DAIMLERCHRYSLER

Tutorial: Die Extensible Markup Language (XML)

Mario Jeckle
DaimlerChrysler Forschungszentrum Ulm
mario.jeckle@daimlerchrysler.com
mario@jeckle.de
www.jeckle.de

Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - Strukturelle Grundkonzepte
  - Namensräume
  - Dokument-Typ-Definitionen (DTD)
  - XML-Schemasprachen
  - XML und das Web
- XML-Standards und -Anwendungen der zweiten Generation ...
- Anwendungen der XML im praktischen Einsatz ...

Gliederung

- Dokumente und Daten ...
- XML-Standards und -Anwendungen der zweiten Generation ...
  - (Dokument-)Verknüpfungen: XML Links
  - Die Lokatorsprache XPath
  - Erzeugung von Präsentationssichten: XML Stylesheets
  - Transformation von XML-Dokumenten: XSL Transformations
  - Die Anfragesprache XQuery
- Anwendungen der XML im praktischen Einsatz ...

Gliederung

- Dokumente und Daten ...
- XML-Standards und -Anwendungen der zweiten Generation ...
- Anwendungen der XML im praktischen Einsatz ...
  - Die Simple API for XML (SAX)
  - Das Document Object Model (DOM)
  - Nahtlose Integration von XML und Hochsprache: XML Data Binding
  - Persistenz: XML und Datenbanken
  - XML-basierter Nachrichtenaustausch und entfernte Methodenaufrufe: XML Protokolle

Gliederung

- Dokumente und Daten ...
- **Einführung und Überblick**
  - Strukturelle Grundkonzepte
  - Namensräume
  - Dokument-Typ-Definitionen (DTD)
  - XML-Schema
  - XML und das Web

Dokumente und Daten ... Einführung und Überblick

Das Datenformat [XML] erleichtert den Informationsaustausch zwischen vernetzten Computern

[DER SPIEGEL, 2000-06-23]

XML schickt sich an in die Fußstapfen von HTML zu treten

[c't 10/2000, p. 200]

Streit um Programmiersprache XML

[F.A.Z., 2001-07-26]

Das XML-Format, [...] das richtige Werkzeug zur Herstellung eigener Webinhalte

[DER SPIEGEL, 2000-06-22]

Sinnliche Suchmaschine [...] existierende Systeme wie [...] XML

[DER SPIEGEL, 2000-06-07]

Alle Dokumente sind gleich

[SZ, 1999-02-16]

Die Extended Markup Language für eCommerce

[F.A.Z.]

Nachfolger für ungeliebte Cookies
[...] Enge Verbindung von Java mit XML, [...] Erweiterung des HTML-Standards

[DER SPIEGEL, 1999-10-0...]

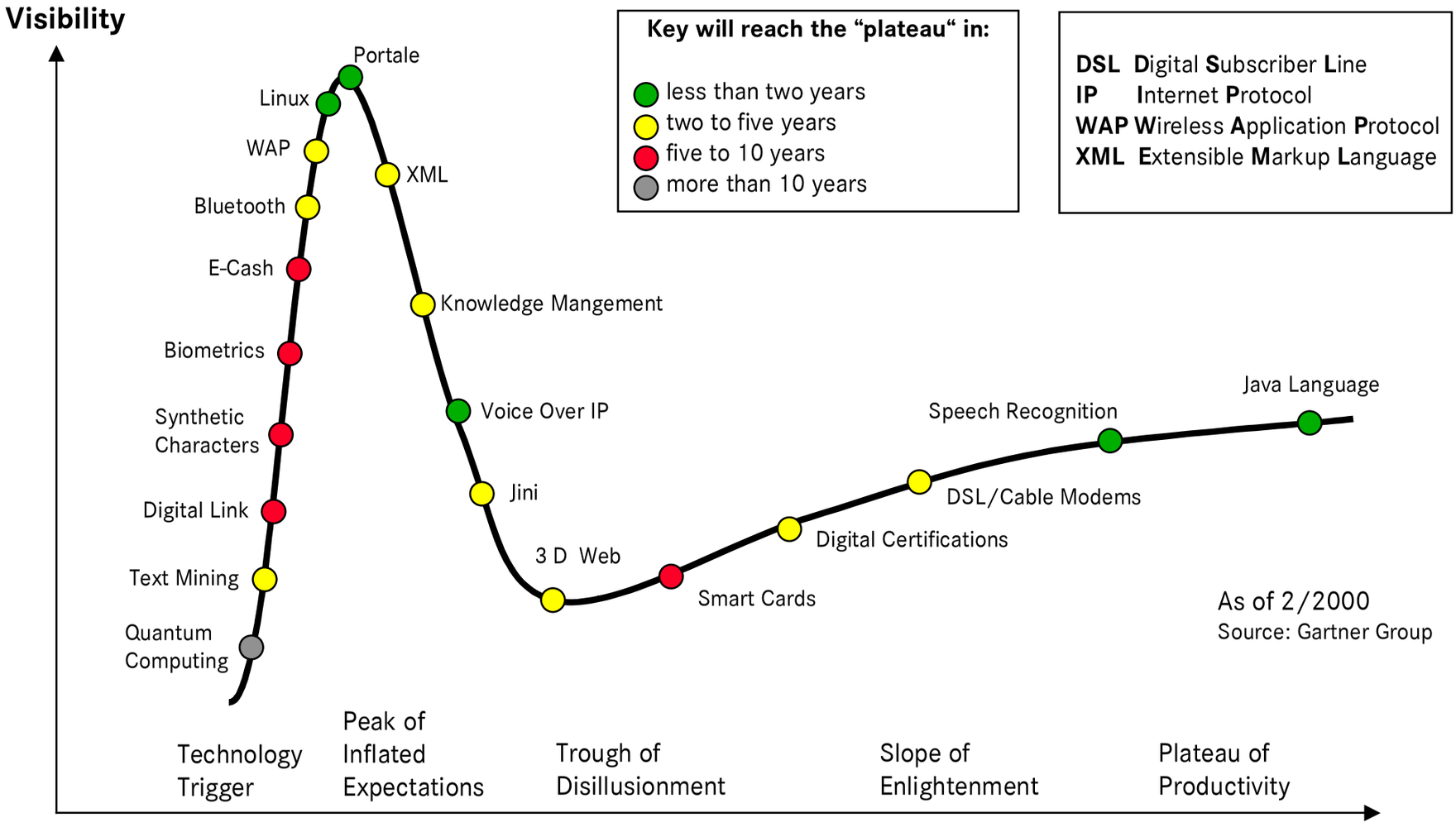
Sortieren statt Stottern
Programmiersprache HTML stößt an ihre Grenzen
XML ist kommender Code im Netz

[Süddeutsche Zeitung, 2000-01-11]

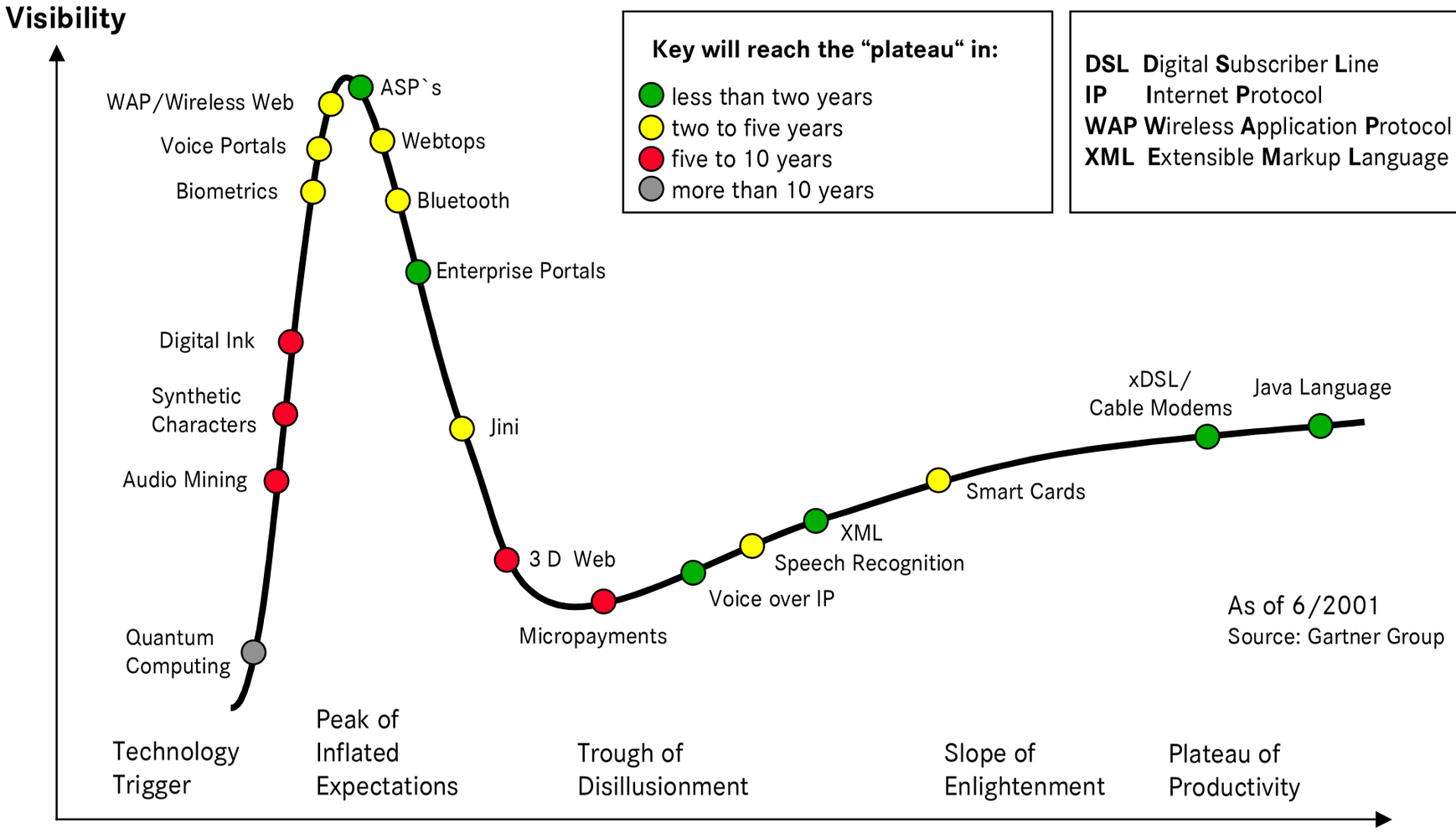
Ein digitales Esperanto für das Internet

[Die Welt, 2000-10-07]

Dokumente und Daten ... Einführung und Überblick



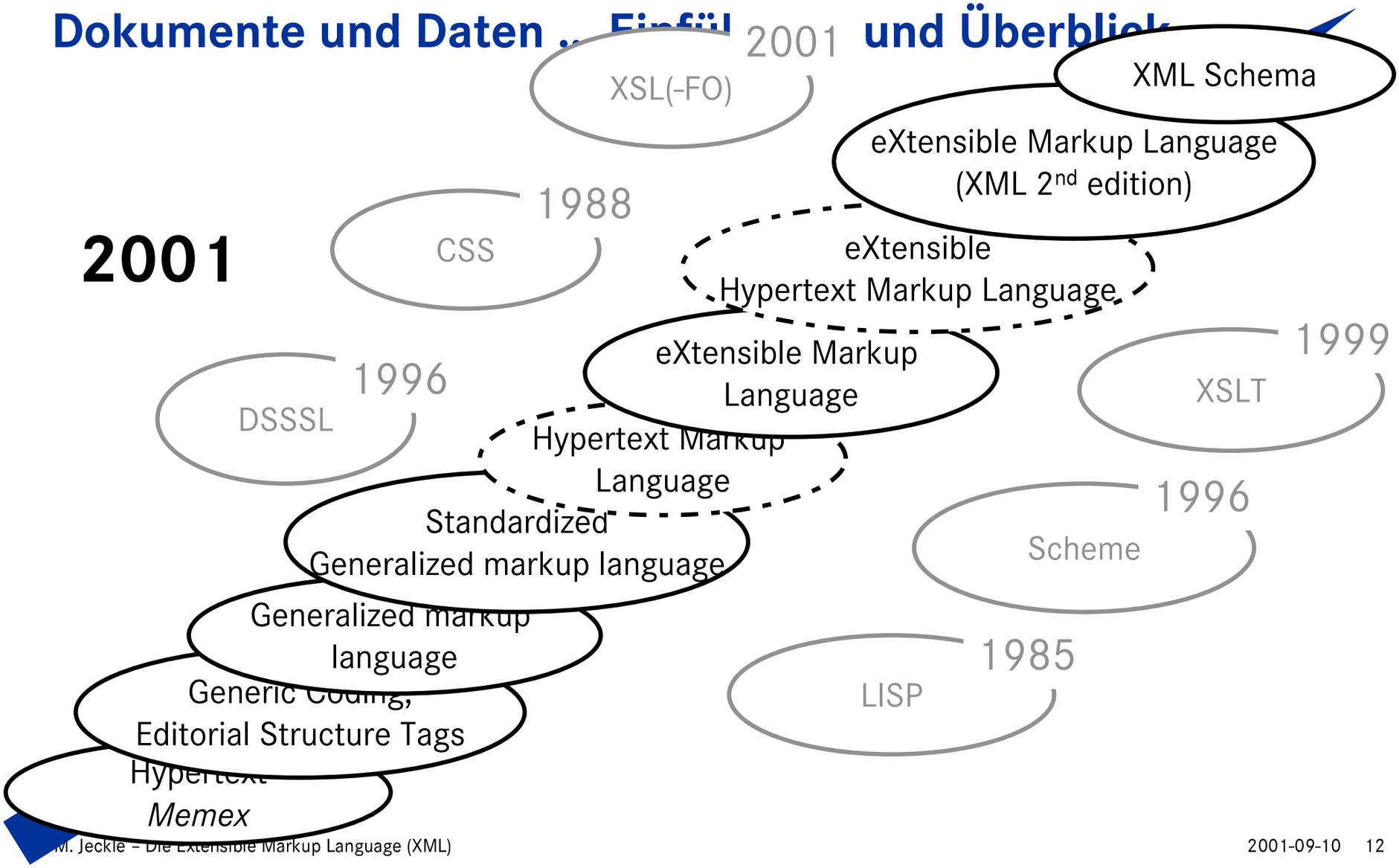
Dokumente und Daten ... Einführung und Überblick



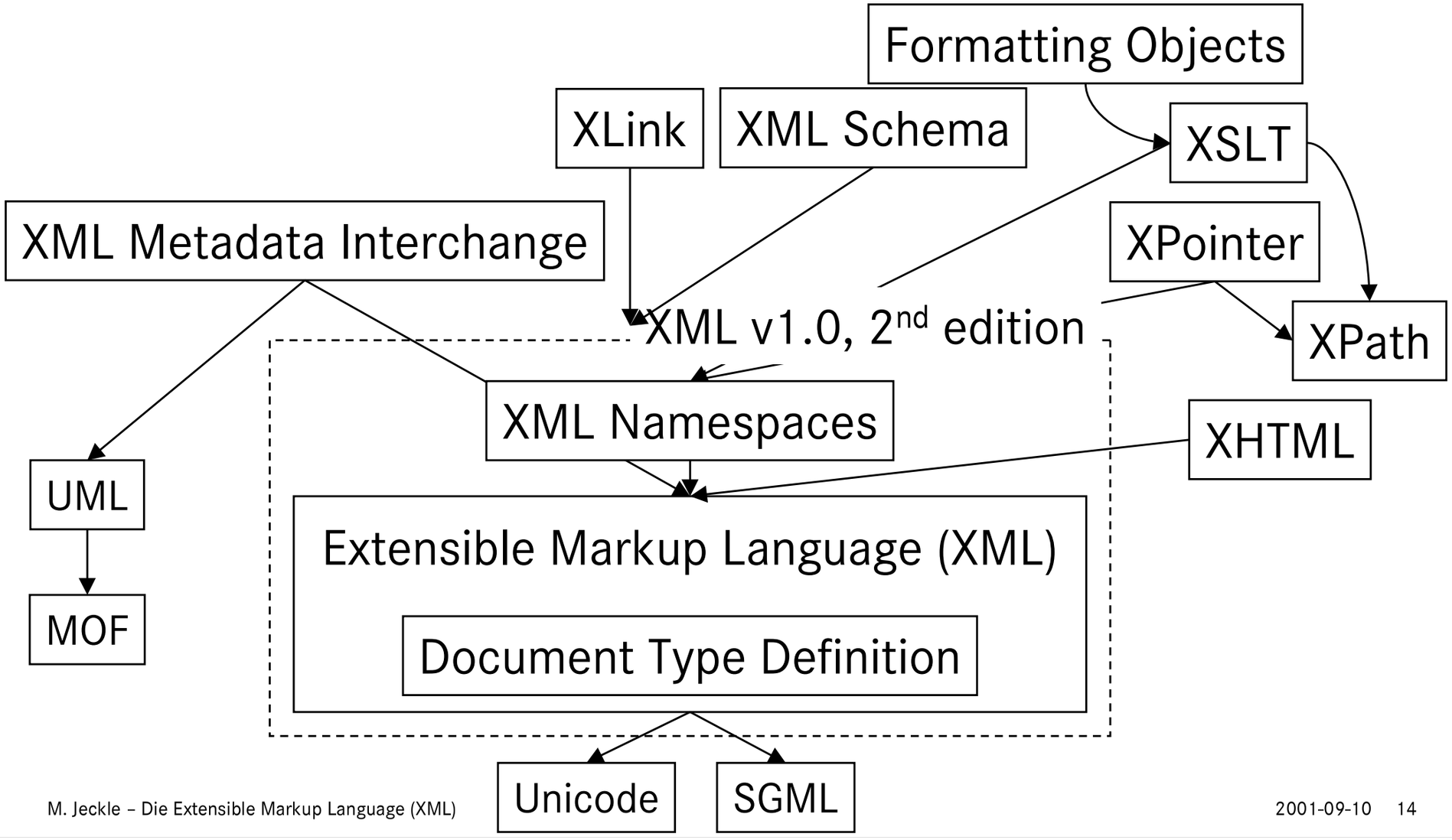
Dokumente und Daten ... Einführung und Überblick

- XML ist eine Methode, um strukturierte Daten in einer Textdatei darzustellen
- XML ist Text, aber nicht zum Lesen
- XML ist ausführlich, was aber kein Problem darstellt
- XML sieht fast aus wie HTML, ist aber kein HTML
- XML ist eine Familie von Techniken
- XML ist eine Untermenge des ISO-Standards SGML
- XML ist international und unterstützt beliebige Alphabete (Unicode)
- XML ist neu, aber nicht so neu
- XML ist lizenzfrei, plattform- und herstellerunabhängig, und gut unterstützt
- XML ist inzwischen weit verbreitet

Dokumente und Daten .. Einfüh[illegible] und Überblick

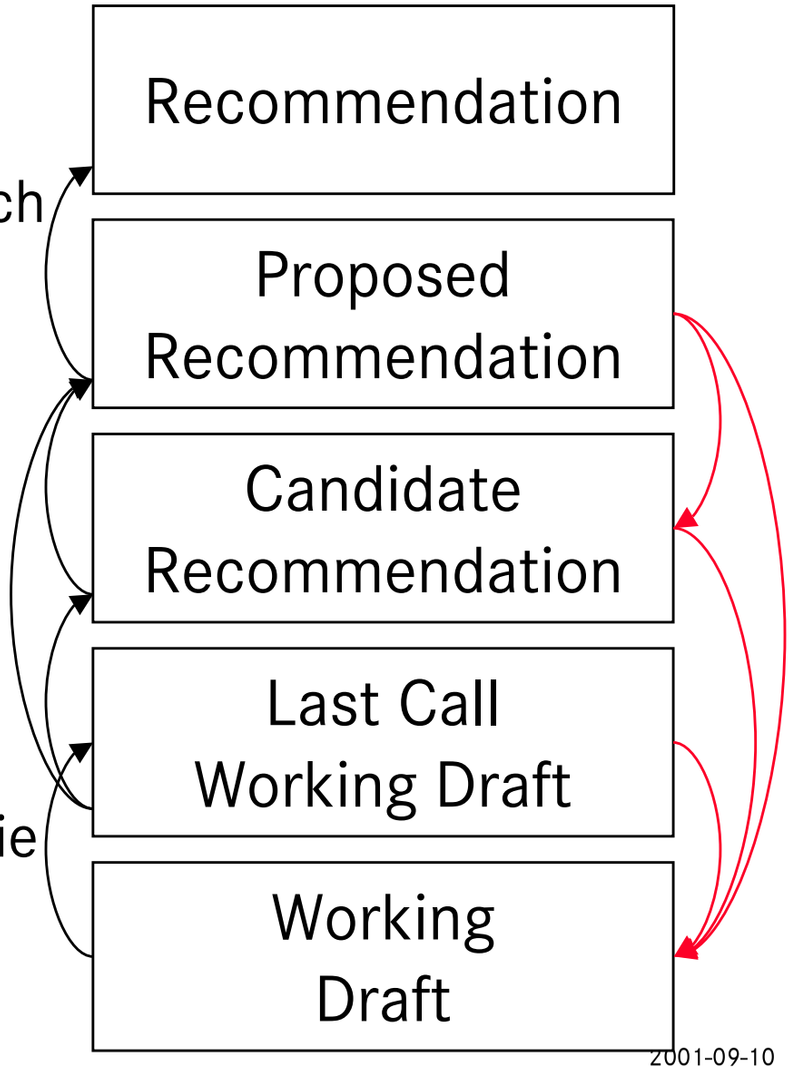


Dokumente und Daten ... Einführung und Überblick



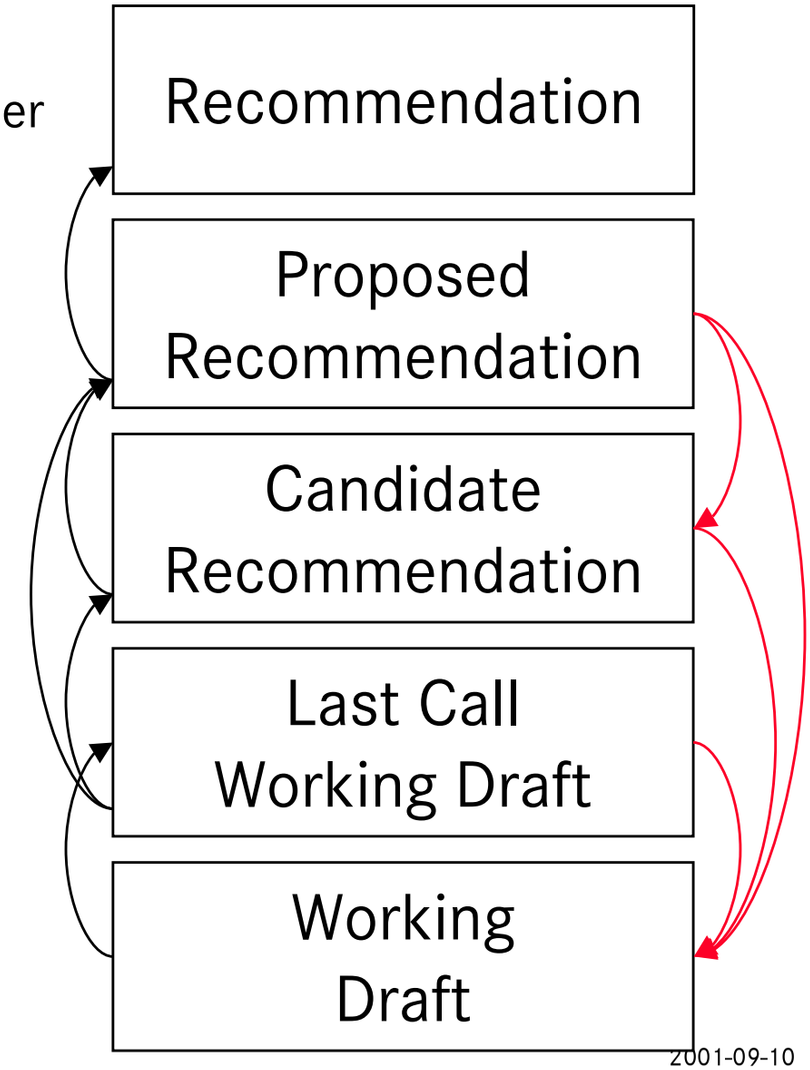
Dokumente und Daten ... Einführung und Überblick

- **Note:** Allgemeines datiertes Dokument zur Vorstellung einer Idee oder neuen Technik; es kann sich auch um einen formalen Kommentar handeln.
- **Working Draft:** Interimsstand einer W3C-Arbeitsgruppe zu einem Thema das durch das W3C weiter bearbeitet wird.
- **Last Call Working Draft:** Markiert die Erreichung der im Anforderungs-dokument festgelegten Ziele.



Dokumente und Daten ... Einführung und Überblick

- **Candidate Recommendation:**
  Direktor des W3C bestätigt die Erreichung der definierten Anforderungen, bzw. erklärt sich mit der Nichterreichung bestimmter Anforderungen einverstanden.
- **Proposed Recommendation (PR):**
  Bestandteile der PR sind umgesetzt. Idealerweise sollte die Arbeitsgruppe zwei interoperable Implementierungen jedes Spezifikationsbestandteils vorweisen können.
- **Recommendation:** Direktor erklärt sich mit dem Grade der Unterstützung durch das Advisory Committee einverstanden und erklärt PR zum Recommendation und damit zum offiziellen W3C-Standard.



Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - **Strukturelle Grundkonzepte**
  - Namensräume
  - Dokument-Typ-Definitionen (DTD)
  - XML-Schema
  - XML und das Web

Dokumente und Daten ... Strukturelle Grundkonzepte

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

- Ein **XML-Dokument** ist ein Datenstrom (der nicht zwingend als Datei vorliegen muß), welcher den Strukturierungsprinzipien der Extensible Markup Language genügt.

- Ein **XML-Prozessor** ist eine maschinelle Komponente (typischerweise: Software), die zum Lesen eines XML-Dokuments eingesetzt wird.

  Er erlaubt Zugriff auf den Inhalt und die Struktur des XML-Dokuments.

Dokumente und Daten ... Strukturelle Grundkonzepte

- Ein XML-Dokument enthält (informal):
  - (optional) die XML-Kopfzeile
  - Wurzelelement
  - Inhaltselemente
  - Attribute
  - Namensräume
  - textuelle Inhalte
  - Kommentare
  - Notationen
  - Entitäten
  - CDATA-Bereiche

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

- *Wohlgeformtes XML-Dokument*:

  Ein textartiges Objekt dessen Inhalt folgenden Anforderungen genügt:

  - Das XML-Dokument nutzt eine DTD,
    oder enthält die Deklaration standalone="yes"
  - Zu jedem Start-Tag existiert ein Ende-Tag.
    Bei leeren Elementen können diese zu einem Tag zusammenfallen.
  - Korrekte Elementschachtelung,
    d.h. Elemente überlappen einander nicht.
  - Genau ein Wurzelelement.
  - Alle Attributwerte sind in einfachen oder doppelten Anführungszeichen.
  - Kein Element enthält zwei oder mehr Attribute desselben Namens.
  - Keine Kommentare oder Processing Instructions innerhalb von Tags.
  - Kommentare beginnen und enden mit genau zwei Bindestrichen.
  - Die Sonderzeichen < und & treten nicht innerhalb von Elementeninhalten oder Attributwerten auf.

Dokumente und Daten ... Strukturelle Grundkonzepte

- **XML Information Set** definiert die strukturellen Grundprimitive eines XML-Dokuments.
  Er bildet das *Metamodell der Extensible Markup Language*
- Ein XML-Dokument enthält (formal):
  - Document Information Item
  - Element Information Item
  - Attribute Information Item
  - Character Information Item
  - Namespace Information Item
  - Comment Information Item
  - ...

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

Document Information Item

character encoding scheme = ISO-8859-1
base URI =
standalone =
version = 1.0

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

Document Information Item

character encoding scheme = ISO-8859-1
base URI =
standalone =
version = 1.0

Element Information Item

namespace name =
local name = Vortrag
prefix =
base URI =

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

Document Information Item

character encoding scheme = ISO-8859-1
base URI =
standalone =
version = 1.0

Element Information Item

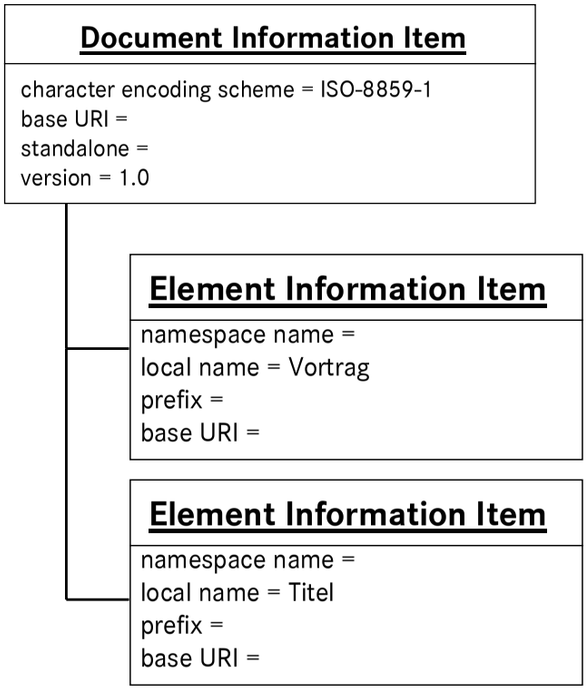
namespace name =
local name = Vortrag
prefix =
base URI =

Element Information Item

namespace name =
local name = Titel
prefix =
base URI =

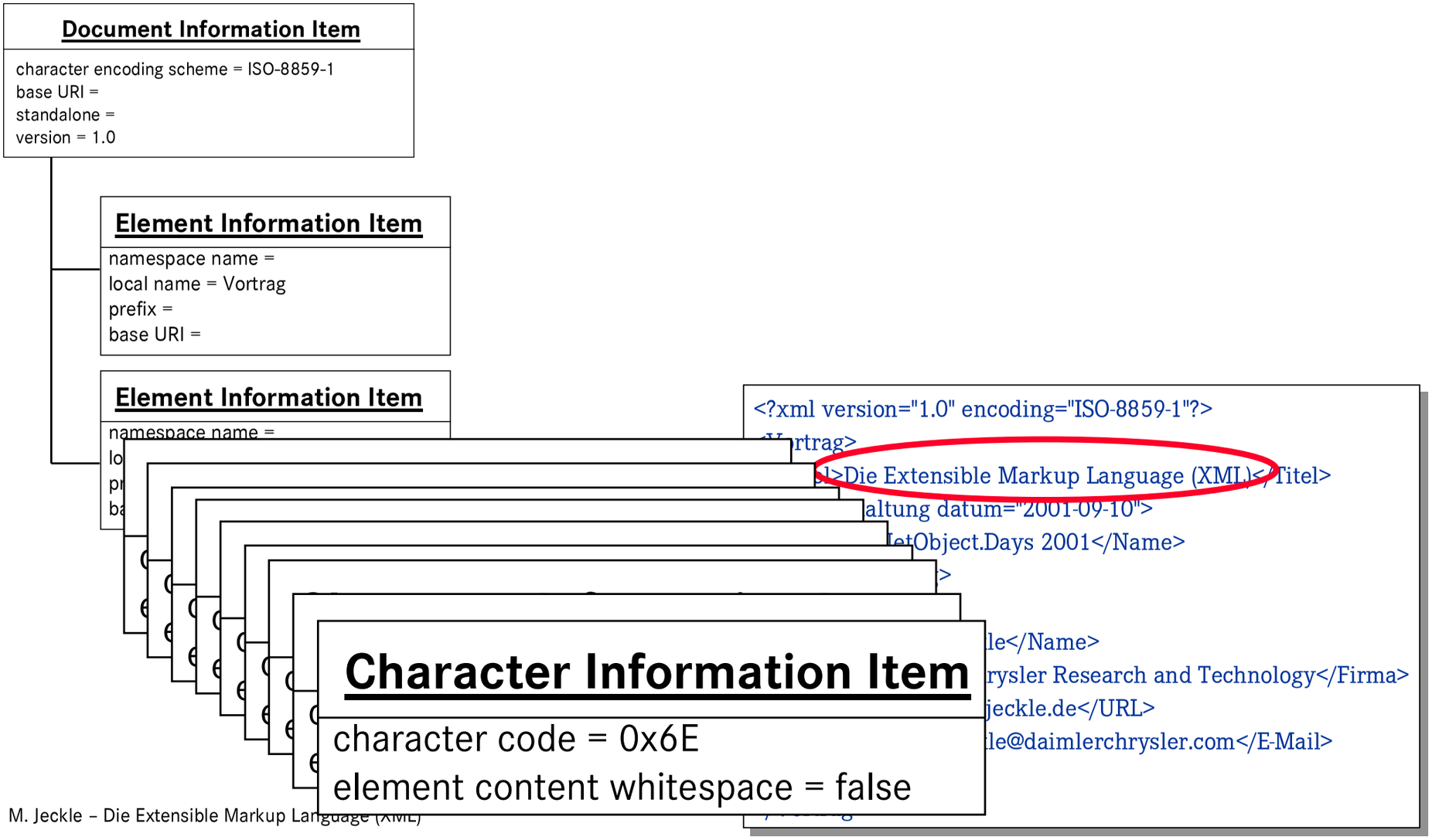
```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte




```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte



Dokumente und Daten ... Strukturelle Grundkonzepte

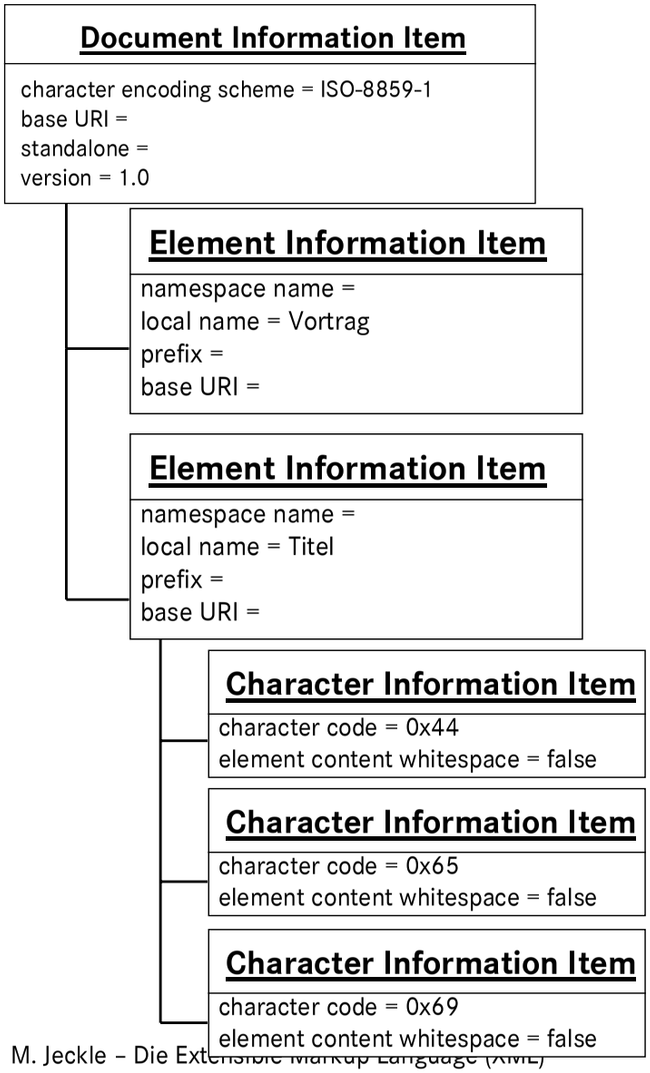
- Exkurs: *XML und Unicode*

```
<?xml version="1.0" encoding="Shift_JIS" standalone="yes"?>
<kougi>
  <title begin="二千一年三月二十一日三時三十分＋九時">選挙義務講義ＸＭＬ</title>
  <organization>アウグスブルグ大学コンピュータ・サイエンス過分、工学と家政と形成の大学</organization>
</kougi>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<أ ؤؤ>آ ألآ ؤؤ إأ ؤئ ى ئ ى خ خ س س س ص ض ض ط ط ط ط ظ ظ ش ط
ص غ ض ك لإ ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى ئ ى خ خ س س س
س ض ز س ض ط ط ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق ي ط ظ آ ألآ
ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض ط ط ط ط ط ظ ش ص غ ض ك لإ
ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض ط ط
ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ
خ س س س س ض ز ض ط ط ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق ي
ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض ط ط ط ط ط ظ ش ص غ
ض ك لإ ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض
ط ط ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى
خ خ س س س س ض ز ض ط ط ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق
ي ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض ط ط ط ط ط ظ ش ص غ
ض ك لإ ك ق م ج ق ي ط ظ آ ألآ ؤؤ إأ ؤئ ى خ خ س س س س ض ز ض
ط ط ط ط ط ظ ش ص غ ض ك لإ ك ق م ج ق ي ظ ظ </أ ؤؤ>
```

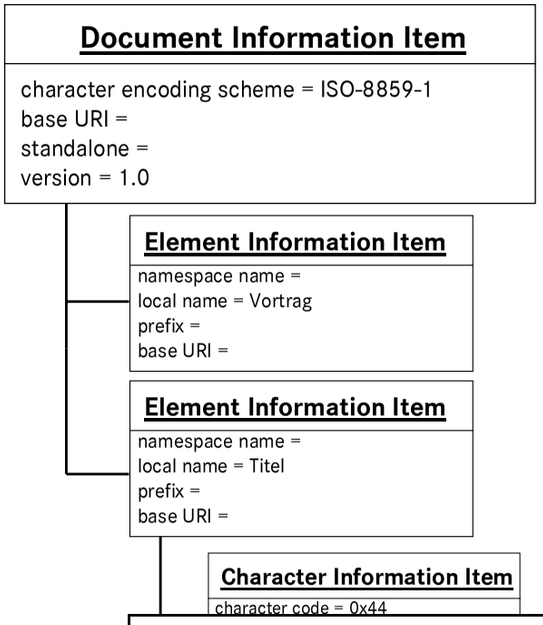
```
ml version="1.0" encoding="ISO-8859-1"?>
rtrag>
'itel>Die Extensible Markup Language (XML)</Titel>
/eranstaltung datum="2001-09-10">
<Name>NetObject.Days 2001</Name>
Veranstaltung>
Referent>
  <Name>Mario Jeckle</Name>
  <Firma>DaimlerChrysler Research and Technology</Firma>
  <URL>http://www.jeckle.de</URL>
  <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
</Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte




```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

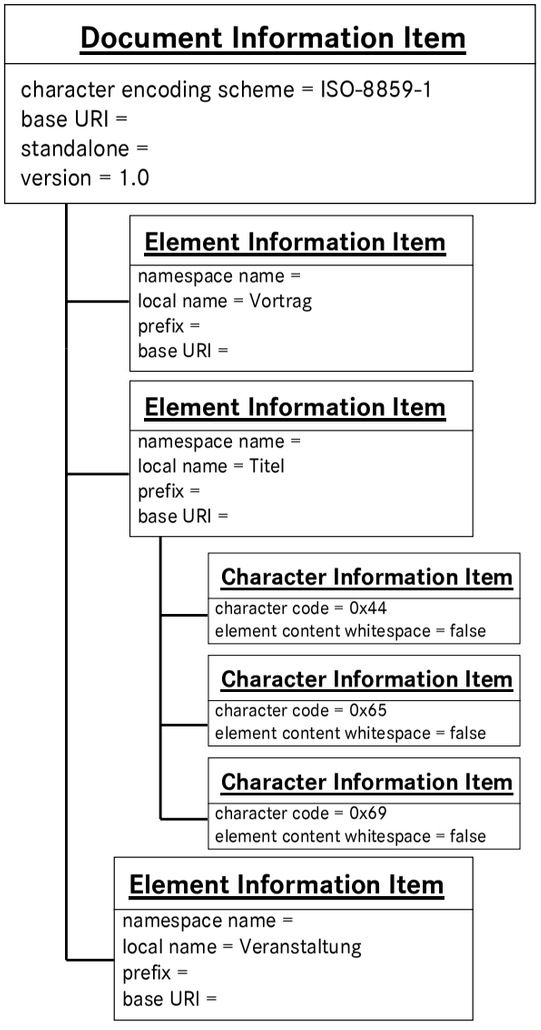


Element Information Item

namespace name =
local name = Veranstaltung
prefix =
base URI =

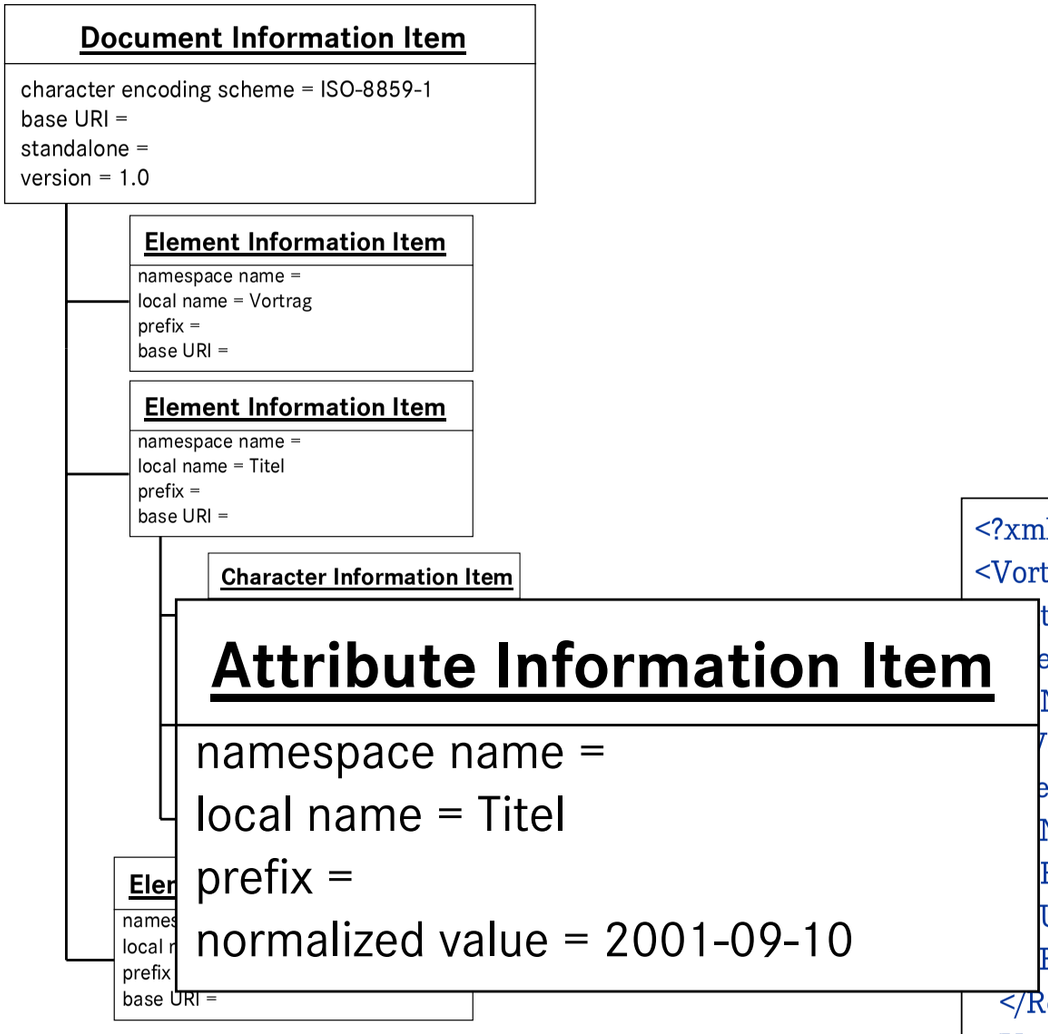
```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte



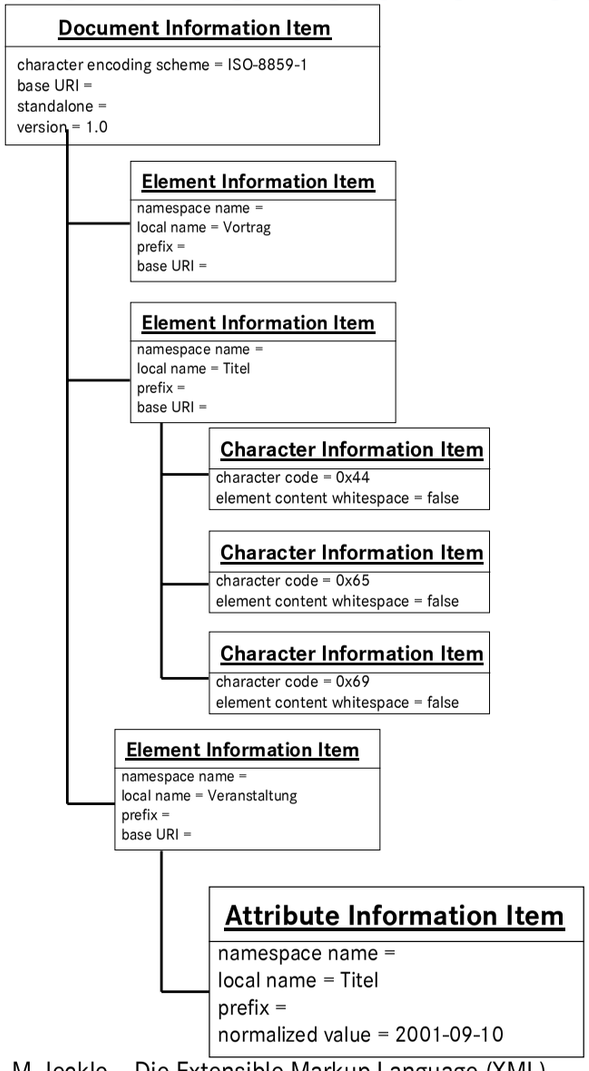

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte




```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  tel>Die Extensible Markup Language (XML)</Titel>
  eranstaltung datum="2001-09-10">
  Name>NetObject.Days 2001</Name>
  /eranstaltung>
  eferent>
  Name>Mario Jeckle</Name>
  Firma>DaimlerChrysler Research and Technology</Firma>
  URL>http://www.jeckle.de</URL>
  E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
 </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte




```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Vortrag>
  <Titel>Die Extensible Markup Language (XML)</Titel>
  <Veranstaltung datum="2001-09-10">
    <Name>NetObject.Days 2001</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

Dokumente und Daten ... Strukturelle Grundkonzepte

- Zusammenfassung: *XML Information Set*
  - Präzisiert die durch die XML Recommendation gegebene pysische Repräsentation von XML-Daten semantisch
  - Abstraktes Modell zur Beschreibung logischer XML-Strukturen
  - Infoset ist im wesentlichen ein Baum
  - Baumknoten werden *Information Item* genannt
  - Liefert Begriffswelt für beliebige wohlgeforme XML-Dokumente
  - Wird durch andere W3C-Standards verwendet
  - Üblicherweise durch einen Parsingvorgang erzeugt

Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - Strukturelle Grundkonzepte
  - **Namensräume**
  - Dokument-Typ-Definitionen (DTD)
  - XML-Schema
  - XML und das Web

Strukturelle Grundkonzepte ... Namensräume

- Problem der (globalen) Eindeutigkeit von Attribut- und Element-bezeichnern in XML
- Lösungsmöglichkeit 1: (*klassisches* Vorgehen in EDI, STEP, ...)
  Standardisierung (d.h. normative Festlegung) der Bezeichner

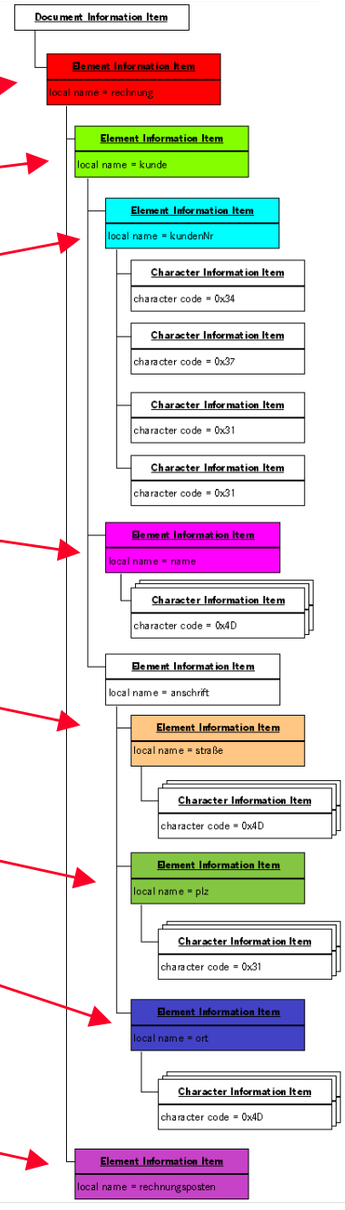
➡ Praktisch unmöglich eine globale Einigung hinsichtlich Semantik und Inhalt zu erzielen

- Lösungsmöglichkeit 2: (ähnlich DNS)
  Prinzipiell freie Vergabe, bei zwingender zentraler Registrierung

➡ Nicht praktikabel, hinsichtlich des abzuschätzenden Aufwandes, sowie der abzusehenden (Rechts-)Streitigkeiten

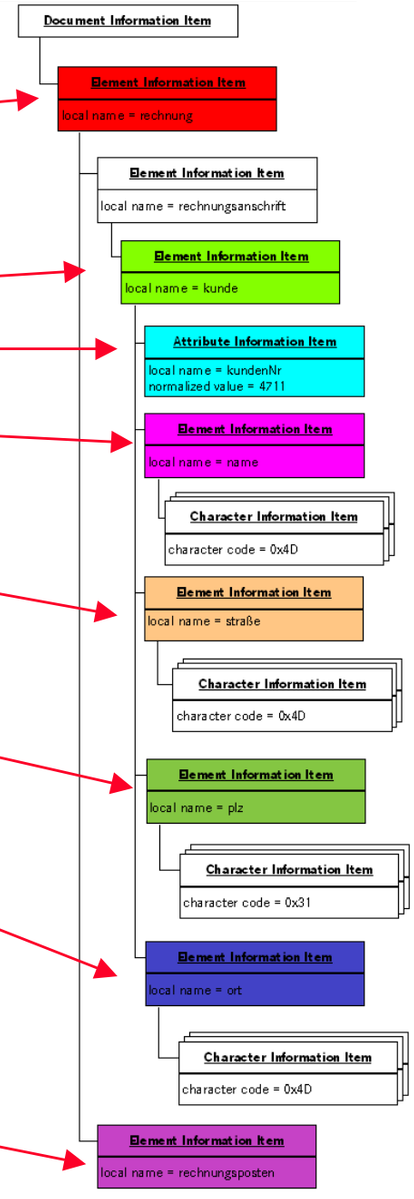
Strukturelle Grundkonzepte ... Namensräume

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<rechnung>
  <kunde>
    <kundenNr>4711</kundenNr>
    <name>Max Mustermann</name>
    <anschrift>
      <straße>Musterplatz 1</straße>
      <plz>12345</plz>
      <ort>Musterstadt</ort>
    </anschrift>
  </kunde>
  <rechnungsposten>
...
<rechnung>
```



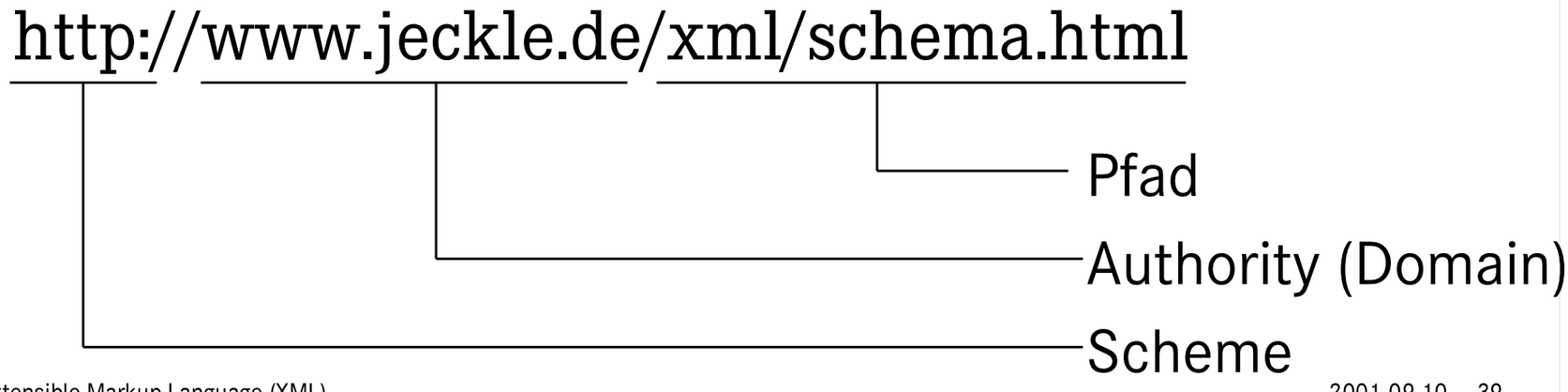
Strukturelle Grundkonzepte ... Namensräume

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<rechnung>
  <rechnungsanschrift>
    <kunde kundenNr="4711">
      <name>Max Mustermann</name>
      <straße>Musterplatz 1</straße>
      <plz>12345</plz>
      <ort>Musterstadt</ort>
    </kunde>
  </rechnungsanschrift>
  <lieferanschrift>
  </lieferanschrift>
  <rechnungsposten>
...
<rechnung>
```



Strukturelle Grundkonzepte ... Namensräume

- Lösungsidee:
  (Wiederbe-)Nutzung der URI als Namenszusatz der Attribut- und Elementbezeichner
  - Zentral verwalteter Domain-Anteil garantiert generell eindeutigkeit
  - Dezentral verwalteter Pfad-Anteil garantiert größtmögliche Flexibilität



Strukturelle Grundkonzepte ... Namensräume

- *Naive Lösung*: Erweiterung der Elementnamen

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<http://www.example.com/sales:rechnung>
  <http://www.example.com/sales:kunde>
    <http://www.example.com/sales:kundenNr>4711</http://www.example.com/sales:kundenNr>
    <http://www.example.com/sales:name>Max Mustermann</http://www.example.com/sales:name>
    <http://www.example.com/sales:anschrift>
      <http://www.example.com/sales:straße>Musterplatz 1</http://www.example.com/sales:straße>
      <http://www.example.com/sales:plz>12345</http://www.example.com/sales:plz>
      <http://www.example.com/sales:ort>Musterstadt</http://www.example.com/sales:ort>
    </http://www.example.com/sales:anschrift>
  </http://www.example.com/sales:kunde>
  <http://www.example.com/sales:rechnungsposten>
...
<http://www.example.com/sales:rechnung>
```

Strukturelle Grundkonzepte ... Namensräume

- *Naive Lösung*: Erweiterung der Elementnamen
  - Umständliche Notation,
    senkt Lesbarkeit dramatisch
  - Vergrößert Speicherplatzaufwand signifikant
    (im Beispiel rund 150%!)
  - Problematisch hinsichtlich Domain- oder Pfadänderungen
  - Verstoß gegen XML-Syntax;
    Pfadtrenner „/“ ist innerhalb Elementnamen nicht zugelassen!

Strukturelle Grundkonzepte ... Namensräume

- Standardisierte Lösung:
  - Zweistufiges Vorgehen:
    1. Bindung einer URI an ein frei wählbares Präfix durch ein standardisiertes Attribut

```
<rechnung xmlns:rch="http://www.examples.com/sales">
```

    2. Qualifizierung der Element- oder Attributnamen mit dem definierten Präfix

```
<rch:rechnung xmlns:rch="http://www.examples.com/sales">
  <rch:kunde>
    <rch:kundenNr>4711</rch:kundenNr>
  </rch:kunde>
...
```

Strukturelle Grundkonzepte ... Namensräume

- Zusammenfassung: *XML Namensräume*
  - Mächtiger, flexibler und einfach zu handhabender Mechanismus zur Gewährleistung der Namenseindeutigkeit von Attributen und Elementen
  - Grundvoraussetzung der XML-basierten *content syndication*
  - Abwärtskompatibel zu Namensraum-losen Dokumenten
  - Parser- und Werkzeugunterstützung verfügbar
  - Eingesetzt in verschiedenen XML-Standards und -Vokabularen
  - W3C-Recommendation seit 1999
  - Vorsicht mit *fragment URIs!*

Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - Strukturelle Grundkonzepte
  - Namensräume
  - **Dokument-Typ-Definitionen (DTD)**
  - XML-Schemasprachen
  - XML und das Web

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Dient zur Festlegung der strukturellen und inhaltlichen Merkmale eines Dokuments
- Eigene (einfache) Syntax
- Untermenge des SGML-Pendants
  - ... daher eher Dokumenten- denn Daten-orientiert
- Grammatik eines XML-Dokuments
- Steigerung der *Wohlgeformtheit* zur *Gültigkeit*
- Wird durch *validierende Parser* verarbeitet
- Werkzeugunterstützung verfügbar

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Verknüpfung zwischen Dokument und DTD

```
<!DOCTYPE ProjektVerwaltung
   SYSTEM "file:///C:/dtds/projektverwaltung.dtd">
<ProjektVerwaltung>
  ...
</ProjektVerwaltung>
```

  - DOCTYPE-Deklaration referenziert wahlweise eine lokale Ressource (SYSTEM) oder eine allgemein bekannte (und potentiell zentral abgelegtes) DTD (PUBLIC)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
```

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Elementdefinition
  - Elementname
  - *Inhaltsmodell* (in Klammern)

```
<!ELEMENT ProjektVerwaltung
   (Person, Projekt)>
```

...im XML-Dokument

```
...
<ProjektVerwaltung>
  <Person>...</Person>
  <Projekt>...</Projekt>
</ProjektVerwaltung>
...
```

...in InfoSet-Darstellung

Element Information Item
namespace name =
local name = ProjektVerwaltung
prefix =
base URI =

Element Information Item
namespace name =
local name = Person
prefix =
base URI =

Element Information Item
namespace name =
local name = Projekt
prefix =
base URI =

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Elementdefinition

```
<!ELEMENT ProjektVerwaltung  (Person+ , Projekt+ )>
<!ELEMENT Person (Vorname+, Nachname, Qualifikationsprofil?)>
```

  - Vielfachheit:
    - \+ mindestens einmaliges Auftreten {1, 2, 3 ...}
    - * beliebiges Auftretten {0, 1, 2, ...}
    - ? höchstens einmaliges Auftreten {0, 1}
  - Gruppierung durch Klammerung, Operatoren "," und "|"

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Elementdefinition

```
<!ELEMENT Projekt EMPTY>
<!ELEMENT Ueberschrift (#PCDATA)>
```

  - Elemente ohne Kindelemente (*„leere Elemente“*) haben Inhaltsmodell EMPTY

  - Unstrukturierter textueller Inhalt: PCDATA
    parsed character data

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Elementdefinition

```
<!ELEMENT Qualifikationsprofil
          (#PCDATA | Qualifikation | Leistungsstufe)*>
```

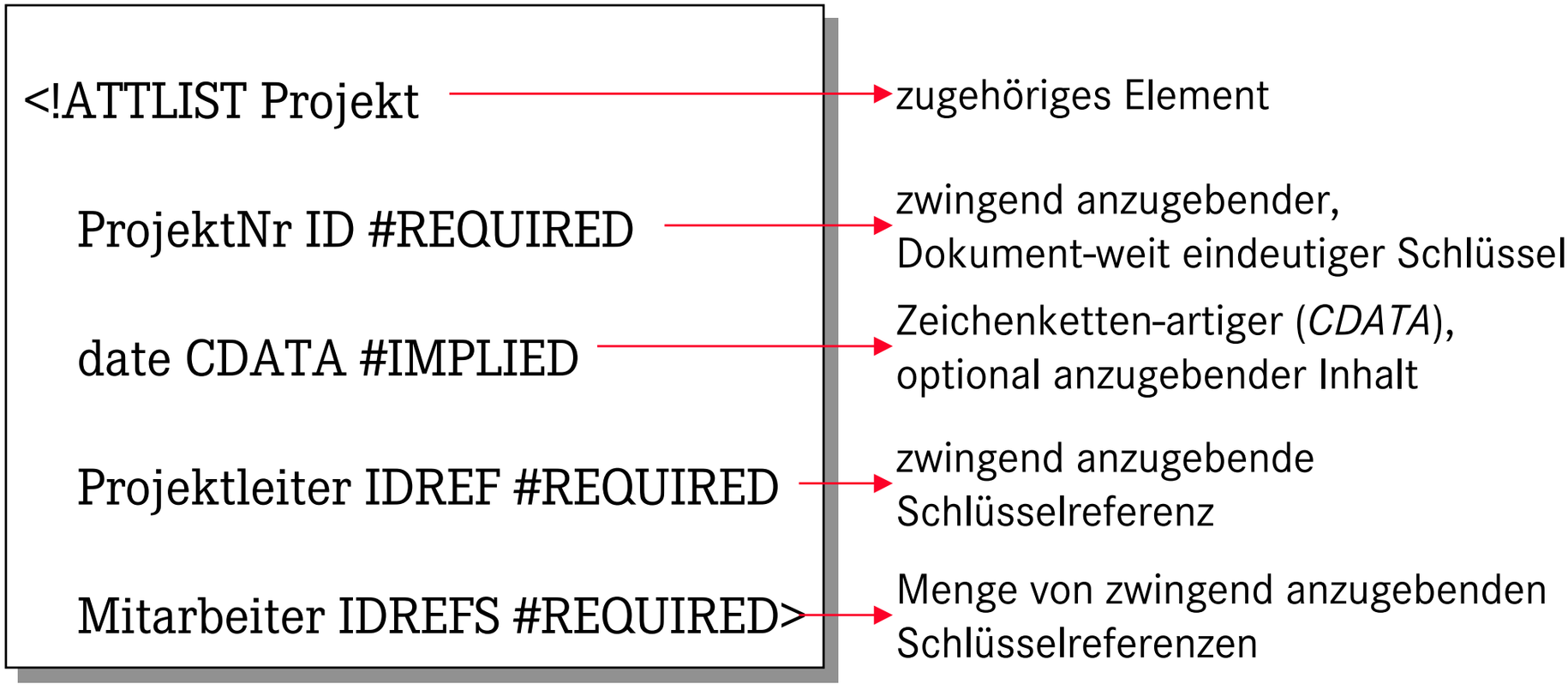
- *Gemischter Inhalt* („mixed content“), kombiniert unstrukturierten textuellen und durch markup strukturierten Inhalt

...im XML-Dokument

```
...
<Qualifikationsprofil>IT-Kompetenz verschiedene Betriebssysteme und
<Leistungsstufe>professionelle</Leistungsstufe>
  <Qualifikation>Programmierung</Qualifikation verschiedener Programmiersprachen
  <Qualifikation>Entwickler</Qualifikation>von 1988-1990
  <Qualifikation>Projektleiterfunktion</Qualifikation> von 1990-93 im X42-Projekt
  in Abteilung AB&amp;amp;C </Qualifikationsprofil>
...
```

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Attributdefinition



Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Attributdefinition

```
<!ATTLIST ProjektVerwaltung

   version CDATA #FIXED "1.0">

<!ATTLIST Projekt

   budget CDATA "10000"

<!ATTLIST Person

   Gehaltsgruppe (1 | 1a | 2) "1a" >
```

→ konstante Belegung

Zeichenketten-artiger (*CDATA*), Inhalt mit Vorgabebelegung. Implizit optional; validierender Parser setzt bei fehlendem Wert automatisch Vorgabewert ein.

Aufzählungstyp mit Vorbelegung

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Entitäten

  ... bekannt aus HTML

```
<!ENTITY auml   "&#228;">
  <!– latin small letter a with diaeresis, U+00E4 ISOlat1 –>
```

  - Einfacher *handlicher* Textersetzungsmechanismus
  - in XML wegen frei wählbarer Codierung nicht mehr von herausgehobener Bedeutung
  - Bei unsachgemäßer Handhabung sehr fehlerträchtig

Dokumente und Daten ... Dokument-Typ-Definition (DTD)

- Zusammenfassung *Dokument-Typ-Definitionen*:
  - Grundlage der Erstellung eigener Vokabulare
  - Von SGML übernommene proprietäre (d.h. nicht XML-)Syntax
  - Dokument-orientiert
    - Rudimentäres (auf Text-Typen beschränktes) Typsystem
    - Nur einfache Strukturierungsmechanismen
    - Keine direkte Unterstützung der XML-Namensräume
  - Gemeinsam mit einem XML-Dokument Eingabe eines validierenden Parsers
  - Werkzeugunterstützung verfügbar

Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - Strukturelle Grundkonzepte
  - Namensräume
  - Dokument-Typ-Definitionen (DTD)
  - **XML-Schema**
  - XML und das Web

Dokumente und Daten ... XML-Schema

- DTDs im Daten-orientierten Umfeld ...
  - Streng hierarchische Sichtweise oftmals zu einschränkend. Gewünscht: Netzartig verknüpfte Elemente und Dokumente
  - Keine Namensraumunterstützung
  - Keine Validierbarkeit der Grammatik
  - Proprietäre Syntax der Grammatiksprache
  - Angebotenes Typsystem unzureichend
  - Angebotene Strukturprimitiven unzureichend

Dokumente und Daten ... XML-Schema

- Erweiterungsforderungen ...
  - Strukturell ....
    - Namespace Integration (und damit XML 2nd edition)
    - Steuerung der Auftretensreihenfolge
      (Beschränkung und kontrollierte Freigabe)
    - Vererbung (Kopier- und Substitutionssemantik)
    - Wiederverwendungsunterstützung
    - Praktisch einsetzbarer Referenzierungsmechanismus
  - Inhaltlich ....
    - Primitive Datentypen (int, float, boolean, ...)
    - Binärstrukturen
    - Komplexe Datentypen (date, Elemente, Strukturen, ...)
    - Eigendefinierte lexikalische Datentypen
    - Konsistenzsichernde Einschränkungen

Dokumente und Daten ... XML-Schema

- Lösungsalternativen ...

Erweiterungen des bestehenden (SGML-/XML-)DTD-Mechanismus

- Data Types for DTD (DT4DTD)

Wissensbeschreibung

- Document Content Description for XML (DCD)
  (*RDF basierte Weiterentwicklung von XML-Data*)

Inspiriert durch XML-API-Entwicklung

- Schema for Object oriented XML (SOX)

XML-Sprachen zur Inhaltsbeschreibung

- Document Definition Markup Language/XSchema (DDML)
- Schematron (XSL-basierte Auswertung der Dokumentstruktur)
- XML-Data/XML-Data Reduced (XDR) *(erster Ansatz noch vor Verabschiedung XML 1.0)*
- Document Structure Description (DSD)

Dokumente und Daten ... XML-Schema

- XML-Sprachen zur Inhaltsbeschreibung ...

  Zwei konkurrierende Philosophien:

  - **Regelbasiert**
    - Definiert Validierungsregeln
    - Bekannteste Implementierung: *Schematron*

  - **Reguläre kontextfreie Grammatik**
    - Definiert alle gültigen Dokumentstrukturen explizit
    - Bekannt aus Programmiersprachen
    - Implementierungen: *XSD, XML-Data, XDR, DSD, ...*
    - Stand 1998/9: Viele verschiedene (konkurrierende) Sprachen
      - Zunehmend kaum noch erkennbare inhaltliche Unterschiede
    - Ab Mai 1999: W3C Arbeitsgruppe entwickelt eigenen Ansatz
      - Trennung zwischen *Struktur* und *Datentypen*
      - Berücksichtigt (nahezu) alle bedeutenden Vorgänger (explizit: DCD, DDML, SOX, XDR, XML-Data)

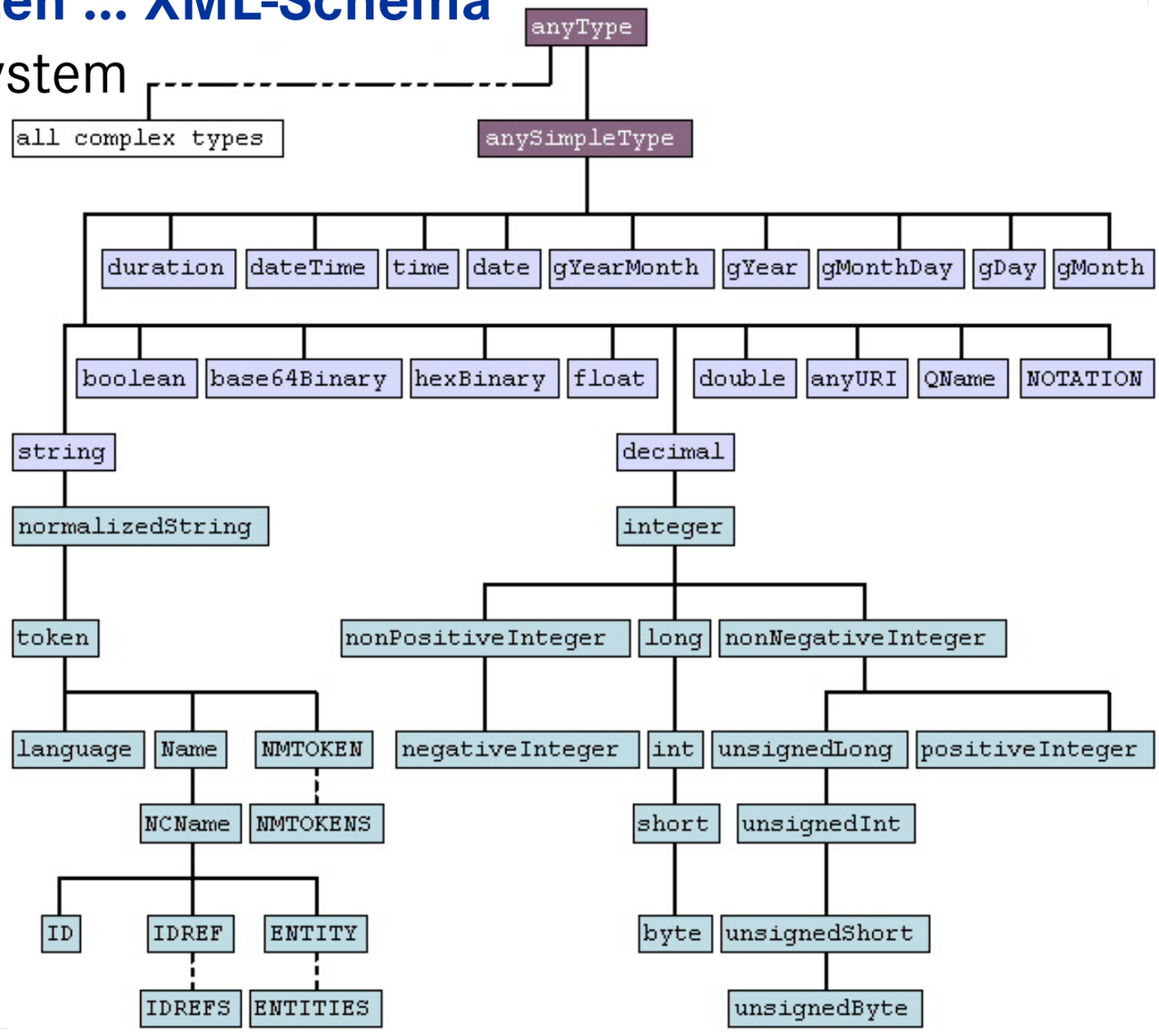
Dokumente und Daten ... XML-Schema

- Mächtigkeit von XML-Schema
  - Attribute und Elemente (wie in DTDs)
  - Namensraum-Unterstützung
  - Atomare Datentypen (int, float, boolean, ...)
  - Anwenderdefinierte
    - atomare Datentypen
      - Einschränkung des Wertebereichs (Domänenrestriktion)
    - lexikalische Muster (reguläre Ausdrücke)
    - Aufzählungstypen
    - Mengentypen
    - komplexe Datentypen (complexType)
  - Vererbung
    - Restriktion und Erweiterung
  - Substitution
  - Erweiterter Schlüsselmechanismus
  - NULL-Werte

Dokumente und Daten ... XML-Schema

- XML-Schema Typsystem

- Ur-typ
- Vordefinierter Primitivtyp
- Vordefinierter abgeleiteter Typ
- Typeinschränkung
- Aggregierter Typ



Dokumente und Daten ... XML-Schema

XML-Schema Typsystem: im Detail

ID	test, XYZ	XSD-Darstellung des DTD-Typen ID. Zugelassen sind alle Ausprägungen der Namespaceproduktion 4 (NCName). ID ist eine einschränkende Spezialisierung des Typs NCName
IDREF	test, XYZ	XSD-Darstellung des DTD-Typen IDREF. Zugelassen sind alle Ausprägungen der Namespaceproduktion 4 (NCName). IDREF ist eine einschränkende Spezialisierung des Typs NCName
IDREFS	test1 test2 test4, test3 test5	XSD-Darstellung des DTD-Typen IDREFS. Zugelassen sind Listen aus white space separierten Ausprägungen der Namespaceproduktion 4 (NCName). IDREFS ist eine nichtleere Aufzählung von IDREF-Ausprägungen
ENTITY		XSD-Darstellung des DTD-Typen ENTITY. Zugelassen sind alle Satzformen die der Produktion NCName der XML-Namensräume entsprechen und als ungeparste Entität definiert sind. ENTITY ist eine einschränkende Spezialisierung des Typs NCName
ENTITIES		XSD-Darstellung des DTD-Typen ENTITIES. Zugelassen sind Listen aus white space separierten Ausprägungen des Typs ENTITY. ENTITIES ist eine nichtleere Aufzählung von ENTITY-Ausprägungen
NOTATION		XSD-Darstellung des DTD-Typen NOTATION. Zur Verwendung dieses Typs in einem Schema muß eine Ableitung von NOTATION durch den Anwender definiert werden.
NMTOKEN	US, Deutschland	XSD-Darstellung des DTD-Typen NMTOKEN. Ausprägungen dieses Typs müssen konform zur Produktion 7 der XML-Spezifikation sein. NMTOKEN ist eine einschränkende Spezialisierung des Typs token
NMTOKENS	US UK Aus, Ger	XSD-Darstellung des DTD-Typen NMTOKENS. Zugelassen sind Listen aus white space separierten Ausprägungen des Typs NMTOKEN. NMTOKENS ist eine nichtleere Aufzählung von NMTOKEN-Ausprägungen

Dokumente und Daten ... XML-Schema

XML-Schema Typsystem: im Detail

string	Hello
World	Jedes beliebige Unicode Symbol gemäß XML-Syntaxproduktion 2
normalizedString	&#20Hello&#20;World	Jedes beliebige Unicode Symbol außer Zeilenvorschub, Wagenrücklauf und Tabulatoren. normalizedString ist eine einschränkende Spezialisierung des Typs string
token	Hello World	Jeder normalizedString, unter Weglassung führender, abschließender und mehrfacher Leerzeichen (#x20), sowie Zeilenvorschüben (#xA) und Tabulatoren (#x9). token ist eine einschränkende Spezialisierung des Typs normalizedString
Name	aName, _helloWorld, :notThatGood	XML Name gemäß Syntaxproduktion 5. Name ist eine einschränkende Spezialisierung des Typs token
QName	xsd:element, xsl:template, foo	Durch Namensraumpräfix qualifizierter Name gemäß Produktion 6 der XML Namespace Recommendation
NCName	aName, _anotherName, X	Name, der keinen Doppelpunkt enthält (non colonized name), gemäß Produktion 4 der XML Namespace Recommendation

Dokumente und Daten ... XML-Schema

XML-Schema Typsystem: im Detail

decimal	-1.23, 12678967.543233, +100000.00, 210	Wertebereich: $i*10^{-n}$, mit i, n aus integer, n>=0 Ein Prozessor muß mindestens 18 Dezimalstellen unterstützen
long	-1, 0, 12678967543233, +100000	Wertebereich: 2^{63} <= long <= 2^{63}-1 long ist eine einschränkende Spezialisierung des Typs integer
int	-1, 0, 126789675, +100000	Wertebereich: -2^{31} <= int <= 2^{31}-1 int ist eine einschränkende Spezialisierung des Typs long
byte	-1, 42	Wertebereich: -2^{7} <= byte <= 2^{7}-1 byte ist eine einschränkende Spezialisierung des Typs short
integer	-1, 0, 12678967543233, +100000	Wertebereich: entspricht der mathematischen Menge der ganzen Zahlen ($\mathbb{Z}$) integer ist eine einschränkende Spezialisierung des Typs decimal

Dokumente und Daten ... XML-Schema

- XML-Schema Typsystem: im Detail

nonNegativeInteger	1, 0, 12678967543233, +100000	Wertebereich: 0 <= nonNegativeInteger nonNegativeInteger ist eine einschränkende Spezialisierung des Typs integer
positiveInteger	1, 12678967543233, +100000	Wertebereich: entspricht der mathematischen Menge der natürlichen Zahlen ($\mathbb{N}$) positiveInteger ist eine einschränkende Spezialisierung des Typs nonNegativeInteger
negativeInteger	-1, -12678967543233, -100000	Wertebereich: {..., -2, -1}, die unendliche Menge der negativen Zahlen negativeInteger ist eine einschränkende Spezialisierung des Typs nonPositiveInteger
unsignedLong	0, 12678967543233, 100000	Wertebereich: 0 <= unsignedLong <= 2^{64}-1 unsignedLong ist eine einschränkende Spezialisierung des Typs nonNegativeInteger
unsignedInt	0, 1267896754, 100000	Wertebereich: 0 <= unsignedInt <= 2^{32}-1 unsignedInt ist eine einschränkende Spezialisierung des Typs unsignedLong
unsignedShort	0, 12678, 10000	Wertebereich: 0 <= unsignedShort <= 2^{16}-1 unsignedShort ist eine einschränkende Spezialisierung des Typs unsignedInt
unsignedByte	0, 126, 100	Wertebereich: 0 <= unsignedByte <= 2^{8}-1 unsignedByte ist eine einschränkende Spezialisierung des Typs unsignedShort

Dokumente und Daten ... XML-Schema

XML-Schema Typsystem: im Detail

time	13:20:00-05:00, 13:20:00.000	Uhrzeit, die täglich wiederkehrt, ausgedrückt im Format gemäß ISO 8601
date	2001-06-11, 20010611	Datumsformat: CCYY-MM-DD, gemäß ISO 8601
dateTime	2001-06-11T14:00:00.000+02:00	Zeitpunkt, ausgedrückt durch Datum und Uhrzeit; beide gemäß ISO 8601 codiert.
gDay	---05, ---31	Darstellung eines wiederkehrenden Tages eines Monats gemäß ISO 8601
gMonth	--03--, --12--	Monatsformat: --MM-- gemäß ISO 8601
gYear	1999, 2000	Darstellung von Jahren des gregorianischen Kalenders gemäß ISO 8601
gYearMonth	2001-05	Darstellung eines Monats eines bestimmten Jahres des gregorianischen Kalenders gemäß ISO 8601
duration	P1Y2M3DT10H30M12.3S Zeitraum von eim Jahr, zwei Monaten, drei Tagen, zehn Stunden, 30 Minuten und 12,3 Sekunden	Nach Größe (Signifikanz) geordnete Koordinate im sechs-dimensionalen Raum aus Jahr, Monat, Tag, Stunde, Minute und Sekunde. Formatdefinition laut ISO 8601

Dokumente und Daten ... XML-Schema

XML-Schema Typsystem: im Detail

float double	-1E4, 1267.43233E12, 12.78e-2, 12, INF	32-Bit-Zahl mit einfacher Genauigkeit gemäß IEEE 754-1985. Wertebereich: $m * 2^e$, wobei m und e integer-Elemente mit $m <= 2^{24}$, und $-149 <= e < 104$ sind.
boolean	true, false, 1, 0	Unterstützung der klassischen zweiwertigen Logik
anyURI	http://www.jeckle.de foo:bar/baz	Jede gemäß IETF RFC 2396 bzw. IETF RFC 2732 gültige URI
language	en-GB, en, de-de	Sprachcodierung gemäß IETF RFC 1766 und XML Recommendation language identification. Die Identifikationsnamen werden durch ISO 639 sowie ISO 3166 definiert. language ist eine einschränkende Spezialisierung des Typs token
base64Binary	aGVsbG8gd29ybGQh	Base64-Darstellung eines beliebigen Binär-interpretierten Inhaltes gemäß IETF RFC 2045. Datenvolumen erhöht sich um Faktor 1,36
hexBinary	0FB7	Hexadezimale Darstellung beliebiger Binär-interpretierter Inhalte. Datenvolumen erhöht sich um Faktor 2
anyType	1, 2.3, aGVsb, 06b8f45, test&#20;für&#20;anyType &#0A; <sentence>the quick brown <animal>fox</animal>...	Allgemeinster Datentyp. Konzeptionell bildet er die Vereinigung aller angebotenen XSD-Typen.

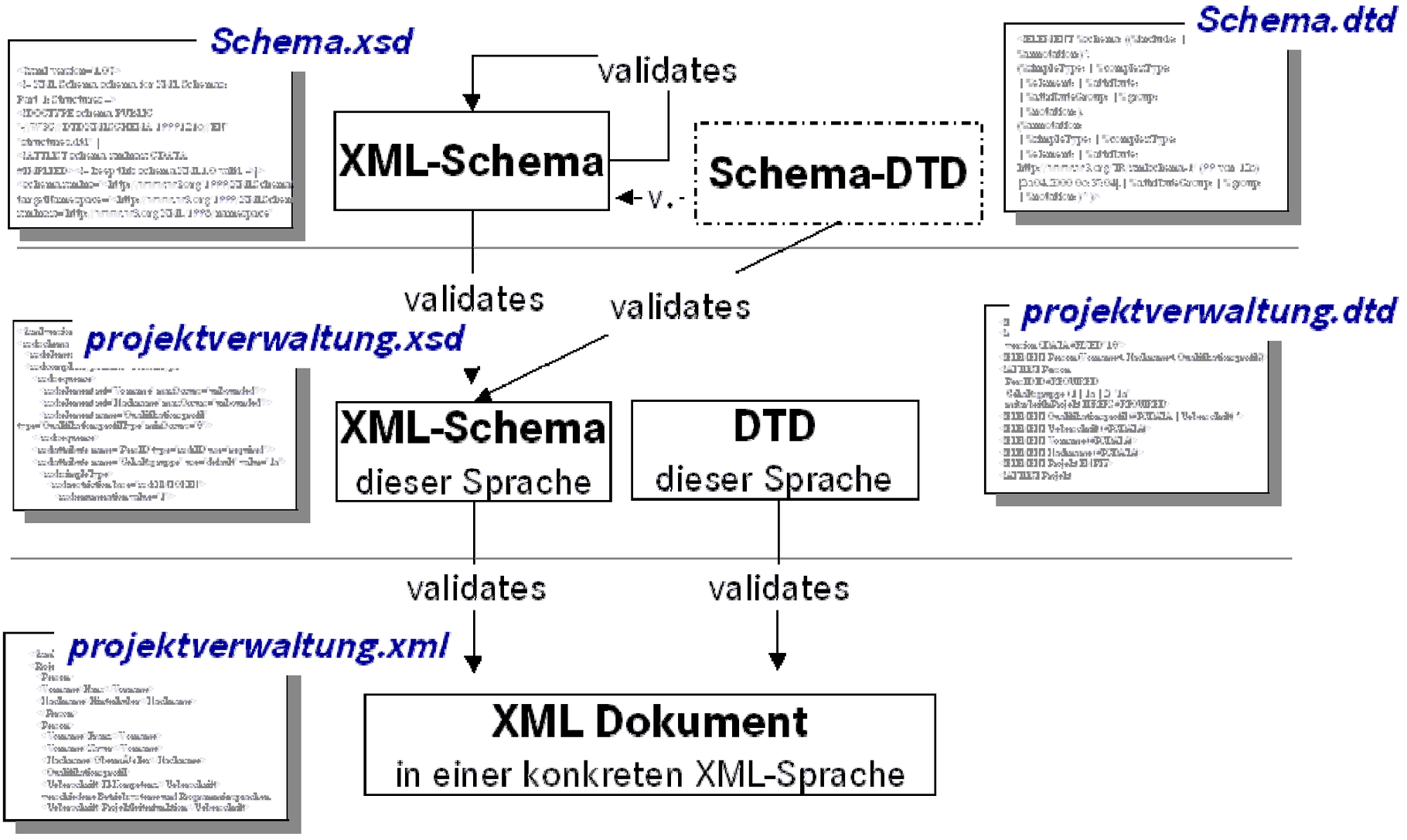
Dokumente und Daten ... XML-Schema

- Schemareferenz
  - Analog der SYSTEM-Variante der DOCTYPE-Deklaration

```
<?xml version="1.0"?>
<ProjektVerwaltung
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.jeckle.de/vorlesung/xml/examples/projektverwaltung.xsd">
...
</ProjektVerwaltung>
```

  - Deklaration des *schema instance* Namensraumes
  - Nutzung des Attributs schemaLocation zur Referenzierung
  - Zu jedem Dokument kann nur maximal ein Schema existieren

Dokumente und Daten ... XML-Schema



Dokumente und Daten ... XML-Schema

- Elementdefinition
  - einfachster Fall:

```
<!ELEMENT Ueberschrift (#PCDATA)>
```

```
<xsd:element
  name="elementName"
  type="elementType"/>
```

  - Zeichenketten-artiges Inhaltsmodell:

```
<xsd:element
  name="Ueberschrift"
  type="xsd:string"/>
```

Dokumente und Daten ... XML-Schema

- Elementdefinition
  - Inhaltsmodelle:
    - explizite benannte Definition:

```
<xsd:complexType
  name="complexTypeName">
  ....
</xsd:complexType>
```

    - implzite anonyme Definition:

```
<xsd:element name="elementName">
  <xsd:complexType>
    ...
  </xsd:complexType>
</xsd:element>
```

Dokumente und Daten ... XML-Schema

- Elementdefinition
  - leeres Inhaltsmodell:

```
<!ELEMENT Projekt EMPTY>
```

    - implzite anonyme Definition:

```
<xsd:element name="Projekt">
  <xsd:complexType/>
</xsd:element>
```

    - Alternativ:

```
<xsd:complexType name="ctEmpty"/>

<xsd:element name="Projekt" type="ctEmpty"/>
```

Dokumente und Daten ... XML-Schema

- Elementdefinition

```
<!ELEMENT Person (Vorname+, Nachname, Qualifikationsprofil?)>
```

  - beliebige Inhaltsmodelle:

```
<xsd:element name="Person">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="Vorname" minOccurs="1" maxOccurs="unbounded"/>
      <xsd:element name="Nachname" type="xsd:string"/>
      <xsd:element name="Projekt" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
</xsd:element>
```

  - minOccurs und maxOccurs erlaubt die Feingranulare Auftretenshäufigkeitssteuerung.
    Reformulierung und Erweiterung von: ?, *, +

Dokumente und Daten ... XML-Schema

- Elementdefinition
  - beliebige Inhaltsmodelle:

```
<xsd:element name="Person">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="Vorname" minOccurs="1" maxOccurs="unbounded"/>
      <xsd:element name="Nachname" type="xsd:string"/>
      <xsd:element name="Projekt" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
</xsd:element>
```

    - sequence erzwingt im XML-Dokument dieselbe Elementreihenfolge, die im Schema angegeben ist.
      Analog dem ","-Operator: (...,...,...)
    - all erlaubt das beliebig geordnete Auftreten der Elemente im XML-Dokument
      Analog: (...|...|...)*. Jedoch mit Häufigkeitskontrolle
    - choice erlaubt das auftreten genau eines Elements der Auswahl.
      Analog: (...|...|...)

Dokumente und Daten ... XML-Schema

- Elementdefinition, *mixed content*

```
<!ELEMENT Qualifikationsprofil
          (#PCDATA | Qualifikation | Leistungsstufe)*>
```

```
<xsd:element name = "Qualifikationsprofil">
  <xsd:complexType mixed = "true">
    <xsd:sequence>
      <xsd:element name = "Qualifikation" type = "xsd:string"
         minOccurs = "0" maxOccurs = "unbounded"/>
      <xsd:element name = "Leistungsstufe" type = "xsd:string"
         minOccurs = "0" maxOccurs = "unbounded"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

Dokumente und Daten ... XML-Schema

- Vererbung
  - Ableitung durch Einschränkung (*derivation by restriction*)
    Der erbende Subtyp gibt eine engere Definition des Supertypen
  - Ableitung durch Erweiterung (*derivation by extension*)
    Der erbende Subtyp erweitert die Definition des Supertypen
  - Ableitung durch Redefinition (*derivation by redefinition*)
    Der erbende Subtyp verändert die Typdefinition des Supertypen

Dokumente und Daten ... XML-Schema

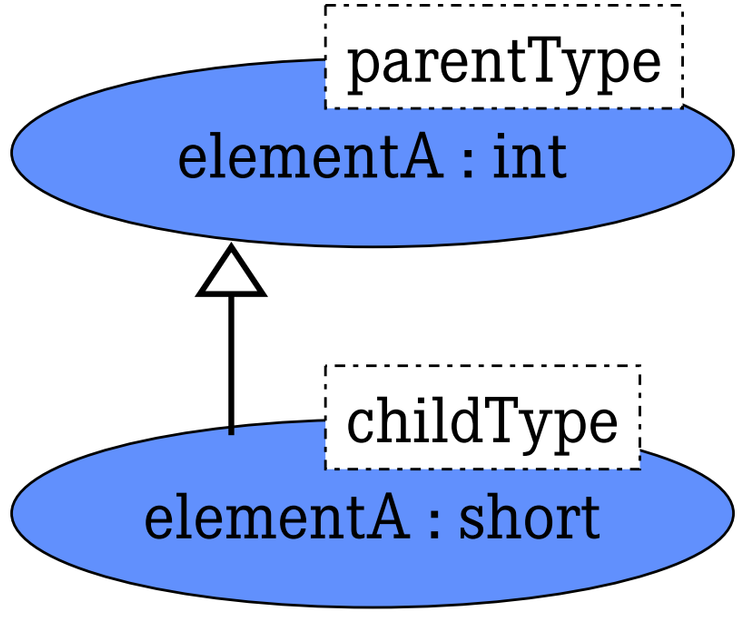
- Vererbung: derivation by restriction

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema
    xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<xsd:complexType name="parentType">
  <xsd:sequence>
    <xsd:element name="elementA" type="xsd:int"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="childType">
<xsd:complexContent>
  <xsd:restriction base="parentType">
    <xsd:sequence>
      <xsd:element name="elementA" type="xsd:short"/>
    </xsd:sequence>
  </xsd:restriction>
</xsd:complexContent>
</xsd:complexType>

<xsd:element name="usage1" type="parentType"/>
<xsd:element name="usage2" type="childType"/>

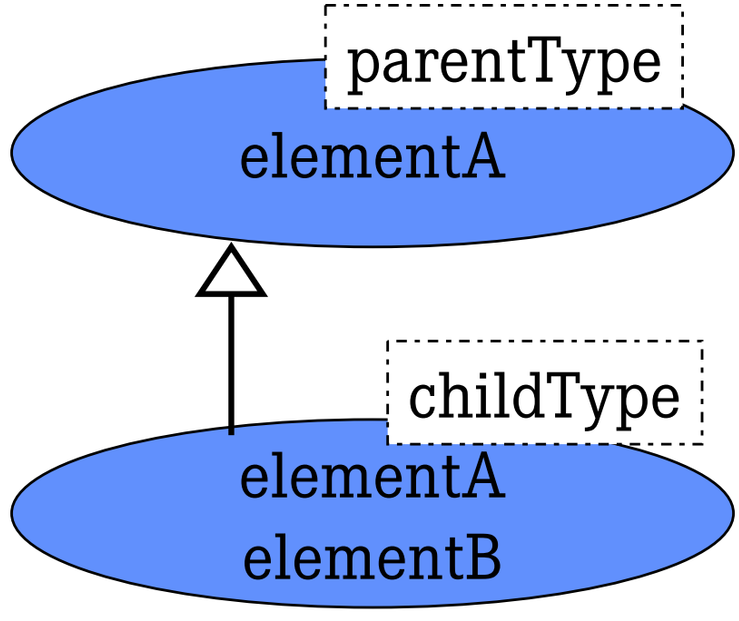
</xsd:schema>
```



Dokumente und Daten ... XML-Schema

- Vererbung: derivation by extension

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:complexType name="parentElement">
    <xsd:sequence>
      <xsd:element name="elementA"/>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="childElement">
    <xsd:complexContent>
      <xsd:extension base="parentElement">
        <xsd:sequence>
          <xsd:element name="elementB"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>
</xsd:schema>
```



Dokumente und Daten ... XML-Schema

- Vererbung: derivation by redefinition

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema
   xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:complexType name="aType">
    <xsd:sequence>
      <xsd:element name="elementA"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:element name="usage1" type="aType"/>
```

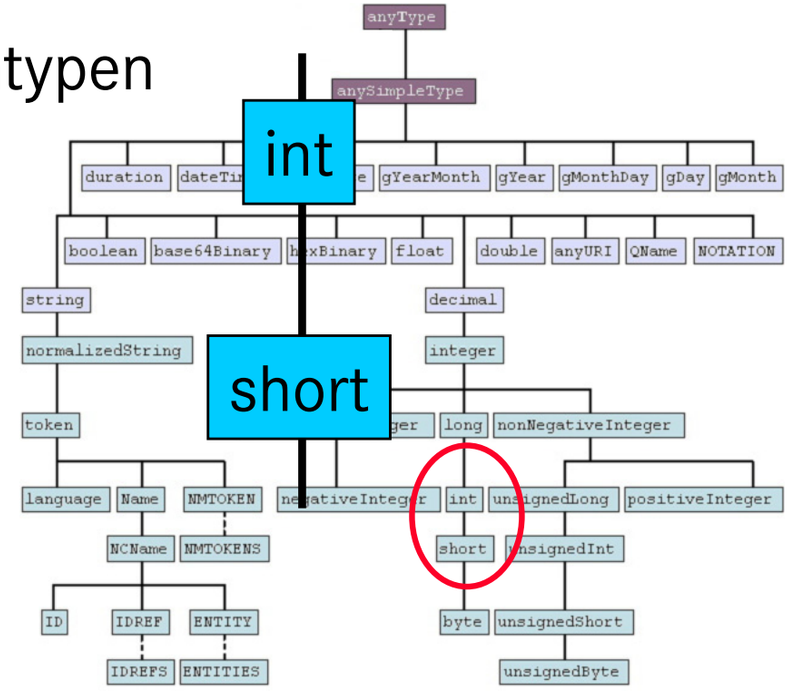
```
  <xsd:complexType name="aType">
    <xsd:complexContent>
      <xsd:extension base="aType">
        <xsd:sequence>
          <xsd:element name="elementB"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>

  <xsd:element name="usage2" type="aType"/>
</xsd:schema>
```

Dokumente und Daten ... XML-Schema

- Erweiterung des Typsystems
  - Definition eigener einfacher Datentypen

```
<xsd:simpleType name="short">
  <xsd:restriction base="xsd:int">
    <xsd:minInclusive value="-32768"/>
    <xsd:maxInclusive value="32767"/>
  </xsd:restriction>
</xsd:simpleType>
```



Dokumente und Daten ... XML-Schema

- Erweiterung des Typsystems
  - Definition eigener einfacher Datentypen
    - Durch vollständige Aufzählung

```
<xsd:simpleType name="ampelfarben">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="rot"/> ...
</xsd:simpleType>
```

    - Durch Typeaggregation

```
<xsd:simpleType name="WarenkorbElemente">
  <xsd:list itemType="xsd:string"/>
</xsd:simpleType>
```

    - Durch Typvereinigung

```
<xsd:simpleType name="termin">
  <xsd:union memberTypes="xsd:date NamenDerWochentage"/>
</xsd:simpleType>
```

Dokumente und Daten ... XML-Schema

- Erweiterung des Typsystems
  - Definition eigener einfacher Datentypen
    - Durch Beschränkung bestehender Typen
      - Längeneinschränkung

```
<xsd:simpleType name="Postleitzahl">
  <xsd:restriction base="xsd:string">
    <xsd:length value="5"/>
  </xsd:restriction>
```

      - ... sinngemäß Minimal- und Maximallänge
      - Lexikalische Einschränkung durch reguläre Ausdrücke

```
<xsd:simpleType name="gerAutoNummer">
  <xsd:restriction base="xsd:string">
    <xsd:pattern value="\p{Lu}{1,3}-\p{Lu}{1,2} \p{Nd}{1,4}"/>
  </xsd:restriction>
</xsd:simpleType>
```

Dokumente und Daten ... XML-Schema

- Zusammenfassung: *W3C's XML Schema*
  - Grammatik für beliebige XML-Vokabulare
  - Part 1 beschreibt Strukturen und Inhaltseinschränkungen
  - Part 2 definiert Datentypdefinition für Schema Part 1 und weitere XML-Vokabulare
  - Signifikante Erweiterung der DTD-Mächtigkeit, wird diese langfristig ersetzen
  - Ist eine XML-Sprache
  - Integriert die wichtigsten konkurrierenden Vorgängeransätze
  - Seit 2001-05-02 W3C Recommendation
  - Basis aller W3C-Standards der 2. Generation (XPath v2.0, XSLT v2.0, XHTML v2.0, SOAP 1.2, ...)
  - Werkzeugunterstützung verfügbar
  - Erster Schritt der Schema-Bestrebungen, weitere werden folgen ...

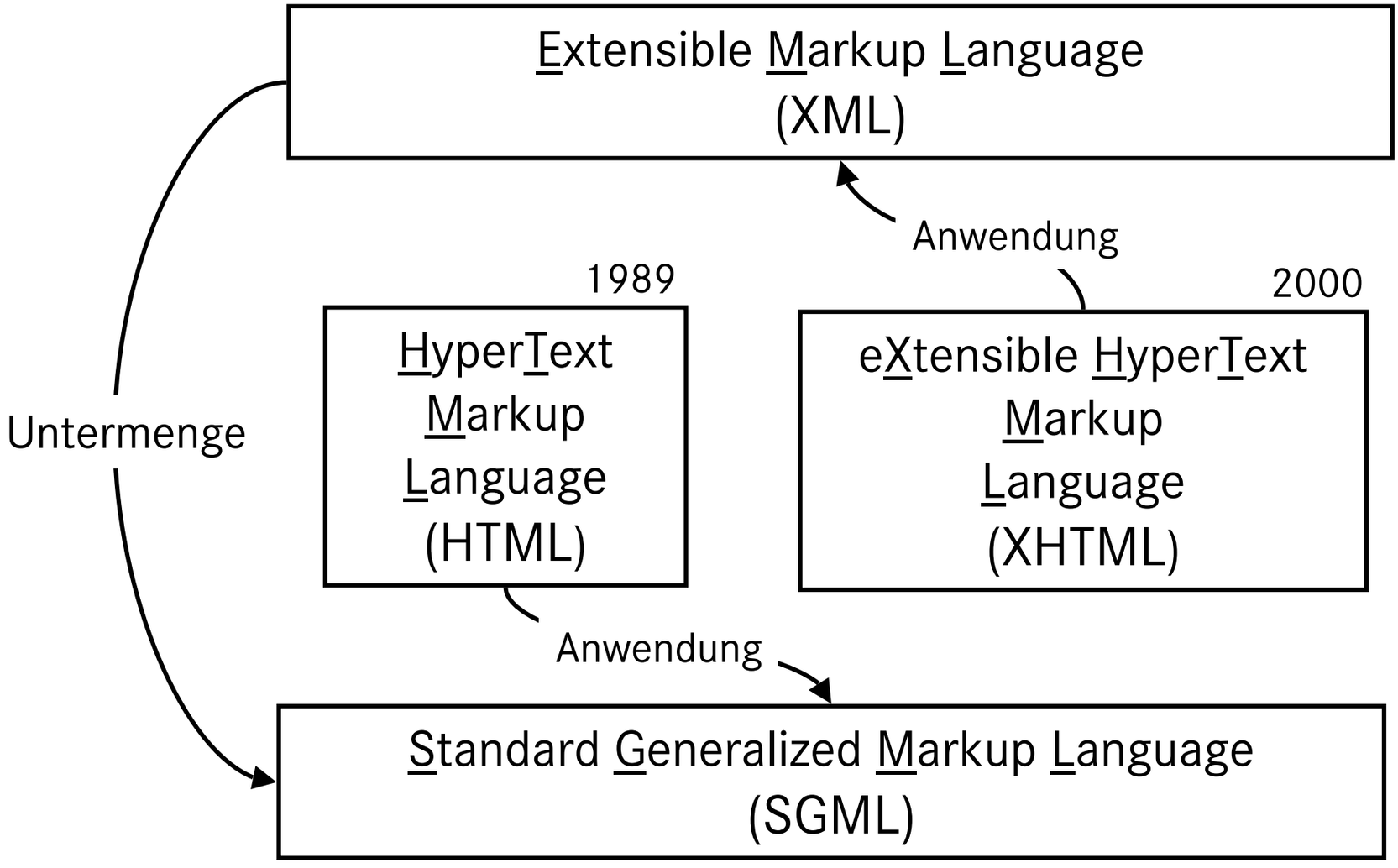
Gliederung

- Dokumente und Daten ...
  - Einführung und Überblick
  - Strukturelle Grundkonzepte
  - Namensräume
  - Dokument-Typ-Definitionen (DTD)
  - XML-Schema
  - **XML und das Web**

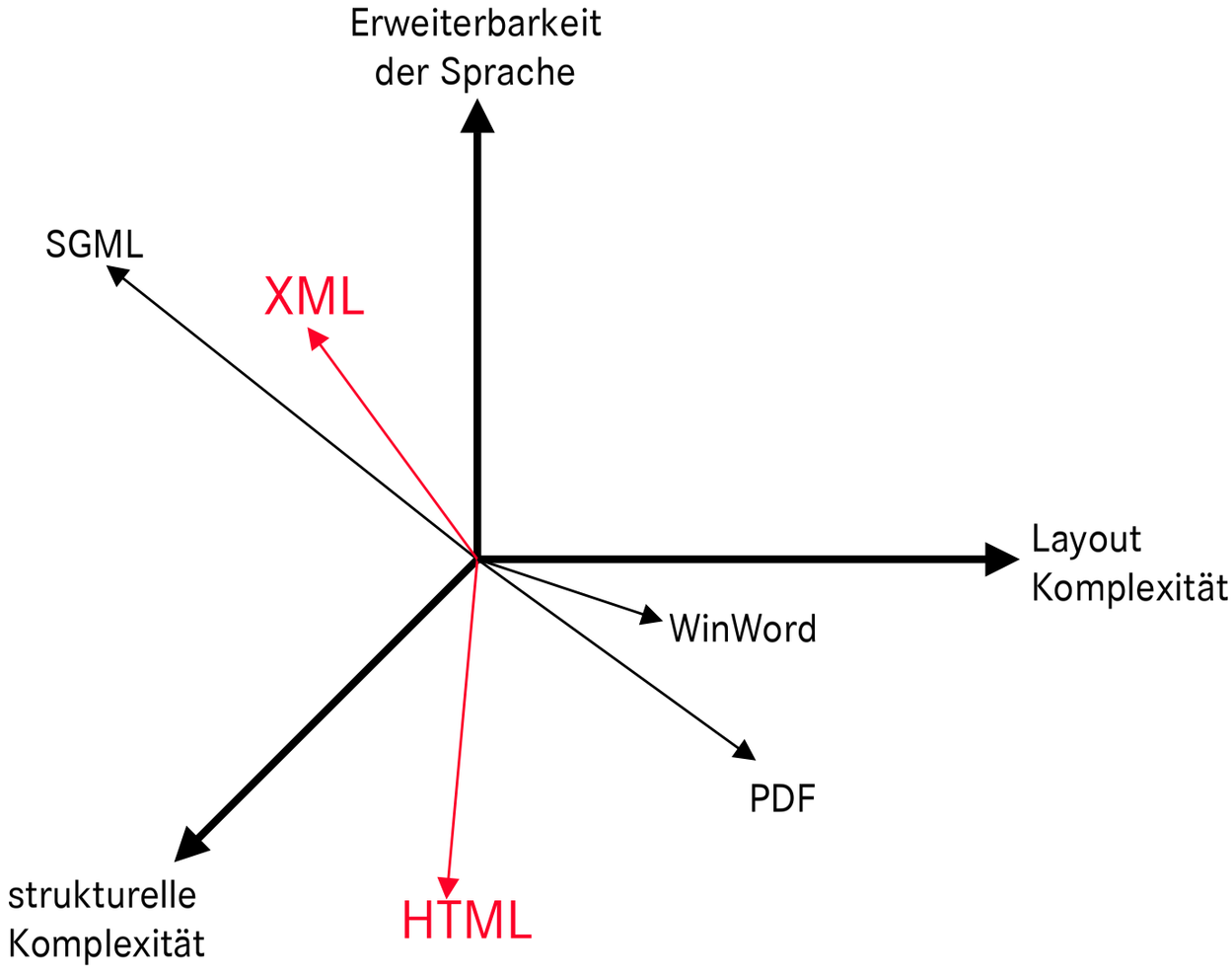
Dokumente und Daten ... XML und das Web

- HTML im Zeitalter von XML
  - von HTML zu XML
  - Aktuelle Entwicklungen
  - was geht heute schon?
- XML im Web-Browser?!
- HTML und eigene präsentationsorientierte Markupsprachen

Dokumente und Daten ... XML und das Web



Dokumente und Daten ... XML und das Web



Dokumente und Daten ... XML und das Web




```
<html>
<head><title>jeckle.de</title></head>
<body bgcolor="#003366" topmargin="0" leftmargin="0"  marginwidth="0" marginheight="0">
<table width="100%" border="0" cellspacing="0" cellpadding="0" height="65%" bgcolor="#f0f0f0">
 <tr>
  <td valign="bottom"> </td>
  <td colspan="2" valign="bottom" align="center">
   <table width="800" border="0" cellspacing="0" cellpadding="0" align="center" height="315">
    <tr bgcolor="#f0f0f0">
     <td height="25"> </td>       <td height="25"> </td>
     <td height="25"> </td>       <td height="25"> </td>
...
```



Tim Berners-Lee

Beginn Fettdruck (*bold*)

Ende Fettdruck

- HTML legt (hauptsächlich) das Präsentationsverhalten fest
- XML definiert die syntaktische Struktur der Information

Dokumente und Daten ... XML und das Web

- Keine überlappenden Elemente

```
<p>here is an emphasized <em>paragraph.</p></em>
```

```
<p>here is an emphasized <em>paragraph</em>.</p>
```

- Korrekte Terminierung

```
<p>here is a paragraph.</p>
<p>here is another paragraph.</p>
```

```
<p>here is a paragraph.
<p>here is another paragraph.
```

Dokumente und Daten ... XML und das Web

- Attributdarstellung

```
table rows="3"
```

```
table rows=3
```

- Attributminimierung

```
<dl compact="compact">
```

```
<dl compact>
```

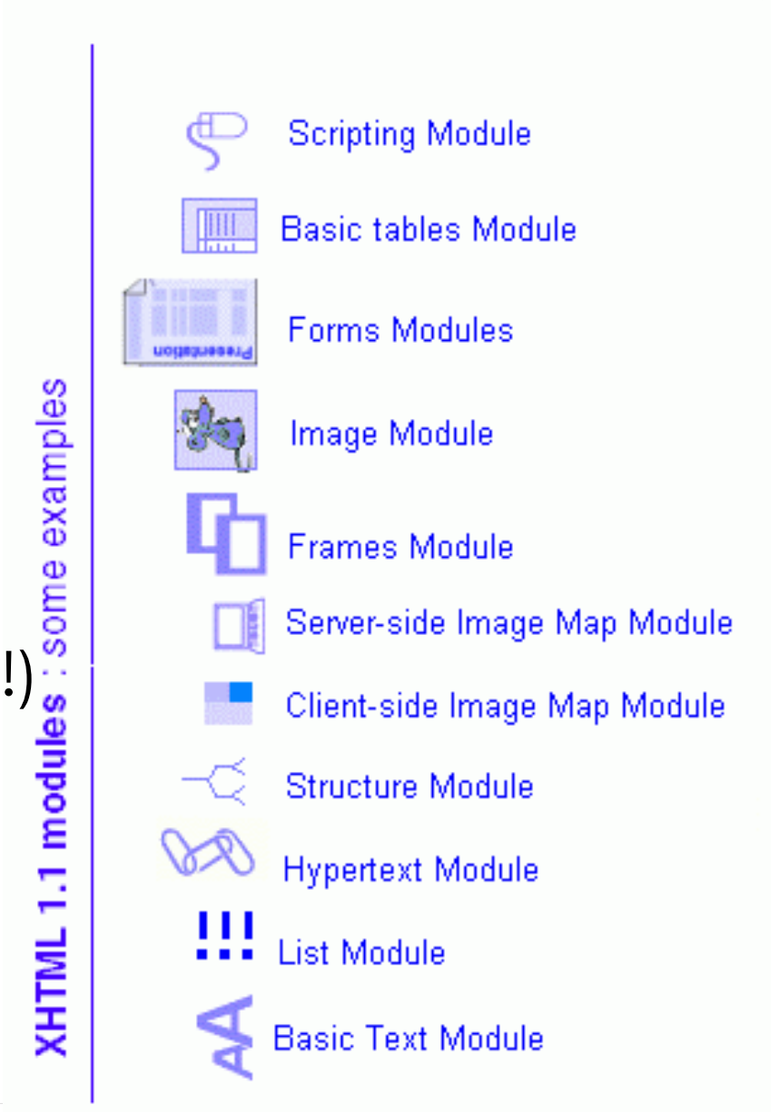
- Leere Elemente

```
<br/><hr/>
```

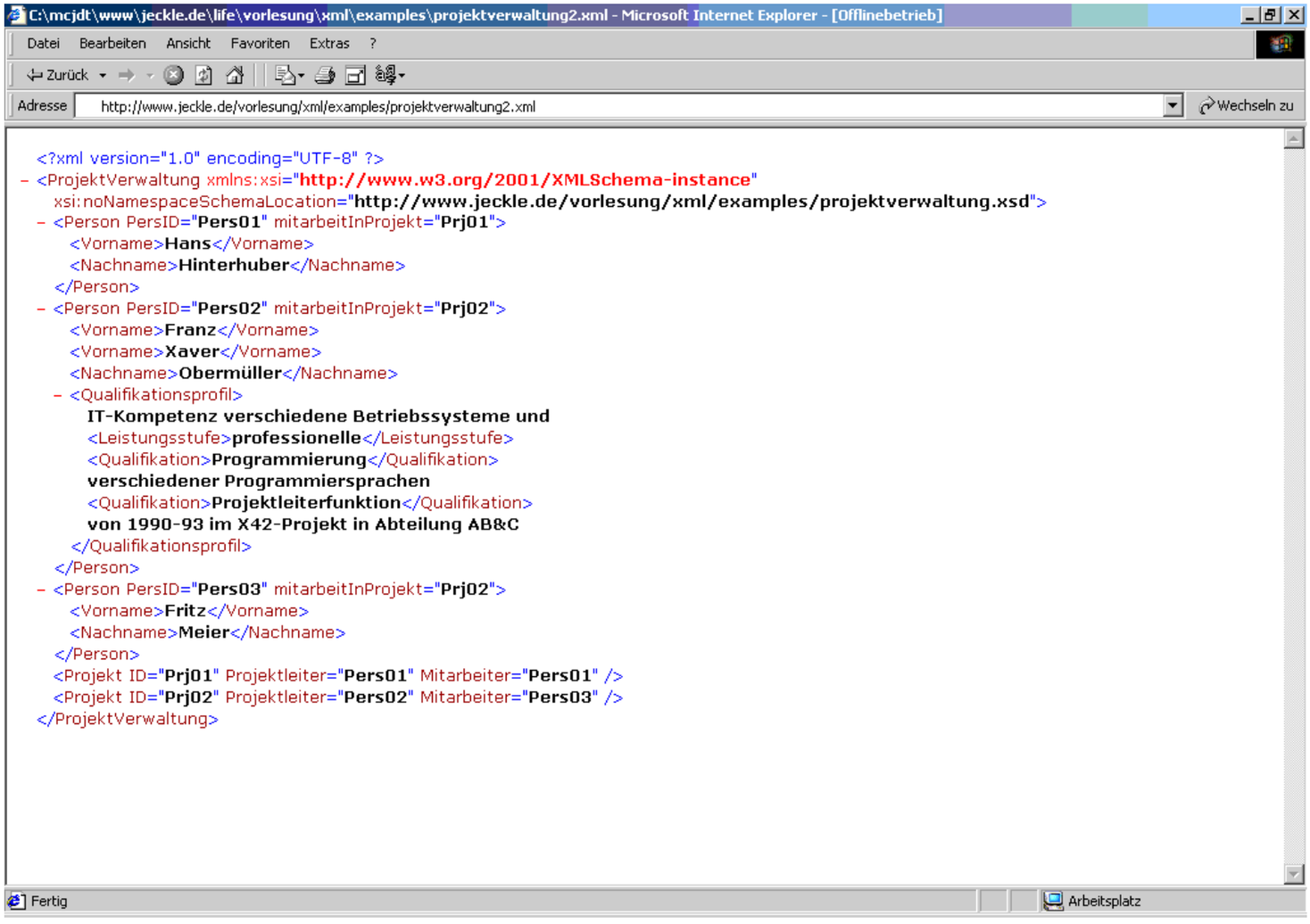
```
<br><hr>
```

Dokumente und Daten ... XML und das Web

- Modulares XHTML ist die funktionale Dekomposition von XHTML.
- Separierung der in XHTML adressierten verschiedenen Problembereiche, um diese
  - getrennt einsetzen zu können (WAP!)
    erweitern zu können, ohne XHTML-Standard zu verletzen



Dokumente und Daten ... XML und das Web



Dokumente und Daten ... XML und das Web

- Einbettung von XHTML-Modulen in eigene XML-Vokabulare
  - Namensräume!
  - XML-Schema!

XML-Dokument:

```
<Qualifikationsprofil>
  <u>IT-Kompetenz</u><em>verschiedene</em> Betriebssysteme und
  <Leistungsstufe>professionelle</Leistungsstufe>
  <em><Qualifikation>Programmierung</Qualifikation></em>
  verschiedener Programmiersprachen
  <em><u><Qualifikation>Entwickler</Qualifikation></u></em> von 1988-1990
  <u><Qualifikation>Projektleiterfunktion</Qualifikation></u>
  von <b>1990-93</b> im X42-Projekt in Abteilung AB&amp;amp;C
</Qualifikationsprofil>
```

Dokumente und Daten ... XML und das Web

- Einbettung von XHTML-Modulen in eigene XML-Vokabulare
  - Namensräume!
  - XML-Schema!

XML-Schema:

```
<xsd:complexType name="QualifikationsprofilType" mixed="true">
  <xsd:choice minOccurs="0" maxOccurs="unbounded">
    <xsd:element ref="Qualifikation"/>
    <xsd:element ref="Leistungsstufe"/>
    <xsd:any namespace="http://www.w3.org/1999/xhtml"
       minOccurs="0" maxOccurs="unbounded"/>
  </xsd:choice>
</xsd:complexType>
```

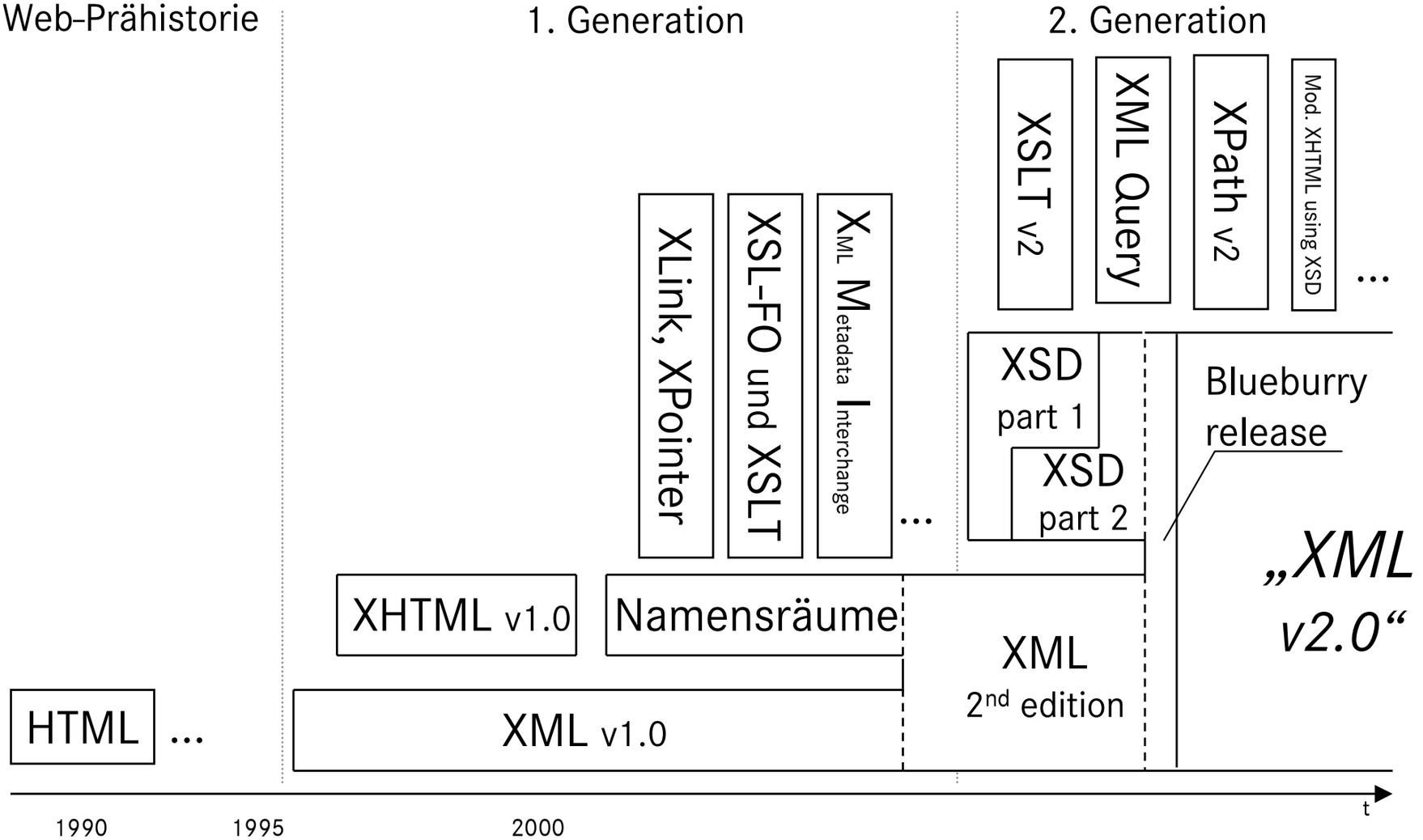
Dokumente und Daten ... XML und das Web

- Zusammenfassung: XML und das Web
  - HTML hat sich von einer Anwendung der SGML, zu XHTML einer Anwendung von XML weiterentwickelt
  - Reformulierung XHTML v1.0 erschließt HTML v4.01 die XML-Sprach- und Werkzeugwelt
  - Durch Namensräume und XML-Schema kann HTML-Semantik in eigene XML-Sprachen übernommen werden
  - Web-Browser-Darstellung von nativem XML ist möglich und wird verschiedentlich unterstützt. Die Hauptintention liegt jedoch nicht in der Präsentation; hierfür existieren Transformationssprachen (->XSL(T))

Gliederung

- **XML-Standards und -Anwendungen der zweiten Generation ...**
  - (Dokument-)Verknüpfungen: XML Links
  - Die Lokatorsprache XPath
  - Erzeugung von Präsentationssichten: XML Stylesheets
  - Transformation von XML-Dokumenten: XSL Transformations
  - Die Anfragesprache XQuery

XML-Standards und -Anwendungen der zweiten Generation

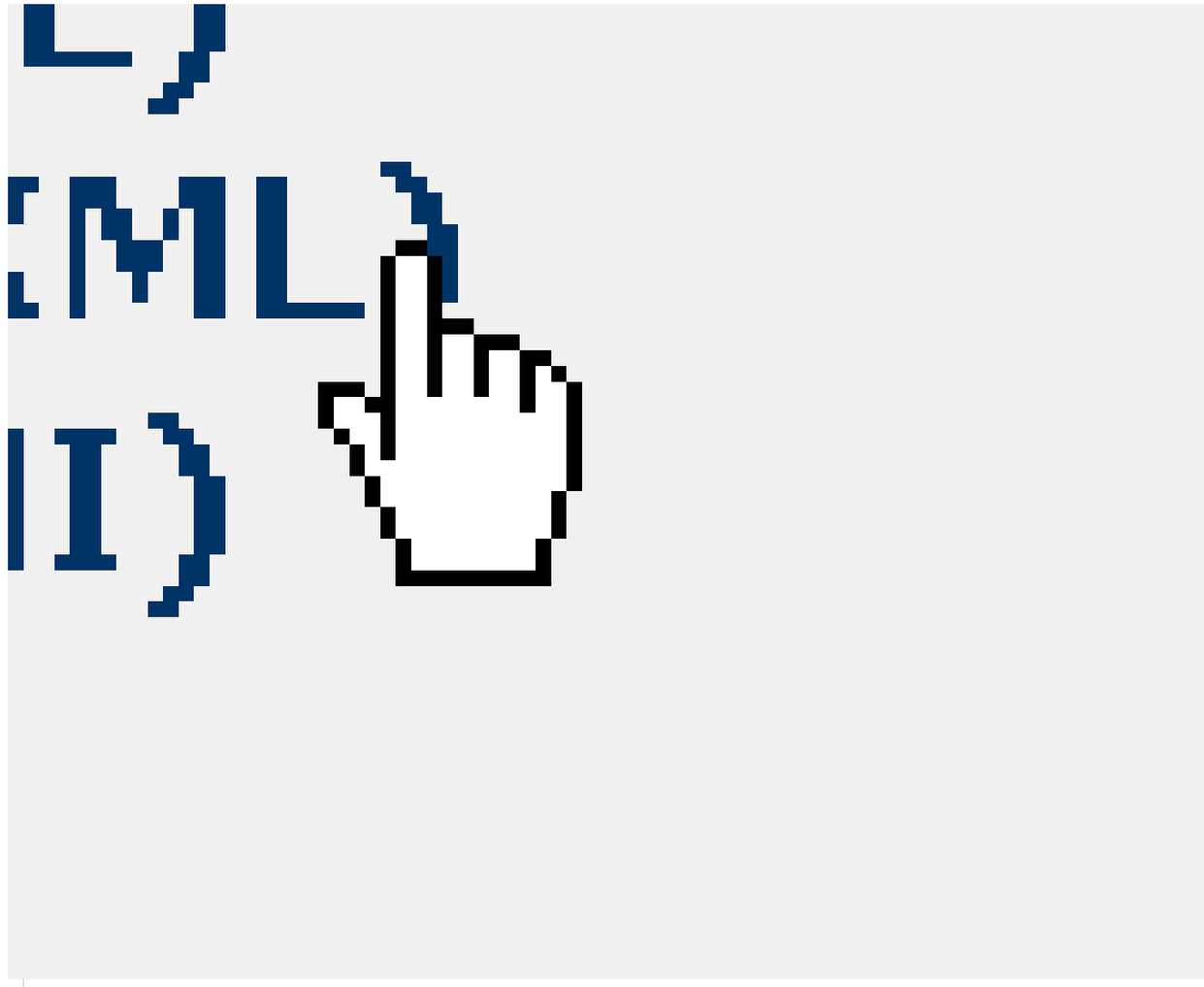


Gliederung

- XML-Standards und -Anwendungen der zweiten Generation ...
- **(Dokument-)Verknüpfungen: XML Links**
- Die Lokatorsprache XPath
- Erzeugung von Präsentationssichten: XML Stylesheets
- Transformation von XML-Dokumenten: XSL Transformations
- Die Anfragesprache XQuery

(Dokument-)Verknüpfungen: XML Links

- HTML-Hyperlinks sind allgegenwärtig
- 404-en auch :-(
- Links sind nur unidirektional
- Verweisen nur auf ein einziges Ziel
- Sind Dokument-intern abgelegt



(Dokument-)Verknüpfungen: XML Links

- Verallgemeinerung des HTML-Hyperlinkmechanismus auf beliebige XML-Dokumente
- Erweiterung der klassischen Möglichkeiten
- XLink definiert hierzu eine Menge von XML-Attributen, die in beliebige eigene Sprache eingebaut werden können
- XLink-Prozessoren *verstehen* nur diese Attribute, und lassen andere Sprachbestandteile unverarbeitet

(Dokument-)Verknüpfungen: XML Links

- *Simpler*-Xlink rekonstruiert Mächtigkeit der HTML href

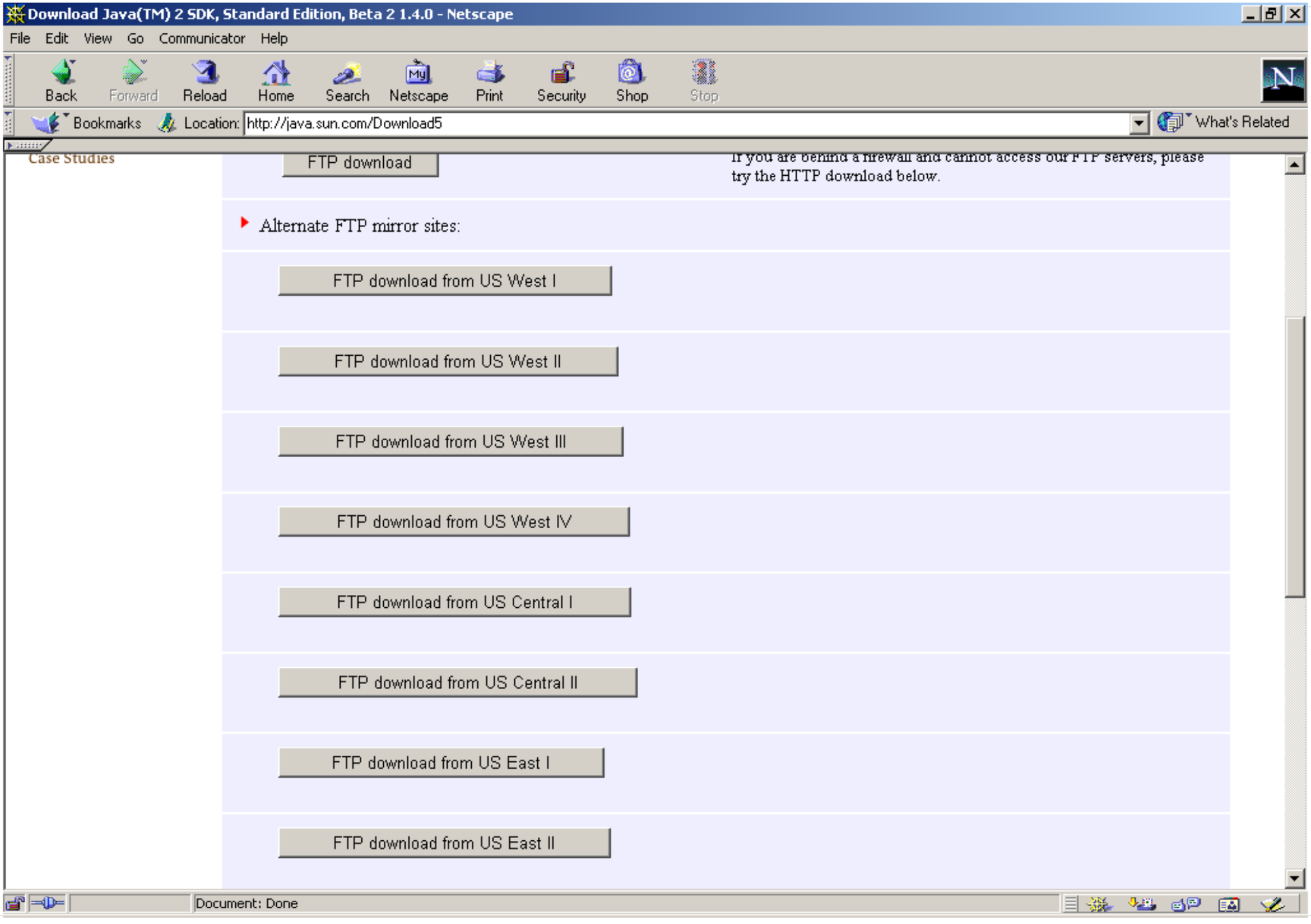
```
<myElementA xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href="http://www.jeckle.de">
  ...
</myElementA>
```

- Vorgehen:
  - Definition des XLink-Namensraums
  - Verwendung der qualifizierten Attribute daraus an beliebiger Stelle

(Dokument-)Verknüpfungen: XML Links

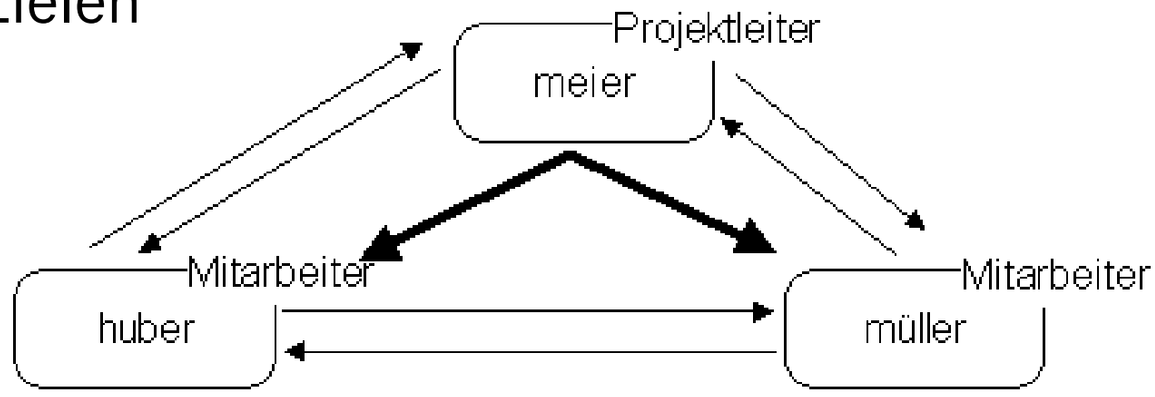
- Bereits HTML bietet zwei Hyperlink-Varianten:
  - Offensichtlich: href
  - Transklutorische Links: Das img-Element
    img verweist auf beliebige, durch URI bezeichnete, Ressourcen.
    Allerdings sind als referenzierte Datentypen nur Bildformate zugelassen
- Verallgemeinerung in Xlink für beliebige Ressourcen:

```
<myElementB
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:href="http://www.w3.org/Icons/w3c_home.gif"
  xlink:actuate="onLoad"
  xlink:show="embed">
  ...</myElementB>
```


Download Java(TM) 2 SDK, Standard Edition, Beta 2 1.4.0 - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://java.sun.com/Download5
What's Related
Case Studies
FTP download
If you are behind a firewall and cannot access our FTP servers, please try the HTTP download below.
Alternate FTP mirror sites:
FTP download from US West I
FTP download from US West II
FTP download from US West III
FTP download from US West IV
FTP download from US Central I
FTP download from US Central II
FTP download from US East I
FTP download from US East II
Document: Done

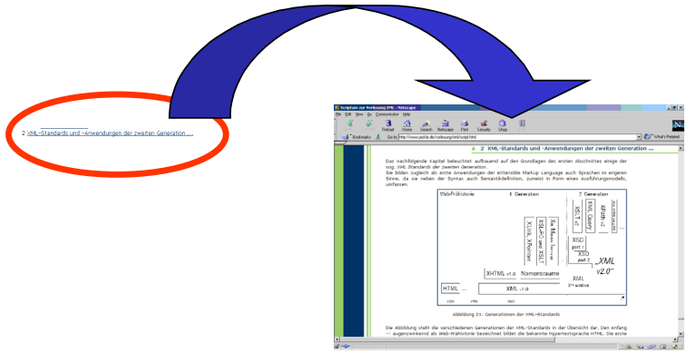
(Dokument-)Verknüpfungen: XML Links

- XLinks zu mehreren Zielen



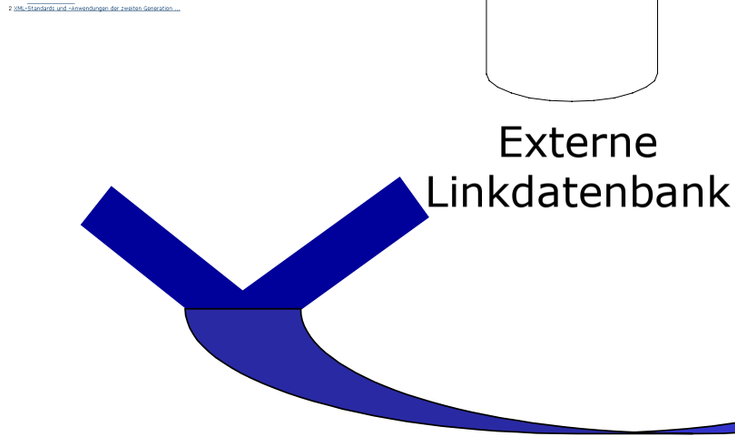

```
<Projekt xmlns:xlink="http://ww
xlink:type="extended">
  <Person xlink:type="locator"
  xlink:href="urn:meier"
  label="ID001"
  xlink:title="Projektleiter"
  xlink:role="http://www.example.com/pl" />
  <Person xlink:type="locator"
  xlink:href="urn:huber"
  label="ID002"
  xlink:title="Mitarbeiter"
  xlink:role="http://www.example.com/emp" />
  <Person xlink:type="locator"
    xlink:href="urn:müller"
    label="ID003"
    xlink:title="Mitarbeiter"
    xlink:role="http://www.example.com/emp" />
  <leader xlink:type="arc" xlink:from="ID001"
    xlink:to="ID002"/>
  <leader xlink:type="arc" xlink:from="ID001"
    xlink:to="ID003"/> </Projekt>
```

(Dokument-)Verknüpfungen: XML Links



Interner Link:

Externer Link:



(Dokument-)Verknüpfungen: XML Links

- Zusammenfassung: *XML Links (XLink)*
  - Führen die Möglichkeit einfacher und komplexer Verweise in die XML-Sprachwelt ein
  - Basiskonzept zur Adressierung und Identifikation ist die URI
  - Orientieren sich konzeptionell an erprobten Lösungen wie: HTML, HyTime, TEI, Dexter, FRESS, OHS, Microcosm, Intermedia
  - Verweise können in XML-Dokumente eingebettet werden, oder in einer separierten *Link Datenbank* versammelt werden

Gliederung

- XML-Standards und -Anwendungen der zweiten Generation ...
  - (Dokument-)Verknüpfungen: XML Links
  - **Die Lokatorsprache XPath**
  - Erzeugung von Präsentationssichten: XML Stylesheets
  - Transformation von XML-Dokumenten: XSL Transformations
  - Die Anfragesprache XQuery

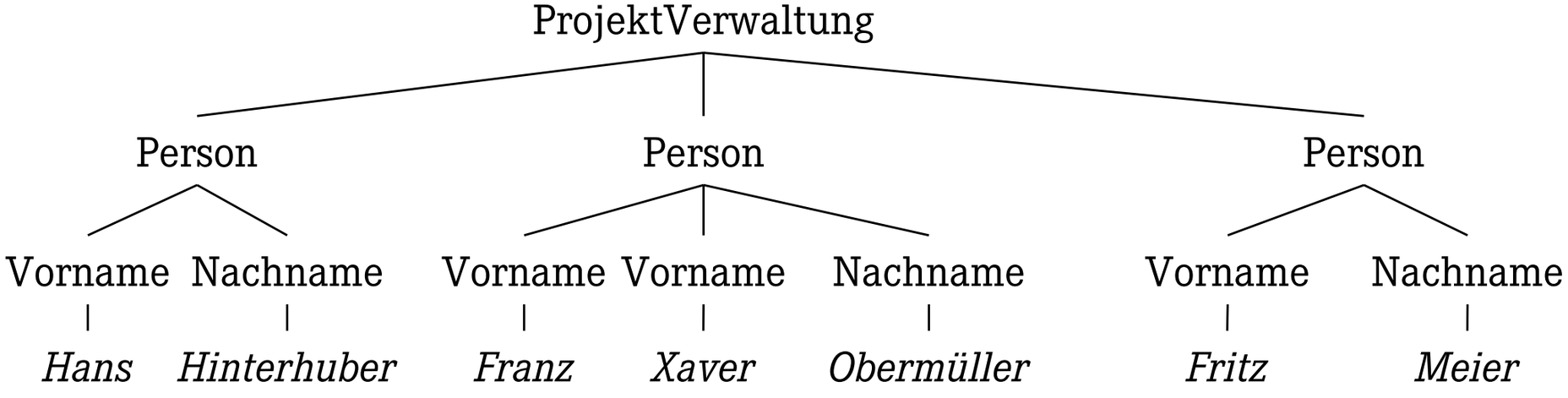
Die Lokatorsprache XPath

- Eigentlich „nur“ Schnittmenge zwischen XSLT und XPointer
- Technisch: Lokatorsprache zur Adressierung beliebiger Knotenmengen eines XML-Dokuments
- Grundlegende Möglichkeiten zur Manipulation von Zahlen, Zeichenketten und Boole‘schen Werten
- Keine XML-Sprache, wohl aber W3C-Recommendation und XML flankierender Standard
- Anfragen werden immer, sofern nicht explizit anders angegeben, relativ zum Kontextknoten ausgewertet

Die Lokatorsprache XPath

- Lokalisierungspfade
  - Hierarchische Struktur wie aus dem InfoSet bekannt
  - Top-down Navigation entlang benannter (typisierter) Knoten

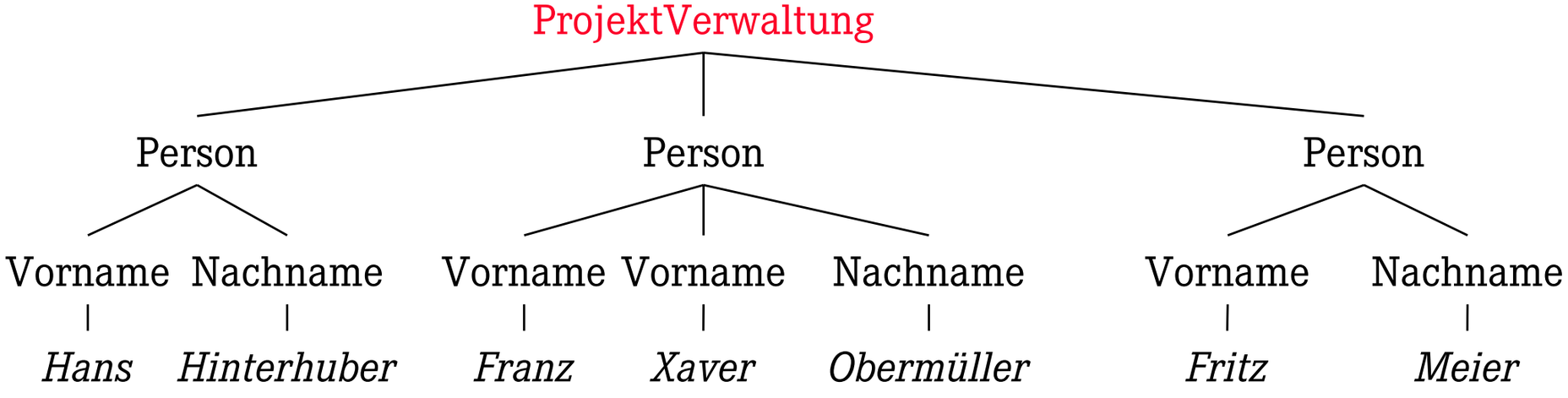
/ProjektVerwaltung/Person/Vorname



Die Lokatorsprache XPath

- Lokalisierungspfade
  - Hierarchische Struktur wie aus dem InfoSet bekannt
  - Top-down Navigation entlang benannter (typisierter) Knoten

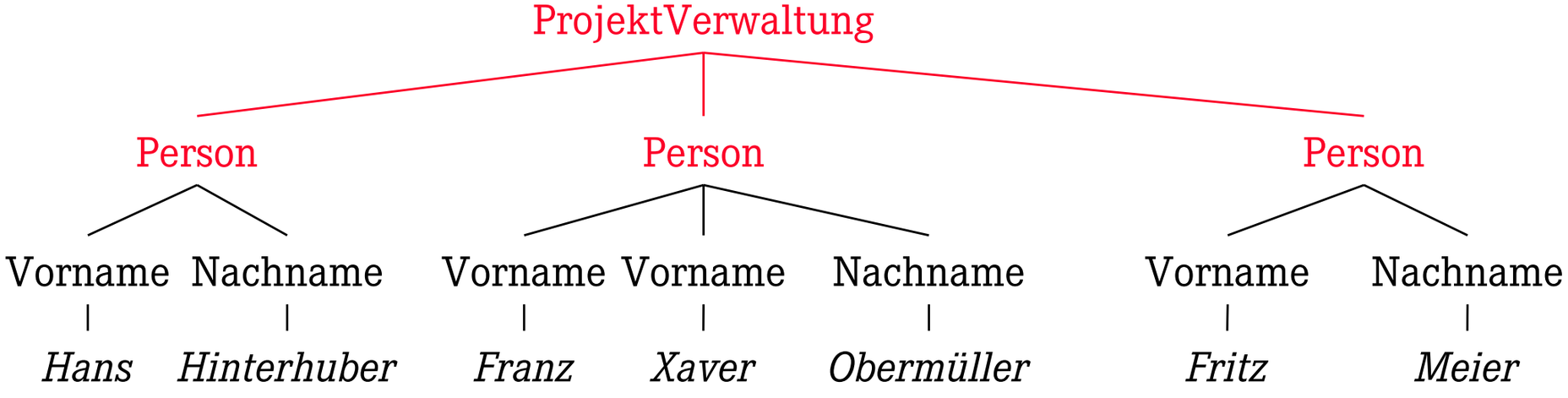
/ProjektVerwaltung/Person/Vorname



Die Lokatorsprache XPath

- Lokalisierungspfade
  - Hierarchische Struktur wie aus dem InfoSet bekannt
  - Top-down Navigation entlang benannter (typisierter) Knoten

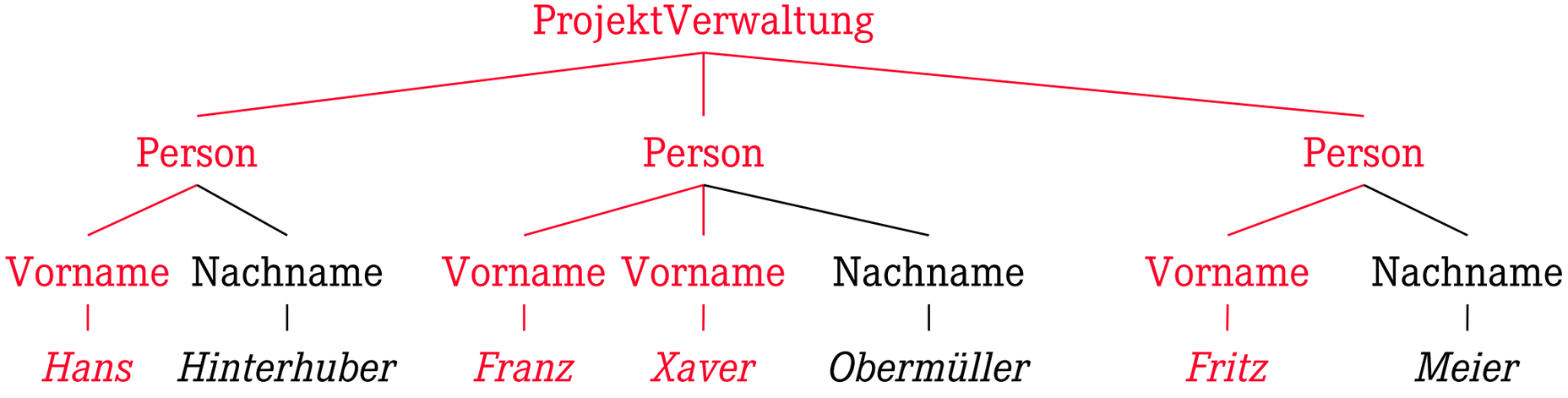
/ProjektVerwaltung/Person/Vorname



Die Lokatorsprache XPath

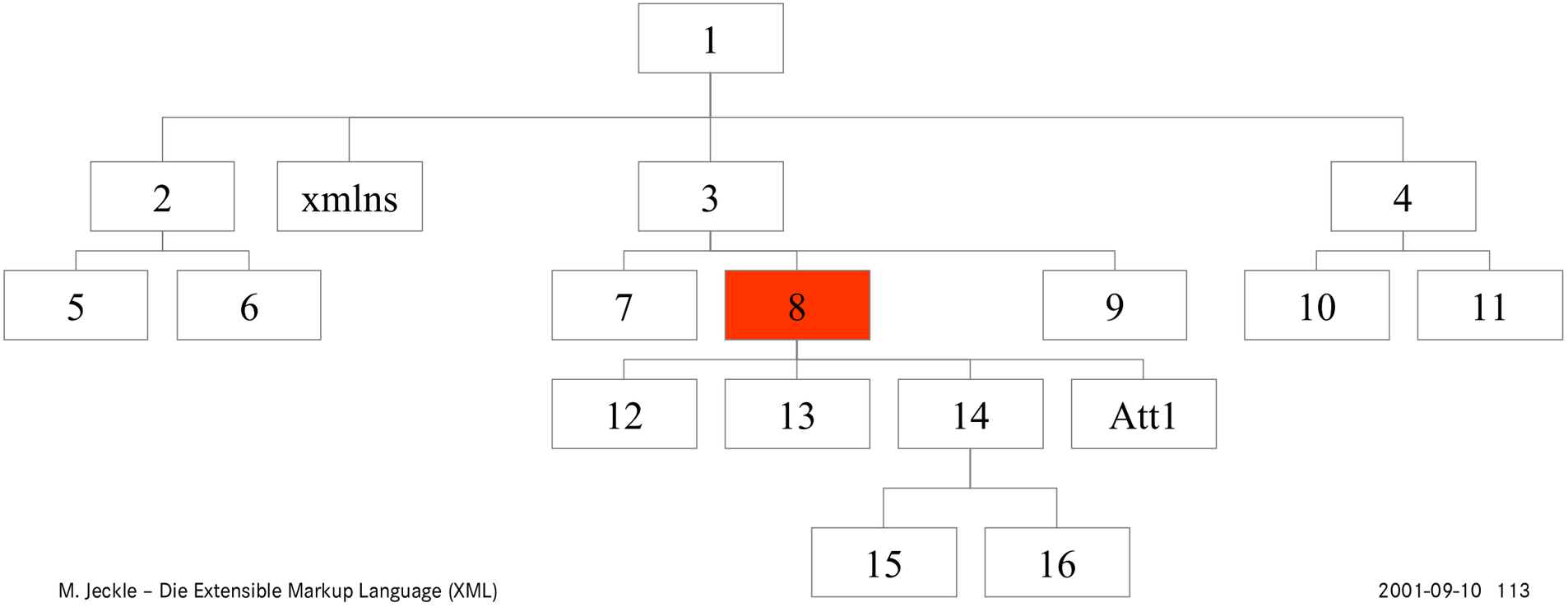
- Lokalisierungspfade
  - Hierarchische Struktur wie aus dem InfoSet bekannt
  - Top-down Navigation entlang benannter (typisierter) Knoten

/ProjektVerwaltung/Person/Vorname



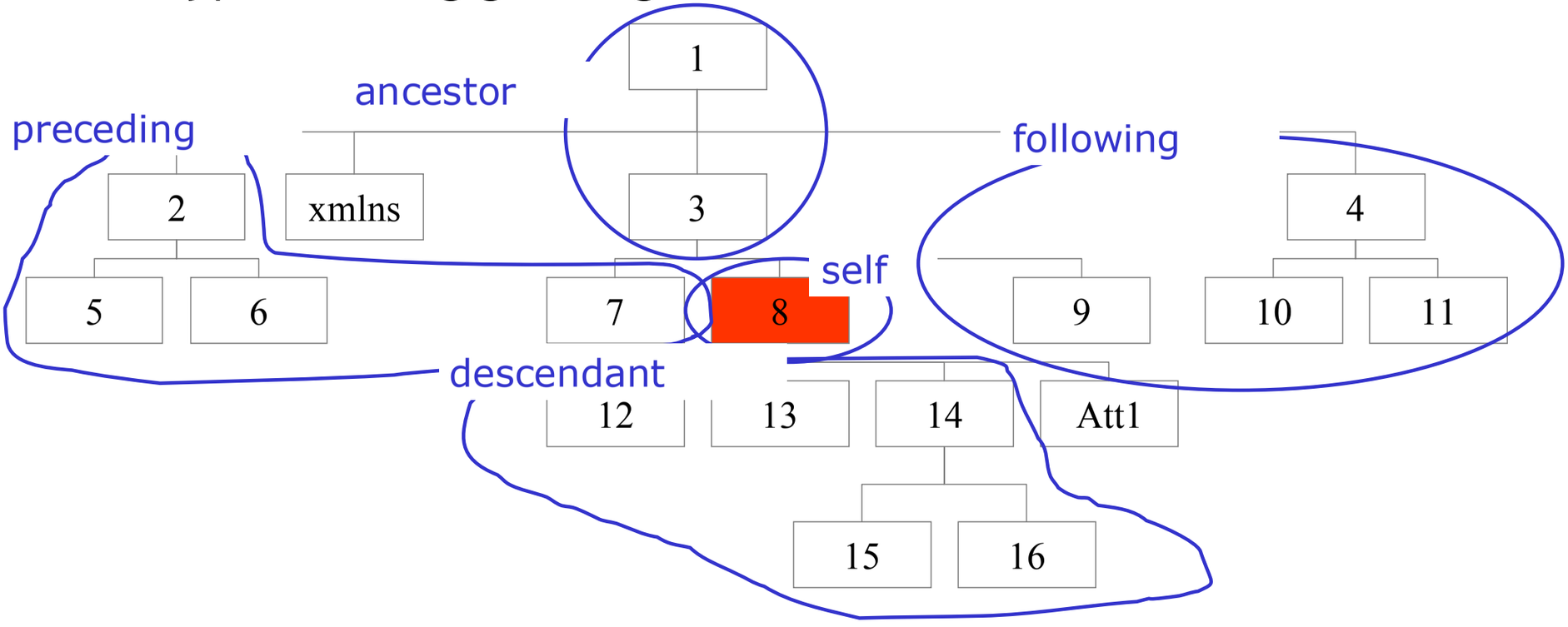
Die Lokatorsprache XPath

- Achsen
  - Ziel: Durchbrechen der rein hierarchisch-absteigenden Traversierung und Einbeziehung *benachbarter* Knoten



Die Lokatorsprache XPath

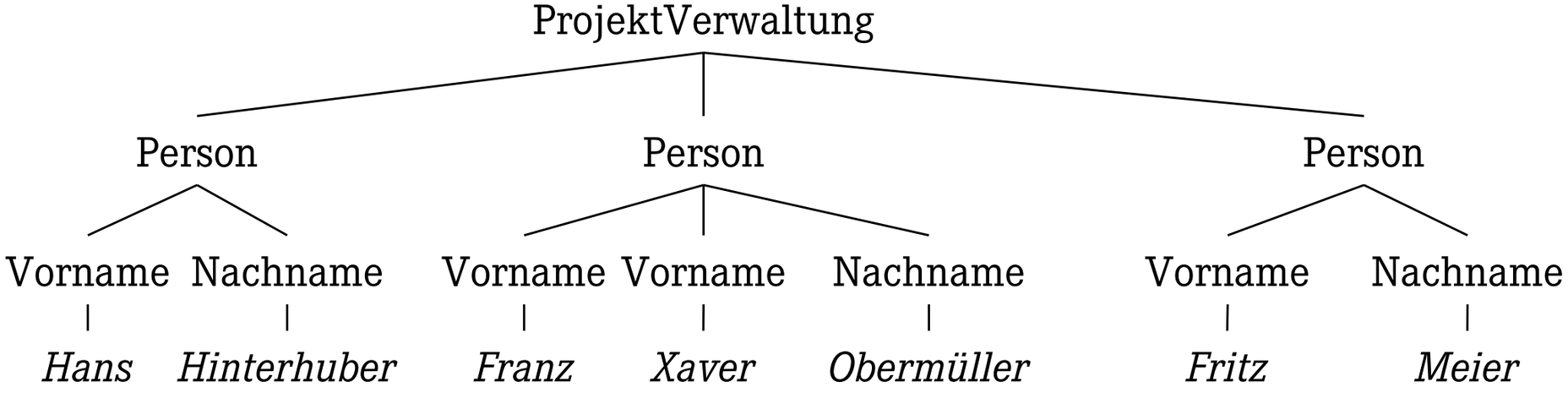
- Achsen
  - Insgesamt existieren 13 verschiedene Achsen zum typunabhängigen Zugriff auf Knoten verschiedenster Position



Die Lokatorsprache XPath

- XPath als Anfragesprache
  - Nutzung der Prädikate

```
//Person[count(Vorname) &gt; 2]/Nachname
```

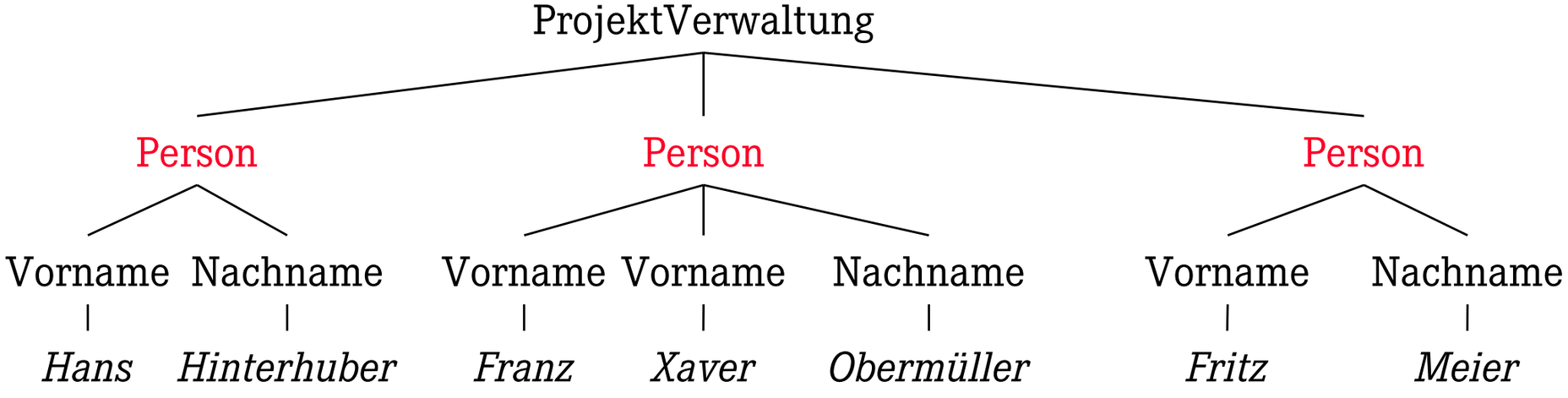


Die Lokatorsprache XPath

- XPath als Anfragesprache
  - Nutzung der Prädikate

```
//Person[count(Vorname) &gt; 2]/Nachname
```

  - Auswahl aller Personen an beliebiger Hierachieposition

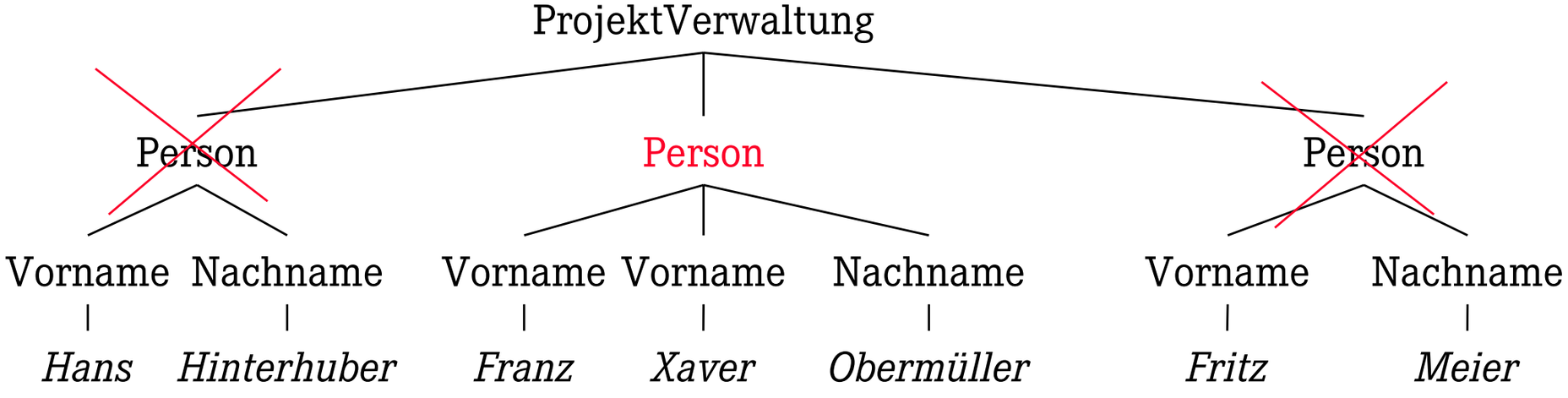


Die Lokatorsprache XPath

- XPath als Anfragesprache
  - Nutzung der Prädikate

```
//Person[count(Vorname) &gt; 2]/Nachname
```

  - Verminderung der Ergebnismenge

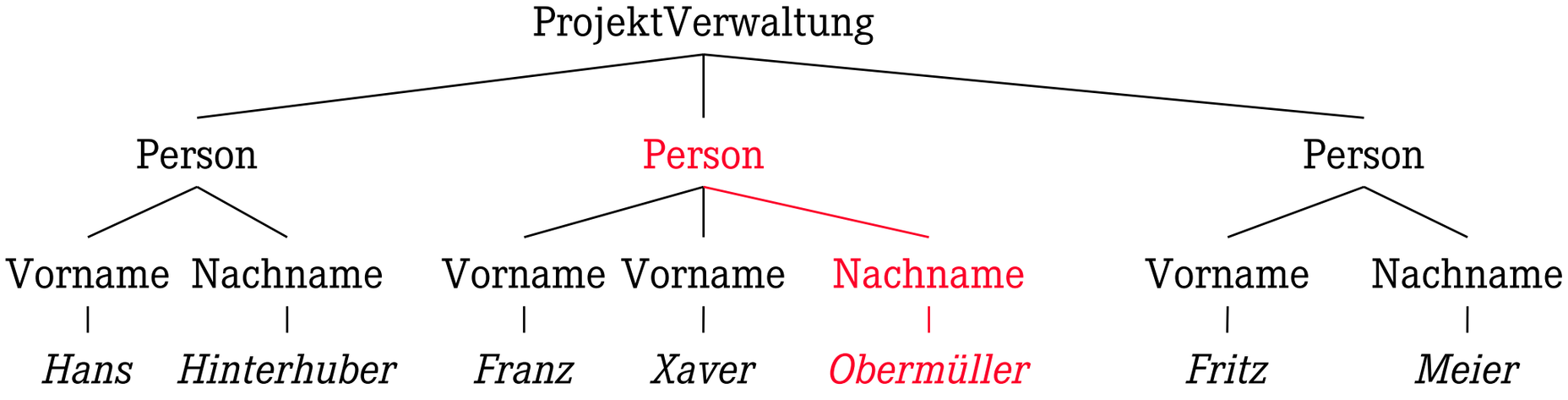


Die Lokatorsprache XPath

- XPath als Anfragesprache
  - Nutzung der Prädikate

```
//Person[count(Vorname) &gt; 2]/Nachname
```

  - (Weiter-)Selektion



Die Lokatorsprache XPath

- Zusammenfassung: *XPath*
  - Mächtige Lokatorsprache mit navigierendem Zugriff
  - Erlaubt die Auswahl beliebiger Knoten(-mengen)
  - Syntax ist String-basiert, nicht XML;
    findet jedoch in den XML-Sprachen XPointer
    und XSLT Verwendung
  - (Teilweise mit etwas Mühe) als Anfragesprache geeignet.
    => XQuery stellt hier eine (bessere) Alternative dar, die
    eine SQL-artige (algebraische) Anfrage auf beliebige XML-
    Dokumente definieren wird

Gliederung

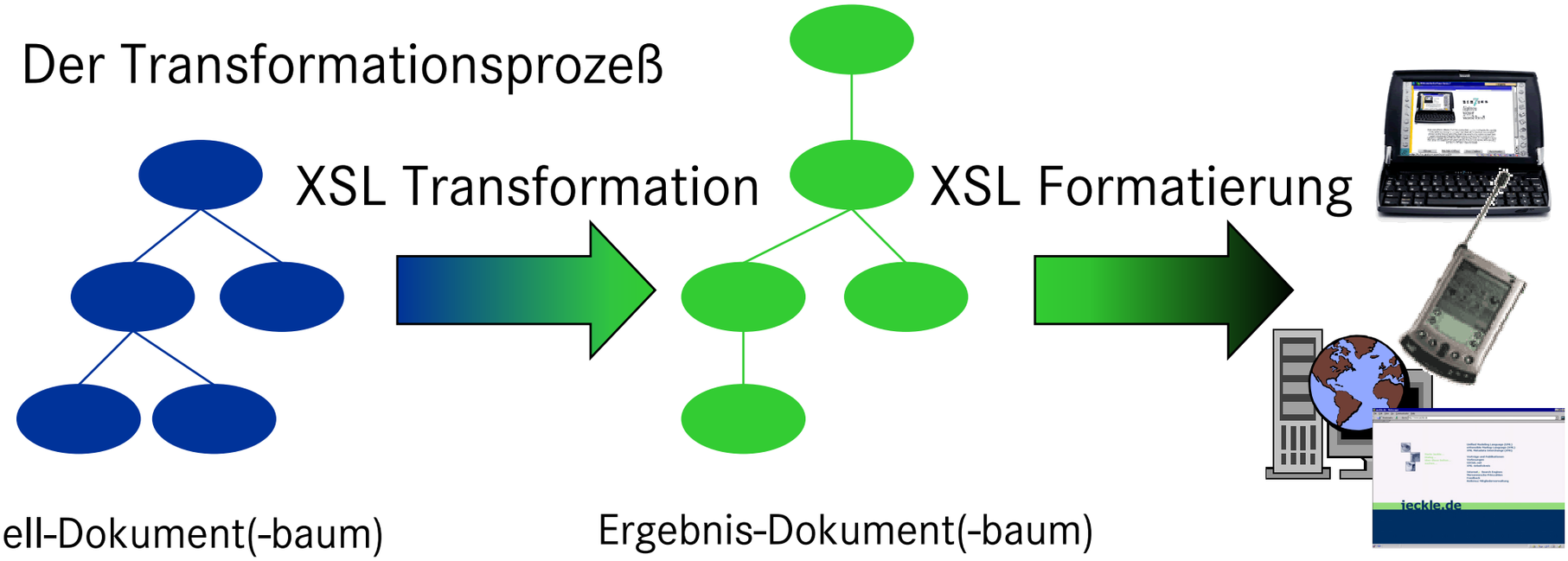
- XML-Standards und -Anwendungen der zweiten Generation ...
  - (Dokument-)Verknüpfungen: XML Links
  - Die Lokatorsprache XPath
  - **Erzeugung von Präsentationssichten: XML Stylesheets**
  - Transformation von XML-Dokumenten: XSL Transformations
  - Die Anfragesprache XQuery

Erzeugung von Präsentationssichten: XML Stylesheets

- Aufgabe: Erzeugung von (beliebigen) Präsentationsdarstellungen aus XML-Dokumenten
- Definiert Vokabular zur Spezifikation von Formattierungs-semantik und Transformationssprache (XSLT) zur Umwandlung von XML-Dokumenten (siehe nächstes Kapitel)
- XSL-Formatierunganweisungen sind Medien-neutral und können für verschiedene Endgeräte unterschiedlich umgesetzt werden
- XSL erweitert die von HTML bekannten *Cascading Style Sheets*

Erzeugung von Präsentationssichten: XML Stylesheets

- Der Transformationsprozeß



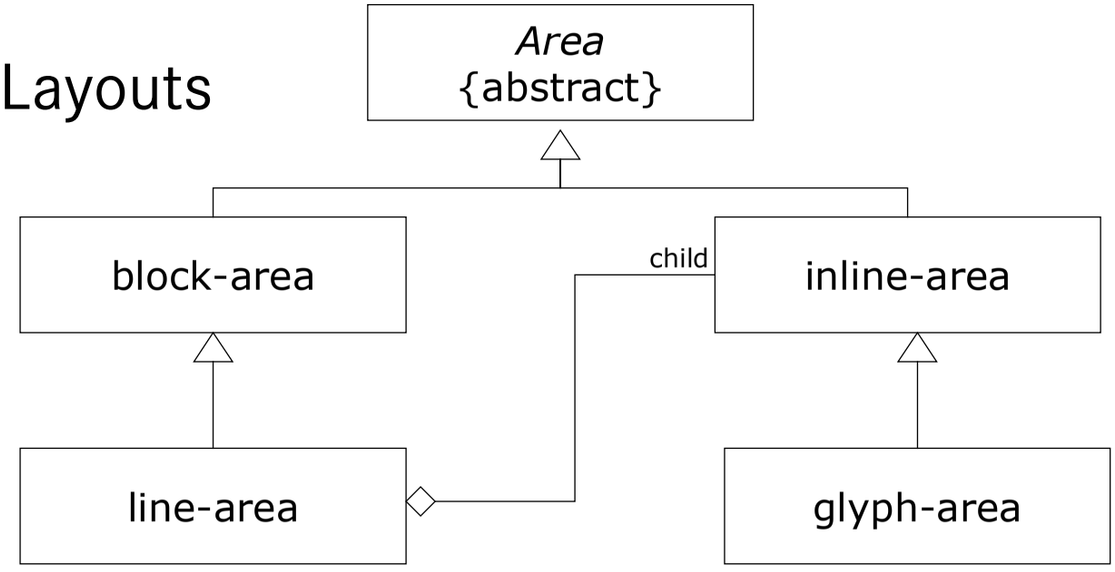
Quell-Dokument(-baum)

Ergebnis-Dokument(-baum)

- XSLT-Prozessor formt XML-strukturierte Eingabe um; Ergebnis: XML-Dokument mit Präsentationsinformation in XML (XSL:FO)
- XSL:FO-Prozessor erzeugt beliebiges (nicht XML-)Ergebnisformat

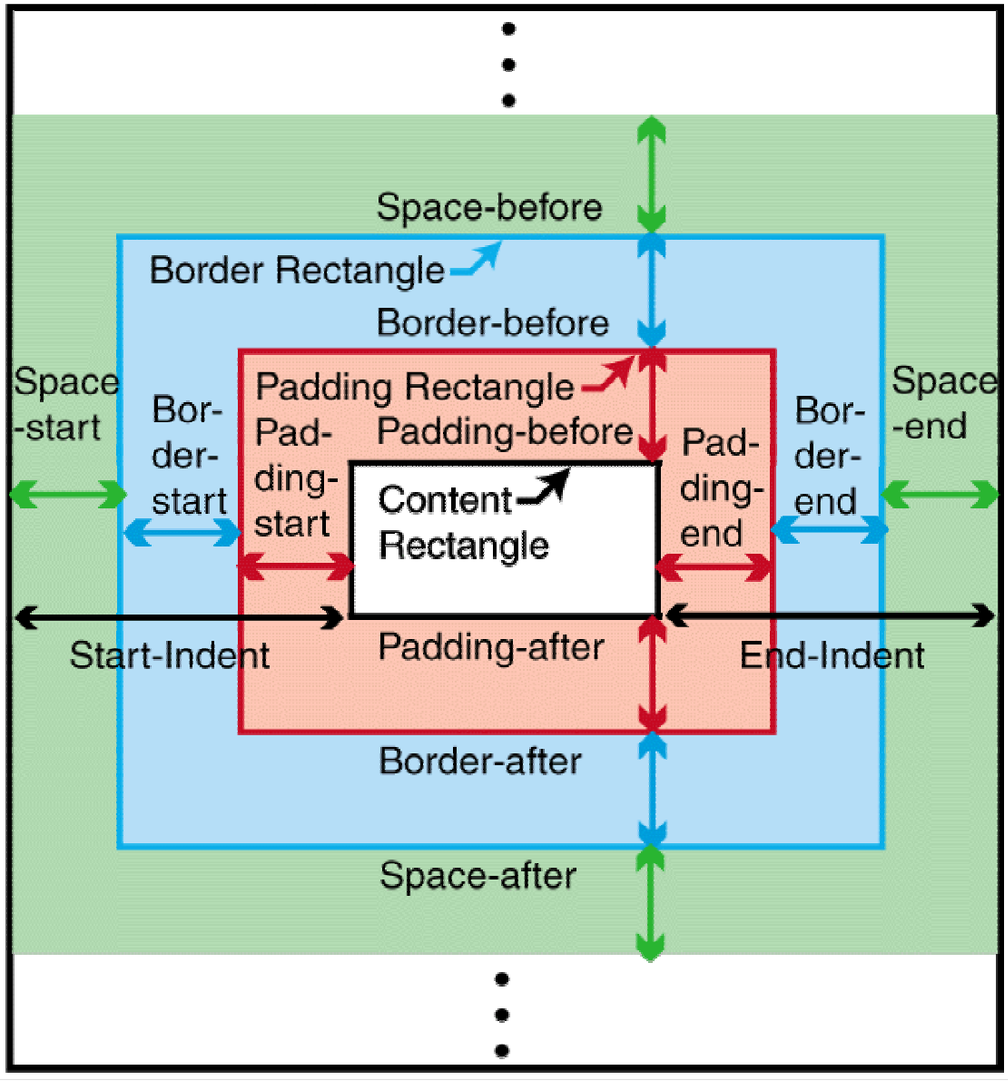
Erzeugung von Präsentationssichten: XML Stylesheets

- Die Grundbausteine des Layouts
  - **block-area:**
    Allgemeinstes Element Rechtwinkliger Layout-Bereich zur Aufnahme von Absätzen, Überschriften etc.

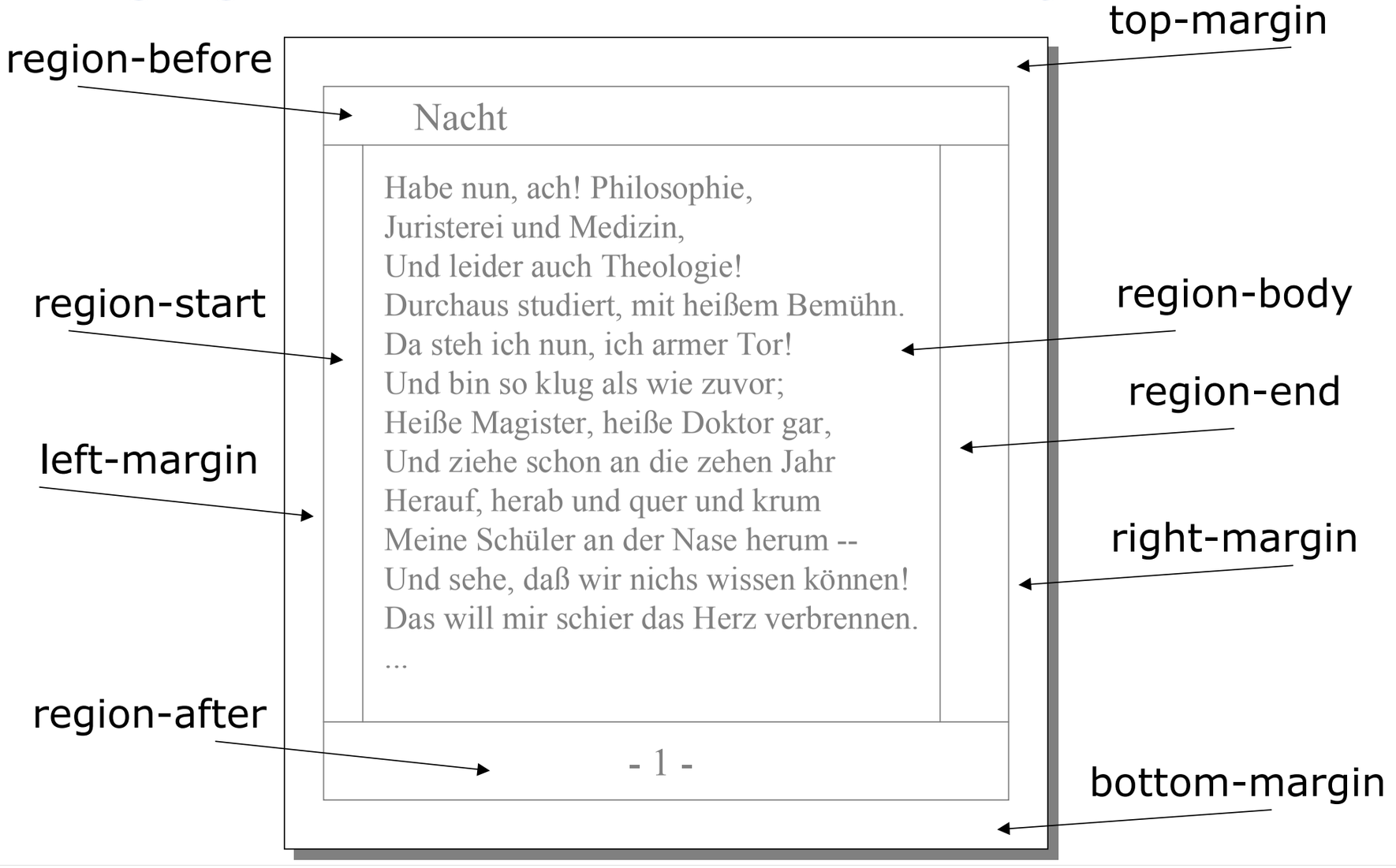


  - **inline-area:** Graphische Hervorhebung einzelner Textbereiche
  - **line-area:** Textbereiche ohne umgebenden Rand
  - **line-area:** Einzelner Buchstabe

Erzeugung von Präsentationssichten: XML Stylesheets



Erzeugung von Präsentationssichten: XML Stylesheets

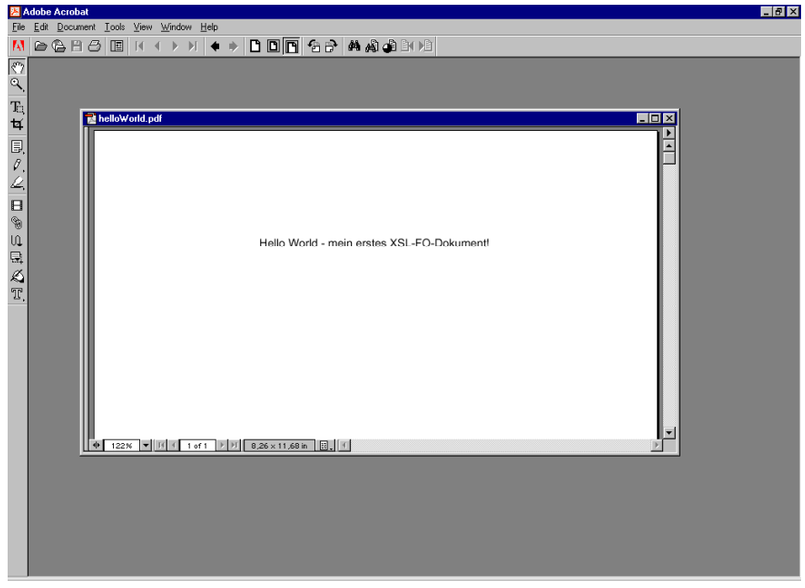


Erzeugung von Präsentationssichten: XML Stylesheets

- Durchgängiges Beispielszenario

```
<?xml version="1.0" encoding="utf-8"?>
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">
  <fo:layout-master-set>
    <fo:simple-page-master master-name="simple"
        page-height="29.7cm"
        page-width="21cm"
        margin-top="1cm"
        margin-bottom="2cm"
        margin-left="2.5cm"
        margin-right="2.5cm">
      <fo:region-body margin-top="3cm"/>
      <fo:region-before extent="3cm"/>
      <fo:region-after extent="1.5cm"/>
    </fo:simple-page-master>
  </fo:layout-master-set>
...
</fo:root>
```

XSL:FO-Prozessor



Quelldokument (XML-FO)

Zieldokument (PDF)

Erzeugung von Präsentationssichten: XML Stylesheets

```
<?xml version="1.0" encoding="utf-8"?>
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">
  <fo:layout-master-set>
    <fo:simple-page-master master-name="simple"
        page-height="29.7cm"
        page-width="21cm"
        margin-top="1cm"
        margin-bottom="2cm"
        margin-left="2.5cm"
        margin-right="2.5cm">
      <fo:region-body margin-top="3cm"/>
      <fo:region-before extent="3cm"/>
      <fo:region-after extent="1.5cm"/>
    </fo:simple-page-master>
  </fo:layout-master-set>
</fo:root>
```

```
<fo:page-sequence master-name="simple">
    <fo:flow flow-name="xsl-region-body">
      <fo:block font-size="12pt"
        font-family="sans-serif"
        line-height="15pt"
        space-after="3pt"
        text-align="center">
          Hello World - mein erstes XSL-FO-Dokument!
      </fo:block>
    </fo:flow>
</fo:page-sequence>
```

Erzeugung von Präsentationssichten: XML Stylesheets

- Internationalisierung von Dokumenten

