# XML

## FÜR ZEITSCHRIFTENVERLAGE

XML ist die perfekte Basis für die Mehrfachverwendung von Inhalten – und damit in Zeiten des Multi-Channel-Publishings für Verlage unverzichtbar.

Von Klaus Kresse, Lehrbeauftragter, Journalist, Medienberater, Appenweier

Warum, bitte, machen wir uns die Welt oft komplizierter, als sie sein müsste? Und warum arbeiten wir meist mehr, als nötig wäre? Die Antwort ist einfach: Das alles ist nur wenigen von uns bewusst. Und wer etwas grundlegend ändern will, gewinnt ohnehin nicht den Beliebtheitswettbewerb. Nicht umsonst singen die Bremser das Hohelied des Beharrens. Tenor: Das haben wir schon immer so gemacht. Punkt. Aus.Kommt jetzt noch hinzu, dass sich das Neue mit einem so unsympathischen Kürzel wie XML schmückt, wird die Lage ganz düster.

Wobei an dieser Stelle schon die erste Korrektur nötig wäre. Denn XML ist eigentlich ein alter Hut. Die drei Buchstaben stehen für Extensible Markup Language, was so viel heißt wie erweiterbare Auszeichnungssprache. Die Geburtsurkunde wurde am 10. Februar 1998 vom World Wide Web Consortium ausgestellt, also vor mehr als 23 Jahren. Und wer in den 90er Jahren mit FrameMaker gearbeitet hat, erinnert sich an SGML, die Standard Generalized Markup Language – wenn man so will, ein Vorgänger und naher Verwandter von XML. Teile der

SGML-Erbschaft finden sich in HTML. XML wird in Expertenkreisen daher als eine Untermenge von SGML bezeichnet. In der Praxis hat XML bereits in vielen Bereichen SGML abgelöst.

## Medienneutrale Datenaufbereitung

Das Konzept hinter XML ist genial. Denn damit wird es möglich, Daten zwischen verschiedenen Computersystemen auszutauschen (siehe Kasten: „Ein Muss für Medienunternehmen“), viele für Verlage typische Prozesse zu automatisieren und die Wertschöpfung auf Basis des Contents deutlich zu erhöhen. Da klingeln den Controllern natürlich sofort die Ohren. Aber hier sei gewarnt: Natürlich ist XML aus der Verlagsbranche inzwischen nicht mehr wegzudenken. Aber der Umstieg aus der Old-School-Welt ist alles andere als trivial. Vor allem in vielen Köpfen müssen eingerostete Hebel umgelegt werden. Denn vielen geht es wie unserem letzten Kaiser Wilhelm II., der nach der Erfindung des Automobils dreist ankündigte: „Die Zukunft gehört dem Pferd!“ Also nähern wir uns dem Thema am besten ganz vorsichtig.

Für jede Führungskraft in einem Verlag ist es mittlerweile eine Binse, dass der Kampf um die Aufmerksamkeit des Lesers (neudeutsch: User) nicht auf einem Distributionskanal gewonnen werden kann. Wer vorne mitmischen will, muss die Formel A hoch 3 zum Leitmotiv machen. A hoch 3 gleich anything, anytime, anywhere. Das hatten sich ja Leser, Hörer und User schon immer gewünscht. Informationen zu einem beliebigen Thema nach Wahl, und das zu jeder Zeit und an jedem Ort. Vor Internet und mobilen Endgeräten war das ein frommer Wunsch. Inzwischen aber ist es Standard. Will heißen: Mal wird der Kanal Print genutzt, dann wieder die Website, der

© MarkStein

Das Publishing-System tango CloudGeneration bietet einen ganzheitlichen Ansatz für das Publizieren von Inhalten über alle Kanäle

E-Mail-Newsletter, ein Blog oder diverse Social-Media-Kanäle. A hoch 3 eben.

Mit redaktioneller Arbeit und Medienproduktion wie dunnemals ist das nicht zu stemmen. Denn Inhalte, wie wir sie früher erzeugten, sind dumm. Und zudem sind sie überfrachtet mit Datenmüll, der nur in einem Kanal sinnvoll sein mag, in allen anderen aber nur stört.

Von Tobias Ott, Geschäftsführer der Tübinger pagina GmbH und anerkannter XML-Experte, gibt es dazu ein anschauliches Beispiel: Stellen Sie sich vor, Sie würden mit Microsoft Word den Text „Zutaten: 1 frisches Huhn“ schreiben und speichern. Dann entstünde eine Datei in einer Größe von rund 20 Kilobyte, also 20.000 Byte. Der reine Text freilich hat nur 24 Zeichen, was lächerlichen 24 Byte entspricht. Woher kommen die anderen rund 20.000

Foto: unsplash (Alexander Sinn, Next Academy)

## EIN MUSS FÜR MEDIENUNTERNEHMEN

Ehrhardt F. Heinold, Gründer und alleiniger geschäftsführender Gesellschafter des Hamburger Beratungs-Unternehmens Heinold, Spiller & Partner, hat eine klare Ansage an die Verlage: „Technologien und Speicherformate wechseln, in XML gespeicherte Inhalte sind ‚Plain Text’ und können von jedem Editor gelesen werden. XML ist kein ‚proprietäres Format’ und gehört keiner Firma (wie Word/Microsoft oder InDesign/Adobe), sondern ist ein offener, internationaler Standard. XML bietet somit das einzige sicher archivierbare zukunftssichere Format. – XML hat sich als eines der wichtigsten Austauschformate entwickelt, das die Brücke zwischen den unterschiedlichsten Programmen bilden kann: So können Word und InDesign XML im- und exportieren, das Internet basiert mit HTML und HTML5 ohnehin auf XML. Wenn Medienunternehmen Inhalte austauschen, nutzen sie fast immer XML, so bietet dpa seinen Newsfeed als NewsML-G2 an. Ein Medienunternehmen muss heute XML-Feeds liefern (und importieren) können, um austauschfähig zu sein.“

Ehrhardt F. Heinold, geschäftsführender Gesellschafter des Hamburger Beratungs-Unternehmens Heinold, Spiller & Partner, hält XML für Medienunternehmen für unverzichtbar

© Heinold, Spiller & Partner

© MarkStein Software GmbH

## GANZ KLAR: CONTENT FIRST

Michael Stühr, Co-Geschäftsführer der Mark-Stein Software GmbH, favorisiert Content First für Verlage, die Inhalte für Zeitschriften und das Internet produzieren

Diplom-Ingenieur Michael Stühr, Co-Geschäftsführer der Mark-Stein Software GmbH in Darmstadt (Redaktionssystem tango) favorisiert eine „Content-First-Strategie“: „Die zielführende Strategie für Verlage, die Inhalte für Zeitschriften und das Internet produzieren, ist ‚Content First’. Hier werden die Inhalte zentral produziert und dann in Print und die digitalen Kanäle verwertet. Wichtig ist hierbei eine einheitliche Aufbereitung der Inhalte, so dass keine Einbahnstraße nach Print entsteht. Unser Publishing-System tango unterstützt den medienbruchfreien Content-First-Ansatz für Print und Online in umfänglicher Weise durch einen neutralen Web-Editor, eine Bibliothek von CSS-Komponenten für Online, Echtdarstellung für Print, durchgängige Strukturierung über Formatgruppen und Nomenklatur, ein einheitliches XML-Schema im Hintergrund und hybride Texte (absatzweise Textnutzung für

Byte? Nun, darin sind all die Informationen codiert, die Word benötigt, um diesen Text auf ein Stück Papier drucken zu können: Format des Dokuments, Größe der Ränder unten und oben sowie rechts und links, Schriftart, -schnitt, -größe und -farbe, Buchstaben und Zeilenabstand. Und. Und. Und.

Das Fatale dabei: So sinnvoll dieser Wust für die Druckausgabe sein mag – sobald andere Kanäle (Web etc.) ins Spiel kommen, hilft er nicht nur nicht, er stört sogar und muss entfernt werden. Erst dann können die für die Darstellung notwendigen Informationen drübergestülpt werden, on the flight sozusagen.

Die Alternative zum dummen Inhalt wird in der angelsächsischen Welt „smart content“ genannt, frei übersetzt also „kluger Inhalt“. Das Fachblatt „Information Week“ hat sich mehrfach mit dem Thema beschäftigt und die Pluspunkte von Smart Content etwa so auf den Punkt gebracht: Kluger Inhalt ist viel leichter aufzufinden (zum Beispiel in Datenbanken und Archiven), er ist viel besser wiederzuverwenden, und er erweist sich als profitabler für den Produzenten und mehr Nutzen bringend für den Kunden (Leser, User etc.).

### Immer komplexere Contentszenarien

Wie das mit XML möglich wird: Wer schon mit HTML in Berührung gekommen ist, kennt Auszeichnungsbefehle in spitzer Klammer, also die Start- und die End-Tags. Wobei die Möglichkeiten mit XML allein deshalb ungleich größer sind, weil hier – wie der Namensteil „extensible“ ausweist – der Befehlssatz individuell erweiterbar ist.

Grundsätzlich gibt es in XML drei Gruppen von Strukturauszeichnungen (Markups). Tobias Ott unterscheidet:
– prozedurale Markups, die „zur maschinellen Verarbeitung eines Textes (in den gewünschten Ausgabeformen) erforderlich sind. Das prozedurale Markup bildet die formalen Strukturen … ab“.
– deskriptive Markups, die „den Text inhaltlich beschreiben. Das deskriptive Markup bildet die semantischen Strukturen ab“ (Semantik gleich Bedeutung)
– hermeneutische Markups, die „den Text (auch konkurrierend) interpretieren. Das deskriptive Markup bildet den Blick des Nutzers auf ein Werk ab“.

Damit bietet XML eine Fülle von Vorteilen. Dies war das zentrale Thema des hochkarätigen CrossMediaForums (CMF), zu dem das auf Medienunternehmen spezialisierte Beratungsunternehmen Heinold, Spiller & Partner im Juli dieses Jahres zum 23. Mal eingeladen hatte. Mehr als 1.220 Gäste aus der Publishing-Welt fühlten sich vom Motto der Veranstaltung („XML First, XML Last – Product First, Product Last: Unterschiedliche Konzepte für die Contentproduktion“) angezogen.

Kein Wunder, gab doch Ehrhardt Heinold (siehe auch Kasten „Ein Muss für Unternehmen“) im CMF-Sonderheft zu Protokoll: „Was sich ebenfalls seit 1988 geändert hat, ist die Vielfalt von Medien, Nutzungsgeräten, Publikationskanälen und vor allem von Formaten und Publikationsformen. Damit hat sich bewahrheitet, was Ende der 90er absehbar war: Verlage müssen immer komplexere Contentszenarien bewältigen, nicht nur für Produkte, sondern vor allem auch für die wachsende Zahl der Marketingkanäle, von der E-Mail über den Newsletter bis hin zu Social-Media-Postings. Neben einer möglichst medienbruchfreien Content-Kette sollten zusätzlich alle Beteiligten so in den Workflow integriert werden, dass die Inhalte strukturiert und automatisiert von einer Station zur nächsten fließen – also beispielsweise von einem freien Redakteur per Smartphone zur Editorin-in-Chief im Homeoffice und von dort in das gedruckte Magazin, auf die Website, in die App.“

Damit ist klar, dass ein Distributionskonzept „Single Source/Multi Channel“ nach solch einem Ansatz förmlich schreit. Aber auch das Thema „Wiederverwendbarkeit“ hat Gewicht. Heinold: „Mit Hilfe von XML können

Inhalte auf Basis der Bedeutung der einzelnen Elemente strukturiert werden, ähnlich wie bei den Formatvorlagen von Word oder den Stilvorlagen bei InDesign – allerdings mit einem wesentlich ausgereifteren Konzept, das zudem keine Mischform aus Layout und Bedeutungsauszeichnung darstellt, sondern originär für eine rein semantische Auszeichnung geschaffen wurde, aber auch Layoutinformationen enthalten kann. Layoutneutrales, inhaltlich strukturiertes XML ist damit die perfekte Basis für die Wiederverwendung von Inhalten."

### Beratung empfehlenswert

Aber wie setzt man das in der Praxis um? Wie kommt man zu einer DocTypeDefinition (DTD), zu einem XLM-Schema und zu well formed XML? Hier ist nicht der Platz, auf diese Details die Lupe zu richten. Dafür sollte sich ein Verlagshaus der Unterstützung externer Experten bedienen, für die so etwas eine Fingerübung ist. Infrage kommen Unternehmen wie beispielsweise die Firma pagina von Tobias Ott, der sich auch an der renommierten Stuttgarter Hochschule der Medien einen Namen gemacht hat.

Wichtiger noch ist es, zuvor die Frage nach dem Warum und dem betriebswirtschaftlichen Nutzen zu beantworten – schließlich ist Medienproduktion keine altruistische Veranstaltung. Hier kommen erfahrene Berater wie Ehrhardt Heinold ins Spiel, der seine Kunden an die Hand nimmt, um Geschäftsmodelle zu entwickeln, die den charmanten Vorteil haben, dass sie nicht nur innovativ wirken, sondern sich auch rechnen.

### XML First oder Last?

Noch ein Weg ist denkbar. Während unter Redaktionssystemen wie K4 über Censhare und Woodwing bis Neo eine klassische DTP-Plattform wie InDesign oder QuarkXPress liegt, die – nicht ganz unkompliziert – über PlugIns angesteuert wird, fährt das Systemhaus MarkStein (siehe Kasten „Ganz klar: Content First") in Darmstadt mit tango einen konsequenten All-in-one-Ansatz. Hier sind Datenbank, Asset-Verwaltung, Blattplanung und Schreib-/Layout-Tool aus einem Guss. Denn unter tango liegt eine MarkStein-eigene DTP-Engine, so dass es keinerlei Brüche gibt. „Das bedeutet: Ein gemeinsamer Datenbestand für alle Publikationsformen. Print ist nicht zwangsläufig führend", heißt es dazu bei MarkStein.

Weshalb sich hiermit auch die leidige Diskussion um XML First oder Last erledigt.

Wobei die auch an einer anderen Front inzwischen leidlich entschärft ist. Ursprünglich galt zwar, XML First eigne sich nicht für layout-betonte Publikationen. Doch das ist inzwischen Geschichte (siehe Kasten „XML: First oder Last?").

Bleibt eigentlich nur noch eine Empfehlung: Nicht endlos über Nutz und Frommen intelligenter Content- und Medien-Produktion sinnieren – einfach machen! ●

## XML: FIRST ODER LAST?

Zunächst einmal: beide Varianten sind möglich. XML first bedeutet, dass die Inhalte in einem XML-Editor erfasst werden und somit von Anfang an sauberes XML vorliegt. Bei XML last hingegen werden die Inhalte zunächst klassisch erfasst. In der InDesign-Welt etwa im Tool InCopy. Hierbei können über den eigentlichen Inhalt hinaus auch diverse Stilvorlagen eingegeben werden, mit denen sich auch – zum Beispiel für einen Index – Bedeutungen (Semantik) kodieren lassen. So können Namen grundsätzlich als „halbfett" ausgezeichnet werden – durch den semantischen Zusatz aber lässt sich unterscheiden, ob es sich um einen Personen- oder einen Firmen-Namen handelt. Was hilfreich ist, wenn man automatisch je ein Personen- und ein Firmen-Register erzeugen will. Auf diese Weise erfasste Inhalte lassen sich (deshalb XML last) später in XML wandeln. Dieser Weg wurde längere Zeit für Layout-betonte Printobjekte bevorzugt. Grund: In den XML-Editoren war es ursprünglich weder möglich, auf Länge zu schreiben noch auf ästhetischen Zeilenfall zu achten. Auch Hurenkinder und Schusterjungen ließen sich im ersten Anlauf nicht vermeiden. Diese Einschränkungen spielen heute jedoch kaum noch eine Rolle, weil XML-Inhalte inzwischen so schnell gerendert werden können, dass präzise Satzarbeit auch beim Einsatz eines XML-Editors kein Problem mehr ist. Deshalb gilt heute überall da, wo die notwendigen technischen Voraussetzungen gegeben sind: XML first hat Vorfahrt. Übrigens: Die Texter müssen nicht zu Programmieren werden. Denn im XML-Editor können sie die jeweiligen XML-Tags (Befehle) bequem über ein Pull-Down-Menü eingeben. Was diese vordefinierten Befehle bedeuten, sollten sie aber schon wissen.