

Factsheet: XML-Workflows in der Verlagsbranche

Verlage im digitalen Zeitalter

Der digitale Wandel der Gesellschaft ist in vollem Gange, ein Ende ist nicht abzusehen. Und auch die Verlagsbranche hat sich längst den Herausforderungen eines veränderten Mediennutzungs- und Leseverhaltens, neuer Märkte und neuer Vertriebskanäle gestellt. Doch in vielen Verlagen haben diese veränderten Rahmenbedingungen noch nicht zu veränderten Produktionsprozessen geführt. Die Schere zwischen der im Verlag vorhandenen Datenqualität und dem technischen Anspruch an den Verlagscontent wird dabei immer größer.

Betrachten wir zunächst diesen Anspruch abstrakt: Ein Verlag muss heute in der Lage sein, flexibel auf seine Inhalte zugreifen und diese in beliebigen Medien mit geringem Aufwand publizieren zu können. Die Voraussetzung hierfür sind auskorrigierte Daten in einem offenen, wiederverwertbaren Format. Hier hat sich XML in der Verlagsbranche und weit darüber hinaus durchgesetzt. XML ist eine weltweit gültige, offene Auszeichnungssprache für Texte (und andere mediale Inhalte), mit deren Hilfe man Inhalte medienneutral vorhalten und mit höchstmöglichem Automatisierungsgrad in die verschiedensten Ausgabeformate ausleiten kann. „Medienneutrale Daten“ sind seit vielen Jahren eine Anforderung der Branche an den Content. Es geht dabei um das Konzept, die Verlagsinhalte nicht nur in einer formatierten Datei (also z.B. in InDesign und PDF) vorzuhalten, sondern in einer Form, die unabhängig vom Ausgabekanal die Inhalte strukturiert und damit maschinenverarbeitbar speichert. Die Trennung von Struktur und Layout ist die Grundidee von XML – und von allen digitalen Ausgabeformaten, allen voran dem Web.

Alle gängigen digitalen Ausgabeformate (Websites, EBooks, Apps) basieren daher unmittelbar auf XML. Da im Produktionsprozess digitaler Medien eigentlich keine „Korrekturläufe“ vorgesehen sind, geht man für die Produktion von bereits auskorrigierten Inhalten aus. In der Produktion von Printmedien ist das häufig anders: Hier werden die Inhalte meist noch während der Satzarbeiten überarbeitet, so dass erst zum Abschluss der Satzproduktion auskorrigierte Inhalte vorliegen. Alle gängigen Satzsysteme können heute mit XML-Daten umgehen – die einen besser, die anderen weniger gut.

Um zu XML-Daten zu kommen, gibt es daher zwei Wege, die sich grundsätzlich unterscheiden: Der „XML-last“-Ansatz beschreibt ein Vorgehen, bei dem der Verlag erst mit Abschluss der Satzarbeiten die Inhalte nach XML konvertiert. Als „XML first“ oder „Content first“ werden die Workflows bezeichnet, bei denen die zu publizierenden Inhalte zuerst nach XML überführt werden und anschließend alle Medienformen – incl. print – aus diesen Daten erstellt werden.

Wir wollen nachfolgend einen Überblick über XML als strategisches Werkzeug geben sowie die beiden Ansätze „XML first“ und „XML last“ gegenüberstellen. Wir hoffen, dass dieses Factsheet den Verlagen, die sich heute an der Schwelle zur digitalen Transformation befinden, eine nützliche Handreichung ist.

Die Vorteile von XML auf einen Blick

1. **Wiederverwendung.** XML ermöglicht die ständige Verfügbarkeit und Verwertbarkeit einmal angelegter Inhalte. XML ist nicht, wie InDesign oder Word, von Versionsaktualisierungen betroffen, sondern jederzeit aktuell.
2. **Kompatibilität.** XML ist ein offenes, nachhaltig lesbares Datenformat, das nicht an einen Hersteller gekoppelt ist. Damit ist XML die Basis für jede Art der Langzeitarchivierung und die Schnittstelle zwischen verschiedensten Systemen.
3. **Crossmedialität.** XML-Daten sind die technische Voraussetzung für die meisten elektronischen Publikationsformate. Hält ein Verlag seinen Content in XML vor, ist meist die Grundlage für digitale Publikationen bereits geschaffen.
4. **Produktion auf Knopfdruck.** Nutzen Sie XML, um Satzkosten zu sparen. Durch vereinheitlichte und hinterlegte Stylesheets ermöglicht Ihnen XML eine Konvertierung nach PDF oder EPUB (fast) auf Knopfdruck.

5. **Schnittstellen.** Es gibt keine Schnittstelle in der gängigen Software unserer Branche, egal ob PIM, CMS, App oder Webshop, die nicht mit XML umgehen kann. Sobald Inhalte in XML vorliegen, können diese in Drittsoftware einfließen. Verlage haben mit XML einen one-fits-all-content und nicht n verschiedene Dateien für m verschiedene Ausleitungen.
6. **Medienneutralität.** Die Trennung von Struktur und Gestaltung ist das Grundprinzip von XML: In XML werden die Inhalte in ihrer Dokumentstruktur abgebildet, die Layoutinformation wird kanalspezifisch zugespielt. Die Verlagsinhalte sind damit medienneutral gespeichert, die Grundvoraussetzung für das digital publishing.
7. **Semantisches Markup.** Eine für Verlage besonders wichtige Eigenschaft von XML ist die Möglichkeit, die Dokumente nicht nur formal (Überschriften, Absätze), sondern semantisch (also nach dem Bedeutungsinhalt) auszuzeichnen. Damit kann ein Kochbuchverlag seine Rezepte nach Zutaten, Zubereitung und Kalorienangabe strukturieren, ein juristischer Fachverlag nach Gesetztestext, Paragraph, Kommentierung etc. Auf dieser Grundlage entstehen neue digitale Geschäftsmodelle, die sich „nur“ aus formatierten Daten nicht ableiten lassen würden. Ein Musterbeispiel für semantisches Markup sind die Metadaten eines Titels (Autor, Titel, ISBN).
8. **Barrierefreiheit.** XML ermöglicht das Anlegen und die Pflege von barrierefreien Daten - Ihre EPUBs oder PDFs, die aus diesem Datensatz entstehen, können dadurch zu 100% barrierefrei sein.
9. **Optimale Datennutzung.** In Kombination mit einem CMS (Content Management System) sorgt XML dafür, dass Ihr gesamter Verlagscontent jederzeit beliebig durchsuch- und kombinierbar ist. Das ist die Grundvoraussetzung für eine effiziente Wiederverwertung von Inhalten.
10. **Qualitätssicherung.** Fester Bestandteil des Konzepts von XML ist eine hinterlegte Grammatik. Sie regelt, für den Verlag, wie die Daten aufgebaut sein sollen, welche Strukturen zulässig und welche Metadaten erforderlich sind. So werden verlagsweit – oder sogar darüber hinaus branchenweit – einheitliche, strukturell 100% fehlerfreie Daten sichergestellt; die Voraussetzung für alle digitalen Geschäftsmodelle.

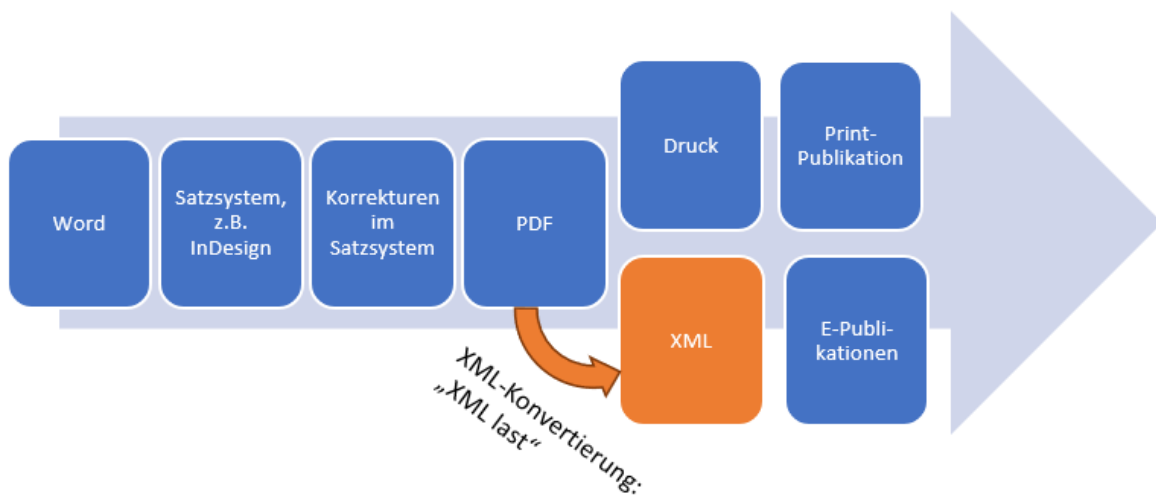
XML first vs. XML last

Sobald die Entscheidung für XML als führendes Datenformat im Verlag getroffen wurde, geht es an die Umsetzung. Eine zentrale Frage dabei ist, an welcher Stelle im Publikationsprozess die XML-Daten entstehen. Hier unterscheidet die Verlagsbranche zwischen sog. „XML first-“ und „XML last-Workflows“.

Bei einem XML-last Workflow werden die XML-Daten erst nach Abschluss der Satzarbeiten aus dem finalen Umbruch erzeugt. Der wesentliche Vorteil dieser Methode ist, dass sie sich leicht in den Verlagsalltag integrieren lässt, denn an den Prozessen der Printproduktion ändert sich nichts, das Lektorat und Teile der Herstellung sind nicht involviert. Diesem pragmatischen Ansatz stehen jedoch erhebliche Nachteile gegenüber, die wir in der unten stehenden Tabelle zusammenfassen.

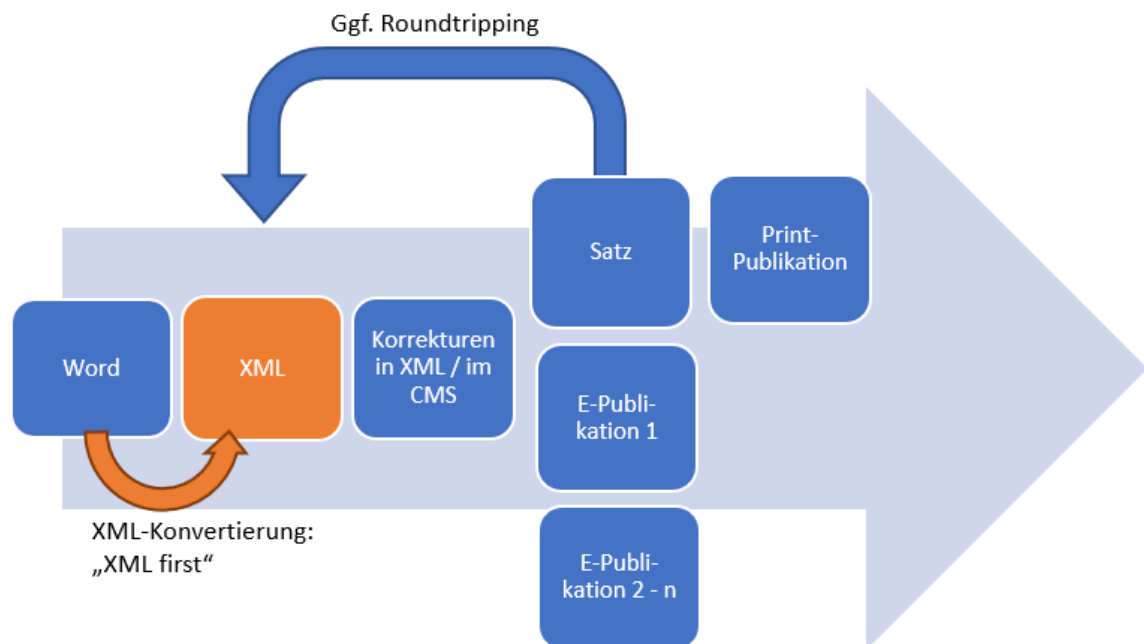
Bei einem XML-first Workflow hingegen werden die Daten vor Beginn der Medienproduktion, idealerweise schon für den redaktionellen Prozess, nach XML überführt (meist: aus Word) oder direkt in XML erstellt (in Redaktionssystemen). Die Einbindung der Lektorate und Redaktionen führt dabei zu deutlich höherwertigen XML-Daten, da nicht nur typographische, sondern auch semantische Auszeichnungen vorgenommen werden können. Neben der höheren Datenqualität sind es v.a. die Einspareffekte im Satz sowie die Möglichkeit, direkt digitale Publikationen ohne den „Umweg“ über eine Print-Produktion zu erstellen („digital first“), die immer mehr Verlage zu dieser Produktionsweise bringen. Dem steht der Nachteil gegenüber, dass häufig ein sog. „Roundtripping“ erforderlich wird: Werden bei einem XML-first-Workflow im Satzprozess noch Korrekturen an den Daten vorgenommen, müssen diese entweder in Doppelpflege auch in den XML-Daten ausgeführt werden, oder das Satzsystem muss in der Lage sein, die (nunmehr auskorrigierten) XML-Daten nach Abschluss der Satzarbeiten wieder zu exportieren. Das aber ist nicht die Stärke der gängigen Layoutprogramme und erfordert den Einsatz von Plugins.

XML last



An erster Stelle steht der Print-Herstellungsprozess. Die inhaltliche Bearbeitung/Korrektur und die medien spezifische Verarbeitung für Print findet in einem gemeinsamen Arbeitsschritt statt. Beides geschieht in der Regel im Satzprogramm beim Satzdienstleister. Erst nach Abschluss der Print-Herstellung werden die Daten in das medienneutrale XML-Format überführt.

XML first



Die (Word-)Daten werden zu einem möglichst frühen Zeitpunkt nach XML konvertiert und qualitätsgesichert. Die inhaltliche Bearbeitung findet idealerweise unabhängig vom Satz in den XML-Daten / in einem CMS statt. Die print-Ausleitung ist eine Medienproduktion von vielen und verliert den Status des „zwingenden“ Prozessschrittes. Wenn im Satz noch Korrekturen anfallen, müssen diese nach Abschluss der Satzarbeiten auch in den XML-Daten enthalten sein. Dieses (automatische) Rückspielen nennt man „Roundtripping“.

Entscheidungsgrundlage

Kriterium	XML first	XML last
Umsetzung v. individuell gestalteten Seiten und Layouts	✓	✓
Automatisierte PDF-Erstellung aus XML (PrintCSS, XSL-FO etc.)	✓	✗
Einfügen v. Nicht-XML-Content	✓	✓
Automatisierung der Satzproduktion / Kostenreduktion im Satz	✓	✗
Optimierung der EBook-Produktion	✓	✓
Barrierefreie Ebook-Produktion	✓	(✓) (erschwert, nachträglicher QS-Prozess zwingend erforderlich)
Einbringen von Semantik in die XML-Daten	✓	(✓) (erschwert, nur durch nachträgliche Einbindung des Lektorats sinnvoll möglich)
Schnellere Erzeugung v. Vorab-Materialien, z.B. für Lektorat o. Vertrieb	✓	✗
Keine XML-Kenntnisse im Verlag erforderlich	✓ ✗ (Grundsätzlich mehr XML-Knowhow erforderlich, aber auch bei XML first ist eine Auslagerung der Arbeiten an Dienstleister möglich)	✓ ✗ (Auslagerung der Arbeiten an Dienstleister ist der „normale“ Weg bei XML last; ohne eigene Kenntnisse ist es aber fast unmöglich, dies zu steuern und zu verantworten)
Einmalige QS der Ausgangsdaten	✓	✗
i.d.R. keine Akzeptanzprobleme im Verlag	✗	✓
Ohne Initialaufwand schnell anwendbar	✗	✓