

# DAIMLERCHRYSLER

## **XML – Standards, Tools, Entwicklungen**

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

- **Dokumente und Daten ...**
  - Der Hype um XML
  - Was ist XML und wo liegen seine Wurzeln?
  - Auszeichnungssprachen
- **Die XML-Technologie ...**
  - Die XML-Sprachfamilie
  - Von HTML zu XML
  - Eigene XML-Vokabulare und -Sprachen
- **Anwendungsfelder und Potentiale ...**
  - Präsentation und Dokumentation
  - Datentransfer und Systemintegration
  - Persistenz und Datenbanken

# Dokumente und Daten ... Einführung und Überblick

Das Datenformat [XML] erleichtert den Informationsaustausch zwischen vernetzten Computern

XML schickt sich an in die Fußstapfen von HTML zu treten

[2000, p. 200]

Streit um Programmiersprache XML

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

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

[DER SPIEGEL, 2000 -06-07]

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

[L, 2000-06-22]

Vom 'zugänglichen' XML [...] im Web kaum noch was zu sehen [...]

[c't,18/2001, 188]

Alle Dokumente sind gleich

[SZ, 1999-02-16]

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

[DER SPIEGEL, 1999 -10-03]

Die Extended Markup Language für eCommerce

[F.A.Z.]

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

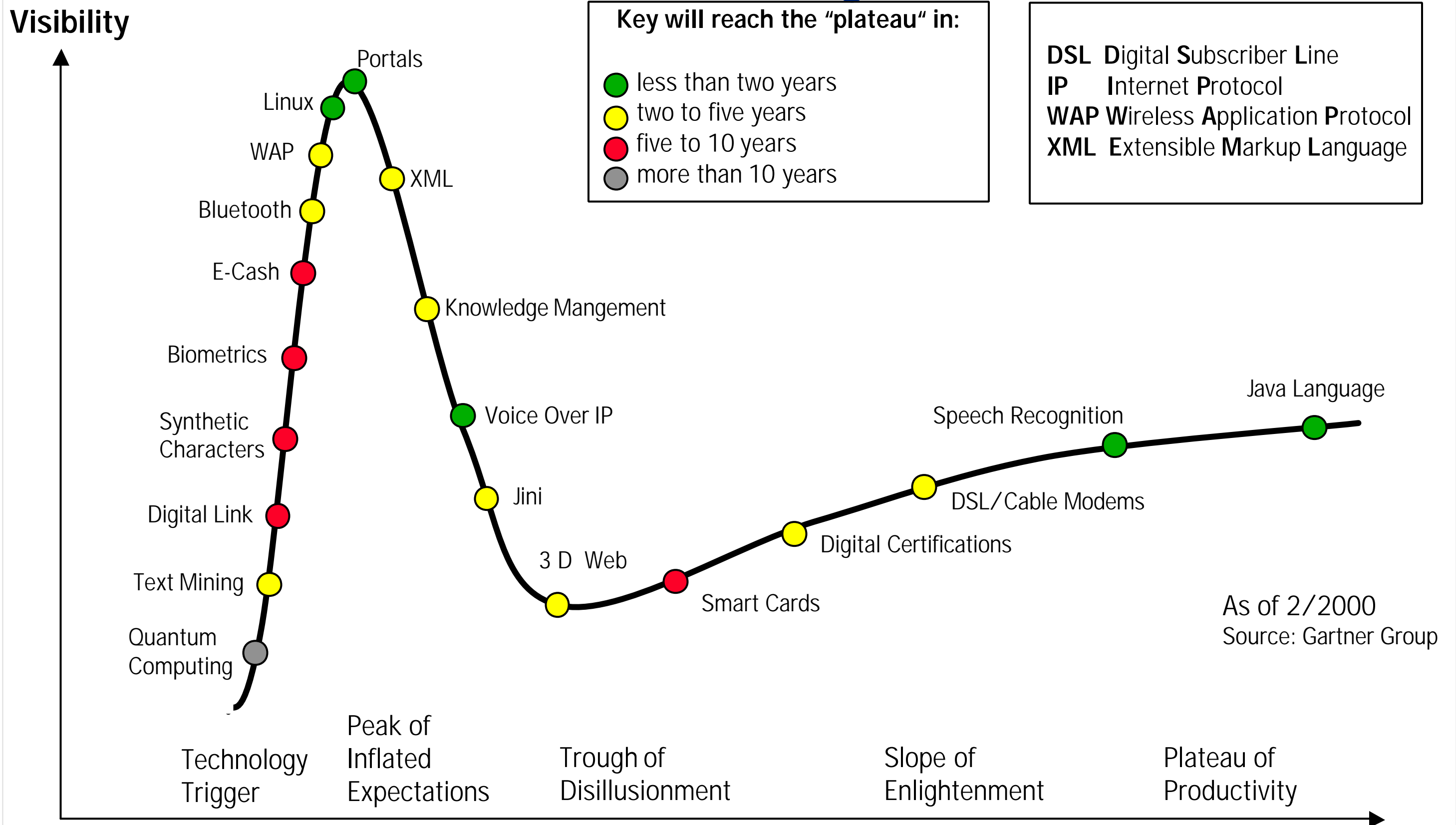
[Süddeutsche Zeitung, 2000-01-11]

Ein digitales Esperanto für das Internet

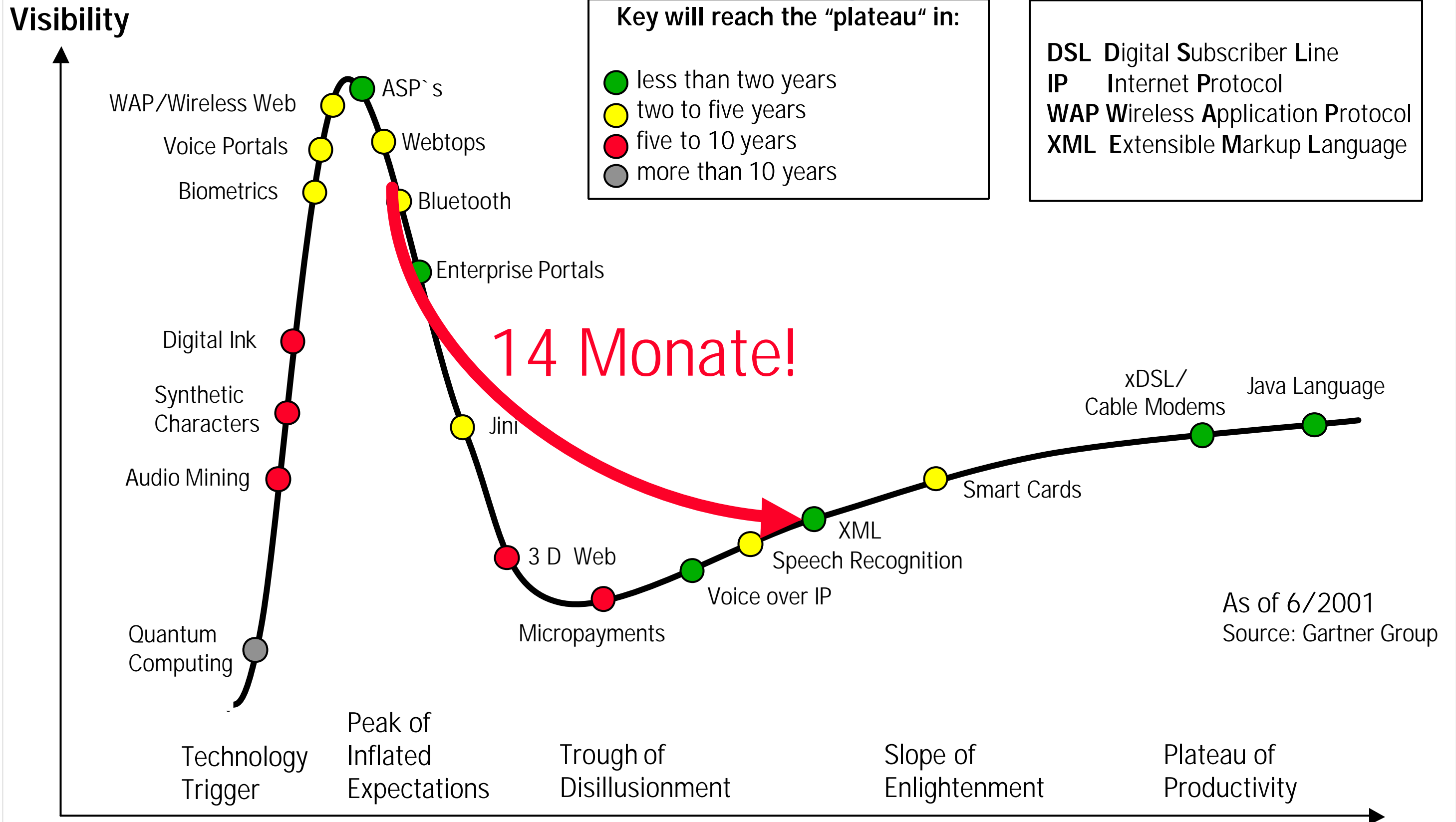
[Die Welt, 2000-10-07]

2001-11-17

# Dokumente und Daten ... Einführung und Überblick



# Dokumente und Daten ... Einführung und Überblick

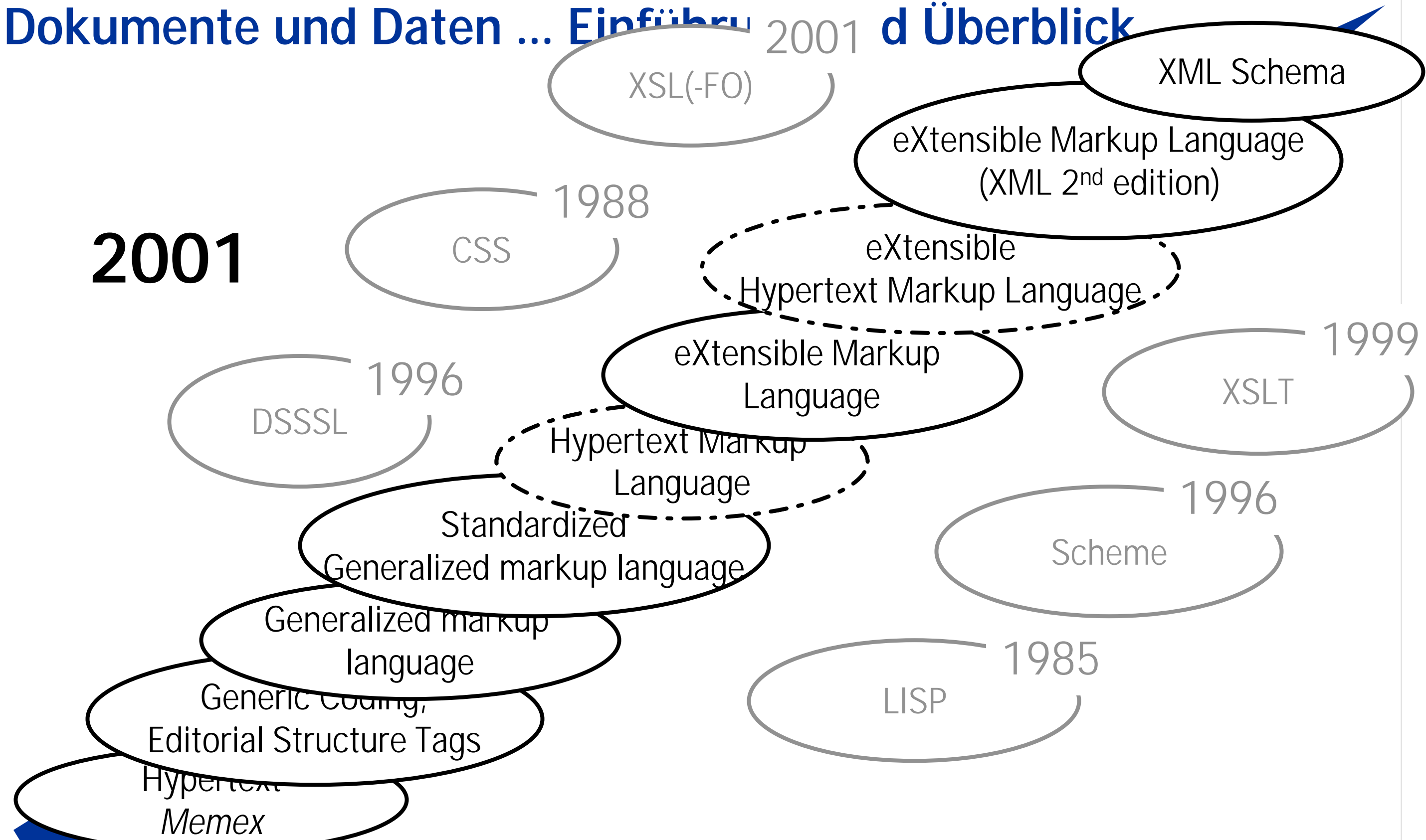


## Dokumente und Daten ... Einführung und Überblick

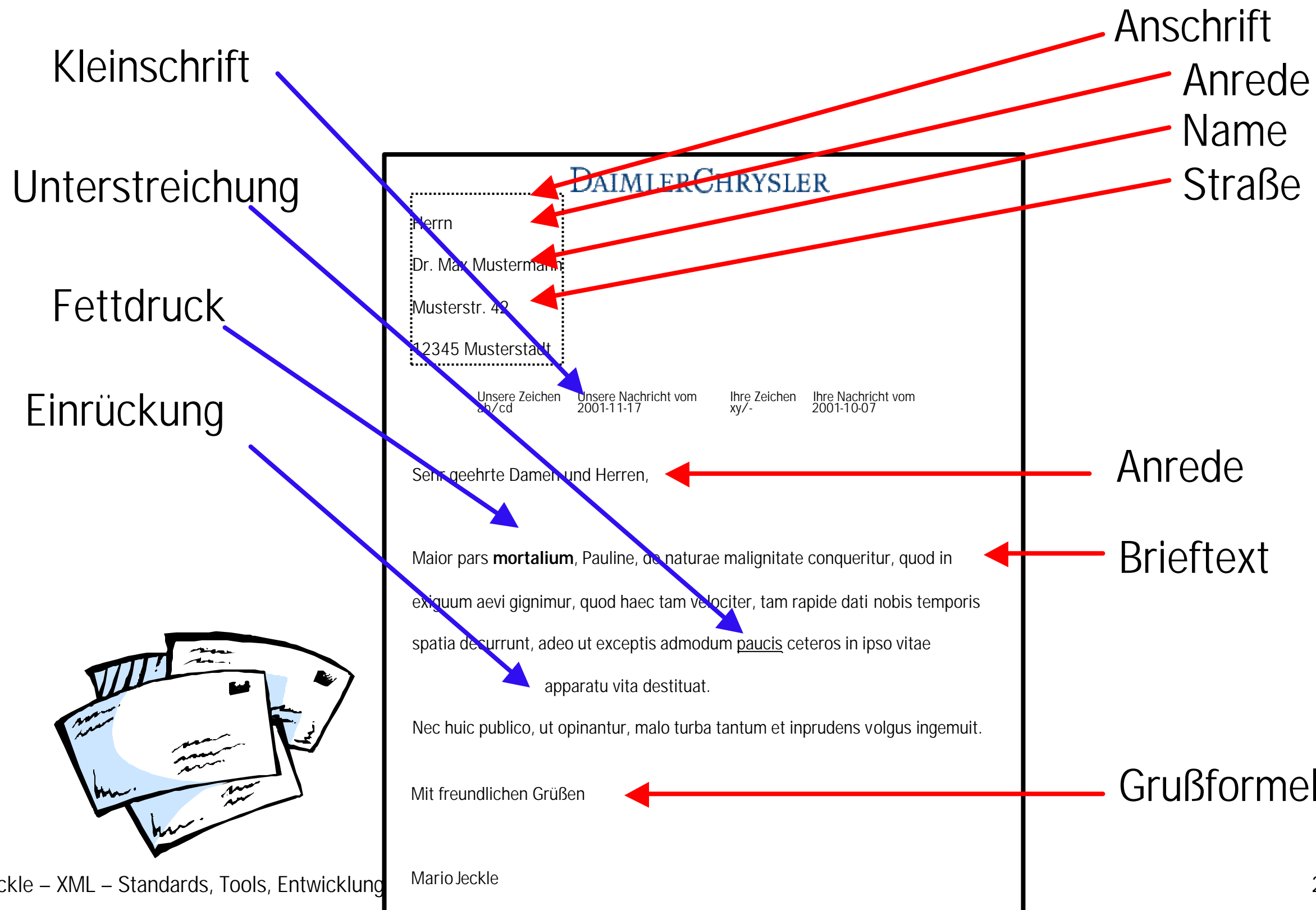
- XML ist eine Methode, um strukturierte Daten in einer Textdatei darzustellen
- XML ist Text, aber nicht zum Lesen
- XML ist ausführlich, was aber kein Problem darstellt
- XML sieht fast aus wie HTML, ist aber kein HTML
- XML ist eine Familie von Techniken
- XML ist eine Untermenge des ISO-Standards 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 ... Einführung und Überblick

2001

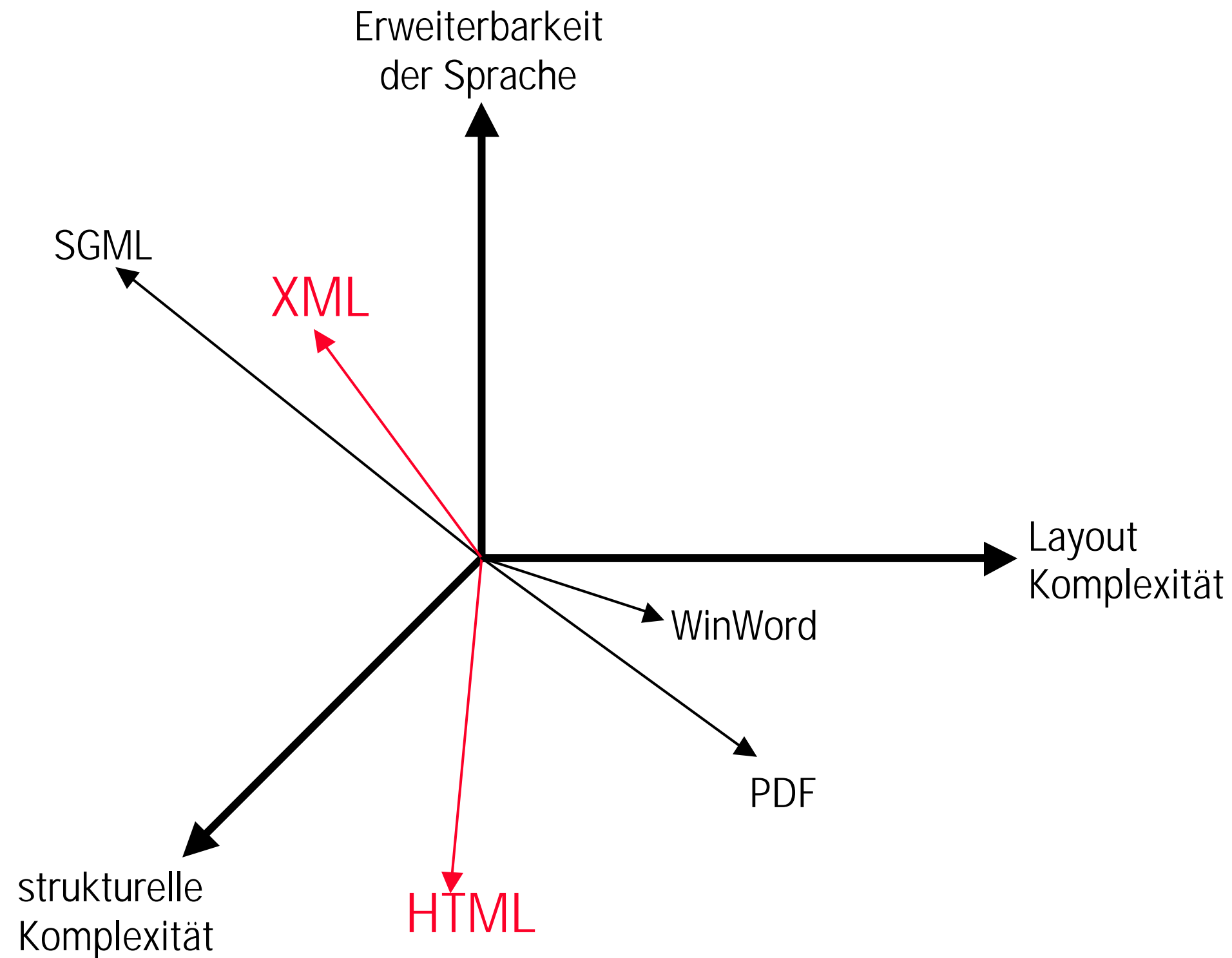


# Dokumente und Daten ... Auszeichnungssprachen





# Dokumente und Daten ... Auszeichnungssprachen



## Die XML-Technologie ...

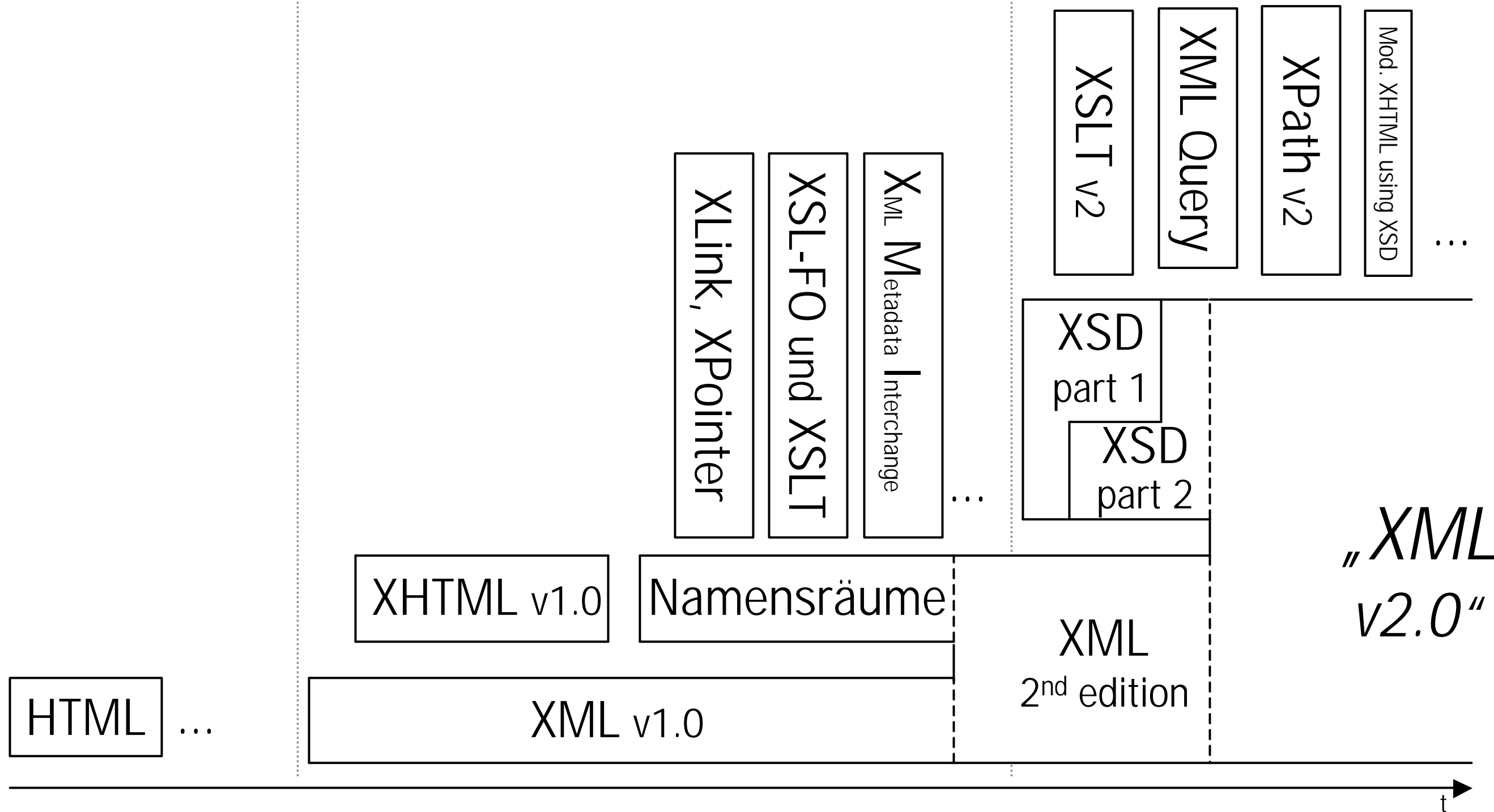
```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>XML – Standards, Tools, Entwicklungen</Titel>
  <Veranstaltung datum="2001-11-17">
    <Name>Universität Ulm</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```

# Die XML-Technologie ... Historische Entwicklung

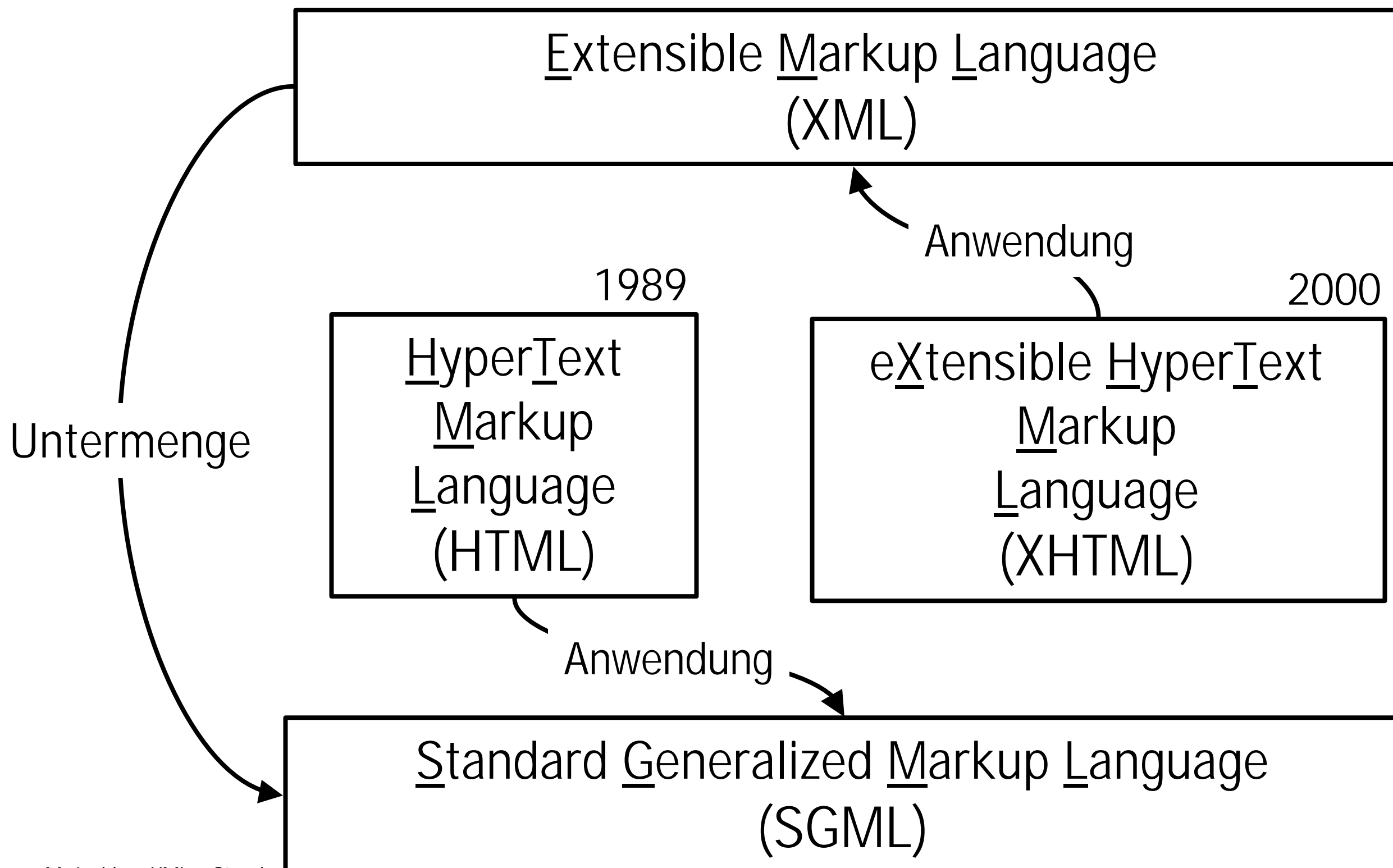
Web-Prähistorie

1. Generation

2. Generation



## Die XML-Technologie ... von HTML zu XHTML



## Die XML-Technologie ... Dokument-Typ-Definitionen

```
<!ELEMENT Vortrag
  (Titel , Veranstaltung , Referent)>
```

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

```
<!ELEMENT Veranstaltung (Name)>
```

```
<!ATTLIST Veranstaltung datum CDATA #REQUIRED >
```

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

```
<!ELEMENT Referent (Name , Firma , URL , E-Mail)>
```

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

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

```
<!ELEMENT E-Mail (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>XML – Standards, Tools, Entwicklungen</Titel>
  <Veranstaltung datum="2001-11-17">
    <Name>Universität Ulm</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>
```

## Die XML-Technologie ... 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

## Die XML-Technologie ... 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 Strukturprimitive unzureichend

## Die XML-Technologie ... XML Schema

- Erweiterungsforderungen ...
  - Strukturell ...
    - Namespace Integration (und damit XML 2<sup>nd</sup> 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



## Die XML-Technologie ... 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

## Die XML-Technologie ... XML Schema

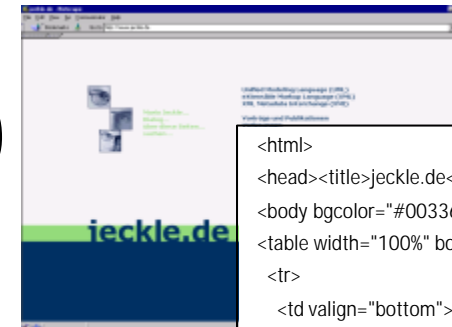
- Zusammenfassend:
  - Grammatik für beliebige XML-Vokabulare
  - Part 1 beschreibt Strukturen und Inhaltseinschränkungen
  - Part 2 definiert Datentypdefinition für Schema Part 1 und weitere XML-Vokabulare
  - Signifikante Erweiterung der DTD-Mächtigkeit, wird diese langfristig ersetzen
  - Ist eine 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 ...

## Anwendungsfelder und Potentiale ...

### ● Präsentation und Dokumentation ...

#### ● XHTML – Darstellung im Web(-Browser)

- Reformulierung von HTML v4.01
- (Halb-)automatisiert überführbar
- Aus beliebigen XML-Vokabularen erzeugbar

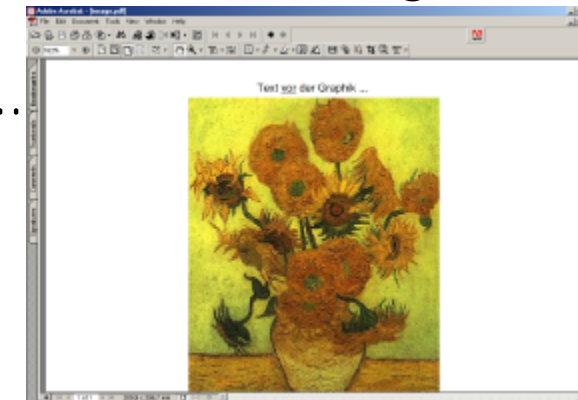


```
<html>
<head><title>jeckle.de</title></head>
<body bgcolor="#003366" topmargin="0" leftmargin="0" marginwidth="0" marginheight="0">
<table width="100%" border="0" cellspacing="0" cellpadding="0" height="65%" bgcolor="#f0f0f0">
<tr>
<td valign="bottom">&nbsp;</td>
<td colspan="2" valign="bottom" align="center">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

```

#### ● XSL Formatting Objects – Erzeugung von druckfertigen Darstellungen

- Erzeugung vielfältiger Zielformate (WYSIWYG) z.B. PDF, PS, GIF, ..
- Erste Implementierungen verfügbar
- XSL-FO kann aus beliebigen XML-Vokabularen erzeugt werden

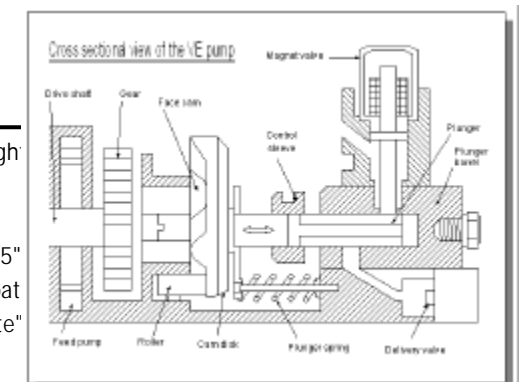


#### ● Scalable Vector Graphics – Graphikdarstellung und -austausch

- XML-basiertes Graphikformat
- Erste Implementierungen verfügbar
- Editierbarkeit und Auswertbarkeit bleibt erhalten

```
<svg xml:space="preserve" width="752" height="752">
<desc>DIAGRAM.PDX</desc>
<defs>
<pattern id="pattern1" x="0" y="0" width="15" height="15" patternContentUnits="userSpaceOnUse" patternTransform="translate(0,0) rotate(0) scale(1,1)"/>
<rect width="15" height="15" style="fill:white" />
<path style="fill:none;stroke:black" d="M 7.5,0 L 0,7.5 M 15,7.5 L 7.5,15"/>
</pattern>

```



# Anwendungsfelder und Potentiale ...

## ● Datentransfer und Systemintegration ...

### ● XML-Vokabulare und ihre Transformation

- Nutzung von (quasi) Standards
- Definition eigener Sprachen
- Transformation durch XSLT

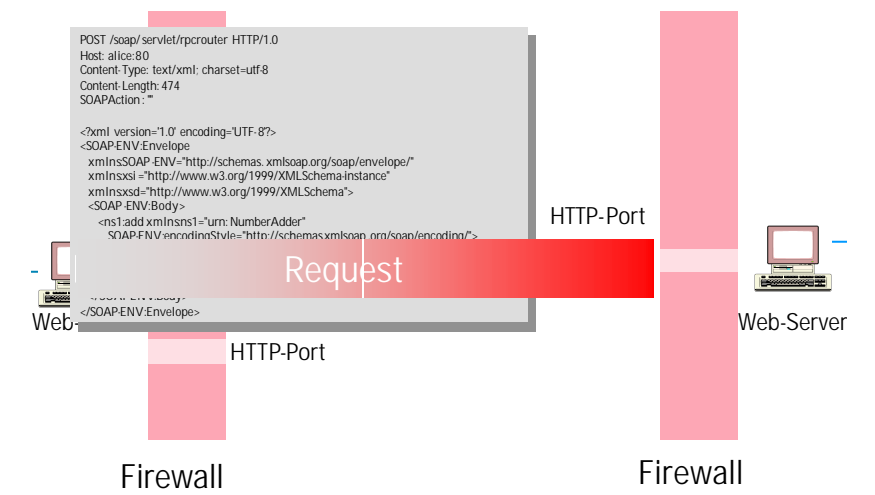
```
<?xml version="1.0" encoding="ISO-8859-1"?>
<bestellung>
  <artikel>
    <nummer>4711</nummer>
    <benennung>Wusch Superfein</benennung>
    <preis>
      <betrag>195.99</betrag>
      <waehrung>DEM</waehrung>
    </preis>
    <kunde nummer="X-363-23"/>
  </artikel>
</bestellung>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<order>
  <customer>
    <custID>X-363-23</custID>
  </customer>
  <itemlist>
    <item>
      <itemNO>4711</itemNO>
      <identification>Wusch Superfein</identification>
      <price currency="EUR">100.20...</price>
    </item>
  </itemlist>
</order>
```



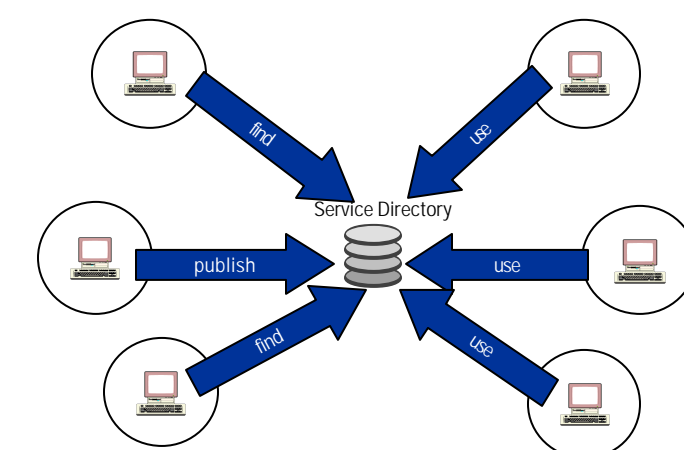
### ● XML-basierte Remote Procedure Calls

- Leichtgewichter Ansatz für RPCs
- Sicherheitsthematik
- SOAP als (de facto und zukünftiger W3C-) Standard



### ● Web Services

- Dienste auf Basis offener RPC-Mechanismen
- Neutrale Schnittstellenbeschreibung
- Veröffentlichung der Dienste in offenen Verzeichnissen



## Anwendungsfelder und Potentiale ...

### ● Persistenz und Datenbanken ...

#### ● XML-Schemarepositories

- Veröffentlichung und Dokumentation existierender Schemata
- Aufbau eigener interner Repositories
- Speicherung und Publikation von Schemata, Transformationen, Beispieldokumenten, ...

#### ● XML-API-Schnittstellen

- Verbreitete (standardisierte) generische Ansätze: SAX und DOM
- Spezialisierte Ansätze: JDOM, DOM4J, .NET, ...
- Integration von Softwareentwicklung und XML-(Schema-)Design: JAXB, XMI, Schema2Java, ...

#### ● XML-Datenbanken

- Umwandlung in relationale- oder netzwerk-Strukturen
- Erweiterungen und Module existierenden (Post-)Relationaler Datenbankmanagementsysteme
- „Native“ XML-DBMS