zarathustra dreißig Jahre alt war, verließ er seine Heimat und den See seiner heimat und ging in das Gebirge. Hier genoß er seines Geistes und seiner Einsamkeit und wurde dessen **zehn** Jahre nicht müde. Endlich aber verwandelte sich sein herz, - undeines Morgens stand er mit der Morgenröte auf, trat vor die Sonne hin und sprach zu ihr also:

"Du großes Gestirn! Was wäre dein Click, wenn du nicht die

Used All

- Annotation
- Citation
- Deletion
- Emphasis
- InlineGraphic
- Insertion
- Link
- Strong
- Subscript
- Superscript
- TT
- ULink
- Underscore

Change Insert Apply

Citation to a bibliography entry

© Softquad

Rules Checking On NUM



## Fazit

zu Editoren vgl.: Stefan Mintert, Erik Wilde,  
Henning Behme: Auszeichnend, iX 8/2002, S.52-58

- Epic und XMetaL sind die komfortabelsten Editoren auf dem Markt und etwa gleich gut. Sie bieten sich vor allem für Word-Umsteiger an. Nachteil: Hoher Preis
- Auf Unix-Rechner ist Emacs zu Erstellung von LaTeX-Dokumenten weit verbreitet. Autoren, die Emacs kennen, könnten ohne großen Aufwand direkt in XML bzw. DiML schreiben.



# XML unterstützende Schreibwerkzeuge

- verändertes Verfahren beim Schreiben : hohe Kenntnis der Struktur ist vorauszusetzen
- anderes Schulungskonzept ist erforderlich
- entsprechende Software ist in den einzelnen Fächern anzuwenden



# Tools für XML

- **Editoren**
  - XMLNotepad, XMLSpy, XMetal, TurboXML, XMLCooktop, EPCEdit, Merlot, XEMACS mit PSGML ...
- **Textverarbeitungssysteme, DTP-Systeme**
  - Abiword, Staroffice, OpenOffice FrameMaker, QuarkXpress + Avenue, InDesign
- **Konverter**
  - Majix, SCHEMAMarkupKit, XMLWriter, ...
- **Layouter / Styleengines Satzsysteme**
  - Powerpublisher, 3B2





Fragen?



069/1525-1703

[korb@dbf.ddb.de](mailto:korb@dbf.ddb.de)

089/6004-3314

[thomas.pieruschka@UniBw-Muenchen.de](mailto:thomas.pieruschka@UniBw-Muenchen.de)

Wir danken für Ihre Aufmerksamkeit!