

XML 誕生秘話



村田 真

日本 IBM (株)
東京基礎研・国際大学研究所
mmurata@trl.ibm.com

1 XML 誕生秘話

□■まえがき

2002年12月、XML 2002というイベントの会期中に、バルチモア市のとあるバーで、W3C XML Working Groupの往時のメンバが久しぶりに顔を合わせた。私もメンバの1人として参加した。このWGこそがXML 1.0¹⁾の生みの親である。議長のJon Bosakは、テーブルを挟んで向かい合っていた。エディタの1人のEve Malerは右隣でビールを手にしており、技術リーダのJames Clarkは右側のテーブルでウイスキーのグラスを傾けていた。久しぶりに会うJean Paoliはその隣にいた。やはりエディタのTim BrayとMichael Sperberg-McQueenが不参加だったのは残念だが、あの素晴らしいW3C XML WGのメンバとまとまって会えたのは嬉しかった。XML勧告案ができたときの感激は私には忘れがたい思い出である。

XMLは、当初のあらゆる期待を遥かに上回る成功を収めた。1996年の段階では、XMLに懐疑的な人も多く、私も半信半疑だったものである。1997年末に勧告案になるころから、国内外でのXMLフィーバが始まった。どの企業も争ってXMLを実装し、製品で採用することを表明した。XMLについての解説記事は多くの雑誌に掲載され、XMLについての雑誌や書籍はいくつも発行された。2003年現在、米国ではXMLは広く普及しており、日本でもある程度の普及をみせている。必ずしもエンドユーザの目に触れるわけではないが、裏方としてXMLが使われていることは多い(たとえばEDIやWebサイト構築)。このような成功を収めたXMLはどのようにして始まったのだろうか。

□■XML 1.0の制定

XMLの制定がW3Cで正式に始まったのは1996年9月である。国際規格SGML(Standard Generalized Markup Language)²⁾から余分な機構を削ぎとって、Webでの利用に適したサブセットを作ることが目標であった。制定活動に参加した人のほとんどは、SGMLに深くかかわっていた。

1996年当時、すでにSGMLのサブセット提案はいくつも存在していた。やはりXML WGの一員であるSteven DeRoseの書籍SGML FAQ Book³⁾には8つの提案が言及されている。これらの提案は基本的な方針ではほぼ一致していた。SGMLの機能のどれが余分で、どうサブセット化すべきかについては、ほぼ合意がとれていたと言っている。XMLの最初の案は簡単にまとめ、1996年11月の会議SGML 96で一般に公開された。1998年2月にW3C勧告となったXML 1.0は、この案の完成度を高めたものである。

W3CにおけるXML制定活動は、SGML ERB(Editorial Review Board)とSGML WGの2つのグループによって行われた。その後、前者は改名されてXML Working Groupになり、後者はXML Special Interest Groupになる(以下では、簡単のためXML WGとXML SIGで統一する)。多くの人々が参加して議論するのがXML SIGであり、実態としては単なるメーリングリストである。一方、XML WGはすべての決定を担い、メールでの議論のほかに毎週の電話会議を行う。決定権を持つのはXML WGであるが、XML SIGでの議論はたいへん充実しており、XML WGにとってきわめて有益であった。

XML WGの電話会議について説明しておく。参加者は、議長のBosakをはじめとするXML WGのメンバで

あり、1企業からは1名までに制限されていた。電話会議は、2時間にわたって毎週開催されており、投票も頻繁に行われた。当時、参加者はむしろ寡黙であり、議長のBosakに促されなければ話し出さないことが多かった。整然として能率のよい電話会議であった。これには、議長の運営が巧みだったこと、XML SIGでの議論によって問題が整理されていたことなど多くの理由がある。しかし最大の理由は、当時のXML WGメンバが互いに知り合いで、意見がほぼ一致し、相互に敬意を持っていたことであろう。後に当時のメンバは、XML WGは最高だったと回顧することになる。

1997年3月にはXMLについての最初の会議がサンディエゴで開かれる。この会議で、Microsoft社のInternet Explorerチームを統括していたAdam BosworthがXMLのサポートについて述べた。後述するCDFがMicrosoft社からW3Cに提出されたのもほぼ同時期である。この頃から、XMLパーサがあいついで発表された。個人が作成したものと企業（Microsoft社やIBM社などを含む）が作成したものがほぼ同数であったように記憶している。10月には、Microsoft社のInternet Explorer 4.0の一部としてXMLパーサが配布される。1997年初頭にはXMLに対して懐疑的な声も聞かれたが、この頃からXMLへの熱狂が始まる。XMLに反対する企業は1つもなく、各社は争ってXMLへの支持を表明した。

1997年12月8日には、XML勧告案（Proposed Recommendation）が完成する。残っている懸案事項に一応の決着をつけなければならないので、最後の追い込みは大変であった。電話会議も週2回に増え、土日にまで開くことになる。声を聞くだけで、疲れきっていることが分かるメンバもいた。最後に、すべての懸案事項の審議を終えたとき、「我々はついにXML1.0の設計をやり遂げた」と誰かが言った。疲労の果てに爽やかな達成感があった。また、あんな瞬間に立ち会ってみたいと本当に思う。

□■日本におけるXML

私とXMLとのかわりかは、発表された仕様を入手したことによって始まる。SGMLの最も上質の部分をうまく切り出していることが直ちに見て取れた。XMLはものになるかもしれないと思い、議長のBosakにメールを送ったのが1996年12月である。BosakとSperberg-McQueenに励まされて、ドラフトを日本語に翻訳することをまもなく決めた。1997年1月にW3C XML SIGに参加した。

XML SIGに入って最初に手をつけたのは重要資料の翻訳である。XML仕様書ドラフト、XML FAQ⁴⁾、Bosak

議長の「XML, Java, そしてWebの将来」⁵⁾を国内のSGML関係者と共同で翻訳し1997年6月までにすべて公開した。これらの翻訳は、SGML Cafe（富士ゼロックス情報システム（株）が設けたSGMLについてのWebページ）に掲載された。XML仕様書ドラフトの日本語訳は、英文版の改定に伴って何度も改定を繰り返した。これらの翻訳作業と並行して、XMLにおける日本語文字の扱いについて検討した。この検討は、マルチメディア/ハイパメディア調査研究委員会のXML特別作業グループおよび富士ゼロックス情報システム（株）が運営した国内用XMLメーリングリストで行った。検討結果はSGML SIGに提出し、そこでさらに議論された。

これらの活動を続けるうちに、議長のBosakからW3C XML WGに参加するよう8月に依頼された。XML WGの超人的な活動には畏怖すら感じていたので、XML WGへの参加依頼は驚きであった。XML WGメンバはXML SIGに流れる膨大な量のメールを読み、XML WGの電話会議に参加しなければならないのである。私には重過ぎる仕事のように思えて逡巡していたが、当時の上司・同僚に励まされて、XML WGメンバを引き受けることにした。

その後は、XML SIGとXML WGに流れる膨大な量の英文メールを読み、毎週の電話会議に参加した。英語力も十分ではない私には、大変な作業であった。また、当時の私はHTTPやMIMEやUnicodeについての知識もまったく乏しかった。

しかし、私にもいくつかの強みがあった。1つは翻訳作業を行っていることである。英文で書かれた仕様書を日本語に翻訳しようとするれば、一言一句まで徹底的に検討することを余儀なくされる。この検討の間にいくつもの誤りを指摘することができた。もう1つは、私が日本人だという単純な事実である。私を通じて日本にXML





を広めたいという思いが、関係者の間にあったようだ。当時のメールを読み返すと、翻訳作業を XML WG のメンバが実によくサポートしてくれていたことが確認できる。メーリングリストで英文メールを読み違えて変なことを書いても大目に見てもらったように思う。もう1つは国際化である。当時の私はまったく無知だったが国際化に関する日本人専門家にはよく助けていただいた。

9月頃から私は、日本語化・国際化にとどまらず、XML や XML 名前空間のすべての点について積極的に意見を述べるようになる。後に Bray は「The Annotated XML Specification」⁶⁾ で、

While Makoto arrived in the WG relatively late in its working life, he had very substantial input into the spec, and was a diligent and effective reviewer.

と書いている。私が XML コミュニティの一員として本当に受け入れられたのはこの頃であろう。

□■ XML 前史

XML 1.0 ができるまでの歩みは、順風満帆だったように思えるかもしれない。しかし、XML 1.0 の制定が W3C で始まるまでには長い不遇の歴史があった。

XML は、20 年以上の歴史を持つ構造化文書の 1 つである。構造化文書には ISO/IEC の ODA (Open Document Architecture) や Xerox の Interscript などもあったが 1995 年にはこれらはほぼ消滅していた。LaTeX も構造化文書であるが、研究者が論文を書くときに限って利用するだけと言ってもあながち的外れではないだろう。一般向けの構造化文書として残っているのは SGML しかなかった。しかし、SGML が利用されるのも、ごく一部の大手企業が大量かつ画一的なマニュアルを作るときなどに限られていた。SGML を採用することのコストと、SGML によってもたらされるメリットを比較すると、普通はコストのほうが圧倒的に上回るという結論が出つつあった。

そこに HTML が現れた。Mosaic や Netscape などのブラウザが無料で配布され、HTML によって書かれたページが星の数ほど出現した。HTML の仕様は拡張を続け、フォームやテーブルまで実現した。Web には HTML があればそれで十分だとほとんどの人は考えていただろう。HTML は、SGML に似た構文を用いているが、SGML そのものではない。特に、利用者が独自タグを定義できるという拡張性は HTML にはない。実質的に、HTML を拡張できるのは、ブラウザを実装している会社に限られていた。

ここで SGML 関係者は悲哀をかこつことになる。

HTML の陰で SGML はひっそり消えていくのではないかという危機感があった。SGML 関係者でなければまず SGML を知らないし、知っている人もすでに過去の技術だと考えていた。1996 年頃に comp.text.sgml などのニュースグループには、SGML の将来についての悲観的な意見が多かったように記憶する。

この頃に、ひっそりと XML の胎動が始まる。SGML を Web で使えるようにしようと、最初に呼びかけたのは Yuri Rubinsky である。彼が中心となって、簡略化された SGML を Web で使おうという動きが始まった。彼の会社である SoftQuad は、SGML 対応 Web ブラウザ (Panorama) を無料で提供した。惜しまれつつ彼は 1996 年に急逝した。私自身は一度しか会ったことがないが、彼の影響を受けた人は XML コミュニティには多い。

Rubinsky の志を継いだのが Jon Bosak である。Bosak は、1996 年 8 月にワークショップ「SGML と DSSSL のインターネットでの応用」を開催し、インターネットで SGML を使うことの利点を検討した。これらの利点は、「XML, Java, そして Web の将来」にまとめられている。インターネットで利用するには、SGML は不要な機能が明らかに多すぎた。Bosak は、業界標準として SGML のサブセットを作ろうと関係者に呼びかけた。その後で W3C チームと折衝し、XML に関する標準化活動を W3C で行うことを認めさせた。

W3C チームの了承を取り付けたことが Bosak の最大の功績だと Tim Bray は発言 (以下の引用を参照)⁷⁾ しており、Bosak もそれを認めている。

It is to Jon Bosak's immense credit that he (like many of us) not only saw the need for simplification but (unlike anyone else) went and hounded the W3C until it became less trouble for them to give him his committee than to keep on saying SGML was irrelevant.

XML が始まったころ、W3C チームは SGML に懐疑的であり、W3C で SGML についての活動することに反対していた。その後も、1997 年 8 月まで W3C チームはほとんど XML に関与していない。Bosak の強談判により、W3C のリソースを使わないことを条件として、XML についての活動がやっと承認されたのである。

□■ XML の普及

元々 SGML の延長の構造化文書として XML は始まった。黎明期から XML にかかわっている人のほとんどは、ボーイング社の飛行機整備マニュアルのような長大な文書を SGML で印刷した経験を持つ。Web 上で SGML の

サブセットを使うことになっても、いきなり別の応用分野を志向したわけではない。SGMLの伝統的なアプリケーションである電子出版をWeb上で実現するためにXMLは作られたのである。その証拠に、XMLの初期においては、今でいうXML 1.0はXML技術の第1部であり、リンク技術（今日のXLinkとXPointer）がXMLの第2部、スタイルシート技術（今日のXSLTとXSL-FO）がXMLの第3部として位置付けられていた。

ところが、XMLの普及は、電子出版とは別の形で始まった。最初のきっかけは、Microsoft社のCDF（Channel Definition Format）の出現である。CDFは、SGMLの伝統的な処理対象である文書ではなく、Webのプッシュ型配送のためのメタデータを表現する。CDFデータはそのまま表示・印刷されて人間の目に触れるものではなく、Webブラウザが内部的に利用するデータに過ぎない。

表示・印刷される文書ではなく、Webアプリケーションが処理するデータのためにXMLを利用した例はその後もいくつも現れている。特に、EDI（Electronic Data Interchange）は、現時点ではXMLの最も成功した応用分野だろう。XMLは電子商取引のためのデータ交換形式であるという説明すら現れるようになったほどである。元々の構想にはない分野からXMLの普及が始まったのは歴史の皮肉である。

一方、W3Cにおいては、HTMLを拡張する体系的な方法としてXMLは位置付けられることになった。まずHTMLそのものをXML化したのがXHTMLである。単にHTMLをXML化するだけならメリットは少ないが、HTMLに新しい機能を取り入れようとするときにXMLは効力を発揮する。たとえば、数式を表現するためにXMLに基づくマークアップ言語であるMathML（Mathematical Markup Language）が制定され、ベクター図形を表現するためにこれもXMLに基づくマークアップ言語であるSVG（Scalable Vector Graphics）が

制定された。XHTML文書の中にMathMLやSVGのタグを埋め込めば、HTMLの数式拡張・ベクター図形拡張が実現できる。現時点ではXHTMLが広く普及しているとは言えないが、今後には十分期待が持てる分野であることは間違いない。

最後に、元々のXMLの用途であった電子出版について述べておこう。XML文書からHTMLへの変換は、XML専用の変換言語であるXSLTやいくつかのプログラミング言語（perlなど）によって広く行われている。しかし、ページやカラムまで含めた組版のための技術であるXSL-FOについてはまだ普及したとは言いがたい。XMLのためのリンク技術はまったく普及しておらず、今後の見通しも明るくない。現状をまとめると、電子出版へのXMLの応用は着実に進んでいるが、その足どりは緩慢であるといえる。ただし、Microsoft社のOffice 11がXMLを採用することによって、電子出版への応用が一気に進む可能性もある。

□■次回以降の予定

今回は、XML 1.0の成立過程を駆け足で説明した。次回以降は、1回ごとにテーマを立て、それに関する事実と私見を述べることにする。第2回は「企業と個人」、第3回は「国際化」、第4回は「情報アーキテクチャとしてのXML」を予定している。

参考文献

- 1) W3C 勧告, Extensible Markup Language 1.0 (1998).
- 2) ISO 8879, Standard Generalized Markup Language (SGML) (1986).
- 3) DeRose, S.: The SGML FAQ Book, Kluwer Academic Pub. (1997).
- 4) Flynn, P.: The XML FAQ, <http://www.ucc.ie/xml/> (1997-2003).
- 5) Bosak, J.: XML, Java, and the Future of the WWW (1997).
- 6) Bray, T.: The Annotated XML Specification, <http://www.xml.com/axml/testaxml.htm> (1998).
- 7) Bray, T.: Re: History, <http://lists.xml.org/archives/xml-dev/199911/msg00642.html> (1999).

(平成 15年 3月 4日 受付)

