

(Web Archive) を国際標準化機構 (International Organization for Standardization : ISO) に提案している。WARCは、ARCを標準的なフォーマットとして改定したものである。WARC、ARCとも、任意の1つのファイルの中に、簡単なテキストヘッダーとコンテンツデータを格納する点は同じだが、WARCではメタデータの付与や重複収集の記録等が可能である。

4. 今後の動向とNDLの役割

IIPC第2期は、アーカイブ技術のさらなる高機能化とともにアーカイブデータの長期保存等を焦点としている。具体的には (1) 収集の障害となるスパム回避や映像コンテンツ収集等Heritrixの性能向上、(2) 新たな全文検索の仕組みの検討、(3) ウェブアーカイブ長期保存のためのガイドラインの策定等を予定している⁽⁴⁾。

NDLデジタルアーカイブシステム⁽⁶⁾では、Heritrixの非欧米言語対応追加機能等を開発の上、採用する計画である。NDLは、これらの開発成果やIIPC結成とほぼ同時期から開始したWARP⁽⁶⁾ (インターネット情報選択的蓄積事業) 等ウェブアーカイブの取組み実績を背景に、IIPCの活動への積極的な貢献を目指している。

(関西館電子図書館課：柴田昌樹^{しばたまさき})

- (1) IIPCの加盟機関、組織、成果物等は次を参照。
International Internet Preservation Consortium "Netpreserve.org". <http://netpreserve.org/about/members.php>, (accessed 2008-05-13).
- (2) 第1、2回IWAの主催者のひとりが、2008年1月当館で講演を行ったマネス (Julien Masanès : 2004年IIPC議長、現ユーロピアンアーカイブディレクター) 氏である。
"講演とディスカッション「ウェブアーカイブの現在と展望 - 国際連携に向けて -」". 国立国会図書館. http://www.ndl.go.jp/jp/publication/proceedings/web_arch08/index.html, (参照 2008-05-13).
- (3) Internet Archive. "Wayback Machine". <http://www.archive.org/web/web.php>, (accessed 2008-05-13).
なお、3.1.4)の "Open source Wayback" は、Wayback MachineをOSS版として改良したものである。
- (4) Carpenter, Kris. "Opportunities for Global Cooperation & Collaboration in Web Archiving". National Diet Library. http://www.ndl.go.jp/jp/publication/proceedings/web_arch08/carpenter.pdf, (accessed 2008-05-13).
- (5) "NDLデジタルアーカイブシステム". 国立国会図書館. <http://www.ndl.go.jp/jp/aboutus/ndl-da.html>, (参照 2008-05-13).
- (6) "WARP". 国立国会図書館. <http://warp.ndl.go.jp/>, (参照 2008-05-13).

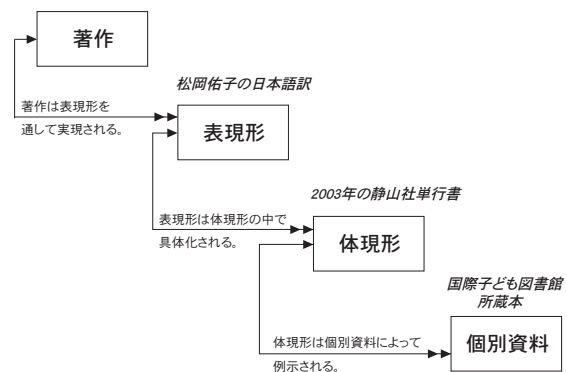
CA1665 OCLCのFRBR化の取組み：xISBNサービスを中心に

1. FRBRの影響とFRBR化

1997年の刊行以来、図書館界内外から注目を集めている「書誌レコードの機能要件 (Functional Requirements for Bibliographic Record : FRBR)」は、近年進行している図書館目録の見直しや高度化に影響を与えている。FRBRは、書誌レコードが持つべ

き機能を、データベース構築で使われる実体関連分析という手法を使って、実体、実体の属性、実体間の関連、利用者タスクにより表した概念モデルである。FRBRでは利用者の関心対象を10の実体として定めているが、その中でもモデルの核心は知的・芸術活動の成果として定義される「著作 (Work)」「表現形 (Expression)」「体现形 (Manifestation)」「個別資料 (Item)」の4実体である (CA1480参照)。「著作」等が重要視されるのは、これらがカッター (Charles A. Cutter) の時代から言われているが未だ実現に至っていない目録が持つべき機能、すなわち (利用者にとっては未知の) 版を特定の「著作」の下に集中させて (collocate) 提示すること、の実現につながるものだからである。「著作」は、その重要性は認識されながらも、曖昧模糊とした抽象的な概念であるため、従来の目録では十分に活用できていなかった。

JK.ローリングのハリ・ポッターと賢者の石



※斜体は、各実体の例を示す。

図 知的・芸術活動の成果としての実体とその例

目録等の検索システムにFRBRを実装することは、FRBR化 (FRBRization) と呼ばれている。現在行われているFRBR化は、図に示したような「著作」等の実体とその間の関連をシステムに適用するもの、具体的に言えば、「体现形」レベルにある既存の書誌レコードを「著作」や「表現形」の単位でクラスター化するものが多い。こうしたFRBR化システムの例には、オーストラリア国立図書館LibraryLabsが開発した総合目録プロトタイプシステム、VTLS社の図書館システム "Virtua"、OCLCの "Fiction Finder" (E588参照) や "Worldcat.org" などがある。(なお、後述するように、これらのシステムは厳密な意味でのFRBR化ではないため、FRBR風 (FRBResque) と呼ばれることもある。) 特にOCLCは、2001年頃からFRBR化の研究に取り組み、その研究成果のいくつかは、2008年3月現在、正式なサービスとなっている。

2. FRBR化に関するOCLCの研究開発

OCLCのFRBR化に関する取組みの基盤は、2002年

に開発したFRBR Work-Setアルゴリズムである。主任研究員ヒッキー (Thom Hickey) が開発した同アルゴリズムは、書誌データベースに含まれる特定の「著作」について、当該「著作」に関連付けられる「表現形」(書誌レコード)を機械的に同定するものである。「FRBR」化ではあるが、「表現形」の同定や「表現形」単位での書誌レコードのクラスター化はできない。

アルゴリズムは(ごく大雑把に言えば)MARC21レコード中の基本記入標目(1XX)やタイトルステートメント(245)等から著者とタイトルを抽出し、それらを組み合わせて「著作」を同定する識別子として用いている。更に、既存の書誌レコードは記述の一貫性が必ずしも保たれていないことを考慮して、同定にあたっては、「Name Authority Cooperative Program (NACO)」の典拠レコードや正規化ルールも使っている。アルゴリズムの詳細はOCLCのウェブサイトで公開され、アルゴリズム自体も無料でダウンロードできるようになっている。

このWork-Set アルゴリズムを使ってOCLCは各種の研究開発を進め、ついには世界最大の総合目録データベースであるWorldcatのFRBR化を、オープンなウェブサービスであるWorldcat.orgにおいて実現させている。Worldcat.orgでは、同一の「著作」に関連付けられる書誌レコード(表現形)をWorksetとしてグループ化し、検索結果一覧の画面では、そのWorksetに含まれる最も所蔵館の多いレコードを代表として示している。更に、書誌詳細表示では、同じグループに属する他のレコードを“Editions”としてタブから見えるようにし、例示されたレコードとは異なる版のレコードに飛べるようになっている。例えばローリング(J.K. Rowling)の*Harry Potter and the sorcerer's stone*は2008年5月現在、66件のレコードが1つのWorksetとなっているが、結果一覧では1998年に刊行された最初の米国版単行本のレコードが例示され、Editionsではその他の版(英国版やオーディオブックなど)が示される。なお、このWorksetは、FRBRでの「著作」の定義とは異なり、テキスト言語が異なるレコード、例えば日本語版の『ハリーポッターと賢者の石』は別のWorksetとしてグループ化されている。

3. xISBNサービス

OCLCのFRBR化に関する研究開発の中で、Worldcat.orgとともに実用化されているのがxISBNサービスである。これは、ISBNを埋め込んだURLをxISBNサーバに投げると、そのISBNと同じ「著作」に紐づけられる他のISBNやメタデータのリストを、XML等の形式で返すものである。利用者は、これにより特定の「著作」に関連付けられる一群のISBN(もしくはメタデータ)が分かるので、ISBNをキーとして、

特定の「著作」に関わる書誌レコードを集めることができる。同サービスで得られるデータはWorldcatから抽出されたもので、xISBNのウェブページによると、17,143,332件のISBNと13,743,771の著作をカバーしている(ともに2008年5月現在)。更に、得られる結果を、特定の図書館の蔵書やWikipediaに含まれるISBNに限定できる機能も提供されている。

当初プロトタイプとして提供されていたxISBNは、2007年2月に正式なサービスとなった。正式サービスと言っても、非営利目的であれば一日500リクエストを超えない範囲で、無料で利用できる。また、APIも公開され、同サービスを使ったアプリケーションとしてxISBNブックマークレット等も提供されている。Web 2.0的なオープンなサービスであるxISBNを検索システムに導入する動きも進んでおり、例えば、インディアナポリス・マリオン・カウンティ公共図書館のOPACや前述したオーストラリア国立図書館のプロトタイプシステム等が導入している。

4. おわりに

本稿で紹介したOCLCの取組みは、FRBRの定義を厳格に適用したものでなく、そのため厳密な意味でのFRBR化とは言えない。FRBRの完全な適用は、目録規則や書誌レコードのフォーマット自体を、FRBRを使って再構築する必要があるため、ハードルがかなり高い。とはいえ既存の書誌レコードを使ったOCLCのFRBR化は、現実の目録においてFRBRが提唱する概念を活用し、これまで不完全であった目録の機能の実現を図っているという点で非常に興味深い。

(総務部企画課：橋詰秋子^{はしづめあきこ})

Ref:

- Pisanski, J. et al. Functional requirements for bibliographic records: an investigation of two prototypes. Program. 2007, 41(4), 400-417p.
 O' Neill, E. "The Impact of Research on the Development of FRBR". Understanding FRBR. Taylor, A, Westport, Library Unlimited, 2007, 59-72p.
 OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records. <http://www.oclc.org/research/projects/frbr/default.htm>. (accessed 2008-03-25).
 xISBN (Web service). <http://www.worldcat.org/affiliate/webservices/xisbn/app.jsp>. (accessed 2008-03-25).
 Worldcat.org. <http://www.worldcat.org/>. (accessed 2008-03-25).