


Was ist XML?

Inhalt

Inhalt	XML 1
1 Was ist XML?	2
2 XML Stammbaum.....	4
3 XML-Szenario.....	6
3.1 Ist-Zustand	6
3.2 Aufgabenstellung.....	7
3.3 Lösungsmöglichkeit 1	10
3.4 Lösungsmöglichkeit 2 (XML).....	11
4 XML statt HTML.....	13
5 XML-Anwendungen	19
6 XML-Sprachfamilie	21


1 Was ist XML?

Was ist XML?



Alle reden und schreiben davon ...
... doch was verbirgt sich wirklich dahinter?

Ist XML nur eine Mode-Erscheinung oder ist
sie wirklich eine bleibende Technologie?

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 2

XML ist eine *Metasprache* für die Definition eigener *Auszeichnungssprachen* ("Markup Language").

→ Universalgrammatik

Der amerikanische Linguist und Philosoph Noam Chomsky vertritt die Meinung, dass alle menschlichen Sprachen im Grunde gleich sind, weil sie über eine gemeinsame, universelle Basis verfügen. So eine "Universalgrammatik" könnte XML für die Kommunikationssprachen in der Datenverarbeitung werden.


Was ist XML?

XML

→ **Extensible** - Erweiterbare

→ **Markup** - Markierungen

→ **Language** - Sprache

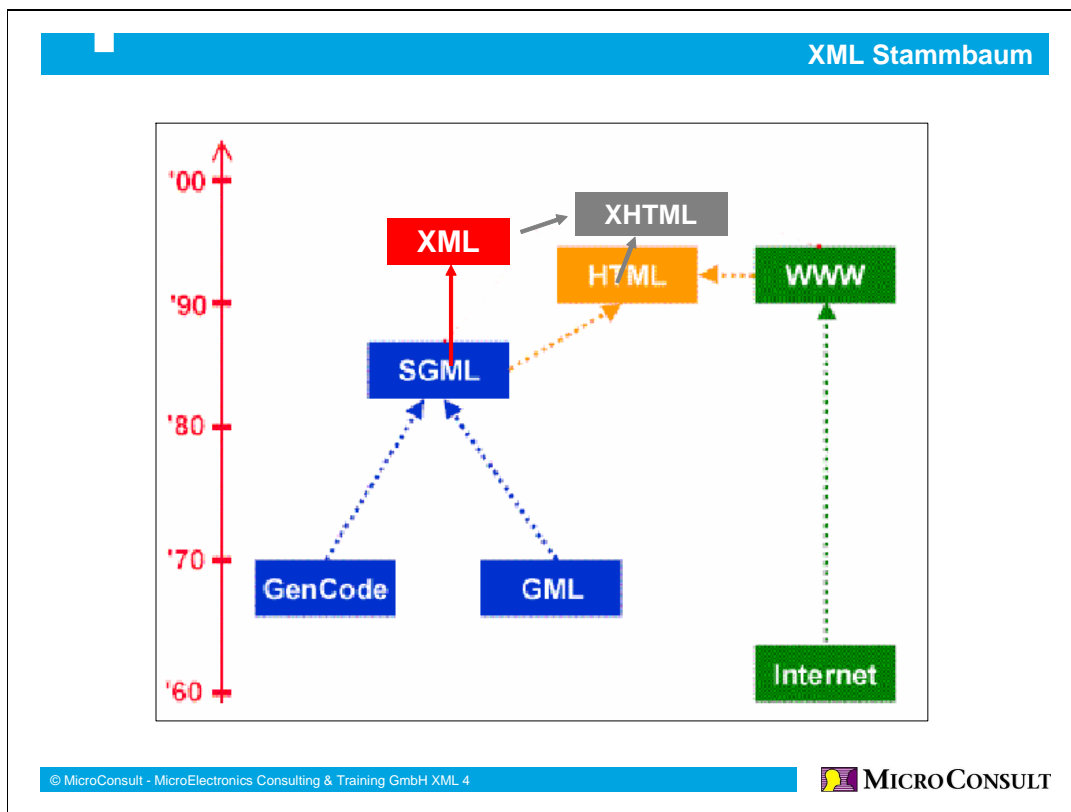
© MicroConsult - MicroElectronics Consulting & Training GmbH XML 3

Extensible → XML bietet keine (!) vordefinierten Tags, sondern Standardvorgaben, um eigene Tags zu entwerfen, damit Daten strukturiert gespeichert und übertragen werden können → erweiterbar.

Zum Vergleich: HTML besteht aus einer fest definierten Anzahl sogenannter Tags zur Präsentation von Webinhalten

"**Markup**" bezeichnet die Folge von Zeichen oder anderen Symbolen, die Sie an bestimmten Stellen einer Textdatei einfügen, um anzugeben, wie die Datei beim Drucken oder Anzeigen aussehen soll, oder um die logische Struktur des Dokuments zu beschreiben. Die einzelnen Markup-Elemente werden als "**Tags**" bezeichnet.

2 XML Stammbaum



Ende der 60er Jahre haben Charles Goldfarb, Edward Mosher und Raymond Lorie bei IBM die Auszeichnungssprache GML (Generalized Markup Language) für Textformatierung entwickelt. Die Sprache wurde für die firmeninterne Dokumentation erfolgreich eingesetzt. Die Aufbereitung der Dokumente erfolgte, wie damals üblich, im Batch-Betrieb. Etwa zur gleichen Zeit wurde bei der GCA-Organisation (Graphic Communications Association) ein Verfahren namens GenCode konzipiert, um generische Formatierungscodes für Satzsysteme verschiedener Hersteller definieren zu können.

Die chronologische Entwicklung der heute eingesetzten Auszeichnungssprachen:

1950: Hypertext als Theorie entwickelt von Ted Nelson

1969: Generalized Markup Language von IBM entwickelt

1986: Standard Generalized Markup Language festgelegt im ISO Standard 8879 (SGML)

1989: Hyper Text Markup Language entwickelt von Tim Berners-Lee im CERN in Genf (HTML)

1994: Hyper Text Markup Language 2.0 erstmals unter Leitung des W3C als Standard verabschiedet (HTML 2.0)

1994: Cascading Style Sheets 1.0 als Ergänzung zu HTML verabschiedet (CSS 1.0)

1996: Hyper Text Markup Language 3.2 verabschiedet nachdem es zu Version 3.0 keine Einigung gab (HTML 3.2)

1996: Extensible Markup Language 1.0 zunächst als Diskussionsvorschlag verabschiedet (XML 1.0)

1997: Hyper Text Markup Language 4.0 im Dezember als Richtlinie vom W3C verabschiedet (HTML 4.0)

1998: Cascading Style Sheets 2.0 Weiterentwicklung des bestehenden CSS-Standards (CSS 2.0)

1998: Extensible Markup Language 1.0 als Standard vom W3C beschlossen (XML 1.0)

1998: Extensible Style Language 1.0 im August vorerst als Arbeitsvorschlag zur Diskussion gestellt (XSL 1.0)

3 XML-Szenario


3.1 Ist-Zustand

XML-Szenario

Ist – Zustand

In gegenwärtigen IT – Umgebungen spielt der Datenaustausch zwischen unterschiedlichen Systemen eine herausragende Rolle

- a) Stichwort : E-Commerce - B2B / B2C
- b) Vernetzung von überregional wirkenden Firmen

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 5

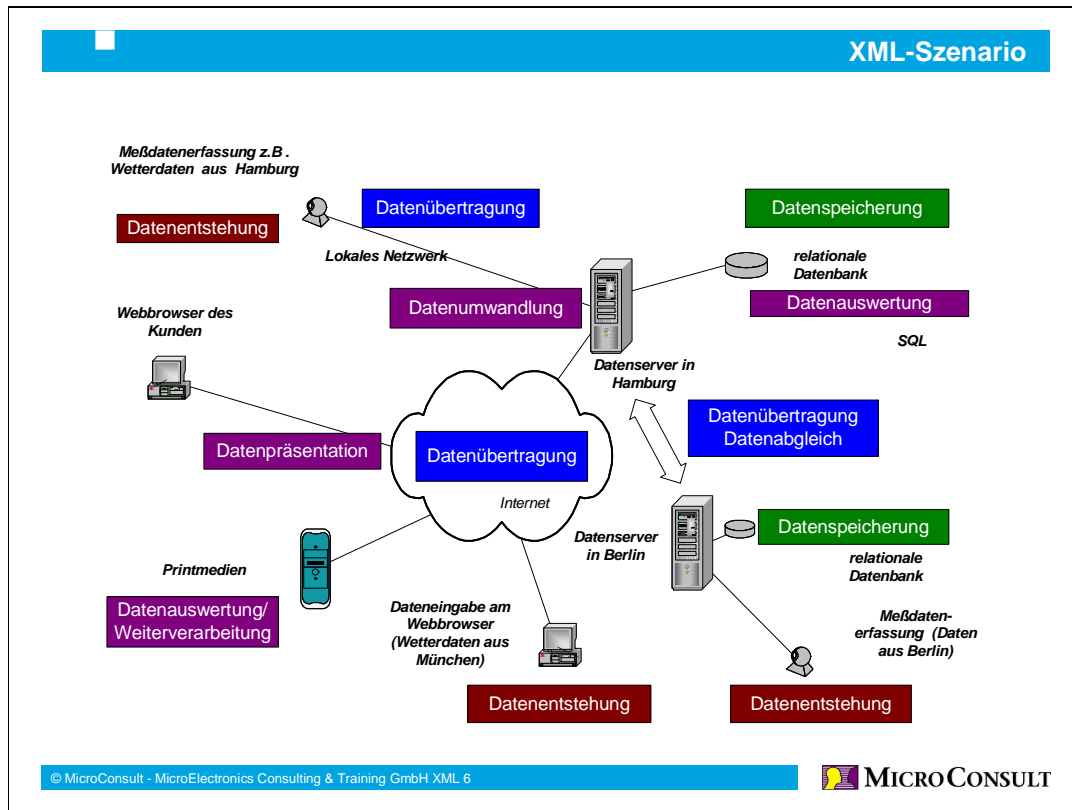
E-Commerce

- Informationsaustausch zwischen Kunden, Großkunden, Händlern, Herstellern (EDI – Electronic Data Interchange)

Vernetzung von überregional wirkenden Firmen

- durch die Bedeutung des Internets ist der Kommunikationsbedarf innerhalb einer Firma stark angestiegen.

3.2 Aufgabenstellung



Szenario : Deutschlandweite Erfassung von Wettermessdaten und Bereitstellung dieser Information für unterschiedlichste Zwecke

In diesem Beispiel werden folgende Fragen deutlich:

- In welchem Format entstehen die Daten an den einzelnen Erfassungspunkten?
- In welchem Format werden die Daten jeweils übertragen?
- In welchem Format werden die Daten auf den Servern gespeichert?
- Mit welchen Mitteln (Programmiersprachen) werden die Daten verarbeitet, ausgewertet oder umgewandelt?
- Mit welchen Mitteln verarbeiten Dritte diese Daten weiter (z.B. Kunden oder Printmedien)?
- Wie werden die Daten im Internet präsentiert?

Gibt es in der Realität für jede dieser Fragen mehrere Antworten bedeutet das:

- ständiges Anpassen der anfallenden Daten an das örtliche Datenformat
- Vorhandensein von hochspezialisierten Entwicklern, die nur Ihr Datenformat beherrschen
- keine Möglichkeit der Nutzung von gemeinsamen Programmressourcen

diese Liste ließe sich beliebig fortsetzen.

XML-Szenario

Deshalb : Wunsch nach ...

- Nutzung eines gemeinsamen standardisierten Datenformates zum Datenaustausch
- Nutzung von standardisierten Werkzeugen die dieses Datenformat verarbeiten können und eine Schnittstelle für andere Formate bieten

Gegebenenfalls könnte eine Nachricht per Email so aussehen:

„Es ist der 20.01.2001 um 12:30 Uhr. In München ist es gerade heiter und es sind kühle 5 Grad Celsius“.

„München“ ist offenkundig der Ort der Messung,
„20.01.2001“ ist das Datum der Messung,
„12:30 Uhr“ ist die Uhrzeit,
„heiter“ ist die Bewölkung und
„5 Grad Celsius“ ist die aktuelle Temperatur.

Nehmen wir an die zentrale Wetterdienststelle unterhält ein Netz von Außenstellen, die in bestimmten Zeitabständen aktuelle Wetterdaten von ihren Standorten übermitteln.

Der Mensch als Empfänger dieser Nachricht kann die semantischen Informationen leicht erkennen und verarbeiten.

- bei dieser Datenübertragung spielt der Mensch bei der Deutung die Hauptrolle
- die Datenübertragung muss im menschlich verständlichen Format erfolgen

Das Problem liegt bei der Automatisierung von Datenerfassung:

Wie soll eine Maschine (oder Software) die Semantik der Wörter erkennen?

Wir müssen ihr sagen, wie die einzelnen Wörter zu interpretieren sind. Dies kann zum Beispiel für optische Texterkennungssysteme durch Anordnung der Wörter auf der Seite (Formulare) oder durch Kennzeichnung der Wörter mitgeteilt werden. Die Bestellung könnte dann aussehen wie in der folgenden Folie:

3.3 Lösungsmöglichkeit 1


XML-Szenario

Lösung 1

Textbasiertes Datenformat:

```
0 | München | 20.01.2001 | 12:30 | heiter | 5.0  
1 | Hamburg | 20.01.2001 | 12:45 | bewölkt | 6.5  
2 | Berlin | 20.01.2001 | 12:50 | regen | 6.0
```

```
0;München;20.01.2001;12:30;heiter;5.0  
1;Hamburg;20.01.2001;12:45;bewölkt;6.5  
2;Berlin;20.01.2001;12:50;regen;6.0
```

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 8

textbasierte Datenformate:

- enthalten keine Semantik, welche die Bedeutung erklären
- die Trennzeichen müssen zwischen Sender und Empfänger vereinbart werden
- für hierarchische Daten werden mehrere unterschiedliche Trennzeichen benötigt
- Erweiterungen der Struktur führen zur Versionsinkompatibilität
- für den gezielten Zugriff müssen Algorithmen programmiert werden

Binärdatenformat:

- Daten können nur mit speziellen Programmen gelesen und erstellt werden.
- Die Daten sind nicht plattformunabhängig.

3.4 Lösungsmöglichkeit 2 (XML)

XML-Szenario


Lösung 2

XML Datenformat {

```
<messdaten ort="Muenchen">
  <datum>20.01.2001</datum>
  <uhrzeit>12:30</uhrzeit>
  <bewoelkung>heiter</bewoelkung >
  <temperatur>5.0</temperatur >
</messdaten>
```


XML – Datenformat

- von Maschinen und Menschen interpretierbar
- standardisiert durch W3C
- standardisierte Werkzeuge

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 9

Jetzt hat die Maschine alle Informationen, um die Nachricht verarbeiten zu können. Die sog. *Marken* zeichnen die Teile des Dokumentes aus und ordnen ihnen damit eine Bedeutung zu. Die Maschine wird natürlich dadurch nicht intelligenter, die Marken lösen nur die einprogrammierten Verarbeitungsschritte aus.

XML Datenformat

- lesbar und interpretierbar durch Datenspeicherung im ASCII – Format
- herstellerunabhängige Standardisierung durch das World Wide Web Consortium (W3C)
- durch offenen Standard Vorhandensein einer Vielzahl von Werkzeugen zur Bearbeitung dieses Formats

Was ist XML?

X M L

**... ist das Datenaustauschformat
nicht nur für das Internet, sondern für
alle Prozesse, wo Daten entstehen,
übertragen und verarbeitet werden!**

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 10

 MICROCONSULT

4 XML statt HTML

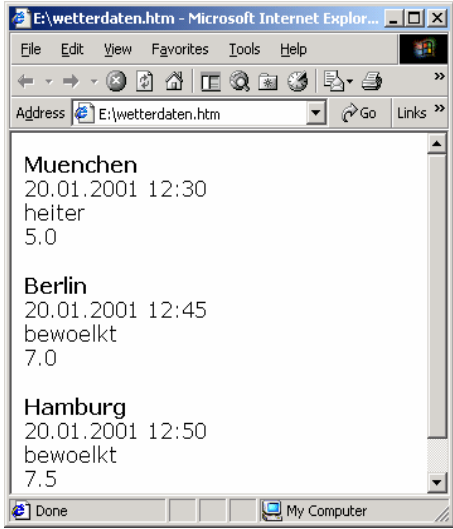
XML statt HTML

Eine **HTML Datei** mit normalen Formatierungen:

Formatierungen

Text

HTML - Seite



The screenshot shows a web browser window titled 'E:\wetterdaten.htm - Microsoft Internet Explor...'. The address bar shows 'E:\wetterdaten.htm'. The page content is as follows:

```


Muenchen
20.01.2001 12:30
heiter
5.0

Berlin
20.01.2001 12:45
bewoelkt
7.0

Hamburg
20.01.2001 12:50
bewoelkt
7.5

```

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 11



HTML → Browser sind in der Lage Textinformationen und standardisierte Formatierungsinformationen (Tags) in eine Präsentationsformat umzuwandeln

Es ist kaum oder nur sehr schwer möglich der Software die in einer HTML-Datei vorhandenen Textinformationen auszuwerten und weiterzuverarbeiten

HTML Quelltext für obige Datei :

```

<html>
  <head></head>
  <body>
    <p><font face="Verdana"><b>Muenchen</b><br>
    20.01.2001 12:30<br>
    heiter 5.0 </font></p>
    <p><font face="Verdana"><b>Berlin</b><br>
    20.01.2001 12:45<br>
    bewoelkt 7.0 </font></p>
    <p><font face="Verdana"><b>Hamburg</b><br>

```

XML statt HTML

```
20.01.2001 12:50<br>
bewoelkt&nbsp; 7.5 <br><br></font></p>
</body>
</html>
```

XML statt HTML

Eine **XML Datei** ohne Stylesheet im Internet Explorer 5.0:

semantische
Strukturierung der
Informationen

Text

XML Datei

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 12

XML → Speicherung von Textinformationen und dazugehörigen semantischen Informationen.

sowohl Mensch, als auch Maschine sind in der Lage den Text anhand der Bedeutungsinformation auszuwerten

Der Quelltext der XML Datei (mit Stylesheet!):

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<?xml-stylesheet type="text/css" href="Kunden.css" ?>
```

```
<DATENBANK>
```

```
  <ADRESSE>
```

```
    <VORNAME>Frank</VORNAME>
```

```
    <NACHNAME>Meier</NACHNAME>
```

```
    <STRASSE>Hauptstrasse 3</STRASSE>
```

```
    <PLZ>80345</PLZ>
```

```
    <ORT>Muenchen</ORT>
```

```
  </ADRESSE>
```

```
  <ADRESSE>
```

```
    <VORNAME>Heike</VORNAME>
```

XML statt HTML

```
<NACHNAME>Schoene</NACHNAME>  
<STRASSE>Am Waldsaum 3</STRASSE>  
<PLZ>80377</PLZ>  
<ORT>Muenchen</ORT>  
</ADRESSE>  
</DATENBANK>
```

XML statt HTML

Eine XML Datei mit Stylesheet
(für Ausgabe im Browser):

Formatierungen

semantische
Strukturierung der
Informationen

Text

XML - Ausgabe

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 13

Um XML zu präsentieren werden standardisierte Werkzeuge bereitgestellt, die es ermöglichen Informationen in Browsern, auf Druckern oder anderen Medien (Sprachausgabe) bereitzustellen.

z. B. die folgenden in späteren Kapiteln beschriebenen Werkzeuge :

- Cascading Style Sheets
- Extensible Style Sheet Language

Inhalt der CSS-Datei für das Beispiel:

DATENBANK { font-family: Verdana ; font-size:10pt }

ADRESSE { font-family: Verdana; font-weight: bold; display: block; margin-top: 12pt}

VORNAME { font-style:italic }

NACHNAME { font-style:italic }

STRASSE { font-style:italic;display: block }

PLZ { font-style:italic}

ORT { font-style:italic}

Gründe für den XML-Einsatz

- XML ist ein offener Standard mit hoher Akzeptanz.
- XML unterstützt eine klare Trennung zwischen Daten und Präsentation.
- XML ist textbasiert.
- XML ist erweiterbar.
- XML ist selbstbeschreibend.
- XML ist plattform- und programmiersprachenunabhängig.
- XML ist maschinell verarbeitbar.
- XML ist geeignet für langfristige Datenablage.
- XML ist leicht transformierbar.

Die Entwurfsziele im Internet:

<http://www.w3.org/TR/REC-xml/>

deutsche Übersetzung :

<http://www.mintert.com/xml/trans/>

XML Vorteile:

XML stellt einen allgemeingültigen Standard dar

XML Dateien sind selbsterklärend (enthalten Metadaten → Daten über Daten)

XML ist auf eigene Bedürfnisse anpaßbar (erweiterbar)

XML bietet die Möglichkeit Gültigkeitsregeln für Daten zu vereinbaren

XML kann mit jedem ASCII Editor erzeugt und gelesen werden

XML kann mit Standardwerkzeugen verarbeitet werden

in XML kann durch Namen auf Detaildaten zugegriffen werden

XML Dateien können erweitert werden ohne daß die Kompatibilität gefährdet wird


5 XML-Anwendungen

XML-Anwendungen

XML-Vokabular / XML-Anwendungen

- Unter einer *Anwendung* von XML versteht man nicht ein konkretes Programm, sondern ein XML-Vokabular für einen bestimmten Anwendungsbereich, das zum Beispiel durch eine DTD oder ein Schema definiert sein kann.
(→ auch bezeichnet als Auszeichnungssprachen)

- Es gibt Anwendungen in jeder Komplexitätsstufe und für fast jedes Gebiet.

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 15

Nach der Veröffentlichung der XML-Spezifikation wurde eine wahre Flut von XML-Anwendungen vorgeschlagen, und es ist noch kein Ende abzusehen.

Liste von XML-Anwendungen unter:


<http://www.schema.net>

<http://www.xml.com>

XML-Anwendungen

Beispiele für XML–Auszeichnungssprachen:

- SMIL - Synchronized Multimedia Integration Language
- CML - Chemical Markup Language
- MathML - Math Markup Language
- CDF - Channel Definition Format
- WIDL - Web Interface Definition Language
- XHTML – Extensible HTML

© MicroConsult - MicroElectronics Consulting & Training GmbH XML 16

SMIL

Synchronized Multimedia Integration Language

Zur Integration und Synchronisation von Multimedia-Inhalten.

CML

Chemical Markup Language

Zur Beschreibung von chemischen Vorgängen, beispielsweise des Molekülaufbaus.

MathML

Math Markup Language

Zur Beschreibung komplexer mathematischer Formeln.

CDF

Channel Definition Format

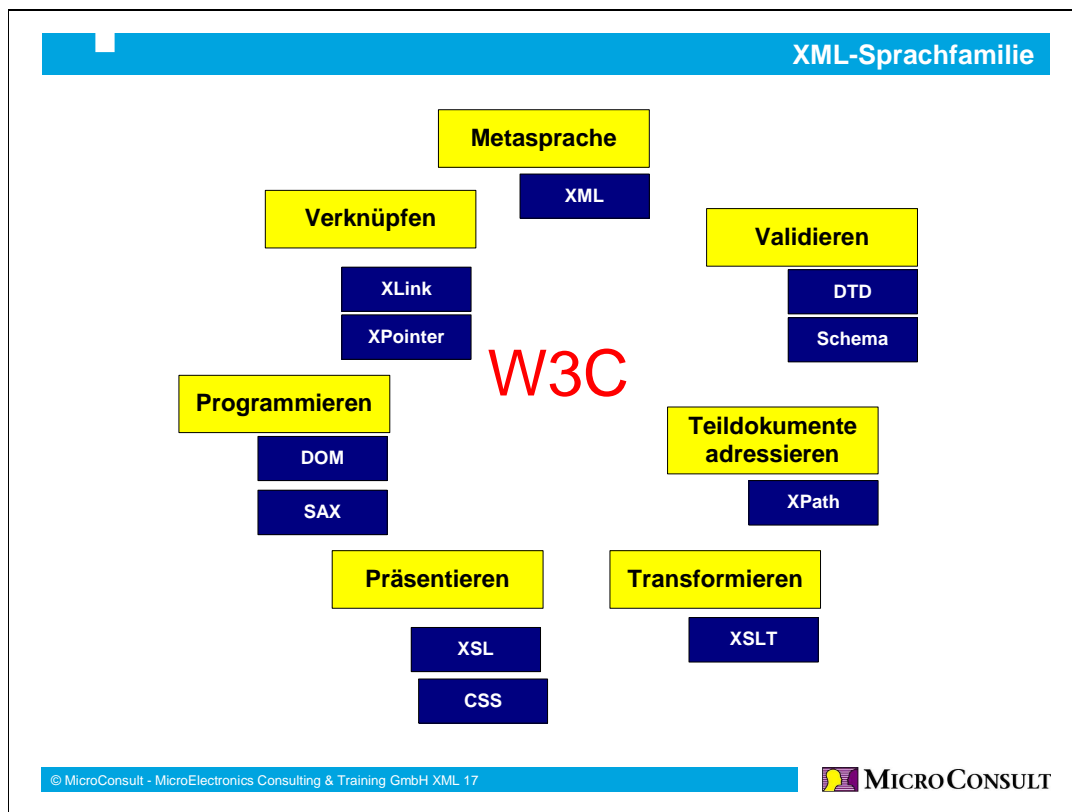
Von Microsoft eingeführtes Format, das die Push-Technologie unterstützt (Datenübertragung vom Server zum Client, ohne daß Client die Daten anfordert).

WIDL

Web Interface Definition Language

Neue objektorientierte Sprache zum Erstellen dynamischer Websites.

6 XML-Sprachfamilie



XML ist mehr als nur eine Metasprache – XML ist inzwischen ein Gattungsbegriff für eine ganze Reihe von Technologien

Abkürzungen :

DTD – Document Type Definition

DOM – Document Object Model

SAX – Simple Api for XML

XSL – Extensible Stylesheet Language

XSLT – XSL Transformation

CSS – Cascading Style Sheets

Informationen zu anderen XML finden Sie unter:

<http://www.w3.org/XML/>