

JEPA日本電子出版協会

辞書フォーマット国際標準提案について

2009年5月13日

JEITA(電子情報技術産業協会) E-BOOK標準化G

主査(IEC TC100/TA10 マネージャー)

植村八潮

委員(同プロジェクトリーダー)

齋鹿 尚史

オブザーバ(デジタルアシスト)

永田 健児

2009/05/13

©JEITA E-BOOK標準化G 2009

1

1. 電子出版と標準化

2009/05/13

©JEITA E-BOOK標準化G 2009

2

標準化 (Standardization) と標準 (Standards)

- 国際標準 (Global standard) とは、製品の品質、性能、安全性、寸法、試験方法などに関する国際的な取り決めのことです。
- そもそも、国際標準は工業化社会が到来し製品が国境を超える交易の対象となって間もなく登場したもので、経済活動が国内取引で完結せず国際貿易に依存するようになったことの必然的結果です。
- 標準化は「行為」であり、特定の標準が「規格」
- 出典：日本工業標準調査会 (JISC) ウェブサイト

2009/05/13

©JEITA E-BOOK標準化G 2009

3

標準の2種類

- デジュール標準 (de jure standard)
 - 仏語の「法にあった」「正式の」の意
 - 公的標準。公的機関で明文化され公開された手続きによって作成された標準。
 - 例「写真フィルム感度」「電池」
 - ITU/ISO/IEC + ISO/IEC JTC1 (Joint Technical Committee)
- デファクト標準 (de facto standard)
 - 仏語の「事実上の」の意
 - 実質標準。実質的に国際市場で採用しているいわゆる「世界標準」。法的根拠はないが市場での競争力で勝ち抜いた標準。
 - 例「Windows」
 - コンソーシアム標準／フォーラム標準
 - 関心のある企業などが集まって結成された「フォーラム／コンソーシアム」が中心となって作成された標準。
 - W3C, ユニコード, USB, Bluetooth

2009/05/13

©JEITA E-BOOK標準化G 2009

4

標準化の目的

- 従来
 - 互換性・インタフェースの整合性確保
 - 生産効率の向上
 - 製品の適切な品質の設定
- 近年
 - 技術の普及
 - 安全・安心の確保(アクセシビリティ)
 - 環境(省エネ, リサイクル)
 - 貿易促進

2009/05/13

©JEITA E-BOOK標準化G 2009

5

ITU

- International Telecommunication Union
- 国際電気通信連合は国際連合の専門機関
- 無線通信と電気通信分野において各国間の標準化と規制を確立することを目的。
- 1865年5月17日にフランスのパリで設立された万国電信連合に端を発する。
- メンバー: バチカン市国及びほぼ全ての国際連合加盟国の191カ国。セクターメンバーは700社以上(2008年4月時点)。

2009/05/13

©JEITA E-BOOK標準化G 2009

6

ISO

- 国際標準化機構(International Organization for Standardization)
- 各国の代表的標準化機関から成る国際標準化機関で、電気及び電子技術分野を除く全産業分野(鉱工業, 農業, 医薬品等)に関する国際規格の作成を行う
- 沿革 1947年に18ヶ国により発足。
- 目的
 - 国家間の製品やサービスの交換を助けるために、標準化活動の発展を促進
 - 知的, 科学的, 技術的, そして経済的活動における国家間協力を発展させる
- 会員その他(2008.1現在)
 - 会員数:157ヶ国(正会員+準会員) 規格数:17,041規格
 - 規格作成委員会数:専門委員会(TC)201

2009/05/13

©JEITA E-BOOK標準化G 2009

7

IEC

- 国際電気標準会議(International Electrotechnical Commission)
- 各国の代表的標準化機関から成る国際標準化機関であり、電気及び電子技術分野の国際規格の作成を行う
- 沿革 1906年に13ヶ国により発足。
- 目的
 - 電機及び電子の技術分野における標準化のすべての問題及び規格適合性評価のような関連事項に関する国際協力を促進し、これによって国際理解を促進すること。
- 会員その他(2008.5現在)
 - 会員数:70ヶ国(正会員+準会員)
 - 規格数:5,794規格(2007年末現在)
 - 規格作成委員会数:専門委員会(TC)92

2009/05/13

©JEITA E-BOOK標準化G 2009

8

標準化をめぐる国際環境の変化

- WTO/TBT協定 1995年発効
 - WTO/TBT (Technical Barriers to Trade) 第2条4項及び付属書3
 - (貿易の技術的障害に関する協定)
- 各国は強制規格や適合性評価手続きの作成や改正を行う際、原則として国際規格(ISO/IEC)を基礎とすることを義務づける
 - 国内規格(日本ではJIS)は国際規格にあわせなければならない。
 - 電気図記号(電気抵抗)の変更
 - JIS C 0301(TBT協定遵守のため廃止)
 - IEC 60617第2版 → JIS C 0617「電気用図記号」制定
 - JIS X 0208(第1第2水準) → ISO/IEC10646 → JIS X 0221(国際符号化文字集合)



2009/05/13

©JEITA E-BOOK標準化G 2009

9

標準化をめぐる国際環境の変化

- WTO/GP協定
 - WTO/GP (Government Procurement)
 - (政府調達に関する協定)
- TRIPS協定
 - Agreement on Trade-Related Aspects of Intellectual Property Rights
 - WTO協定付属書1Cとして規定
 - (知的所有権の貿易関連の側面に関する協定)
- 中国のWTO加盟 2001年

2009/05/13

©JEITA E-BOOK標準化G 2009

10

政府調達における国際標準の影響

- 非接触ICカードをめぐる対応
- ICカード規格
 - ISO/IEC 14443-2 Type A (Philips) Type B (Motorola)
 - Type C (SONY FeliCa方式) 標準に入れず
- 2001年10月 JR東日本がFeliCa方式の調達計画
- Motorola異議申立「WTO政府調達協定違反」
- 2003年3月 ICカード規格ではなく汎用通信規格として国際標準化 (ISO/IEC18092)

2009/05/13

©JEITA E-BOOK標準化G 2009

11

デジュール標準をめぐる争い

- ODFとOOX (Open XML) の戦い
 - ODF (OpenDocument)
 - オフィスソフトOpenOffice.orgで使われていたファイル形式を拡張し、標準化団体のOASISが標準化。ISO/IEC 26300:2006
 - OOX (Office Open XML)
 - Office 2007から採用 (デフォルト) の文書ファイル形式。2006年12月に国際標準化機関Ecma Internationalによって標準化
 - 2008年4月 ISO/IECで標準として承認。ISO/IEC 29500:2008
 - HTML, PDF, ODFと並んでISOとIECによって認められたオープンな文書フォーマット規格となった (MS発表)
- 無線LAN規格の国際標準化争い
 - IEEE Wi-Fiがデファクト標準化
 - 中国 2003年 WAPI (互換性なし) を策定。輸入販売の禁止

2009/05/13

©JEITA E-BOOK標準化G 2009

12

『世界市場を制覇する国際標準化戦略』

原田節雄 著

『世界市場を制覇する国際標準化戦略』

東京電機大学出版局, 2008



- なぜ, JR東日本は中国に新幹線を建設するのか
- なぜ, 東京電力は中国に超高压送電網を展開するのか
- なぜ, HD-DVDを推す東芝はBDを推すソニー/パナソニックを相手に戦ったのか
- なぜ, ソニー/パナソニックの競争相手が日立製作所や三菱電機でなく東芝だったのか
- なぜ, 非接触型ICカードの開発企業がオムロン, 東芝, 日本信号でなくソニーだったのか
- なぜ, マイクロソフトはOOXMLのISO/IEC国際標準化に固執したのか

2009/05/13

©JEITA E-BOOK標準化G 2009

13

IEC TC100の紹介

- IEC/TC100は, オーディオ, ビデオ, マルチメディアシステムおよび機器の技術分野に関連する国際標準化を行う。民生用分野・業務用分野の機器の性能, 測定方法およびマルチメディアシステムの応用, システムと機器間のインターオペラビリティなどの規格化を推進
- TC100は, 他のTCにおけるSC(Sub committee)に相当する組織であるTA(Technical Area)から成り, 迅速かつ柔軟に対応できる組織運営を行い, 各分野に対して業界共通のインフラ作りに取り組んでいる。

2009/05/13

©JEITA E-BOOK標準化G 2009

14

IEC TC100/TA10の紹介

- Multimedia e-publishing and e-book
- 2006年5月 ヘルシンキAGM会議設立決定
- 2006年9月28日 第1回TA10国際会議がベルリンで開催
- 2007年10月11日 第2回アルザス会議
- 2008年11月19日 第3回サンパウロ会議
- 2009年10月 第4回テルアビブ会議(予定)

2009/05/13

©JEITA E-BOOK標準化G 2009

15

IEC TC100/TA10

- TECHNICAL COMMITTEE TC100: Audio, video and multimedia equipment and systems
- SUBCOMMITTEE TA10: Multimedia e-publishing and e-book
 - TAM: 植村八潮, TS: 向井幹雄(ソニー)
- TA10国内対応委員会
 - JEITA(電子情報技術産業協会) E-BOOK標準化G
- 現在までのプロジェクト
 - IEC/TS 62229 Ed. 1.0
 - IEC 62448 Ed. 2.0
 - IEC 62524(Reader's format)
 - PT 62571(Digital Audiobook File Format)
 - PT 62605(Interchange Format for E-dictionaries)

2009/05/13

©JEITA E-BOOK標準化G 2009

16

IEC/TS 62229 Ed. 1.0

- Multimedia systems and equipment - Multimedia e-publishing and e-book - Conceptual model for multimedia e-publishing
- 電子書籍と電子出版のための概念モデル
- 電子書籍フォーマット関連の標準化において方向性を示した。

2009/05/13

©JEITA E-BOOK標準化G 2009

17

Conceptual model for multimedia e-publishing

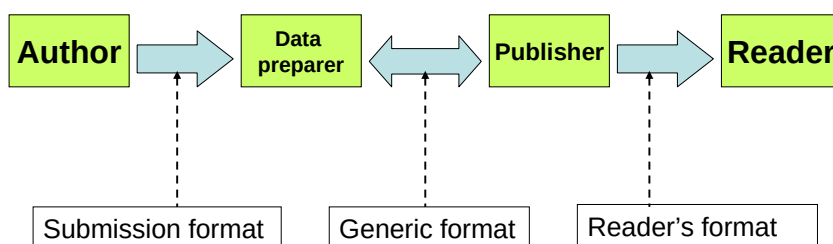


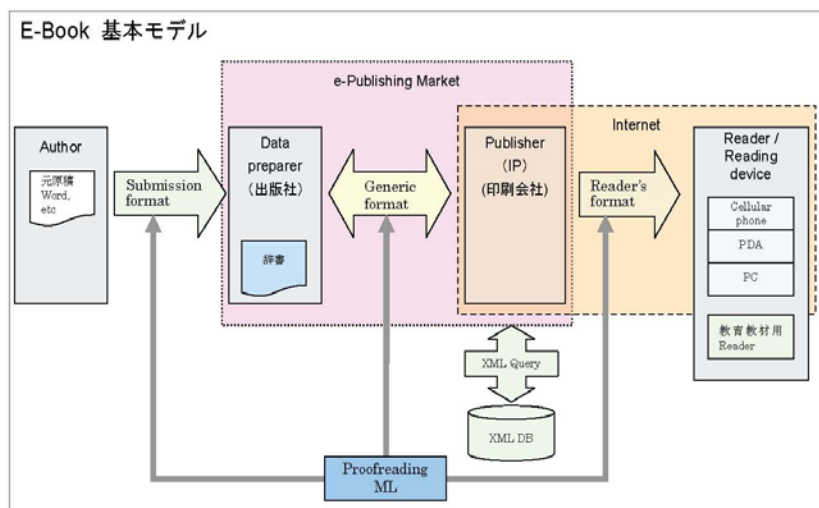
Figure 1 - Contents creation/distribution model

2009/05/13

©JEITA E-BOOK標準化G 2009

18

E-Book 基本モデル



2009/05/13

©JEITA E-BOOK標準化G 2009

19

IEC 62448 (Generic format) 2nd edition

- Multimedia systems and equipment - Multimedia e-publishing and e-books - Generic format for e-publishing
- IEC 62448 (Generic format) 2007年4月IS発行
- 直ちにamendment 1として検討開始
- 2nd edition として2009年2月IS発行

2009/05/13

©JEITA E-BOOK標準化G 2009

20

IEC 62524 (Reader's format)

- 2005年9月 TC100/AGSにて提案
- Reader's formatに求められるもの
 - Formatting Objects(not programming Language)
 - Light-weight (軽い)
 - Non-revisable (変更不可能)
 - Equipment-adaptive (端末適応性)
 - Application-adaptive (アプリケーション適応性)
 - Legibility (可読性)
 - OS-independent
 - Open Standard
- 2009年2月IS発行

2009/05/13

©JEITA E-BOOK標準化G 2009

21

PT 62571 (Digital Audiobook File Format)

- 米国からの提案。
- 2007年9月21日 NP発行
- 2009年4月3日 CDV案をIEC中央事務局へ提出

2009/05/13

©JEITA E-BOOK標準化G 2009

22

PT 62605(電子辞書交換フォーマット)

- Interchange Format for E-dictionaries
- 日本からの提案。
- 2008年5月16日 NP発行
- 2009年3月24日 CD案をIEC中央事務局へ提出

2009/05/13

©JEITA E-BOOK標準化G 2009

23

その他の電子出版標準化

- The International Digital Publishing Forum (IDPF)
- <http://www.idpf.org/>

2009/05/13

©JEITA E-BOOK標準化G 2009

24

2. 電子辞書交換フォーマットの国際標準化

PT 62605 (Multimedia systems and equipment -
Multimedia e-publishing and e-books - Interchange
format for e-dictionaries)

2009/05/13

©JEITA E-BOOK標準化G 2009

25

2.1 経緯と現在の状況

2009/05/13

©JEITA E-BOOK標準化G 2009

26

2.1.1 経緯と現在の状況 (1)

電子辞書が普及・・・各社、コンテンツごとに異なるフォーマットが使われていることが障害になる。

- JEITA E-Book 標準化Gスタート時(2005)から議論あり。
- 2007年からJEITA E-Book 標準化G, IEC TC100/TA10 で議論
- IEC 会合 (2007/10) で、日本側から説明
- 同会合で、NP提案 (新規作業項目提案; 次ページ)を行うことを奨励された。
- NP提案を日本から提出(2008/4) プロジェクトリーダー・・・齋鹿(シャープ)
- JEPA 次世代辞書研究委員会との協力の下、JEITA E-Book 標準化Gで CD (委員会原案;次ページ)をまとめ、提出(2009/3)

日本がイニシアティブを取って標準化を進行

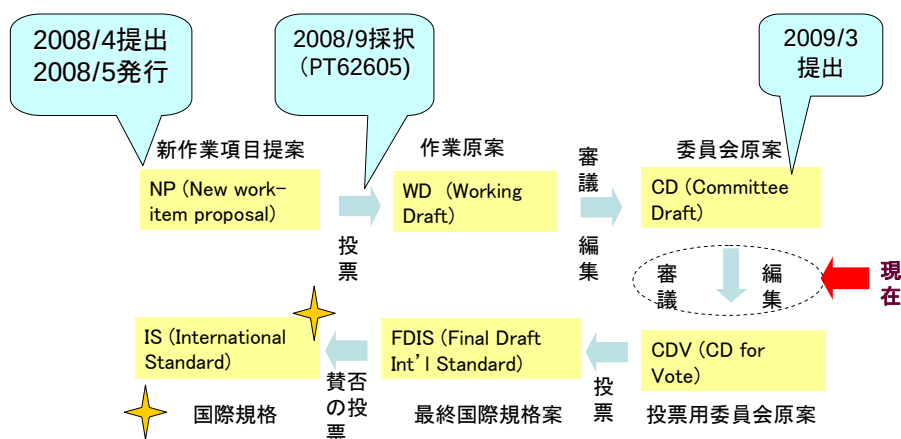
出版社側(JEPA)との連携の下、仕様策定

2009/05/13

©JEITA E-BOOK標準化G 2009

27

2.1.2 経緯と現在の状況 (2)



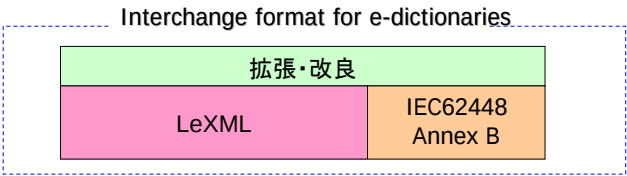
2009/05/13

©JEITA E-BOOK標準化G 2009

28

2.1.3 方針

- ・ LeXML のタグ, 属性名をベースにする。
- ・ 国際標準化になじまない仕様は調整する。
- ・ JEITA E-Book 標準化G + IEC TC100/TA10での審議に基づいて, 機能, 仕様を追加する。
- ・ LeXML にIEC62448 Annex B (XMDFベース)のタグを追加し, 拡張・改良したものとなっている。
- ・ 各国特有の内容(禁則文字など)は, Appdendix にlocalization 項目としてまとめる。



2.2 CD(委員会原案)の内容

2.2.1 Scope (扱う範囲)

出版社, コンテンツ作成者, 電子辞書メーカーの間の交換フォーマットを規定。

以下は扱わない。

- ・ 閲覧装置のためのデータフォーマット
- ・ 印刷にのみ関係する要素
- ・ 物理的デバイスに関連するレンダリングの問題
- ・ DRMなどのセキュリティの問題

※あくまで「交換フォーマット」である。

2009/05/13

©JEITA E-BOOK標準化G 2009

31

2.2.2 Interchange format の位置づけ

Author <--(1)--> Data preparer <--(2)--> Publisher --(3)--> Reader



Interchange format がカバーする範囲

※"though it can be used as a reader's format" という但し書きがあり, (3) として使われることを否定はしていない。

2009/05/13

©JEITA E-BOOK標準化G 2009

32

2.2.3 Interchange format が備えるべき要件 (Requirements)

- 以下を規定

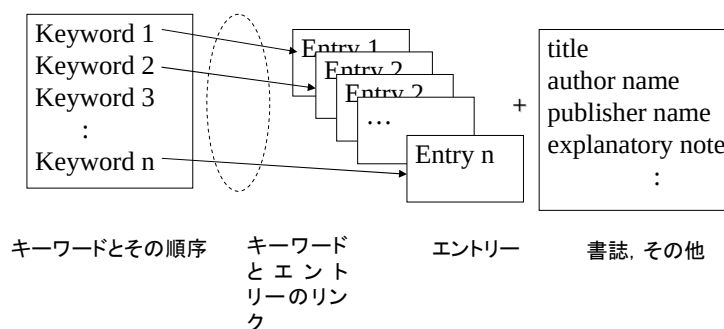
- キーワードの記述, 順序の記述およびキーワードとエントリ(各見出し語の定義)との間のリンク (Keywords and their order, Link data)
- エントリの記述(テキスト, 画像, マルチメディア機能を含む) (Entry Data)
- 書誌(著者名, 題名など), その他の情報(補足など) (Bibliographical data, etc.)
- いろいろな言語で書かれたコンテンツが表現できること。

2009/05/13

©JEITA E-BOOK標準化G 2009

33

2.2.4 Interchange format が備えるべき要件 (Requirements) (2)



ここまでは一般論。以下, フォーマットの詳細が記載される

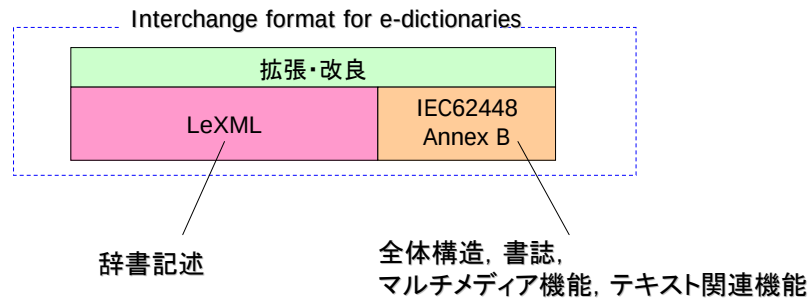
2009/05/13

©JEITA E-BOOK標準化G 2009

34

2.2.5 標準化フォーマットの内容(1)

- XMLで記述
- タグ・属性はLeXMLのものを優先採用し、存在しない機能については追加
- 日本語のタグ名などは変更。



2009/05/13

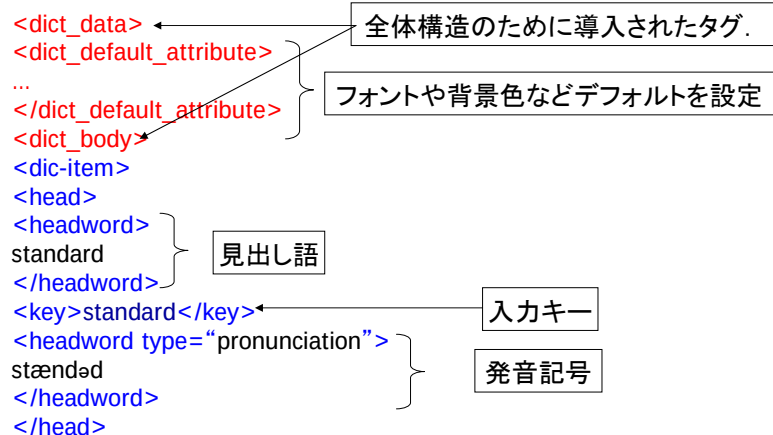
©JEITA E-BOOK標準化G 2009

35

2.2.6 標準化フォーマットの内容

辞書記述例

※あくまでも一例



青文字・・・LeXML仕様から採用 赤文字・・・今回追加・変更

2009/05/13

©JEITA E-BOOK標準化G 2009

36

2.2.6 標準化フォーマットの内容 (cont'd)

辞書記述例(続き)

<meaning level="1" no="1">標準</meaning>

<example>

He set a standard.

彼はお手本となった。

</example>

<meaning level="1" no="2">

知名度が高く、よく歌われる歌曲

</meaning>

</dic-item>

<dic-item>...</dic-item>

<dic-item>...</dic-item>

⋮

</dict_body>

</dict_data>

語義

用例

青文字・・・LeXML仕様から採用

赤文字・・・今回追加・変更

2009/05/13

©JEITA E-BOOK標準化G 2009

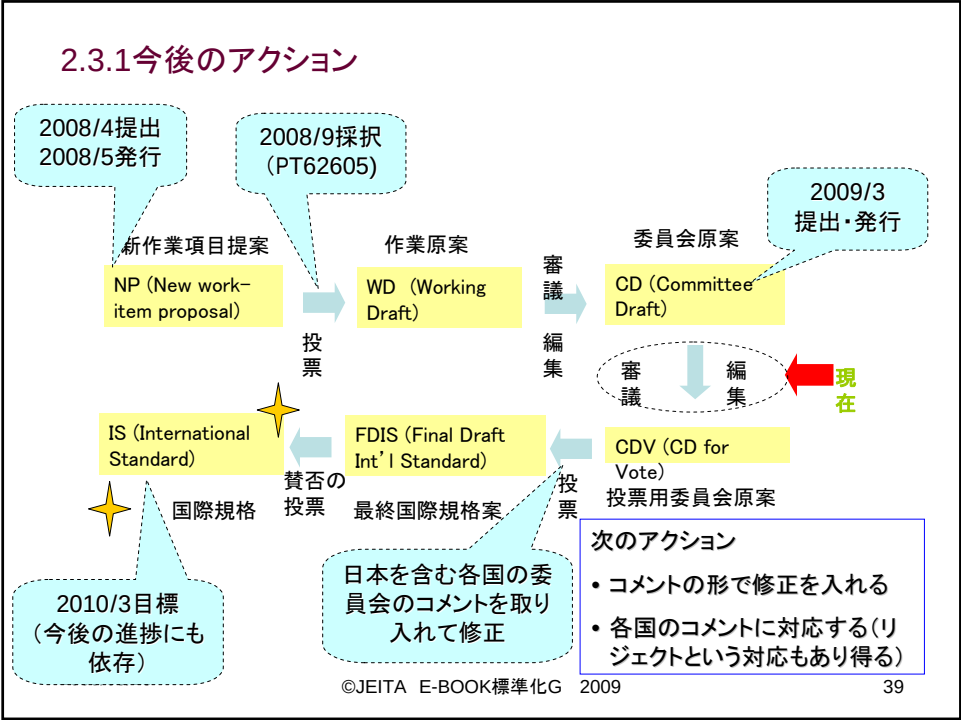
37

2.3 今後の予定

2009/05/13

©JEITA E-BOOK標準化G 2009

38



3. LeXMLについて

LeXMLとは

- **LEX**icographical **eX**tensible **M**arkup **L**anguageの略
- 辞書および事典類の“本体”部分の構造化を目的としたXML仕様
- 2001年からデジタルアシスト社が策定・管理
- 欧文系外字・記号類の記述形式「LeXMLエンティティ」とセットで、辞書コンテンツの統一的なデータ記述を実現
- 外部提供用途だけではなく、編集マスタや派生企画のベースとしても利用可能

2009/05/13

©JEITA E-BOOK標準化G 2009

41

LeXMLの実績

- 各社電子辞書のほか、Web/ケータイでの検索サービス、CD-ROM等、複数のデバイス向けの提供フォーマットとして、いろいろな種類のレファレンス系コンテンツに対応

2009年6月末時点の予定

LeXML化したコンテンツ		電子辞書搭載	検索サービス提供
国語系辞書	43点	15点	20点
英語系辞書	52点	28点	29点
第二外国語辞書	40点	15点	3点
事典・用語集	125点	29点	69点
その他	54点	39点	2点
合計	314点	126点	123点

2009/05/13

©JEITA E-BOOK標準化G 2009

42

今後の各フォーマットとPT62605

- 【既にLeXML化済みのコンテンツ】
必要時に、比較的簡単に、PT62605にコンバートすることが可能
(変換ツールを用意する予定)
- 【他の形式でデジタル化済みのコンテンツ】
個別に変換フィルタを作成する必要がある
- 【デジタル化がこれからのコンテンツ】
シンプルな構造のものであれば、直接PT62605形式でデータ作成
しても良い

2009/05/13

©JEITA E-BOOK標準化G 2009

43

今後のLeXMLとPT62605

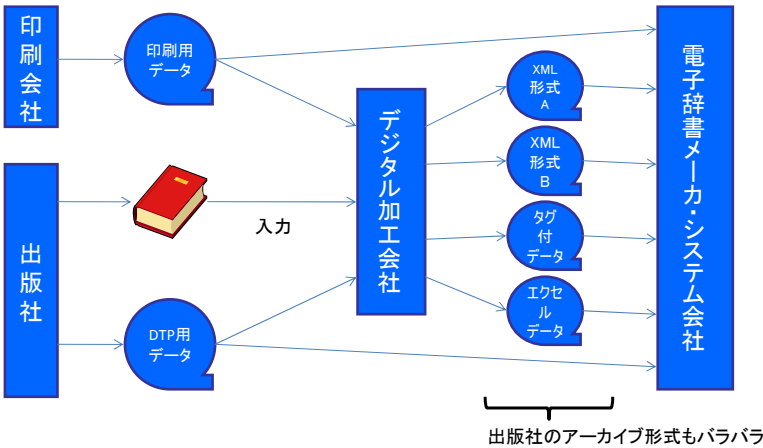
- 「PT62605」は「LeXML」はその基本構造・主要タグにおいてほぼ
同一である。
- 編集マスタとしての用途も想定する「LeXML」は、交換フォーマット
としては特に必要とされない内容・専門的なタグや編集メンテナ
ンス用のタグも保持しつつ、今後も独自にバージョンアップの予定。
※汎用的で利用価値の高いものについては、適宜、PT62605に
組み入れ提案を行っていく(というわけで今後も連動予定)。

2009/05/13

©JEITA E-BOOK標準化G 2009

44

従来の辞書系コンテンツの流れ



統一フォーマットの採用

