

# Whitepaper parsX

## Konzept und Module

Stand: Januar 2021

Liebe Leserin, lieber Leser,

seit fast zehn Jahren entwickeln wir parsX – ein Set von Softwaremodulen rund um die zentrale Idee, Publikums- und Wissenschaftsverlagen ein **einheitliches, optimales Datenformat** anzubieten. Wir sind der Überzeugung, dass das »Publizieren 4.0« nur gelingen kann, wenn ein Verlag die technologischen Grundlagen nicht immer wieder neu entwickeln muss, sondern auf den Erfahrungen einer Community aufsetzen kann und mit dazu sein eigenes Wissen wieder in die nächsten Entwicklungsstufen einfließen lässt.

parsX ist das einzige erprobte XML-Format für Publikumsverlage im deutschsprachigen Raum. Er ist ausgereift und zehntausendfach erprobt. Eine immer größer werdende Verlagscommunity setzt parsX als zentrales Datenformat ein und stellt gemeinsam mit uns sicher, dass der Herstellerstandard stets alle Anforderungen abdeckt, die ein moderner Verlag an seine Daten stellen kann.

Es war an der Zeit, aus diesen Erfahrungen einen offenen Herstellerstandard abzuleiten und der Verlagsbranche frei zugänglich zu machen. Wir wollen damit unseren Beitrag leisten, dass die digitale Transformation in der Verlagsbranche gelingen kann.

parsX ist kein vorgeschriebener Prozess zur Buch- und E-Book- Produktion. parsX besteht vielmehr aus einzelnen **Softwarebausteinen für die automatisierte Produktion sämtlicher Medienformate**. In jedem Fall entstehen neben den Publikationen auch **optimal strukturierte XML-Daten**, mit denen Sie bestens für alle Anforderungen der Zukunft gerüstet sind. Und zwar branchenweit im gleichen Format. Den daraus resultierenden Effizienzgewinn spüren wir jetzt schon im Gespräch mit den über 30 Dienstleistern, die parsX-Daten erstellen, verarbeiten und Qualitätssichern.

Wir laden Sie heute ein, Teil dieser Bewegung zu werden. Profitieren Sie von den bestehenden Entwicklungen und lassen Sie Ihre eigenen Ideen und Wünsche mit in unsere Arbeit einfließen! Mit diesem Whitepaper wollen wir Ihnen parsX vorstellen – die Ideen, die uns antreiben und die Technologie, die dahintersteht.

Tobias Ott, Geschäftsführung pagina GmbH

*// Zentraler Ansatz bei  
parsX ist das Schaffen  
eines offenen Daten-  
formats für die Verlags-  
branche sowie ein  
moderner Kooperations-  
gedanke.*



## Die Ideen hinter parsX

### 1. Unsere Mission: Die Etablierung eines offenen Datenformats für Publikums- und Sachbuch-/ Fachbuchverlage

In Industrie und Wissenschaft hat sich längst die Erkenntnis durchgesetzt, dass der Einsatz eines branchenspezifischen Datenstandards der Schlüssel zum Erfolg in der digitalen Zeit ist. Ob Automobilbranche, Chemieindustrie, universitäre Forschung oder Börsenhandel: Erst die Etablierung eines Branchenstandards erlaubt schnellen Datenaustausch, die Entwicklung von Standardsoftware und den Eintritt in das „Industrie 4.0-Zeitalter“. Auch für zahlreiche Verlagssparten gibt es mittlerweile XML-Standards: *TEI* für die inhaltliche Arbeit an Texten, die *NLM-DTD* für naturwissenschaftliche Publikationen, *DITA* für technische Dokumentationen und vieles mehr.

**Nur für die Anforderungen der Publikums- und geisteswissenschaftlichen Fachverlage gab es bislang keine erprobte und bewährte XML-Struktur**, obwohl die Anforderungen täglich steigen. Es gibt keine schnell implementierbaren Softwarelösungen von Drittanbietern, keine kostengünstige Implementierung in CMS- oder MAM-Systeme.

Mit parsX wollen wir das ändern.

parsX ist der Herstellerstandard für Publikums- und Sachbuch-/Fachbuchverlage. Das Schema wurde in enger Abstimmung mit großen, innovativen Verlagshäusern entwickelt und beinhaltet die Architekturen für alle Anforderungen eines crossmedial ausgerichteten Verlags. Der modulare Aufbau stellt sicher, dass Sie dabei nur die Elemente angeboten bekommen, die zu Ihrem Verlagsprogramm passen. In der Regel sind keinerlei Anpassungen erforderlich, um Ihr Verlagsprogramm optimal abzubilden.

Immer mehr Softwareentwickler und Dienstleister im In- und Ausland bieten Lösungen und Dienstleistungen auf der Grundlage von parsX-Daten an. Und immer mehr Anbieter von CMS/MAM-Systemen integrieren die parsX-Module in ihre Lösungen.

### 2. Unsere Überzeugung: Partnerschaft statt Konkurrenz - Systematische Einbindung Ihrer Dienstleister

Wer die Medienbranche verändern will, darf keine Angst vor Mitbewerbern haben. Wir sind der Überzeugung: **Eine Softwarelösung, die einen Verlag an einen Dienstleister bindet, ist unzeitgemäß und strategisch falsch**. Deshalb stellen wir Ihnen und Ihren Dienstleistern das XML-Schema von parsX kostenlos zur Verfügung – mit den vorbereiteten Modulen, um das Schema an Ihre Bedürfnisse anzupassen.

Und auch unser Lizenzierungsmodell für die optionalen, kostenpflichtigen parsX-Softwaremodule zahlt auf diese Überzeugung ein: Wenn Sie als Verlag parsX lizenzieren, erwerben Sie damit automatisch das Recht, die Software an Ihre Vorstufendienstleister kostenlos weiterzugeben.

Ein Qualifizierungsprogramm stellt das XML-Knowhow und den souveränen Umgang mit parsX bei Ihren Dienstleistern sicher. So wollen wir dazu beitragen, dass der Wandel in der Branche tatsächlich gelingt und nicht am „Klein-klein“ im Konkurrenzverhalten zwischen Ihren Satzpartnern behindert wird.

Schon über 30 Dienstleister arbeiten heute nach dem parsX-Standard, ständig kommen neue hinzu. Selbstverständlich steht auch pagina Ihnen als Konvertierungs- und Satzdienstleister mit langjähriger XML-Erfahrung zur Verfügung.

### 3. Unser Ziel: Der Verlag 4.0

Immer wieder erlebt unsere Branche technologische Erneuerungen. Während sich aber alle bisherigen Technologiesprünge stets auf den Beschaffungsmarkt bezogen und damit eingekauft werden konnten, erlebt die Branche derzeit den ersten großen Wandel im Absatzmarkt – mit allen offenen Fragen. Wir von pagina sind uns sicher: Das Verlagsgeschäft der Zukunft kann nicht darin bestehen, statt einem Medium – dem Buch – künftig eben zwei Medienformen – Buch und E-Book – in Auftrag zu geben. Vielmehr sind, ohne das Kerngeschäft aus den Augen zu verlieren, **innovative Vertriebsmodelle**, flexibles Agieren im Autorenmarkt, E-Book-first-Konzepte und vieles mehr gefragt.

Mit parsX wollen wir Antworten geben. Mit Modulen, die nicht nur die optimierte Satz- und E-Book-Produktion betreffen, sondern darüber hinaus eine **automatisierte Leseproben-Produktion** ermöglichen, Multimediale E-Books erstellen, die Seitenwechsel aus InDesign ausspielen (zur Erzeugung von EPUB-PageLists) oder auch beliebige Leseproben zu einem E-Book für den Vertrieb aggregieren. Und selbstverständlich stellen wir auch die Barrierefreiheit der Produktionen sicher. Automatisches Splitten von Romanen in einzelne Serials, ein Qualitätssicherungskonzept, das keine eigenen XML-Kenntnisse erfordert, eigene Copy-Editing-Module in Word für einen **E-Book-first-Workflow**, individuell hinterlegbare E-Book-Profile, vollautomatische Satzvorschau und vieles mehr: Alle parsX-Module werden mit dem Ziel entwickelt, konkrete Lösungen für Aufgabenstellungen in einem modernen, crossmedialen Verlag anzubieten. Und alles zum Einsatz sowohl im Verlag als auch bei Ihren Dienstleistern. Dabei soll die Umstellung eines Verlags auf eine XML-basierte Produktion – inklusive aller Schulungen – innerhalb weniger Monate möglich sein.

Wenn Sie selbst Anregungen für ein neues Modul oder eine Erweiterung haben, sprechen Sie uns an. Wir prüfen und kalkulieren jede Anfrage und versuchen sie so zu konzipieren, dass sie Teil von parsX wird und Ihnen somit nur ein Bruchteil der Entwicklungskosten entsteht.

## parsX – das offene Schema, die einzelnen Module

Um allen Erfordernissen der Verlagskunden hinsichtlich des Verlagsprogramms, den geplanten Prozessen und den geplanten digitalen Produkten gerecht zu werden, haben wir parsX modular aufgebaut. Das XML-Schema mit dazugehöriger Dokumentation ist frei zugänglich und kostenlos. Es besteht keinerlei Verpflichtung, die kostenpflichtigen Module zu lizenzieren.

### Das XML-Schema

In jedem XML-Workflow ist die XML-Grammatik das zentrale Element. Hier wird festgelegt, in welchen Strukturen Ihre Daten künftig vorgehalten werden, was als korrekt und was als fehlerhaft angesehen werden soll. Mit Fördermitteln vom Bund haben wir das Projekt gestartet, aus dem bestehenden parsX-Schema einen de-facto-Standard für Publikums- und Wissenschaftsverlage abzuleiten. Keine andere verfügbare XML-Architektur ist speziell auf die Bedürfnisse von Belletristik- und Sachbuchverlagen ausgelegt, die crossmedial publizieren wollen. In die aktuelle Version sind die Erfahrungen aus über 25.000 produzierten Titeln und viele Wünsche lizenznehmender Verlage eingeflossen. Auch künftige Weiterentwicklungen werden in enger Abstimmung mit der Anwenderschaft vorgenommen. Das parsX-Schema ist modular aufgebaut, so dass Sie immer nur die Elemente verwenden müssen, die für Ihren Verlag relevant sind: So setzen Sie z.B. das Dramenmodul nur dann ein, wenn in Ihrem Verlag auch tatsächlich Bühnenstücke verlegt werden.

### Das Dramensatz-Modul

Mit dem Dramensatz-Modul können Belletristik-Verlage auch komplexe Bühnendialoge, Regieanweisungen etc. auszeichnen und über den E-Book-Konverter im E-Book ausgeben. Die Satzunterstützung ist in allen gängigen XML-fähigen Werksatzsystemen getestet.

Die einzelnen Softwaremodule, die auf dem XML-Standard aufsetzen, werden kontinuierlich von pagina weiterentwickelt und sind als lizenzpflichtige Software erhältlich. Das **Lizenzmodell schließt Ihre Dienstleister automatisch mit ein**, so dass Sie die gesamte Produktionskette genauso weiterlaufen lassen können wie bisher.



### Das oXygen-Framework (Teil der parsX-Essentials)

Für die Bearbeitung Ihrer XML-Daten und als „Schaltzentrale“ für den Aufruf der Module setzen wir auf den bewährten XML-Editor oXygen unseres Technologiepartners SyncroSoft. Das parsX-Framework für oXygen macht aus dem unkonfigurierten Editor eine angepasste parsX-Oberfläche. Mit graphischer *wysiwym* („what you see is what you mean“)-Oberfläche und einfacher Schaltflächen-Steuerung für alle parsX-Module ist der oXygen-Editor Ihr Haupt-Arbeitswerkzeug. Auch Ihre E-Books und Lese-proben werden künftig mit wenigen Mausklicks aus dem parsX-Framework im oXygen produziert.

### Das parsX-Schematron (Teil der parsX-Essentials)

Valide XML-Daten gemäß Schema sind die wichtigste Voraussetzung für jede XML-Verarbeitung. Aber: Validität alleine reicht oft nicht aus. Jeder Verlag hat eigene Anforderungen an seine Daten, die sich nicht abprüfen lassen, sogenannte *Konventionen*. Daher haben wir in parsX einen Schematron-Konventionsprüfer eingebaut, der zahlreiche Plausibilitätsprüfungen standardmäßig durchführt, sich darüber hinaus aber auch leicht an Ihre spezifischen Bedürfnisse anpassen lässt.



### Der parsX-DokuChecker (Teil der parsX-Essentials)

Unseren DokuChecker werden Sie lieben. Er ist neben dem Schema und Schematron der dritte Qualitätssicherungsbaustein von parsX – und er verfolgt einen völlig anderen Ansatz. Der DokuChecker ist für XML das, was ein *preflight-check* für PDF-Daten ist: Eine Prüfung Ihrer Daten, ohne dass der Anwender Kenntnisse von XML haben muss. Abgeprüft und dokumentiert werden hier über einen Online-Dienst neben Plausibilitäten sämtliche strukturellen Besonderheiten der XML-Datei, jedes einzelne Zeichen, aber auch die von Ihnen gewünschte Mikrotypographie (Festabstände). Das Ergebnis ist eine PDF-Datei mit freisprachlichen Fehlermeldungen und Dokumentationen. Und da Ihr Dienstleister auf denselben Online-Dienst zugreifen kann, sieht er die Fehlermeldungen noch bevor Sie eine Eingangsprüfung der Daten machen und kann sie korrigieren.



### Der parsX-E-Book-Konverter (Teil der parsX-Essentials)

Neben dem InDesign-PlugIn ist der E-Book-Konverter das Herzstück von parsX. Mit gut 200 Konfigurationsmöglichkeiten definieren Sie ein E-Book-Profil, das den Ansprüchen Ihres Verlages bis ins Detail entspricht: *Wo steht das Impressum? Wie lautet eine Titelei für eine Leseprobe? Wie viele Hierarchien soll das generierte Inhaltsverzeichnis enthalten? Sollen die Seitenwechsel aus dem gedruckten Buch in eine PageList überführt werden? Wie soll mit den Abbildungen und Tabellen umgegangen werden? Sollen Werbeseiten zugespielt werden?* Dieses Profil kombinieren Sie mit beliebigen, hinterlegbaren E-Book-Layouts im CSS-Format – und erstellen ab sofort auf Knopfdruck komplexeste E-Books (EPUB3), die genau Ihren Vorstellungen entsprechen. Nach modernsten Barrierefreiheitsstandards, so weit es die Daten hergeben.

Eingebaute Qualitätssicherungstools wie der *pagina-GlyphenChecker* (prüft jedes Zeichen daraufhin ab, ob es in dem von Ihnen angegebenen E-Book-Font vorhanden ist) und der *pagina-FontSubsetter* (erstellt zur Laufzeit der E-Book-Produktion einen eigenen Font, der nur die benötigten Zeichen enthält) stellen die maximale Qualität Ihrer E-Books sicher.



### Das parsX-InDesign-PlugIn

Mit dem parsX-InDesign-PlugIn wird Ihr InDesign (und das Ihrer Dienstleister) zu einem Roundtripping-fähigen XML-Satzsystem für parsX-Daten. Eigene XML-Import- und -Export-Mechanismen und zahlreiche Erweiterungen der Funktionalität beheben die Schwächen, die InDesign standardmäßig im Umgang mit XML-Daten hat. Dank der mitgelieferten Muster-Templates, die den Leistungsumfang der

parsX-DTD bereits aufnehmen, produzieren Sie im Satz so schnell wie nie zuvor – ohne dabei auf Teile des Leistungsumfangs von InDesign verzichten zu müssen.



### Der parsX-PDF-Previewer und Leseproben-Konverter

Mit dem PDF-Previewer erzeugen Sie vollautomatisiert Voransichten und echten, vollautomatischen Satz Ihrer XML-Daten über den neuen „PrintCSS“-Standard. Genau wie beim E-Book-Konverter lassen sich beliebig viele Profile („Templates“) anlegen – nutzen Sie den PDF-Previewer zur Generierung spezifischer Ausgabeformen für die Bearbeitung im Lektorat, für eine rasche Umfangsschätzung oder um den Text in verschiedenen Reihenlayouts zu testen.

Mit unseren vorgefertigten PrintCSS-Stylesheets kann jeder, der die CSS-Syntax kennt, perfekte Buchlayouts erstellen – das daraus resultierende Satzbild hält (soweit es sich überhaupt automatisieren lässt) jedem Vergleich stand und wird von immer mehr Verlagen als günstige, vollautomatische Satzalternative eingesetzt – z.B., um Titel auch im Großdruck barrierefrei verfügbar zu haben oder um einen nur als E-Book verfügbaren Titel auch als *Print-On-Demand*-Titel anbieten zu können.



### Die CMS-Schnittstelle / Kommandozeilen-Schnittstelle

Über die CMS-Schnittstelle lassen sich alle parsX-Module auch außerhalb des oXygen-Frameworks aufrufen und steuern. parsX lässt sich damit vollständig in Ihre bestehende oder künftige CMS/MAM-Umgebung integrieren.

Die Anwendung dieser kommandozeilenbasierten Steuerung ist aber selbstverständlich nicht auf Datenbanken beschränkt, sondern erlaubt Ihnen zum Beispiel die EPUB- oder PDF-Erstellung auch im Backend Ihrer Website oder Ihres Webshops.



### Der pagina-Word-Konverter und das Word-Satzmodul

Der Word-Konverter ist kein integriertes Modul von parsX, sondern ein eigenständiges pagina-Produkt, das wir hier in einer auf die parsX-Schema vorkonfigurierten Fassung auch für parsX-Kunden anbieten. Es enthält die pagina-Wordtools, eine Sammlung von ca. 20 Word-Erweiterungen zum systematischen Copy-Editing, zum Bereinigen und Vorbereiten Ihrer Word-Daten als Grundlage für eine automatisierte Weiterverarbeitung. Der anschließende Konvertierungsprozess überführt Ihre Word-Daten nach parsX.

## Weitere Module

### Das Multimedia-Modul

Das Multimedia-Modul ist eine Erweiterung des E-Book-Konverters, um multimediale Inhalte (Video, Audio, Bildergalerien) in EPUB-3-Dateien einbinden zu können. Erstellen Sie ein eigenes Profil für *enriched E-Books* und produzieren Sie diese genauso einfach wie Ihre normalen E-Books.

### Das Generated Content Modul

Mit dem EPUB- und PDF-Konverter können Sie aus getaggtten Registerbegriffen umfangreich konfigurierbare Indices erzeugen. Neben dem Erzeugen von lebenden Registern, ermöglicht das Modul auch das Generieren von Inhaltsverzeichnissen, Glossaren sowie Abbildungs- und Tabellenverzeichnissen.

Sie erhalten außerdem die Möglichkeit zur Pflege von Verweisen mit dynamisch generierten Referenzen, die medianabhängig ausgegeben werden (z.B. „siehe Seite 12“ in der Print Ausgabe und Link „siehe Kapitel ‚Vorwort‘“ für die E-Book-Ausgabe).

### Das Fremdleseprobenmodul

Mit dem Import-Modul können andere Werke in parsX-XML-Struktur dateisystembasiert referenziert und im EPUB als Leseprobe angehängt werden.

### Das Sammelbandmodul

Das Sammelband-Modul stellt ein alternatives Inhaltsmodell für ein parsX-Projekt zur Verfügung: Dazu bauen Sie eine Werkstruktur auf, in der mehrere andere parsX-Werke referenziert und zu einem „Sammelband“ zusammengestellt werden. Besonders komfortabel erledigen Sie diese Arbeit im Zusammenspiel mit dem neuen „CMS-Modul“.

**Alle parsX-Module sind „BPMN-ready“, lassen sich also in moderne Workflowmanagementsysteme einbinden und über Schnittstellen ansprechen.**

### Exkurs: XML-first oder XML-last?

*parsX schreibt Ihnen nicht vor, in welchem Prozess Sie Ihre Daten entstehen lassen. Es enthält alle erforderlichen Bausteine, um einen durchgängigen XML-first-Workflow im Verlag komfortabel zu implementieren – aber das ist keine Bedingung. Je nach Verlagsbereich, vielleicht sogar je nach Mitarbeiter, Terminlage etc. entscheiden Sie, ob Sie ein Werk XML-first oder XML-last produzieren wollen. Ihre E-Book-Qualität wird immer dieselbe sein. Auch wenn Sie derzeit einen XML-last-Prozess mit anderen Partnern etablieren, ist parsX die ideale Ergänzung, um die E-Book- und Leseprobe-Produktion zu standardisieren, die Qualitätssicherungsmechanismen zu nutzen und somit eine datentechnisch einheitliche Backlist aufzubauen. Die meisten parsX-Kunden entscheiden sich, mit einem XML-first-Workflow beim Dienstleister zu beginnen und die Softwaremodule im Verlag erst nach und nach (meist in der Herstellung, seltener im Lektorat) zu nutzen.*

*pagina empfiehlt einen XML-first-Workflow.*

---

## Kontakt

pagina GmbH Publikationstechnologien  
Herrenberger Straße 51 72070 Tübingen

Tobias Ott, Geschäftsleitung

+49 (0) 7071 9876-54

[tobias.ott@pagina.gmbh](mailto:tobias.ott@pagina.gmbh)

Anja Gehring, Produktmanagerin parsX

+49 (0) 7071 9876-40

[anja.gehring@pagina.gmbh](mailto:anja.gehring@pagina.gmbh)

[www.parsx.de](http://www.parsx.de)    [www.pagina.gmbh](http://www.pagina.gmbh)