



Dr. Victor
Wang

E-Books mit ePUB

Von Word zum E-Book mit XML

Vorwort

Nach über zehn Jahren intensiver Beschäftigung mit Konzepten des medienneutralen Publizierens war es für den Autor naheliegend – man könnte fast sagen zwingend –, das Thema E-Books näher zu beleuchten. Anfang 2009 – als die Idee zu diesem Buch Gestalt annahm – war absehbar, dass hier der ePub-Standard eine wichtige Rolle spielen würde. Erstaunlicherweise fand sich nirgends eine Publikation zu diesem E-Book-Standard, so dass er quasi über die Spezifikationen erarbeitet werden musste. Der Autor war dabei überrascht festzustellen, dass es einerseits relativ leicht war, ein gültiges E-Book im ePub-Standard zu erstellen, andererseits die dahinter liegenden Substandards und deren Konzepte dennoch einen gewissen Tiefgang aufwiesen. Viele Grundprobleme des elektronischen Publizierens werden in ePub mit relativ ausgereiften Konzepten gelöst.

Auf der anderen Seite steht die inzwischen fast unüberschaubare Anzahl an E-Book-Lesegeräten, die – gemessen an dem inzwischen gewöhnten Bedienungskomfort von PC-basierten elektronischen Produkten – teilweise noch recht spartanisch anmuten. Häufig fühlte sich der Autor an die Anfangszeit der Browserentwicklung erinnert, als es bei der Entwicklung von HTML-Seiten darum ging, möglichst viele der aktuellen Webbrowser zu testen, da sich ein und dieselbe Seite abhängig von Browserhersteller und -version unterschiedlich darstellte. Ähnliches gilt heute wieder (oder hoffentlich eher noch) für die E-Book-Entwicklung.

Umso wichtiger erscheint es, den ePub-Standard als gemeinsamen Nenner der Entwicklung gut zu kennen. Neben dem praktischen Umsetzen eines ePub-E-Books soll dieses Buch daher eben diesem Dokumentationszweck des ePub-Standards dienen. Der Autor ist überzeugt davon, dass es aktuell keinen E-Book-Standard mit vergleichbar guten Zukunftsaussichten gibt. Dies wird auch durch die Tatsache bestärkt, dass aktuell intensiv an der Weiterentwicklung von ePub (»ePub 3.0«) gearbeitet wird.

Natürlich ist ein Buch über E-Books kaum ohne praktische Umsetzung denkbar und daher liegt der zweite Schwerpunkt des Titels auf der technischen Realisierung eines »echten« ePub-E-Books. Der Autor hat es sich hierbei bei der Wahl des Ausgangsformates Word nicht unbedingt leicht gemacht. Andere XML-Formate wie DocBook oder XHTML wären eine geeignetere Grundlage für eine glatte und einfache Konvertierung gewesen. Da die Datengrundlage jedoch in der Praxis nur selten so aussieht, wurde

das »Allerweltsformat« Word gewählt, das seit der Version 2007 ebenfalls ein XML-Datenformat (Office Open XML) verwendet.

Bei der Darstellung der Umsetzung kam dem Autor sicherlich die intime Kenntnis von Verlags- und Satzworkflows zustatten, da viele Detailprobleme in jedem XML-basierten Herstellungsalltag auftauchen. Das Ziel war daher auch, einen Musterworkflow zur E-Book-Erstellung vorzustellen, der in der Regel bei den so genannten »Schnellanleitungen« meist fehlt. Das Ziel ist hierbei immer, nicht nur den Einzelfall zu betrachten, sondern möglichst automatisierte Wege zu finden, die schnell und rasch einsetzbar sind. Davon wird letztlich aber auch der Einzelfall eines E-Books profitieren.

Der Autor ist – und das mag vielleicht überraschen – ausgesprochen bibliophil und weiß gut gemachte Bücher sehr zu schätzen. Daher hofft und wünscht er, dass der vorliegende Titel auch dazu beiträgt, für E-Books ähnliche Maßstäbe zu finden.

Zu guter Letzt gehört es zu den angenehmen Pflichten eines Autors, den beteiligten Personen für ihre Unterstützung zu danken. Dank gilt zunächst meine beiden Lektoren Frau Schulz und Herr Dralle, die mich nicht zuletzt durch das sehr rasche Aufgreifen meiner Buchidee bestärkt haben. Zahlreiche inhaltliche Anregungen verdanke ich den Gesprächen und Mails mit Herrn Dr. Marko Hedler, Herrn Manuel Montero und Herrn Sebastian Lemke. Die letzteren beiden haben auch Anteil an Korrekturvorschlägen ebenso wie meine Mutter Frau Dr. Klinkhardt. Wie so oft wäre eine Autorschaft nicht möglich gewesen ohne die Unterstützung durch meine Frau Eliza und letztlich auch meinen drei Kindern, die über ein Jahr geduldig die Schreibwut des Vaters ertragen haben und sich nun über den Abschluss des Titels besonders freuen.

Heddesheim, November 2010

Einführung

e.1 Warum ein Buch über E-Books?

Das Thema E-Books beherrscht spätestens seit den letzten zwei Jahren die öffentliche Wahrnehmung. Auf der einen Seite als Hype angesehen, dominiert das Thema nicht nur die öffentliche Meinung, sondern beschäftigt auch immer stärker das Fachpublikum. Zahlreiche Konferenzen und Seminare greifen das Thema E-Books auf und spiegeln den Druck wider, der auf die Verlags- und Medienwelt einwirkt. Die Frankfurter Buchmesse 2010 hat E-Books zum Themenschwerpunkt erhoben und mit zahlreichen Fachvorträgen und Spezialistenforen bedacht.

Woher kommt dieser plötzliche Druck auf die Branche, wie kann es sein, dass ein technisch weder neues noch revolutionäres Konzept den Markt dermaßen durcheinanderwirbelt? Von den vielschichtigen Gründen seien nur einige genannt: die Entwicklung einer neuer Generation von Lesegeräten, die zunehmende Affinität eines breiten Publikums, mobile Geräte einzusetzen, die zunehmende Bereitschaft, online Konsum- und Medienprodukte zu erwerben, und nicht zuletzt das Vorhandensein von ausgereiften, erprobten Technologien und Standards.

Unter den zahlreichen möglichen E-Book-Formaten haben sich im Wesentlichen zwei Standards etabliert: zum einen das von Amazon getriebene Mobipocket/AZW-Format für die Kindle-Geräte, zum anderen das im Mittelpunkt dieses Buches stehende ePUB-Format. Während die Amazon-Formate proprietär sind, stellt ePUB (Akronym für »electronic Publication«) einen offenen Standard dar, der unter Beteiligung der (hauptsächlich amerikanischen) Software- und Medienindustrie seit Ende der 1990er Jahre entwickelt wurde. Die Dominanz von ePUB wurde in den letzten zwei Jahren immer deutlicher, zahlreiche Verlage bieten das Format wenigstens als Variante an, einige propagieren sogar den ausschließlichen E-Book-Vertrieb nach der Devise »ePUB only«. Auch in der Softwareindustrie macht sich diese Tendenz bemerkbar, so bezeichnet beispielsweise Adobe ePUB mittlerweile als die »lingua franca for reflowable eBooks«. Spätestens seitdem auch Apple Anfang 2010 mit der Reader-App iBooks auf dem iPhone bzw. iPad auf den Markt gekommen ist, gibt es in der Branche keine Zweifel mehr, dass an ePUB niemand wirklich mehr vorbeikommt.

Das ePUB-Format basiert wesentlich auf der Technologie der Extensible Markup Language (XML) und daher muss sich ein Buch über ePUB zwangsläufig damit auseinandersetzen. Jedes E-Book im ePUB-Format muss aus Quelldaten erzeugt werden, die in XML vorliegen. Daher ist die erste Frage immer, in welcher Form die Quelldaten verfügbar sind. Für die breite Masse von potenziellen Buchinhalten dürften die Ausgangsdaten in Microsoft Word erstellt oder weiterverarbeitet worden sein. Dies ist, man kann zu Microsoft stehen, wie man möchte, die erste der Ausgangsüberlegungen. Das aktuelle docx-Datenformat, das seit Word 2007 eingeführt wurde, beinhaltet andererseits eine kleine Revolution: War bislang das Dateiformat proprietär und von außen nicht zugänglich, so setzt Microsoft nun auf ein gut dokumentiertes XML-Datenformat (Office Open XML) auf, das für einen XML-Prozess in Richtung E-Book zugrunde gelegt werden kann.

Der vorliegende Titel schlägt also zwischen zwei XML-Datenformaten die Brücke: Als Ausgangsdaten werden Office-Open-XML-Daten angenommen, als Zieldaten werden die XML-Formate von ePUB erzeugt. Technisch gesehen wird die Brücke mit einem weiteren XML-Standard realisiert, zum Einsatz kommen wird weitgehend die eXtensible Stylesheet Language Transformations (XSLT).

Die Kenntnis der Standards und des technischen Werkzeugkastens genügt jedoch in der Praxis nicht, vielmehr müssen diese in einen geeigneten Ablauf eingebettet werden. Benötigt wird eine Produktionskette, an deren Ende ein frisch erzeugtes und nach bestimmten Kriterien geprüftes E-Book steht. Ein solcher Musterablauf wird im Rahmen dieses Titels ebenfalls vorgestellt werden.

Zur E-Book-Produktion gehören schließlich auch wichtige inhaltlich-konzeptionelle Fragen: Welches Konzept steht hinter dem E-Book? Handelt es sich einfach um eine Kopie aus der Printwelt im Kleid eines neuen Mediums oder sollen Inhalte ergänzt oder neu aufbereitet werden (so genannte »enriched« E-Books)? Dieses Thema ist besonders spannend und zukunfts-trächtig, da es bei Weitem noch nicht abgeschlossen ist und viele E-Book-Produzenten erst am Anfang der Entwicklung stehen.

Angesichts der klaren Dominanz von ePUB mutet es merkwürdig an, dass es vor allem im deutschen Sprachraum noch keine fundierte Publikation über diesen Standard gibt. Im Internet sind knappe Kurzanleitungen und vor allem zahllose Blogeinträge zu speziellen Aspekten von ePUB zu finden¹. Seit August 2010 existiert ein englischsprachiger Titel von Elizabeth Castro, der die ePUB-Produktion speziell für iBooks auf dem iPad be-

schreibt². Der nach E-Books händierend suchende Markt scheint bislang keine Zeit für eine vertiefte Beschäftigung mit dem ePUB-Standard gefunden zu haben.

Diese Lücke möchte das vorliegende Buch versuchen zu schließen. Der Autor vertritt hierbei die Ansicht, dass im ePub-Standard zahlreiche wichtige Konzepte verankert wurden, die bislang kaum bekannt, geschweige denn verwendet wurden. Insofern soll der vorliegende Titel helfen, diese vielfach noch unentdeckten Schätze zu heben. Zugleich möchte er auch einen Beitrag leisten, die Qualität von E-Book-Produktionen zu verbessern, die häufig unter dem Zeitdruck und fehlendem Know-how gelitten haben³.

Der primäre Fokus der Darstellung auf dem ePUB-Format bringt auch mit sich, dass a priori kein spezielles Endgerät/Reader bevorzugt wird. Vielmehr ist der Anspruch, dass ein sauberes ePUB nach wie vor geräteunabhängig funktionieren sollte und keine spezielle »Tricks« angewendet werden, die nur auf einer bestimmten Klasse von Readern laufen. Insbesondere soll der im Moment (Sommer 2010) zu beobachtende iPad-Hype nicht übernommen werden, der nach wie vor die iPad/iBooks-Szene beherrscht⁴. Der iBooks-Reader von Apple hat sicherlich in seinem Auftreten neue Maßstäbe gesetzt, er entfernt sich aber teilweise auch deutlich von der offiziellen ePUB-Spezifikation. Die Erfahrung der letzten zwei Jahre hat eher gezeigt, dass es wenig sinnvoll ist, sich auf bestimmte Endgeräte festzulegen, da diese recht rasch vom Markt wieder verschwinden können. Soweit der aktuelle ePUB-Standard hierfür noch keine optimalen Möglichkeiten bietet, beispielsweise die Unterstützung von Videoinhalten oder Animationen betreffend, so sollte auf die in Arbeit befindliche neue Version von ePUB gewartet werden. Erst dann lohnt es sich, diese in einen Produktionsprozess einfließen zu lassen.

Schließlich sei ein Wort noch zur Schreibweise von ePUB gesagt. Merkwürdigerweise besteht hier keine ganz feste Konvention, es kommen quasi alle Varianten in Bezug auf die Groß-/Kleinschreibung vor: »epub«, »EPUB«, »Epub« und »ePUB«. Der Autor hat sich für die offizielle, auch in den Logos des IDPF verwendete Schreibweise mit kleinem »e« und großem »PUB« entschieden⁵.



e.2 Worum es in diesem Buch geht

Der Inhalt des Buches ist in vier Hauptteile gegliedert.

Teil 1: E-Publishing

Der erste Teil legt den Grundstein für das Verständnis des elektronischen Publizierens, zu dem auch E-Books zweifellos gehören. Vor allem geht es dann um eine konkrete Begriffsbestimmung des E-Books und insbesondere der begrifflichen Abgrenzung von Lesegeräten und Formaten (Kapitel 1). Im Anschluss wird dann Hintergrund und Herkunft von ePUB sowie von den wichtigsten aktuellen E-Book-Formaten näher beleuchtet; außerdem werden die Grundlagen des Themas Digital Rights Management vorgestellt (Kapitel 2). Am Ende dieses Teils folgt schließlich ein kurzes Praxiskapitel (Kapitel 3), das auf anschauliche Weise am Beispiel des Formates Mobipocket illustriert, wie ein E-Book mit einem Autorenwerkzeug erzeugt werden kann.

Teil 2: XML-Grundlagen für das E-Publishing

Der zweite Teil dringt dann tiefer in die benötigte XML-Technologie ein. Er vermittelt die technischen Grundlagen, die später für den Praxisteil benötigt werden. Neben einer kurzen Einführung in die eXtensible Markup Language (XML) werden die wichtigen Schlüsseltechnologien knapp erläutert, die ePUB verwendet (Kapitel 4). Da für die Verarbeitung von Word-Daten eine Kenntnis von Word-XML gebraucht wird, folgt dann in Kapitel 5 eine Einführung in die wesentlichen Strukturen des aktuellen docx-Format (Office Open XML/WordProcessing Markup Language). Am Ende des zweiten Teils steht dann die XML-Transformationsprache (XSLT), die im Praxisteil als Werkzeug für die Erzeugung von ePUB verwendet wird (Kapitel 6).

Teil 3: ePUB – ein offener Standard

Hinter ePUB stehen eigentlich drei Substandards, denen der dritte Teil gewidmet ist. Ziel dieses Teils ist es, die Merkmale und Besonderheiten der ePUB-Substandards – das Open Publication Format (OPF, Kapitel 7), die Open Publication Structure (OPS, Kapitel 8) sowie das Open Container Format (OCF, Kapitel 9) – zu dokumentieren. Am wichtigsten ist hierbei sicherlich das Inhalteformat (OPS), für das es gemäß der Spezifikation zwei gültige Sprachvarianten gibt. XHTML ist die in der Regel aus dem Internet bekannteste Sprache für E-Book-Inhalte. Weitgehend unbekannt ist die Alternative, die Digital-Talking-Book-Sprache (DTBook), eine für Buchinhalte recht geeignete Struktur. Beide Varianten werden in ihrer Anwendung gegenübergestellt.

Teil 4: ePUB in der Praxis

Nach den Grundlagen-Teilen folgt im vierten Teil die Praxis. Zunächst werden die Werkzeuge für einen ePUB-Workflow ausgewählt und vorgestellt (Kapitel 10). Das letzte und umfangreichste Kapitel 11 stellt dann einen kompletten Ablauf von Word zu ePUB vor. Der Fokus in der Darstellung besteht darauf, die einzelnen Konvertierschritte verständlich und schrittweise vorzustellen. Am Ende des Ablaufs hält der Benutzer ein selbst erzeugtes E-Book in Händen.

e.2.1 Warum Sie dieses Buch lesen sollten

Es mag verschiedene Gründe geben, warum Sie diesen Titel in Händen halten: Möglicherweise ist Ihnen das Format ePUB bereits begegnet und Sie haben schon länger einen vertiefenden Titel darüber gesucht. In diesem Fall kann ich Sie beglückwünschen, Sie dürften im Zuge der Lektüre die meisten Ihrer Fragen beantwortet bekommen. Wahrscheinlicher ist, dass Sie auf den Titel »E-Book« aufmerksam geworden sind und sich nun fragen, inwieweit Ihnen hierbei neue Erkenntnisse geboten werden. In der Annahme, dass Sie einer der drei folgenden Zielgruppen angehören, möchte ich Ihnen die folgende Reihenfolge der Lektüre empfehlen:

Der Entscheider/die Entscheiderin

Sie wollen sich einen Überblick über das Thema E-Books verschaffen und sollten daher auf jeden Fall die Kapitel 1 bis 3 zur Hand nehmen. Wenn Sie außerdem noch näher wissen wollen, wie ePUB funktioniert, und vor allem, welche E-Book-Konzepte mit ePUB umgesetzt werden können, so bietet sich zum Abschluss die Lektüre der ePUB-Kapitel (Kapitel 7 bis 9) an.

Fachmann/frau mit wenig XML-Kenntnissen

Sie haben zwar nicht tiefer mit XML gearbeitet, wollen sich aber über ePUB schlaumachen. Meinen Glückwunsch, Ihnen wurde das Buch quasi auf den Leib geschneidert und Sie können daher beruhigt den ganzen Inhalt von der ersten bis zur letzten Seite lesen. Die Grundlagen, die speziell für XML nötig sind, finden Sie in Kapitel 4.

Fachmann/frau mit XML-Kenntnissen

Abhängig davon, welche XML-Kenntnisse Sie bereits haben, können Sie nach der Lektüre der E-Publishing-Grundlagen den zweiten Teil verkürzen oder sogar überspringen. Für Sie ist dann vermutlich die Dokumentation von ePUB in Teil 3 und natürlich die Praxisumsetzung in Teil 4 besonders wertvoll.

Anmerkungen

- 1 Exemplarisch: Ainsworth, Harrison: Epub Format Construction Guide, http://www.hxa.name/articles/content/epub-guide_hxa7241_2007.html; Diedrich, Oliver: Der kleine E-Autor. E-Books im ePUB-Format selbst erstellen, in: c't 2009/Hf. 25, S.146–151; Meinike, Thomas: epubMinFlow. Ein minimaler Workflow zur automatisierten Umsetzung von E-Books im EPUB-Format, <http://datenverdrahten.de/epubMinFlow>, 2010.
- 2 EPUB Straight to the Point: Creating ebooks for the Apple iPad and other ereaders, Bercheley (Peachpit Press) 2010
- 3 Kritik beispielsweise: <http://www.e-leseratte.de/2009/04/03/deutschland-viele-fehlerhafte-drm-epubs>.
- 4 Exemplarisch: Castro, Elizabeth: Apple kills fonts in iBooks, strikes blow to standards, Blog 23.6.2010, <http://www.pigsgourdsandwikis.com/2010/06/apple-kills-fonts-in-ibooks-strikes.html>; Coles, Stephen: What the iPad Is Missing (No, It's Not a Camera), Blog 14.4.2010, <http://www.fastcompany.com/1616121/what-the-ipad-is-missing-no-it-s-not-a-camera>.
- 5 Die Diskussion der korrekten Schreibweise dauert auch im Internet noch an, siehe beispielsweise eine Befürwortung der vom Autor favorisierten »ePUB«-Schreibweise (<http://www.epubbooks.com/blog/20080412/epub-exactly-how-do-we-write-that/>) oder auch unter www.mobileread.com zu den verschiedenen Varianten der Schreibweise des »eBook« (<http://www.mobileread.com/forums/showthread.php?t=21786&highlight=PVI+e-ink>).