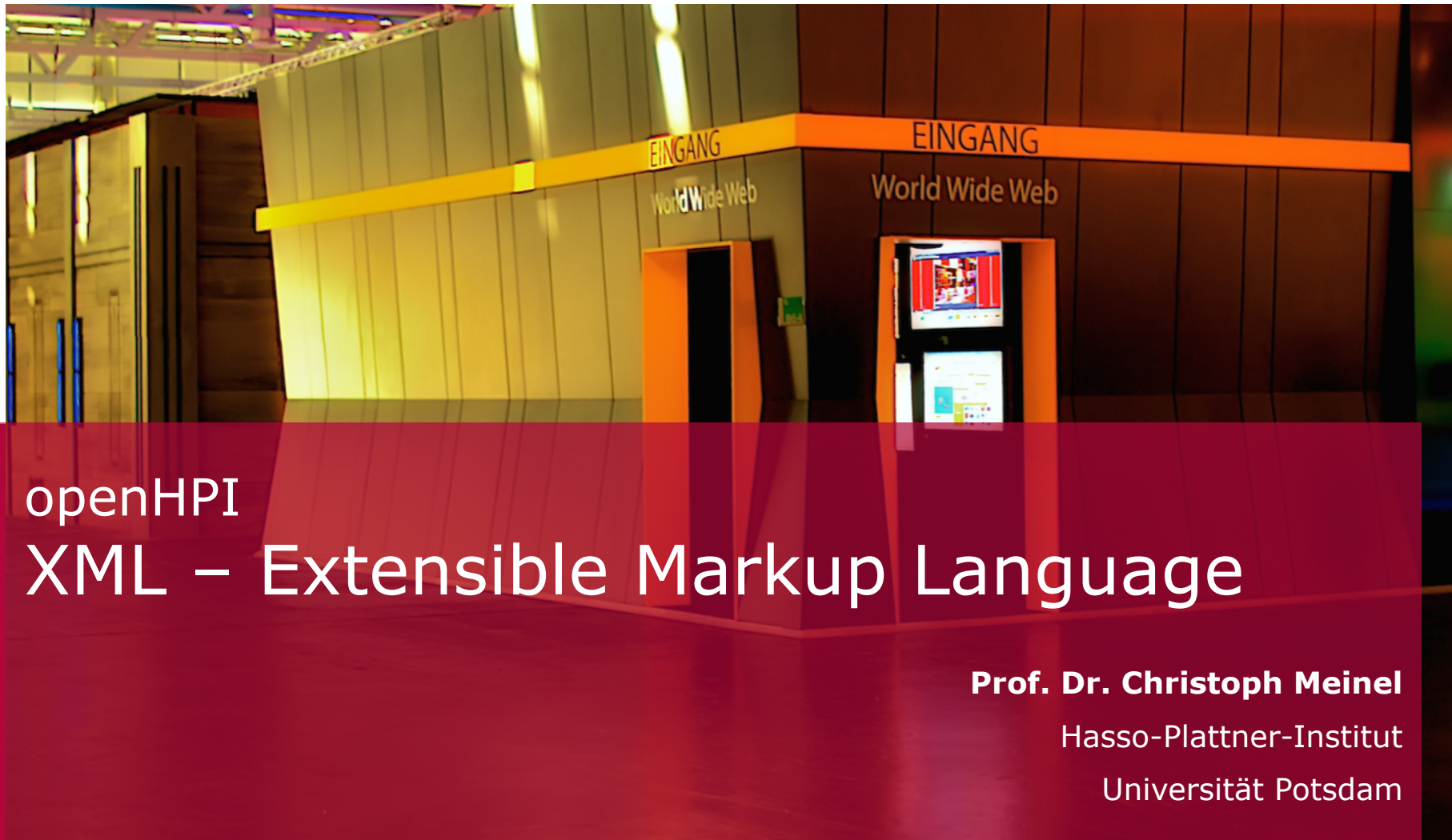




openHPI XML – Extensible Markup Language

Prof. Dr. Christoph Meinel
Hasso-Plattner-Institut
Universität Potsdam

Einführung (1/2)

- Eine große Kritik an HTML war immer ein Mangel an Flexibilität
- Der Grund dafür ist die in der **HTML Document Type Definition** (DTD) festgelegte Syntax von HTML
- Autoren von HTML-Dokumenten können HTML-Elementen keine eigene Bedeutung zuweisen oder selbst neue Elemente definieren
 - ist großes Hindernis für die Interaktion mit externen Anwendungen
- Genau hier kommt die **Extensible Markup Language (XML)** ins Spiel
 - erlaubt Einführung von neuen und aussagekräftigen Elementen durch die Definition von eigenen DTDs

Einführung (2/2)

- Vorlage für die Entwicklung von XML war, genau wie bei HTML, die komplexe Meta-Markupsprache SGML (Standard Generalized Markup Language), mit der jedes Dokument unterschieden und strukturiert werden kann
- Da XML primär für Webanwendungen entwickelt, wurde nur Teilmenge von SGML zur Definition von XML genutzt, um Interpretation von XML-Dokumenten, Autorenwerkzeuge, Browser, usw. so einfach wie möglich zu halten
- In XML wurden nur absolut notwendige Bestimmungen von SGML umgesetzt, um die vollständige Flexibilität von SGML auch für das WWW zu nutzen

XML – Extensible Markup Language (1/2)

- **XML-Spezifikation** beschreibt nur die Struktur von Dokumenten
- **XML-Dokument** das den XML-Erzeugungsregeln [1] entspricht, ist auch syntaktisch korrekt, z.B.:
 - Dokument hat genau ein Wurzelelement
 - Elemente haben Start- und End-Tag
("<xy></xy>", Kurzform "<xy />")
 - ...
- XML-Dokument das der allgemeinen XML-Spezifikation entspricht und die Regeln und Definitionen in der Grammatik einhält, wird **valide** genannt
- Definition der Grammatik kann z.B. durch ein **XML-Schema** oder eine **DTD** festgelegt werden

[1] <http://www.w3.org/TR/REC-xml/#sec-well-formed>

XML – Extensible Markup Language (2/2)

XML

- Ersetzt HTML nicht, aber ergänzt es in Bereichen, wo HTML Schwächen hat, wie z.B. in der semantischen Unterstützung von Interaktionen mit externen Anwendungen
- Allgemeines Objektmodell erlaubt die Entwicklung von Anwendungen mit maschinellem Zugang und zur Manipulation der Daten

➤ **Document Object Model – DOM**

- Bietet offenen Standard zur Beschreibung beliebiger Datenstrukturen, was den Austausch über das WWW vereinfacht und eine integrierte Verbindung zwischen Dokument und Anwendung schafft

XSL – Extensible Stylesheet Language

- XML beschreibt nur die Struktur; die Formatierung eines Dokuments wird durch eine Dokument-Formatbeschreibungssprache in Form von Stylesheets bestimmt ähnlich wie bei HTML
- **Extensible Stylesheet Language – XSL** (eigentlich XSL-FO Formatting Objects) dient zur Beschreibung des Layouts von XML-Dokumenten
- XSL wird oft in enger Verbindung mit **XSL-Transformation – XSLT** genutzt
- Typische Verwendung:
 - XML-Dokument wird umgewandelt durch XSLT in ein XSL-Dokument, um vom FO-Prozessor als lesbares/druckbares Dokument dargestellt werden zu können

DTD – Document Type Definition (1/3)

- XML bietet Autoren Möglichkeit zur maßgeschneiderten Definition ihrer Dokumente mittels eigener Dokumenttypbeschreibung-
Document Type Definitions – DTDs
- Um die **Validität** eines XML-Dokuments überprüfen zu können, wird eine Grammatik benötigt, z.B. definiert als DTD oder XML-Schema
- **XML-Parser** wird benötigt, um XML-Dokumente gemäß zugehöriger DTD zu interpretieren – Deklaration und Regeln der XML-Syntax
- **XML-Schemas** sind komplexer, bieten aber zusätzlich
 - Vererbung,
 - XML-Parsing,
 - ausführlichere Beschreibungen, ...

DTD – Document Type Definition (2/3)

Beispiel für eine DTD – Zeitung

```
<!DOCTYPE newspaper [  
    <!ELEMENT newspaper (article+)>  
    <!ELEMENT article (title, introduction, text)>  
    <!ELEMENT title (#PCDATA)>  
    <!ELEMENT introduction (#PCDATA)>  
    <!ELEMENT text (#PCDATA)>  
  
    <!ATTLIST article author CDATA #REQUIRED>  
    <!ATTLIST article lecturer CDATA #IMPLIED>  
    <!ATTLIST article date CDATA #IMPLIED>  
  
    <!ENTITY newspaper "New York Times">  
    <!ENTITY publisher "Arthur Ochs Sulzberger Jr.">  
]>
```

Hinweis: Entitäten erlauben die Definition von Konstanten
(Sonderzeichen oder häufig verwendete Textteile)

DTD – Document Type Definition (3/3)

Beispiel eines XML-Dokuments entsprechend der newspaper.dtd

```
<?xml version="1.0" ?>
<!DOCTYPE newspaper SYSTEM "../newspaper.dtd">
<newspaper>
  <article author="Hasso Plattner" date="2015-01-07">
    <title>User Oriented</title>
    <introduction>Introduction text ...</introduction>
    <text>Main article text</text>
  </article>
  <article author="Christoph Meinel" date="2015-01-09">
    <title>New Student Welcome</title>
    <introduction>Introduction text ...</introduction>
    <text>Main article text</text>
  </article>
</newspaper>
```

Ist valides XML-Dokument in Bezug auf newspaper.dtd

XML – Verknüpfungskonzept (1/2)

Erinnerung: HTML Verknüpfungskonzept mit dem Hyperlink-Element `<a>` ist sehr eingeschränkt:

- HTML-Hyperlink verlinkt nur zwei Ressourcen
- HTML-Hyperlink ist immer unidirektional
- HTML-Hyperlink ist Teil der HTML-Form wo der Link herkommt
- HTML-Hyperlink beeinflusst nicht die Darstellung des Link-Ziels im Browser

XML – Verknüpfungskonzept (2/2)

Für XML wurden zwei separate Sprachstandards entwickelt, um HTML-Verknüpfungskonzept zu verallgemeinern und für die Definition von weiteren Verbindungen zwischen Informationsressourcen:

- **XML Linking Language – Xlink**

- definiert auch multidirektionale Verbindungen

- **XML Pointer Language – XPointer**

- erweitert die XPath-Spezifikation um auf Teile von XML-Dokumenten zu verweisen
- Beispiel:
 - `xlink:href="cds.xml#xpointer(/ABBA/Waterloo)"`

Bereichsspezifische Beschreibungssprachen

Mit dem XML-Metasprachkonzept zur Definition eigener DTD ist der Weg für Entwicklung eigener bereichsspezifischer Beschreibungssprachen frei, z.B.:

- Mathematical Markup Language – **MathML**
- Chemical Markup Language – **CML**
- Synchronized Multimedia Integration Language – **SMIL**
- Scalable Vector Graphics – **SVG**
- ...

Da XML mehr ein Werkzeug für Spezialisten ist, kommt der Endnutzer nur in Kontakt mit XML bei verschiedenen Anwendungen oder mit XHTML als XML-konformem HTML