

DAIMLERCHRYSLER

W3C's XML Schema -- Ein neuer Beschreibungsmechanismus für (Meta-) Daten- Strukturen

Mario Jeckle

DaimlerChrysler Research and Technology
Dept. product development process chain (FT3/EK)

www.jeckle.de
mario.jeckle@daimlerchrysler.com
mario@jeckle.de

Informatik Kolloquium Universität Würzburg – 2000-06-26

- Einführung
 - Strukturelle Eigenschaften von XML
 - Familie der XML-Sprachen

- Der Mechanismus der *Document Type Definition* zur XML-Sprachbeschreibung
 - Mächtigkeit
 - Grenzen
 - Notwendige Erweiterungen

- XML Schema
 - Notwendigkeit
 - Anforderungen
 - Ansätze
 - W3C's *XML Schema Part 1 & Part 2*

Strukturelle Eigenschaften der eXtensible Markup Language

Start Tag

```
<vortrag datum="2000-06-26">
```

Tag Name

Attribut Name

Attribut Wert

Attribut

W3C's XML Schema –
Ein neuer Beschreibungsmechanismus für
(Meta-) Daten-Strukturen

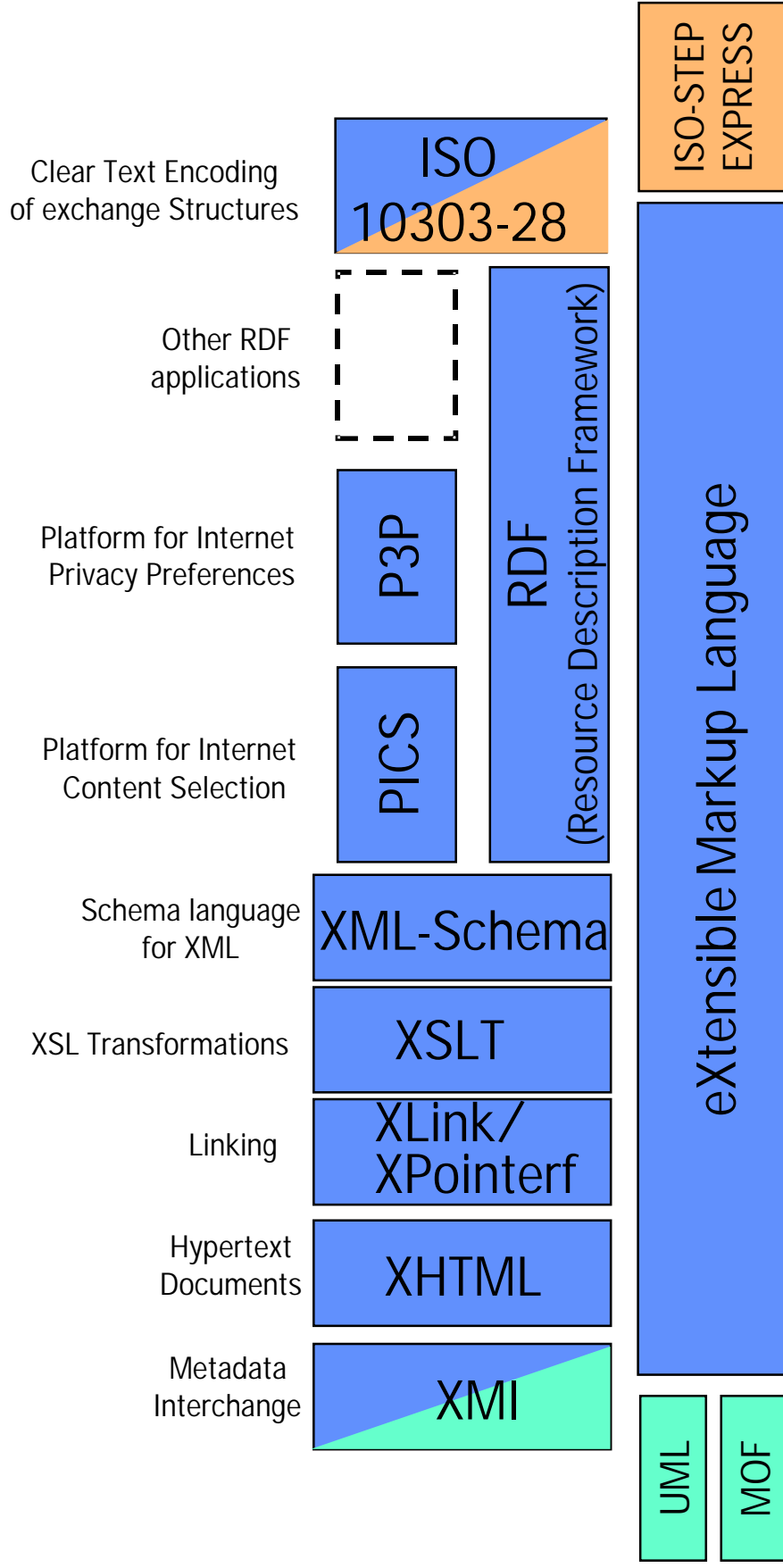
```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

End Tag

Element

Die Familie der XML-Sprachen



Ein XML-Dokument...

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<!DOCTYPE vortrag SYSTEM "vortrag.dtd">
```

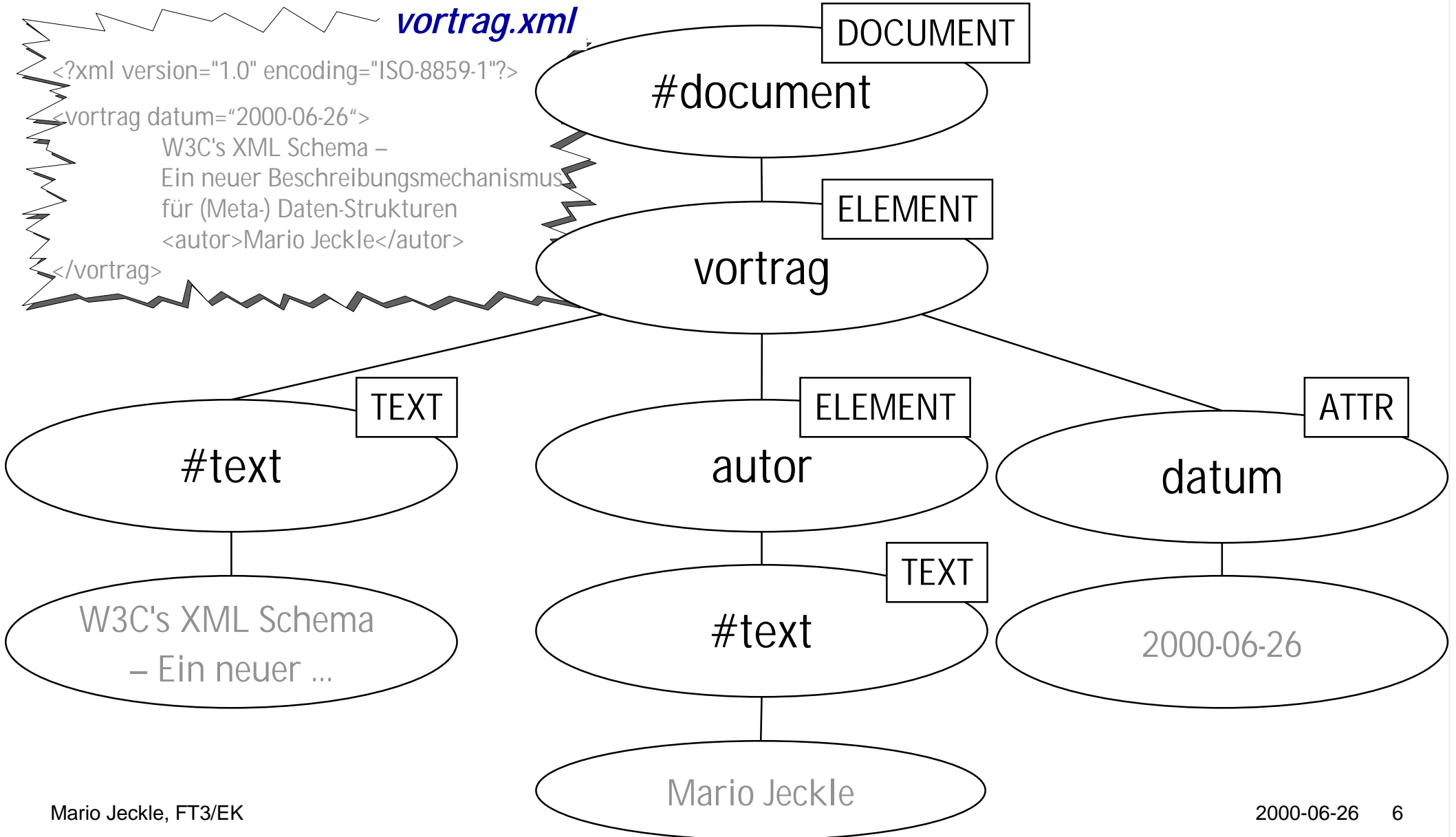
```
<vortrag datum="2000-06-26">
```

W3C's XML Schema – Ein neuer
Beschreibungsmechanismus
für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

Hierarchische Struktur eines XML-Dokuments



Die *Document Type Definition*

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<vortrag datum="2000-06-26">
```

W3C's XML Schema – Ein neuer Beschreibungsmechanismus
für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
    datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```

Die *Document Type Definition*

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<vortrag datum="2000-06-26">
```

W3C's XML Schema – Ein neuer Beschreibungsmechanismus

für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
  datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```


Die *Document Type Definition*

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<vortrag datum="2000-06-26">
```

W3C's XML Schema – Ein neuer Beschreibungsmechanismus
für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
    datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```

Der DTD-Mechanismus

- streng hierarchisch
- Im Wesentlichen *ELEMENTS* und *ATTRIBUTES* als Knotentypen
- Keine Datentypen (abgesehen von CHAR-Data)
- Rudimentärer Referenzierungsmechanismus (ID(S), IDREF(S))
- Selektionstyp
- Vorgabewerte
- Dokumentstruktur u.U. nur eingeschränkt ausdrückbar
- Dokumentenzentriert
- DTD ist nicht XML

Grenzen der DTD und notwendige Erweiterungen

Structural schemas

Mächtigkeit annähernd ähnlich des bestehenden DTD-Mechanismus um Dokumentstruktur (Reihenfolge, Auftrettsvielfachheit von Elementen und Attribute) zu beschreiben.

Insbesondere sollen folgende Erweiterungen verwirklicht werden:

- Namespace Integration
- Definition von Einschränkungen für Elementinhalte
- Integration Strukturschema und primitive Datentypen
- Vererbung: DTD unterstützt nur *part-of*-Beziehungen
- Erweiterter Referenzierungsmechanismus (URI)
- Kommentare (*human* und *machine readable*)
- Nutzung der bestehenden XML Technologie

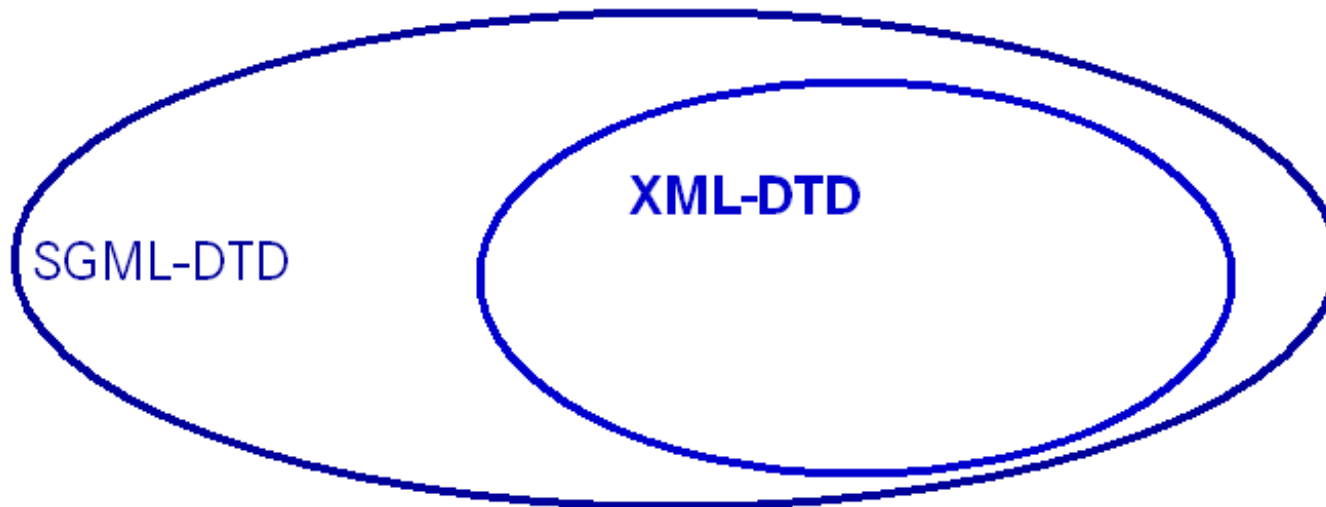
Grenzen der DTD und notwendige Erweiterungen

Primitive Datentypen ...als Ersatz für #PCDATA

- „klassische“ atomare Datentypen, ergänzt um SQL-artige, wie *integer*, *date*.
- Programmiersprachen-übliche (typischerweise Java-artige) *build-in types*
- uninterpretierte Binärstrukturen
- (durch Anwender) erweiterbares Typsystem
- lexikalische Definitionen
- Einschränkungen an Typen

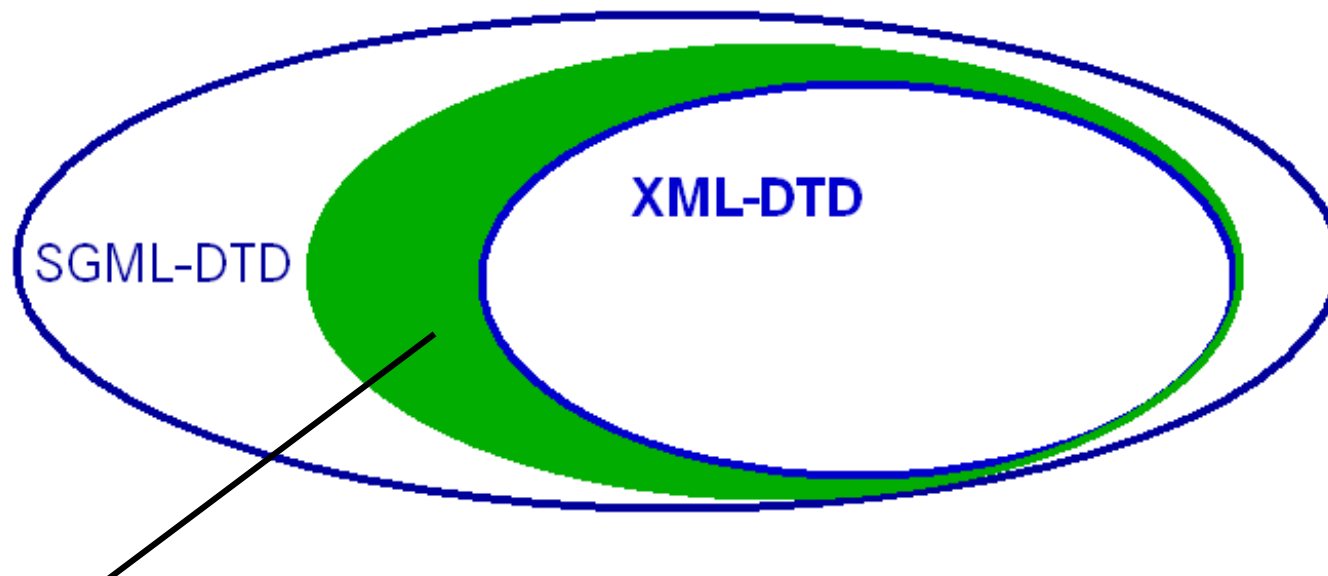
Optionen bei der Definition eines XML-Schemamechanismus

Status quo



Ausdrucksmächtigkeit der XML-DTD bildet eine (echte) Untermenge des SGML-DTD Mechanismus.

Optionen bei der Definition eines XML-Schemamechanismus



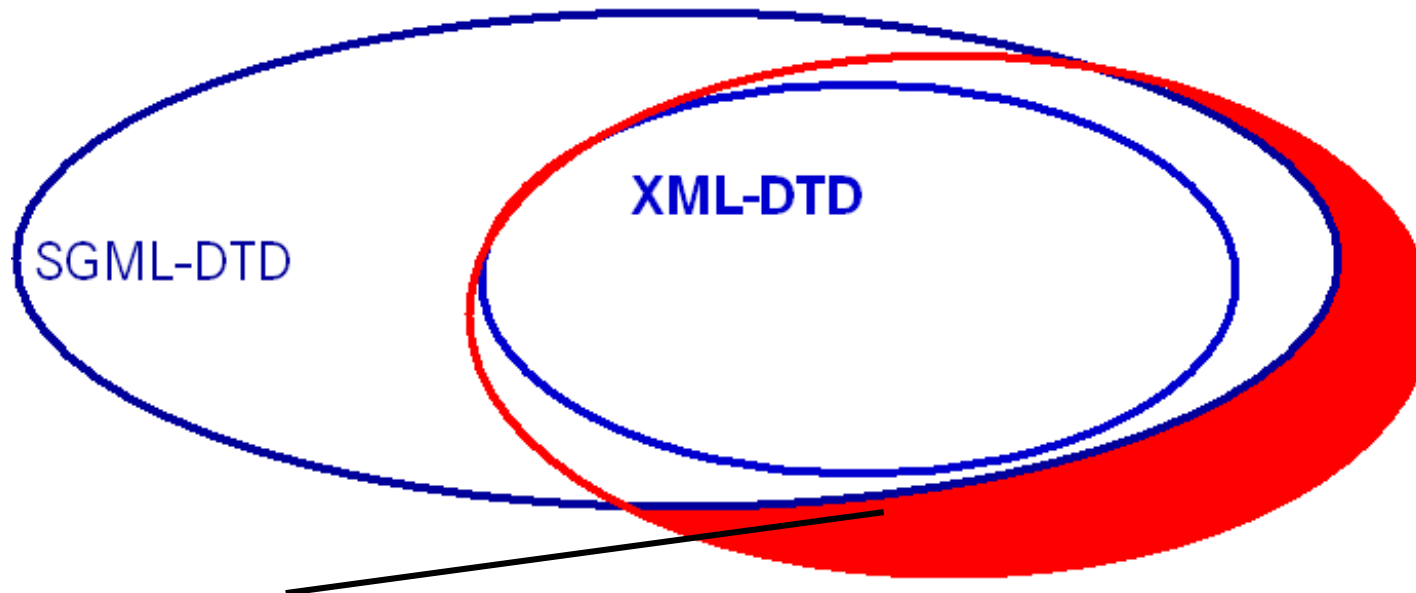
Erweiterung des XML-DTD Mechanismus um weitere Elemente der SGML-DTD.

(+) Ausdrucksmächtigkeit nähert sich (wieder) der von SGML an

(-) ...die Komplexität auch

(-) Ausdrucksmächtigkeit kann die von SGML niemals übertreffen

Optionen bei der Definition eines XML-Schemamechanismus



Erweiterung des XML-DTD Mechanismus um Elemente, die *nicht* mit SGML-Mitteln ausdrückbar sind.

- (+) Freiheitsgrad hinsichtlich beliebiger Erweiterungen
- (-) XML-Grundforderung nach Untermengenbeziehung zu SGML entfällt
- (-) immernoch zwei verschiedene Sprachen für Inhalt und Schema

Existierende Ansätze

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)

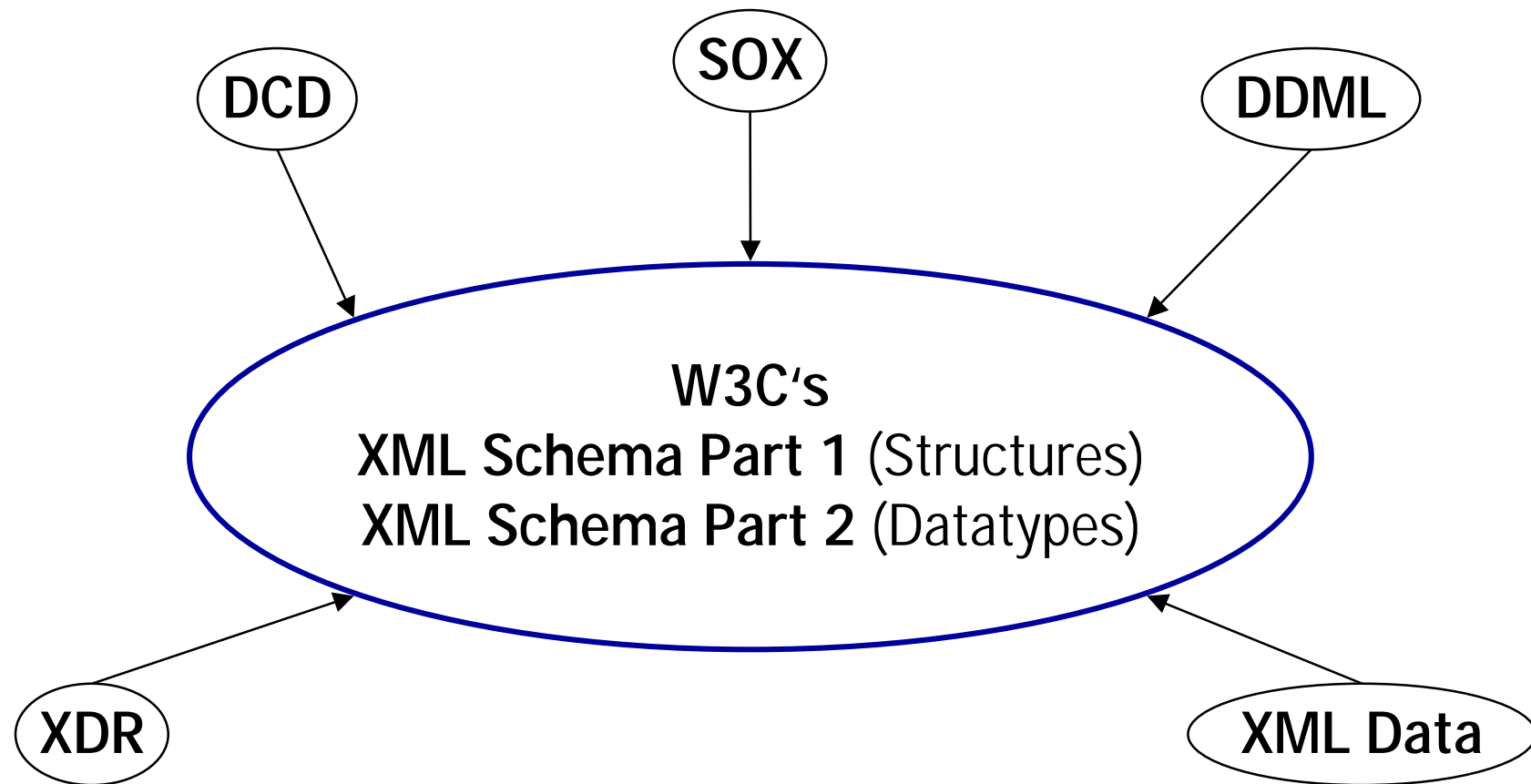
Inspiziert durch XML-API-Entwicklung

- Schema for Object oriented XML (SOX)

XML-Sprachen zur Inhaltsbeschreibung

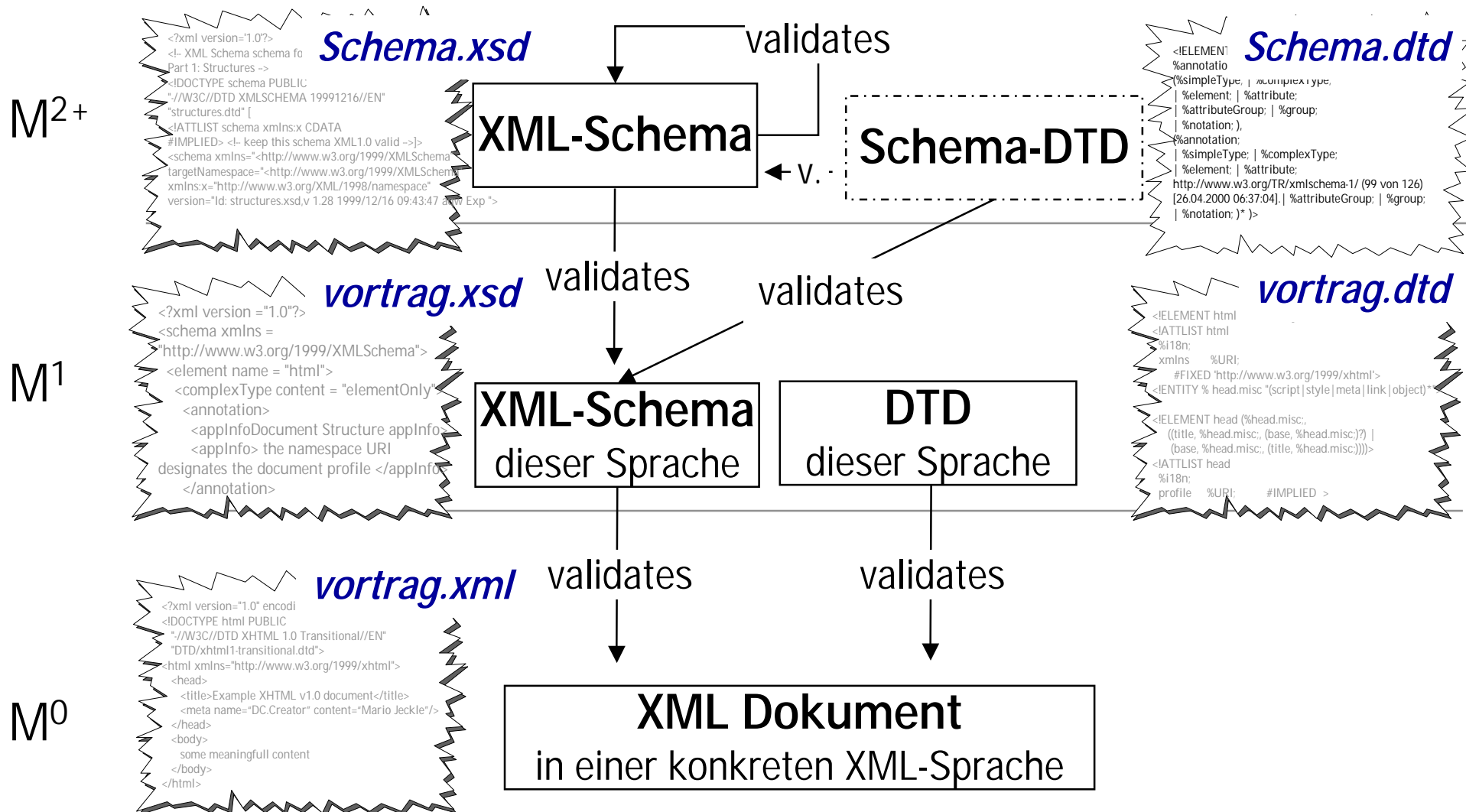
- Document Definition Markup Language/XSchema (DDML)
- Schematron (XSL-basierte Auswertung der Dokumentstruktur)
- XML-Data/XML-Data Reduced (XDR) *(erster Ansatz noch vor Verabschiedung XML 1.0)*
- Document Structure Description (DSD)
- **W3C's XML-Schema**

W3C's XML Schema



W3C's XML-Schema konsolidiert und integriert die wesentlichen existierenden Ansätze und entwickelt sie fort.

W3C's XML Schema Technologie -- Metamodellierung



Inhaltliche Charakteristika W3C's XML Schema

Gegenüber DTD ...

- stark erweiterte build-in Typen
 - Basistypen angelehnt an ISO 11404, SQL und Java
 - Datums- und Zeittypen angelehnt an ISO 8601
- Erweiterbares Typsystem
 - Typamplifikation und -restriktion
 - lexikalische Definition
- Typbeziehungen (äquivalente Typen)
- „echte“ XML-Sprache
- Explizite Beziehungskardinalität
- Kontextabhängige Eindeutigkeit (Schlüssel und Referenzen)
- Elemente gleichen Namens mit verschiedenen Inhalten (Polymorphie)

Schema Part 1 (Structures) -- *complex types*

```
<vortrag datum="2000-06-26">  
  W3C's XML Schema –  
  Ein neuer Beschreibungsmechanismus  
  für (Meta-) Daten-Strukturen  
  <autor>Mario Jeckle</autor>  
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
  datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```

```
<?xml version="1.0"?>  
<xsd:schema xmlns:xsd="http://www.w3.org/1999/XMLSchema">  
  <xsd:complexType name="vortragType" content="mixed">  
    <xsd:element name="autor" type="xsd:string" minOccurs="1" maxOccurs="1"/>  
    <xsd:attribute name="datum" type="xsd:date" minOccurs="0" maxOccurs="1"/>  
  </complexType>  
  <xsd:element name="vortrag" type="vortragType"/>  
</xsd:schema>
```

Schema Part 1 (Structures) -- *elements*

```
<vortrag datum="2000-06-26">
```

W3C's XML Schema –

Ein neuer Beschreibungsmechanismus
für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
    datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```

```
<?xml version="1.0"?>
```

```
<xsd:schema xmlns:xsd="http://www.w3.org/1999/XMLSchema">
```

```
<xsd:complexType name="vortragType" content="mixed">
```

```
<xsd:element name="autor" type="xsd:string" minOccurs="1" maxOccurs="1"/>
```

```
<xsd:attribute name="datum" type="xsd:date" minOccurs="0" maxOccurs="1"/>
```

```
</complexType>
```

```
<xsd:element name="vortrag" type="vortragType"/>
```

```
</xsd:schema>
```

Mario Jeckle, FT3/EK

Schema Part 1 (Structures) -- *attributes*

```
<vortrag datum="2000-06-26">
```

W3C's XML Schema –

Ein neuer Beschreibungsmechanismus
für (Meta-) Daten-Strukturen

```
<autor>Mario Jeckle</autor>
```

```
</vortrag>
```

```
<!ELEMENT vortrag ANY>
```

```
<!ATTLIST vortrag  
    datum CDATA #IMPLIED>
```

```
<!ELEMENT autor (#PCDATA)*>
```

```
<?xml version="1.0"?>
```

```
<xsd:schema xmlns:xsd="http://www.w3.org/1999/XMLSchema">
```

```
<xsd:complexType name="vortragType" content="mixed">
```

```
<xsd:element name="autor" type="xsd:string" minOccurs="1" maxOccurs="1"/>
```

```
<xsd:attribute name="datum" type="xsd:date" minOccurs="0" maxOccurs="1"/>
```

```
</complexType>
```

```
<xsd:element name="vortrag" type="vortragType"/>
```

```
</xsd:schema>
```

Mario Jeckle, FT3/EK

Schema Part 2 (Datatypes) angebotene Datentypen...

Build-in datatypes

- ID
- IDREF
- IDREFS
- ENTITY
- ENTITIES
- NMTOKEN
- NMTOKENS
- Name
- QName
- NCName
- NOTATION
- integer
- nonPositiveInteger
- negativeInteger
- long
- int
- short
- byte
- nonNegativeInteger
- unsignedLong
- unsignedInt
- unsignedShort
- unsignedByte
- positiveInteger
- boolean
- float
- double
- decimal
- string
- date
- time
- timeInstant
- timePeriod
- month
- year
- century
- recurringDate
- recurringDay
- timeDuration
- recurringDuration
- binary
- uriReference
- language

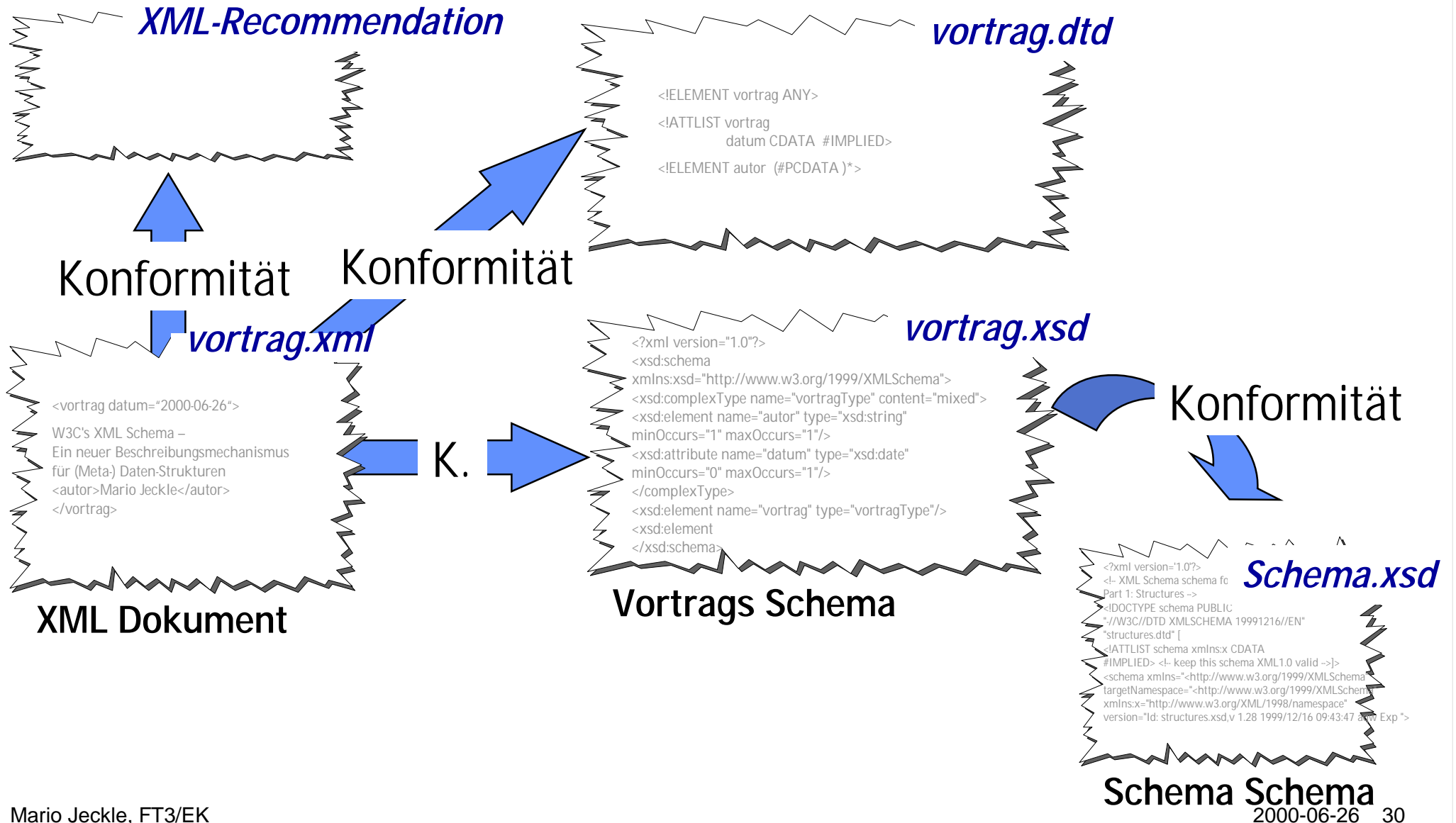
Schema Part 2 (Datatypes) angebotene Datentypen...

• integer	-1, 0, 7683, +555	
• nonPositiveInteger	{..., -2, -1, 0}	
• negativeInteger	{..., -3, -2, -1}	
• long	$-2^{63} \leq \text{long} \leq 2^{63}-1$	
• int	$-2^{31} \leq \text{int} \leq 2^{31}-1$	
• short	$-2^{15} \leq \text{short} \leq 2^{15}-1$	
• byte	$-2^7 \leq \text{byte} \leq 2^7-1$	
• nonNegativeInteger	{0,1,2,...}	
• positiveInteger	{1,2,3,...}	
• unsignedLong	$0 \leq \text{unsignedLong} \leq 2^{64}-1$	
• unsignedInt	$0 \leq \text{unsignedInt} \leq 2^{32}-1$	
• unsignedShort	$0 \leq \text{unsignedShort} \leq 2^{16}-1$	
• unsignedByte	$0 \leq \text{unsignedByte} \leq 2^8-1$	
• boolean	{true, 1, false, 0}	
• float	32-Bit Fließkommazahl gemäß IEEE 754-1985	-1E4, 12.64E8, 12e-2, INF
• double	64-Bit Fließkommazahl gemäß IEEE 754-1985	
• decimal	-12, 8, 3.14151592, +1.0	

Schema Part 2 (Datatypes) angebotene Datentypen...

• string	ISO 10646 und Unicode	"hello world"
• date	ISO 8601	2000-06-06
• time	ISO 8601	09:00:00+2:00
• timeInstant	ISO 8601	2000-06-06T09:00:00+2:00
• timePeriod	ISO 8601	P7M
• month	ISO 8601	P1M
• year	ISO 8601	P1Y
• century	ISO 8601	19
• recurringDate	ISO 8601	PT24H
• recurringDay	ISO 8601	P1M2D
• timeDuration	ISO 8601	POY0M0DT1H30M
• recurringDuration	<i>abstrakter Supertyp von duration und period; nicht direkt verwendbar!</i>	
• binary	<i>abstrakter Typ; nur verwandbar durch Ableitung und encoding-Spezifikation</i>	
• uriReference	IETF RFC2396	http://www.jeckle.de
• language	IETF RFC1766	de-de, en-uk, x-klingon

Well-formed & valid & schema valid ...



W3C's XML Schema – Status und Resümee

- Status: *last call working draft* (WD 2000-04-07)
Vor der Verabschiedung als *recommendation*
- Der zukünftige Standard zur XML-Sprachdefinition
- Werkzeugunterstützung verfügbar
 - "Schemas are coming: Start using them!"
 - *Tim Berners-Lee, 1999-11-05*
- Umstellung von DTD-Repositories auf Schema zu beobachten
- Sprachgenerierung (XMI!) statt (manueller) Spracherstellung
- XML-Schema ist keine Datenmodellierungssprache
- XML-Schema ist eine XML-Sprache
- Jedes Schema ist ein (schema) valid XML-Dokument

References

W3C's XML-Schema:

www.w3.org/TR/NOTE-xml-schema-req (*XML schema requirements*)

www.w3.org/TR/xml-schema-0

www.w3.org/TR/xml-schema-1

www.w3.org/TR/xml-schema-2

Alternativvorschläge:

www.w3.org/TR/dt4dtd

www.w3.org/TR/NOTE-dcd

www.w3.org/TR/NOTE-ddml

www.brics.dk/DSD/

www.w3.org/TR/NOTE-SOX/

www.ascc.net/xml/resource/schematron/schematron.html

www.w3.org/TR/1998/NOTE-XML-data-0105

References

Sekundärliteratur:

www.w3.org/TR/schema-arch

www.lindamann.com/xml/XML%20Schemas%20NG%20Guide%20HTML.htm

xml.com/pub/2000/02/23/xmldeviant/index.html?wwwrrr_20000223.txt

www.w3.org/TR/NOTE-xml-schema-req

www.iso.ch/cate/d19346.html (*ISO 11404*)

Werkzeuge:

www.alphaworks.ibm.com/formula/xml

xml.apache.org (*Xerces*)

www.extensibility.com (*XML Authority*)

Tangierte und weiterführende XML-Literatur:

www.w3.org/TR/REC-xml (*XML v1.0 Recommendation*)

www.w3.org/TR/REC-xml-names

www.w3.org/TR/xml-infoset

Dieser Vortrag und weiterführende Information:

www.jeckle.de