

Barbara Barth

## Was bedeutet eigentlich... XML?

### Am Anfang war das Dokument...

Ausgehend von der Vorstellung eines an der Schreibmaschine sitzenden Schriftstellers, der in seinen Texten nicht viel mehr Formatierungsmöglichkeiten als die des Unterstreichens zur Verfügung hatte, über den mit meist nicht sichtbaren Steuerzeichen kämpfenden Schreiberlings an einem PC war die weitgehend identische Darstellung von Texten auf Papier und Bildschirm mit der Einführung einer graphischen Oberfläche ein großer Entwicklungsschritt.

Unter den Stichworten "What you see is what you get" (WYSIWYG) und "Desktop Publishing" (DTP) konnte nun am Layout, an der Formatierung und an der Darstellung von Texten auf den verschiedensten professionellen Ebenen gearbeitet werden.

Dieser formatorientierte Ansatz ist wenig erfolgreich, wenn außer Papier noch andere Ausgabemedien genutzt werden. Wer kennt nicht die Formatierungsverluste bei einem einfachen "Kopieren und Einfügen"?

Betrachten wir exemplarisch die formalen Bestandteile eines Textdokumentes:

- Inhalt: Text, Bilder, Tabellen, Videos usw.
- Struktur: Aufteilung des Textes in Kapitel, Abschnitte, Überschriften usw.
- Präsentation: Formatierung, Layout, die gewählte Schriftart usw.

Für eine Weiterverarbeitung von Textdokumenten ist der Verlust all der logischen Informationen, die in der Struktur des Textes vorgegeben sind, sicherlich nicht sinnvoll. Legen wir den Schwerpunkt bei der Textverarbeitung also nicht auf das Aussehen sondern auf dessen logische Struktur.

Diese schon mehr als 30 Jahre alte Idee ist im ISO-Standard 8879 im Jahre 1986 als Standard Generalized Markup Language (SGML) veröffentlicht worden.

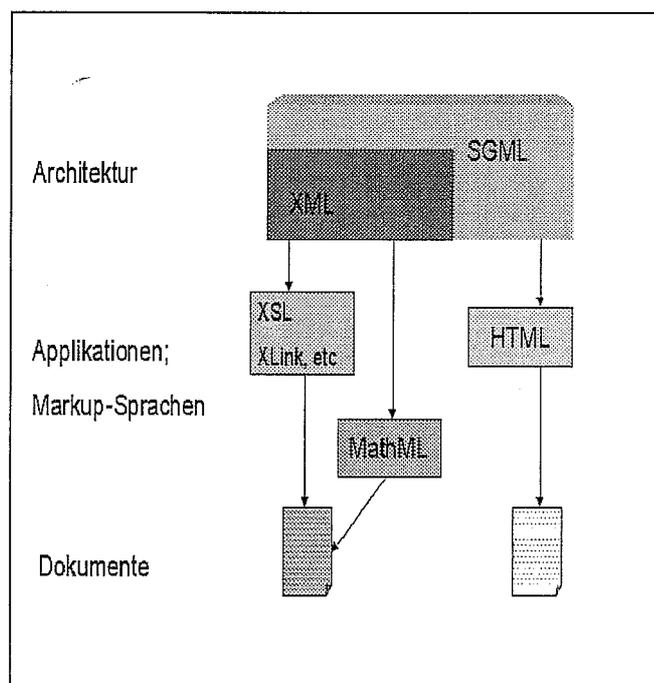
SGML ist eine grundlegende Architektur - eine Metasprache - mit der man Markup-Sprachen definieren kann. Die wohl prominenteste Anwendung von SGML ist die Internetsprache Hypertext Markup Language (HTML). Diese durch ihre Ein-

fachheit bestechende Auszeichnungssprache hat den Siegeszug des World Wide Web Mitte der 90er Jahre maßgeblich beeinflusst. Doch den steigenden Anforderungen des für immer komplexere Anwendungen genutzte WWW kann diese Sprache nicht gerecht werden.

Zur Lösung dieses Problems gründete das sogenannte WWW-Konsortium (W3C) 1996 eine Arbeitsgruppe, die auf der Basis von SGML eine Web-taugliche, universelle Markup Sprache entwickeln sollte. Es entstand die Extensible Markup Language (XML), deren Spezifikation im Februar 1998 vom W3C als "XML 1.0 Recommendation" veröffentlicht wurde.

XML ist also eine Teilmenge von SGML mit wesentlich geringerer Komplexität. Den 33 gedruckten Seiten für die formale Definition der Metasprache XML stehen 500 für SGML gegenüber.

Veranschaulichen wir uns das in der folgenden Abbildung (aus dem Aufsatz "XML" von Stefan Gillert, Universität Würzburg):

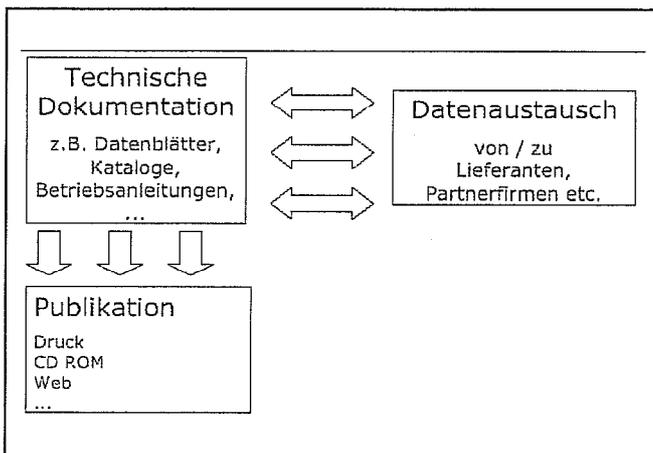


Ohne auf technische Details einzugehen lassen sich gemäß W3C folgende Eigenschaften von XML formulieren:

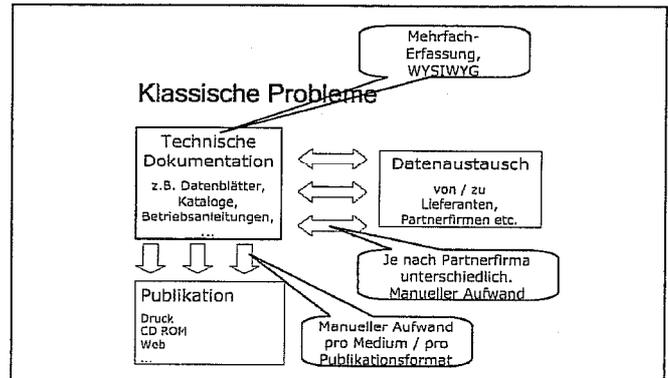
- XML ist eine Methode, um strukturierte Daten in einer Textdatei darzustellen;
- XML sieht fast aus wie HTML, ist aber kein HTML;
- XML ist Text, aber nicht zum Lesen;
- XML ist vom Design her ausführlich; XML- files sind größer als binäre Formate;
- XML ist eine Familie von Techniken;
- XML ist lizenzfrei, plattformunabhängig und gut unterstützt.

Die Metasprache XML erlaubt die Beschreibung, den Austausch, die Darstellung und die Manipulation von strukturierten Daten, so dass diese Daten von einer Vielzahl von Anwendungen genutzt werden können. Dabei ist das Grundkonzept die Trennung von Inhalt, Struktur und Layout. XML-Dokumente unterscheiden sich auf den ersten Blick nicht wesentlich von HTML-Dokumenten. Auch XML-Dokumente bestehen aus durch Tags ausgezeichneten Inhalten. Während die Anzahl und Benennung der Tags für HTML aber vorgegeben sind, können für XML-Dokumente beliebig viele und frei ("semantisch") benannte Tags verwendet werden.

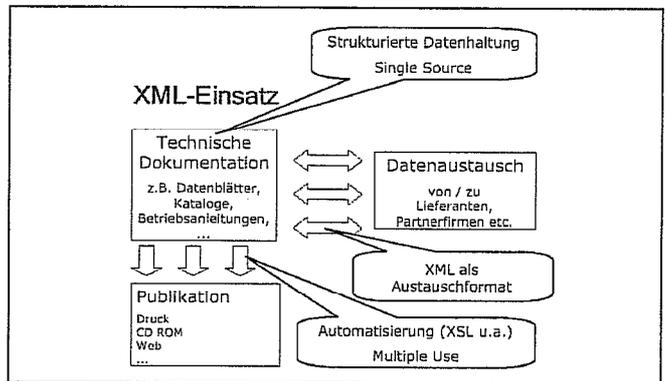
Typischerweise haben wir in der Praxis das folgende Szenario ("XML in der Praxis, Ulrike Schäfer, www.infotakt.de):



mit den klassischen Problemen:



Der Einsatz von XML für den Dokumentenaustausch und für Dokumentation und Publishing, liefert einen guten Ansatz zur Problemlösung:



## Was ändert sich?...

Für den normalen HTML-Anwender, der lediglich einfach strukturierte Informationen über das Netz anbieten möchte und dem die Gestaltungsmöglichkeiten von HTML keine Grenzen setzen, wird die Einführung von XML keine Sensation bedeuten.

Anwender mit speziellen Anforderungen an das Layout eines Dokumentes werden in Zukunft Standards haben, die einen einfachen, plattformunabhängigen Datenaustausch beinahe beliebiger Daten möglich machen werden. Für das Management vieler und großer Dokumente bietet XML eine einheitliche, weitverbreitete und akzeptierte Architektur.