日本の電子ジャーナル製作に関する諸考察と、NLM-DTD XMLを利用した電子ジャーナル出版

Study for e-journal production in Japan and e-publishing with NLM-DTD XML

林 和弘¹ | 中谷 敏幸¹ | 太田 暉人¹

HAYASHI Kazuhiro¹; NAKAYA Toshiyuki¹; OHTA Teruto¹

1 日本化学会 (〒101-8307 東京都千代田区神田駿河台1-5) Tel : 03-3292-6165 E-mail : hayashi@chemistry.or.jp ; nakaya@chemistry.or.jp ; t-ota@chemistry.or.jp

1 The Chemical Society of Japan (1-5 Kanda-Surugadai Chiyoda-ku, Tokyo 101-8307)

原稿受理 (2009-01-26)

(情報管理 51(12):902-913)

1. はじめに

インターネットの発展とともに電子ジャーナルの浸透が進んで久しく、電子ジャーナル自身の、あるいはデータベースなどに関連する情報提供サービスは研究者にとっての必要不可欠な存在となった。出版者も当然のように冊子の出版から電子ジャーナル製作を前提とした出版・製作体制に変更し、電子ジャーナル製作を促す電子化はさまざまな場面で起きている。まず、著者が作成したデジタル原稿は、電子投稿、電子査読を経て評価され出版の可否が問われる。掲載可となった原稿は電子出版によって時には印刷前に全文PDFもしくはhtmlが全世界（の購読者）に向けて発信され、多くの場合、全文検索も可能になる。今では引用文献へのリンクや被引用論文からのリンクもほとんどの大手出版者で実現し、公開後に個々の論文情報はほどなくして2次情報DBや引用DB等に登載、もしくはインデックスされ、さまざまなデータベースを通じて1次情報への多数のアクセスルー

トを確保する。このようにして、電子ジャーナルは冊子体と比較して、より早い、より広い世界の読者向けの、より便利なサービスを実現した¹⁾。

さて、この電子ジャーナルサービスを支えているのが、検索やリンクのためのデータであり、今回話題にするXML等のタグ付けされた構造化文書が利用されることが多い。今現在はこのような電子ジャーナル用の基礎データ（慣習的に出版用メタデータとも呼ばれる）が、PDFを主とする本文用ファイルと一緒に用意されて初めて多様な電子ジャーナルサービスが実現すると言ってもよい。また、このような基礎データは本来どのようなものでもよく、実際に、SGML、XMLに限らず、分野や雑誌によってはTeXや、ラベル付きのテキストファイルの場合もある。いずれにせよ、このような電子ジャーナル用基礎データを作成することが電子ジャーナルサービスを実現するために不可欠であり、その基礎データの質が提供できるサービスの質と比例する。また、電子ジャーナル事業を安定化させるためには基礎データ作成のコスト考

察が不可欠となっている。

日本化学会でも1989年より英文論文誌Bulletin of the Chemical Society of Japan (BCSJ) 誌の電子化に取り組み, 2001年までは独自の電子ジャーナルを, 2002年よりJ-STAGEを利用した本格的な電子ジャーナルを実現してきたが²⁾⁻⁴⁾, これらの運用体制を構築する際, 種々の出版スタイルの試行錯誤を経て現在の出版体制に至っている。その結果, Chemistry Letters誌では世界最速クラスの発行体制を整えることができ, 電子ジャーナル課金, COUNTER準拠, RSSやGoogle Scholarとの連携など, 日本では数少ない国際標準的な電子ジャーナルサービスを実現し, さらに, 電子ジャーナル事業としても安定化させた⁵⁾⁻⁷⁾。このような電子ジャーナルサービスの基礎となる出版体制を発展させるに至った経緯を日本と欧米の電子ジャーナル製作手法の違いとともに紹介し, 2009年から始まったNLM-DTDを用いた新しいXML出版体制⁸⁾について考察を加える。

2. 古くて新しいXML (SGML) 学術出版

まず, XMLについて簡単に触れると, XML自体はSGMLの発展系として1998年に勧告され10年が経っている構造化文書もしくはマークアップ言語である。この10年の間にいわゆるデータ型のXMLとしては金融システムなどのデータベース間のやり取りを円滑に行う中間ファイルの形式として世界中に浸透した⁹⁾。

一方, テキスト型のXML, つまりXMLを使って書籍や雑誌等の文章を出版する世界ではというと, 主に辞書, マニュアル, カタログなどの分野で浸透してはいるものの, こと学術出版に関して言えば, 特に日本ではほとんど浸透していないと言ってよい。その理由は, 主に文書の構造化と誌面の自由度のバランスが難しいことと, 費用対効果が

図1 文章構造化とレイアウト力からみた各学術出版ソリューションの位置づけ

出しにくいことであった。

まず, 前者は, SGMLやXMLはデータや文章の構造化を設計する言語であって, その内容をWeb上や誌面で表現する手段に乏しい。あるいは, その文書構造化の設計の制約でレイアウト表現が制限されることが多い。一方, 誌面は究極的には手書き原稿のオフセット出版が示すように元来自由なものである。そして, 少なくとも今現在までの著者や読者は見た目を重要視した。その結果はSGMLよりも, そのサブセットでありWebページを記述するhtmlの方が圧倒的に世の中に受け入れられてWebの浸透を押し進めたことで一目瞭然であった。しかしながら, 先に述べた電子ジャーナルのサービス実現のためには, 文章のある程度の構造化とタグ付けが必須である。ここに, 文章の構造化とレイアウトのバランス問題が発生した。今現在では, 構造化文書を編集するソフトがなるべく自由な表現を提供する方向と, ワードプロセッサやDTPソフトが構造化文書に対応する方向, この両面からさまざまな検討がなされている。理想的には, 高度な構造化文書を持ち, かつ, 自由な組版ができる出版ソリューションを模索しているが, 現時点ではその完全な両立は難しくなんらかの妥協や機能限界を持ちながら出版を行っている。図1は文書構造化とレイアウト対応を軸にとりて各学術出

版ソリューションを位置づけたマトリックス図である⁹⁾。

続いて後者の問題は、端的に言えば、タグ付けのコスト問題である。XMLにせよ他の規格にせよ、これが著者で、これが主題でとわかるラベル付けが必要であって、さまざまな自動化が試みられてはいるものの、最終的には人の目で確認する作業を必要とする場合がまだまだ多く、労働集約型の作業が加わることになる。さらに、タグが多い方が、つまり、付加情報が多い方がさまざまなサービスに発展させることを可能にするが、作業負担は当然タグ付けの量と比例する。そして、不必要なタグを入れることはコスト的には無駄となるが、どこまでのタグが必要かの線引きが難しく、そもそものタグ付けコストと合わせて事業的に見た場合は慎重にならざるを得ない。

3. 欧米のSGML (XML) 出版と日本の対応

欧米の大手学術出版者はこれらの問題を、組版ソフトへの継続的な投資と労働集約作業を海外へアウトソースすることで解決した。まず、著者の原稿は労働人件費の安い国の企業に送られ、ほとんどの大手出版者が、SGMLあるいはXMLのタグ付けを行っている。得られたXML類はワンソースマルチユースの概念に従って利用され、その当時で便利な組版ソフトに流し、誌面を構成していった。あるいは、XML類から全文htmlなどの電子ジャーナルを作成した。このような変革に対応するにはそれまでの冊子を主体としたワークフローを根本的に変えなければならなかったが、紙の出版と電子出版のワークフローを融合させて作業とコストの効率化を図った¹⁰⁾。

一方、日本の多くの学術出版者である学会、もしくはそれと密接に連動している印刷会社の多くは、結果として既存の冊子出版の体制を保持した

まま、後付で電子ジャーナルデータを作成する手法をとった。DTPやCTPの発達によって、テキストPDFが作成できるようになっていれば、そのデータを抽出してタグ付けすることは可能であり、既存のインフラをなるべく修正することなく、電子ジャーナルを製作するには確かに効率のよい方法であった。また、後にも述べるが、この後付手法で今のところは多くの電子ジャーナルサービスが実現できる。しかし、この手法は、主に二つの大きな欠点を持っている。

その一つは出版者版の印刷前公開が難しいことである。電子ジャーナルの利点を生かし、論文単位で情報が固定されれば冊子の号集約作業や印刷工程を待つことなくすぐにWeb上で公開することがワークフローの構造上難しく、少なくとも、欧米のような冊子体発行の1か月以上前の公開はまず不可能である。むしろ、冊子作成後、時には何か月もかけて電子ジャーナルデータを作成して公開している雑誌もあると聞く。多くの研究者は自分の成果をいち早く世の中に広めたいにもかかわらず、その利点を生かしていない。一部、著者最終版の原稿をそのまま早期公開している雑誌もあるが、研究業績として固定させたい場合、研究者が必要としているのは現時点ではあくまで出版者版の本文ファイルであることは分野を問わない現状であり、それをいかに早く作成し、公開できるかが出版側の研究者への訴求ポイントである。

続いては、コストの問題である。後付手法では、あくまで冊子の製作コストはそのままにかかった上に、電子ジャーナルデータ作成のための後付コストがほぼそのまま追加される。したがって、そのコストはどこかで回収せざるを得ないが、残念ながら日本では電子ジャーナルでビジネスを確立しているところは非常に少なく、また、昨今の図書予算の減少や特定パッケージへの予算の集中化も相まって増収は極めて難しい状況にある。

さらに、以上の日本の状況を電子ジャーナルプラットフォームから見ることも重要である。日本の電子出版の多くは英文誌に限れば独自のプラットフォームかJ-STAGEを利用している。独自のプラットフォームでは独自のスタイルの電子ジャーナルデータを作成すればよいが、J-STAGEでは統一された電子ジャーナル用のデータフォーマットが必要であった。本来J-STAGEはJSTが設計したSGMLを利用した電子ジャーナル登載と公開サービスを想定し、SGMLや同等の構造化文書の利用を促していたが、SGMLや同等の高いレベルの構造化文書を用いた出版に対応できたのはごく一部の学会、印刷所であった。そして先の印刷会社の事情を含めて、結果的には多くの日本の出版者がBIB-CITと呼ばれる、テキスト形式の電子ジャーナル用データをこれまでの冊子発行の作業の後に作成することになった(図2)。BIBファイルは書誌事項や全文情報を、CITファイルは引用文献情報を記述したデータである。この形式はいずれも、いわゆるラベル付きのテキストファイルであって、原則冊子用のデータが出来上がっていることを前提に後付でデータを作成することを促す形式でもある。今ではJ-STAGE登載誌の7割以上がこのBIB形式を利用して電子ジャーナルを掲載しており、そのデータ作成を促すツール類が配布されている。また、CIT形式を作成する代わりに、引用文献のテキストの書式から引用リンク生成に必要な情報を自動解析するツールも開発され、一定の成果が出てもある。また、J-STAGEを見る限り、この形式で公開された電子ジャーナルで基本的な公開のサービスとリンクは実現している。しかしながら、この形式では先のような早期公開を行うことができず、また、印刷会社の負担増とコスト増を招いた。さらに、この形式では特殊記号に対する対応が低く、全文htmlを作成するには製作作業をさかのぼってやり直す必要があるなど、拡張発展的なサービスに対応し

5. 書誌事項表示情報ファイル(BIB)の例

次に書誌事項表示情報ファイルの具体的な例を示します。わかりやすいように、項目名は太字にしてあります。

```

<?xml-stylesheet href="http://www.jstage.jp/jstage/jstage.dtd" type="text/xml" />
<start-ver=1>
  <id-journal=sample>
    <no-vol=152>
      <no-issue=1>
        <article-no=1>
          <start-page=1>
            <end-page=7>
              <dt-received=2005/10/25>
                <dt-revised=2005/12/01>
                  <dt-accepted=2006/01/26>
                    <dt-pub-year=2006>
                      <dt-pub=1>
                        <en-article=Original Article>
                          <kn-article=原著論文>
                            <kn-subject=1>
                              <en-title=Importance of Treatise Magazine Internet Public Presentation>
                                <kn-title=学術論文誌インターネット公開の重要性>
                                  <en-subtitle=In the case of J-STAGE Preview -->
                                    <kn-subtitle=J-STAGE Preview の場合 -->
                                      <en-abstract=In order to maintain and develop Japan's science and technology research at an international level, it is important to disseminate outstanding research and development results to the world instantaneously. To that end, it is important to computerize bulletins of academic societies and research papers that are currently appearing on paper by user organizations and release them to the appearance on the Internet. In order to support the information transmission function of user organizations, the "Japan Science and Technology Information Aggregator, Electronic" (J-STAGE), developed by Japan Science and Technology Agency (JST), set up the hardware and software necessary for electronic journal release within JST to provide services 24 hours a day, 7 days a week. By taking advantage of the hardware and software, the user organizations are able to computerize bulletins of academic societies and research papers currently appearing with ease and at low cost. Computerized documents can be accessed from anywhere in the world with this system.
                                      <kn-abstract=日本の科学技術研究を国際的なレベルに確立発展させていくためには、優れた研究開発成果をいち早く世界に向けて発信していくことが重要です。そのためには、現在利用機関で公開されている紙の原稿としての学会誌、論文誌の発行を電子化し、インターネット上で公開していくことが重要です。この独立行政法人科学技術振興機構 (JST) が構築した「科学技術情報発信・流通総合システム (J-STAGE)」では利用機関の情報発信機能を実現するため、電子ジャーナル出版に必要なハードウェア・ソフトウェアを JST 内に用意し、24 時間年中無休で運用をおこなっています。利用機関はそのハードウェア・ソフトウェアを利用して、現在発行中の学会誌、論文誌を容易に、かつ低コストで電子化できます。電子化した論文はこのシステムにより、世界中どこからでもアクセスできるようになります。
                                  <en-copyright=1>
                                    <kn-copyright=1>
                                      <en-auct-name=Hiroko YABUCHI>
                                        <en-auct-addr=YABUCHI>
                                          <en-auct-mail=Hiroko>
                                            <kn-auct-name=山田花子>
                                              <kn-auct-addr=山田>
                                                <kn-auct-mail=花子>
                                                  <auct-affil=JST>
                                                    <en-auct-name=Taro YABUCHI>

```

15

01-2001-STD-EJAM-MAN-021-01

図2 BIBファイルの例

にくい。BIBファイルによってこれまでの印刷発行体制でも後から比較的簡便に電子ジャーナルデータが作成できたのであり、これがJ-STAGEの登載誌数を500誌まで押し上げた一つの要因とも言える。その果たした役割は最大限尊重しながらも、もはや世界的に見ても大規模と言えるようになったJ-STAGEプラットフォームが用いる電子ジャーナル基礎データが少なくとも世界のスタンダードと比較した場合、残念ながら水準が高いとは言えない点を解決する必要があると考える。

4. 日本化学会のSGML出版とTeX-3B2出版

4.1 紆余曲折した日本化学会の電子ジャーナル出版体制

さて、日本化学会では既報通り、欧米に引けをとらない出版体制の開発と維持を目指してさまざまな試みを行った。日本化学会の出版体制の変遷

表1 日本化学会BCSJ誌出版体制の変遷

開始	終了	メタデータ	組版
1993	2000	SGML	TeX組版システムをベースにしたもの
2001	2002	SGML	FrameMaker をベースにしたもの
2003	2008	TeX	3B2をベースにしたもの
2009	-	XML	3B2をベースにしたもの

Frame Maker: Adobe社の組版ソフト

3B2: 旧Advent社の組版ソフト 現在の名称はArbortextのAdvanced Print Publisherとなっている

について、BCSJ誌を中心に述べると、1989年にSGML出版の検討を開始し、1993年から、つまりインターネットが本格的に浸透する前より、実際にSGML出版の運用を開始した。この運用体制は2001年に若干の変更を加えて2002年まで続いた。この時点までの結果を踏まえて、2003年からTeXとAdvent社の3B2（現Arbortext社のAdvanced Print Publisher）を用いた出版体制に移行しBCSJ誌では2008年まで続いた。このTeX-3B2出版では、TeXはあくまで構造化文書としてのタグ（コマンド）を利用した電子ジャーナル用基礎データの利用にとどめ、レイアウト情報を持たせていない点が特徴である²⁾。また、化学会SGML出版が多くのタグ付けを行って将来性、拡張性を確保した結果、出版コストが高くなった点を踏まえて、当時の電子ジャーナルサービスで必要なタグ付けにとどめ、また、タグ付けの労働集約作業は海外出版者と同様に国外にアウトソースした。その結果、組版のレベルを上げながら組版コストを30%程度抑えることに成功した³⁾。また、J-STAGEの利用とアーカイブの登載にあたっては、先に述べたBIBファイルやDiovと呼ばれる形式の作成も経験し、結果的にさまざまなメタデータを取り扱った。

4.2 SGML出版でわかったこと

さて、過去のSGML出版の経験で得られた最も大事なことは、組版データと電子ジャーナル基礎

データの同期を取ることが重要であることだった（図3、4）。この同期がうまく取れないと、誌面の訂正とデータの訂正をそれぞれ修正が発生する度に繰り返すことになり、ミスの誘発ならびにコスト増の要因となった。また、SGMLではその設計（DTDの設計）が自由であったが、独自設計に基づく出版はそのスケールメリットを生かすように多くの雑誌に展開し、また改善し続けられない限り、かえって手間がかかる。また、SGMLのタグ付けが労働集約的であり、結果的にコスト増要因になった²⁾。なお、このような問題はXMLになっても大きく変わってはいない。

4.3 TeX-3B2出版でわかったこと

一方、現行のTeX-3B2システムは現在もChemistry Letters誌製作で非常に安定的に動いており、また先のSGML出版と比較して組版コストを大幅に削減できたことなど事業効率も高い。さらに、J-STAGEで実現しているサービスのほぼすべてに対応できる基礎データを出力することもできる。しかしながら、化学の分野で現在MS-Wordでの提出率が98%を超え、TeXを利用する著者は非常に少ないために、MS-WordからTeXへの変換工程が必要となり、この点で労働集約的な作業を行っている。また、もともと3B2システムはSGMLやXMLを展開す

図3 電子ジャーナル用データ作成の概念と運用図

図4 3B2システム上での版下データとメタデータの同期の様子

© Arbortext Advanced Print Publisher

ることを想定したシステムであった。逆に、TeXはもともと数式を表現することを目的として開発された経緯があり、また、インターネットの本格的発展の前に成熟したこともあって、電子ジャーナルおよびそれに関連するサービスを前提として設計されたものではなく、例えばWeb上での簡便な表現手法を持たない。それでも少ないリソースで多くの表現が可能なTeXの潜在的能力を最大限に生かしたのが現行のシステムとも言え、海外アウトソースを利用して必要なだけのタグ付けを低コストで行って電子ジャーナル公開を行う、現時点で非常に有力な手法である。

4.4 電子ジャーナル製作に伴うCopy Editの変化

これらの出版体制の変化はCopy Editにも一定の変化をもたらした。冊子の出版の時代から変わらず、多くの学術出版では、Copy Editorがその雑誌の

規則に従い各論文のスタイルを統一してきた。具体的には、タイトルにおける単語先頭の大文字の扱い、著者の記述スタイル、引用と引用文献の表記スタイルなどであり、理工系では、単位記述の統一、特殊記号の扱いの指示や化合物命名法の統一などを行い、論文の中身には影響の無い範囲で雑誌としての統一感や整合性をもたせて読みやすくするのがCopy Editorの仕事である。かつては紙の原稿から打ち直していたために、その製作を助けるために時にはイタリックやボールドなどあらゆる書式の変化に対して、通称「赤入れ」と呼ばれる追記やコメントが加えられることもあった。そして、著者からデジタル原稿をもらうようになった今でも、多くの出版者では原稿を打ち出した紙に追記・修正や指示を入れることが多い。また、電子ジャーナルを製作するようになってから、この赤入れにタグ付けに関する指示を入れることが加わり、欧米の大手の学会などでは著者のファイルを直接開いて赤入れ相当分を修正する試みも行わ

図5 Copy Edit（赤入れ編集）された原稿の一例

れている。いずれにせよ、今もCopy Editorがデジタル原稿を効率的に利用して赤入れを行い、また、電子ジャーナル製作に必要な情報を製作用に与えており、この作業が雑誌の一定の高い品質を保証する一方で、その作業のコスト対策と効率化が求められている。図5にその一例を示す。

5. 世界の状況の変化—標準XMLとツールの出現

このように電子ジャーナル製作における状況が変化していく中で、先に述べたSGML出版に関しては欧米も基本的に同様の問題を抱えていた。特にアーカイブとして電子ジャーナル基礎データを保存する際に、出版者ごとに違ったDTDのまま保存することは望ましくなく、業界は標準の共通規格を模索した。この動きを後押ししたのが、National Library of Medicine (NLM) とアンドリュー・W・メロン財団であり、2003年にNLM-DTDとして公開された¹¹⁾。現在は2008年11月にリリースされたバージョン3.0が最新のものとなっている。また、この動きと呼応する形で、2005年より大手出版社が協

力して進めたアーカイブプロジェクトであるPortico¹²⁾が開始し、このNLM-DTDをアーカイブファイルの共通仕様として採用した。すなわち、このDTDに沿ったXMLを作成しておけば、アーカイブや電子ジャーナルプラットフォームへの一定の親和性が保証されることとなり、学術雑誌出版のための事実上のXML標準が整ったと言ってもよい状況となった。さらに、XMLのタグ付けや編集を簡単にするソフトやMS-WordからXMLを出力するさまざまな出版者向けツールも現れた。

6. 新しいXML出版体制

6.1 WordからXMLを出力

以上の学術出版業界の状況下、日本化学会の論文誌では著者の圧倒的な支持を得ているMS-Wordから直接XMLを、できればNLM-DTDに基づくXMLを作成する方法が、なるべく人海戦術を伴わないあるいは、コストのかからない方法で実現できないかを調査した。その結果2004年に開発されたeXtended Stylesというソフトウェアを用いるとWordから

6.2 eXtylesとカスタマイズ

- ・ 余計な改行など不要なデータの削除
- ・ パラグラフ単位でのスタイル付けとタグ付け
- ・ パターンマッチングによるCopy Edit校正と修正

- ・リファレンス記述の解析とタグ付け
- ・PubMed, CrossRefを介したリファレンスリンクの
チェック (図6)
- ・図表引用, リファレンス番号の整合性チェック

表2 パターンマッチングによる校正ルールを記述した表の例

修正日	修正前	検索パターン	修正後	修正パターン	修正したxsbファイルの名前	備考	
2008/3/25	, Japan¥n	¥(¥[¥a¥d¥]¥)(¥w¥)Japan¥n	削除	¥1	Test.xsb	CSJYtest.xsb は新規作成所属の最後が「Japan」で終わっているときは、その「Japan」は削除する。instのスタイルのみに適用	済
	Chikusa,	¥(¥w¥)Chikusa¥(¥w¥)	Chikusa-ku	¥1Chikusa-ku¥2	Test.xsb	instのスタイルのみに適用	済
		¥(¥[¥a¥d¥]¥)(¥[¥]¥)¥n		¥1	Test.xsb	メールアドレスの末尾にピリオドが入っている場合は削除する	済
	coworker(s)	¥(¥[cC¥]¥)oworker¥(s¥)¥?¥)	co-worker	¥1o-worker¥2	TestDictionary.xsb	All	
	Between ... and ...		—		CSJBetweenfrom.xsb	Between ... and ... に関する置換を無効化	済
	r.p.m.	¥(¥(¥w¥)¥?¥)(r¥[¥]p¥[¥]m¥[¥]¥)(¥(¥w¥)¥?¥)	rpm	¥1rpm¥3	Test.xsb	「r.p.m.」を「rpm」に	済
2008/3/26	Fig.		Figure		CSJFigCitations.xsb	Fig. をスペルアウトするように変更	済
	Fig.		Figure		CSJFigLegends.xsb	Fig. をスペルアウトするように変更	済
2008/3/27	boiling point		—		CSJTermsAbbrev.xsb	boiling point → b.p.にする置換を無効化	済
	<p>数字	¥p1¥(¥w¥)	—	One¥1	CSJSoun.xsb	パラグラフの最初がアラビア数字だったとき、“One”のように英語にする置換を無効化	済
	Hydrogen bond interaction	¥(¥[hH¥]¥)ydrogen¥(¥w¥)bond¥(¥w¥)interaction	Hydrogen-bond interaction	¥1ydrogen-bond interaction	TestDictionary.xsb	All	済

作業を軽減し、出版をさらに早くすることが可能となった。このツールの導入にあたっては、日本化学会論文誌の体裁とスタイルとパターンマッチングの校正ルール（表2）をeXtyleに覚えこませる必要があるが、そのルールを設定ファイルに記述するための作業を繰り返し、半年程度かけてカスタマイズを行った。

以上の作業の後には直ちにNLM-DTDに準拠したXMLファイルを書き出すことが可能で、このファイルを組版システムのArbortext Advanced Print Publisher（旧3B2）に展開すると自動組版が開始される。自動組版では足りない修正は後に人の手で行う。ここでいったん著者校正用のPDFを作成し、著者校正を行う。著者校正の内容を反映させ、論文の情報が固定すると、その他の出版掲載の情報と合わせて最終的なXMLが出力される。この修正の際は、誌面の情報とXMLの情報は連動しているため、前述で指摘した修正に関する人為ミスは発生しない点が重要である。

6.3 運用体制と各オペレーションの見直し

また、eXtyleの導入と合わせて、出版ワークフローそのものの見直しも行った。学会、印刷会社の各オペレータが一堂に会してそれぞれの役割を見直し、既存のワークフローにeXtyleを合わせるのではなく、eXtyleでXMLを効率よく作成できるように、また、作成したXMLを使って効率よく版下と電子ジャーナル用データを作成できるような体制とオペレーション手順ならびに基本出版スケジュールを検討した（図7）。

その結果、

- ・ eXtyleの機能を最大限利用するためには、事前に行う前処理とそれをサポートする編集作業での赤入れが必要。
- ・ 数式をXMLではなくTeXのままで保持しておく

効率がよい。

ことがわかった。

特に数式に関してはすでに印刷会社にてTeXで素早く組むノウハウが蓄積されているので、早い段階でMathML等のXMLに変換するのではなく、最終版下完成後にXMLとして書き出すこととした（図8）。

6.4 新しいXML出版体制のメリットとデメリット

以上の準備を経て、新しい体制による出版をBCSJ誌の2009年の1月号発行から開始した。実際の処理の様子と、処理前後の原稿の違いについては電子付録の動画と関連ファイルを参照されたい（<http://joi.jlc.jst.go.jp/JST.JSTAGE/johokanri/51.902>）。以上の運用体制で得られたメリットは、

図7 XML出版各担当関係図

図8 XML作成および整形フロー図

- ・手入力では非常に非効率と言えるXMLタグ付けが容易になりNLM-DTDに基づいた質の高いXMLの作成が可能となった。
- ・初期XML出力までが非常に早く、著者の原稿の内容次第では著者の原稿から数時間以内にXMLを出力することも可能となった。
- ・XMLを有していることでhtmlやxhtmlへの変換が容易になり、全文ファイルやWeb上の目次作成が容易になった。上記と組み合わせてxhtmlでのこれまでよりさらに早い早期公開も可能である。
- ・Wordの利用となるのでTeXと比較すると運用のためのヒューマンリソースの確保がしやすくなった。
- ・Copy Editorの単純校正修正作業が減った(リファレンス表記の姓名順入れ替えなど)。

である。一方、デメリットとして考えられるものは、

- ・著者原稿がWordでない場合は変換する必要がある。
- ・10Sの1ツールに依存するリスクがある。
- ・著者原稿の製作的な視点から見た質にばらつきがあるため、前処理に負担がかかる場合がある。
- ・TeX-3B2システムと比較すると、海外アウトソース分が無いことによる相対的なコスト増の可能性はある。

などである。

7. 今後の課題と展望

以上のメリットとデメリットを考慮しながら、当面BCSJ誌の発行体制の安定化に努めつつ、現在号単位で進めているWeb公開を論文単位あるいは

入稿論文単位で公開する体制に移行することを検討する。そして、XML出版に移行して出版が期待通り早くなるか、赤入れ編集作業の効率化をどう実現するか、あるいはコスト面でTeX-3B2システムと互角になるかどうかを見極めていくことになる。また、Chemistry Letters誌にも展開し、その速報性と品質をさらに高めることを目指す予定である。さらに、可能であれば、化学や化学に限らず他の分野の雑誌などに展開することも時機をみながら検討することも可能である。特に和文誌への発展は欧米の出版者が活動しにくい市場であり展開が期待される。いずれにせよ、本格的なXMLを用いた学術出版が日本で浸透するかどうかを、一つの手法を通じてではあるが見極めることになる。

また、XMLを作る理由は、検索や引用リンクなど既存のサービスを効率よく実現するためだけではなく、そのXMLを使って付加価値の高い新しいサービスを実現することが期待されるからでもある。例えばイギリス化学会のプロジェクトProspect¹³⁾では、論文本文に含まれる化合物を物性データベースなどとリンクさせ、これまでにない電子ジャーナルサービスを実現しており、ひとたびXMLを持ったわれわれの論文誌でも多少の拡張を行うことで同様なサービス提供が可能となる。ただし、これらは当然、サービス提供にかかる費用とユーザーに提供できる価値など期待される効果をコスト回収策とともに勘案して実現すべきものであり、研究者のニーズと日本化学会として使える人的・経済的リソースを正しく把握しながら対応することになる。

8. おわりに

冒頭で述べたように、ドキュメントXMLはデータXMLほど浸透していないのが一般的な現状であ

るが、学術論文誌に限った場合は欧米では手法はかなり確立されたといえる。そのような中で、日本でも同様の質の高い出版体制が構築できた意義は大きいと考えられる。

一方、これまでは審査終了時に事実上原稿を改めてもらい直して出版していたが、電子投稿査読システムの発達により、投稿審査時の情報がそのまま出版に結びつくワークフローが進みつつある。さらに言えば、blogや機関リポジトリの動向によって、学術情報流通の仕組み自体が変わりうる可能性も否定できないが、研究者の昇進と研究費

の配分を決定付けるメディアは依然として学術論文誌であり、このメディアで質の高い出版を行えるインフラを整えることが今しばらく重要であるとする。

謝辞 この原稿を執筆するにあたっては、JST、印刷会社、Inera社他多数の方のご協力をいただきました。また、電子付録用の原稿に関しては東京大学大学院工学系研究科の加藤隆史教授のご協力をいただきました。ここに謝意を表します。

参考文献

- 1) 林和弘. 理工医学系電子ジャーナルの動向—研究情報収集環境と事業の変革—. 科学技術動向. 2007, p. 17-29.
- 2) 林和弘, 門條司. 日本化学会論文誌の状況と電子ジャーナルの運用における考察. 情報の科学と技術. 2002, vol. 52, no. 2, p. 94-99.
- 3) 林和弘, 太田暉人, 小川桂一郎. 日本化学会の電子ジャーナル化: J-STAGE利用の現状と課題. 情報管理. 2003, vol. 46, no. 6, p. 373-382.
- 4) 林和弘, 門條司. 日本化学会での学術情報発信と流通. 情報の科学と技術. 2003, vol. 53, no. 9, p. 441-447.
- 5) 植松利晃. J-STAGEの現状とJ-STAGE2の開発. 情報管理. 2003, vol. 46, no. 8, p. 536-545.
- 6) 林和弘, 太田暉人, 小川桂一郎. 売れる電子ジャーナルをめざして: 日本化学会の取り組み. 情報の科学と技術. 2005, vol. 55, no. 3, p. 145-149.
- 7) 林和弘, 太田暉人, 小川桂一郎. 日本の電子ジャーナル事業の課題と展望: 日本化学会での取り組み. 情報管理. 2005, vol. 48, no. 2, p. 87-94.
- 8) 林和弘, 中谷敏幸, 太田暉人. NLM-DTD XMLを利用した電子ジャーナル出版. 第5回情報プロフェッショナルシンポジウム予稿集, 2008, p. 25-28.
- 9) 第2回 SPARC Japanセミナー 2008「学術出版とXML対応—日本の課題」.
<http://www.nii.ac.jp/sparc/event/2008/20080624.html>, (参照2009-01-10).
- 10) 第3回 SPARC Japan連続セミナー 2007「メタデータPublishingの現在—電子ジャーナル主体の製作・出版に必要なもの」.
<http://www.nii.ac.jp/sparc/event/2007/1102.html>, (参照2009-01-10).
- 11) NLM Journal Archiving and Interchange Tag Suite. <http://dtd.nlm.nih.gov/>, (accessed 2009-01-10).
- 12) 時実象一. 電子ジャーナルの長期保存—LOCKSSとPortico. 情報の科学と技術. 2008, vol. 58, no. 2, p. 84-88.
- 13) Project Prospect. <http://www.rsc.org/Publishing/Journals/ProjectProspect/>, (accessed 2009-01-10).

著者抄録

電子出版では電子ジャーナルサービスを支える、XMLなどに代表されるメタデータを作成することが必須である。日本の主な学術ジャーナル出版では冊子製作後に電子ジャーナル用データを作成することがまだ多く、速報性に欠けるなどいくつかの電子ジャーナルの利点を生かせていない。日本化学会では国際標準的な電子出版を目指してSGML, TeXなどのメタデータを利用したさまざまな手法を経験した。それらのメリット, デメリットと国際状況を考察し, その結果を踏まえて改良し2009年から開始した国際的にも通用する新しいXML出版体制を紹介して運用と課題を考察する。この体制では, eXtyleというツールを利用してMS-Wordから直接NLM-DTD準拠のXMLを作成し, できたXMLを利用して版下を作成後, 著者校正が終わり次第すぐに電子ジャーナル公開が可能になる体制となった。

キーワード

電子出版, 電子ジャーナル, メタデータ, SGML, XML, 3B2 (Advanced Print Publisher), eXtyle, リファレンスリンク, NLM-DTD, コピーエディット

Author Abstract

For e-publishing of scholarly journals, it is necessary to prepare metadata represented by XML, which allows for many electronic functions like searching and linking. In Japan many publishers still prepare metadata after printing production and hence are unable to take advantage of many beneficial aspects such as early web release. The Chemical Society of Japan has explored e-publishing with SGML, TeX, and other markup languages with respect to international compatibility. In this paper, new XML based e-publishing and related frameworks initiated from 2009 are introduced. These frameworks have been improved based on consideration of international trends of e-publishing and the strengths and weaknesses of previous systems. In this new framework, NLM-DTD based XML can be easily produced from MS-Word by using the tool eXtyle and galley proofs are semi-automatically produced with XML. Papers can then be released online directly after author proofreading.

Key words

e-publishing, e-journal, meta-data, SGML, XML, 3B2 (Advanced Print Publisher), eXtyle, reference linking, NLM-DTD, copyedit