

第2回 SPARC Japan セミナー 2008

「学術出版と XML 対応 - 日本の課題」

XML の到達点と学術出版 -XML の現状への雑感

福重 青史（株式会社デジタルコミュニケーションズ）

国立情報学研究所

講演要旨

XML の実用化の歴史を個人的な経験を基に俯瞰してみる。SGML の基本理念の壮大さと実用化の取り組みに比べて、現時点の WEB や XML 等のハイパーメディアの到達点はまだその目標に遠く及ばないものがある。一定の普及期に到達した XML には 2 つの潮流があり、データ系の XML だけが普及し、平文系 XML は HTML レベルで低迷している。このことが理念実現の遅れの要因となっていた。現在はメタ XML の利用が促進されセマンティック Web の展開を急いでいるが、その進化過程で平文 XML の実運用が必須課題となることは明白である。学術出版はまさに平文 XML とメタ XML の実用化によって、合理性、利便性を提供できる時代に入る時となっており、本論考では実用化のためのツールも準備されて来たことを説明したい。

講演者プロフィール



福重 青史

株式会社デジタルコミュニケーションズ
代表取締役

1980 年写植版下制作会社に在籍時、写研電算をワープロ、PC9801 で代用するシステムで運用。入力の不合理性を解決するために DB による情報処理を検討し、汎用マークアップ言語による自動組版システムを開発した。ほぼ同時期に DTP と SGML が日本で紹介され、思想的に同じであったので、DTP と SGML の実用化に着手。1994 年に DTP によるプリプレスのデジタル化をいち早く完了させ、SGML、ハイパーテキスト（HTML）への転換を事業化する。1998 年 SGML の専門企業として（株）デジタルコミュニケーションズを設立し、SGML - XML による自動組版、ハイパービューシステムの開発事業を開始。銀行業務規定システム、自動車メーカー SGML マニュアル、塾テスト問題作成システムなどを開発。2002 年ごろより Word2XML の製品化を開始。官庁、自治体での通達作成システム、条例作成システムに展開。Word2XML による論文査読システムの開発運用も開始。現在は XML の Knowledge 利用を促進するメタデータ XML（RSS）による文書管理システムを Word2XML、査読システムと統合的に開発し、利用促進を目指している。

はじめに

簡単に、私の会社「株式会社 デジタルコミュニケーションズ」を紹介します。(図1)

平成10年、XMLはまだ勧告になっていませんでしたが、SGMLのドキュメントを専門にする会社を設立したいと考え、この会社を立ち上げました。4名でスタートし、今でも従業員は10名程です。XMLだけを専門にしています。

XMLには、データ系のXMLとコンテンツ系のXMLという二つの潮流があります。

データ系のXMLは、ビジネスとしてそれなりに成功もしていますし、一般にもかなり多く使われている状態になっています。

一方、ドキュメント系のXMLは、学術出版も含め、先にスタートしながら、一般にはあまり広く使われていないというのが今日の状況です。そうした現状を、反省も踏まえて紹介します。

私どもの会社は、XML、SGML系の受託、開発を中心に行ってきました。当初は、SGML、XMLのシステム事業に参入している大手のSIなどから仕事を受注することが多かったのですが、ソフトバブルがはじけて直接仕事を頂くことが減少したため、自分たちで商品を開発し、提供していくビジネスに切り替え、このような商品群を現在出しています。(図2)

この中で一番メインの商品はWordからXMLに変換する『Word2XML』です。極端に言えば、この商品だけを10年間やり続けてきたとも言えるかもしれません。また最近、全体のコンテンツ管理をする『文織RSS』というシステムも提供し始めました。

(株)デジタルコミュニケーションズのご紹介

- ・会社名 株式会社デジタルコミュニケーションズ
- ・代表取締役 福重 秀史
- ・設立 平成10年1月22日
- ・資本金 9000万円
- ・従業員 10名 外部スタッフ30名
- ・本社 神奈川県横浜市港内2-2-9
- ・事業本部 東京都港区芝大門1-1-33 三洋ビル2F

1. XMLコンテンツ制作サービス
2. XMLソリューション開発サービス
3. XMLドキュメント管理システム開発サービス
4. e-learningコンテンツ制作/ソリューション開発
5. DTP自動組版システム開発/DTP制作サービス
6. WEBコンテンツ制作/システム開発サービス
7. XMLシステムの企画コンサルティング/運用管理サービス
8. XMLツール企画/開発/販売 ⇒ 事例集

【最近のトピック】

- ・Ayaori-RSS【文織】
- ・Word2FLASH等のWord2XML製品のアプリケーション

Copyright(C) 2008 DIGITAL COMMUNICATIONS Co.,Ltd All Rights Reserved. 2

(図1)

Web2.0コンテンツ・マネージメント
ソリューション

(1)コンテンツXML変換・オーサリングソフト

- ①XML変換ソフト
 - ★Word2XML
 - ★Excel2XML
 - ★PPT2XML
 - ★Quark2XML
- ②XML入力エディタ
 - ★Word2XML
 - ★Excel2XML
 - ★XML-Square(Documentor-TM)
- ③XMLオーサリングツール
 - ★Word2FLASHXML2FLASH(新商品)
 - ★Word2FOXML2FO
 - ★Word2Web
 - ★Word2Indesign(XML2Indesign)
- ④印刷系コンテンツ変換ツール
 - ★Excel2Illustrator
 - ★Excel2Indesign
 - ★Quark2Indesign移行支援SDK
 - ★Indesign画像名表示ツール

(2)Ayaori-RSS【文織】

- ★文書カンバン方式を実現した経理法文書管理システム

(3)業種・業務別コンテンツ連携ソリューション

- ①教育機関向けコンテンツDBシステム
- ②出版印刷業向けコンテンツ作成管理システム(技術論文作成管理システム)
- ③旅行業向けコンテンツ作成管理システム
- ④自治体向け広報作成管理システム
- ⑤金融業向け貸借・約款作成管理システム
- ⑥医薬品・化粧品・食品情報作成管理システム
- ⑦製造業向け製品情報作成管理システム
- ⑧製造業向け品質文書作成管理システム
- ⑨建設業向け業務文書作成管理システム(条例、規約、規則等の改定履歴管理)
- ⑩Webコンテンツ作成管理システム(Net百科事典作成管理システム)

(4)企業向けXMLアドバンス教育

Copyright(C) 2008 DIGITAL COMMUNICATIONS Co.,Ltd All Rights Reserved. 3

(図2)

(A) XML への歩み

会社を始める以前は印刷会社に勤務しており、写研、電算写植のコーディングを行っていました。それをもっと楽にやりたいと思い始めたのが、この道にかかわり始めたきっかけです。(図3)

1980 年前半、日本では PC ブームでしたが、アメリカでは既に DTP ブームが起きていました。

当時、日本では SGML を知らなかったのが、独自の汎用マークアップ言語 (ML) で DB 自動組版を行っていました。そして、80 年後半になって SGML が紹介されたのです。

『BugNews』という DTP の啓蒙紙が出版されたのですが、今読み返しても、今日の時代をまさに予想した極めて魅力的な雑誌だと思います。

今だにこの半分も実現できていないくらいですが、これを読んで真剣に何とかしようと思ったわけです。その中に、SGML の紹介もありました。

SGML を手に入れたいと、発行元に連絡して問い合わせましたが、当然ですがそんなものはないわけで、自分たちで SGML の製品を作り、加工しなければいけないと、これに取り組み始めました。

90 年代前半には、SGML を実用化させ、自動組版や DB へ応用していきました。当時、私がいた印刷会社では、日本のレコードジャケット系の大体 60% の制作印刷を行っており、カラオケ DB も懸命に作りました。その後、政府系の SGML のシステムも作りました。『戦史叢書』という SGML システムは、全 100 巻の太平洋戦争の歴史で、靖国神社に収蔵されているのですが、それをフル SGML で実装した日本初のものでした。

90 年代後半に入ると、HTML が出てきました。その頃、アメリカに SGML の実態調査に行きました。この年は約 70~80 人の人たちがアメリカへ行っったと思います。アメリカで、政府出版系がすでに SGML でバリバリ行われていることを目の当た

りにして、日本でも何とかならないかと思いました。

そうした中で、前述の政府系の SGML 実装や、銀行規定書の SGML 化を始めました。その他、IETM という電子的にインタラクティブを行うマニュアルなど、これらを一生懸命チャレンジに作っていた面白い時代でした。

SGML だけでビジネスができる時代が来るかもしれないと、デジタルコミュニケーションズを設立したのが 98 年です。SGML 等のハイパードキュメント専門をメインにしました。

同時期、インフォテリアも SGML 専門企業として創立されました。インフォテリアは、データ系の SGML 専門企業としてスタートし、我々はドキュメント系です。

会社設立後すぐ、一太郎 8 の SGML バージョンというエクステンションの開発に協力させていただきました。同時に XML の勧告がなされ、SGML から XML に全面転換をすることになります。そういう時代に会社を作ってしまったという状態でした。その後の実績については、当社のホームページをご覧ください。

A-1. プリプレスの進化過程と XML

(図4) は、全体的にまとめて広い視野で見たものです。

プリプレスのシステムは、活版、電算写植、ワープロ DTP、そして Windows DTP へと流れてきて、製版システムも大きく変わりました。

80 年後半から 90 年にかけて、印刷会社のデジタル化は、全業種的にもトップレベルにあったと思います。その後の 10 年間がかなり遅れてしまったという感じがします。SGML、HTML、XML が、ほぼこの 10 年の間に誕生します。

XML は誕生して 10 年が経ち、かなり使われるようになりま

(A) XML への歩み	
1980 年前半: PC ブーム、米国での DTP ブーム、電算写植による組版	⇒ 独自 ML による DB 自動組版
1980 年後半: DTP とフルデジタル制作の開始	⇒ SGML の紹介 (1986 年: BugNews DTP 啓蒙紙)
1990 年前半: SGML の実用化 (独学)	自動組版、DB へ応用
	※ カラオケ DB への利用
	※ 政府系 SGML 支援、実装 (戦史叢書)
1990 年後半: 米国事情調査、HTML 対応	銀行規定書 SGML 化、IETM (マニュアル) SGML
1998 年	(株) デジタルコミュニケーションズ創立
	SGML 等のハイパードキュメント専門の企業
	※ インフォテリア創立
	データ系 SGML 専門企業
	一太郎 8 SGML (開発協力)
	・ XML 勧告
	・ SGML から XML 全面転換
	CSS、XSL 対応のブラウザの出現
	自動組版はどちらでも同じ
	※ 以後は DC の実績年表を参照下さい。
	http://www.sgml-xml.jp/company/chronology.html

(図 3)

A-1. プリプレスの進化過程と XML			
プリプレスシステム	製版システム	ハイパーメディア	概略年代
活版	活版		1950
写植	フィルム手作業 スキャナー		1970
電算写植 (CTS)			
ワープロ	トータルスキャナー CEPS	GML SGML	1980
DTP	イメージセッター CEPS → DTP 連携	HTML	1990
Windows DTP		XML	
パソコンソフト	CTP 高速プリンタ		2000

(図 4)

したが、技術的な進化があったかという、そうでもありません。

A-2. プリプレスとSGML

(図5)は、個人的な体験を整理したものです。

初めに触ったものは、モリサワ、リョービ、写研の手動写植機です。

手動写植機は、プロポーションを持った文字組版はそのまま自動では打てないので、一生懸命手で計算しながら端数を詰めて打っていきます。その後やっと自動写植が出てきて、マイコンが組み込まれました。80年代になって初めて、電算写植が誕生し、この頃は写研の一人勝ちという時代でした。

この頃ワープロも全盛期でしたが、OASYSが100万円、150万円する時代で、数十の製品が乱立していました。

80年代の前半から後半にかけてパソコンが普及しました。しかし、活版もまだ残っていました。私は一部、活版のための手伝いもしていました。そういう時代に、アメリカでは既にDTP革命なるものが起きていて、一方ではSGMLも規格化されているということになかなか気が付きませんでした。紹介されて初めてそういうことに気が付いたわけです。

日本は独自路線で、電子組版、富士通IPS、EZPSなど、様々なものがこの時代にたくさん出ました。当時を経験されている方も多いと思います。

DTPに関する情報が日本にあふれ出し、DTPが紹介されました。当時は、「Japan as No.1」と言っている時代で、バブル崩壊直前です。印刷業は、コンピュータ化では全産業で恐らくトップを走っているだろうという状態でした。

国産のDTPも出てきました。EDIAN、EDICOLOR、様々なものが出来、あわせてポストスクリプトDTPが誕生し、それに移行します。

その中で、SGMLの製品も出てきました。『インターリーフ』というユニックスのDTP製品です。当時、2000万円もするよ

うなDTPですが、それを一生懸命売ろうとしている会社もあって、私も購入しました。DTPを含めてこうしたプリプレスを行っている時代に、なぜSGMLをやったのかというと、当時、私がいた音楽産業の制作事情が背景にあります。

A-3. プリプレスからハイパーメディアへ

当時の音楽産業は、一つの楽曲が出ると、EP、LP、8トラ、カセット、ビデオ(VHSとベータ)、それからVHDとLD、CD、通信カラオケ、MD、CD-ROM、DVDなどが、5年の間にほとんど併存しながら、一気にDVDまで変わってきたという時代です。

EPからCDまでのメディアに対しては、一つの楽曲に対してマルチに全て同じ情報を展開するという時代でしたので、それまでは、同じ作業の繰り返し、同じデータの再入力ということを行わなければならませんでした。これにSGMLを利用できるということがあったのです。(図6)

また、この頃DTPが急成長し、DTPの中にもハイパーカードがありましたので、これを見て、データベースをきちんとやりたいと思いました。

RDBも、80年代前半に出てきたのですが、日本のRDBは2000文字も打つとパンクするというようなものでしたので、独自のランゲージを自分で作りました。しかし、SGMLはそれを半分以上解決してくれるわけで、SGMLをやらなければいけなかったのはこうした事情によるものでした。

その後、パソコンが普及し、LANの時代に入っていきます。

A-2. プリプレスとSGMLを適合へ
福重の体験

(1)福重が自分でも触ったもの

①手動写植……1975ー
モリサワ、リョービ、写研

②自動写植……1977ー
リョービ、写研。一部マイコンを組み込み。

③電算写植……1980ー、写研一人勝ち。

※ワープロ全盛期へ(OASYSが¥100万時代)。数十の製品が乱立。
※パソコン大普及時代に(NEC9800の名機が出現)。
※活版……文庫本はまだ、活版で残っていた。修正作業、タイプライタが残っていた。
※ほのかに米国でのDTP革命なるものの噂が聞こえて来た。
★SGMLがISOで規格化 ⇒ 情報が日本へも。一部の人が注目

④電子組版……1987ー。
富士通IPS。
★DTPに関する情報が日本へあふれ出して来た。
★JAPAN AS No1を謳歌。バブル崩壊直前(1991)
★印刷業コンピュータ化で全産業でトップを走る。

⑤国産DTP……1989ー
EDIAN、EDICOLOR

⑥ポストスクリプトDTP……1991
PageMaker、QuarkXPress

⑦インターリーフ(SGMLの大実験)

Copyright (C) 2008 ORGITAL COMMUNICATIONS Co., Ltd All Rights Reserved.

(図5)

A-3. プリプレスからハイパーメディアへ
福重の体験

(1)プリプレス作業の矛盾と解決方法

①同じ作業の繰り返し
②同じデータの再入力
③メディアの多様化1970ー1990
音楽産業の流通でメディアの流通とマルチ化を体験
・EP、LP(アナログレコード)
・8トラ、カセットテープ
・ビデオ(VHSとベータ)
・VHDとLD
・CD
・通信カラオケ
・MD
・CD-ROM
・DVD

④DTP急成長
RDB、ハイパーカードの出現
パソコン時代、LAN

★SGMLがISOで規格化 ⇒ 情報が日本へも。一部の人が注目

⑤インターネットの出現と超ブーム1994ー1996
⑥マルチメディアブーム(超短命の超ブーム)1996ー1997
★インターネットの超ブームと改革/ITブーム到来とバブル
☆[株]デジタルコミュニケーションズの設立(1998)
XML時代の到来を予測: 同時期設立の(株)インフォテリアは昨年、上場。

Copyright (C) 2008 ORGITAL COMMUNICATIONS Co., Ltd All Rights Reserved.

(図6)



(図7)

(図7) はカラオケの早見表ですが、一つのSGML から、早見表の中身一つを取っても、いろいろな出し方 (レイアウト) があります。

当時 20 社くらいあった音楽会社さんのほとんどカラオケの早見表は、私の方の自動処理で行いました。バッチ処理ですから、入力だけ正しく行えば、一番上のデータ部分をSGMLで作成してクライアントさんには新規部分だけをSGMLではない見やすい形で校正してもらい、完成させれば、あとは自動で全て組めます。

当時の工賃で1時間当たりの売り上げが3万円を超えますので、そういう意味ではいい時代だったと思います。

もちろん、クライアントにも多くの利便性を提供しています。例えば、カラオケ早見表は200ページに4000曲くらい入っており、毎回組み直しのため、それを全部チェックしなければいけないということが、一番の悩みでした。しかし、このシステムですと、新規に作った10曲分なら10曲分の前後左右を見てもらえばいいわけですから、校正する手間が非常に少なくて済みます。それ以外の校正はこちらの責任ですから、お客さまにも非常に大きな利便性があります。

A-4. SGML からインターネットへ

その後「SGML からインターネットへ」という時代が来ました。(図8)

SGML は、自分の中では自動組版としてはうまく使ったのですが、ハイパーメディアとしては実はあまり使いこなしていなかったという思いがありました。

SGML の普及が進まなかった理由は、当時のパソコンでは、力不足で全く使いこなせなかったのです。ユニックスは高過ぎますし、SGML のツールがありません。我々のコンピュータスキルも低いということもあります。仕様の複雑さが原因と今でもおっしゃる方がいますが、そうではなく、当時のコンピュータスキルやコンピュータのパワー不足でできなかったのだと思います。

今、XML は、この複雑さをなくしたと言いつつ、逆に複雑なところに戻ってきていると言ってもいいのではないのでしょうか。

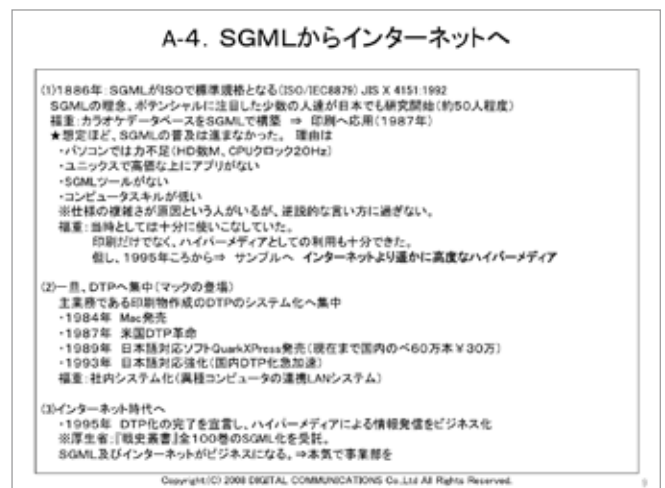
この頃、印刷業はMacのDTPに翻弄されました。実は私もこれを一生懸命やりました。

MacのDTPは、組版に関していうと非常に後退をもたらしたと言って間違いないと思いますが、製版に関していうと非常に効果がありました。

当時、写植やDTPをやって、1ページ当たりの単価が4000~5000円だったとしても、フィルムにA4を1枚出すだけで7000~8000円も費用が出るのであれば、マックDTPをカラーでやらない手はありません。ですから、数年の間にカラーDTP化を一気に進めました。

そうこうしているうちに、インターネット時代の到来です。94年にMosaicを見て非常に驚きました。95年になって日本でも大ブームが起こるなどということは予想もできませんでした。96年には、マイクロソフトがブラウザを無料化し、本当に予想を上回る状態でインターネット時代に入りました。

その頃、時代的にはもう少しさかのぼりますが、『戦史叢書』のSGML化を本格的に受託することになったので、それ



(図8)

業務年表 (DC設立以前: 原点)	
1987	SGML 研究開始 辞書管理システム構築 マークアップ言語システム構築
1989	構造化DTPシステム構築 自動索引システム構築 バーコード作成システム構築
1990	MacDTPシステム化開始 通信カラオケサービス構築
1991	SGML-写真出版システム構築 SGML-I PS出版システム構築
1992	SGML-インターリーフ出版システム構築 SGML-TeX出版システム構築
1993	ISO9000対応研究 医薬品データ管理システム (SGML) 運用 SGML-DTP出版システム構築
1994	ISO9000対応マニュアルシステム構築 SGML情報統合化プラン構築 SGML-ISO9000セミナー開催 エデュテイメントCD-ROMタイトル制作/販売
1995	官庁文庫SGML化 (戦史文庫) DTP画像情報系基本LAN構築 各社DTP化対応コンサルティング 日本語SGMLブラウザの実用化
1996	各社/官庁SGMLパイロットシステム構築多数 各社/官庁ホームページ多数 画像データベースシステム構築 DTPカラーマネージメントシステム実用化/通信系オンデマンド印刷システム構築

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd. All Rights Reserved.

10

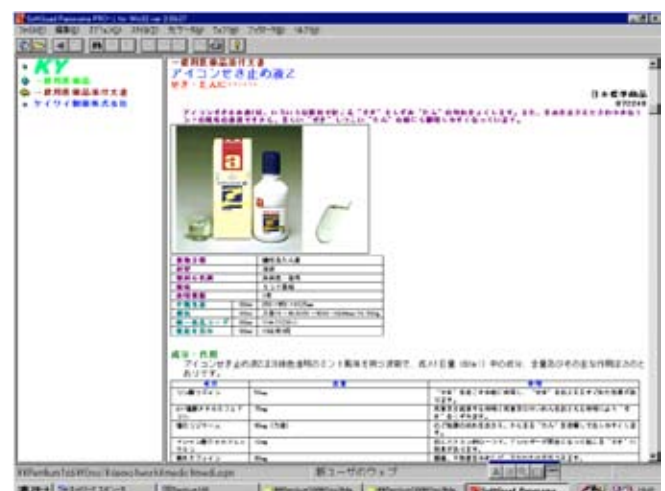
(図 9)



(図 11)



(図 10)



(図 12)

ではということで SGML をやり始めたのです。(図 9) は、関連する今までの実績を一覧にまとめたものです。

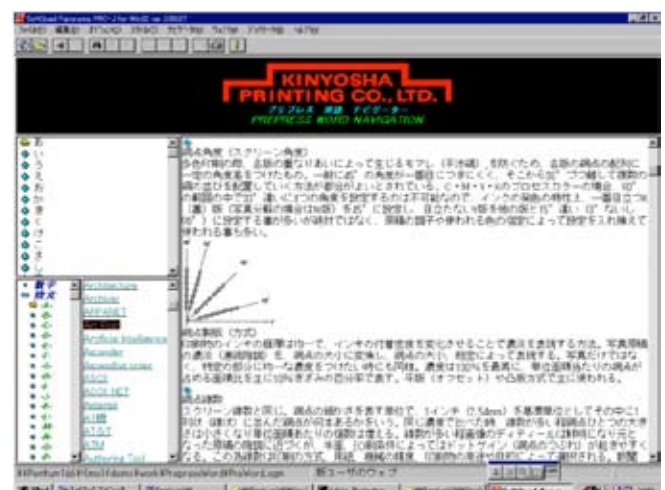
(図 10) はある塾向けの英語の教材で、サンプルなのですが、前の会社でいた 10 年以上前にこれを作っていました。

(図 10) は HTML 版で作ったものですが、(図 11) は SGML でフルに作ったものです。SGML にすることによって、各パートはダイナミックなリンクによってダイナミックに表示することを実現しました。当時、このような形で情報を全て流し込むということが HTML ではできませんでしたので、SGML ブラウザを日本語化して作ったものがこの製品です。

(図 12) も同じようにして作りました。

辞書系の商品も何点が作りました。(図 13)

これはプリプレスの用語ナビゲーターと言いますが、社内でも使っていますし、ある外部の会社でも使っていただいています。



(図 13)

(図14)は実際に受託して、その出版社さんが販売された用語辞書です。これも、フルSGMLで作ったものです。

(図15)は大実験と言えるIETMの一つで、色が変わったり線を引いたりしているものは修正履歴を併用しており、以前の情報と新しい情報をそのまま出して見ることができます。また、新規だけにしたり、旧文書を表示したりを動的に切り替えられます。この頃は企業の方でもチャレンジとして、お金を出してこうした実験作品を作るといったことがありました。ある自動車メーカーの内部の部品管理システムも、このときに作りました。今は自分たちで作っておられます。

A-5. SGML/XML 利用の発展

(図16)は、これまでの印刷業とハイパーメディアの関係を年代ごとに整理したものです。

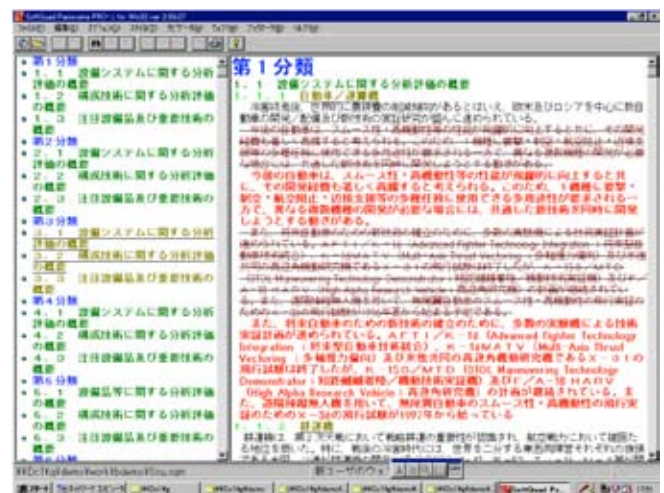
ハイパーテキストの概念は、実は1945年から出ており、いまだにXMLをもってしてハイパーテキストとし、ハイパーメディアを縦横無尽に使いこなす時代にはなっていないというのが、私の正直な感想です。これを本当に使いこなせれば、紙表現を超えた表現が本来できるはずだと思っています。紙表現にあえてこだわらなくても、十分な表現ができさえすれば、コストの問題はどこかで片付くのではないかと思います。10年、この会社をやっています。しかし、多少の芽生えは生まれつつあるかもしれませんが、残念ながらいまだにできておりません。

産業界がどんどん進んだ理由は、88年のCALSコンセプト、94年の日本のNCALSにあります。政府から270億円もの予算が出たのですが、結局大企業がそのお金を使って研究したというだけに終わってしまいました。

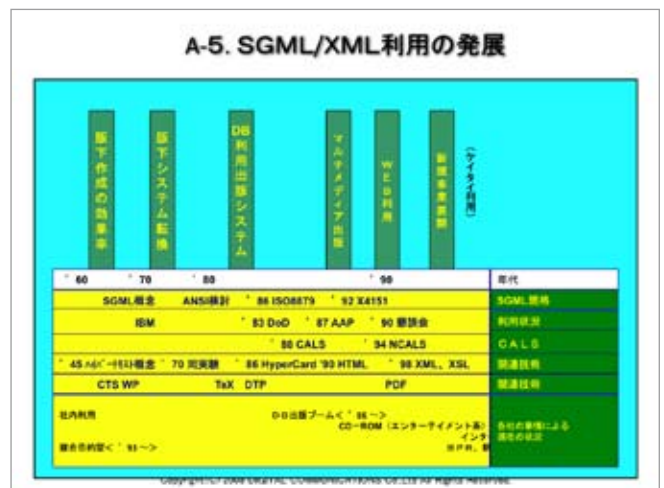
SGML懇談会もできましたが、94年に一度解散しました。その中で、この80~90年にかけての間に、印刷業はDTPとかPDFなどが出て大発展しましたが、95年以降の十数年間の印刷系の発展が少し落ちていていると感じます。



(図14)



(図15)



(B) XMLの到達点

こうしたことをやりながら会社を作ってきた私の主観で、今のXML到達点を整理するとこのようになります。(図17)

XMLには二つの路線があります。vsと書いてあるのは、対立しているという意味ではなく、違う路線があるということです。

平文系のXMLは、我々印刷出版系のXMLです。もう一つの路線はデータ系、つまりCSV系のXMLです。

視点を変えると、メタXMLとインスタンスXMLがあります。XML自体もメタデータといいますが、そのXMLをさらにメタで包むという意味でのメタです。一方のインスタンスXMLとは情報実体を指します。

このように出版、印刷業界系で使うドキュメントのXMLという二つの観点からXMLを見ておく必要があると思います。

また、W3Cでの標準化は、ものすごい勢いで進んでいます。標準化は進まないと言われていたのですが、大企業が実はかなりしっかりやられており、業界団体での標準化も進められています。これには、ヨーロッパでは、排ガスや化学物質、品質保証などのさまざまな規制が行われていますが、それを乗り越えるためにXML化が普通に行われているということが背景にあります。

これに対して、印刷出版系は少しのんびりしていると感じます。

ドキュメントXMLの課題については、まだ答えを出せませんが、我々がやってきたことを含めて説明します。

B-1. ドキュメント(平文)XML vs データ(CSV)XML

まず、ドキュメントXMLとデータ系XMLについてですが、XMLはSGMLから生まれたものなので、そもそもドキュメントの視点を持っています。しかし、XMLが出てきた時点で、Web表示という観点が強調され、我々が望むような意味でのドキュメント系の視点は非常に薄くなっています。(図18)

唯一、組版系の規格があるとすれば、XSL-FO(フォーマットオブジェクト)という規格くらいしかありません。日本から組版規格を出そうという動きもありますが、まずそこが第一の問題です。

HTMLにおいても、Webサイトが進化したことで企業はただホームページを作ればいいという時代から、ECサイトをしっかり作る時代になりました。

ECサイトといっても、コンシューマー向けのECサイトだけではなく、企業間で、部品や様々な情報のやり取りをするためのECサイトで、それが非常によくできています。そのために仲介をする専門の会社があったり、業界団体が会社を作ったりもしています。それでうまくいかなければ別の会社に移行すればいいし、いろいろとチャレンジしながら行われています。そういう観点からも、HTMLというWebサイトも、データ利用の視点が大きいと言えます。

しかし、HTMLも平文ドキュメントに関してはすべて手作りです。

HTMLツールはいろいろありますが、HTMLを直接書かないというだけで、DTPと同じようにデザインしながら作っているという点では、現時点でも全く生産性の上まらないツールを使って、手作業で生成している形です。

本来、XSLT、CSSを使ってきっちりデザインすれば、かなり

(B) XMLの到達点(現状認識)

(XMLの現状評価について独自の見方で整理)

(1)ドキュメント(平文)XML vs データ(CSV)XML

(2)メタXML vs インスタンスXML

(3)標準化と業界適用の状況

(4)ドキュメントXMLの課題 ⇒ 学術出版の課題

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd All Rights Reserved.

(図17)

B-1. ドキュメント(平文)XML vs データ(CSV)XML

- ① XMLはそもそもSGMLから派生したものであるが、Web表示視点が強調されドキュメント(平文)の視点が薄くなっている。
- ② HTMLからの要求としてもデータ利用の視点が強調
- ③ HTMLもXMLもドキュメント(平文)に関しては
⇒現時点でも手作業による生成、個別表示にとどまり、費用と再利用性に疑問を持たれたまま!
- ④これに対し、データ(CSV)のXML化の対応は
元データが既にStructured Data(構造化データ)として存在していることが特徴。
※SGML、XMLは米国ではSemi Structured Dataと呼ばれている。
その生い立ちから生成には問題がなく、費用も発生しない。
変換、流通、再利用が特徴的である。
(発生してもシステム予算や製造予算に組み入れられる)
アプリケーション、DB間連携、統合する Enterprise application integration (EAI)
やBIとしてツールビジネスが成功
全体的にはXMLと言えば、データXMLの方が普及した。

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd All Rights Reserved.

(図18)

のレベルまで行くはずですが。とはいっても、印刷出版系の立場から言えば「気に入らない」ことがどこまでもつきまといますが、費用をかけ、再利用性に問題があると言いながら10年たった今でも、それを継続しています。

これに対してデータ系のXML化は、前述のように企業のECサイトを含めて圧倒的に進んでおり、さまざまなシステムの中でも使われています。この背景には、企業元データがStructured Data（構造化データ）として存在していることがあります。

Structured Data（構造化データ）とは、データベースに入っているデータのことです。日本ではSGMLやXMLを作ったデータのことをStructured Dataと呼ぶという間違いがありました。アメリカではSGML、XMLに関しては、**Semi Structured Data**と呼んでいます。

元データのStructured Dataから、XMLないしSGMLを介して、次のStructured Dataに変換していくというプロセス、パイプラインとも言われていますが、そういったものとして使われています。データそのものがメタデータであるということの意味合いをフルに利用しているシステムです。

わかりやすく言うと、紙幣や貨幣もメタデータです。

我々は商品と商品の間を交換するための手段として、誰もが共通に使える貨幣を使っていますが、XMLをまさにそのように使っているわけです。

例えば、A社さんの商品情報データベースとB社さんの商品情報データベースをつなぐために、XMLを使っています。これによって、情報が非常にうまくつながるようになります。

すでにデータベースに入っているのだから、費用は発生しませんし、変換するためのツールを導入すればいいだけです。たとえ費用が発生したとしても、システム予算や製造予算の中に組み込まれるので、企業にとっては必須の投資ですから、数百万円、数千万円でしょうが、当然行われます。それが印刷出版系コンテンツのやり方との違いとなっています。

こうしたアプリケーションやDB間の連携を行うため、EAIやBIといったツールビジネスも成功しました。インフォテリアが出したEAIツールは、日本でも400社くらいの導入があるそうです。

そういう意味で、全産業的に言うとデータXMLの方が普及しています。

一方のドキュメント（平文）XMLは、いまだに自動組版に我々自身が満足しないという状況の中で行っている状態です。

B-2. メタXML vs インスタンスXML

次に、メタXMLとインスタンスXMLについてですが、この呼び方は私が勝手に使っているもので、世間一般的な呼び方ではありません。（図19）

最近では、Web2.0や、大手企業の中ではEnterprise2.0が注目されています。

Enterprise2.0は、企業内のナレッジを活用するためにWeb2.0のツールを使って情報が蓄積するように作られたものです。

今、ブログやSNS、Wikkiといったものを企業内でも使えるよう、その導入に躍起になっており、企業内、企業間での戦いも、製品レベルからナレッジレベルの戦いになっています。

これらの特徴は、メタデータやタギング（分類等）です。

データXMLの通有性を高めるために、実体データとともに、それを受け渡すためのエンベロップデータをつけます。つまり、エンベロップ（封筒）に宛名を書いて渡すようにこれに包み込んで中のデータを渡します。私は、この部分のデータをメタXMLと言っていますが、企業では、これをさらに進化させ、Web2.0のツールを使ってメタデータをもっと作り、ナレッジ管理（KM）をもっと行おうとしています。こうしたラッピングデータが本来の意味でのメタXMLの代表です。

ナレッジマネジメントに失敗したというトラウマから、この言葉（KM）を使うと嫌がる会社もあるので、今はWeb2.0で、例えばKnow-Whoというような言葉を使うこともあります。

メタXMLの典型的な事例が、RSSです。

RSSに関してはご存じの方も多いでしょうが、W3Cが今、一番力を入れているところは、このメタデータ規格、RDFの規格、オントロジーを作るためのOWLの規格です。このために、グーグルやアマゾンなどはこのRSSを活用して運用しており、SNSもブログも、実はRSSが通知しています。そういったものがどんどん使われる時代になっています。

これに対して、その情報実体であるインスタンスのXML化はかなり停滞していると言えるでしょう。これは、ドキュメントXMLの停滞とほぼ同期しながら進んでいる状態だと思います。

危ぶまれるのは、ラッピングするメタデータだけが先行すると、内部実体のXML化を飛ばして、メタデータだけで管理すればいいのではないかと思われがちなこと。本来そうあってはならないことですが、一時的にはそうするしか方法がないのかもしれない。「内部情報は検索エンジンやマイニングツールで一次しのぎ的に処理」できるので、それで済んでしま

B-2. メタXML vs インスタンスXML

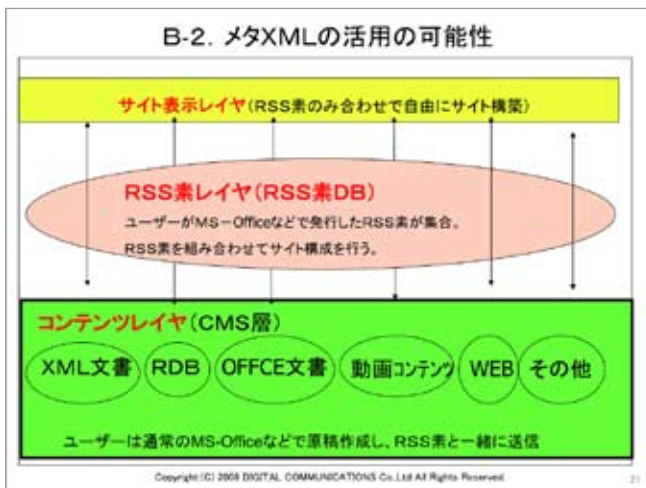
最近の特徴はWeb2.0、Enterprise2.0と言われる、
メタデータやタギング（分類等）に注目
・データXMLの通有性を高めるためのラッピングXMLがメタXMLの代表
・WebではRSSが急速に普及
・グーグルやアマゾンでのCGMなど

これに対し、情報実体であるインスタンスのXML化は停滞。
⇒ドキュメントXMLの停滞と同期
・内部情報は検索エンジンやマイニングツールで一次しのぎ的に処理
※Enterprise Search エンジンを購入する企業多い。
Concept Base3000社とも

XMLは急速に利用されているが、
ドキュメント（平文）対応はアマゾンでも実現できていない。

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd All Rights Reserved. 20

(図19)



(図 20)



(図 21)

うのです。

Enterprise2.0 のもう一つのツールが Enterprise Search と言われているものです。例えば、グーグルがやっているのも基本的な機能はサーチです。グーグルも、今後は Enterprise Search を提供するでしょう。

また、ジャストシステムさんの Concept Base は、企業系 3000 社に導入されているといわれています。そんなものを使わなければ実体の検索ができないのであれば、やはり XML 化するべきではないかと思います。XML 化すればもっと楽に簡単なツールで検索できるはずだと思いますが、実際にはそうはならず、前述したような状態です。

XML なしでも内部情報の検索ができる、マイニングができるということになれば、メタデータの部分だけ XML でやればいいではないかということになりかねません。それでこの数年が終わってしまい、そのうち日本だけが何もしないことになるかもしれないという危険性があると感じています。

アマゾンでは、書誌データだけではなく、中身のデータも提供するように要請しているそうです。すでに提供されている出版社もあるようですが、アマゾンでさえできないということは、中身の XML 化に非常に手間とお金がかかるという現状を表していると思います。

これはやはり、情報を作るプロセスに問題があるのだと思います。それを解決しない限り、メタデータだけで実体を管理することでよとなる危険性があります。

メタ XML の活用に関しては、デジタルコミュニケーションズにはコンテンツシステム管理の製品（文織 RSS）がありますが、この製品はコンテンツ管理とともに、もう一つ、HTML を 1 行も書かないホームページ作成も可能です。（図 20）

コンテンツレイヤに、XML 文書、RDB、OFFICE 文書や動画コンテンツ、そのほか画像などを入れれば、RSS レイヤに RSS ができます。

この RSS に、ブログのように「この商品をこのようにしてこんな人を買ってほしい」というようなことを書けばいいのです。

RSS は XML ですから、適当な表現のデザインの CSS と

XSLT があれば、あとはサイト表示レイヤが勝手にやってくれてホームページができてしまう、というようなことができると思います。

ここで言う中身も XML 化ができないという危険性があるものの、ホームページそのもののお金をかけずに情報を投げ込むだけでできてしまうのです。

手書きブログのように、コンピュータで絵を描く必要はなく、紙に自分で描いた絵をスキャンしてもブログができ上がります。表現したいことは手書きでも何でもかまいません。

出版系は皆そうですが、表現という部分が手作業なのです。そこを自動化しようということがそもそも間違いなのだと思います。何を自動化するかという部分と、コンピュータに何をさせるかということをやうまく整理させたら、もう少し良いことができるような気がします。

さらに、これはまだ実装ができていないわけではありませんが、コンテンツ管理システムの中に情報を入れ、これを共有して仕事をこなしていきまると、ドキュメントに関するメタデータが集まります。それをどういう視点で見るかを決めると、情報をその視点から見ることもできます。メタデータだけで見られるので「文書共有から知識共有へ、あなたの会社のナレッジが見えます」というキャッチコピーをつけていますが、このようなものが作れるようになります。（図 21）

B-3. 標準化と業界適用

次に、標準化と言われている XML の概略です。(図 22)

XML がただの規格だということをなかなか理解されない方がいて、「XML 化すれば自動組版ができるんですよね」と言う人がいまだに多いのですが、XML に関して W3C がやっていることはただの規格制定作業であって、規格をどう使うかはユーザーに任されています。ですから、それをうまく利用すれば世界中の皆さんと一緒に仕事ができるということです。

中でも、主要な規格としてXMLスキーマ、DTD、XSLTなどの重要性をあえて強調したいと思います。

XMLとは何か、何を勉強すればいいかと、よく皆さんに聞かれます。私の会社に入社してくる人も、XMLを全部理解していなかったり、言葉だけ知っているという人が来ます。そういう人には、「1週間時間をあげるのだから一番難しいXSLTの本を1冊読んで理解できたら入社できる」と言っています。

XSLTをきっちり理解すれば、XML とは何なのかという概念理解は別にして、ほとんど実用的に使いこなすことができますので、ぜひここは勉強していただきたいと思います。

また、X Query という言語が、勧告案か、勧告準拠まで来ています。これには注目しなければいけません。

今は XML のネイティブデータベースといわれているものを日本だけが購入している状態です。XML や SGML はそもそもデータベースですから、Query 言語があればいいだけで、アメリカでは XML データベースなるものを購入したりはしません。

しかし、今度、DB2 や Oracle11 など、XML の DB と、RDB を加えたハイブリッドデータベースといわれているものが

出現しました。それに呼応して X Query が勧告になり、DB に X Query が使えるインターフェースを備えて出てくるようになったわけです。IBM は既に本も出しています。こういうものを使いこなすようになれば、とても楽になると思います。

データベースは高いと思っている方がいらっしゃいますが、DB2 などの express バージョンの無償版もあります。また、Oracle も、ワンユーザライセンスなら 1 人 1 万 2000 円程度です。1 万 2000 円で OracleDB が入手できるわけですから、チャレンジブルにやったほうがいいと思います。

X Query が出てきて、2つのデータベースが実装したということは、印刷出版系に対してうかうかしてはいけないという警鐘です。

これが今までの W3C 側の動きですが、これに対し、業界でもさまざまなことを行っています。(図 23)

(図 23) の右側に、特に重要と思われるものを並べています。

WSDL、SOAP などが Web サービスとされています。これらは、前述の EC サイトや情報公開を通して使う道具としては、もう普通に使われています。実は、私どもの会社で先日作ったある塾のフランチャイズ向けの問題作成システムも、すべて WSDL と SOAP を使ってやり取りするシステムになっています。そういう意味で、この辺りはしっかり見ておく必要があると思います。

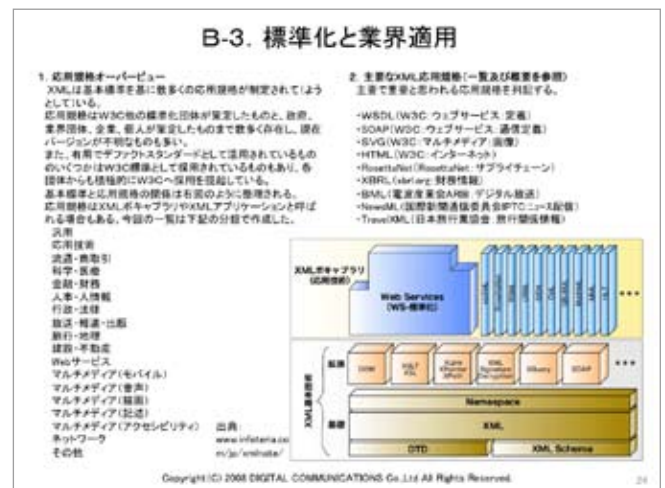
XBRL は、日本銀行に採用されており、国際会計基準はこれで全て行うことになっているようです。

それから、地デジ対応の BML などがあります。

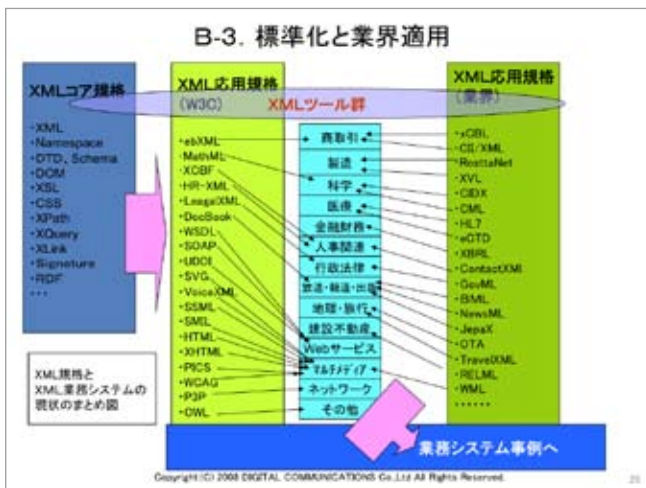
このようにさまざまな業界ごとにきっちりやっている状況の中で、印刷出版向けの日本標準、世界標準をどうするかを考える必要があります。



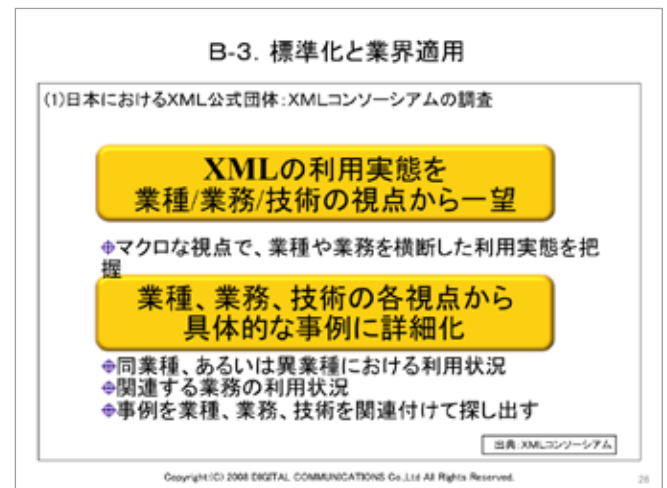
(图 22)



(图 23)



(図 24)



(図 25)

(図 24) は、標準化と業界適用をまとめたものです。
W3C 標準でのコアな規格と、W3C の応用規格と言われるものがあり、これに対して、それぞれの業界ごとの標準規格が作られています。この中に、前述した XBRL や BML などがあったり、医療業界でも HSM や MML の 2 つの陣営が争っていたりと、そうした問題も現実には起こっています。

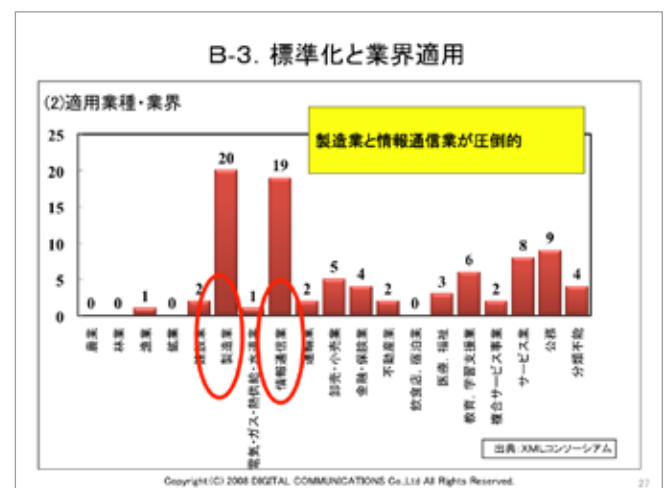
W3C に向けて XSL-FO の日本規格を出そうとしているのが、次に説明する XML コンソーシアム、日本のコンソーシアムです。
(図 25)

こういう動きはもっと積極的に行うべきだと思います。印刷出版系の皆さんは、特にかかわってやられたらいいと思います。JEPAX という出版協会の日本標準の DTD もありますが、0.9 のまま何年もバージョンアップしていません。1.0 にしたくても、怖くて誰も手が出せないという状況なのです。

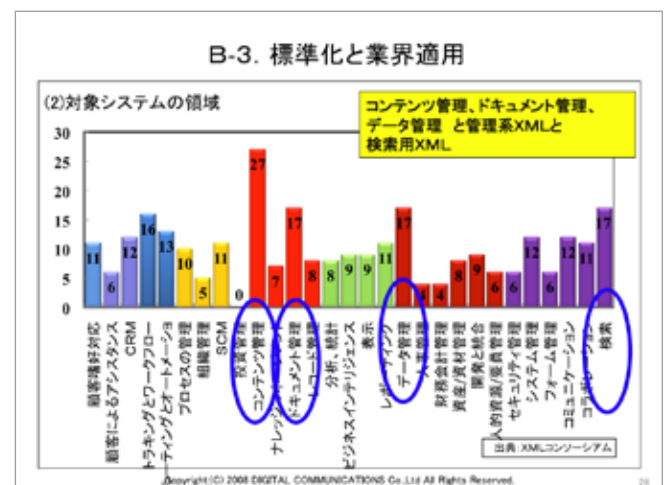
これらは、XML コンソーシアムが、それに向けて XML の利用を調査しようということで作ったものです。

どこに適用されているかという、圧倒的に製造業と情報通信業です。(図 26)

(図 27) は対象のシステム領域を調べたものですが、この表からはコンテンツ系とドキュメント系という印象がありますが、コンテンツ管理とドキュメント管理をどちらで判断するか、つまりインスタンス側の XML と見るか、管理するためのメタデータと見るかで、実は観点が違って来るわけです。次に多いのがデータ管理ですから、これも同様です。あとは検索に XML を使っていることになります。



(図 26)



(図 27)

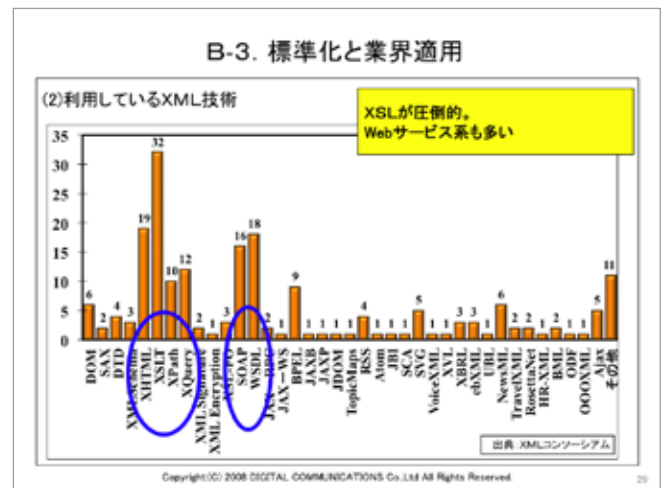
(図 28) は、実際に使っている XML 技術は何かを調べたものです。

これを見ると、圧倒的に XSLT を使っています。ですから、XML を使いこなすには、XSLT を使いこなすことが大事だと思います。XSLT の本来の主要な目的は、XML を HTML に変換するということですが、アウトする側のデータ規格は別に何でも構わないのです。XML を、例えば Indesign 用のデータに変換したり、XML をほかの組版システムに変換したりということができます。そのように考えれば、XSLT は非常に便利な万能ツールですから、一生懸命やる価値はあると思います。

XML 専門家ではないとしても、XSLT をやらずに XML に手を出すのはなかなか大変だと思います。しかし、自分では書けなくても、どういうものを書いてほしいかという要求を出していただければ、もちろん私たちは書きますし、ちょっと分かる人なら書けると思いますので、ぜひ検討してみてください。

Web サービス系は、SOAP と WSDL が多いようです。

SOAP は、データ交換をするために使われているので、そう考えれば、XSLT とこの Web サービス系の技術を使うことは当然かもしれません。これは、印刷出版業でもきちんとやるべきところだと思います。



(図 28)

(C) XML・学術出版の課題と対策

最後に、学術出版系、印刷出版系についてもう少し整理します。(図 29)

結局、我々に残っているものは、自動組版という問題です。

自動組版問題とは、つまりデザインの問題です。これがやはり解決しません。

数式も作れないのではないかとされていますが、それについてはベンダーさんの努力次第というところもあると思います。例えば、モリサワのツールは、MathML から数式が組めます。もう少し調べれば、いろいろなツールで MathML も組むことができますと思います。

化学式を組むツールは、残念ながら私もまだつかんでいませんが、数式を組むツールは基本的にできていますから、この辺も含めて何か考えなければいけないと思います。

原稿も問題です。執筆段階から XML データを作らなければなりません。要するに、XML がなければ自動組版にしたいとできない、何も始まらないわけです。データベースがなければ自動組版にできないのと同じで、コストをかけずにいかに XML データを作れるかというプロセスを作らなければなりません。やはり、執筆段階からの XML データ、私は Word2XML を提唱しています。

また、校正の問題も全く同じです。(1) と (2) の 2 つは自律的な問題ですが、(3) の校正問題は、正直言って難しいところがあるかもしれません。程度問題で、どこまでこれを減らせるかということについての意識改革が必要だと思います。

もう 1 点は、電子ジャーナルへの移行です。

ハイパーメディアを使って、十分な表現のできる電子ジャーナルを作れるようにすることが必要です。そうすることで、紙出版との統合も考えられし、双方の価値を高めていく方法を考えていくべきだと思います。紙出版は読むだけの価値はありますが、データとして使いこなすことができません。電子ジャー

ナルとして PDF を置いていたのでは、これも価値がありません。本格的な電子ジャーナルを紙と統合して出していくことが必要だと思います。

もちろんコストの問題もあります。

学術論文系ですから、査読システムやコンテンツ管理があります。査読システムはいろいろな会社さんがチャレンジされていますが、コンテンツ管理の問題もしっかり押さえていかないと、電子ジャーナルへ移行するプロセスの間に、もっと苦勞がつかまうでしょう。

C-1. 自動組版のデザイン問題

まず、自動組版のデザイン問題ですが、XML は自動組版システムではありません。それを誤解している方がいて、XML で自動組版ができるのではないかと、できないのはデザインが悪いなどと文句を言われる方がありますが、これは XML の責任ではありません。

この解決には、「微細なデザイン調整をやめる」という方法です。(図 30)

企業系のマニュアルは、この方法で一応進んでいます。それができない場合は、手動修正と混合制作ラインを作るべきでしょう。これは運用問題ですから、程度問題でもあります。

コストのラインをどこに置くかによって、手動で修正する分の兼ね合いで運用することはできると思います。

私たちが今やっている大型のもので言うと、ある大手の通販総合カタログがあります。それは、XML から Indesign で自動組版していますが、売る側のお客さまから言うと、それだけでは訴求ポイントが弱いので、Indesign の中に入り込んで、部分的に大きくしたり、小さくしたり、写真を入れ替えたりすると

(C) XML・学術出版の課題と対策

- (1) 自動組版問題
デザイン問題: 数式等
- (2) 執筆段階からの XML データ作成
XML 自動組版するには XML が事前に必要
- (3) 校正問題
前述
- (4) 電子ジャーナルへの移行と紙出版の統合
- (5) コスト問題
- (6) 査読システムとコンテンツ管理

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd. All Rights Reserved.

30

(図 29)

C-1. 自動組版のデザイン問題

- (1) 組版・デザイン問題が解決できていない
- ① XML は本来、下図のように体系に対し、固執しない立場
 - ② XML で自動組版ができると幻想
 - ③ いずれも、フォーマッター (DTP 等) の問題であって、XML の問題ではない
- ⇒ 解決方法は?
- ① 微細なデザイン調整をやめる ⇒ 企業系マニュアルで適用
 - ② 手動修正と混合制作ライン
 - ③ 「構造」と「表現」の分離という固定概念は
「構造」と「表現」を保有した XML から使用時に分離する。
という「デザイン XML」の機能を提供 ⇒ DTP 通上、OpenXML
- ★ フォーマッタの進化を期待 Indesign の進化
★ XSLT の習得 (ユーザーの努力)
- (2) 校正問題が未解決
仮に自動組版ができて、真実まで赤字が入る制作体系では、最終的な XML の生成が困難
完全原稿の作成は不可能、校正プロセスの改革。
- (3) そもそも XML データが作成できない
⇒ Word2XML で解決
- (4) XML データができて、1 種の利用しかできない
(4) マルチに利用するには開発コストがかかる

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd. All Rights Reserved.

31

(図 30)

いうことができるようにし、そのデータを、さらにまた XML に戻すというプロセスとセットにしています。その代わり、修正の仕方に制限があります。

Indesign が本来持っている XML の内部的なリンク情報を壊さないように、しっかり運用する必要があり、日本の DTP 制作会社ではできなかったのが、中国に持って行って 5 日間教え、中国と通信でやり取りしながら運用しています。中国では、DTP は文化産業で非常にハイソな業種になっているので、皆さんの意欲が非常に高いということもあると思います。

要は、手動の部分と自動の部分とうまく混合して、制作ラインを作っていくことです。

それから、これはうまくいきませんでした。XML は構造と表現の分離という概念を持っています。しかし、だからといって XML データの中から直接分離しなければいけないということはありません。

XML の中にデザインと構造を一緒に持っても構わないではないかということを、私は一時、主張していて、Quark から出てくるデータや、Word から出てくるデータを、デザイン情報を持ったまま XML にしました。そうすれば、少なくとも、XML と QuarkXpress の間はネイティブに移行しますし、Word2XML、XML と Word の間は移行したのですが、皆さんに受け入れてもらえませんでした。実は 2001 年くらいから実践してみたのですが、途上のままで終わってしまいました。

人が何か作業をした痕跡というものは、そこにナレッジが込められている可能性が十分にあるので、それをいかに引き出すかということをテーマに考えていきたいと、執念を持っているわけです。それが WordML なのです。嫌いな人もたくさんいると思いますが、マイクロソフトの OpenXML は、まさにデザイン XML という機能を、ある意味、完璧に提供しているわけで、これを使わない手はないと思います。

そして、フォーマッタの進化に期待したいところです。

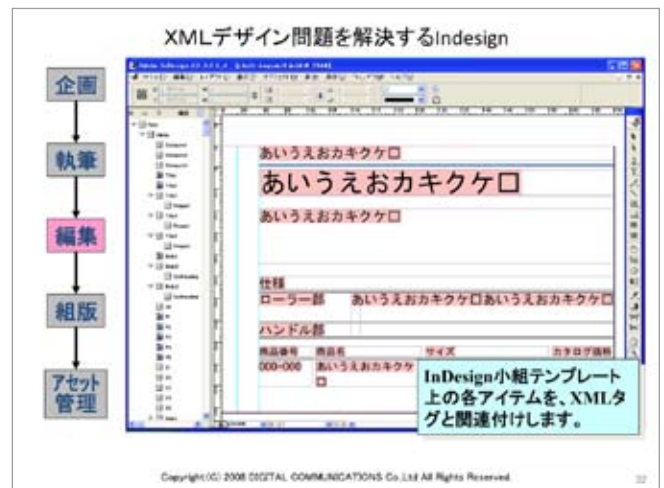
例えば、Indesign は前述のように、少し使いづらいところがありますが、XML 2 Indesign 2 XML の循環的利用は完全に解決できます。

あとは、校正問題や、XML データが作れないという問題のほか、印刷出版業特有の常に as is だけな使い方で、マルチに使っていないという問題があります。マルチに使うことを何とかしないと、コストがかかるだけで、問題は解決しません。

(図 31) は Indesign の中身です。

Indesign には、左側に XML インポートの機能が付いているので、正しく読み込むことができます。ですから、それに対してマッピングをする、つまり、そこにはこれは、ここにはこれということ、デザインしながらやっていきます。そこは手作業ですが、一旦作ってしまえば自動化ができます。

これは、Word2FO に XSL を使うと自動組版ができるということを表している図です。私どもの製品の中からのサンプルですが、左側にメニュー欄があり、ここで微細な修正などが可能です。こうしたものが、PDF のリンク付けまで含めて自動的に作れるようになります。(図 32) (図 33)



(図 31)



(図 32)



(図 33)

C-2. コストと技術問題

次に、コストの問題です。

XML を行うには、このようなツールを全部自前で揃えたうえ、XML と組版システムを準備しなければいけませんので、その部分は技術や努力でカバーする必要があります。それらを全部できなくても、例えば、やりたいことを明確に伝えていただければ、適正なコストで開発のご協力ができると思います。

C-3. ドキュメント XML の作成問題(Word2XML)

XML には 2 つの機能があり、印刷出版系の皆さんはデータを作るほうに目がいきがちですが、データの再利用や、データをいろいろなところにディストリビューションするためには、アトリビュートを付けなくてはなりません。それを付けることによって価値も増します。(図 35)

今までは、私たち専門家が両方行っていました、それはとても大変なのです。左側の XML データ作成 (エレメント) は何とかなっても、右側のアトリビュートを付けるのはやはり大変です。この作業を分けなければならないと思います。

分けてしまえば、左側のほうは Word で作れます。右側も、XML 専門の人たちがやるのではなく、業務、つまりナレッジをよく知っている人たちがやるべき仕事です。つまり、XML の専門家の出番はもうないということです。

私たちは道具だけを用意して、あとは皆さんのほうでやってくださいということです。この 2 つを一緒にやっているために、無理が生じているのです。

今までうまくいかなかった最大の理由は、ここにあります。

C-2. コストと技術問題

3. XMLシステムに必要なツール

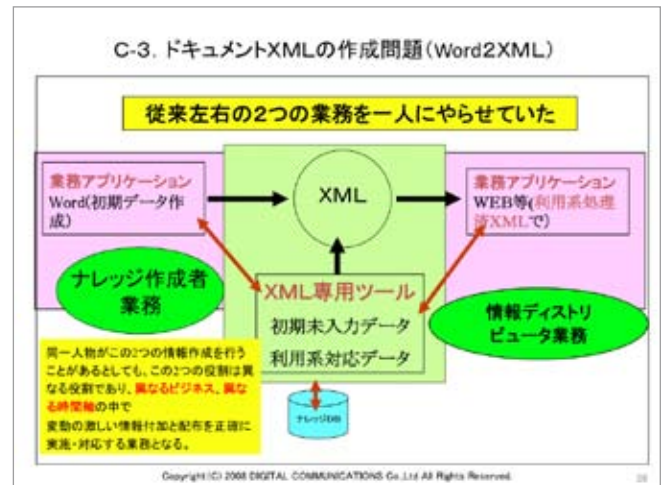
XMLを操作したり、システム化には下記のようなツールが必要。

- XML文書(XMLエディタ、XMLタグ生成ツール)ツール
XML文書そのものを作成するツール。専用のツールやMS-Officeから変換するもの、タグのみを扱うツールなど多数ある。⇒Word2XML
- DTD(XMLスキーマ)設計ツール
DTDやSchemaはかなり専門性の高い技術であり、専用の高価なツールもあるが、直接的に作成するケースが多い。
- XMLパーサ(XML Parser)
XML文書をロードし、構文チェックを行い、XML文書を他のプログラムやシステムへツリー化して引き出すツール。ツールのレベルによってスキーマ妥当性のチェックも行う。代表的なツールとしてMS-XML等がある。
- XSLTプロセッサ
XSLTによって指定された変換仕様を実現するコンバータの一種で、XMLのファイルを含く他の構造や形式のファイルに変換することができる。サーバサイドでXMLをHTMLに変換、Web ブラウザ上のものなどがある。
- XML生成ツール
XML専門性が高く、適当なツールが少ない。直接的に作成するケースが多い。
- XMLデータベース
ツリー構造をそのまま扱うXML専用のDBがある。XML処理のためのAPIが揃っている。その他、RDBのXMLオプションで対応できるようになっている。
- XMLブラウザ
XMLを表示するXSLプロセッサのユーザーアプリ。Internet Explorerは代表的なブラウザである。

+XML自動組版システム

Copyright (C) 2008 DIGITAL COMMUNICATIONS Co., Ltd. All Rights Reserved. 23

(図 34)



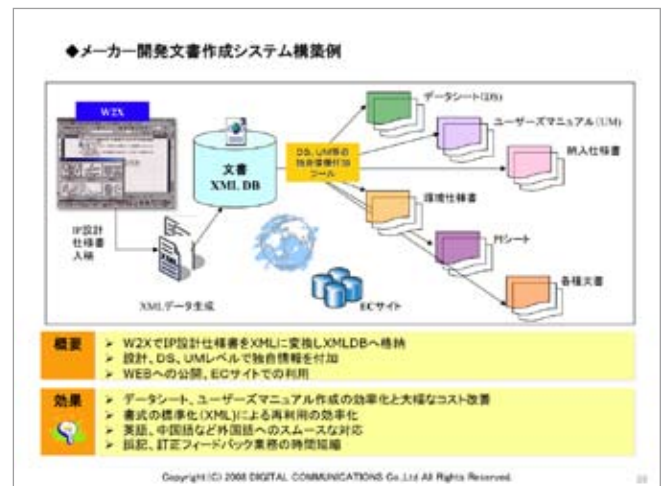
(図 35)

事例集 :Word2XMLで何ができるのか？

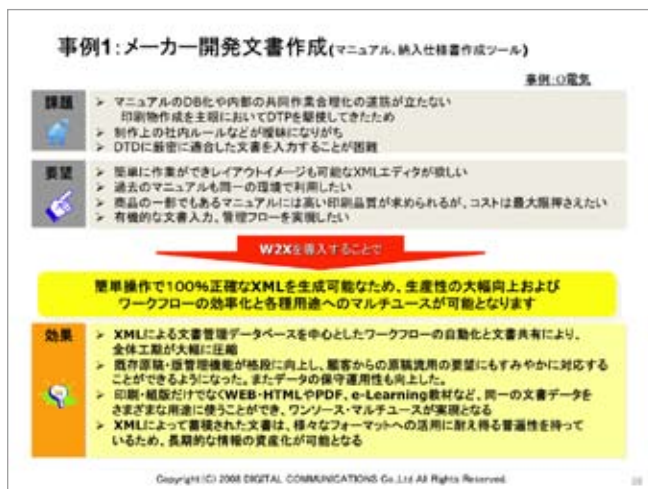
ここで、Word2XML を使うとこんなことができたという事例集を挙げます。

(図 36、図 37) はメーカー系の開発文書です。

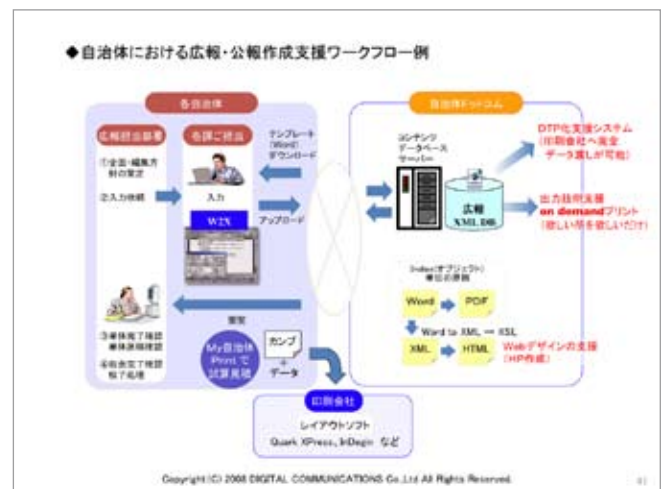
(図 38 ~ 40) は日本最大の自治体の広報・公報です。実は、この条例作成システムを、モリサワと一緒に3年前にWord2XMLを使用して作りました。Word2、XML2、MCB2という形で、縦組みの組版で条例が全てできます。これですと、約2000人の庁内の人たちに、Word2XMLを使って原稿を書いていただき、MCB2で自動組版できます。それを印刷会社で印刷してもらっています。



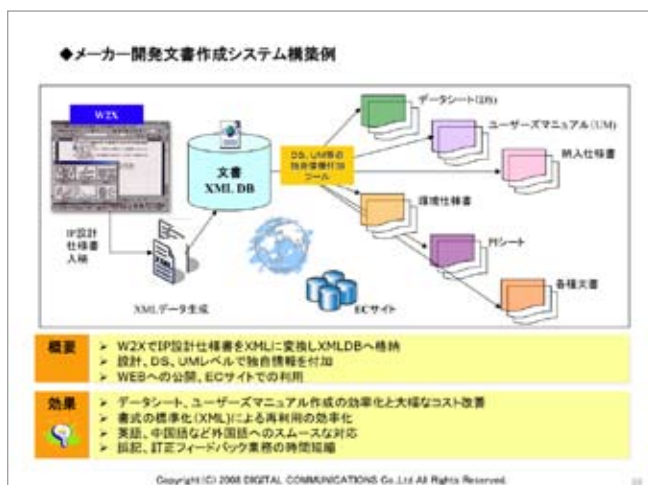
(図 38)



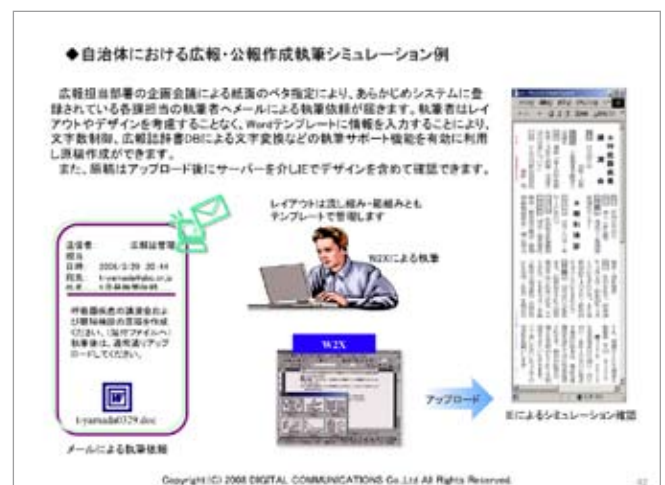
(図 36)



(図 39)



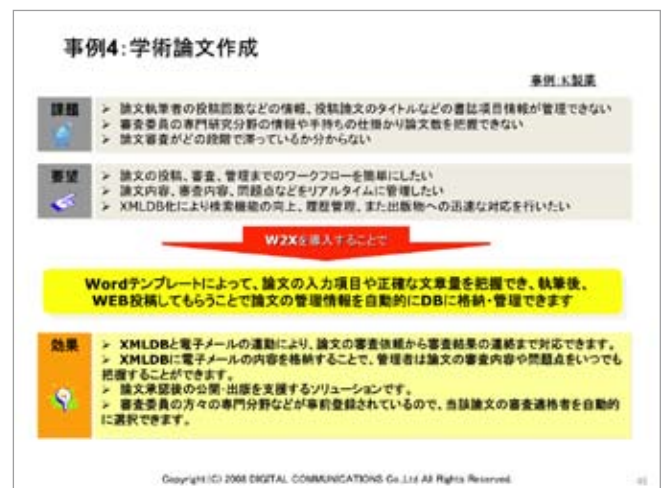
(図 37)



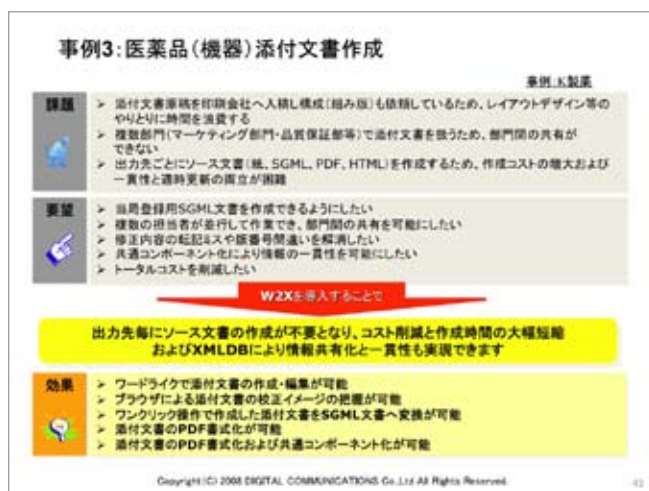
(図 40)

(図 41、図 42) は医薬品添付文書です。医薬品医療機器もちろんそうですが、SGML は必須になっています。これも、印刷会社からはデータを作ってほしいという要求があるのですが、今は私たちがメーカーさんに直接このデータツールを納入しています。

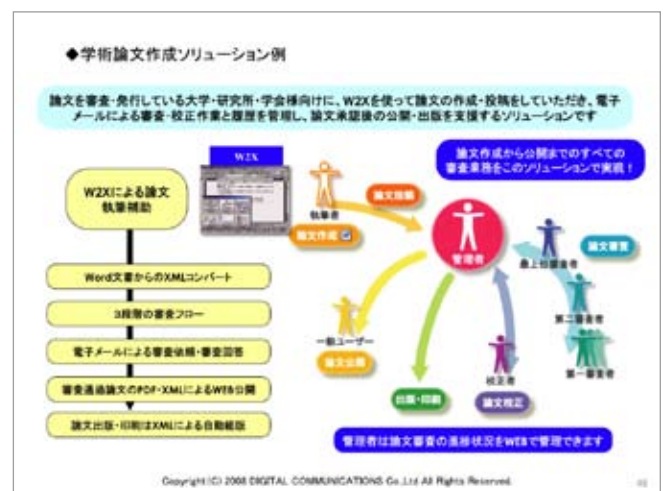
(図 43 ~ 47) は学術論文の Word2XML を使った事例です。当時は 2000~2003 を含めて行いました。当時は Indesign がなかったので、FrameMaker で組む自動組版で行いました。



(図 43)



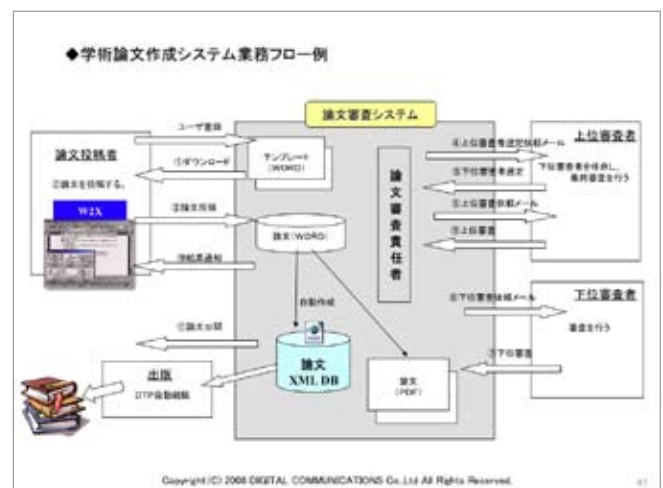
(図 41)



(图 44)

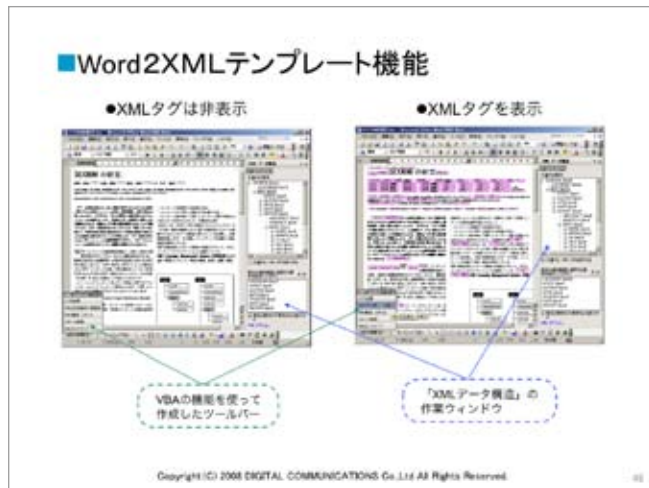


(図 42)

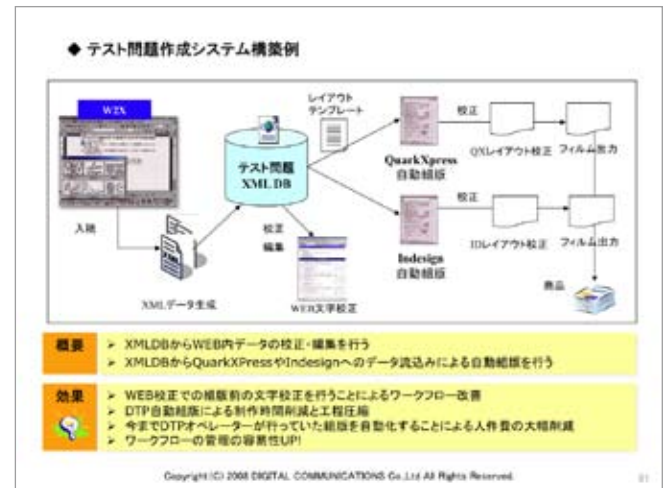


(図 45)

(図48～50)は予備校系の問題作成システムです。ある塾では、塾の前にマンションを一つ借りて、東大の学生にアルバイトで作らせているそうです。Word2XMLで問題を作成し、それを塾の教材に使っています。



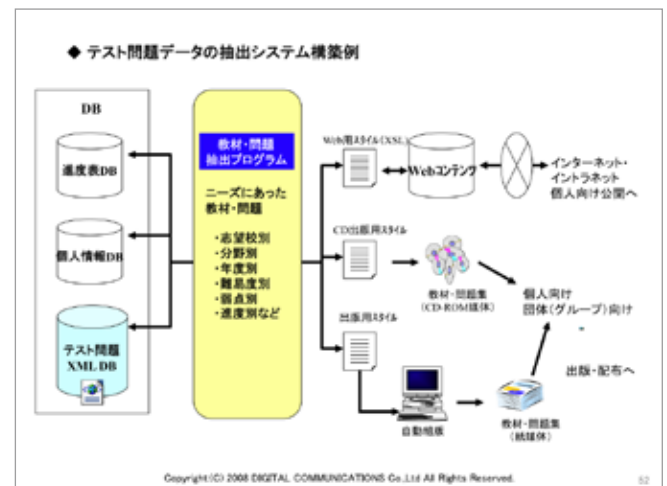
(図46)



(図49)



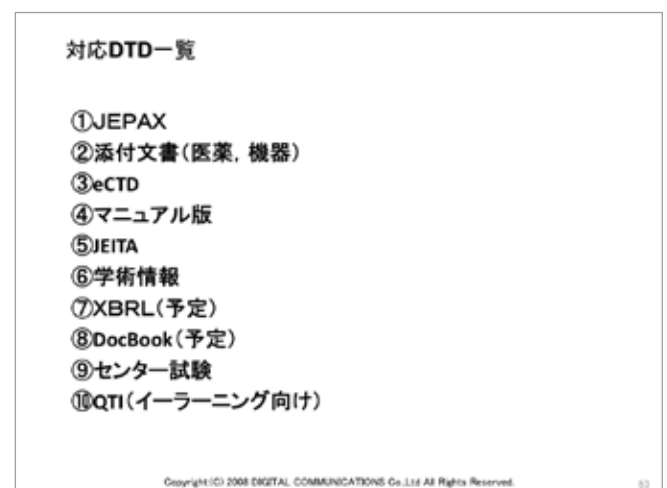
(図47)



(図50)



(図48)



(図51)

C-4. ドキュメントXMLの作成のもう一つの問題点

XMLの作成のもう一つの問題点とは、前述のアトリビュートの作成です。(図52、図53)

WordからXMLを作ることは、100%間違いなくできますが、ナレッジ層をきちんとやる、アトリビュートをきちんと付けるという仕事は非常に難しいのです。印刷出版系から見ればなくてもいいと思うかもしれませんが、XMLDBを採用する側(顧客)は、アトリビュートがどうしても必要です。私どもではXMLの中に書き込むことをやめて、外側に開放しました。そうすることで、教材の専門家の方が自分で書き込むことができるようになります。アトリビュートは自動的に付けられるようになります。

さらにもう一点、世界的な流れとして、DITAという技術文書系の規格があります。

出版系DocBookなどで、アメリカでは行っていますが、技術文書系を、DITAでは実体ファイル側だけではなく、メタ側を活用して作成しています。概念的にはRSSと全く同じですが、ポイントはこれをW3Cが標準として策定したことです。

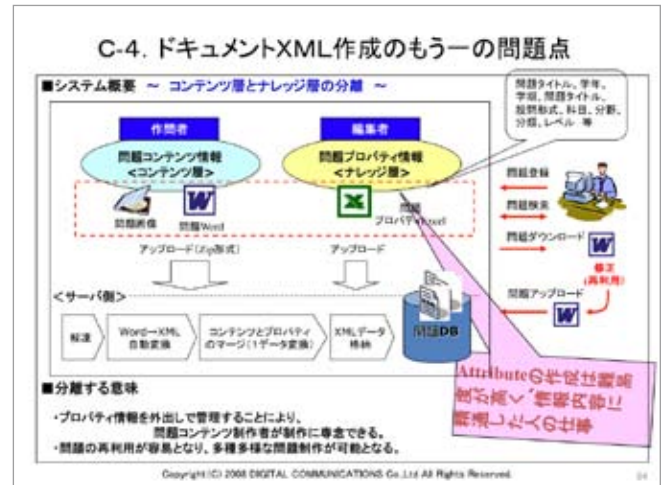
このDITAという規格もIBMが最初に行っていて、それをOASISに無償提供しました。これが標準として勧告され、アメリカでもユーザー会ができるなど、本格的にDITAに移行しようとしています。これは、再利用性を高めることが目的です。

左側がメタデータで、右側が文書実体です。(図54)

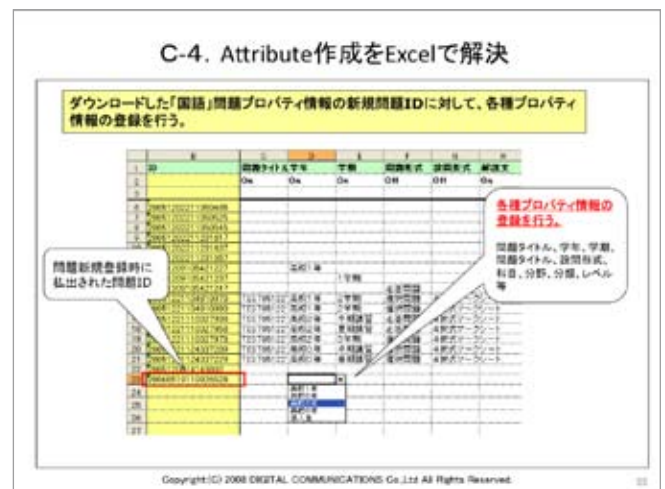
日本では、XMLを長大なデータで作ってしまいがちですが、ここでは文書実体といっても、一つのファイルは短文です。一つが意味を成す最小単位に、全部分割して作ります。分割して作ることで、作業する人も数多くできます。

例えば、文書の一つ一つをそれぞれ専門家が書かなければいけない内容があるとすると、ここはAさんが書けてもBさんは書けない。別のところではBさんしか書けない。そんなときに、AさんもBさんも一つのものが出来上がるのを待ちながら作っていくのは時間の無駄です。分割していれば、一斉に10人でも20人でも同時進行で、それぞれ自分の専門業務に合ったドキュメントを、例えばWord2XMLで、書くことができます。左側の一つをトピックと称して、そのトピックに対するメタデータを作っていくことができるわけです。

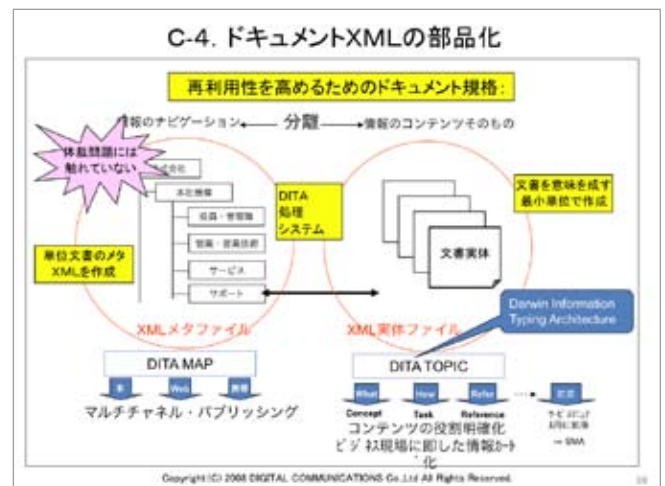
そうすると、メタデータらをどう組み合わせるかもいろいろ



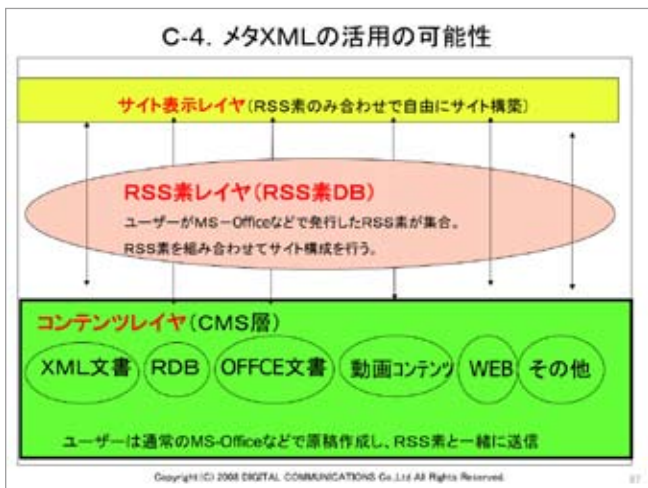
(図52)



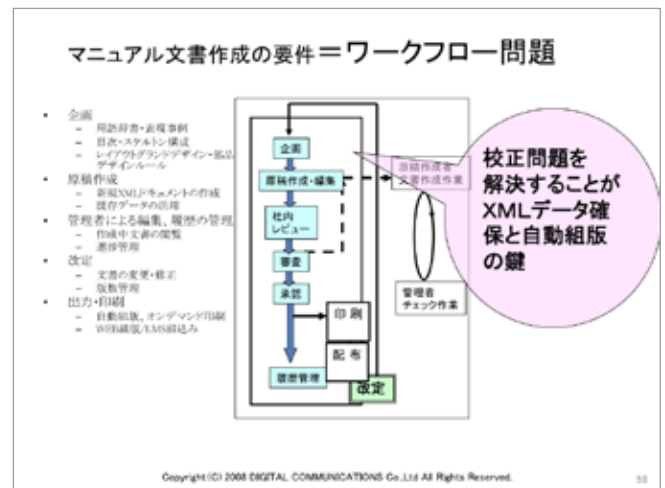
(図53)



(図54)



(図 55)



(図 56)

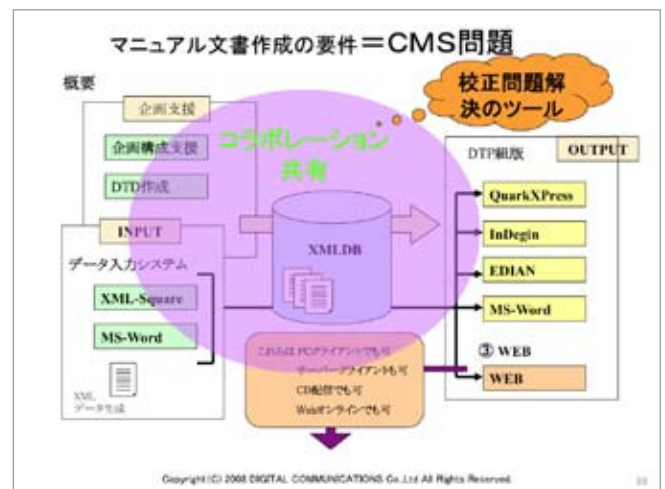
な選択が考えられます。組み合わせ方も自由に選べて、一つのドキュメントを作ることができるのです。例えば、今回のドキュメントはこれとこれ、次のドキュメントはあれとそれというように、活用することができます。(図 55)

要するにXMLは、もう普及期に入ってきているといえるということです。

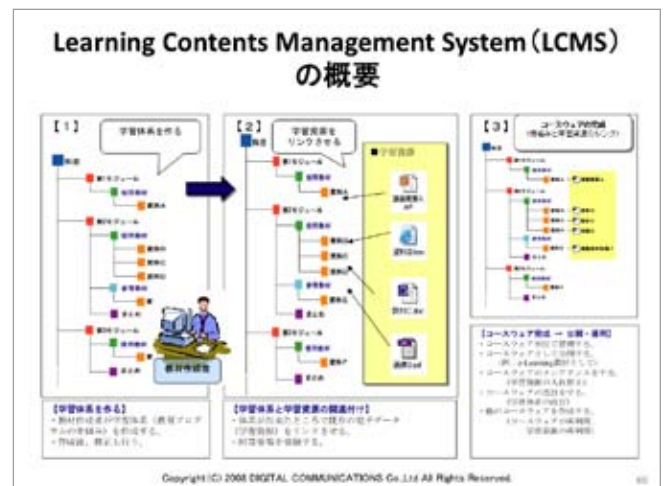
データ系、ネット系の進化に負けない、ドキュメント系XMLの対応を何とかしないと、ビジネスチャンスがなくなってしまうかもしれないと思います。

XML生成は、Officeツール、ここではあえてMS-Officeと書いていませんが、MS-Officeでいいと思います。そのOfficeツールで、ほとんどはXMLの出力が可能で、そのXML、WordMLが気に入らないと言う人もいますが、私は良いデータ規格だと思っています。それをうまく処理すればいいだけです。

このように、コストをかけずにXMLは作れます。校正問題を解決することで、自動組版もユーザーサイドのスキル対応も可能です。(図 56～58)



(図 57)



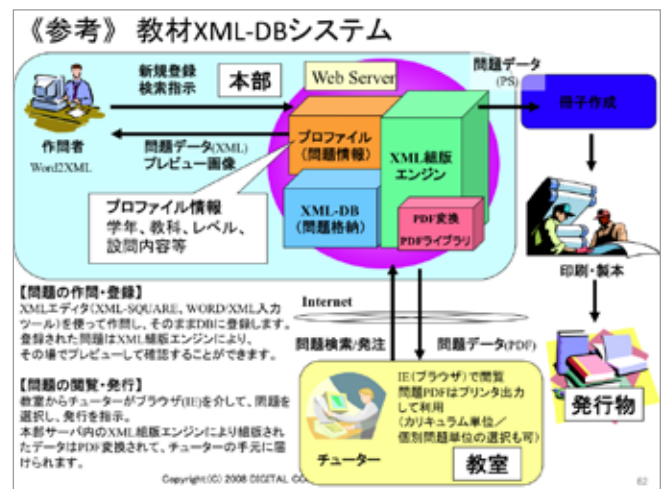
(図 58)

ただし、メタデータの活用を行い、それを高度利用することによってもう少し価値を高めない、コストカットだけでは、どうしても勝てないだろうと思います。価値を上げ、付加することで、逆の意味でのコストカットになるという提案をしていきたいと思っています。

参考までに、私どもの作成した WEB 百科事典サイトや教材 DB システムを添付します。(図 59～61)



(図 59)



(図 60)



(図 61)