

# XML-Workflows im Verlag

Ein Positionspapier der Peergroup Produktion  
in der IG Digital

Stand: Juni 2022



# XML-Workflows im Verlag

## Inhalt

1. Verlage im digitalen Zeitalter	3
2. Die Vorteile von XML auf einen Blick	4
3. XML first vs. XML last	6
4. Die Workflowalternativen im Überblick	8
5. Entscheidungsgrundlage: Welcher Prozess eignet sich für meinen Verlag?	13

### Versionierung

Datum	Ergänzungen / Änderungen
Juni 2022	[to come]

## 1. 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 sind die Produktionsprozesse noch nicht an diese veränderten Rahmenbedingungen angepasst. Die Schere zwischen der im Verlag vorhandenen Datenbasis und dem Anspruch an die technische Qualität des Verlagscontents wird dabei immer größer. Nicht zuletzt bringen die ab 2025 verpflichtenden Vorschriften des European Accessibility Act (EAA) Anforderungen an die Barrierefreiheit digitaler Publikationen mit sich, sodass Verlage sich systematisch mit XML-basierten Publikationsprozessen auseinandersetzen müssen.

Betrachten wir den Anspruch an die „technische Qualität des Verlagscontents“ zunächst abstrakt: Ein Verlag muss heute in der Lage sein, flexibel auf seine Inhalte zuzugreifen und diese in beliebigen Medien und in unterschiedlichen Produktausprägungen mit geringem Aufwand zu publizieren. 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 wiederverwertbar vorhalten und mit höchstmöglichem Automatisierungsgrad in die verschiedensten Ausgabeformate ausleiten kann. Mit diesem sogenannten „medienneutralen Vorhalten“ des Verlagscontents werden die Verlagsinhalte nicht nur in einer formatierten Datei (also z.B. in InDesign und PDF) gespeichert, sondern in einer Form gesichert, in der die Inhalte unabhängig vom Ausgabekanal strukturiert und damit automatisiert verarbeitbar sind.

Die Trennung von Struktur und Layout ist die Grundidee von XML – und von allen digitalen Ausgabeformaten, allen voran dem Web. Auch HTML als lingua franca des Internet basiert auf XML-Technologien. Doch nicht nur dem Web, sondern allen gängigen digitalen Ausgabeformaten (E-Books, Apps) liegt intern ein XML-Format zugrunde. Ziel eines zukunftsweisenden Produktionsprozesses ist es nun, diese digitalen Ausgabeformate möglichst automatisiert zu erstellen. Da sich verschiedene XML-Formate technisch sehr gut ineinander überführen („transformieren“) lassen, besteht die Aufgabenstellung darin, für die Content-Strategie des Verlages eine XML-Struktur zu wählen, aus denen sich alle digitalen Zielformate automatisiert ableiten lassen – der „kleinste gemeinsame strukturelle Nenner“ sozusagen.

Neben der technischen Datenqualität ist dafür allerdings ein zweiter Punkt notwendige Voraussetzung: Für eine automatisierte Medienproduktion muss man von bereits

auskorrigierten Inhalten ausgehen. Es geht also nicht nur um die technische, sondern auch um die inhaltliche Datenqualität, die einem automatisierten Produktionsprozess zugrunde liegt. 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.

Viele Wege führen zu auskorrigierten Daten als wichtigste Voraussetzung für die Produktion digitaler Medien. Sie sind so unterschiedlich, wie die Verlagsprogramme und -sparten: Während im Zeitungs- und (Publikums-) Zeitschriftenmarkt längst XML-basierte Redaktionssysteme üblich sind, haben fast alle Buchverlage mit klassischen Herstellprozessen zu tun, die üblicherweise mit einem Word-Manuskript beginnen. Und auch innerhalb der Buchverlage sind die Anforderungen eines Publikumsverlages völlig anders als die eines Schulbuch- oder Ratgeberverlages. Das geht bis hin zu der Frage, ob man bei der Content-Strategie wirklich von echter „Medienneutralität“ sprechen kann, also ob sich die Inhalte also überhaupt in identischer Form für die Ausspielung in verschiedene Medienkanäle eignen.

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

## **2. 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 – technisch gesehen – jederzeit aktuell.
- 2. Kompatibilität.** XML ist ein offenes, nachhaltig lesbares Datenformat, das nicht wie z.B. InDesign an einen Hersteller gekoppelt ist. Damit ist XML die Basis für jede Art der Langzeitarchivierung und der Schnittstellen zwischen den verschiedensten Systemen.
- 3. Crossmedialität.** XML-Daten sind die technische Voraussetzung für die meisten elektronischen Publikationsformate. Hält ein Verlag seinen Content in XML vor, ist die Grundlage für digitale Publikationen oft 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 jeweils an die einzelnen Ausleitungen angepasste, verschiedene Dateien.
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“, d.h. unabhängig vom Ausgabekanal gespeichert. Diese „Medienneutralität“ ist Grundvoraussetzung für das Digital Publishing. Dabei gilt es freilich zu bedenken, dass der Begriff „Medienneutralität“ suggeriert, dass die Daten redaktionell unverändert in allen Kanälen gleich ausgegeben werden können. Das gilt für viele Verlagssparten (v.a. in der Belletristik und im Sachbuch), ist jedoch nicht in jedem Fall gegeben (z.B. häufig Schulbuch/Lehrwerke).
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 Gesetzestext, Paragraph, Kommentierung etc. Auf dieser Grundlage entstehen neue digitale Geschäftsmodelle, die sich aus „nur“ 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. Für die barrierefreie Aufbereitung von Publikationen sind typischerweise Zusatz-Informationen wie Alternativtexte für Bilder, semantische Struktur oder Sprach-Merkmale nötig, die für eine reine Print-Produktion nicht nötig sind. Deswegen sind DTP-Applikationen oft nur schlecht oder gar nicht in der Lage, diese Zusatz-Informationen zu verwalten. Über eine medienneutrale XML-Struktur können Barrierefreiheits-Informationen dagegen bereits bei der Erfassung mit aufgenommen und flexibel im Prozess verarbeitet werden. 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.

### 3. XML FIRST VS. XML LAST

Wir wenden uns hier an Verlage, bei denen am Anfang eines Publikationsprojektes ein Autorenmanuskript steht. Etwas vereinfacht bezeichnen wir die Manuskript-Phase im Folgenden als „Arbeit in Word“. Und damit räumen wir auch gleich mit einem der großen Missverständnisse auf: Die Umstellung eines Verlages auf XML bedeutet nicht, sich von den Word-Manuskripten zu verabschieden und den Autor:innen ihrer gewohnte Arbeitsumgebung zu nehmen! Vielmehr geht es darum, den bestmöglichen Zeitpunkt im Publikationsprozess zu finden, um zu XML-Daten zu kommen.

Grundsätzlich werden dabei zwei Verfahren unterschieden:

Der **„XML-last“-Ansatz** beschreibt ein Vorgehen, bei dem die Inhalte erst nach Abschluss der Satzarbeiten im Layoutprogramm nach XML konvertiert werden. Der wesentliche Vorteil dieser Methode ist, dass sie sich leicht in den Verlagsalltag integrieren lässt, denn an den bewährten Prozessen der Printproduktion ändert sich nichts: Das Lektorat und Teile der Herstellung sind nicht – oder nur marginal – involviert. Diesem pragmatischen Ansatz stehen jedoch Nachteile gegenüber, die wir in der untenstehenden Tabelle zusammenfassen.

Als **„XML first“** oder **„Content first“** werden hingegen alle Workflows bezeichnet, bei denen die zu publizierenden Inhalte vor dem Satz (präziser: zu Beginn der Arbeiten an der ersten Publikationsform) zunächst strukturiert (nach Möglichkeit weitestgehend automatisiert), dann nach XML überführt werden und anschließend alle Medienformen – inkl. Print-PDF – aus diesen Daten erstellt werden. Die Einbindung der Lektorate und Redaktionen kann dabei zu deutlich höherwertigen XML-Daten führen, 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. Werden jedoch 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 (sogenanntes Roundtripping). Das aber ist nicht die Stärke der gängigen Layoutprogramme. Es dennoch zu ermöglichen, erfordert gut geschulte Satzverantwortliche, die sich an vereinbarte Standards halten und den Einsatz von Software-Erweiterungen.

Auch eignen sich nicht alle Verlagsprodukte gleichermaßen gut für einen XML-first-Workflow. Als Faustregel kann festgehalten werden: Je mehr individuelles Seitenlayout erforderlich ist bzw. je weniger sich der Umbruch durch Templates automatisieren lässt, umso geringer ist der Nutzen und desto höher der Aufwand für XML first.

### XML first / XML last im Vergleich

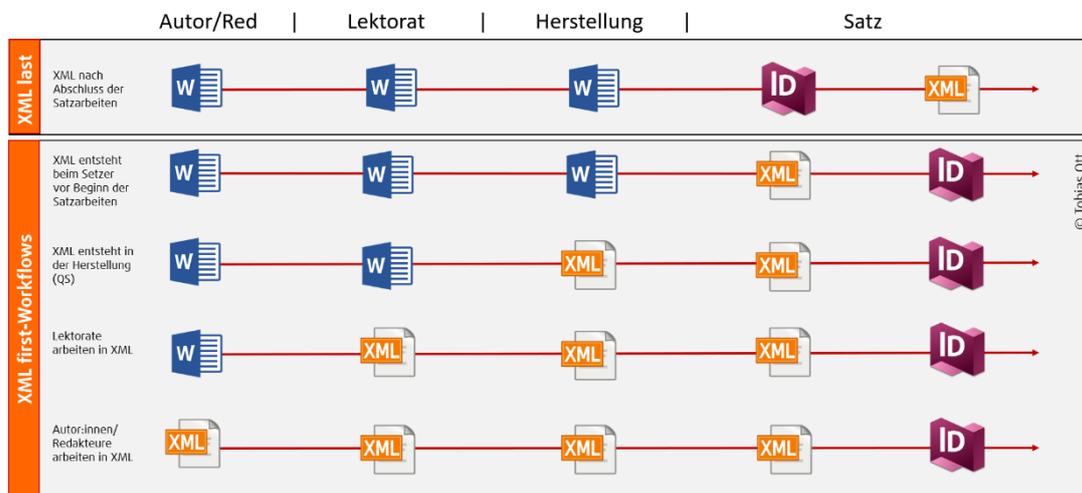


Abb. 1: XML-first-/XML-last-Workflows im Vergleich.

Alle gängigen Satzsysteme können heute mit XML-Daten umgehen – die einen besser, die anderen weniger gut. XML first muss also nicht heißen, dass schon das Lektorat mit XML-Daten arbeitet. Wo im Prozess der Wechsel von Word nach XML stattfindet, sollte individuell bezogen auf die Situation im einzelnen Verlag entschieden werden.

Eine gute Orientierung gibt die nachfolgende Matrix, die eine grobe Landkarte der Verlagssparten zeichnet. Mit ihren zwei Dimensionen bezieht sie sich allerdings nur auf „klassische“ Abgrenzung zwischen XML first und XML last, bei der der Fokus auf der eingesetzten Satz-Technologie liegt.

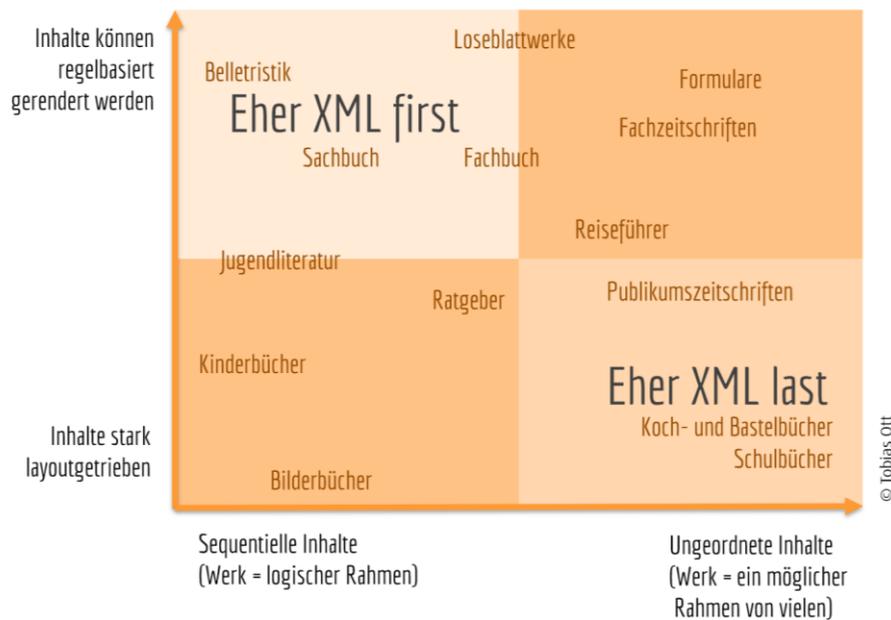


Abb. 2: Orientierungsmatrix XML first/XML last

Für die Bestimmung der bestgeeigneten Arbeitsweise sollte unbedingt noch ein drittes Kriterium beachtet werden: Der Anspruch des Verlages an die inhaltliche Auszeichnung der Inhalte. Für viele digitale Anwendungen macht es einen maßgeblichen Unterschied, ob ein Inhalt lediglich als „Absatz“ oder als „Warnhinweis“, „Rezept“ oder „Tipp“ ausgezeichnet wird. Semantisch reichhaltige XML-Daten sind daher der „Goldstandard“ für Verlagscontent. Dieser Anspruch bzw. die Anforderungen an digitale Produkte führt derzeit z.B. bei vielen Bildungsmedien-Verlagen dazu, dass trotz den damit einhergehenden Erschwernissen im Satz dennoch eine XML first-Strategie verfolgt wird. Für digitale Lernkonzepte sind semantisch ausgezeichnete Verlagsdaten (Codierung der Inhalte als „Aufgabe“, „Lösungsweg“, „Lösung“ etc.) schlicht alternativlos.

In diesen Fällen – wenn kleinteilige Inhalte und layoutlastige Produktion auf einen hohen Anspruch an die Datenauszeichnung treffen – lässt sich dieser scheinbare Widerspruch am besten durch den Einsatz eines XML-basierten Content Management Systems lösen.

#### 4. DIE WORKFLOWALTERNATIVEN IM ÜBERBLICK

So unterschiedlich wie die Verlage (hinsichtlich Größe, Verlagsprogramm etc.) sind, so unterschiedlich können auch tragfähige und zukunftsweisende Produktionsprozesse sein. Wir wollen nachfolgend die wichtigsten Eckpunkte der verschiedenen Ansätze „kein XML“, „XML last“ und „XML first“ zusammenfassen.

## **Kein XML-Workflow**

Für eine ganze Reihe von Verlagen ist eine Arbeit in XML nicht sinnvoll. Das gilt insbesondere dann, wenn keinerlei digitale Zweitverwertungen geplant sind UND eine Automation der Abläufe durch XML nicht möglich oder nicht angestrebt wird. Beispiele dafür sind Bilderbücher, Faksimile-Bände etc. Aber auch viele andere Verlage, bei denen der Einsatz von XML möglich und technologisch sinnvoll wäre, tun sich mit dem Umstieg auf eine XML-basierte Produktion schwer, solange sich Aufwand und Nutzen in keinem überzeugenden Verhältnis bewegen.

Der Verlag arbeitet weiterhin gänzlich traditionell. Inhalte in Form von Textdaten werden oft allenfalls rudimentär und meist rein optisch von Autor, Lektor oder Hersteller im Textverarbeitungsprogramm ausgezeichnet. Anpassung der aktuellen Abläufe sind nicht notwendig.

Die Überführung in zumindest rudimentär strukturierten Inhalten ist dem Satzbetrieb überlassen. Falls dieser nur wenig systematisch arbeitet, unterbleibt die Strukturierung u.U. ganz. Bei Arbeit mit unterschiedlichen Dienstleistern erhält der Verlag in der Regel auch unterschiedliche, zueinander inkompatible Daten zur Archivierung zurück. Wenn die Datenhoheit nicht beim Verlag liegt, sondern nur beim Dienstleister archiviert werden, droht Datenverlust z.B. bei Geschäftsaufgabe des Dienstleisters.

Der Zugriff auf archivierte Daten wird schnell zur Glückssache. Können alte Dateiformate die an konkrete Softwaresysteme gebunden sind (QuarkXPress, Pagemaker, 3B2/APP etc.) nicht mehr geöffnet werden, droht oft schon nach wenigen Jahren ebenfalls Datenverlust bzw. erheblicher Aufwand zur „Datenrettung“.

Die Überführung so entstandener Daten in digitale Formate erfordert immer erhebliche Nacharbeit, um den Anforderungen an valide Ausgabeformate oder Barrierefreiheit zu erfüllen.

Bei klassischen Produktionsprozessen ohne XML oder ohne valide strukturierte Daten hat zwingend die Pflege mehrerer Dateien zuzufolge, sobald ein Medienwechsel bzw. Layoutwechsel erfolgt. Während des Produktionsprozesses entstehen also mehrere Masterdateien, die synchronisiert werden müssen.

## XML-last-Workflows

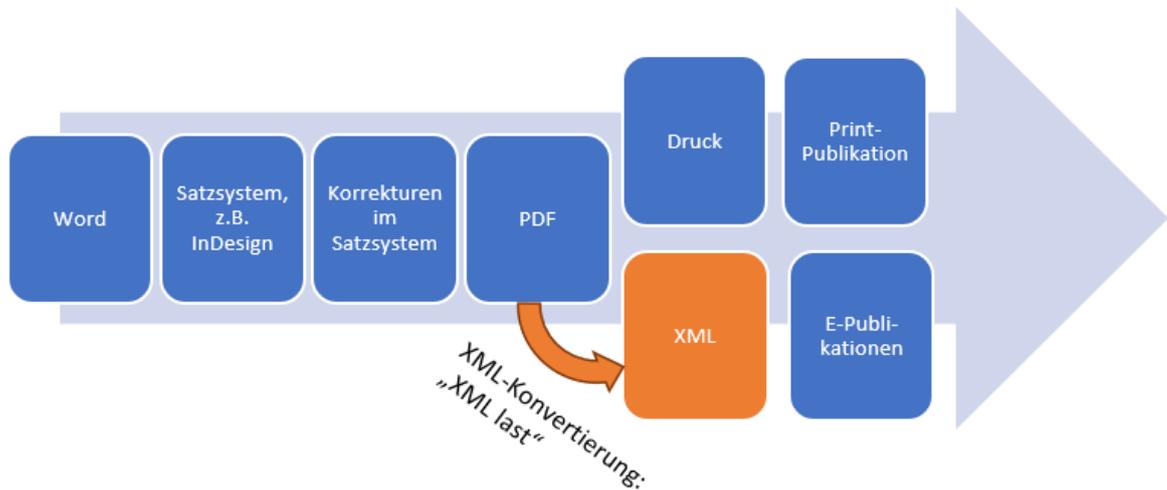


Abb. 3: Produktionsschema XML last

An erster Stelle steht der Print-Herstellungsprozess. Die inhaltliche Bearbeitung/Korrektur und die medienspezifische 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. Häufig wird anschließend auch eine redaktionelle Nacharbeit der XML-Daten notwendig, insbesondere um dem Anspruch an barrierefreie E-Publikationen gerecht zu werden (z.B. das Nachpflegen von Bildbeschreibungen).

Die aus dem Satzsystem exportierten XML-Daten müssen unbedingt separat qualitätsgesichert werden – es besteht immer die Gefahr, dass z.B. durch nicht verknüpfte Textrahmen Teile des Contents in den exportierten XML-Daten fehlen. Erkundigen Sie sich daher bei Ihrer Setzerei, ob sie mit einer XML-last-Produktion vertraut ist. Es gibt Setzereien, die sich auf XML-last-Workflows spezialisiert haben.

### XML last im Überblick

XML-last-Prozesse eignen sich insbesondere für Inhalte, die layoutlastig und kleinteilig sind, bei denen also der Einspareffekt durch Satzautomation geringer ausfällt oder der Einsatz von XML die Satzproduktion gar verteuert.

In der Regel sind die XML-Daten, die in einem XML-last-Prozess entstehen, semantisch nicht so reichhaltig ausgezeichnet. XML-last-Prozesse eignen sich daher insbesondere für Verlage, bei denen der Anspruch an die Strukturtiefe ihrer Daten nicht so hoch ist.



„zwingenden“ Prozessschrittes. Wenn im Satz noch Korrekturen anfallen, muss sichergestellt werden, dass nach Abschluss der Satzarbeiten der letzte Datenstand wieder aus dem Satzsystem (z.B. InDesign) in eine XML-Datei zurückgespielt werden kann. Dieses (automatische) Rückspielen nennt man „Roundtripping“. Alternativ kann in Einzelfällen eine Doppelpflege der Daten (Korrekturen im Satz und parallel in den XML-Quelldaten) die günstigere Variante sein, bei der dann das Roundtripping entfällt.

### **XML first im Überblick**

XML first-Prozesse eignen sich besonders gut für sequenzielle Inhalte, die einem systematischen Aufbau folgen und gut automatisiert hergestellt werden können. In diesem Fall sind Kosteneinsparungen im Satz möglich, die die (zusätzlichen) Kosten für die Konvertierung von Word nach XML kompensieren. XML first-Prozesse erlauben eine semantische Reichhaltigkeit der Daten. Diese kann idealerweise schon während der Lektoratsphase geschaffen werden. Auch alle Strukturen, die für barrierefreies Publizieren erforderlich sind, lassen sich zu diesem Zeitpunkt in die Daten einbringen.

Die Anforderungen an die Barrierefreiheit lassen sich generell mit XML-first-Prozessen leichter erfüllen.

Wenn in einem XML first-Prozess noch Korrekturen während der Satzarbeiten anfallen, muss das „Roundtripping“ sichergestellt werden, damit nach Abschluss der Satzarbeiten auskorrigierte XML-Daten vorliegen.

Soll in einem Verlag eine „E-Book-first/E-Book-only“-Produktion etabliert werden, führt an einem XML first-Prozess kein Weg vorbei. Die Daten können dann nicht erst im Satz auskorrigiert werden, da es ja gar keinen Satz der Daten gibt.

Die Entscheidung für ein Content Management System bringt in aller Regel auch die Entscheidung für einen XML first-Workflow mit sich.

## 5. ENTSCHEIDUNGSGRUNDLAGE: WELCHER PROZESS EIGNET SICH FÜR MEINEN VERLAG?

Welcher Prozess eignet sich für meinen Verlag?		
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 E-Book-Produktion	✓	✓
Barrierefreie E-Book-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	(✓) ✗	✓ (✗)
Einmalige QS der Ausgangsdaten	✓	✗
i.d.R. keine Akzeptanzprobleme im Verlag	✗	✓
Ohne Initialaufwand schnell anwendbar	✗	✓

# **Ihre Branche** **Der Verband**