

# DAIMLERCHRYSLER

## Datenorientiertes XML mit XML-Schema

Mario Jeckle

DaimlerChrysler Forschungszentrum Ulm

[mario.jeckle@daimlerchrysler.com](mailto:mario.jeckle@daimlerchrysler.com)

[mario@jeckle.de](mailto:mario@jeckle.de)

[www.jeckle.de](http://www.jeckle.de)

# Gliederung

## I Motivation

- Verändertes Einsatzszenario textbasierte Auszeichnungssprachen
- SGML ...

## II SGML-DTD-Mechanismus

- Mächtigkeit und Grenzen
- Erweiterungsoptionen

## III W3Cs XML-Schema

- Typsystem
- Strukturen

## XML

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Datenorientiertes XML mit XML-Schema</Titel>
  <Veranstaltung datum="2001-03-11">
    <Name>Fachgruppentreffen der GI-Fachgruppe
      Datenbanksysteme</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>
```

## XML und XML-Grammatiken (in den 1990er Jahren)

```
<!ELEMENT E-Mail (#PCDATA)>
<!ELEMENT Firma (#PCDATA)>
<!ELEMENT Name (#PCDATA)>
<!ELEMENT Referent (#PCDATA)>
<!ELEMENT Vortrag (#PCDATA)>
<!ELEMENT Titel (#PCDATA)>
<!ELEMENT Datum (#PCDATA)>
<!ELEMENT Firma (#PCDATA)>
<!ELEMENT E-Mail (#PCDATA)>
<!--
<!ATTLIST Vortrag
  Titel CDATA #REQUIRED
  Veranstaltung CDATA #REQUIRED
  Referent CDATA #REQUIRED
-->
```

**Dokument Typ Definition**



```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Datenorientiertes XML mit XMLSchema</Titel>
  <Veranstaltung datum="2002-03-11">
    <Name>Fachgruppentreffen der C...</Name>
  </Veranstaltung>
  <Referent>
    <Name>M...</Name>
    <Firma>...chnology</Firma>
  </Referent>
  <E-Mail>...@daimlerchrysler.com</E-Mail>
</Vortrag>
```

**XML-Dokument**

- Untermenge des SGML-ISO-Standards (8879)
  - Definiert Struktur einer Familie von Dokumenten
  - Möglichkeiten zur Spezifikation erlaubter Inhalte (Typvorrat im wesentlichen Zeichenketten)
  - Aus Sicht XML: proprietäre Syntax
  - Notwendig um, ausgehend von wohlgeformten, *gültige* XML-Dokumente zu erhalten
- 
- Enthält (im wesentlichen) Freitext
  - Unicode codiert
  - Strukturierung durch (frei benennbare) Auszeichnungsmarken
  - Untermenge des SGML-ISO-Standards (8879)
  - Syntax durch XML Recommendation festgelegt
  - Sprach-Semantik nur rudimentär bestimmt, später durch *XML Information Set* definiert
  - Wohlgeformtheits-Regeln emanzipieren XML von strukturiertem Text

# XML-Grammatiken (in den 1990er Jahren)

## DTD-Charakteristika

- Festlegung der Dokumentstruktur
  - Elemente
  - Attribute
  - Notationen
  - Entitäten (Textmakro auf Dokumentebene)
    - Parametrisierte Entitäten (Textmakro auf DTD-Ebene)
- Festlegung des Inhaltsmodells
  - Datentypen
    - Zeichenketten-artig
    - Vorgabewerte
    - Auswahltypen
  - Schlüssel/Referenzen

## XML-Grammatiken (in den 1990er Jahren)

### DTD-Charakteristika

- Baumartige hierarchische Organisationsform
- *Die-ganze-Welt-ist-ein-Dokument-Sicht*  
(von SGML übernommen)
  - ... mit entsprechenden Auswirkungen auf Referenzen
- Keine (praktisch verwendbare) Unterstützung für Namensräume
- DTD ist keine XML-Sprache
  - => zusätzliche Werkzeuge notwendig
- es existiert keine DTD für DTDs
  - => Grammatik kann nicht validiert werden
- Angebotene Datentypen unzureichend
- Angebotene Strukturprimitiven unzureichend
  - => Notwendige Konstrukte zum Ausdruck mächtigerer Semantik müssen aufwendig und proprietär realisiert werden

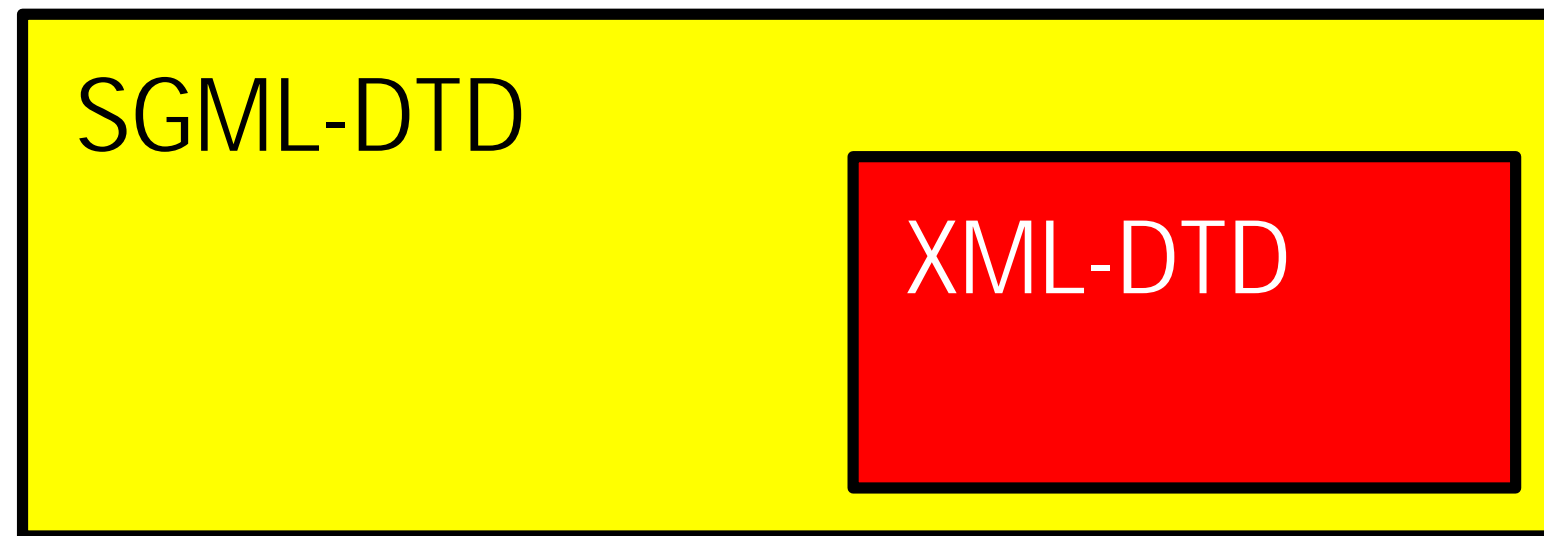
## XML-Grammatiken (in den 1990er Jahren)

### DTD-Charakteristika

- Strukturell
  - Namensraum-Unterstützung
  - Etablierung eines Typ-Begriffs
  - Steuerung der Auftretensreihenfolge
  - Vererbung
  - Wiederverwendungs-Unterstützung
  - Erweiterter Referenzierungsmechanismus
- Inhaltlich
  - Klassische Primitivtypen (z.B. int, float, boolean)
  - Erweiterte Basistypen (z.B. date)
  - Binarstrukturen (z.B. base64)
  - Komplexe Datentypen
  - Erweiterbares Typsystem
  - Unterstützung konsistenzgarantierender Einschränkungen

## XML-Grammatiken (in den 1990er Jahren)

Weiterentwicklungsoptionen für die Document Type Definition

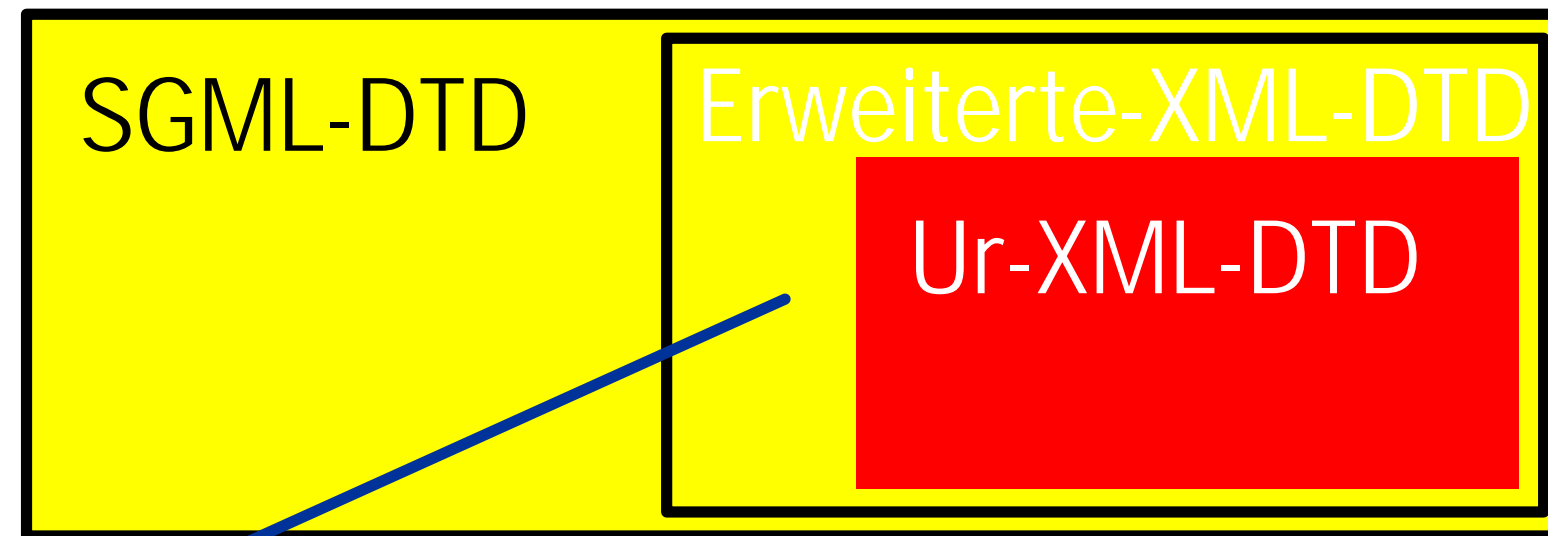


XML-DTD bildet syntaktisch und hinsichtlich der Mächtigkeit eine echte Untermenge des SGML-DTD-Mechanismus



## XML-Grammatiken (in den 1990er Jahren)

Weiterentwicklungsoptionen für die Document Type Definition

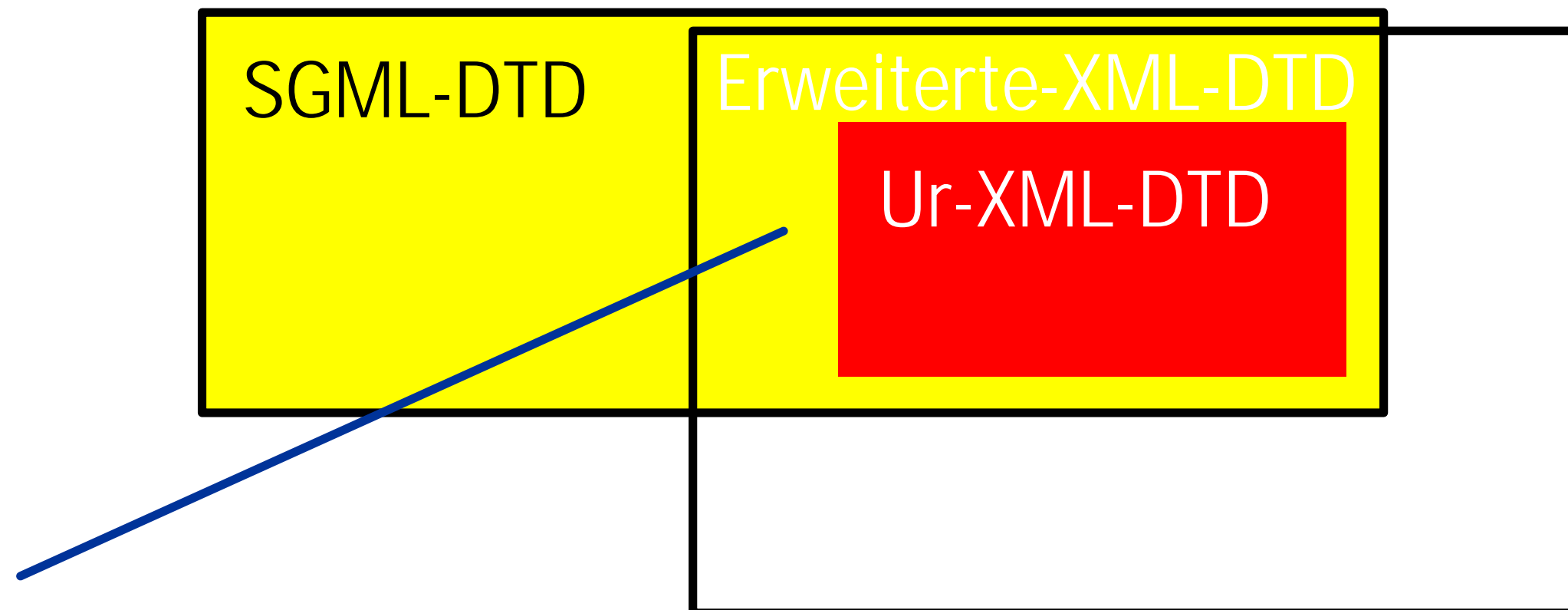


Erweiterung der XML-DTD-Mächtigkeit um weitere (bereits existierende) Primitive des SGML-DTD-Mechanismus:

- (+) (Unbestreitbarer) Mächtigkeitsgewinn
- (-) (unvermeidlicher) Komplexitätszuwachs
- (-) SGML-Mächtigkeit bildet natürliche Obergrenze

## XML-Grammatiken (in den 1990er Jahren)

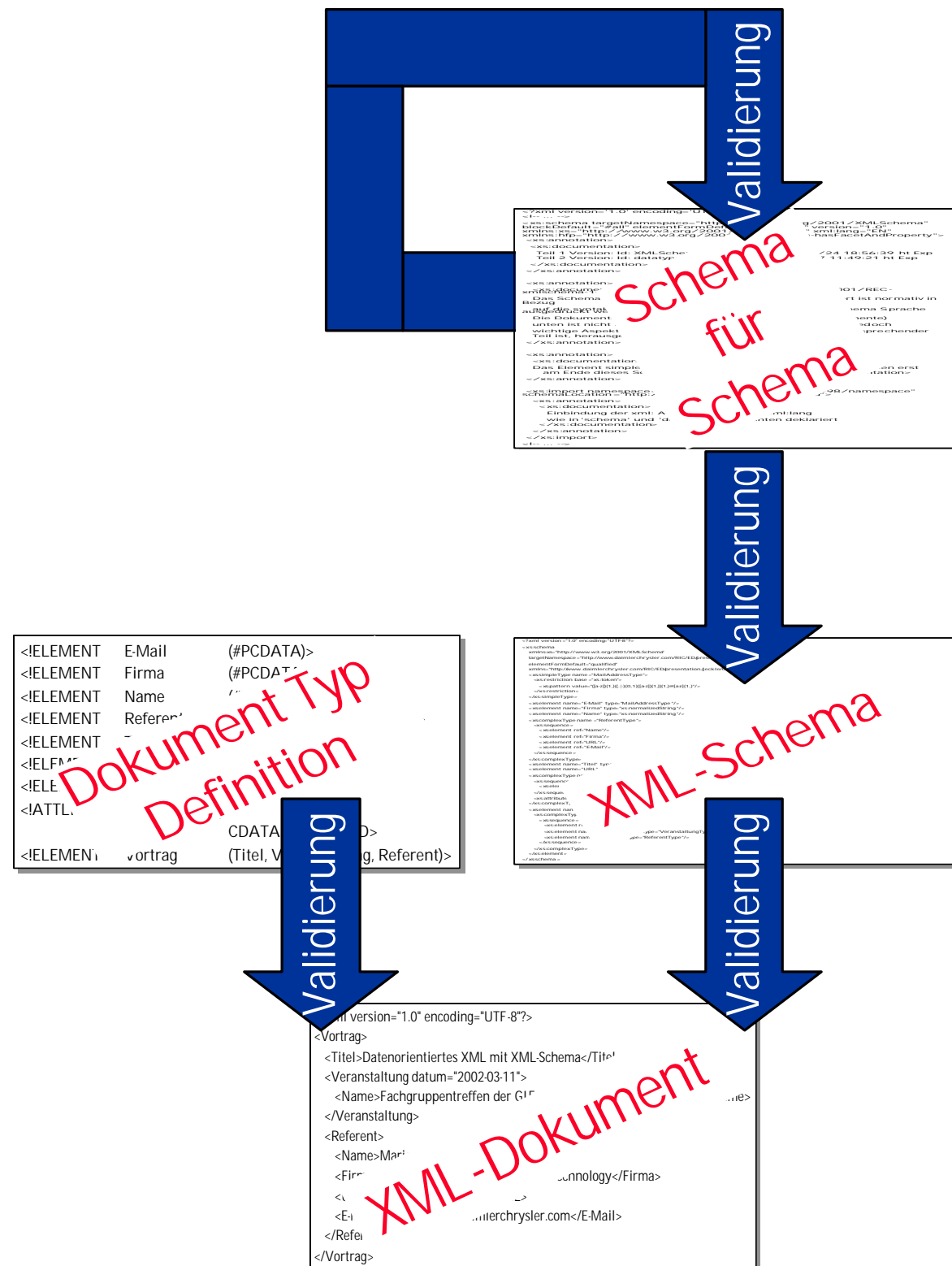
Weiterentwicklungsoptionen für die Document Type Definition



Erweiterung der XML-DTD-Mächtigkeit um Primitive die über den SGML-DTD-Mechanismus hinausgehen:

- (+) Freiheitsgrad hinsichtlich beliebiger Erweiterungen
- (-) XML-Grammatiken keine Untermenge von SGML mehr

# W3Cs XML-Schema



- Grammatik für beliebige XML-Vokabulare
- Part1 beschreibt Struktur und Inhaltseinschränkungen
- Part2 beschreibt Typsystem
- Signifikante Erweiterung des DTD-Funktionsumfanges
- Als XML-Sprache realisiert
- Integriert verschiedene Vorgängeransätze
- W3C Recommendation
- Grundlage der XML-Sprachen der 2. Generation
- Werkzeugunterstützung verfügbar
- Erster Schritt der Schemabestrebungen, weitere werden folgen ...

## W3Cs XML-Schema -- Mächtigkeit

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

# W3Cs XML-Schema -- Datentypen

Atomare Typen

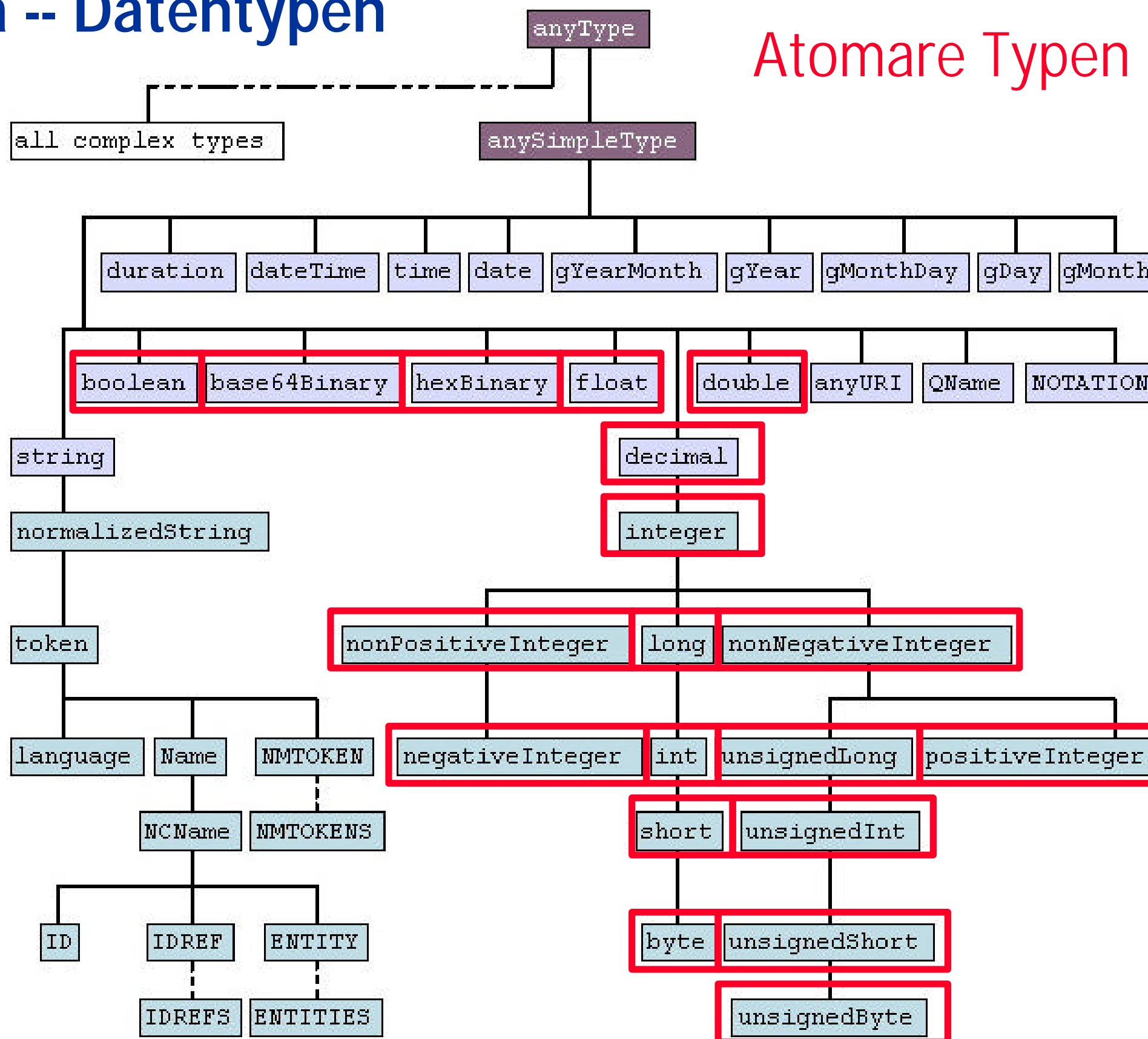
■ Ur-typ

■ Vordefinierter Primitivtyp

■ Vordefinierter abgeleiteter Typ

— Typeinschränkung

----- Aggregierter Typ



Aus: XML-Schema Part 2: Datatypes, W3C-Recommendation 2001-05-02, chap. 3

## W3Cs XML-Schema -- Primitivtypen

Typ	XML - Schema	Microsoft CLS	Java	SQL92/SQL3
$-2^7 \leq t \leq 2^7$	byte	-	byte	tinyint
$0 \leq t \leq 2^8 - 1$	unsignedByte	byte	-	
$-2^{15} \leq t \leq 2^{15} - 1$	short	short	short	smallint
$-2^{31} \leq t \leq 2^{31} - 1$	int	int	int	int, integer
$-2^{63} \leq t \leq 2^{63} - 1$	long	long	long	byte
Unicode-Zeichen(kette)	string	string	String	clob, char, character(n), varchar(n)
IEEE754	float double	float double	float double	float, real, double
{true,false,0,1}	boolean	bool	boolean	boolean
N	decimal	decimal	BigDecimal	numeric(p,s), dec, decimal(p,s)

# W3Cs XML-Schema – anwenderdefinierte Typen

## Komplexe Typen

```
<xs:complexType name="ReferentType">
  <xs:sequence>
    <xs:element ref="Name"/>
    <xs:element ref="Firma"/>
  </xs:sequence>
</xs:complexType>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Datenorientiertes XML mit XML-Schema</Titel>
  <Veranstaltung datum="2001-03-11">
    <Name>Fachgruppentreffen der GI-Fachgruppe
      Datenbanksysteme</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>
```

```
<xs:element name="Firma" type="xs:normalizedString"/>
<xs:element name="Name" type="xs:normalizedString"/>
```

## Benannte komplexe Typen:

- Verwendungsunabhängige Inhaltsmodellformulierung
- Basis der Wiederverwendung

# W3Cs XML-Schema – anwenderdefinierte Typen

## Typerweiterung

```
<xs:complexType name="ReferentTypeExt">
  <xs:complexContent>
    <xs:extension base="ReferentType">
      <xs:sequence>
        <xs:element
          name="E-Mail-Adresse"
          type="MailAddressType"/>
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Datenorientiertes XML mit XML-Schema</Titel>
  <Veranstaltung datum="2001-03-11">
    <Name>Fachgruppentreffen der GI-Fachgruppe
      Datenbanksysteme</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>
```

## Spezialisierung komplexer Typen:

- Ausschließlich extensionsional
- Keine Substitutionssemantik



# W3Cs XML-Schema – anwenderdefinierte Typen

## Definition eigener Primitivtypen

```
<xs:simpleType
  name="MailAddressType">
  <xs:restriction base="xs:token">
    <xs:pattern
      value="([a-z]){1,}((.|-){0,1}([a-z]){1,}){1,}@([a-z]{1,})"/>
    </xs:restriction>
  </xs:simpleType>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Datenorientiertes XML mit XML-Schema</Titel>
  <Veranstaltung datum="2001-03-11">
    <Name>Fachgruppentreffen der GI-Fachgruppe
      Datenbanksysteme</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>
```

## Definition atomarer Basistypen:

- Als parametergesteuerte Einschränkung bestimmter Eigenschaften
- Durch reguläre Ausdrücke