

日本語表現と求められる標準化

電子書籍ファイルフォーマットの標準化と交換フォーマット

植村 八潮*

標準化の背景

日本語書籍における組版規則は、日本語表現と出版文化形成に大きな役割を果たしている。欧米の書籍と比較すればわかるように、縦組み、ルビなどの日本語特有の組版規則、多数の文字、さらに多様なフォントなど、いくつもの特徴を有している。その結果、日本語の電子書籍の制作においては、手間とコストがかかる傾向にある。

一方で、「電子書籍元年」と呼ばれる熱狂的な電子書籍ブームの到来である。先頃開催された東京国際ブックフェア（東京ビックサイト、7月7～10日）では、過去最高の来場者となり、中でも電子出版関連のコーナーに多くの見学者が押し寄せることとなった。

日本での電子書籍市場は、574億円（インプレスR&D「電子書籍ビジネス調査報告書2010」）となり、出版市場（1兆9356億円）に対して3%程度と十分な市場を形成するに至った。2010年後半には、日本語対応電子書籍端末の販売が予想されており、さらに成長が期待されている。

このような状況で、出版界や印刷業界は、電子書籍コンテンツの制作と流通対応が急務となっている。そのためには電子書籍コンテンツの生産性を向上し、さらに制作した電子書籍を多種多様なプラットフォーム・端末において利用し、提供できる環境作りを行う必要がある。日本語電子書籍ファイルフォーマットの標準化が、従来から求められてきた理由である。

以上のような背景を受けて、総務省、文部科学省、経済産業省による「デジタル・ネットワーク

社会における出版物の利活用の推進に関する懇談会」（以下、三省デジ懇）が設立され、喫緊の課題やいくつかの論点整理を行っている。このほど、報告書を公表し、いくつかの提言を行った^{*1}。

この中でも、とくに電子書籍の中間（交換）フォーマットの統一規格を検討する「電子出版日本語フォーマット統一規格会議」の設置が注目されることとなった。筆者は、三省デジ懇の傘下に設けられた「技術に関するワーキングチーム」の構成員として報告書案の作成に参加し、ファイルフォーマット標準化の重要性を主張した。

報告書の中では、

日本語表現に実績のあるファイルフォーマットである「XMDF」（シャープ）と「ドットブック」（ボイジャー）との協調により、出版物のつくり手からの要望にも対応するべく、我が国における中間（交換）フォーマットの統一規格策定に向けた大きな一歩が踏み出された。

としている。

結果的に電子書籍の中間（交換）フォーマットの統一規格が報告書に盛り込んだのは、これまでIEC（国際電気標準化会議）や国内メーカー団体のJEITA（電子情報技術産業協会）において、ファイルフォーマットの国際標準化が進んでいたからである。

もともと民間活動の中で取り組まれてきたことが、改めて政府の懇談会で取り上げられたことになったのである。とはいえ日本が主導して国際標

* UEMURA, Yashio
東京電機大学出版局
局長

〒101-8457 東京都千代田区神田錦町 2-2
yashio@jim.dendai.ac.jp

※ 1：2010年6月28日報告書公表、http://www.soumu.go.jp/main_content/000072064.pdf

準化が進んでいることについて、政府内や電子書籍関係者に知られていなかったのも事実である。再評価されたことで、実証実験の予算化検討も含め、標準化に対する政府の支援が期待されている。

IEC TC100 / TA10 での国際標準化活動

電子書籍に関する国際標準は、IEC TC100 / TA10 が担当しており、現在まで、以下に示す4つの国際標準が発行されている。IEC 62571を除いて、すべて日本提案である。カッコ内は発行年月。

1) IEC/TS 62229 Ed. 1.0 (2006年7月)

マルチメディア電子出版及び電子書籍の概念モデルを示した。

2) IEC 62448 (2007年4月第1版, 2009年2月第2版)

中間(交換)フォーマットである。annexAとしてソニーのBBeBをベースに標準化し、改訂版でannexBとしてシャープのXMDFをベースに標準化した。

3) IEC 62524 (2009年2月)

配信・閲覧に用いられるリーダーズフォーマット。

4) IEC 62571 (2010年1月)

米国提案によるデジタルオーディオフォーマット。

また、検討中のプロジェクトは次の通りである。

1) PT 62605

電子辞書フォーマットで、国内電子辞書のデファクト標準に近いデジタルアシスト社のLeXMLとIEC 62448 Annex B (XMDFベース)のタグを追加し、拡張・改良したものである。2010年中に国際標準の発行が見込まれている。

2) PT 62665 (Texture map for auditory presentation of printed text contents)

日本でデファクト化しつつある、印刷物の音声プレゼンテーションのための表示方法を定義したもので、ユーザビリティ、アクセシビリティの点から期待されている。

電子書籍ファイルフォーマット概念

電子書籍を製作するには、大きく分けて2つの工程がある。1つはDTPデータをもとに電子書籍フォーマットに加工する方法であり、もう1つは印刷物をスキャニングして画像データやPDF形式とする方法である。前者は、文字中心のコンテンツに多く、後者はDTP導入以前の書籍や、図表の多い学術専門書、さらに現在でも版下によって入稿されているマンガに多い。

書籍は、冊子体という物理的な構造がほぼ共通であっても、開いて見ればわかるようにレイアウトは多様である。章、節、項という体系的な見出しや、本文、図表の関連など、コンテンツの構造をレイアウト表現に転化しているからである。紙面を構成する要素が多ければ多いだけ、構造は複雑になり、統一したフォーマットで表現することは困難になる。その結果、画像データなどでの電子書籍化が図られることになる。

一方、文芸などの文字中心のコンテンツであれば、組版ルールという壁は残されるものの、ある程度、統一したフォーマットにまとめることができる。市場規模の大きい文芸コンテンツの流通促進を考慮すれば、電子書籍の専用フォーマットであることが求められる。

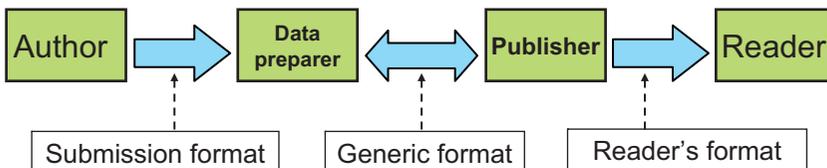


図1 Contents creation/distribution model (コンテンツ生成と流通モデル)

電子書籍コンテンツが、出版社、コンテンツプロバイダを経由して、エンドユーザ（一般消費者）によって閲覧されるまでには、いくつかの段階がある。そこで、IEC/TS 62229 では、この概念モデルとして図1に示すような Contents creation/distribution model を定義している。電子書籍フォーマット関連の標準化においては、これを参照して、どの部分のフォーマットに対応するのか、明らかにすることが行われている。中間（交換）フォーマットは、図1の Data preparer と Publisher の間で用いられるフォーマットで、ここでは Generic format と呼んでいる。図1では、Author が 著 者、Data preparer は 出 版 者、Publisher は コンテンツプロバイダー、Reader は 読 者 お よ び デ バ イ ス（端 末）と 考 え て い た だ き たい。

具体的な例で説明しよう。製作過程では、著者、出版社、製作会社（印刷会社）の間でのデータ交換や、異なるシステム間での変換を保証する必要がある。また本文の文字情報などに加え、ルビや段組、縦中横、脚注といった頁組版情報や画像・

音声といったデジタルならではの表現形式の取り扱いを規定していく必要がある。これらの条件に應えるのは、XMLのような構造化文書となる。

テキストデータ形式であることからデータ量は大きい。印刷会社内での利用や、出版社と印刷会社における閉じたネットワーク間でのやりとりであり、市場流通するものではないので問題とはならない。

一方、電子書籍の読書にはケータイからパソコンまで多様な読書端末装置が使われている。画面サイズ、カラー表示、音声や画像処理、入力インタフェースなどや、処理能力にもかなりの違いがある。そこで流通し、読まれるテキストコンテンツはブログや掲示板などで入手できるテキスト情報やケータイメールなど、必ずしも対価を必要としないものが多い。これに対し電子書籍のコンテンツは、原則的に情報収集に対価を必要としている。このため電子書籍はコンテンツの管理や著作権管理が必要であり、コンテンツ同士も販売競争が常に行われている。また流通上の制約として、データ量が小さい方が好ましい。さらに流

表1 電子書籍ファイルフォーマットの区分

各種フォーマット (オーサリング・ビルダー・ビューワー)	中間(交換)フォーマット 記述フォーマット (テキストデータ)	閲覧フォーマット (テキスト/ バイナリーデータ)	流通フォーマット・ 配信サービス(DRM)
	オーサリングツールにより編集が可能で、フォーマット間の交換・再利用性が高い	低スペックCPU、少ないメモリでも高速に解釈する実行フォーマットもある	配信時のセキュリティ確保、専用ビューワーが必要
HTML	HTML (タグ付きテキスト)	HTML(+CSS)	【DRM不可】
T-Time : ポイジャー	TTX (HTML)	.book	.book(DRM付き)
X MDF : シャープ	X MDF 記述フォーマット(XML)	X MDF	X MDF(DRM付き)
IEC TC100/TA10 国際標準	IEC62448 Generic format(XML)	IEC62524 Reader's format	—
ePub : IDPF	ePub (XHTML)	ePub(XHTML+CSS) →zip圧縮	ePub(ACS4などによるDRM付き)、AZW
Acrobat : アドビ	【交換不可】	PDF	Keyring PDF etc

通適性を考慮すると暗号化や DRM（著作権管理システム）情報を含む必要があり、表示ファイルはバイナリーデータ形式となる。

このように制作過程など中間段階でのファイル形式（Generic format）と、読者へ配信して表示するファイル形式（Reader's format）では、本質的に異なることになる。つまり現実的な標準化として、両者を一つに統一する必要はない。

中間フォーマットの統一の目的

現在、ブームとなっている電子書籍は、文芸などの文字中心コンテンツであり、日本語表現に実績のあるファイルフォーマットとして、前述報告書のように「XMDF」（シャープ）と「ドットブック」（ボイジャー）がある。そこで、IEC 62448 の第 3 版として、ドットブックとも交換可能な中間（交換）フォーマットの策定を目指すこととした。

このような「日本語フォーマット」は日本企業による「ガラパゴス」標準を決めるだけで「世界から孤立するだけだ」という俗耳に入りやすい見方がある。このような誤解が生じている理由の 1 つとして、制作段階に応じて、いくつかの異なるファイル形式が存在していることが理解されていない点がある。

具体的には、表 1 に示したように、HTML のように記述形式（タグ付きテキスト）で書かれた「中間（交換）フォーマット」がある。これを専用端末や携帯電話で閲覧するために実行形式（バイナリーデータ）としてデータ量を小さくした「閲覧フォーマット」。不正な複製を防ぐ目的もあって DRM がかけられた「配信フォーマット」。さらに、電子書籍ファイルを閲覧する「ビューワーソフト」や、「オーサリングツール」と呼ばれる制作するための開発システムも存在する。たとえば「XMDF」と呼ばれるのは、これらのファイル形式や開発システムを総称して呼んでいるもの

である。

今後とも配信フォーマットは各社、各サービスの競争にゆだねられている。もちろん、見やすい「ビューワーソフト」や使いやすい「オーサリングシステム」も同様である。

各社の競争により多様なファイルフォーマットが存在するのはやむを得ない点でもある。米国などは、コンテンツホルダーではなく、IT 企業の主導によって事実上の標準化（デファクトスタンダード）となる傾向にある。一方で、多様なファイルフォーマットに対応することで電子出版制作の非効率性が生じることや、ファイルフォーマットの違いを通じた電子出版端末・プラットフォームでのコンテンツの囲い込みなどは、避けるべきである。

そこで三省デジ懇の報告書では、「様々なプラットフォーム、端末が採用する多様な閲覧ファイルフォーマットに変換対応が容易に可能となる、中間（交換）フォーマットの確立」が求められているとした。このように交換フォーマットを標準化することで配信フォーマットへの変換にも対応しやすくなるだろう。これにより「ワンコンテンツ・マルチファイル」（1 つの作品に対していくつものファイルを作らなくてはならない状況）から「ワンコンテンツ・ワンファイル・マルチプラットフォーム」の実現を目指すものである。 ■

