

# ONIX から MARC へのマッピング

キャロル・ジーン・ガッドビー

OCLC 研究員



OCLC Research 出版物

Mapping ONIX to MARC  
Carol Jean Godby, for OCLC Research

This report is copyright ©2010 OCLC Online Computer Library Center, Inc.  
All rights reserved  
April 2010

OCLC Research  
Dublin, Ohio 43017 USA  
www.oclc.org

ISBN: 1-55653-380-2 (978-1-55653-380-8)  
OCLC (WorldCat): 588972377

関連するクロスウォークがクリエイティブ・コモンズ Attribution 3.0 (CC-BY) ライセンス  
<http://creativecommons.org/licenses/by/3.0/> で公開されている。

連絡は直接下記へお願いします:

Carol Jean Godby  
Research Scientist  
godby@oclc.org

この文書の引用例:

Godby, Carol Jean. 2010. Mapping ONIX to MARC. Report and crosswalk produced by OCLC Research. Available online at <http://www.oclc.org/research/publications/library/2010/2010-14.pdf> (report) and <http://www.oclc.org/research/publications/library/2010/2010-14a.xls> (crosswalk).

## 目次

|                             |    |
|-----------------------------|----|
| 要旨                          | 6  |
| 謝辞                          | 7  |
| はじめに                        | 8  |
| 1.0. 図書館と出版社のためのメタデータ       | 8  |
| 1.1. ONIX 2.1 と MARC 21     | 8  |
| 1.2. ONIX と MARC の往復変換      | 11 |
| 1.3. ONIX から MARC へのクロスウォーク | 13 |
| 1.4. OCLC の実装               | 13 |
| 2.0. マップの詳細な調査              | 14 |
| 2.1. 識別子                    | 15 |
| 2.2. 補足テキスト                 | 16 |
| 2.3. 主題                     | 17 |
| 2.4. 状況の確認                  | 19 |
| 2.5. 難問                     | 19 |
| 3.0. 教訓                     | 24 |
| 参考文献                        | 27 |

## 表一覧

以下の表は、<http://www.oclc.org/research/publications/library/2010/2010-14a.xls> で公開されている ONIX-MARC クロスウォークの同名のシートに該当する。

- 表 1. ONIX
- 表 2. ProductForm
- 表 3. ProductContent
- 表 4. ContributorRole
- 表 5. Edition Type
- 表 6. Extent
- 表 7. BISAC SH
- 表 8. Audience Code
- 表 9. Audience Range
- 表 10. Description
- 表 11. Measure
- 表 12. Country Codes
- 表 13. State Provinces

## 図一覧

|                                      |    |
|--------------------------------------|----|
| 図 1.1. ONIX レコードと MARC レコードの対応       | 9  |
| 図 2.1. ONIX と MARC で表現された EAN        | 15 |
| 図 2.2. ONIX と MARC で表現された補足テキスト      | 16 |
| 図 2.3. MARC と ONIX における主題標目          | 17 |
| 図 2.4. BISAC コードと BISAC 件名標目が関係するマップ | 19 |
| 図 2.5 a, b. ONIX と MARC における貢献者の関係   | 20 |
| 図 2.6. ONIX と MARC におけるタイトル          | 22 |
| 図 2.7. ONIX と MARC における形態事項          | 23 |

## 要旨

本文書は、OCLCが開発し、OCLCのウェブサイトとEDItEUR<sup>1</sup>で公開されているONIX 2.1からMARC21へのクロスワークを説明するものである。図書館界と出版業界で広く使用されている2つの標準間で書誌的記述データを交換する方法を記述したクロスワークにおける主要な改善点を示す。実際の出力を検討し、この戦略的に重要な関係の次の抜本的な見直しに向けた理論的、概念的な課題を特定する。

---

<sup>1</sup>EDItEUR. <http://www.editeur.org/>.

本文書で説明するクロスウォークは OCLC の同僚 レニー・レジスターとボブ・ピアソンが作成し、私のプロジェクトチームが実装したものである。ただし、本文書の説明に誤りがあった場合はすべて私の責任である。

## 1.0. 図書館と出版社のためのメタデータ

ONIX (ONline Information eXchange) はコンピュータシステム間でコミュニケーションをとるための一連の国際標準であり、EDItEUR により開発・管理されている。ONIX 標準は、あらゆる種類、あらゆるメディアの出版物の記述、メディアコンテンツの知的財産権、利用の許可と制限、著作とその体现形に対する標準識別子の登録、などを行う幅広いアプリケーションをカバーしている。ONIX for Books は最初に登場し、もっとも広く採用されている。また、ニューヨークを本拠地とする Book Industry Study Group (BISG)、英国の Book Industry Communication (BIC) をはじめとする様々な国の数多くの実装グループにより支持されている。ONIX for Books は出版社と書店の間で電子データの交換を推進するよう設計されている。標準的なレコードには、タイトル、著者、主題、出版者、固有識別子などの要素を持つ書誌的記述が含まれている。他にも、販売・頒布権、価格情報、在庫状況を記述する要素が含まれている。

ONIX データは主に供給連鎖(サプライチェーン)を通じた書籍の移動を容易にするために使用されているが、オンライン書店のウェブサイトでは書籍の内容紹介や書評を読んだり、書店を訪れて主題別に陳列されている書籍や DVD をブラウジングしたりする際に、顧客の目にも触れる機会がある。図書館の来訪者は同様な利便性の多くを MARC (Machine Readable Cataloging) レコードから得ている。MARC は既に41年もの間使用されている古い標準である。当初は目録カードの電子版として作成されたが、現在ではオンライン閲覧用目録(OPAC)の基礎として、その派生版が数多くの国の図書館界で採用されている。

### 1.1. ONIX 2.1 と MARC 21

1988年に出版されたミステリ小説を例に、対応する ONIX レコードと MARC レコードを下の図1.1に示した。ONIX レコードでは書誌的記述に使用される要素のみを示している。MARC レコードと重なるのはこの部分だけだからである。表示を省略したのは、地域販売権、販売制限、在庫状況、電子商取引に関するその他の情報を記録する ONIX 要素である。

図には、ONIX から MARC へのマッピングで得られる詳細な結果を視覚化するために第2章で使用する以下の表記法を採用している。

```
<Product>
<RecordReference>0892962844</RecordReference>
<ProductIdentifier>
  <ProductIDType>03</ProductIDType>
  <IDValue>9780892962846</IDValue>
</ProductIdentifier>
<ProductForm>BB</ProductForm>
<Title>
  <TitleType>01</TitleType>
  <TitleText>McBain's Ladies</TitleText>
</Title>
<Contributor>
  <ContributorRole>A01</ContributorRole>
  <PersonNameInverted>Hunter, Evan</PersonNameInverted>
</Contributor>
<Subject>
  <SubjectSchemeIdentifier>02</SubjectSchemeIdentifier>
  <SubjectHeadingText>
```





図 1.1. ONIX レコードと MARC レコードの対応

- ⑩ 赤色のテキストは、せいぜい若干の変更だけで ONIX から転写できるユニークなデータ値を表している。たとえば、<RecordReference>要素と<ProductIdentifier>要素はそのまま対応する MARC フィールドに転写されており、<TitleText>要素は異なる大文字化規則でフォーマットされて MARC レコードに転写されている。
- ⑩ 青色のテキストは、2つのレコードの単なる構造的な違いを示している。ONIX レコードの構造は ONIX 2.1 の XML スキーマにより規定されている。MARC レコードは複数の構造で表現することができるが、図 1.1 では人が理解しやすい ISO-2709 構文のタブ区切り版を使用している。MARC レコードを左から右に読んでいくと、次の3種類の標識が存在する (LC 2009)。第1は、リーダー (Leader) を除き 001 から 999 までの3桁の数字である タグにより命名されたフィールドである。第2は、2つで1組のインジケータであり、フィールドに関連付けられたデータを説明または操作するコード化された指示を含んでいる。第3は、\$a などの1つ以上のサブフィールドで、データ値を含んでいる。
- ⑩ 緑色のテキストは、そのまま転写されるのではなく、マッピングが行われるコード化データを表している。マッピングが ONIX から MARC コードへの単なる1対1の対応になる場合もある。たとえば、ISBN-13 は ONIX では<ProductIDType>要素の値が値 03 の<ProductIdentifier>要素で表現されるが、MARC においても第1インジケータの値が3の024フィールドで表現される。しかし、このように単純に行かない関係の方が多い。たとえば、ONIX の<ProductForm>要素の値 BB はこのレコードが書籍の記述であるとタグ付けしているが、同じ情報が対応する MARC レコードではリーダー、008フィールド、300フィールドの3つの値の組として伝達されている。これらの関係の詳細は第2章で論じるが、ここでは単に、ONIX レコードと MARC レコードで書誌的記述を行っている部分はおおよそ同じであり、図書館界と出版業界で使用されているアプリケーションでおそらく共有可能であることを示し

ておきたい。

ONIX と MARC には明らかに類似性がある。両標準とも最終的には出版された作品と読者をつなぐ働きをするからである。しかし、両者は同じ標準ではなく、一方が他方から由来したのものでもない。2つの標準は異なるニーズと利用者を対象としているので、構造的にもセマンティクス的にも異なっている。ONIX は出版された商品に関するデータを伝達するための国際標準であり、受け手によりデータがどのように使用、拡張、操作されるかについて何も想定していない。MARC は図書館システムのサポートや図書館界での情報交換に国際的に使用されている、密接に関連した一群の標準である。

MARC レコードは図書館界で開発される様々なアプリケーションの期待に応えなければならないパッケージとして使用されることが意図されているので、独立したデータ転送で必要とされる以上の内容とフォーマットを持っている。第1に、MARC レコードは国際標準書誌記述、すなわち ISBD (IFLA 2007) で定義された区切り記号法に従ってフォーマットされる。図1.1では、245(タイトル)フィールドの最初の単語だけを大文字で始めること、300 \$a フィールドに記入するページ数に略語 "p." を追加することを ISBD は規定している。第2に、MARC レコードは特定の言語的・地理的地域で定義されている目録規則に従う。米国とカナダでは、通常、*英米目録規則第2版*、すなわち AACR2 (JSC-AACR 1988) に従っている。これにより、このレコードを OPAC で使用することが適切だと保証される。MARC レコードではアクセスポイントの作成に特に注意が払われている。アクセスポイントは OPAC の利用者が検索問い合わせを行って書誌レコードを検索できるようにしたものである。図1.1に示した MARC レコードのアクセスポイントは、主たる貢献者を列挙する100フィールド、タイトルを示す245フィールド、主題を示す650フィールドである。ただし、これらのフィールドが ONIX においてこれらの概念を表す <Contributor>、<Title>、<Subject> 要素に比べて複雑であることを示すにはこの例のデータは単純すぎるものである。アクセスポイントは、図書館目録において特別な機能を持っているため、複数の ONIX 要素からマッピングされるデータを含む場合がある。たとえば、245フィールドの記録媒体記述や100フィールドのタイトルである。詳細は次の章で詳しく説明する。

まとめると、ONIX と MARC のマッピングの問題は一見したところ簡単のように思われる。図1.1の赤色で示されたデータを調べて、ONIX の <Subject> 要素が MARC の650フィールドに対応し、ONIX の <Contributor> 要素が MARC の245フィールドに一致することなどを発見することは容易である。ただし、使用する用語と抽象化のレベルに関して明確にしておくことが重要である。そして、この点になると、問題がはるかに複雑になる。第1に、ONIX や MARC という語は、複数の構文表現やバージョンを持つ可能性がある、標準の非公式な呼称である。本文書では、ONIX は ONIX for Books 2.1を表す語であり、MARC は MARC21の ISO-2709表現を意味するものとする。もう1つの問題は、ISBD 区切り記号など、変換元の ONIX にはない特別なフォーマットを変換先の MARC は持つ場合があり、これを変換過程のどこかで導入する必要があることである。したがって、2つの標準の基本的な関係を異なるバージョン、構文、フォーマット規約を通じて保存する、または、まったく新しい変化が導入された際に簡単に改定することができるデータモデルと変換ソフトウェアを設計することが技術的課題となる。

しかし、図書館機能を支援するデータベースにレコードを入力することが目標である場合は、もう1段階必要である。すなわち、必須のデータ要素の有無をチェックしてレコードのセマンティクス上の正しさを検証しなければならない。残念ながら、セマンティクスの検証方法には世界共通の標準が存在しない。OCLC が行っている MARC レコードのセマンティクス検査は、これまでの実践に従い AACR2に基づいているが、これには英国寄りの観点が含まれている。そこで、OCLC は AACR2に代わるものとして提案されている RDA (Resource Description and Access) の進捗状況に注目している。RDA はより国際的な適用範囲を約束しており、MARC 以外の標準との高い相互運用性を可能にする (RDA 2010) からである。さらに、OCLC ではニューヨークを本拠地とする BISG の勧告に基づいて ONIX レコードのセマンティクス検査も行っており、英国の BIC の勧告を取り入れる計画も持っている。それでも、セマンティクス

の検証は我々の変換モデルでもっとも未解決な部分であり、これを改善する標準化コミュニティの専門家からのアドバイスを歓迎する。

ONIX と MARC は構造的にもセマンティクス的にも異なっているが、両者の関係は戦略的に重要である。一例を挙げると、図書館は書店から多くの出版物を購入しており、出版市場の約12%を占めている。出版物を受け入れたら、図書館は添付されている出版界が作成した記述データを図書館データベースに統合しなければならないからである。

## 1.2. ONIX と MARC の往復変換

出版社と図書館のメタデータを統合するという目標を達成するためには、ONIX データを MARC に対応付けなければならない。これを効率的に行うには作業が自動化されるべきである。これが実現すれば、ONIX の提唱者が言う電子データの交換という目標が出版業界の垣根を超えて異なる標準を持つコミュニティにもそのまま拡大することを意味する。ただし、出版社と図書館の間のデータ交換は必ずしも一方とは限らない。出版社や書店が作成したレコードには、タイトル、著者、出版社名、ISBN などの固有識別子に関する信頼できる情報だけでなく、関連著作を探して書店のウェブサイトを訪問する顧客に周辺情報を提供する、書評や概要、目次、サンプルページなども含まれている。一方、図書館のレコードも、分類や詳細な件名標目、著者名から本人を同定する情報源へのリンクなど、独自の価値ある情報を作成している。また、図書館のレコードは、新旧の様々な版やハードカバーとペーパーバックなど、同一の知的著作の様々な表現形を形式的に関連づける書誌記述モデルを反映している場合もある (IFLA 2010)。したがって、図書館と出版者のデータを統合できれば、基幹的な情報を2つのフォーマット間で容易に受け渡すことができ、結果的に両コミュニティが共に有用だと考える高品質なレコードが生まれることになる。したがって、2つの標準の共通基盤を同定し、ONIX と MARC の両レコードを自動的に統合するソフトウェアソリューションを考案することが急務である。

これらのアイデアの一部は、2009年中頃に公開され、現在も開発が続いている OCLC の *Metadata Services for Publishers* (OCLC 2009) プロジェクトで検討されている。このプロジェクトの目標は、出版社や図書館向けの高品質なレコードを作成し、出版社の供給連鎖におけるメタデータ管理の非効率性を解決することである。完全に自動化されたプロセスが、米国の出版社や業者から ONIX レコードを入手し、MARC21 (多くの主要国、特に英語圏の国で採用されている MARC の1バージョン) に変換し、何千万件もの書誌レコードを持ち、同種のコレクションでは世界最大の OCLC WorldCat<sup>®</sup> データベースに存在する類似のレコードから取り出したデータで高度化している (OCLC 2010b)。その結果作成される MARC レコードは図書館に公開されているが、これとは別の経路で、同じレコードが ONIX に逆変換され出版業界に配布されている。

このプロジェクトは、ONIX から MARC21 への堅牢なクロスウォーク、すなわち、セマンティクス上の対応セットを必要とした。このようなクロスウォークは以前にも提案されていたが、*Metadata Services for Publishers* プロジェクトチームは、それらは不完全で、一貫性がなく、時代遅れであると判断した。そのため、プロジェクトの初期の成果物の1つは、議会図書館から公開されていた ONIX から MARC へのクロスウォークと対応するレコード構築アルゴリズム (LC 2000) を大幅に改良したものであった。なお、現在 LC からはもっと包括的なバージョンが公開されている。最新のクロスウォークは、現在のところ、ONIX for Books 2.1 (EDItEUR 2009a) を対象に定義されているが、ONIX 3.0 (EDItEUR 2009b) を対象とした改訂が現在進行中である。OCLC ではこのクロスウォークを *Metadata Services for Publishers* プロジェクトやその他のメタデータ管理プロジェクトで過去6ヶ月にわたって使用している。図書館や出版業界で検証され、できれば改良されるように、このクロスウォークを現在 EDItEUR で公開している。

我々の目標は、ONIX と MARC の両レコードに含まれる書誌的記述を対応付け、次の3つの要件を満たすことができる、商用品質のシステム用の知能を定義することであった。

第1に、システムにより作成される MARC レコードが MARC および AACR2 の検証規則をパスすること。

第2に、システムにより作成される ONIX レコードが EDItEUR が公開している XML スキーマに対して構文上妥当であり、BISG のベストプラクティスガイドライン(BISG 2005)に基づく検証検査をパスすること。このガイドラインでは、識別子、タイトル、作成者、出版社、主題、形態事項など、書誌的記述におけるコア要素が存在することを求めている。最後に、システムにより作成された ONIX と MARC の両レコードが往復の変換、すなわち、ソースからターゲットへ変換し、さらにそれをソースに戻した場合に、できる限り保存されること。確実な往復変換は、OCLC *Metadata Services for Publishers* プロジェクトにおける本格的なビジネスニーズを満たすものである。なぜなら、出版社の供給連鎖から入手した ONIX レコードは、通常 MARC レコードからマッピングしたデータで高度化した上で、ONIX レコードとして提供業者に返す必要があるからである。なお、往復変換は、データ値の脱落や変化を禁止するので、変換処理の整合性をテストすることにもなる。

### 1.3. ONIX から MARC へのクロスウォーク

クロスウォークは、人が読むことができる Microsoft Excel のスプレッドシートで表現されており、ONIX から MARC に変換するソフトウェアの実装に役立てることができる。14枚 訳注のシートで構成されているが、最も重要なシートは ONIX と名づけられたシートである(以下、表1と呼ぶ)。それ以外のシートはいくつかのマッピングを完全に実装する際に必要とされる詳細情報を含んでいる。通常、これらのシートは、ONIX コードを複数の MARC フィールドにマッピングする方法を規定したり、簡潔に述べるできない複雑な条件付のマッピングを詳細に説明するものである。

クロスウォークは単独で使用できるように設計されているので、議会図書館の *MARC 21 Format for Bibliographic Data* (LC 1999) や EDItEUR の ONIX 2.1 文書 (EDItEUR 2009a) などの参考資料に当たらなくても大部分の指示が解釈できるはずである。本文書で説明するほとんどが対象としている表1で、最初の4つのカラムは、マッピングの対象となる ONIX の単独要素または階層的に構成された複合要素の概略図を示している。1行には1つの要素を示し、要素は人が読み取れる ONIX 参照要素名を使って識別している。5番目のカラムは要素のセマンティクスの簡潔な説明である。次のカラムはその ONIX 要素が、単独要素の場合は BISG の勧告 (BISG 2005) に、複合要素の構成要素の場合は ONIX 2.1 スキーマ (EDItEUR) に基づいて、必須であるか、任意であるかを示している。

最後のカラムは対応する MARC 要素であり、ONIX ソースからの変換コードを作成するための一連の指示が添えられている。指示の中には、ONIX の <NotificationType> 要素から MARC のリーダーフィールドのコード化レベル表示サブフィールドへのマッピングに必要な1対1のコード変換のように単純なものもある。また、比較的わかりやすいブール論理を示している指示もある。たとえば、ONIX の <LanguageCode> 要素は、その言語が主たる言語であるか副次的な言語であるかにより、MARC では 008 フィールドのテキストのデフォルト言語サブフィールドか 041 フィールドのいずれかにマッピングされる。しかし、非常に複雑で、セマンティクスがあいまいで、補助的シートを参照しなければならない指示もある。たとえば、ONIX の <ProductForm> 要素は、最大6つの MARC フィールド、15のサブフィールドへのマッピングが考えられる。

### 1.4. OCLC の実装

ONIX から MARC へのクロスウォークは OCLC のクロスウォークウェブサービスにおいて完全に実装されている。その技術的な詳細は他で説明されている (Godby, et al. 2008a and 2008b) が、ここで、変換の定義や様々なバージョンの ONIX と複数のバージョンや構文表現の MARC とを調整する際に発生すると思われるばらつきの管理を容易にしている設計上の1つの特徴を述べておくに役立つだろう。変換モデルの基本概念はマップ(対応付け)である。これはメタデータ標準の専門家が定義したクロスウォークの1行におおよそ対応する。関連する一連のマップは通常 マッピングと呼ばれている。たとえば、識別

訳注: 原文には15枚とあるが、実際のクロスウォークに合わせて修正した。以下、断らずに同様の修正を行っている。

子のマッピングには10桁 ISBN のマップが含まれる。このマップのソース(変換元)は ONIX の (<ProductIDType>要素の値が01 の) <ProductIdentifier>要素であり、ターゲット(変換先)は MARC 020 \$a である。人間用のスプレッドシートで示されているように、マップはソースとターゲットを持ち、条件付きロジックを持つ場合もある。しかし、マップはその他の点では自己完結している。別のマップから参照されることがないからである。この特徴があるので、次章で行うようにマップを個別に検討することが可能である。

自己完結しているというマップの性質は、より本質的には、クロスウォークだけでなくそれを使った機械処理可能な変換も、扱うタスクにより大小さまざまな順不同のマップの組から動的に構成できることを暗示している。たとえば、このクロスウォークに定義されているすべてのマップを実装した変換プログラム使って、ONIX レコードを網羅的にアーカイブ品質の MARC レコードへ変換したり、その逆の変換をしたりすることができる可能性があり、どちらもセマンティクス検証工程にかけられるものになるだろう。ただし、どちらのレコードも変化が生じる可能性のあるデータストリームに由来するものなので、アトミックなマップは漸進的変化の論理的にわかりやすい担い手として解釈することもできる。

たとえば、*Metadata for Publishers* の工程では、MARC レコードから抽出した件名標目を対応する ONIX レコードに適用している。また、ONIX レコードから入手した目次などの補足データを対応する MARC レコードに適用している。これらの処理には、ONIX の<Subject>要素と<OtherText>要素に関するマップだけが関係しており、マップ全体が関係しているわけではない。また、ONIX 3.0は ONIX 2.1との後方互換性がないが、マップという概念は ONIX 3.0で必要となる変更を実装する際の明確な概念枠組みを提供する。幸運なことに、クロスウォークを完全に書き直す必要はない。新しい要素にはマップを追加し、非推奨となる要素のマップは削除し、その意味は変わらないが ONIX レコードの構造上の位置が変更される要素のマップの多くは容易に変更することができるからである。

次章では、ONIX と MARC の書誌的記述に必要な最も重要なマップについて説明する。ただし、説明は包括的なものではなく、「レコード構築」アルゴリズムの提供を意図したものでもない。その目的は、マップがなぜその形になっているかを明らかにする解釈を提供することである。説明は、簡単な例から始めて、徐々に複雑で問題を含む例に進めていく。これにより、おそらくは MARC や ONIX の標準化委員会が関与することになる将来の標準化作業で解決しなければならない問題を明らかにする。

OCLC プロジェクトチームは、多対1の関係や条件付ロジックを認める十分に詳細なマップの定義が与えられれば、ONIX を MARC にマッピングすることは可能であるという仮説を持って作業を開始した。ソースとターゲットの要素間に等価性が認められると、レコードに含まれる関連データが発見・複写され、アルゴリズムに従って操作され、定数データとして生成されるか対応する表からマッピングされる。これらの操作は、商品品質の実装システムが持つ構文検証またはセマンティクス検証にかけられるレコードを作成できるほど高度なものである。ただし、これらの操作は基本的に構文上および語彙上の等価性を立証するものなので、ONIX と MARC のセマンティクスに大きな違いがある場合はあまり有効でない。

## 2.0. マップの詳細な調査

ONIX と MARC のマッピングは可能である。2つの標準には互換性があり、いくつかの点で相補的だからである。しかし、これは偶然ではない。ONIX 標準は MARC が図書館で初めて採用されてからおよそ30年後に提案されたので、MARC の利用者から学んだ貴重な経験が活かされているのである。北米において ONIX 標準を普及する目的でベストプラクティスを勧告している、ニューヨークを本拠地とする諮問機関である BISG のメタデータ委員会には図書館関係者が参加している。勧告の1つに、リストに挙げられた必須の ONIX 要素を含むレコードを提供しているデータプロバイダを認定するという提案がある。必須要素のほとんどは図書館で作成される書誌的記述の中核をなすものであり、固有識別子、タイトル、作者、主題、資料種別、形態事項、コンテンツの言語が含まれる。その他の要素は、図書や DVD が書店の供給連鎖を通じて移動する際に発生する重要な事象に合わせて指定されており、返品条件、地域販売権、一箱梱包商品数、供給者名などを示す ONIX 要素が該当する。要素の一覧と BISG の勧告は、

*Product Metadata Best Practices for Senders* (BISG 2005)にまとめられている。この章では、書誌的記述の中核を占める ONIX 要素が MARC へのマッピングでどのようになるかを調べる。

## 2.1. 識別子

図2.1は、小売店が在庫リストの管理と点検に使用している EAN-13 商品番号にクロスウォークが定義しているマップを適用した結果を説明する ONIX と MARC 両レコードの一部を示している。基本的な情報は、ONIX-MARC クロスウォークの ONIX シート(表1)の19行カラム E に載っている(OCLC 2010a)。この行と直前の3行は ONIX の<ProductIdentifier>要素と MARC の024フィールド間の対称的な関係を示している。

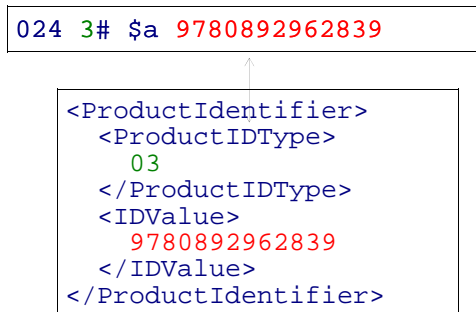


図 2.1. ONIX と MARC で表現された EAN

MARC レコードの断片は図の上の箱の中に示されており、対応する ONIX のデータはその下に示されている。データの色コードは図1.1と同じ意味である。すなわち、構造的な違いは青色、転写されたデータ値は赤色、マッピングされたデータ値は緑色で示されている。この例では、必須の子要素<ProductIDType>と<IDValue>を持つ ONIX の<ProductIdentifier>要素が MARC 024フィールドにマッピングされている。緑色で示されている<ProductIDType>要素の値'03'は MARC の1番目のインジケータの'3'にマッピングされている。また、赤色で示されている ONIX の<IDValue>要素の値は MARC の\$aフィールドに転写されている。

<ProductIdentifier>は BISG のベストプラクティスガイドラインによれば必須要素である。ONIX の<ProductIDType>要素の値は、EAN、UPC、OCLC 番号、ISBN、その他9つの識別子を指定するコードである。対応する情報は MARC では020または024フィールドと1番目のインジケータの値で表現される。これは図書館界でも出版業界でも重要な情報だとみなされているので、マッピングによりこの情報のほとんどが保存されることは幸運である。唯一の例外は10桁 ISBN と13桁 ISBN の違いである。ONIX の表現ではこの違いを<ProductIDType>に別の値を持つことで区別している。取引のためには区別することが重要だからである。現在取引されている商品が2つの形式の ISBN を持たねばならない期間が長期に及ぶ場合は特にそうである。しかし、2つの種類の ISBN は MARC 020フィールドの\$a では同じ表現を持つ。これらを区別するために、OCLC の変換ソフトウェアではデータを調べて桁数を数えるエレガントではない解決策を採らなければならなかった。

## 2.2. 補足テキスト

図2.2は、表10 *Description* シートに示されている補足テキストのマッピング指示で作成されるレコード断片を示している。これらの指示は、目次、書誌的注記、書評、見返しのコメント文、裏表紙のコピー文、要約を含む ONIX 要素をマッピングする。この例では、必須の子要素である<TextTypeCode>と<Text>を持つ ONIX の<OtherText>要素が MARC 520フィールドにマッピングされている。ここで、緑色で示されている第1インジケータの値はテキストの種類が要約であることを指定しており、赤色で示されている<Text>要素のデータは転写されたテキストである。

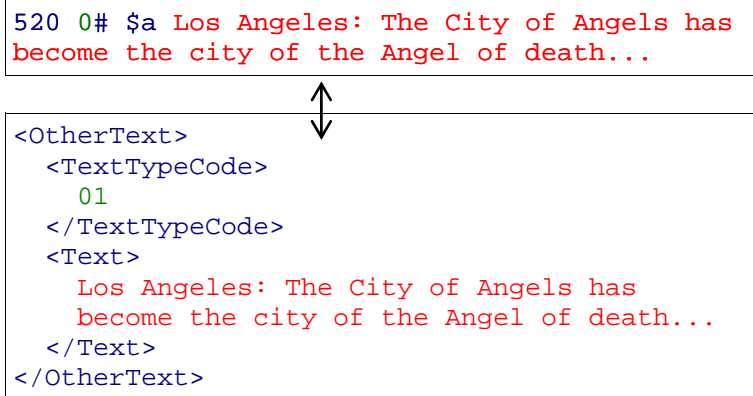


図 2.2. ONIX と MARC で表現された補足テキスト

補足テキストは、ONIX レコードを使って MARC を高度化する重要な情報源であり、MARC の 500, 505, 520, 545 フィールドにマッピングされる。多様な MARC フィールドがターゲットになるが、このマップの複雑性は <ProductIdentifier> 要素に関するマップと同等であり、テキストを参照する必要がないことだけが異なる。その結果、補足テキストのマップの方が少しだけ簡単であるが、テキストの表現方法については何も仮定することができない。たとえば、目次の章見出しと節見出しのインデントがハードコーディングされているかもしれない。フォーマット表示するために正式なマークアップである XML の <CDATA> 要素でテキストが明示的に囲まれているかもしれない。あるいは単純な文字列かもしれない。テキストがローマ字以外の文字を含む場合、ソフトウェアが文字コードを間違いなく特定することは難しいだろう。補足テキストに関してはもう一つの問題がある。読むことができるテキストリソースと画像やオーディオファイルなどのテキスト以外のメディアで表現されたリソースを ONIX 標準は区別していることである。この区別は現在では ONIX 設計者も恣意的なものだと認めている。非テキストメディアへの参照は <MediaFile> 要素にエンコードされるが、これは講演録音について ONIX レコードの作成者に難しい判断を強いるものであり、MARC との関係性を多対多の複雑なマッピングにするものでもある。おそらくはこの問題のため、ONIX for Books 3.0 では <OtherText> と <MediaFile> の 2 つの要素について再検討され、少し異なる区別を行っている。すなわち、要約やカバーの宣伝文、書評の引用などの書かれたテキストは <TextContent> 要素として伝達され、雑誌記事やラジオ放送など、任意のメディアで発表された第三者のコンテンツは <CitedContent> 要素からリンクされることになった。もちろん、この ONIX の主要要素の重大な再設計に歩調をあわせて MARC 標準を変更することはできないので、この改訂された MARC との関係もさらに複雑な多対多のマップとして表現されなければならないことになる。

### 2.3. 主題

図 2.3 のレコード断片は主題要素の ONIX と MARC 間のマッピングを示しているが、これも比較的簡単な関係の一例である。関連するマップは表 1 (ONIX シート) の 107 行から 111 行に定義されている。この例では、<SubjectSchemeIdentifier> 要素のコード '04' は、<SubjectHeadingText> 要素のデータが LCSH であることを示している。識別コードの一覧には、MeSH やシアーズ件名標目表、DDC など図書館界で一般に使用されている主題スキームや出版社や書店で使用されている主題スキームが載っている。



```

<Subject>
  <SubjectSchemeIdentifier>
    04
  </SubjectSchemeIdentifier>
  <SubjectHeadingText>
    Chronic pain—Health.
  </SubjectHeadingText>
</Subject>

```

図 2.3. MARC と ONIX における主題標目

青緑色で示した対応関係はこの例に特有のもので、2つの MARC サブフィールドが、ONIX の <SubjectHeadingText>要素にマッピングされる際に件名標目の ISBN 区切り法に従ってダッシュでつながれて1つのデータ値に折り畳まれることを示している。クロスウォークが示すように、他の LCSH サブフィールドにもこの変換が行われる場合がある。その結果、LCSH の規約に沿ってフォーマットされた文字列になるが、再び<SubjectHeadingText>要素を MARC に戻す際には、データからサブフィールドの情報が失われている。言い換えれば、人も機械処理も *Health* が \$x から来たものなのか、別の MARC サブフィールドから来たものなのかを判断できない。

OCLC の *Metadata for Publishers* プロジェクトにおけるデータ処理では、BISG が管理している BISAC 標目から MARC へのマッピングも必要である。BISAC を処理する必要があるので、クロスウォークの 111行に定義されているように、ONIX と MARC の関係がセマンティクス上少し複雑になっている。これらすべての指示に従うことにより、ONIX から MARC へ、さらに MARC から ONIX へ情報を失うことなく変換することができる。出版社は、電子データの交換を容易にするために供給連鎖で使用されている *BISAC コード*と、宣伝文書の作成や主題別の店舗陳列を行うために使用される *BISAC テキスト*を区別している。この区別は、ONIX データが MARC にマッピングされる際に保持されるが、別の意味が付加される。上で述べたように、BISAC テキストは MARC レコードにマッピングされる際に件名標目に対応するが、BISAC コードは主たる主題とみなされ ONIX の<BasicMainSubject>要素に記録された場合は分類として解釈される。

件名標目と分類コードは出版社の供給連鎖では論理的に区別できないが、図書館員は世界中で記録された情報を発見・入手可能にする目的で組織化するための関連はあるが2つの異なるツールであると考えている(Chan 2007, 3)。OCLC の分類専門家によれば、BISAC は2階層の分類体系だと解釈することができる。BISAC コードの最初の3文字は、*歴史、詩、犯罪実録、学習参考書、聖書*など、最上位の51の主題カテゴリ・ジャンルを示している(BISG 2009)。これに続く6桁の数字コードは、*アフリカの詩や精神的または宗教的な詩*などの下位主題を示している。BISAC コードを従来からある図書館の分類体系に一致させるために、3文字の BISAC プリフィックスを主題カテゴリコードを記入する MARC フィールド072 の \$a に、6桁の拡張コードを \$x にマッピングする。

これらの詳細は図 2.4 (次ページ) の左側の MARC 断片に示されている。ONIX の <BASICMainSubject>要素に記録されている BISAC 主題コードは、上で説明したように、対応するテキスト値に変換され、出力の MARC レコードでは650フィールドとして表現されている。図2.4の右側の MARC 断片は図2.3で示したものと基本的に同じであるが、標目が BISAC 由来であることを示すために第2インジケータに‘7’を、\$2サブフィールドに定数データ‘bisacsh’を設定していることに注目されたい。

BISAC コードは、*レコードの種類、対象利用者、文学形式*など MARC 008フィールドを埋めるための付加的情報として解釈することもできる。詳細は表7 (*BISAC SH*シート)で説明されている。



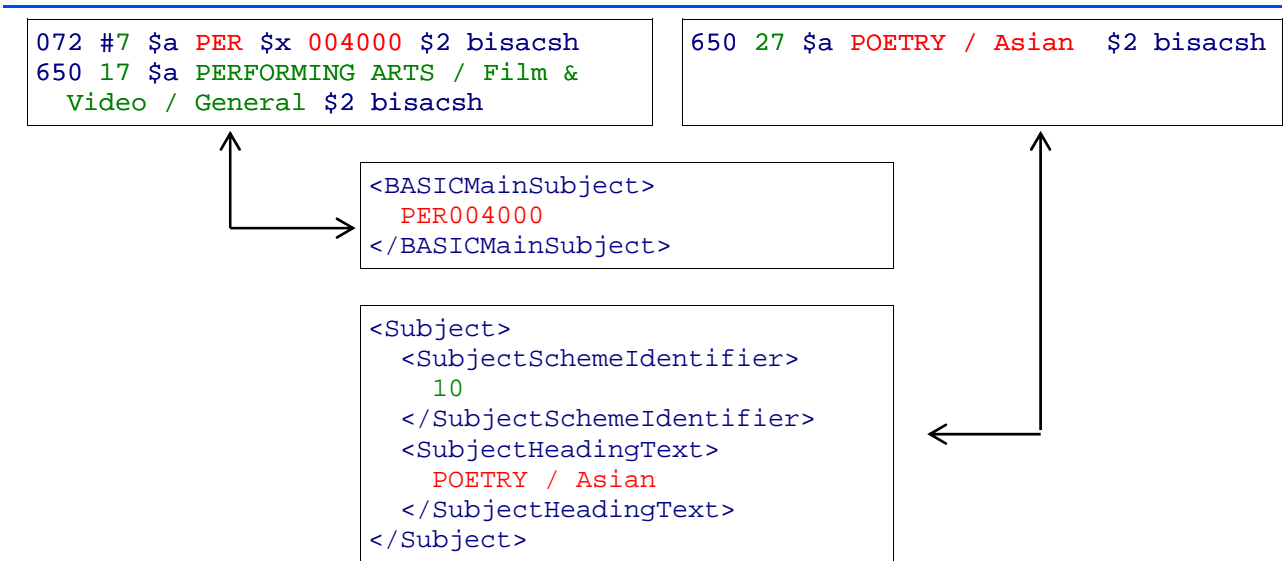


図 2.4. BISAC コードと BISAC 件名標目が関係するマップ

## 2.4. 状況の確認

問題のあるマップを検討する前に、ここで ONIX と MARC の関係について少し掘り下げて検討しておくが良いだろう。これまでに説明したマップはすべて BISG が勧告した必須要素に関係している。また、対象となった要素は MARC への変換後も業務用の処理に使用できる品質を保っている。これはおそらく、要素のセマンティクスが基本的に保存されているという事実の結果である。たとえば、ISBN や ISSN、OCLC 番号などの固有識別子は ONIX でも MARC でもまったく同じものを意味する。また、目次や著者の経歴、書評、その他<OtherText>要素で伝達されるデータも、表示や機械対機械の処理で時に問題を引き起こすという事実はあるものの、同様である。

しかし、たとえ最も簡単な例においても、データの参照や場合によっては変更を行う必要があり、往復変換を行うと情報の消失を招く可能性があることが心配の種である。たとえば、対象となる要素が件名標目に関係する場合、ソースが MARC で複数のサブフィールドが存在する時は ISBD の区切り記号が取り除かれているはずである。しかし、このサブフィールドを元に戻す方法に関する情報は、作成された ONIX データをその後再び MARC にマッピングする場合には失われている。このような要素間の粒度のわずかな違いを浮き彫りにする問題は2つの標準の調整作業を行う専門家により慎重に検討されるべきである。これは ONIX と MARC の関係に広く当てはまる問題であり、最後の章でもう少し取り上げる予定である。

## 2.5. 難問

ONIX と MARC の関係で最も重要だが最も複雑なマッピングが3つ存在する。ONIX の<Contributor>要素、<Title>要素、<ProductForm>要素や<Measure>要素などの形態的特徴に関する詳細情報を記録する ONIX 要素である。この複雑性の原因は、複雑な構文関係、セマンティクスの大きな違い、両標準に見られる使われなくなった古い要素に帰することができる。

### 2.5.1. 著者と貢献者

図2.5は、表1(ONIX シート)の53行から78行で定義されている、ONIX と MARC の責任表示に関する要素の関係を示す2つのマッピングの結果を示している。このマッピングでは2つの問題を解決する必要がある。第1に、AACR 規則では、100フィールドに記載される著作の知的または芸術的内容の創作に主たる責任を持つ個人と、700フィールドに記載される二次的な貢献者を基本的に区別している。この区

別は ONIX には存在しない。ただし、脚本家、作詞家、イラストレーター、序文著者などの <ContributorRole>コードを指定しているコードリスト17の値により近似させることができる。すなわち、値 A01(テキスト著作の著者)は、MARC レコードでは主たる著者とみなして、100フィールドにマッピングする。一方、その他の値を持つ<Contributor>要素は700フィールドにマッピングする。ONIX レコードに値 A01の<ContributorRole>要素を持つ<Contributor>要素が複数含まれる場合は、最初の要素を100フィールドへ、残りの要素を700フィールドにマッピングする。この規則の効果は図2.5に示した2つのレコード断片で見ることができる。Abraham Smith はテキスト著作の主たる著者であり、MARC では100フィールドに記載され、監督の Peter H. Hunt は700フィールドに記載されている。

第2の問題は、名前の構文またはフォーマットである。MARC の100および700フィールドで必須の\$a サブフィールドには少なくとも姓を記載しなければならないが、通常は、姓と名をこの順で表した文字列を記載する。\$a に記載した名前の完全形、家名や王朝名、肩書き、所属などはオプションの MARC サブフィールドに記載する。これらの情報のすべては ONIX の<Contributor>要素からマッピングすることができる。OCLC が実装した ONIX から MARC へのクロスワークでは入力レコードの<PersonNameInverted>要素をチェックし、この要素がない場合は、タグ付けされた個々の名前要素から姓名順の名前を作成している。2つの名前フォーマットは図2.5の a,b に示されている。変換後の MARC レコードにおける名前の文字列は基本的にソースである ONIX 要素によらず同じである。

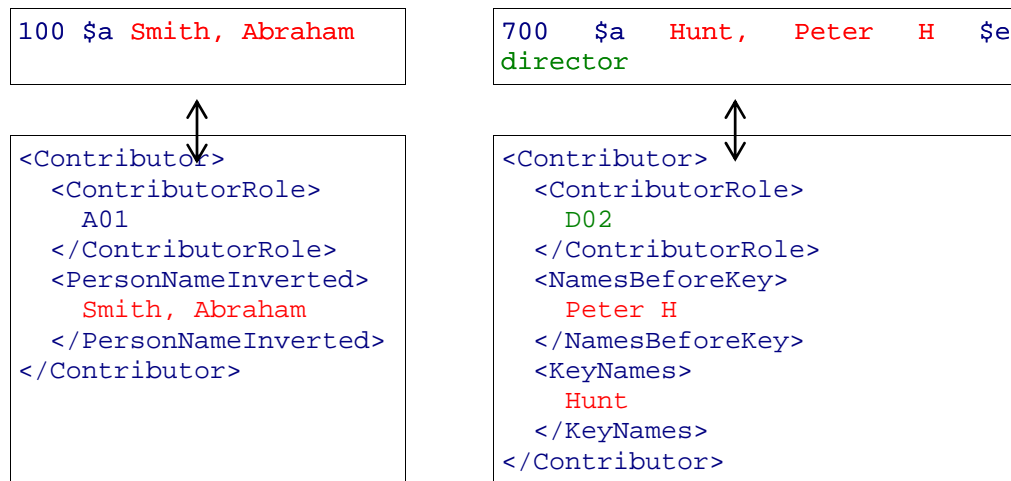


図 2.5 a, b. ONIX と MARC における貢献者の関係

このロジックで ONIX と MARC の貢献者要素の詳細はほとんどが説明できる。しかし、それがなぜそれほど複雑なのかを理解するためには、表面的な構造上の違い以外の原因を知ることが役に立つ。複雑性の一部は、ONIX は通常データ値の句読法や書式には関知しないという原則から<PersonNameInverted>要素が外れているという事実に帰することができる。この要素の場合、要素名がわかりやすく示しているように、データ値に必須の書式を持っている。

さらに重要であるが、複雑性は両標準の貢献者要素の定義におけるセマンティクスの相違に帰することができる。ONIX では<Contributor>要素が一番重要である。したがって、この要素は自己完結的で独立しており、最上位ノードである<Product>要素の直接の子要素である。BISG のベストプラクティスガイドでは、<Contributor>要素が書店にとって最も重要な理由を次のように明確に述べている。

...本の著者は、消費者が知っている我々の商品のもっとも分かり易い「ブランド」である。  
ある種の主題カテゴリにおいては、タイトルや出版社、シリーズ名などの他の重要なデータ

要素も商品に貢献した人の名前の重要性に比べたらほとんど取るに足りないものである。たとえば、ジョン・グリシャムの新作小説のタイトルは、その本を売るためのデータの一つにもならない(BISG 2005, 23)。

一方、MARC レコードでは *貢献者* データは、コンピュータシステムで書誌レコードを検索する際に指定する名前や語、コードである *アクセスポイント* というメタ概念の下に置かれている。MARC レコードにおける主たるアクセスポイントは100フィールドであり、このフィールドには名前の標準形、生没年、著作のタイトル、出版日など、主たる著者またはその著作に関する最適なデータを記載する。原則的に、共著者や翻訳者、イラストレーターなどの二次的な貢献者に関する同様な情報は700フィールド、すなわち、*副出記入* に記録される。ONIX レコードにはアクセスポイントに相当するものがないので、MARC レコードが ONIX にマッピングされるとこの概念は完全に失われる。また、データ自体にも問題がある。なぜなら、MARC において貢献者やタイトルの記述を記載するこれらのフィールドのサブフィールドは ONIX の該当要素に比べて詳細でないからである。さらに、これらの問題は団体著者や統一タイトルを記録する基本記入や副出記入でも繰り返される。

*貢献者* 要素に関係するマップに見られるさらに多くの複雑性は、これは ONIX も MARC も共に持続的に進化していることを証明するものではあるが、両者が過去の形式を引きずっているという事実に戻ることができる。たとえば、ONIX では著者経歴の記述は <OtherText> と <Contributor> のどちらの要素にも現れる可能性がある。そのため、クロスウォークは両方の可能性を考慮しなければならない。同様に、主たる著者、二次的な著者、イラストレーター、翻訳者、編者などといった貢献者の役割に関する注記は、MARC では100と700のいずれかのフィールドの、コードは \$e に、テキストは \$4 に現れる可能性がある。また、両標準とも非推奨となった要素を含んでいる。この要素は、最新バージョンに準拠するようにレコードを変換する場合は無視しても問題ないが、そうでない場合は削除される恐れがある。そして、ONIX 3.0の改訂ではこれらの問題が対処されていないので、両標準が独立に進化を続けるとマッピングがますます複雑になる可能性がある。

## 2.5.2. タイトル

表1 (ONIX シート) の43行から48行で定義されている ONIX と MARC の *タイトル* 要素では、*貢献者* 要素と同じ問題が2つ現れる。1つは、*タイトル* という概念が ONIX レコードでは第一義的なものであるのに対して、対応する MARC フィールドは異種のデータからなる *アクセスポイント* であると解釈されることである。たとえば、図2.7の真ん中に示した245フィールドに注目してほしい。このフィールドは *タイトル* テキストを含む \$a サブフィールドと録音物を表すメディアコードを含む \$h サブフィールドを持っている。

もう1つは、責任表示要素の場合と同じように、ONIX の *タイトル* 要素には、フォーマット済みのテキストを収める要素と未フォーマットのテキストを収める要素があることである。たとえば、<TitleText> 要素は、規定の句読法や書式なしにタイトルを伝達するが、<TitleWithoutPrefix> 要素にはアルファベット順リストで使用する形のタイトルを記入する。両者の形のタイトルをマッピングすることになる。図2.6は、クロスウォークに投入されたヘルマン・ヘッセの小説に関する記述から生成された固有のタイトルとその翻訳タイトルと別タイトルを含む ONIX と MARC の断片を示している。この例では、MARC 245 \$a に記録されている主たるタイトルの元になった ONIX のデータは特別なフォーマットを施された ONIX フィールドで構成されている。それ以外の MARC フィールドは ONIX の <TitleText> 要素からマッピングされている。これらの例は、<TitleType> 要素に指定するためにリストから選択したコードから <TitleText> 要素とそれに対応するフォーマット済みテキスト要素に収められた文字列を主たるタイトル(01)、翻訳タイトル(06)、別タイトル(05)として識別する方法を示している。この情報は <TitleText> 要素の値を242から247の範囲の MARC フィールドにマッピングする。246フィールドへのマッピングだけが問題になる。なぜなら、*頒布者が付けたタイトル*、*タイトルの頭字語*、*表紙の別タイトル*、*裏表紙の別タイトル* などの別タイトルを ONIX は区別しているが、MARC はこの区別を行わないので変換の際にこの情報が失われるからである。

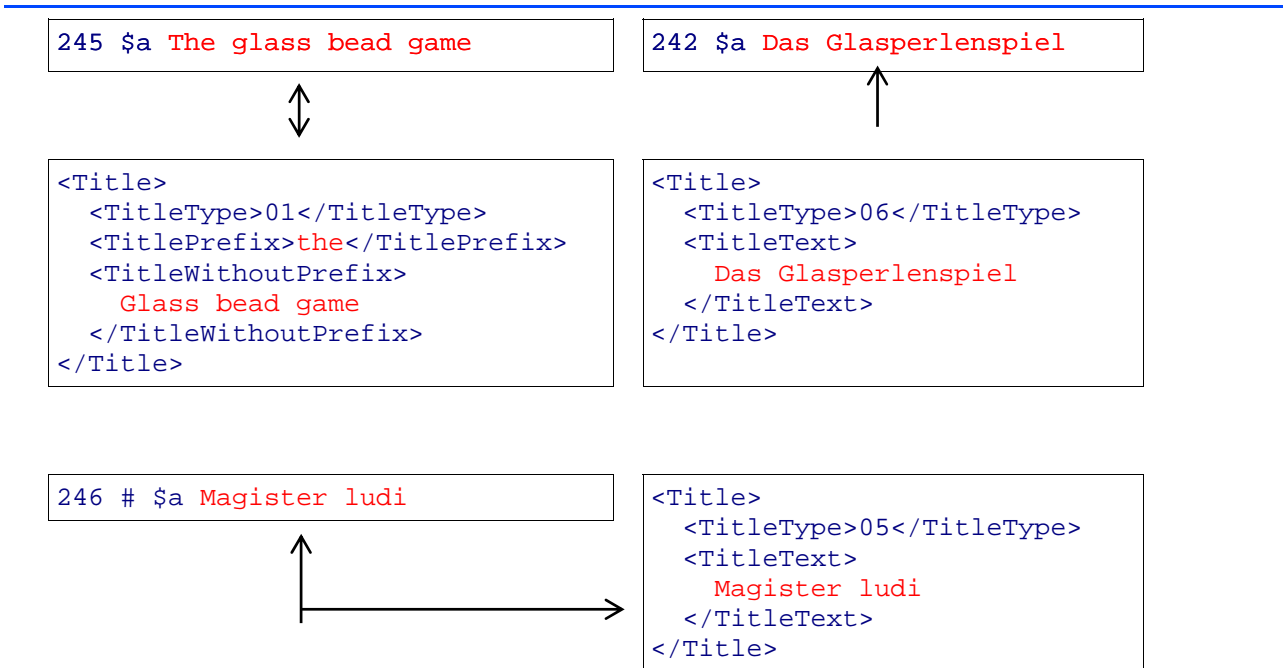


図 2.6. ONIX と MARC におけるタイトル

## 2.5.3. 形態事項

ONIX から MARC への最も複雑なマッピングには形態事項を記入する要素がある。この情報は出版社の供給連鎖では非常に重要である。なぜなら、売買される商品のほとんどは、重さ、寸法、価格、適正な郵便料金、パレットへの積載量、倉庫の保管量など測定すべき物理的次元を持っているからである。さらに重要なことは、消費者は本の大きさやページ数、イラストの有無、ハードカバーかペーパーバックか、箱入りか否かに基づいて本の購入の判断を行うことである。同様に、DVD を購入する場合、ブルーレイディスクを注文したのに古い形式のディスクが届いては要求を満たすことができない。図書館員も、保管・発見されなければならない資料に対して同様な関心を持っているが、形態事項は長期的な保管目的にも役立っている。形態事項の記述は、様々な版を見分ける特徴を捕捉でき、機械可読メディアの場合はコンテンツリーダーを判別できるくらい詳細なものでなければならない。そして、これら2つの要件は永久に満足されなければならない。

図2.7は、様々な演奏家によるプッチーニのオペラからアリアの録音を集めた CD を記述する ONIX と MARC の対応する断片を通じて、これらの問題の多くを示している。ONIX において重要なデータ要素は、<ProductForm>要素の値‘AC’であり、この ONIX レコードが音楽 CD を記述していることを示している。これまでの図と同じように対応する値は緑色で示されている。表2( ProductForm シート) が示しているように、<ProductForm>要素の値‘AC’は6つの MARC フィールドに対応付けられる1対多のマップである。そのうちの1つ、MARC 007フィールドは12のサブフィールドを持っている。これらのサブフィールドの値は記述されている物体が、秒速1.4メートルの一定の線速度で再生することができる、デジタル形式で格納されたデジタル録音を含む大量生産されたプラスチックと金属でできた音盤であることを示している。MARC リーダーフィールドの「書誌レベル」サブフィールドの値‘m’は、記述対象が個々の独立したアイテムであり、コレクションやシリーズの構成要素ではないことを示している。また、その左の位置にある値‘j’は、この物体が録音物であることを示しており、この指定は245 \$h サブフィールドにテキスト値として繰り返されている。

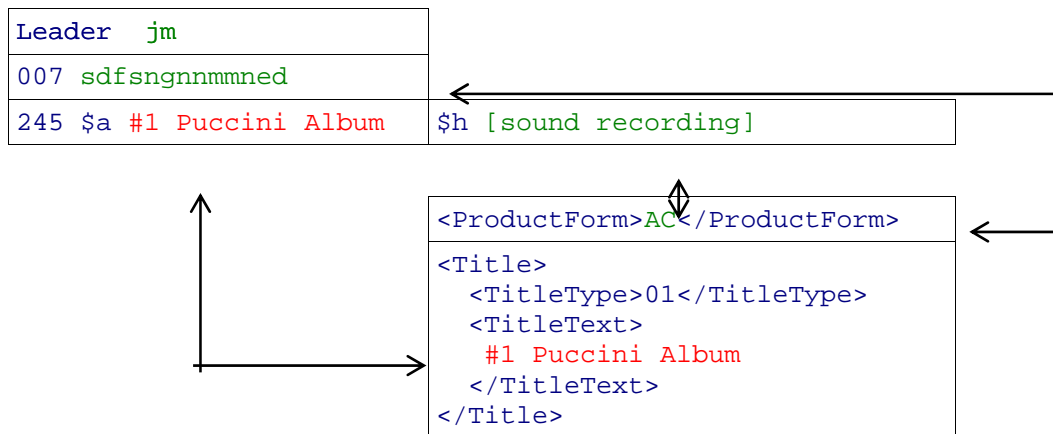


図 2.7. ONIX と MARC における形態事項

利用者の発見ということを考えると、どうして MARC の記述にも ONIX の記述にもエンドユーザが使用する 'CD' という語が見られないのか、また、そのことがプッチーニの aria を録音した CD を探している購入者や図書館利用者にどんな結果をもたらすかという疑問を注意深い読者は持つだろう。もちろん、ONIX の *Audio CD* という用語は、MARC の *sound recording* という用語や 007 フィールドの一連のコードよりも近いものである。しかし、統制語彙の提唱者は、通称の CD はあまりにも不正確で一過性の名前であり、ONIX の *Audio CD* という用語は統制用語でもエンドユーザ用語でもないので使用される見込みはまったくないと主張するだろう。ONIX ではコード 'AC' だけが統制されているのはこのためである。関連する *Audio CD* という用語は注釈に過ぎず、様々な言語やユースケースでそのまま使用されることは保証されていない。しかし、MARC 007 フィールドの記述はおそらくあまりにも細かすぎるだろう。それは、2004 年に米国で製造された特定の録音を正確に記述するかもしれないが、技術的仕様は時がたてば必ず変化するものであり、今でも地域的な差異が存在するだろう。もしそうであれば、いくつかの MARC サブフィールドの様々な値について、このマップを何回も繰り返さなければならないことになる。この問題は、出版物の形態事項に使用する重要な多くの語彙が標準化も定義もされておらず、すべての関係者に役立つような子要素に分解されていないという事実によるものである。

技術は進歩を続けるので、書誌レコードにおける形態事項要素は今後も複雑になる一方であろう。実際、ONIX 3.0 が策定された第一の動機は電子ブックを記述する機能の向上であった。ONIX 3.0 による電子ブックの記述は、現在そうであるように、ファイルサイズを持つことになるが、これは購入判断に使えるほど十分な情報なのだろうか。本が詳細で包括的なものかどうかを知りたい読者とあまり長い本には興味のない読者がいたとしたら、電子ブックの元になった資料のページ数や大きさといった記述はこれらの読者を満足させるだろうか。満足させたとして、印刷本が本当に電子ブックの元になったものであり、単なる相当品でないことが目録作業で確認されていなかったとしたら、図書館界はこの情報を受け入れるだろうか。ファイルフォーマットの違いはどのように記述されるべきなのか。電子ブックリーダーは現在のところ互換性がないので、これは不可欠な情報である。また、たとえ内容が同じものであっても、各フォーマットは違う ISBN を持つことになるのか。これらの疑問に対する回答<sup>2</sup>は ONIX と MARC の関係に影響を与えるが、明らかにそれは関係を単純化するものではないだろう。

### 3.0. 教訓

前章で説明した ONIX から MARC へのマップに含まれている ONIX データ要素のすべては BISG の

<sup>2</sup>これらの課題は2009年10月7日にBISGの後援によるウェブキャストで議論されている。スライドは次のURLで公開されている。<http://www.bisg.org/events-0-469-bisg-webcastonix-for-books-30supporting-new-metadata-for-ebooks.php>.

*Best Practices Guidelines for Data Senders*に挙げられた必須要素である(BISG 2005)。これらのマップは、書誌的記述の相互運用性の促進に重点をおく商用品質のメタデータ管理システムが処理しなければならない最も重要なものである。作成されたクロスウォークは業務用にも使えるものであるが、それは、変更管理に最適化されているソフトウェア環境で作成していても維持するのが困難になるほどの、大量の知的作業を表している。

この章では、何が必要以上にマップを複雑で不安定なものにしているかを改めて取り上げたい。考察にあたっては言語学者の考え方を援用し、2つの標準の構文、セマンティクス、語用論がいかに問題の一因となっているのかを示し、同時にそれが問題解決に向けた生産的な方法を示唆しているかを順に検討する。

**変換にはデータを見なければならないので、マッピングの問題は構文レベルでは解決することができない。**

たとえば、ONIX の<Subject>要素は MARC 650と等価であり、ONIX の<ProductIdentifier>要素は MARC 024と等価である、など、語彙の等価性に基づいてメタデータ管理のための変換モデルを構築することは魅力的である。このモデルは、変換が完全な同義性に基づいて行われ、入力から出力を得るには単純な(最悪でも複雑ではあるが管理可能な)構造的操作が必要とされるだけであることを暗示している。そのようなモデルでは、ONIX から MARC へのマッピングは、語彙上のラベルの付け替えによる構造的変換に等しいものであり、それは古い ISO-2709 MARC 構文から MARC-XML を作成するために必要な変換より一段階複雑な変換にすぎない。

この単純なモデルは、ONIX から MARC へのクロスウォークでメタデータソースからそれに密接に関係するターゲットにマッピングを行うルーチンワークのほとんどを処理できる出発点である。しかし、これまでの説明で示したように、関連データを調べる必要がないマップは1つでも見つけることは困難である。これらのデータは、変換にあたっての追加条件やデータ消失の可能性、意味にわずかな変化をもたらすからである。

**標準間のセマンティクス上の違いは調整されるべきである。**

ONIX と MARC では書誌的記述を行うための要素は多くが重なっているが、2つの標準はセマンティクス的に異なる。2.3節の主題要素の説明で指摘したように、ONIX では分類と件名標目の区別がない。とはいえ、ONIX の<BASICMainSubject>要素は MARC レコードの分類コードの代替として妥当であり、BISAC テキスト文字列は件名標目として解釈することができる。同様に、ONIX 標準には MARC におけるアクセスポイントという概念がないが、通常それは再現可能である。既に述べたように、これらの要素が関係する変換では粒度が失われる場合があるが、両標準においてセマンティクス上の妥当性検査にパスするレコードを作成するという大きな目標にとって、これは致命的なものではない。

より広く見られる問題は、両標準で定義されている概念が基本的に同じものであるか否かがほとんどの場合ではっきりしないことである。これは、5.3節の形態事項の説明で示したように、新しい標準はほとんどの場合旧来の標準を尊重しなければならないという変わりゆく目録作業の歴史を反映して、MARC における資料の形態事項などの概念が重複かつ冗長に表現されているために生じる場合もある。しかし、一般的には、データ値の表現のわずかな違いが、マップにおけるソースとターゲットの要素が等価であるか否か、事実上同じ値を持つか否かの決定に対してほとんど克服できない障害となっている。

たとえば、OCLC *Metadata Services for Publishers* の作業工程で、入力された ONIX レコードを OCLC WorldCat データベースの既存レコードと突き合わせ、2つの情報源のデータを統合し、高度なレコードを作成する次の局面で何が起きるかを考えてもらいたい。入力の ONIX レコードには目次が記録されている<OtherText>要素があり、対応する MARC 505フィールドにも同じデータが記録されている目次フィールドがある。ただし、ONIX の記述では章と節の記述がダッシュか数字で分けられており、MARC の記述ではインデントが直接書かれているとする。この場合、2つのフィールドは統合されず、出力は人間だけが容易に発見できる重複した情報を持つことになる。同じことを、著者やタイトル、主題、出

版社など、フォーマットが明確に指定されていないテキストを含む書誌的記述フィールドにも指摘することができる。

幸いなことに、これらデータ表現の問題の多くは、国立図書館と図書館目録のための勧告を策定する標準化機関のイニシアティブである RDA (Resource Description and Access) の提案者により、現在検討されている (JSC-RDA 2006)。RDA の目標は、図書館界でもっとも広く普及しているセマンティクス標準である英米目録規則を「もっとも重要な要素のための一貫したデータ符号化」を勧告する現代的な標準で置き換えることである (Delsey 2009)。RDA の支持者により提案されている2つの変更が ONIX と MARC の関係、および本文書で説明した問題に特に関係している。

第1に、いくつかの記述は単に改良される。たとえば、ONIX と MARC の双方で使える、形態記述のための RDA 語彙の原案が既に公開されている (Dunsire 2007, JSC 2006b)。両標準がこの語彙を使うようになれば、セマンティクスは正確に同じになるので、適当な出力構文を作成するのに必要な構造的操作だけが必要になることになる。しかし、実現するにはまだ困難な作業が数多く残っているので、これは長期的ビジョンである。

第2の長期的ビジョンは、書誌的記述要素の多くをリンクトデータとして扱うことである (Heath 2009 and Hillmann, et al. 2010)。この変更により、レコードがマッピングされる際にデータが転写されなくなるので、ONIX と MARC の間の関係が非常に単純になる。2つの標準に準拠するレコードは、貢献者名や著作タイトル、標準的識別子、件名などのオーソライズされた同一の情報源を単に参照することになるので、本文書で指摘したような誤記や記述における意味や粒度のわずかなずれが発生する可能性がすべてなくなる。

セマンティクスの調整プログラムに残されたものは、RDA の目標では直接扱われていない問題の一扫である。たとえば、補足資料の種類や貢献者の著作の性格、物理的メディアやキャリアの特徴の指定に使用されている ONIX と MARC のコードが、もし同一でなければ、少なくともその適用範囲や記述力において同等であることを保証するために、常に警戒が必要である。また、RDA は、時に非常に長くなる補足資料をいかに符号化するかという問題にも対応していない。ただし、目次や書評は2つの標準で同じ意味を持っているので、この問題は単に両コミュニティの要件を満たす明確なマークアップスキームを設計する必要があるだけである。

#### **2つの標準に見られる実用上の相違を管理する必要が常に存在する。**

ONIX と MARC のセマンティクスの調整は前進するための大きな一歩になるはずであるが、図書館と出版社が書誌データを使用する方法には常に相違が存在するので、この作業が終了することはないと思われる。しかも、実用上のニーズは変化をもたらす第一の要因である。前進を続けるには、この変化の影響を受けるすべてのコミュニティに対するインパクトを十分に理解した上でこの変化を管理する必要がある。

## 参考文献

すべての URL は2010年3月18日にアクセスしている。

- BISG (Book Industry Study Group). 2005. Product metadata best practices for data senders. [http://www.bisg.org/docs/Best\\_Practices\\_Document.pdf](http://www.bisg.org/docs/Best_Practices_Document.pdf).
- . 2009. BISAC subject headings list, major subjects . 2009 edition. <http://www.bisg.org/what-we-do-0-136-bisac-subject-headings-list-major-subjects---2009-edition.php>.
- Chan, Lois Mai. 2007. *Cataloging and classification: An introduction*. Lanham, Md: Scarecrow Press.
- Delsey, Tom. 2009. Look before you leap: RDA compared to AACR2. Presentation given at the American Library Association Midwinter meeting, January 10, 2009, Denver, Colorado, USA. <http://presentations.ala.org/images/1/10/LLL-Delsey-ALA2009.ppt>.
- Dunsire, Gordon. 2007. Distinguishing content from carrier. *DLIB Magazine*, 13:12. <http://www.dlib.org/dlib/january07/dunsire/01dunsire.html>.
- EDItEUR. 2009a. ONIX releases prior to 2.1. [http://www.editeur.org/95/Releases\\_prior\\_to\\_2.1\\_rev03#Release%202.1%20rev%202](http://www.editeur.org/95/Releases_prior_to_2.1_rev03#Release%202.1%20rev%202).
- . 2009b. ONIX: About Release 3.0. <http://www.editeur.org/12/About-Release-3.0/>.
- Godby, Carol Jean, Devon Smith and Eric Childress. 2008a. Toward element-level interoperability in bibliographic metadata. *Code4Lib Journal*, Issue 2. <http://journal.code4lib.org/articles/54>.
- . 2008b. Encoding application profiles in a computational model of the crosswalk. *International Conference on Dublin Core and Metadata Applications, DC-2008*. <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/914>.
- Heath, Tom. 2009. Linked data: Connect distributed data across the web. <http://linkeddata.org/>.
- Hillmann, Diane, Karen Coyle, Jon Phipps, and Gordon Dunsire. 2010. RDA vocabularies: Process, outcome, use. *DLib Magazine*, 16:1/2. <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html>.
- IFLA. 2007. International standard bibliographic description. <http://www.ifap.ru/library/book264.pdf>.
- . 2010. Functional requirements for bibliographic records. <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>.
- JSC-AACR2 (Joint Steering Committee for Revision of AACR). 1988. *Anglo-American cataloging rules, second edition. Revisions 1985*. Chicago: American Library Association.



JSC-RDA (Joint Steering Committee for the Development of RDA). 2006a. RDA: Resource description and access. <http://www.rda-jsc.org/rda.html>.

———. 2006b. RDA/ONIX framework for resource categorization. <http://www.loc.gov/marc/marbi/2007/5chair10.pdf>.

Library of Congress. 2000. MARC 21 record builder. *Network Development MARC Standards Office*. <http://www.loc.gov/marc/marc2onix.html>.

———. 1999. MARC 21 format for bibliographic data. *Network Development MARC Standards Office*. <http://www.loc.gov/marc/bibliographic/>.

———. 2009. What is a MARC record, and why is it important? <http://www.loc.gov/marc/umb/um01to06.html>.

OCLC. 2009. OCLC metadata services for publishers. <http://publishers.oclc.org/en/metadata/default.htm>.

———. 2010a. ONIX-MARC mapping. Excel spreadsheet. <http://www.oclc.org/research/publications/library/2010/2010-14a.xls>.

———. 2010b. WorldCat. <http://www.worldcat.org/>.

RDA (Resource Description and Access). 2010. Constituency review. <http://www.ifap.ru/library/book264.pdf>.