



図書メタデータワークフローの合理化

Judy Luther 著



102 West Montgomery Ave. #B
Ardmore, PA 19003

2009年6月30日

ホワイトペーパー

米国情報標準化委員会 (NISO)
および
OCLC (Online Computer Library Center)
に提出

出版者:

NISO
One North Charles Street
Suite 1905
Baltimore, MD 21201

および

OCLC, Inc.
6565 Kilgour Place
Dublin, Ohio 43017-3395

ISBN 978-1-880124-82-6

© 2009 NISO and OCLC, Inc. All rights reserved. For noncommercial purposes only, this publication may be reproduced or transmitted in any form or by any means without prior permission in writing from the publisher, provided it is reproduced accurately, the source of the material is identified, and the copyright status is acknowledged.

"Streamlining Book Metadata Workflow" (ISBN 978-1-880124-82-6) の電子複製または資料の電子的使用の許可は、www.copyright.com にアクセスするか、著作権処理センター (CCC: 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400) に連絡されたい。CCCはライセンス提供および各種利用者登録を行う非営利組織である。

OCLC と WorldCat は、OCLC Online Computer Library Center, Inc の商標/サービスマークである。本文書に現れる製品、サービス、商号は、各々の所有者の商標/サービスマークである。

NISO ホワイトペーパーについて

NISO ホワイトペーパーは、標準策定に影響を及ぼす問題を扱う寄稿または依頼文書である。ホワイトペーパーは前標準化活動だとみなすことができる。NISO ホワイトペーパーでは、正式な標準化作業が開始される前にすでに見られる問題を定義・調査する場合もあれば、標準策定が行われる可能性のある分野を特定したり、想定される検討方法を提示する場合もある。すべてのホワイトペーパーはNISOのウェブサイト (www.niso.org) に掲載されている。

目次

はじめに.....	1
各ステークホルダーの様相	2
出版社	2
メタデータ提供者	3
出版取次.....	4
書店	5
国立図書館	5
地方図書館	6
Google	7
メタデータワークフロー	8
ライフサイクル	8
標準	8
メタデータサイロ	11
品質	12
合理化の機会	14
識別子	14
主題スキーム	15
ベストプラクティス	17
将来の展望	18
リンク付き 頭字語集	20
図	
図 1: メタデータサイロ.....	11
図 2: 図書メタデータの交換図.....	19

はじめに

メタデータは、読者がオンライン上で図書を確認して入手できるようにするものであるが、同時に出版社の供給連鎖(サプライ・チェーン)や図書館が図書の出版前後に行う作業にも多様な機能を発揮する。電子書籍のコンテンツが急速に増え、フォーマットの種類が増加するにつれ、世界中の読者や関係者がコンテンツを発見、販売、利用するために必要なメタデータは幾何級数的に増加している。

逆説的ではあるが、ウェブ上で膨大な量のメタデータに容易にアクセスできると、どんな種類のデータも本の発見に利用できるという一種の信仰となり、メタデータの重要性を低下させることになる。しかし、本の購入判断に使用するには、メタデータが効率的に伝達され、正しく表示されなければならない。そのため、その基礎となるメタデータの構造と交換が出版社の供給連鎖や図書館の円滑な業務にとってより重要なものになる。メタデータの交換が図書の選択や販売、円滑な事業活動に不可欠なものであることに気付くようになるにつれ、これらに関係するステークホルダーがメタデータの交換を支援する標準やシステムの開発や採用に参加するようになってきた。

現在の図書発見環境では、何が見つからないかを測定することは困難であり、また、貧弱または不完全なメタデータやメタデータの不在が、最終的には消費者に影響を与えることになる事業や収集分析判断に及ぼす影響を定量化することは極めて困難である。オンライン上で見られる図書情報に期待することは、出版社の内部システムにあるメタデータが伝達されウェブ上に表示されることであり、特定の読者向けの販促資料用に編集されたメタデータが見られることではない。このようにマーケティングとコミュニケーションを進化させるには、供給連鎖の関係者へのメタデータの的確な流通が必要である。

これらの要因と経済低迷が結びついて、出版社や出版取次、書店、メタデータ提供者、図書館には、「今すぐやれ」、しかも、もっと効率的にやれという異常な圧力がかけられている。しかし、この分野で働いている者はそれが「言うは易し、行うは難し」であることを知っている。このプロセスを改善するためにはシステム全般にわたるメタデータのワークフローを総体的に見る必要がある。NISOとOCLCは、様々なステークホルダーのシステム間でメタデータがシームレスに交換される環境にいうビジョンを共有しており、データやコンテンツの供給連鎖のすべての参加者におけるメタデータ交換に要する経費の削減に重点的に取り組んでいる。

ステークホルダーの間の議論を支援するために、業界を概観する本文書では、各ステークホルダーの役割を定義し、メタデータの利用を分析し、現況を記述する。この業界を代表する30名以上にインタビューを行い、各自が管理するメタデータの性質や量、そのライフサイクルの変化、現状の問題点と課題、改善のためのアイデアについて話を聞いた。インタビューから得られた意見や洞察は本報告書、ならびに複雑な環境を図示し、重複作業の特定や書籍業界供給連鎖における効率性の向上を行う機会を提供する「図書メタデータ交換図」の基礎となっている。

各ステークホルダーの様相

書籍業界の供給連鎖に関係する各ステークホルダーの様相は、メタデータを作成、高度化、維持、配布、利用する各自の多様な役割を反映している。各ステークホルダーが作成・管理している図書メタデータの量は大きく異なる。1年に1タイトルしか作成しない小規模な出版社もあれば、2万タイトルを作成する大規模な出版社もある。メタデータ提供者や書店は1,400万レコード近くを管理している。米国議会図書館は世界最大の図書館で3,200万タイトルの図書を所蔵している。WorldCatはOCLCの参加図書館が所蔵する1億3,800万件の図書目録レコードを持ち、10秒毎に新しいレコードが追加されている。

図書館の冊子体コレクション、パブリックドメインのタイトル、出版社の既刊書は急速にデジタル化されている。これは、供給連鎖に現れるタイトル数が増加するだけではない。電子書籍や様々なメディアフォーマットには各々異なるISBNが必要なため、タイトル毎に複数のレコードが必要になる。そのため、メタデータ提供者や書店が管理しなければならないレコードの量は現状の8倍から10倍へと大幅に増加する。

メタデータがこれだけの量になると、購入者が自分の望むフォーマットを選択できるように、タイトルだけでなく、それを読んだり聞いたりするためのオプションについても明確に提示することがきわめて重要になる。増大するメタデータ量に対する不安は様々なステークホルダーの間で繰り返し現れるテーマであった。

出版社

現在のプリント・オン・デマンド技術は棚卸しや保管場所に経費をかけることなく大量のタイトルを提供可能にするので、出版社は既刊書のデジタル化を行っている。デジタルファイルの保管にかかる経費は小さく、注文時の複製も容易なので、もはや「在庫切れ」や「絶版」という状態はありえない。既刊書を変換する際に、章レベルのアクセスを提供したり、将来の収入機会を準備するために、タイトルレベル以下の要素にメタデータを与え始めた出版社も存在する。

約20社の超大規模出版社が毎年何千冊ものタイトルを発行する一方で、何百社かある中規模出版社は毎年12タイトルから数100タイトルを発行している。何千社も存在する小規模出版社はロングテールを形成し、年間数タイトルを発行するにすぎない。ONIX for Booksは、先導的な出版社が大量の記述・管理データを構造化して、供給連鎖の参加者にXML形式で提供するのに使用される。しかし、中小出版社におけるメタデータの編集方法は、Excelファイルや独自のフォーマット、印刷カタログが一般的である。これら無数の、一貫性のないフォーマットへの対応という課題は、多種多様なメタデータ作成者により管理されるメタデータの難しさを助長している。

大量のファイルを処理している出版社は、ここ10年間でONIXによるメタデータの処理に大きな進歩が見られたことを認めている。ONIXは柔軟だが複雑なフォーマットであり、大規模出版社にとってはやりがいがあるかもしれないが、中小出版社にとってはやる気をなくさせる可能性がある。転送可能なフォーマットによるメタデータの提供は組織の上層部により支持されなければならないが、ほとんどの出版社はまだ、購入者や供給連鎖の参加者が自社のコンテンツを発見、販売、利用する際のメタデータの影響力を正当に評価していない。出版社はONIXによるメタデータの配布管理を、AnkoやNetRead、Booksonix、Firebrandといった出版ソフトウェアベンダーやウェブベースのサービス提供ベンダーに任せられる。

ONIX 標準の使用により利用可能なデータ量が増加し、その配布が容易になったが、ONIX をフォーマットとして使用してもメタデータのタイムリーさや正確性は保証されないことを出版社もベンダーも認めている。米国の BISG (Book Industry Study Group)、英国の BIC (Book Industry Communication)、Booknet Canada は、ONIX ファイルを評価して、出版社がデータの品質とタイムリーさにおいて基準を満たしていることを認定する出版社認証プログラムを提供している。残念ながら、このプログラムに参加する出版社は限られている。その理由の1つは、既存の業務システムやコンテンツ管理システムの更新には多額の基盤経費が必要であり、そのため、新しい標準や標準の変更に対応できる出版社が限られるのかもしれない

メタデータ提供者

様々な組織がメタデータの収集、高度化、再配布を行っている。これらの機関には、登録認定機関 (Bowker や Nielsen Book など)、目録サービス機関 (英国の BDS など)、会員制組織 (OCLC や CrossRef など) がある。国立図書館や出版取次もこれらの機能を果たしているが、それが本務ではない。

商用のメタデータ提供者は供給連鎖の中心に位置する。あらゆる情報源から様々なフォーマット (ONIX、Excel、印刷物) のデータを収集して、自社システムに取り込み、特定の標準に基づいてデータの追加を行い、MARC や ONIX in XML、その他のフォーマットのレコードとして配布する。出版社にはレコードの収集と配信を行い、図書館にはデータを修正・追加して MARC レコードを作成し、書店にはしっかりとしたレコードを供給する。相互運用性の必要性を正確に理解しており、標準の策定や適用において指導者的な役割を演じている。たとえば、Nielsen BookData は、タイムリーで正確な情報に与えられる英国 BIC の Product Data Excellence Award (商品データ優秀賞) に認定申請した出版社の審査プロセスの不可欠な要素となっている。

Bowker が毎年追加している約30万件のレコードのうち、50%は ONIX で、45%は Excel などの電子フォーマットで受け取っており、紙による提出もまだほんの少しだけ残っている。

Bowker も Nielsen Book もシリーズタイトルを標準化し、典拠コントロール機能を果たし、出版社レコードへのリンクを提供している。両社は、第1章の内容やメディア紹介文、表紙画像、目次、書評の引用、ベストセラー情報、著者紹介、受賞歴、対象読者層、読者の評価、格付けや推薦などの追加コンテンツも提供している。

Bowker は米国とオーストラリアの、Nielsen Book は英国とアイルランドの ISBN 機関の役割を果たしている。この立場で両社は出版社に出版社コードを与えている。2008年に Bowker、Nielsen Book、CISAC、IFFRO は、「テキスト作品」を識別して同一作品の様々なフォーマットを一つにまとめるシステムの普及、調整、指導を行う ISTC (国際標準テキストコード) 機関を設立した。

英国の商用目録サービス機関である BDS (Bibliographic Data Service) は、英国図書館が CIP (Cataloging in Publication) プログラムを外部委託している業者である。BDS は毎年約75,000件の CIP レコードを作成し、ONIX データから MARC 21へのクロスワークを作成している。同社の社員はほとんどが社内教育を受けており、新入社員は1年かけてデューイ十進分類法、議会図書館分類法、議会図書館著者名典拠 (NACO)、議会図書館件名標目 (LCSH) を学習している。

図書メタデータワークフローの合理化

OCLC は112カ国の図書館が参加する会員制の組織である。WorldCat データベースは1億3,800万件のレコードを持ち、これは米国議会図書館と英国図書館の全所蔵を合わせたものに匹敵する。Vendor Record Contribution Program (業者によるレコード貢献プログラム)を通じて、およそ6万件から12万件のレコードの提供を Brodart、Ingram、Baker & Taylor などから受けている。OCLC は、米国とカナダで70名を越すメタデータと目録の専門家を雇用して、特殊コレクションに含まれる様々なフォーマットや言語の資料のレコード、新規受入資料のレコード、出版社やメタデータ業者向けのレコードを作成している。

OCLC は、Google を通じてアクセスできる同社の世界規模の OPAC である WorldCat において、関連する資料のレコードを同時に表示するために FRBR (書誌レコードの機能要件) で示された原則を使用している。OCLC の NextGen Pilot には、ONIX と MARC 間のクロスワークを開発する出版社との共同作業が含まれ、高度化した ONIX データを出版社や供給業者に提供して各社のシステムを強化するサービスを開始している。

図書メタデータは今や、しっかりとした雑誌メタデータファイルを持つ2つの組織における小さいながらも成長を続ける要素になっている。CrossRef は出版社による会員制組織であり、160万件の圖書の DOI (デジタルオブジェクト識別子) を持っている。Serials Solution の Knowledge Base は、100万件以上の電子書籍レコードを持っている。これらのサービスは元来雑誌文献を支援するために発展したものであり、引用文献からフルテキストへのリンクを提供している。

出版取次

最大手の出版取次 Baker & Taylor と Ingram は、昨年10%以上の成長があったデータベースを持っている。毎年の新刊はおよそ20万タイトルであると推定されるが、デジタル形式で利用できる図書やメディアの増加により、出版取次が扱う新規レコードは約70万件にもなる。レコードは様々なフォーマットや版のために作成されるため、各タイトルは平均で2-3レコードを持つことになる。新刊タイトルのレコードは3回から7回更新されるが、そのほとんどは出版前に行われるようである。

ONIX や Excel でデータを提供する出版社は新規レコードの95%を占める。取次業者は CIP をベースにレコードの高度化を行うが、出版社のデータを優先する。印刷された図書を受け取ると、すべてのレコードの正確性、スタイル、一貫性を検査する。

徐々に、図書館は業者が MARC フォーマットの書誌レコードを付けて図書を納品することを期待するようになっており、最近では資料があらかじめ処理され「すぐに配架できる」ことが要求されることも多い。図書館予算の逼迫は、受入処理を合理化したいという欲求とあいまって、この傾向に拍車をかけている。

米国の学術市場で商売をしている Casalini、Harrassowitz、Pupil などの国際企業は、PCC (共同目録プログラム) に参加して自国で出版される圖書の高品質な MARC レコードを作成するために社員を教育している。MARC レコードの作成、特に外国語のタイトルの作成を各図書館で行うのは大変であるので、言語の専門家を擁する業者がこれらの圖書の MARC レコードを配布することを図書館は期待している。

書店

Barnes & Noble は、記述メタデータが販売を増加させ、良いメタデータがないと販売経費がかさむことを

図書メタデータワークフローの合理化

知っている。毎年、何千冊ものタイトルの注文がキャンセルされ、その総計は何百万ドルにもなるが、それは、メタデータが「在庫切れで納入日未定」だと示しているか、取次リストに載っていないからである。記述メタデータの使用と正確でタイムリーな在庫情報の更新を促進するために、Barnes & Noble は出版社に44のデータ要素の提供を求めた。これは、BISG (Book Industry Study Group) の認証で要求される30のコア要素より14も多い。これに応じた出版社はそのメタデータの品質を認定した"A"リスト、または"B"リストに掲載される。これらのリストは、書籍販売を増加させる、より良いメタデータを促進するために業界内で共有されている。

販売という観点でいえば、書店は図書の価格の変更や在庫状況に関する正確でタイムリーなデータを持たなければならない。ONIX は広範な書誌データを提供するのに有効だと考えられており、粒度や定義としては最高のものである。一方、EDI (Electronic Data Interchange) は軽量な標準伝送フォーマットであり、価格や在庫などの変化しやすい要素の更新を効率的に行うことができる。出版取次は記述メタデータの提供は不得意であるが、価格や在庫はすばやく、時には1日に2度や1時間毎にでも、提供することができる。Nielsen Book が昨年1年に処理した4,300万件のレコードのうち、86%が価格と在庫の更新であった。

電子書籍販売業者は図書館提供サービスの一環として MARC レコードを使用している。すなわち、選書や発注用に MARC レコードへのアクセスを提供し、購入後は電子書籍へのアクセス権と共に図書館に MARC レコードを納品している。電子書籍のメタデータは、特定タイトルの識別、引用のエクスポート、図書館システムのためのタイトルリストの作成、図書館における希望図書の受入支援、再販業者のタイトル発注など多様な応用に使用されている。

Amazon などの大規模ウェブ書店は、数多くの様々なソースからメタデータを受け入れているので、標準やベストプラクティスが大きな影響を与えるには幅広く採用されていなければならない。北米の大多数の図書には ONIX レコードが作成されているが、Amazon で販売されているオーディオやビデオフォーマットの図書には適用されていない。

Wal-Mart や Target などの大規模小売店では、全販売量に対する書籍の占める割合はわずかであり、ONIX は使用されていない。供給連鎖の世界標準規格を調整している GS1 と書籍業界の代表との間で事前協議が開始されている。

国立図書館

国立図書館は、データ交換に関する標準策定の合意を得る際に自国において重要な役割を果たしている。議会図書館 (LC) は国立図書館として指定されておらず、資金も提供されていないが、事実上、米国の国立図書館としての機能を果たしている。2008年に LC は、共同、分散、国際、ウェブベースというメタデータ作成のビジョンを提示した『書誌コントロールの未来に関する作業部会報告』を公開した。R2 Consulting が管理している次のフェーズでは、北米における MARC レコード市場の分析を LC に提供する報告書が公開される予定である。

LC は、分散・共同作業による高品質な MARC レコードの費用対効果の高い作成を支援するガイドラインの策定や教育の提供において国際的に主導的な役割を果たしている。PCC (共同目録プログラム) の参加機関により高度化されたレコードは、オーソライズされた名称や件名標目の使用により高く評価されて

図書メタデータワークフローの合理化

いる。PCCは、今後はMARCなどの単一標準の使用が廃れ、書籍業界や権利管理団体、図書館・情報部門で使用・再利用されている多様なメタデータ標準の相互運用に向かうことを認識している。メタデータは徐々に人ではなく機械の間で交換されるようになっており、これは手作業による配布に比べてエラーを減少させ、効率を高めるものである。

英国図書館(BL)もLCもCIPデータプログラムを管理している。このプログラムは、図書館には発注や作業用として刊行前の出版物データをMARCフォーマットで提供し、出版取次業者には出版より通常3ヶ月から6ヶ月前に新タイトルを知らせるといふ、二重の役割を果たしている。CIPレコードは各図書の表題紙の裏に印刷されるが、このメタデータはMARCレコードとして電子的に配布されている。

著作権法は国により異なるが、BLもLCも自国で出版された著作の「法定納本」を受け入れている。BLは英国内にデポジットされたデジタル資料を自発的枠組みの中で受け入れる最適な方法を検討している。現在、BLは多数の出版社が作成したデジタル資料を入手しているが、XMLフォーマットの電子ジャーナル論文のフルテキストはNLMのDTDに変換し、PDF、Word、ePubフォーマットの図書はオリジナルフォーマットのまま残している。XMLフォーマットを使用することにより、出版社は資料からコード化された要素を抽出してONIXレコードに変換できる可能性がある。BLはフルテキストXMLやテキストベースのPDFファイルからメタデータを抽出する実験も行っている。その目的は、最終的にはMETSという「ラッパー」と共に記述メタデータのフォーマットとしてMODSを使用することである。

BLは受け入れたすべての資料について、小説は簡略レコードで迅速に処理し、学術的著作は完全なレコードを作成するなど、資料の適切な目録レベルに基づいて様々なワークフローを採用している。レコードがLCから提供されることが予想される場合、BLはそのレコードを使用するか、データ追加のベースとなる簡略レコードを探すことになる。昨年、LCで処理された35万タイトルの80%、BLで処理された26万タイトルの55%は、目録専門家により作成またはデータの追加が行われた。OCLCのWorldCatレコードの65%は簡略レコードであり、著者やシリーズ名について典拠作業が必要であり、注記や要約、目次、主題標目の追加が必要だと推定されている。

地方図書館

地方レベルでは、多くの場合パッケージ化されてコレクションとして販売されている電子ジャーナルや図書を図書館は受け入れている。そして、徐々に、このセット販売のコンテンツと共にMARCレコードを提供することを出版社や書店に期待するようになってきている。しかし実際には、図書館はこれらのレコードを発注の段階で必要としている。受入に使用している統合図書館システムがMARCレコードフォーマットを必要とするからである。仮レコードは後で個々の資料やコレクションを利用者に公開する際に、まとめて置き換えたり、個別にデータを追加したりすることが可能である。

図書館における作業は、個々のタイトルの目録を作成することから、他の情報源から入手したレコードを管理して必要に応じてデータを追加することに移行している。その結果、目録担当者が処理する資料の割合は大幅に低下しており、大規模研究図書館でも年間に受け入れる10万タイトルのうち目録を作成するのは30%以下であろう。ただし、外部情報源から受け取ったレコードを各館で定義した目録法にあわせるために多大な努力が依然として払われている。

多くの図書館は作業を合理化し、費用対効果を高めるために作業工程の見直しを行っている。このような

図書メタデータワークフローの合理化

図書館では受入管理に1人、目録管理に1人の図書館員とOJTで仕事を覚えた専門職補佐職員で、購入した、または共有の基本レコードの修正・追加をしていると思われる。同時に、今後5年間に退職する目録担当者に代わる目録担当者がほとんどいないことも最近の風潮である。

過去10年にわたる研究でオンライン目録のMARCレコードが電子書籍の利用に直接影響を与えることが示されているが、Google経由でコンテンツにアクセスできるフルテキストデータベースに件名標目を持たせることに疑問を持つ図書館員が存在する。電子書籍の利用パターンが電子ジャーナルと同じであれば、読者の大多数は、オンライン目録やデータベースのトップページを経由せず、検索結果のリンクからコンテンツに直接アクセスすることになるからである。

2009年に発表された英国のRIN(Research Information Network)のためにKen Chadが行った研究では、冊子体および電子媒体の図書と雑誌のメタデータの流れを調査し、ネットワーク化された環境の使用を減らし、より広く共有されるメタデータを支援するよう提案した。

Google

何百万冊もの図書のデジタル化や支払処理を行うBook Rights Registryの作成により、Googleも図書メタデータに関するおなじみの問題に直面することになった。この場合のメタデータのユースケースは著作を特定し、著作間の関係を確立することである。現在のところ、シリーズと多巻物の著作に課題が見られるが、取り組みは開始されている。

デジタル化されている図書の多くはMARCレコードを持っているが、1970年代にISBNが開発される以前に出版されたのでISBNは持っていない。新しい本については、Google Book Searchは可能な限り良いメタデータを収集するためにONIXとMARCの両レコードを取り込んでいる。そして、貧弱なONIXより高品質なMARCを、低品質なMARCよりリッチなONIXを優先している。ONIXはデータ要素が欠けている場合があり、一方でMARCはデータは存在するかもしれないが、データの理解という意味で機械にやさしくない。

メタデータ問題に対処するためにGoogleでは多くの図書館員が働いている。また、GoogleはOCLCとも共同で作業を行っている。Googleは標準を支援し、新たな標準が成長しやがて採用されるまでフォローしている。Googleは関連の著作を区別する課題を解決する可能性のあるアルゴリズムの開発も行っている。

メタデータワークフロー

ライフサイクル

メタデータの流れは出版社に始まり、購入者または読者で終わる。その間、メタデータは供給連鎖に関係する多様なステークホルダーの間で多くの目的を果たす。図書メタデータの交換図(19ページの図2)は主要なステークホルダー、メタデータの流れ、高品質なメタデータを促す外部品質管理が行われる地点(たとえば、BISG の認証プログラム、Barnes & Noble の格付けシステム、PCC 研修など)を示している。

刊行前の出版データは流動的である。ONIX レコードは通常、出版の3ヶ月から6ヶ月前に公開される。この最初の公開から出版までの間にタイトル、サブタイトル、著作のサイズなどが変更されることがある。図書が出版されると、書誌データは安定する。出版取次や図書館は、図書を受け取るとすぐにすでに入手していた ONIX レコードや MARC レコードに誤りがないかを検証するためにメタデータのチェックを行い、レコードの更新やデータの追加を行う。メタデータの流れはこの地点で自然と遅くなり、その後の変更は主に価格と在庫状況について行われることになる。このように変更が何回も行われることから、メタデータ提供者や出版取次、書店は各レコードを平均で3回から5回修正する必要がある(昨年 Nielsen Book が変更した4,320万件のレコードのうち86%は価格と在庫状況の変更であった)。

CIP データは出版前に作成されるので、出版された現物に合わせてページ数を更新するだけでなく、記述情報の変更も行う必要がある。出版後に、著者の没年や新たな件名標目を追加して MARC レコードを更新する図書館もある。

参加機関の所蔵データを WorldCat を通じて一般に公開し、ILL サービスを提供している OCLC から多くの図書館は MARC レコードを入手している。MARC レコード付きのデジタルコレクションの販売により、目録作成サービスは事実上図書館から出版社へと上流に押しやられている。

標準

長い時間をかけて、出版供給連鎖と図書館コミュニティは各自の目的を果たす標準やベストプラクティスを開発してきた。表1は各コミュニティが同一の問題に対処するために使用している、または策定中の標準を示している。この表では2つのデータ要素、すなわち、識別目的で著作を記述する書誌データと各ステークホルダーにとって重要な業務に重点をおく管理データを区別している。MARC は書誌メタデータを重視しており、ONIX はより新しい標準として、より多くの書誌メタデータに加えて、供給連鎖の関係者に必要とされる管理データも含んでいる。

ISBN 標準は両コミュニティで図書の識別に使用されているが、図書の内容を記述する主題については図書館と出版社は非常に異なるスキームを採用している。たとえば、図書館は分類や検索に関連して、詳細で階層的な件名標目(LCSH は30万の用語を持つ)を採用しているが、出版社は店舗展示用に設計されたはるかに軽量なツール(BISAC は3,000用語しかない)を使用している。これらは各々が意図した具体的な応用に基づいて開発されたものである。

ISNI(国際標準名称識別子)に関する最近の作業は共同プロジェクトの1例である。ISNI は登録機関で管理されている各機関独自のデータを、VIAF(仮想的国際典拠ファイル)に参加する各国の国立図書館で

図書メタデータワークフローの合理化

管理されている図書館典拠データにリンクするブリッジ識別子であると考えられる。OCLC は複数の実体をリンクするリッチなメタデータアルゴリズムを使用するための方法論を示す概念実証システムを開発している。

表1: 同一の問題に対処するために2つのコミュニティで開発された異なる標準

出版コミュニティ			図書館コミュニティ	
標準	用途	データ種別	用途	標準
ISBN	特定の出版物	書誌データ	フォーマット	ISBN
ONIX	販売情報/発見		利用目的の発見	MARC
BISAC、BIC	主題	書誌データ	主題	LCSH、SACO、Wilson
ISNI	著者/著作権使用料	書誌データ	著者/貢献者	NACO、VIAF
ISTC	関連著作	書誌データ	関連著作	FRBR
ONIX、EDI	価格	管理データ	助成コード	統合図書館システム(ILS)
ONIX、EDI	在庫状況	管理データ	所蔵	ILS または OCLC

ONIX

ONIX は、図書の販促に使用されるマーケティング情報を含むリッチな商品メタデータを出版社が出版取次や書店に提供するために1990年代末に策定された。ONIX は、構造的には詳細なコード値リストを持つスキーマであり、メッセージ構造として XML を使用している。ONIX は各機関の内部システムに適応できる十分な柔軟性を持っている。これは ONIX を実装するシステムが各出版社用にインポートプログラムをカスタマイズできることを意味し、これによりフィールド化されたデータに関する質問を40%減少させることに成功している。

出版社における ONIX データの典型的なワークフローは次の通りである。

- ⑩ 基本的なメタデータは編集処理で作成され、Excel ファイル、出版社データベース、ホスト型ソリューションのいずれかに格納される。出版管理に使用しているシステムの副産物として作成される場合もある。
- ⑩ マーケティングデータと製作データが追加され、著作のフルテキストを保管するコンテンツ資産データベースと関連付けられる。
- ⑩ 校正刷りが LC に送られ、表題紙の左頁または裏に印刷される CIP レコードの作成に使用される。
- ⑩ 販売情報と目録データが、必要に応じて様々なフォーマットで出版取次やメタデータ提供者に送られる。印刷本には ONIX フィードが作成され、電子書籍では Excel ファイルが使用されることが多く、XML 形式のメタデータが抄録索引サービスに送られる。
- ⑩ 大容量の画像ファイルは別に送られ、ONIX レコードからリンクされる。
- ⑩ 出版後に出た書評が図書データに関連付けられ、メタデータ提供者や出版取次、書店に配布されている ONIX フィードの価格と在庫状況が更新される。
- ⑩ 自社システムから MARC を生成できる出版社は少ないので、ほとんどの出版社はメタデータ提供者やコンサルタントと提携して、コンテンツと共に配布する MARC レコードを作成している。

最近公開された ONIX 3.0は、過去の版との後方互換性はないが、デジタル商品の取り扱いの改善、レコードの部分的更新の許容、同一著作の様々な商品を関連付ける ISTC の使用、新たなマーケティング情報の規定、シリーズの取り扱いの改善など、いくつかの重要な機能強化を行っている。出版社と業者は移行期間の間2つのバージョンを維持する必要がある、これは現在の経済情勢では難題になるかもしれない。ただし、後方互換性を維持するにも経費はかかる。

MARC

1970年代に開発された MARC(機械可読目録)は、タグ付けされたフィールドを持つデータ交換フォーマットとして設計され、その後、一群の標準、規則、指針へと発展している。目録の作成は、図書館コレクションの記述、および読者が資料を発見できる確実なアクセスポイントの作成という2つの役割を果たす。現在では MARC21が各国の独自 MARC へのクロスウォークを持つ国際的な標準となっている。

国立図書館レベル

出版社が国立図書館(議会図書館や英国図書館に委託されたBDS)に校正刷りを送り、出版される著作に含めるCIP(Cataloging in Publication)レコードを速やかに受け取るというように、MARCレコードは出版物のライフサイクルの早い段階で作成される。CIPプログラムは図書館が受け入れると思われる図書にオーソライズされた標目(著者、タイトル、主題)と分類標数(デューイまたはLC)を提供するために創設された。適用範囲外として指定されている除外カテゴリ資料の基準はLCとBLで異なり、適用範囲の図書のすべてが提出されるわけでもない(LCの除外リストは <http://cip.loc.gov/scope.html> を、BLの除外リストは <http://www.bl.uk/bibliographic/exclude.html> を参照されたい)。BLとLCが作成するCIPレコードの量はおよそ5万件で、これは両図書館で作成される目録の20%以下である。

LCにおけるCIPのワークフローは次の通りで、平日10日以内に完了する。

- ⑩ 前付け、著作権表示、表題紙、索引、第1章または最終章を含む著作の一部の電子コピーがLCに送付される。
- ⑩ 最大手の出版社10社は55,000件のCIPタイトルのほとんどについてONIXレコードを送付しているが、ONIXデータはCIPの作成には使用されない。
- ⑩ LCはONIXから要約と目次を抽出し、MARCからリンクする。
- ⑩ 著者や貢献者を一意に同定する典拠作業が行われる。およそ全タイトルの1/3で新規典拠レコードが作成される。適切な件名標目が付与される。
- ⑩ ほとんど完成したMARCレコードが作成される。
- ⑩ 製作中の図書に含められるようにCIPデータが出版社にメールで送付されるが、XML形式では提供されず、出版社のONIXデータストリームに投入されることもない。(ONIXデータからMARCへの変換処理で生じる問題を解決するパイロットプロジェクトが2009年にLCで行われた。)
- ⑩ 目録配布サービス(Cataloging Distribution Service)を通じてMARCレコードがメタデータ提供者、書店、出版取次に配布される。
- ⑩ 出版されるとすぐに、CIPとして作成されたMARCレコードは更新される。

英国図書館はCIPの作成を、ONIXからMARCへのクロスウォークを作成しているスコットランドのBibliographic Data Services(BDS)に外部委託している。LCはCIPレコードをタイムリーに作成するた

図書メタデータワークフローの合理化

めに国立農学図書館(NAL)、国立医学図書館(NLM)、そして複数の大学図書館の支援を受けている。

出版社が CIP を入手するインセンティブは、出版前に書籍業界に対して告知を行うことである。図書館は、近刊書の注文に使用できる MARC レコードを持つことにより恩恵を受ける。国立図書館もこれらのレコードを使って、著作権法で規定されている法定納本がされていないタイトルを識別することができる。

地方図書館レベル

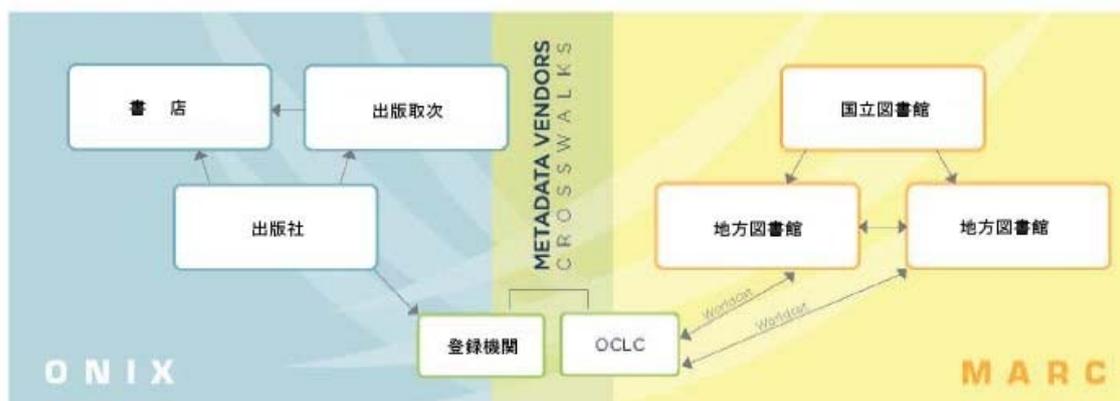
コンテンツを受け入れる図書館における MARC データのワークフローは次の通りである。

- ⑩ 発注処理を始めるために、図書館は出版社や出版取次から受け取った簡略レコード、LC の CIP、OCLC レコードを入力する。
- ⑩ 図書が届いたら、前項のレコードを更新して、所蔵データを追加する。
- ⑩ 出版社が作成した付加的な記述データをメタデータ提供者から定期購読している図書館もある。
- ⑩ 利用者が作成する主題タグを収集する新しいオンライン目録を使用している図書館もある。

ONIX は選書をオンラインで行う場合に必要となる付加的な記述情報(目次や著者紹介)を含んでいる場合が多い。Bowker や Nielsen Book、BDS などのメタデータ提供者を通じて利用できる、このリッチな記述データを入手している図書館が増加している。

メタデータサイロ

ONIX と MARC の間でクロスウォークの作成が必要だということは、これまでいかに両コミュニティが独立して各自の問題に取り組んだかを示している。しかし、オンライン環境では相互運用が必要であり、メタデータを管理するもっと効果的かつ効率的なアプローチの可能性について、図書館と出版供給連鎖関係者の双方が参加する全体的な視点が提案されている。図1は各ステークホルダーのメタデータサイロを示し



ている。

図 1: メタデータサイロ

品質

メタデータの品質を測る最良のものさしは、購入者や読者が図書を必ず発見できるようにするというステークホルダーの目標をどのくらい効果的に支援できるかである。この目標を達成するために必要なデータ要素は、記述メタデータ(著者紹介、表紙画像、図書要約、目次)が選書作業をオンラインで行う際に中心的役割を演ずるオンライン環境に対応するために拡張し続けている。

メタデータ交換処理の品質は供給連鎖の他のステークホルダーへのメタデータの配布とその有用性により決定される。GDSN(Global Data Synchronization Network)は高品質のメタデータの特性を次のように定義している。

- ⑩ (レコード間およびファイル間に)一貫性があり、
- ⑩ コア要素がすべて揃っており、
- ⑩ 合意された標準に準拠しており、
- ⑩ タイムリーに配布されている。

GDSNは、メタデータの唯一の「所有者」として更新データで配布されるデータをコントロールして、すべての参加者のシステムが同時に一貫したデータを持てるようにすることにより、需要と供給の連鎖を国際的に支援している。「データの品質を向上させることにより、取引相手の経費を削減し、生産性を高め、売買の速度を加速する。高品質のデータはCコマースやグローバルなデータ同期化の基礎である。」

大手出版社は処理に週末一杯かかるような何千件ものレコードや更新データを含むファイルを毎週配布している。各タイトルのメタデータ量の増加により、手作業による検査や処理を最小限に抑えた機械対機械のインターフェースが必要となる。その際に生じる問題は、定義(たとえば、「シリーズ」とは何かの明確化)か実装(シリーズはONIXレコードに含まれているか、どこにあるか)のいずれかになると思われる。

品質の諸要素

データ要素を同定する1つの方法は、それが内在的なもの(著作自体を記述する)であるか、外在的なもの(著作に付随する)であるかを検討することである。

- ⑩ **内在的なデータ要素**は、タイトルや著者、出版社、出版日、目次などの著作を記述する要素と、シリーズや主題、分類、MARCレコードにおける著者の典拠作業など、著作を識別するために使用される要素である。これらの要素は書誌メタデータの基礎を形成する。
- ⑩ **外在的なデータ要素**は、様々なステークホルダーのマーケティングや管理作業を支援するために使用される要素である。出版社や書店にとっては価格や発行日、書評などのデータが、図書館にとっては助成コードや所蔵などのデータがこれにあたる。

デジタル製品がXMLで作成されている場合、内在的なメタデータの多くはコンテンツから抽出することができる。メタデータは出版後の分析により付与または添付されるものではなく、タグ付けされた著作から直接引き抜くことができるものだからである。その結果、メタデータはXML形式のコンテンツの代理品となる。

品質向上プログラム

出版社も図書館も標準やベストプラクティスの採用を促進する手段を開発している。

出版社のプログラム

- ⑩ 米国のBISGは、出版社がXMLまたはExcelで送ることができる230のデータ要素のうち30要素の使用を必須とするPDCP(商品データ認証データプログラム)を提供している。
- ⑩ 英国のBICは、データの品質や一貫性に関する評価を出版社に与える商品データ優秀賞

図書メタデータワークフローの合理化

(Product Data Excellence Awards)を提供している。この賞のガイドラインは各国の様々な書籍販売の商習慣に基づいて変更することができる。

- ⑩ Booknet Canada は、評価用のサンプルを監査員に提供した出版社や業者に金・銀・銅の認証レベルを提供している。
- ⑩ Barnes & Noble は、BISG の30の必須フィールドと、図書の販売に影響を及ぼす表紙画像などのデータを含む16の追加データフィールドに基づいて出版社を格付けしている。同社の目標は上位1,000社が A+の格付けを獲得することである。

図書館のプログラム

- ⑩ 議会図書館が管理する PCC (共同目録プログラム)は、目録担当者向けの研修を提供しており、これを受けた者は、標準(図書では BIBCO、逐次刊行物では CONSER)に従い、著者(NACO)や主題(SACO)について適切な典拠コントロールを行ったレコードの投稿を行う権限が与えられる。
- ⑩ 多くの図書館で使用できる高品質な MARC レコードを提供して、各図書館レベルで重複した作業をしなくて済むようにするために、最近、PCC 研修プログラムが国際的な出版取次業者にも拡張された。これらの業者には、イタリアの Casalini Libri、ドイツの Harrassowitz、アルゼンチンの Garcia Cambeiro、スペインとメキシコの Puvill Libros、中国と韓国の東アジアの業者が含まれる。
- ⑩ ドイツとフランスの国立図書館が参加し、OCLC の支援を受けている LC の最近のプロジェクトが VIAF (仮想国際典拠ファイル)である。これは、個人に関するデータを統合し、関連レコードをリンクし、各国語で名称を表示する、780万件の名称典拠情報源レコードの相互参照ファイルを作成することを目標としている。

これらのプログラムの利用は徐々に広がっており、供給連鎖に関係するすべてのステークホルダーが利用できるメタデータの品質に対してプラスの影響を与えている。供給連鎖の中でメタデータの品質を向上させるために様々なステークホルダーが行っている活動は図書メタデータの交換図(19ページの図2)に示されている。

合理化の機会

本文書の執筆中、OCLCは出版社と図書館員が参加してメタデータのニーズと実践を検討するシンポジウムを開催した。シンポジウムでの議論や本文書を作成するために行ったステークホルダーとのインタビューの中で現れたアイデアや疑問は、メタデータ供給連鎖の状況やその有効性を高め経費を削減する合理化の方法に対する彼らの最近の考えを反映している。現在行われている活動や合理化のためのアイデアは大きく、識別子、主題スキーム、ベストプラクティスという3つのトピックにまとめられる。

メタデータの経済性を考えると、集中管理された書誌レコードを各コミュニティに公開される分散型の管理レコードで徐々に高度化するというモデルを提案できるだろう。もう1つ考えられるのは、変更のないデータ要素を収めるフィールドと可変データへのリンクを収めるフィールドを含むレコードを持つ方法である。様々なステークホルダーが単一のメタデータソースに依存するモデルには懸念を示していた。データを均質化するより、データを共有する方法を均質化の方が可能性があるだろう。

出版社が徐々に供給連鎖におけるしっかりとしたタイムリーで正確なメタデータの重要性を認識し始めたように、図書館も次第に個別にメタデータを作成する経費が無駄であることに気付くようになり、出版社や書店がメタデータをMARCレコードで提供することを期待するようになった。デジタル作品は生まれつきメタデータの作成をサポートしている。システムを全体的に見れば、その後のメタデータの高度化処理を合理化するのに役立つだろう。

識別子

識別子は、アイテムのあいまいさの除去と関連アイテムの一括提示という2つの目的で使用される。すなわち、識別子は、著者や著作を同じものとして扱うか、別のものとして扱うかを決定する。

著者

ISNI(国際標準名称識別子)は著者識別子の候補としてほとんどの場合で言及されているが、国立図書館の典拠ファイルなど、様々なステークホルダーが持つ既存のファイルを利用するものである。2008年に新たに発行された著作の1/3について図書館は典拠レコードを作成したが、これは、これらの著書が新たな著者により作成されたことを示している。生没年を典拠レコードに使用することは著者の識別に役立つと思われるが、書店が使用する様々なシステムで表示が統一されない恐れがあるため、ONIXレコードに生没年を含めることに出版社は懸念を示している。

著者を正しく識別する様々な方法が供給連鎖の各所で開発されており、データ要素に対する要件は出版社やステークホルダーで異なる。RRO(Reproduction Rights Organization)や小説本を発行する出版社の中には、複数のペンネームを持つ著者の印税や著作権使用料を効率的に管理するために別名を管理できるシステムを必要とするものがある。GoogleのBook Rights Registryは支払を正確に行うために著者の明確な識別を求めている。科学出版社は世界中に存在する著書を識別する必要があるため、また、敬称を保持するフィールドを必要としている。著者識別子システムが現れたら、貢献者、特にデジタル化されている非常に多くの古い著作の貢献者の情報を整理する標準が必要になるだろう。

個々の著作

様々なフォーマットの電子書籍へ付与されたことで ISBN の数が急増すると、すべてのフォーマットの電子書籍に同一の ISBN を与えたいという反対意見が一部の出版社から現れるようになった。タイトルレベル以下の図書の内容へのアクセスを提供する様々なモデルを出版社が実験した際、章レベル以下の内容の識別には Actionable ISBN や DOI が使用された。

ISBN は確立された標準であるが、現在の応用に関していくつか警告がなされている。古い著作がデジタル化される際には新しい ISBN が与えられるので、図書館は既に所蔵するタイトルとの重複を避けるよう注意しなければならない。提出されたタイトルが市場の需要調査用に使用された場合など、ISBN が付与されたタイトルの中には実際には出版されなかったものもある。販売予定のないデジタルスキャンプロジェクトで作成されたアイテムに ISBN を与えることが適切なのかという疑問も提起されている。

書店や大規模小売店で実際に扱われるタイトルは EAN または UPC で識別される。どちらも、供給連鎖において在庫の移動を追跡するのに使用される GS1 承認の GTIN(Global Trade Item Number) であり、バーコードの形で表示されている。

シリーズ

シリーズは問題があると言われることが多い。なぜなら、出版業界におけるシリーズの名称はマーケティングの都合で決められるものであり、セットやシリーズを定義する図書館の構造的な定義に沿って決められるものではないからである。シリーズタイトルが図書のタイトルより重要であったり、出版社が個々のタイトルのグループ化を希望したために、シリーズタイトルをタイトルフィールドに格納したり、サブタイトルにしたりする場合がある。また、旧システムを使用しているためタイトルの設定方法に制限がある出版社もあるかもしれない。これは、多くの場合、シリーズデータ用のフィールドが不正確または不十分であることを意味する。ONIX バージョン3.0はシリーズの問題に対処しており、出版社と読者が直接つながるオンライン環境では、シリーズはより普遍的な視点に移行できることが期待される。

関連著作

ISTC(国際標準テキストコード)は同一著作の様々な表現形を関連付け、利用者が著作のすべての表現形を発見できるように設計されている。ISTC 機関は今では公開されたスキーマを持つ法人組織である。2009年に終了予定のパイロットプロジェクトでは、既刊書リストのリンクグを作成して発見を最適化し、ISTC が現行の書誌メタデータに組み込まれるようにするために出版社とのワークフローを開発している。採用が拡大すれば、おそらく OCLC や各国の国立図書館も既存の著作に遡及的に ISTC を付与するようになると思われる。

主題スキーム

出版社は店内展示用に図書を組織化する目的で主題スキームを使用しているが、図書館ではオンライン目録におけるアクセスポイントとして主題スキームを使用している。オンライン販売が急速に増加するにつれ、もっと詳細な主題用語が役に立つようになるに違いないと考える出版社が出てきた。たとえば、主題を単なる「料理」ではなく「フランス料理」としたタイトルは、どちらで検索しても結果集合に現れるからである。

図書メタデータワークフローの合理化

出版社は、対象読者コードや対象読者の年齢範囲など MARC レコードにはない ONIX 独自のフィールドを使用している。刊行される図書がコード化される際に矛盾すると思われる複数のカテゴリに入れられる場合がある。たとえば、大人向けと子供向け、あるいは小説と歴史というように、矛盾する2つのタグが付けられた図書である。これは、異なる市場区分の双方に商品を置くために出版社が意図的に行った可能性がある。このデータを図書館で使用する際にはデータを作成した動機を理解する必要があるだろう。

出版社の主題スキーム — BISAC と BIC

BISAC は、3,000以上のカテゴリを持つ比較的変更の少ないリストであるが、上位レベルの50のカテゴリは出版されるタイトルや利用者ニーズの分析に基づいて毎年再検討されて改訂されている。英語で書かれた図書向けに設計されており、米国の BISAC コードと英国の BIC コードは改訂が揃うように調整されている。OCLC は BISAC 件名標目からデューイ十進分類法 (DDC) への対応付けを開始している。

これらのコードは元来店内展示用であったので、オンライン検索ができるようになるまで商業出版社の間で広く採用されることはなかった。しかし、今では、主要な多くの供給連鎖関係者が BISAC コードや BIC コードを各タイトルに適用することを必要としている。これらのコードは、ウェブサーチの設計に組み込まれ、カテゴリによる POS 統計の追跡メカニズムに不可欠なものになり、出版取次が作成する書店や図書館市場向けの選書リストに使用され、今では複数の BI (Business Intelligence) アプリケーション製品にも統合されるなど、商品の店内展示をはるかに超えた利用がなされている。

図書館の主題スキーム — LCSH, Sears, MeSH

図書館では主題用語は図書の発見にとって重要だと考えられているので、著者などの主たるアクセスポイントと同レベルの典拠コントロールが行われている。議会図書館件名標目 (LCSH) には30万以上の用語があり、これは3,000 件の BISAC の100倍のサイズである。(Baker & Taylor の一部である) YBP のような出版取次や ebrary のような電子書籍業者は図書館と共に選書目的のプロファイルを作成したり、オンラインブラウジングを支援するために LCSH や請求記号を使用している。

主題スキームに関するアイデア

- ⑩ 米国では BISAC や BIC、LCSH など様々な主題スキームが使用されており、ドイツの図書館システムではまた違うコードが使用されている。どうしたら、この重複部分を最大化して無駄な繰り返しをなくすることができるか。クロスウォークは実現可能なのか。各コミュニティが各々のコードを持つことには価値がある。どうしたら、その利用を広げ、その価値を高めることができるか。
- ⑩ どうしたら、利用者が作成したメタタグを移植可能にして、Primo などの今あるオンライン目録システムから読者がフィードバックを行うことができる未来のオンラインパブリック目録に移行することができるか。
- ⑩ デジタルコンテンツではキーワード検索が広く利用されているが、サーチャーには今でもキーワードより統制リストの方が利用されているのか。この疑問についてさらに詳しく調べる研究が必要である。
- ⑩ ONIX からの追加データが検索処理の絞込みに利用できたら図書館で使用されるようになるだろうか。

ベストプラクティス

処理を改善するアイデアは、より大きな問題の特定から実現可能なアイデアの提案にまで及ぶ。ベストプラクティスの開発やその幅広い採用には通常数年の時間が必要なので、技術革新により対応できる変更を検討する際には業界動向の変化を判断することが役に立つ。

図書メタデータの使用が広がり、それを配布する ONIX が開発されたことにより、出版社の内部管理システムでの作業の開始から図書供給連鎖の参加機関へと続くメタデータの流れがもたらされた。さらに ONIX と MARC 間のクロスウォークを使うことにより図書館が ONIX データにアクセスすることが増え、図書館と出版社双方にとって有用になるとと思われるデータが交換されるようになれば、両社の間で大局的見地からの対話が促進される。

図書の発注という初期の段階で MARC レコードを入手できれば、図書館は後で時間とリソースがある時にデータの追加・修正ができる基本的な書誌レコードを持つことになる。出版社と書店は、出版社から書店のシステムへの更新データの迅速な伝達をデータセンターやメタデータアグリゲータが支援するためのベストプラクティスを検討している。

ベストプラクティスに関するアイデア

- ⑩ CIP の作成を容易にし、MARC データの XML フィードを出版社に提供するために ONIX と MARC 間のクロスウォークを使用する。
- ⑩ データベースのアクセス権限が拡大されれば、出版後の CIP レコードの追加・修正作業を OCLC の会員図書館が共同で行うことができるだろう。
- ⑩ 各種標準に幅広く準拠し、図書館の重複作業を削減するために、MARC レコードに関する研修をより多くの国際企業に拡大する。
- ⑩ 提供者が信頼できる情報源であることが確定されたら、手作業の点検を省略できる属性が公開されていれば OCLC レコードの点検作業は自動化できるだろう。
- ⑩ 目録担当者が追加する一意なデータ要素(オーソライズされた名称、件名、シリーズ、分類標数)をシステムに組み込むことが出版社にとって価値があるか否かを調査する。
- ⑩ 電子的供給の一環として FTP で受け取る大きなデータファイルには、ファイルを開けなくても何が含まれているかを知ることができるように、コンテンツを特定する「マニフェスト」が必要である。
- ⑩ シリーズや目次といった記述メタデータを統合することにより、コレクションレベルからタイトルレベル、論文レベル、画像レベルへとナビゲーションできる、より洗練された検索結果表示をサポートすることができるだろう。
- ⑩ 章レベルでコンテンツをタグ付けする方法を検討している科学出版社では、商品レベル以下の高品質なメタデータを作成・維持するためのガイドラインと社内専門家が必要である。
- ⑩ ロングテールにある出版社には、一貫した方法でデータを提供する方法の選択肢としてベストプラクティスや簡単な助言が役に立つだろう。
- ⑩ 経済情勢を考えると入力メカニズムの簡素化が必要である。ベストプラクティスは品質管理目標を定義することだろう。
- ⑩ 図書館が出版社や書店から受け入れる個々のタイトルのパッケージを表現するコレクション識別子に対するニーズが存在する。
- ⑩ NISO のデジタルライブラリおよびデジタルコレクションに関する思想的指導者会議(NISO

図書メタデータワークフローの合理化

Thought Leader Meeting on Digital Libraries & Digital Collections)は、出版社が自らのメタデータの品質を自己診断できるツールの開発を勧告している。

⑩ 供給連鎖を(出版社からコミュニティまたは業者から図書館だけでなく)より多角的にするために、交換、更新頻度、フィードバック機構、再利用に関するベストプラクティスを確立する。

⑩ 最近公開された ICTC(国際標準テキストコード)やまもなく公開予定の ISNI(国際標準名称識別子)を既存のワークフローに組み込み、その採用を促進するための調査を行う。ICTC は著作間の関連付けの作成に、ISNI は著者の典拠コントロールに利用できるだろう。

将来の展望

現在の経済情勢は、複数のフォーマットで入手可能なコンテンツの急速な増加を管理する必要性とあいまって、メタデータの使用法とその関連経費を評価するようステークホルダーを駆り立てている。成果を向上させ、効率を高める共同作業は、コンテンツを発見、販売、利用するためのメタデータの流れを合理化することができる。

コミュニティにとっての課題の1つは、供給連鎖においていかに価値が高められるかを認識するために、ステークホルダーの間でメタデータの流れに関する大局的な視点を構築することである。シンポジウムのモデルに従えば、すべてのステークホルダーが会話できる機会を広げる一連の会議は、幅広いコミュニティが一堂に会してメタデータを見るためのレンズの役目と他のコミュニティが提供するメタデータを各コミュニティがいかに活用できるかを調べるための場を提供する。

近い将来、異なるシステム間のクロスワークによりメタデータをステークホルダー間で共有することが可能になるだろう。標準間の変換を可能にする長期的なベストプラクティスはこの業界のすべてのステークホルダーの役に立つだろう。

図書メタデータワークフローの合理化

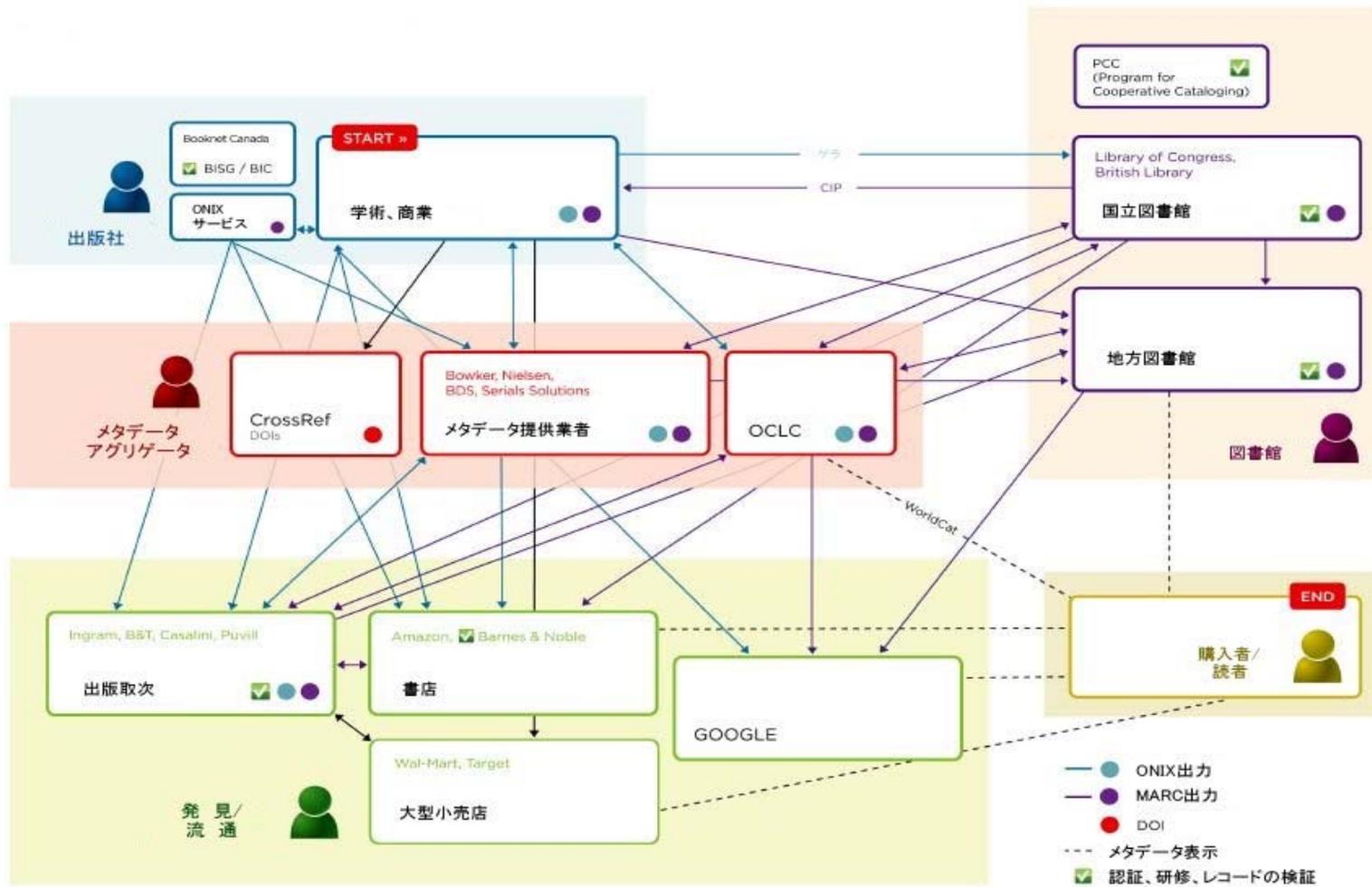


図 2: 図書メタデータの交換図

リンク付き頭字語集

BIBCO — Monographic Bibliographic Record Program (図書書誌レコードプログラム)

<http://www.loc.gov/catdir/pcc/bibco/>

PCCを見よ。

BIC — Book Industry Communication

<http://www.bic.org.uk/>

BICは、電子商取引ならびに標準的なプロセスと手順の活用を通じて書籍関連のあらゆる業種における供給連鎖の効率性の向上を目的に、出版協会、書店協会、図書館・情報専門家協会、英国図書館の後援により設立された英国の独立機関である。

BISAC — Book Industry Standards & Communication

<http://www.bisg.org/bisac/subjectcodes/index.html>

BISAC 主題コード委員会は、件名標目リストとマーチャンダイジング主題リストを維持・管理しており、出版社や書店などがこのリストを実装・利用するための手引きを提供している。BISAC メタデータ委員会は ONIX について、識別子委員会は ISBN-13 について作業を行っている。

BISG — Book Industry Study Group (書籍産業研究グループ)

<http://www.bisg.org>

BISG は政策、標準、研究のための米国の書籍産業事業者団体である。会員は、出版社、印刷・製本業者、サプライヤー、出版取次、書店、図書館員、その他印刷や電子メディア事業に携わる者から成る。書籍コミュニティの向上を図る目的で、30年以上にわたって、すべての業種の専門家が集まり、効率的に課題や懸案事項を解決するフォーラムを提供している。

Booknet Canada

<http://www.booknetcanada.ca/mambo/index.php>

Booknet Canada はカナダにおける書籍供給連鎖の革新をめざす非営利機関である。監査プロセスを使用する独自の認証プログラムを持っている。

CIP — 議会図書館の Cataloging in Publication

<http://cip.loc.gov/>

Cataloging in Publication レコード (別名、CIP データ) は、まだ出版されていない図書のために議会図書館が作成する書誌レコードである。出版社は図書を出版する際に CIP データを著作権表示ページに含めることができ、これを使うことで、図書館や書籍販売業者は図書の処理が容易になる。

CDS — Cataloging Distribution Service (目録情報配布サービス)

<http://www.loc.gov/cds/index.html>

CDS は、MARC レコードや目録作成ツールという形で、図書館や書籍提供業者に書誌製品や書誌サービスを提供する。

CISAC — International Confederation of Societies of Authors and Composers (著作権協会国際連合)

<http://www.cisac.org>

CISAC は1926年に設立され、著作権団体の世界的ネットワークと連携することにより、創作者の権利の認知と保護の向上に向け努力している。

CONSER — Cooperative Online Serials (逐次刊行物オンライン共同プログラム)

<http://www.loc.gov/acq/conser/>

PCC を見よ。

CrossRef

<http://www.crossref.org/>

CrossRef は出版社による会員制の団体で、学術・専門出版物向けの公式な DOI[®]リンク登録機関としての機能を果たしている。CrossRef は出版社を横断する引用リンクシステムを運用している。このシステムを使うと、研究者はある出版社のシステムに掲載された引用文献をクリックすることにより、別の出版社のシステムにあるその文献の内容に(アクセス権があれば)直接アクセスすることができる。

DDC — Dewey Decimal Classification (デューイ十進分類法)

<http://www.oclc.org/dewey/>

メルヴィル・デューイが1870年代に考案し、1988年から OCLC が所有している DDC は、図書館コレクションを編成するための動的な構造を提供するもので、世界で最も広く使用されている図書館分類システムである。

DCMI — Dublin Core Metadata Initiative (ダブリンコア・メタデータ・イニシアティブ)

<http://dublincore.org/dcmirdataskgroup/>

メタデータ用語のコアセットの策定と管理は、情報の発見、共有、管理を容易にするシンプルな標準を提供する DCMI の起源と歴史の一部である。

EAN — GS1の European Article Number (欧州品目コード)

<http://www.gs1.org/ecom/overview.html>

EAN は、完全な ONIX レコードにリンクしており、バーコードで表現される。

EDI — Electronic Data Interchange (電子データ交換)

<http://www.x12.org/>

EDI は、合意されたメッセージ規格を使って、人手を介すことなくコンピュータ間で構造化データを交換することである。EDI は世界中の大部分の電子商取引で使用されているデータフォーマットである。米国の EDI 規格は、ANSI 認可の標準開発機関である ASC (Accredited Standards Committee) X12により策定されている。

EDItEUR

<http://www.editeur.org/>

EDItEUR は、17カ国の90機関が会員となり、書籍および逐次刊行物業界における電子商取引の標準基盤を開発している。中でも、EDI、書誌情報、関連規格、デジタル出版、RFID タグ、権利管理に重点をおいている。EDItEUR は、国際 ISBN 機関の管理サービスも提供している。

FRBR — Functional Requirements of Bibliographic Records (IFLA の書誌レコードの機能要件)

<http://www.ifla.org/VII/s13/frbr/>

FRBR は、IFLA (国際図書館協会連盟) により開発された概念的な実体-関連モデルであり、任意の実体

または関連の発見、識別、選択、入手という利用者タスクを定義している。

GDSN — Global Data Synchronization Network(グローバルデータ同期化ネットワーク)

<http://www.gs1.org/productssolutions/gdsn>

GDSN は、GS1グローバル・レジストリ[®]、GDSN 認定データプール、GS1データ品質フレームワーク、GS1グローバル商品分類により構成され、これらが一体となって正確なデータを安全かつ継続的に同期化する強力な環境を提供する。

GS1 — The global language of business

<http://www.gs1.org/>

GS1は世界中のあらゆる産業部門にまたがる需要・供給連鎖の効率性と視認性を向上させるために各種標準とソリューションの設計と実装を行うグローバル組織である。GS1システムが取り組む領域には Barcodes、eCom、GDSN、RFID の4つがある。GS1は、ヘルス・ケア、防衛、輸送と物流において活発なプログラムを持っている。

IFRRO — International Federation of Reproduction Rights Organizations(世界複製権機構)

<http://www.ifrro.org>

IFRRO は、創作者や出版社の独自の活動を補完する目的で、各国の RRO を通じて効率的な権利の共同管理を促進することにより、テキストおよび画像ベースの著作権を有する作品の合法的な使用を増加させ、不正な複製を排除するための活動を行っている。

IFLA — International Federation of Library Associations and Institutions(国際図書館協会連盟)

<http://www.ifla.org/>

IFLA は1927年に設立された、図書館と情報の専門家にグローバルなフォーラムを提供する団体と個人から成る組織である。

ISBN — International Standard Book Number(ISO の国際標準図書番号)

<http://www.isbn-international.org>

ISBN(ISO 2108)は、特定の出版社により出版または生産された単行形式の出版物の各製品形態または版に与えられる一意な国際的識別システムである。ISO 2108は ISBN の構成法、ISBN の付与・使用規則、ISBN 割当てに関連するメタデータ、ISBN システムの管理を規定している。

ISNI — International Standard Name Identifier(ISO の国際標準名称識別子)

<http://www.isni.org/>

ISNI 標準(ISO/DIS 27729)案は、図書、テレビ番組、新聞記事などのメディアコンテンツの貢献者の公的アイデンティティを一意に識別する方法である。ISNI は、他の人と混同する可能性のある名称の曖昧性を除去するツールを提供し、メディア業界のあらゆる部門で収集・使用されている名称に関するデータを相互に関連付ける。

ISTC — International Standard Text Code(ISO の国際標準テキストコード)

<http://www.istc-international.org/>

ISTC 標準(ISO 21047)は、テキスト作品のためのグローバルな識別システムを定義する。主に出版社、書誌サービス、書店、図書館、権利管理機関での使用を想定している。各 ISTC は中央登録システムが

テキスト作品に付与した一意な「番号」であり、異なる出版社やフォーマットで出版されている同一コンテンツの識別に使用される。

LCSH — Library of Congress Subject Headings (議会図書館件名標目表)

<http://www.loc.gov/cds/lcsh.html>

「赤本」として知られている LCSH は30万件の項目を持つもっとも包括的な件名標目リストを提供する。LCSH は世界中の図書館で使用されている。

Library of Congress Working Group on the Future of Bibliographic Control (書誌コントロールの将来に関する議会図書館作業部会)

<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

この賛否両論のあった先進的な報告書は書誌レコードをより効率的に作成するために共同作業を勧告している。

MARC — MACHine Readable Cataloging (機械可読目録)

http://en.wikipedia.org/wiki/MARC_standards または <http://www.loc.gov/marc/umb/>

MARC レコードは、標準的な書誌データの交換を支援する議会図書館の活動を通じて1960年代に開発され、1) 著作の記述、2) 検証済みの著者名、3) 標準ベースの件名標目、4) 分類標数 (DDC と LC 分類) から成る。MARC レコードは、レコード構造、内容指定、レコードのデータ内容の3要素を含む。1997年、MARC の米国・カナダ版は現在の版である「MARC 21」になった。

MeSH — Medical Subject Headings (医学件名標目表)

<http://www.nlm.nih.gov/mesh/>

MeSH は、国立医学図書館の統制語彙シソーラスである。定義語はアルファベット順と階層構造の2種類で配列されている。

METS — Metadata Encoding and Transmission Standard

<http://www.loc.gov/standards/mets/>

METS スキーマは、デジタルライブラリのオブジェクトに関する記述、管理、構造の各メタデータを符号化し、W3C の XML スキーマ言語で表すための標準である。

MODS — Metadata Object Description Schema

<http://www.loc.gov/standards/mods/>

MODS は、書誌要素セットのためのスキーマで、様々な目的で使用できるが、特に図書館アプリケーション用に設計されている。

NACO — Name Authority Cooperative Program (名称典拠共同プログラム)

<http://www.loc.gov/catdir/pcc/naco/naco.html>

PCC を見よ。

NISO — National Information Standards Organization (米国情報標準化機構)

<http://www.niso.org>

NISO は出版社、図書館、システムプロバイダを集めて、情報供給連鎖のあらゆる分野に影響を及ぼす自発的合意基準を策定している。その範囲は、紙の耐久性や性能評価、検索機能性から、識別子、メタデータフォーマット、保存にまで及ぶ。NISO は ANSI から情報と文書化に関する ISO 技術委員会46の米国技術諮問グループ (TAG) の管理者に指名されている。NISO は ISO 技術委員会46識別と記述に関する

る小委員会9の事務局にもなっている。

OCLC — Online Computer Library Center

www.oclc.org

OCLC は112カ国の71,000図書館による会員制の組織であり、世界中の情報へのアクセスを促進し、図書館資料の検索、受入、目録、貸出、保存に要する図書館経費の上昇を軽減することを目指している。OCLC の WorldCat データベースは、世界最大の図書館資料とサービスのネットワークに存在する書誌レコードに対するオンラインアクセスを提供している。

ONIX for Books — EDItEUR の Online Information Exchange for Books

<http://www.bisg.org/documents/onix.html>

ONIX は、出版社が取次、電子書店、一般書店、他の出版社、その他書籍販売に関係する者に自社発行の書籍に関する電子情報を配布する際に使用できる標準フォーマットである。(注: ONIX という名前は、購読制の商品を収める ONIX for Serials のように、書籍以外の標準にも使用されている。しかし、本報告書で使用されている場合、ONIX は ONIX for Books を表している。)

PCC — Program for Cooperative Cataloging (議会図書館と PCC により調整された共同目録プログラム)

<http://www.loc.gov/catdir/pcc/>

PCC は、1992年に開始された、世界中の図書館が互いに合意した標準に準拠した、有用でタイムリーかつ費用対効果の高い目録を提供することにより、図書館コレクションに対するアクセスの拡大をめざす国際的共同作業である。PCC には以下のプログラムがある。

- ⑩ NACO — 名称典拠共同プログラム
- ⑩ SACO — 主題典拠共同プログラム
- ⑩ BIBCO — 図書書誌レコードプログラム
- ⑩ CONSER — 逐次刊行物オンライン共同プログラム

PCN — 議会図書館の Preassigned Control Number Program

<http://pcn.loc.gov/>

議会図書館は注文できるように、目録カードに一意的な識別番号を与えている。目録カードは図書が出版される前に作成されることもある基本的な目録レコードである。この番号は LC 請求記号として見なされている。

PDCP — Product Data Certification Program (BISG の商品データ認証プログラム)

http://www.bisg.org/documents/certification_productdata.html

PDCP は出版社(とその他一部の組織)が BISG に商品情報のファイルを提出して、明確に定義された基準に基づいて客観的な評価を受けることができる自発的なプログラムである。

RDA — Resource Description and Access (RDA 策定合同運営委員会による資料の記述とアクセス)

<http://www.rda-jsc.org/>

RDA は AACR2により確立された基礎の上に構築され、あらゆる種類のコンテンツとメディアを対象とする資料の記述とアクセスに関する包括的な指針と指示を提供することを目的としている。まだ策定中であるが、2009年11月末に公開される予定である。

RIN — Research Information Network(研究情報ネットワーク)

<http://www.rin.ac.uk/>

RIN は英国の4つの高等教育助成機関、3つの国立図書館、7つの研究会議により構築された。RIN は研究者の情報ニーズの理解と推進に重点を置いている。

SACO — Subject Authority Cooperative Program(主題典拠共同プログラム)

<http://www.loc.gov/catdir/pcc/saco/>

PCC を見よ。

UKSG — United Kingdom Serials Group(英国逐次刊行物グループ)

<http://www.uksg.org>

図書館、出版社、出版取次、技術提供者で構成される会員制組織であり、冊子体出版と電子出版ならびに学術コミュニケーションに関する意見交換に重点を置いている。

VIAF — Virtual International Authority Files(仮想国際典拠ファイル)

<http://www.oclc.org/research/projects/viaf/>

この共同プロジェクトは当初、米国議会図書館、ドイツ国立図書館、フランス国立図書館で開始され、その後、他の国の国立図書館が参加して拡大した。VIAFは、3つの機関すべての名称典拠ファイルを仮想的に結合して、各国の言語で著者の名前を示す、単一の名称典拠サービスを構築する方法を研究している。プロトタイプがOCLCで開発され、<http://viaf.org/> で公開されている。

XML — Extensible Markup Language(W3C の拡張可能なマーク付け言語)

<http://www.w3.org/XML/>

XML は、SGML(ISO 8879)由来のシンプルで非常に柔軟性の高いテキストフォーマットである。XML は元来、大規模電子出版の課題に対応するために設計されたが、ウェブ上などに存在する幅広い種類のデータの交換においても次第に重要な役割を果たすようになってきている。