

「XML の勘所」特集号について

丸山 宏

1. はじめに

1998年2月10日に、World Wide Web Consortium (W3C) によって Extensible Markup Language (XML) 1.0 勧告が制定されてから12年が経った。この12年の間に、XMLはIT産業全体に浸透し、文書のマークアップ、メタデータの表現、メッセージ交換のフォーマット、ソフトウェアの構成ファイルの構文などに、広く使われている。

XMLは簡潔な仕様であるが、その自由度の高さゆえに、うまく使うにはそれなりにノウハウを必要とする。この特集では、XMLを実務の第一線で使っている方々に、実践から得られた知見について論じていただく。

2. 業界別のデータ交換フォーマットをいかに制定するか

XMLは、もともと文書マークアップ言語である Standard Generalized Markup Language (SGML)の簡素化したものとして企画されたが、XMLの仕様制定が最終段階に入った1997年の末ころ、IT業界のアナリストたちが、XMLはアプリケーション間のデータ交換フォーマット、特に企業間をまたがる、ビジネス・ツー・ビジネス、いわゆる B-to-B のメッセージフォーマットとして使えるのではないかと、いう主張をしはじめたことをきっかけとして、XMLはデータ交換の形式としても注目を浴びるようになった。

B-to-Bのデータ交換フォーマットでは、業界ごとに標準ボキャブラリが決められていくことになる。ITベンダが中心となって制定されるインフラ系の標準とは異なり、業界のデータ交換フォーマットは旅行業界、電子部品業界、銀行業界など、業界が中心となって制定する。つまり、XMLやWebサービスの専門家ではなく、ITのユーザ企業を中心とする活動である。

本特集の最初の2論文は、このようなユーザ主導のデータ交換フォーマット制定についての議論である。

(1) 「防災気象情報を一人一人に届けるために一気象庁防災情報XMLフォーマットの策定」

気象庁の長田氏によるこの論文では、防災情報を様々なユーザに伝えるための標準フォーマットの策定と、そ

の過程から得られた知見について議論する。防災情報フォーマットを標準化するためには、まず気象庁で行っている関連する様々な業務の関連を調べ、共通化できるところは共通化し、そうでないところは業務ごとの独自の拡張を加えなければならない。このような、ステークホルダ間の調整をいかに行うかが、データフォーマット標準化の大きなポイントであることをこの論文は教えてくれる。

(2) 「データレポートのためのXBRL」

XMLはデータ交換の構文を規定するが、データのセマンティクスには踏み込まない。主に財務関連のレポート記述言語として設計されたXBRL (eXtensible Business Report Language)は、項目の合計が一致しなければならない、などの、データ間の整合性等のセマンティクスを導入した標準である。著者の和田氏は日本銀行勤務時代にXBRLの制定と実装に貢献した経験から、このようなセマンティクス導入に関する勘所について議論する。

3. XMLデータをいかに格納・検索・利用するか

XMLの利用が広まるにつれ、XML形式のメッセージを交換するだけでなく、これらのXMLデータをデータベースに格納し、検索し、利用するための技術の必要性が高まってきた。XMLはもともと木構造データを表現するデータフォーマットであり、そのままでは、表形式のデータを仮定する既存の関係データベースにはなじまない。XML文書全体をバイナリ形式のオブジェクトとして格納するだけでは、文書の内容や構造に基づく検索ができない。

このため、XMLデータをXMLの木構造を保ったままデータベースに格納しようという、XMLデータベースの技術が発展してきた。次の2本の論文は、この新しい技術をどのように利用したらよいかの知見を、具体的な実践例を通して与えてくれる。

(3) 事例に基づくXMLデータベースの適用検討

著者の服部氏は、XMLデータベースの開発者であり、実際に商用のXMLデータベースを設計した。この論文では、XMLの木構造を保ったままデータベースに格納するネイティブXMLデータベースの設計と実装について

議論する。近年のXMLデータベース技術の発展に伴い、テラバイトのオーダーのデータを扱うことができるようになり、実用性が増している。具体的な文書管理アプリケーションを通して、実際の利用に関する知見についても議論している。

(4) OOXML の管理手法—ハイブリッド型 XML データベースへの格納—

XMLはその柔軟性から、木構造データも、表形式のデータも同じ文書の中で扱うことができる。少し複雑なデータになれば、同じXML文書の中に、論理的に複数の形式のデータが混在することは稀ではない。小西氏のこの論文では、Microsoft Officeの文書形式であるOOXMLを題材にして、このような複雑なデータをどのようにデータベースに格納するかについて議論する。同じXML文書でも、もともとが表形式のデータは表に、木構造のデータはXMLネイティブのデータベースに格納するというハイブリッド型のアプローチについて、その得失を議論する。

4. スクリプト言語によるXMLの処理

XML普及の初期から、JavaやCなどのメインストリームのプログラミング言語には、XML向けのツールが提供され、またそれらに基づくプログラミングのテクニックについても多くの書籍やWebサイトからふんだんに情報を手に入れることができた。

XMLがいわゆるエンタープライズシステムだけでなく、様々なWeb上のサービスで使われるようになり、XMLをスクリプト言語で処理することも多くなってきている。スクリプト言語には、それぞれの言語仕様の特徴があり、それらの特徴をうまく活かすとXML処理のプログラミングスタイルも変わってくる。

(5) RubyによるXML処理

著者の大場氏らは、プログラミング言語Rubyの普及に力をいれている。この論文では、動的なオブジェクト指向言語であるRubyならではのXML処理の方法について議論する。特に、動的な型付けや、イントロスペクションによる動的なクラスの生成は、CやJavaのプログラマでは思いつかない、新たなプログラミングスタイルを可能にしている。XMLの様々なツールも、このように新しいプログラミング言語のスタイルに合わせて進化していくべきであろう。

5. XMLの普及

XMLは1990年代の終わりに導入された新技術であり、

その後10余年で広く定着した。このような新技術の導入と、その普及には様々なドラマがある。

(6) XMLの技術者教育の歴史と克服してきた課題

著者の平野氏は、1998年にXML専門メーカーとしてのベンチャーを立ち上げた方である。XMLの普及がまたビジネスの成功につながるということで、XMLの技術者認定制度の制定など、XMLの技術者教育に多くの力を注いできた。この論文ではその取り組みと、それから得られる知見について議論する。今後また同様な技術を導入し、普及させるときに、この論文の知見が役立つことがあるだろう。

(7) インタビュー

XML教育に関する平野氏の論文に加えて、このインタビューでは、日本におけるXML普及に尽力された二人の方に、その歴史を語っていただいた。新しい技術が導入され、普及されるまでには、過剰な期待を受け、それから幻滅期に入り、それを乗り越えて普及する、といういわゆる「ハイブ・カーブ」があるとと言われる。XMLコンソーシアムの副会長として活躍された、平野氏、田原氏は、まさにこのハイブ・カーブを肌で感じてこられた方々である。IT産業に携わるものとして、彼らの経験から学べることは多いのではないだろうか。

6. 特集号の編集を終えて

デジタル・プラクティスには、実務・実践を通して多くの知見を得ているが、あまり論文を書く習慣のない技術者の方に、そのノウハウを公開・共有していただく場としての役割がある。

では、日々実務で忙しいこのような方々が論文を書くモチベーションは何だろうか？アカデミアにいる研究者であれば、学術論文を発表すればそれは自分の業績になり、博士号の取得や将来のアカデミアでの就職に有利になることだろう。しかし、多くの実務家の方は学位取得や、業績リストに1項目を加えることに興味はないだろう。基礎研究など一部の部門を除けば、企業のほとんどの部署では、学会での論文発表は評価の対象にならないだろうし、そもそも企業のノウハウや人材が流出する懸念から、社員が学会での外部発表を行うのを嫌う企業もある。

にも関わらず、技術者にとって、自分の考えたこと、やったことを文章の形でまとめることは、自分に対する投資対効果が高い、ということを描きたい。技術者としての人生を20年、30年歩むとするならば、その間、非常に多くの仕事をするだろう。しかし、それらの仕事の結果、あるいはその仕事の結果得られた知見を文書の

形でまとめておかないと、いずれそれは散逸してしまう。仕事の中で、「この問題は以前解いたことがあるな」ということを思い出しても、その詳細が手に入らなければ、また同じことを一からやることになるだろう。あるいは、以前解いた問題について、他人から聞かれたときに、うまく説明できないだろう。自分の仕事の知見を文章の形で残せば、時間と空間を越えて、思っても見なかったような聴衆にリーチすることができる。

論文とは、ある知見についての完結した記述である。適切な参考文献を参照すれば、時間や組織の文脈を超えて、著者の知見を共有することができる。したがって、自分の仕事の成果を記録しておきたいのであれば、論文の形をとった文章にするのは大きな意味がある。

加えて、技術者個人としては、学会等、外部の仕組みを使って発表することが重要であることも指摘したい。これからの変化の大きい社会では、技術者がそのキャリアのすべてを一つの会社で過ごすことはまずないと言ってよいだろう。たとえ定年まで働けたとしても、定年後のキャリアがいずれ必要になってくる。しかし、会社の中でどんなによい成果を出し、文書化したとしても、会社を離れた後にはそれらの記録にアクセスする術がなくなってしまう。学会論文として公開されたものであれば、未来の自分はいつでも読み返すことができる。この意味で、論文は、将来の自分に対する投資と言えるのではないか。

今回の特集でも、それぞれ第一線で活躍されているが、あまり論文を書いたことのない著者の方々をお招きした。それぞれの方が、今回の特集号の論文の、ご自分にとっての価値を、いずれ認識してくださることを願いたい。