

DAIMLERCHRYSLER

Praktischer Einsatz und Potentiale der XML-Technologie

Mario Jeckle

DaimlerChrysler Forschungszentrum Ulm

mario.jeckle@daimlerchrysler.com

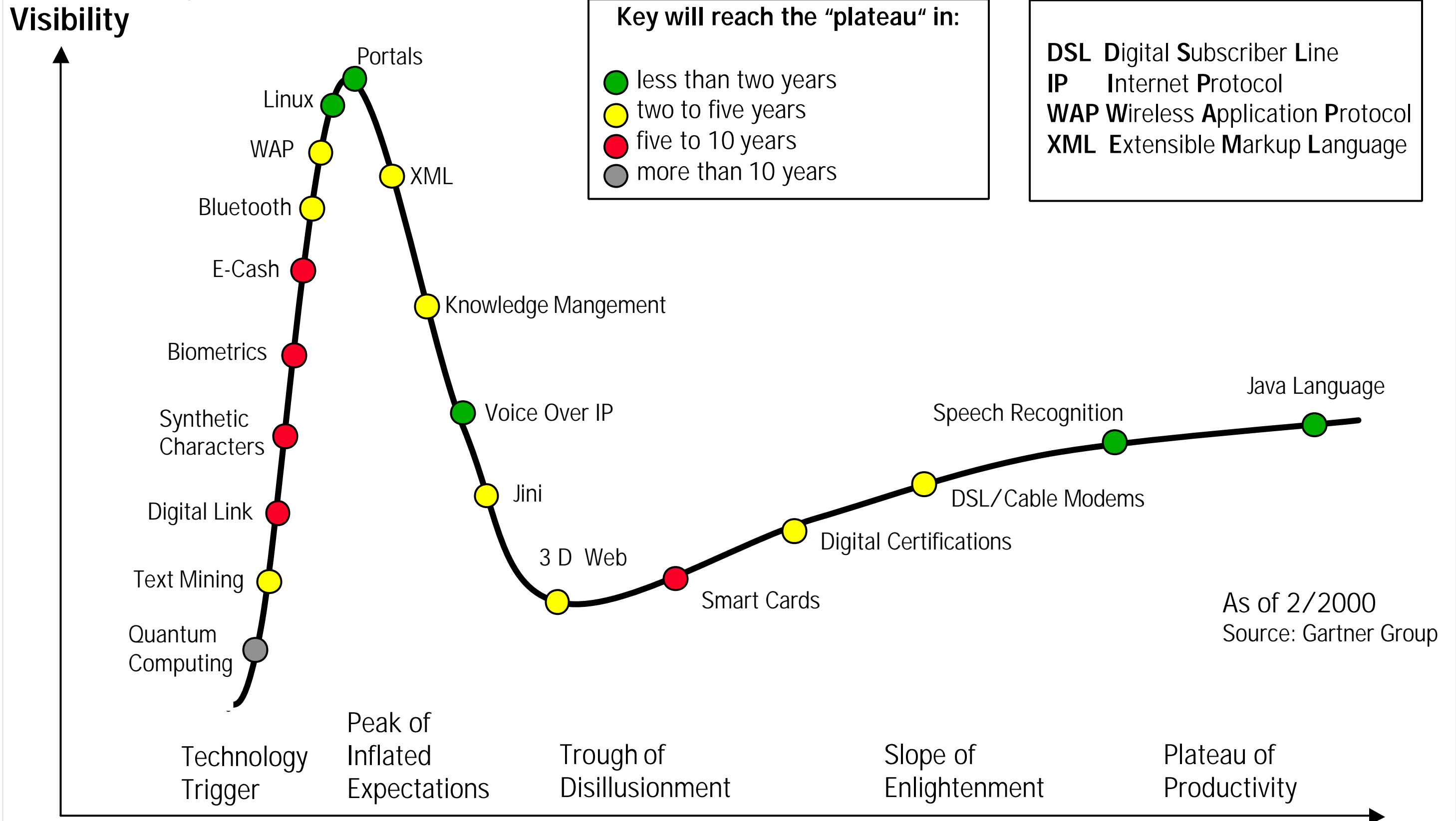
mario@jeckle.de

www.jeckle.de

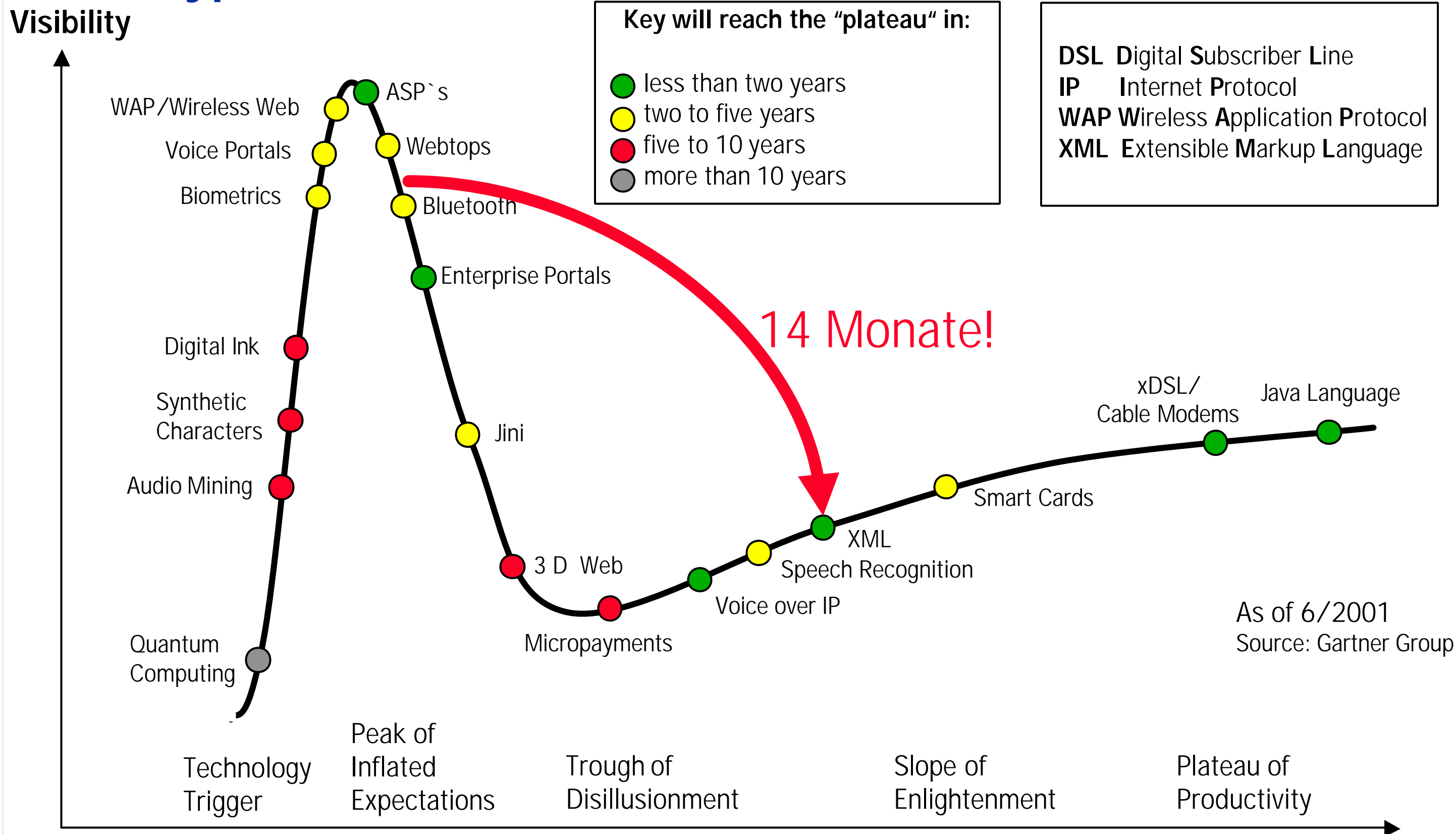
Gliederung

- Der Hype und die Wirklichkeit ...
- Das *XML Universum*
- Architekturelle Aspekte
- XML-Artefakte und die zugehörige Informationslogistik
- XML im Einsatz: Praktische Empfehlungen
- XML-Anbindung an Programmiersprachen
- Standards und Standardisierung
- Persistenz und Datenhaltung
- Middleware und Enterprise Application Integration
- XML und Sicherheit
- XML-Werkzeuge

Der Hype und die Wirklichkeit ...



Der Hype und die Wirklichkeit ...



Der Hype und die Wirklichkeit ...

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

Alle Dokumente sind gleich

[L, 2000-06-22]

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

[c't,18/2001, 188]

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

[SZ, 1999-02-16]

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

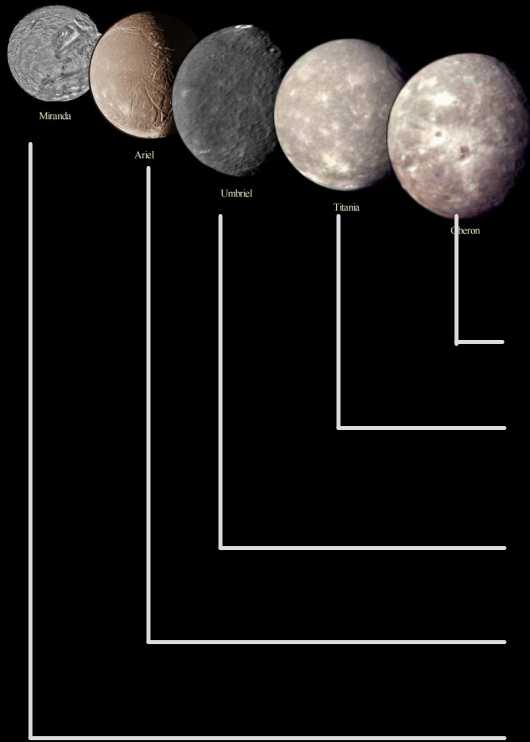
[Die Welt, 2000-10-07]

Der Hype und die Wirklichkeit ...

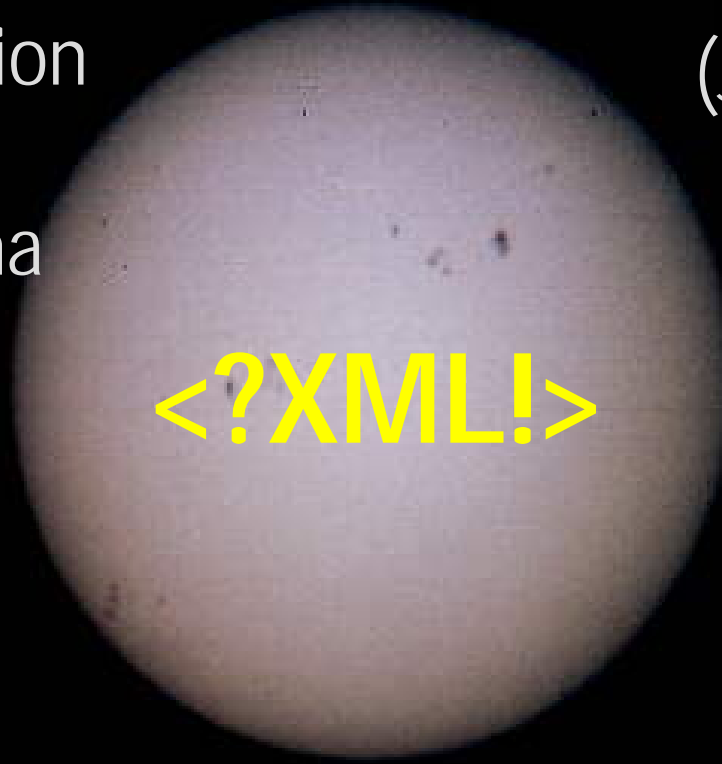
- XML ist allgegenwärtig ...
 - ... als Schnittstelle in verschiedensten Softwareprodukten
 - ... als Standardsprache des Web (HTML)
 - ... als Datenaustauschformat
 - ... als Marketing-Term
- XML ist billig
- ... leicht einzusetzen
- ... geringer Schulungsaufwand
- Technologisch durchaus anspruchsvoll

Das XML-Universum

Architekturen



Transformation
DTD
XML- Schema
Vokabulare
Schichten



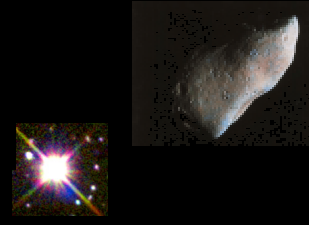
JAXB

Applikationserstellung

(J)DOM

Datenbanken

SAX



Anfragesprachen

OASIS



W3C

Editoren



Parser



Standards

Web-Services

SOAP

Werkzeugunterstützung

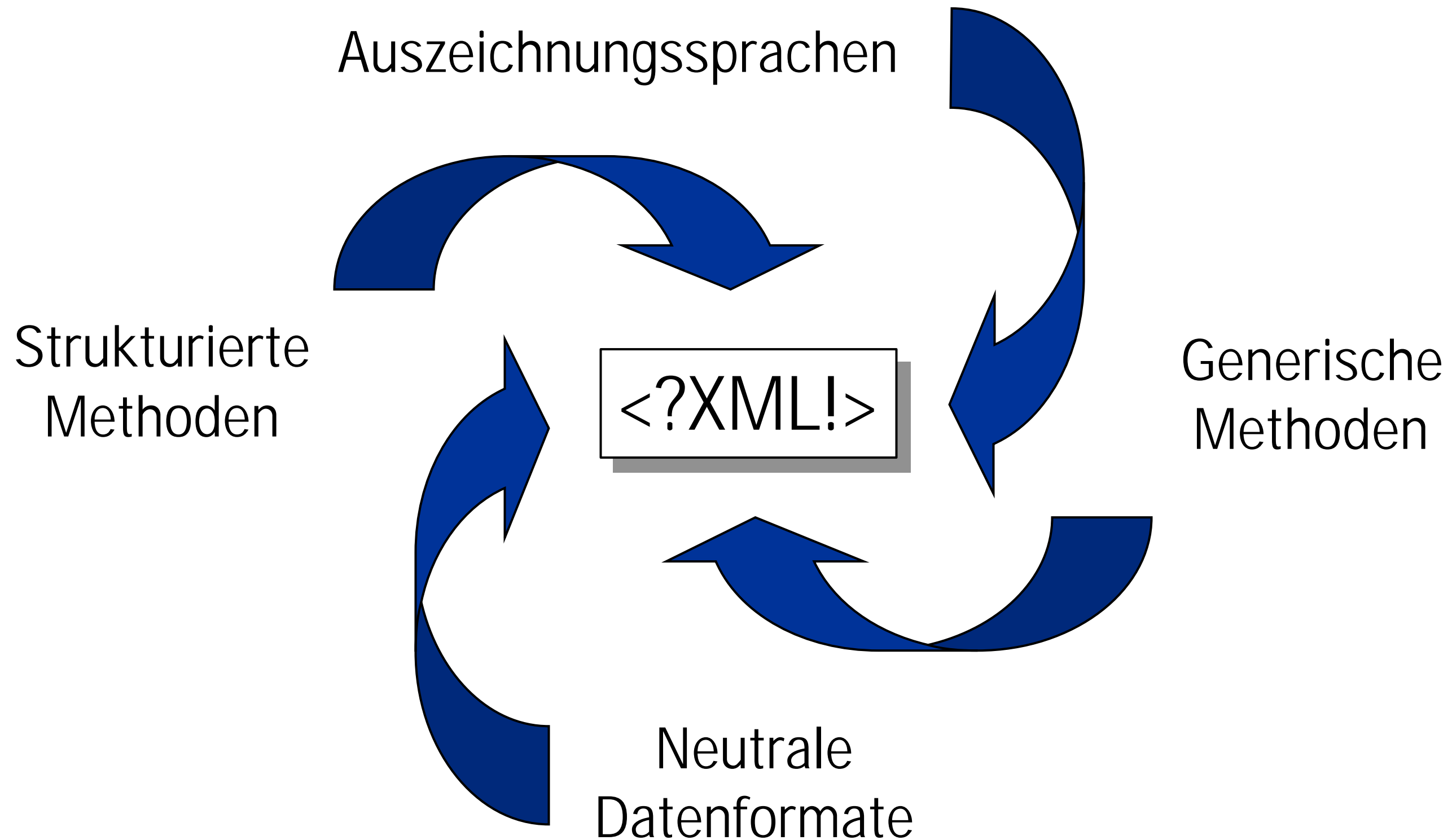


ISO

Integration

Stylesheet-, Linkingprozessoren, ...

Klassische Vorläufer und Einflußfaktoren



Architekturelle Aspekte

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10-24">
    <Name>Management Circle</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>
```

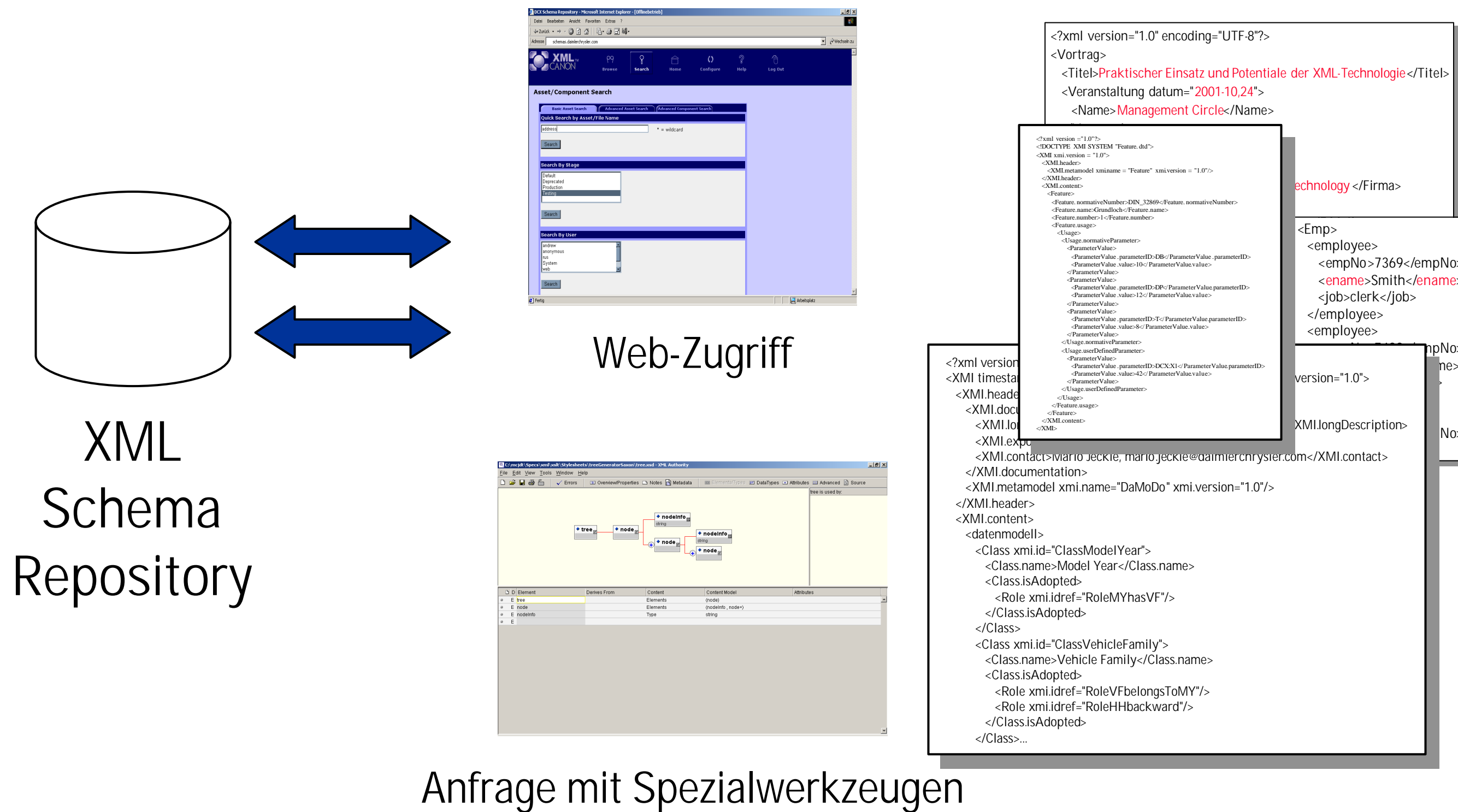
Architekturelle Aspekte

- Klärung der Sinnfälligkeit des XML-Einsatzes
- Festlegung der einzusetzenden XML-Technik pro Applikationsschicht
- Format- oder Standardauswahl
- Folgenabschätzung (Infrastruktur, Prozeßkette, ...)
- Werkzeug- und Implementierungsplattform
- Sinnvolle Begrenzung
- XML-Architektur vs. Architektur die XML enthält
- Offene Architektur um weitere Elemente des XML-Universums einsetzen zu können

Architekturelle Aspekte -- XML-Schema

- Erlaubt die Erstellung regulärer kontextfreier Grammatiken für 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 zweiten Generation (XSLT v2.0, XHTML v2.0, SOAP v1.2, ...)
- Werkzeugunterstützung verfügbar
- Erster Schritt der Schema-Bestrebungen, weitere werden folgen ...

XML-Artefakte und die zugehörige Informationslogistik



XML-Artefakte und die zugehörige Informationslogistik

- Recherche existierender XML-Vokabulare vor Definition eigener
- Redundanzvermeidung in Speicherung und Vermeidung unnötiger Doppelarbeit
- Wiederverwendung bestehender Artefakte
- (zentrale) Versionierung
- Qualitätssicherung
- Stärkung der Interoperabilität
- Effiziente Verwaltung
- Zugriffskontrolle
- Integration externer Partner

XML-Anbindung an Programmiersprachen

- Hauptherausforderung: Verhinderung des *impedance mismatch*
- Möglichst transparente Verarbeitung XML-codierter Daten
- Abbildung von XML-(Baum-)Strukturen in den Hauptspeicher
- Abgleich zwischen Datenmodellierung und Vokabularerstellung
Idealerweise: Wiederverwendung vorhandenen (Meta-)Wissens
- Verwendung von XML als Serialisierungsformat für Speicherstrukturen
- Gibt es ein XML-spezifisches Programmierparadigma?
- Rolle von Java, C# (.NET) -- impliziert XML inhärent eine neue Architektur?

XML-Anbindung an Programmiersprachen

Linear-Ereignisbasierte XML-Verarbeitung (SAX)

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10,24">
    <Name>Management Circle</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>
```

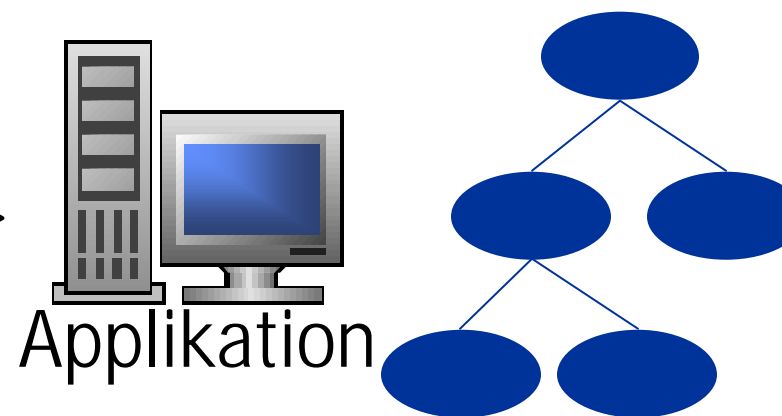


Start Document
Start Element
...
End Element
End Document



Abbildung in generische Hauptspeicherstrukturen (xDOM)

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10,24">
    <Name>Management Circle</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>
```



```
{
  ...
  Document document =
    builder.parse( args[0] );

  Element theRootElement =
    document.getDocumentElement();
  ... }
```

Konfluenz von XML- und Applikationsdatenstrukturen (z.B. JAXB)

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10,24">
    <Name>Management Circle</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 version = "1.0" encoding = "UTF-8"?>
<xsd:schema xmlns:xsd =
"http://www.w3.org/2001/XMLSchema"
  elementFormDefault = "
  <xsd:element name="
  <xsd:complexType base="
  </xsd:complexType>
```

Schema



Generierung

```
public class Vortrag
extends MarshallableObj
implements Element
{
  public
  {
  }
  ...
}
```

Datenstruktur

Standards und Standardisierung

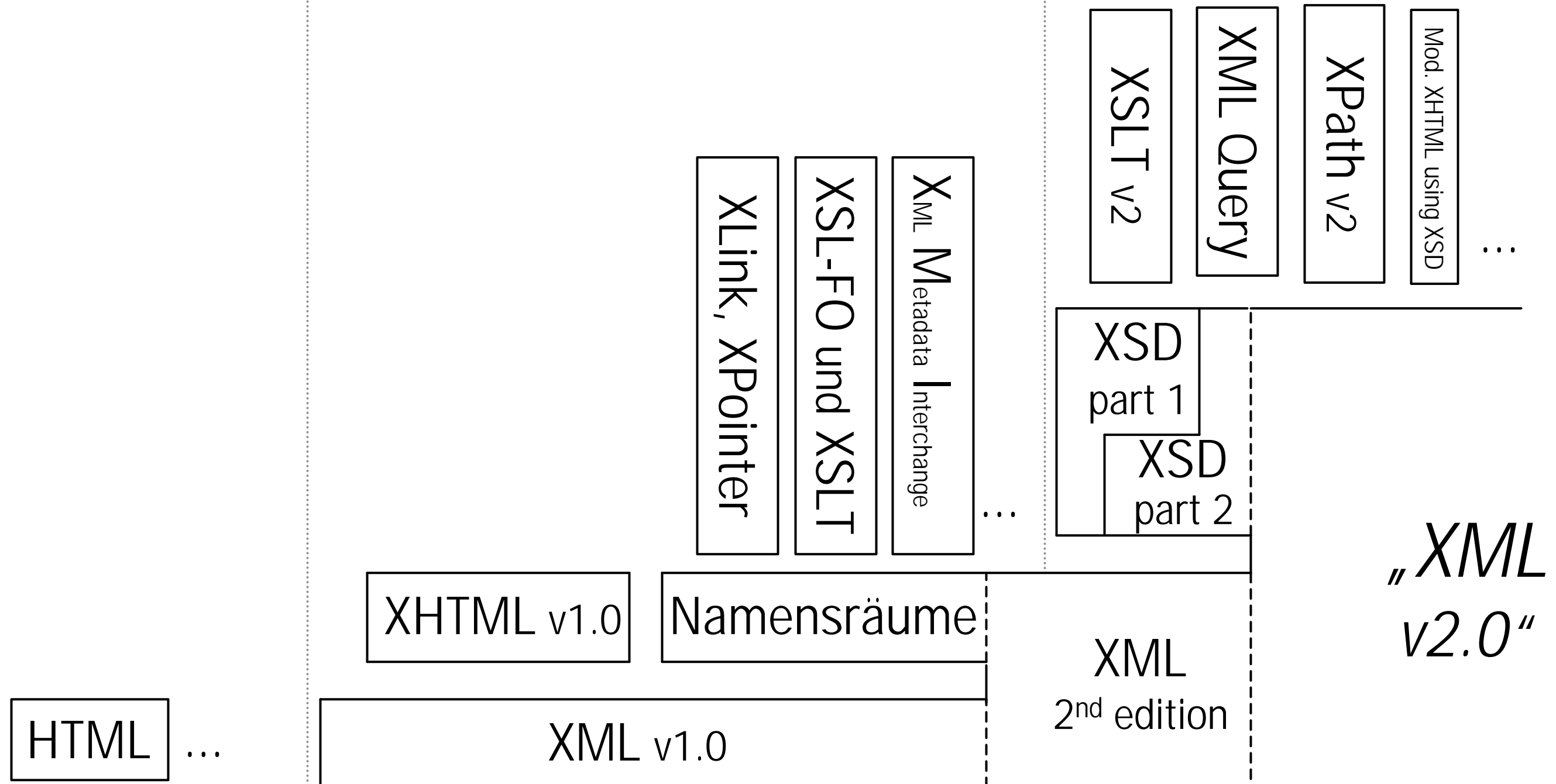
- „Hauptstandardisierungslinie“: World Wide Web Consortium (W3C)
 - Offene non for profit Organisation
 - Herstellergremium
 - Definiert keine *Normen* sondern Industriestandards (*Recommendation*)
 - Verantwortlich für: XML, XLink, XPointer, XPath, (X)HTML, XQuery, XSL(T), ...
- Weitere Gremien und Initiativen
 - xml-dev-Mailingliste. Kommunikationsplattform der XML-Entwickler.
Ursprung der SAX-Schnittstelle
 - Organization for the Advancement of Structured Information Standards (OASIS).
Definiert Sekundärstandards auf der Basis von SGML und XML
 - Weitere: ebXML, BizTalk-Initiative, OAGIS, ...

Standards und Standardisierung

Web-Prähistorie

1. Generation

2. Generation



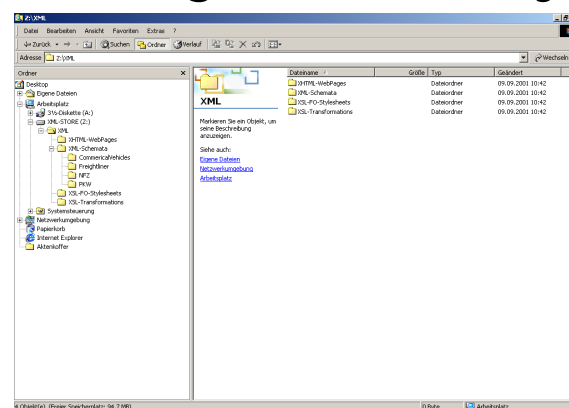
Persistenz und Datenhaltung

- XML scheint strukturinhärent eher eine Baum-artige denn eine mengenorientierte Speicherung zu bedingen
- Geschwindigkeitsgewinn beim Zugriff durch Indexierung
- Einschränkende Sichten auf XML-Daten
- Aggregation und Integration
- Erweiterung post-relationaler Datenbankmanagementsysteme
- Erweiterung semi-strukturierter DBMS
- „Native“ XML-DBMS

Persistenz und Datenhaltung

Naheliegend: XML-Speicherung im Dateisystem

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10-24">
    <Name>Management Circle</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>
```



- Ineffizient
- Eingeschränkte Verteilbarkeit
- Kaum Zugriffsunterstützung

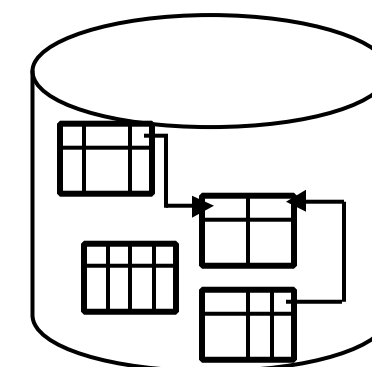
Klassisch: XML-Speicherung in Datenbanksystemen

```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10-24">
    <Name>Management Circle</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>
```



Abbildung auf logische DB-Strukturen

- u. U. Ineffizient
- Impedance mismatch
- Kaum DBMS-Unterstützung

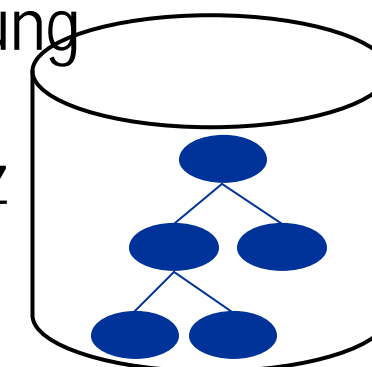


„Neu“: „native“ XML-Speicherung

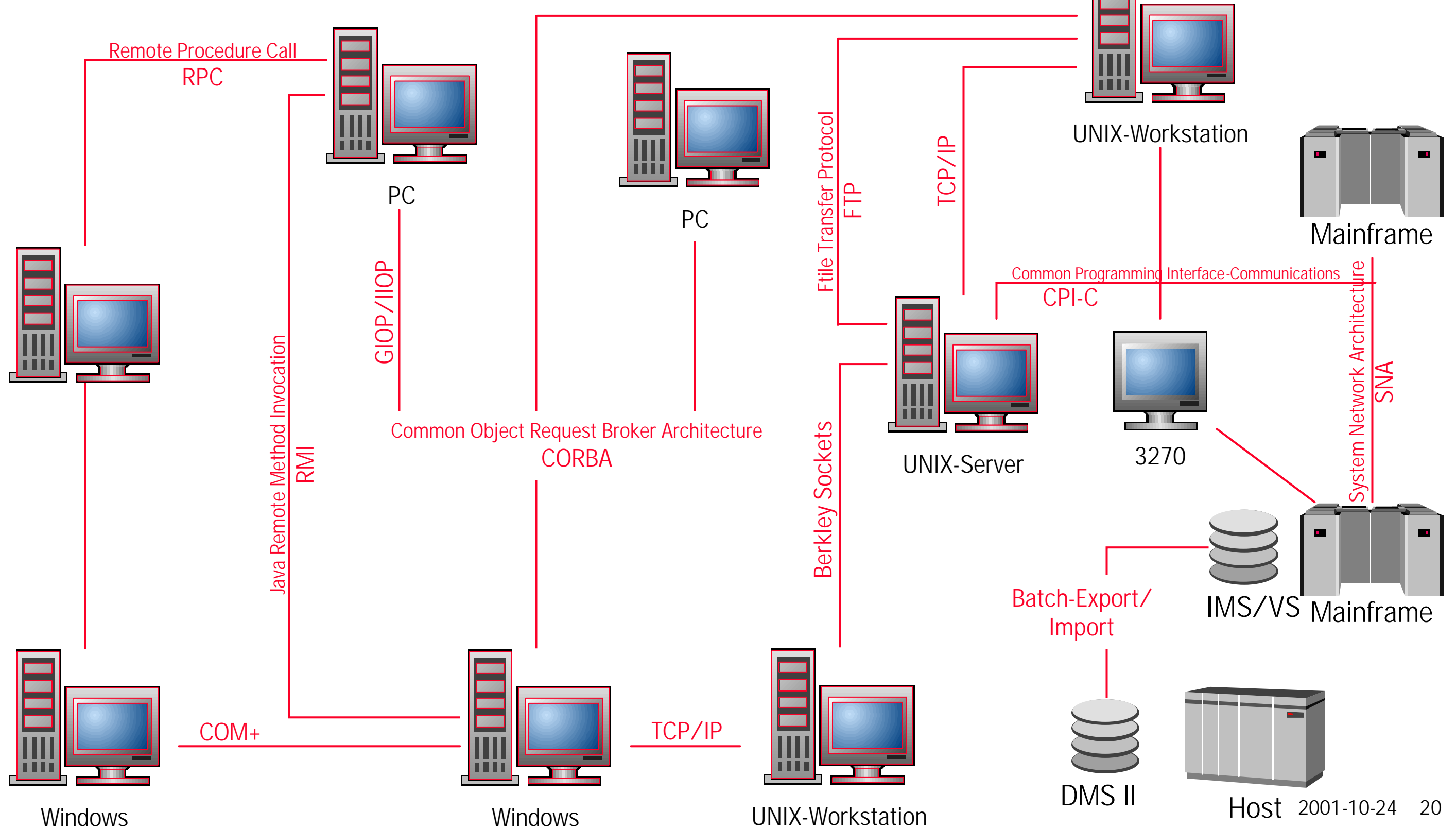
```
<?xml version="1.0" encoding="UTF-8"?>
<Vortrag>
  <Titel>Praktischer Einsatz und Potentiale der XML-Technologie</Titel>
  <Veranstaltung datum="2001-10-24">
    <Name>Management Circle</Name>
  </Veranstaltung>
  <Referent>
    <Name>Mario Jeckle</Name>
    <Firma>DaimlerChrysler Research and Technology</Firma>
    <URL>http://www.jeckle.de</URL>
    <E-Mail>mario.jeckle@daimlerchrysler.com</E-Mail>
  </Referent>
</Vortrag>
```



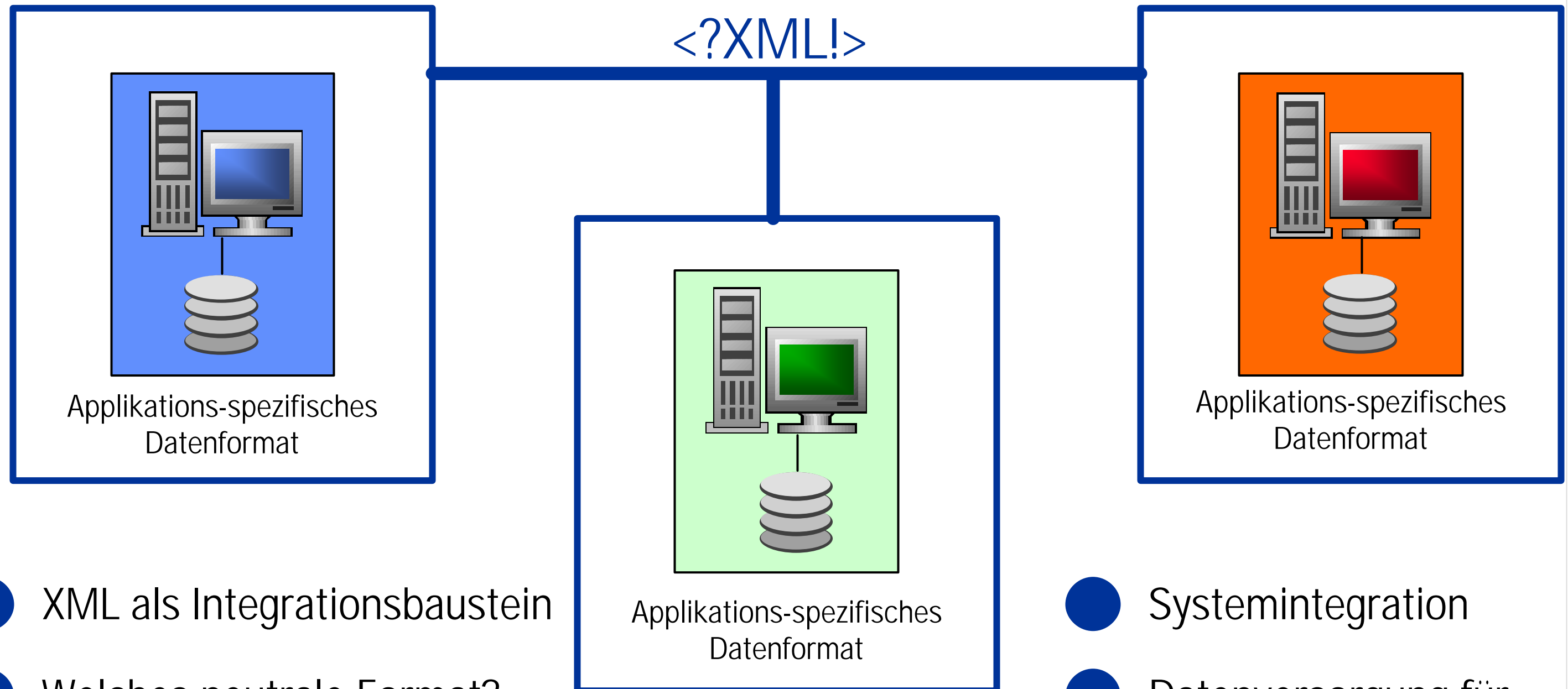
- Baumorientierte Speicherung
- Höhere lineare Performanz
- Im Entstehen begriffen



Middleware und Enterprise Application Integration



Middleware und Enterprise Application Integration



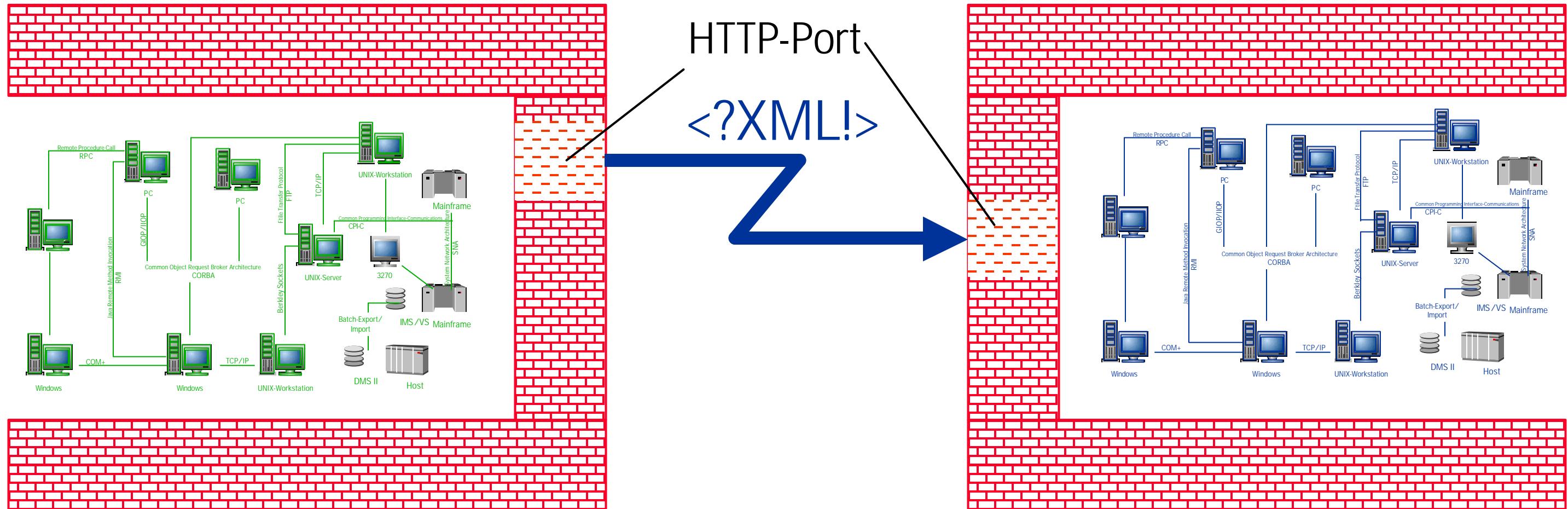
- XML als Integrationsbaustein
- Welches neutrale Format?
- Physische Datenübertragung?

- Systemintegration
- Datenversorgung für Portale
- Offene Schnittstellen

XML und Sicherheit

- XML ist lesbarer Unicode-Text
 - Praktisch für Fehlersuche und Analysezwecke.
„unpraktisch“ für Vertrauliches oder gar Geheimes
- XML definiert keinen inhärenten Sicherheitsmechanismus ...
... ist aber offen für beliebige Sicherungsverfahren
- Besondere Herausforderung: XML-Versandt über Internet-Wege
 - Sichere Übertragung
 - Digitale Signaturen
 - Kryptographie
 - Public Key Infrastruktur

XML und Sicherheit



- XML-basierte Protokolle (z.B. SOAP) lassen Sicherheit absichtsvoll unberücksichtigt
- W3C definiert „XML Digital Signatures“ (Authentifizierung, Datenintegrität)
- W3C definiert „XML Encryption“ (Verschlüsselungsrahmen).
Zum effizienten Einsatz asymmetrischer Verfahren PKI notwendig!
- Sichere Kommunikation auf Leitungsebene durch SSL

XML-Werkzeuge

- Zumeist vergleichsweise preisgünstig verfügbar
- Oftmals gute Open-Source Implementierungen.
Vorsicht: Lizenzbestimmungen!
- Standardkonformität prüfen!
- Auf Basis der Standardschnittstellen leicht eigene Werkzeuge umsetzbar
- Werkzeuge sind nicht alles!
A fool with a tool remains a fool
- Möglichst nur ein einziges Werkzeug einer Werkzeugklasse einsetzen

XML-Werkzeuge

● Werkzeugklassen

● DTD-/Schemaeditoren

XML-Spy

kommerziell

XML-Authority

kommerziell

EditML

...

● XML- oder Instanzeditoren

XML-Spy

kommerziell

XML-Instance

kommerziell

Merlot

...

● XML-Parser

XML4J

Apache Xerces

MSXML

● XSLT-Prozessoren

Saxon

Apache Xalan

Infoteria

kommerziell

...

● XSL-FO Prozessoren

Apache FOP

RenderX XEP

kommerziell

● Datenbanken

Oracle

kommerziell

DB2

kommerziell

Tamino

kommerziell

DBDom

dbXML