

ISSN 1882-0468

ISSN-L 1882-0468

NDL 書誌情報ニュースレター

2014 年 4 号(通号 31 号)

目次

国際的な書誌調整への参加 —世界図書館・情報会議(IFLA 第 80 回年次大会)、 IFLA セマンティック・ウェブ研究会・情報技術分科会共催サテライト・ミーティング、 VIAF 評議会会議(報告) (関西館 図書館協力課 大柴忠彦)	1
国際的なデータ提供 —日本の書誌データと典拠データを世界に (収集・書誌調整課 吉村風)	7
OCLC の多言語書誌構造化の取組み —利用者にとって最適な表示を目指して (収集・書誌調整課 柴田洋子)	11
コラム:国内発行オンラインジャーナルの ISSN 書誌データが国立国会図書館サーチ で検索できます (逐次刊行物・特別資料課 整理係)	18
コラム:書誌データ利活用(5) —テキストエディタを使った NDL-OPAC ダウンロード ファイルの加工 (収集・書誌調整課 書誌サービス係)	21
掲載情報紹介	27

国際的な書誌調整への参加

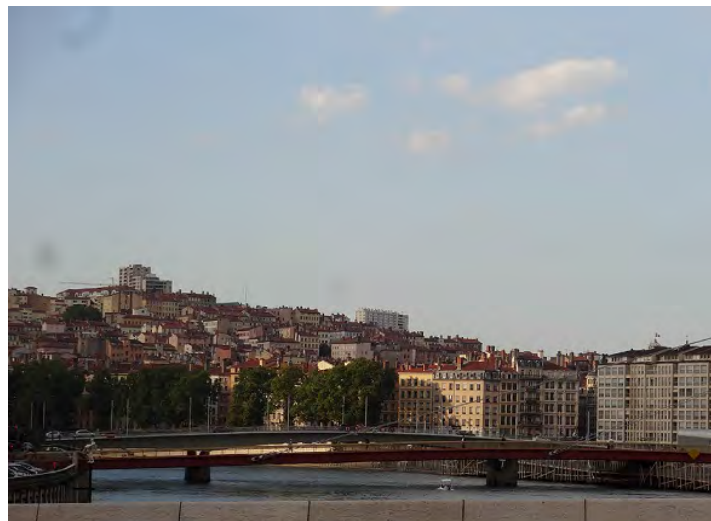
—世界図書館・情報会議(IFLA 第80回年次大会)、
IFLA セマンティック・ウェブ研究会・情報技術分科会共催サテライト・ミーティング、
VIAF 評議会会議(報告)

【はじめに】

「世界図書館・情報会議(国際図書館連盟(IFLA)第80回年次大会)」が2014年8月16日から22日にかけて、フランスのリヨンで開催され、国立国会図書館代表団の一人として参加しました。筆者は、書誌分科会常任委員会に常任委員として出席したほか、関連する分科会の常任委員会やオープン・セッションにも参加しました。

また、大会開催前々日の8月14日にパリで開催されたIFLA セマンティック・ウェブ研究会・情報技術分科会共催サテライト・ミーティングに出席し、[国立国会図書館典拠データ検索・提供サービス\(Web NDL Authorities\)](#)についてペーパー発表を行いました。8月15日には、リヨンで開催されたバーチャル国際典拠ファイル(VIAF)評議会会議に出席しました。

以下に概要を報告します。



リヨンの街並み

1. 世界図書館・情報会議(IFLA 第80回年次大会) [1]

(1) [書誌分科会](#)常任委員会

・全国書誌に係る指針の改訂

書誌分科会の主たる任務は全国書誌に関することです。ここ数年は、2009年刊行の全国書誌に係る指針[2]の改訂について検討を進めています。改訂版は“Best Practice for National Bibliographic Agencies in a Digital Age”と題し、指針に則した各国の全国書誌の事例を集積し、[書誌分科会のウェブサイト](#)にオンライン資料として掲載します。

会議では、改訂版の編集作業の進め方について確認しました。常任委員会内の編集作業チームが作成したプレ

ビュー版を元に、今後、内容の拡充を進めていくこととし、また、常任委員会外の専門家にも執筆やレビューを依頼することとしました。2015年7月には常任委員会メンバーによる確認を終え、8月公開が目標です。公開後も、ウェブページであることの特性を生かして、情報の更新を随時図っていきます。

また、全国書誌における典拠コントロールの重要性をふまえ、狭義の書誌のみならず、典拠に関する内容の充実を図ることも検討していきます。

・“National Bibliographic Register”

書誌分科会では、各国の全国書誌の現況が簡便に把握できるよう、分科会のウェブサイトに [“National Bibliographic Register”](#)（「全国書誌登録簿」）というページを設けて公開しています。この登録簿は、上述の全国書誌の“Best Practice”において利用事例としても位置付けられるべきものであり、両者の関連付けについて今後検討していくこととしました。

また、今期は新規の登録がありませんでしたが、来年はIFLA年次大会開催地が南アフリカ共和国のケープタウンであることに鑑み、特にアフリカ諸国の全国書誌作成機関へ登録を呼びかけていくこととしました。

なお、日本の全国書誌については、2014年3月25日からの全国書誌(電子書籍・電子雑誌編)の提供[3]に伴い、登録簿の情報を更新しています。

・2015年のオープン・セッション

電子書籍等オンライン資料の全国書誌収録促進も、書誌分科会における課題です。収録促進のためには、各国の納本制度におけるオンライン資料への対応[4]が必要です。そこで、2015年のIFLA年次大会のオープン・セッションについて検討し、‘electronic legal deposit’と全国書誌をテーマにプログラムを組むこととしました。また、実施にあたっては、国立図書館分科会へ共催を働きかけることとなりました。

(2) [目録分科会](#)常任委員会および[分類・索引分科会](#)常任委員会

書誌分科会に密接に関連する目録分科会常任委員会および分類・索引分科会常任委員会にオブザーバとして参加しましたが、今年は新しい動きはありませんでした。

目録分科会では、FRBR、FRAD および FRSAD の3モデル[5]の整理統合(Consolidation)作業、「[国際目録原則覚書](#)」の見直し等について、経過報告がありました。また、『国際標準書誌記述』(ISBD)の見直しに関しては、各国の国立図書館等へISBD適用等に関するアンケート調査を行いました。米国ほかいくつかの国立図書館から回答が得られていないため、調査を継続することとなりました。

なお、目録分科会は、大会に先立つ8月13日に、[“Resource Description and Access” \(RDA\)](#)の現状と今後をテーマとしたサテライト・ミーティングをドイツ国立図書館にて開催しました[6]。

分類・索引分科会では、昨年、ジャンル形式用語に関する検討を開始することとしましたが、具体的な進展はまだ見られませんでした。また、そもそも、分科会の名称‘Classification and Indexing’ (分類・索引)が古び

ており、たとえば‘Subject Access’（主題アクセス）という名称に変更してはどうか、という意見が出ていました。

(3) UBC に関するオープン・セッション

今年の大会では、IFLA におけるすべての書誌関連分科会等（目録分科会、書誌分科会、分類・索引分科会および UNIMARC 戦略プログラム）が共催して、“Universal Bibliographic Control in the Digital Age: Golden Opportunity or Paradise Lost?” と題し、国際書誌調整(UBC)をテーマに5時間にわたるオープン・セッションを行いました。

UBC をめぐるさまざまなテーマのペーパー10本の発表がありました。セッション冒頭の OCLC(OCLC Online Computer Library Center)からの基調報告では書誌コントロールにおける典拠の意義が強調され、また、フランス国立図書館からの報告でも「典拠コントロールは戦略的ツール」であると主張される等、全体として書誌調整における典拠の重要性について再認識させられるものでした。

ところで、UBC をテーマにこのようなオープン・セッションを行った背景には、UBC が実質的に IFLA のコア・プログラムではなくなったことに対して、昨年、書誌分科会を中心に [UBC に関する声明](#) をまとめ、IFLA のウェブサイトに掲載したということがあります。すなわち、この声明発表を踏まえて、UBC の重要性をさらに広く訴えていこうというねらいがありました。

しかしながら、それが功を奏したとは言い難いように感じました。セッション全体については散漫な印象を受けましたし、質疑応答もほとんどありませんでした。オープン・セッション後に行われた書誌分科会常任委員会における評価では、ヨーロッパおよび米国からの発表ばかりで(イランからの発表が予定されていたが、当日は都合によりキャンセル)、その地域的な偏りを指摘する意見がありました。

2. IFLA セマンティック・ウェブ研究会・情報技術分科会共催サテライト・ミーティング[7]

(1) ミーティング概要

サテライト・ミーティングは、“Linked Data in Libraries: Let’s make it happen!” と題し、図書館と Linked Data をテーマにフランス国立図書館で行われました。ヨーロッパを中心に、米国、カナダ、チリからも参加があり、参加者は188名(参加者名簿による)にのぼりました。

ミーティングは、午前に2部、午後に2部の4部構成で行われました。第1セッションでは図書館における Linked Data 提供の実践例について、第2セッションでは Linked Data 発展のためのアプリケーション開発について、第3セッションでは Linked Data モデルで使われる語彙の作成・維持管理について等、全体を通して10本のペーパー発表がありました。最後のセッションでは、Linked Data サービスとソフトウェアについて、OCLC およびシステム・ベンダー2社によるオープン・ディスカッションが行われました。

(2) ペーパー発表

筆者は、第1セッションにおいて、典拠データの Linked Data 提供の実践例として、“Web NDL Authorities: Authority Data of the National Diet Library, Japan, as Linked Data” と題し、[Web NDL Authorities](#) につ

いてペーパー発表を行いました。

当館の著者名典拠および件名標目表(NDLSH)について、冊子体や MARC フォーマット等による頒布から Linked Data としての提供へと至った歴史的背景を説明した後、[Web NDL Authorities](#) について、その機能および特徴、また、[VIAF](#) や米国議会図書館件名標目表(LCSH)との連携を紹介しました。最後に、図書館利用者のみならずウェブ環境における利用者へも典拠データを提供することの意義、また、その提供方法としての Linked Data の重要性を述べ、当館として引き続き Linked Data への対応を進めていくことを説明しました。

参加者からは、西洋諸国・言語外での実践例として興味深くとらえられ、特に Web NDL Authorities における NDLSH と LCSH とのリンクについては好評を得ました。さらに、Europeana、ドイツ国立図書館、フランス国立図書館の方々から、欧州図書館の MACS (Multilingual Access to Subjects) [8] との連携可能性を示唆するコメントをいただきました。



サテライト・ミーティングでの当館発表の様子(左端が筆者)

3. VIAF 評議会会議[9]

(1) 議長選挙

次期議長にフランス国立図書館のヴァンサン・ブレ氏(Vincent Boulet)、議長候補(兼副議長)に米国議会図書館のビーチャー・ウィギンズ氏(Beacher Wiggins)が選出されました。

(2) 現況報告および今後の展開

まず、[VIAF](#) の現況について OCLC から報告がありました。VIAF 参加機関は、2014年7月現在、29か国34機関となりました。そのうち国立図書館は24館で、直接の参加ではなく、コンソーシアム等を通じたデータ提供も合わせると、35の国立図書館が貢献しています。

その後、今後の VIAF の展開について議論しました。まず、VIAF への参加基準について検討しました。VIAF のメンバーシップを、国立図書館や国際機関などデータを提供するだけでなく VIAF 評議会へも参加する資格がある

“VIAF Contributors” と、データ提供によって分野や言語等の観点から VIAF の品質向上への貢献を期待できる “Other Data Providers” との二つに分け、それぞれについて参加基準や参加手順を定めることとしました。

また、VIAF データの品質管理に関して、特に、個人名典拠データとして人物の国籍および職業を含めることについて議論しました。特に、国籍の記録方法については、VIAF が多言語データベースであることを考慮すると、文字列よりもコード化情報(国名コード)を入力した方がよいとの意見がありました。

VIAF と ISNI(International Standard Name Identifier、創作者等の名称に関する国際標準識別子)との連携の継続についても確認しました。

(3) Linked Data 戦略

最後に、Linked Data 戦略を議事として採り上げました。OCLC からの発表に続き、評議会前日にパリで開催された IFLA セマンティック・ウェブ研究会・情報技術分科会サテライト・ミーティングのレポートがありました。サテライト・ミーティングでも、VIAF についてはしばしば言及されていました。サテライト・ミーティングでの議論をふまえ、Linked Data を支えるものとしての信頼できる典拠データの必要性が確認されました。

なお、上述のとおりこのサテライト・ミーティングで当館から Web NDL Authorities についてペーパー発表を行ったことを、この場でも報告しました。ここでも、Web NDL Authorities と MACS の連携可能性が話題になりました。VIAF が扱うのは、今のところ、個人名や団体名などの固有名典拠ですが、一方で、MACS が件名典拠版 VIAF となる可能性も示唆されました。

【おわりに】

今年の IFLA 大会では、昨年の大会[10]同様、典拠の重要性について再認識させられました。典拠の重要性は、UBC に関するオープン・セッションのみならず、書誌分科会常任委員会における議論においてもしばしば強調されました。

当館の [Web NDL Authorities](#) はウェブ環境に適した典拠データの提供として、先駆的な実践です。サテライト・ミーティングにおけるこの Web NDL Authorities についてのペーパー発表は、昨年の VIAF 評議会会議[11]における発表に続き、少なからぬ反応がありました。Web NDL Authorities を国内外にさらにアピールするとともに、そのデータ内容やシステム機能をより充実させていくべきものと考えます。

また、当館は、東アジアで最初の参加機関として VIAF に参加し、定期的に典拠データを提供しています。VIAF 評議会会議にも毎年出席し、発表等も行ってきました。こうして、当館は、特に典拠の側面から、国際的な書誌調整に着実に参加してきているといえます。

Web NDL Authorities と VIAF を手掛かりとして、IFLA 等を通じて、今後も国際的な書誌調整に参加していくことができますし、また、積極的に参加していくべきだと考えます。

大柴 忠彦

(おおしば ただひこ 関西館 図書館協力課)

- [1] 大会プログラム、発表ペーパー等は、次に掲載されています。
<http://conference.ifla.org/past-wlic/2014/ifla80.html>, (参照 2014-11-10).
- [2] 国立国会図書館収集書誌部訳「デジタル時代の全国書誌：指針および新しい方向性」を当館ホームページで公開しています。
<http://www.ndl.go.jp/jp/library/data/kokusai.html#anchor04>, (参照 2014-11-10).
- [3] 全国書誌(電子書籍・電子雑誌編)の提供については、本誌2014年2号(通号29号)でご紹介しています。
http://dl.ndl.go.jp/view/download/digidepo_8684166_po_2014_2.pdf?contentNo=1#page=2,
(参照 2014-11-10).
- [4] 当館の対応については、以下のページをご覧ください。
<http://www.ndl.go.jp/jp/aboutus/online/index.html>, (参照 2014-11-10).
- [5] FRBR(Functional Requirements for Bibliographic Records)、FRAD(Functional Requirements for Authority Data)、FRSAD(Functional Requirements for Subject Authority Data)
- [6] 大会プログラム、発表ペーパー等は、次に掲載されています。
<http://www.dnb.de/DE/Standardisierung/International/iflaSatelliteMeeting.html>, (参照 2014-11-10).
また、このミーティングについては、以下の記事に部分的な紹介があります。
柴田洋子. ウェブで広がる図書館のメタデータを目指して—RDAとBIBFRAME. カレントアウェアネス. 2014, (322), <http://current.ndl.go.jp/ca1837>, (参照 2014-12-20).
- [7] 大会プログラム、発表ペーパー等は、次に掲載されています。
<http://ifla2014-satdata.bnf.fr/>, (参照2014-11-10).
また、このサテライト・ミーティングについては、以下に報告があります。
竹鼻和夫. 図書館におけるLinked Data：実現させよう！<報告>. カレントアウェアネス-E. 2014, (268), <http://current.ndl.go.jp/e1618>, (参照 2014-11-10).
- [8] 英語(LCSH)、ドイツ語(SWD)、フランス語(RAMEAU)の件名標目表をリンクさせるプロジェクト。以下の記事などを参照。
多言語シソーラスの構築と開発のためのガイドライン. カレントアウェアネス-E. 2009, (146), <http://current.ndl.go.jp/e904>, (参照 2014-11-10).
- [9] 会議の議事次第、資料等は、次に掲載されています。
<http://www.oclc.org/events/2014/viaf-ifla-2014.en.html>, (参照 2014-11-10).
- [10] 2013年のIFLA大会については、本誌2013年4号(通号27号)にて報告しています。
http://dl.ndl.go.jp/view/download/digidepo_8379163_po_2013_4.pdf?contentNo=1#page=5, (参照 2014-11-10).
- [11] 2013年のVIAF評議会会議については、本誌2013年4号(通号27号)にて報告しています。
http://dl.ndl.go.jp/view/download/digidepo_8379163_po_2013_4.pdf?contentNo=1#page=10,
(参照 2014-11-10).

国際的なデータ提供

—日本の書誌データと典拠データを世界に

【はじめに】

国立国会図書館の重要な任務の一つに、海外の図書館や関係機関に対して協力を行うというものがあります。たとえば、海外図書館からのレファレンスへの回答や当館刊行物の送付、そして国際図書館連盟(IFLA)への参加やアジア各国の図書館との業務交流など、さまざまな事業を行っています。また、2012年には、「[世界の国立図書館、議会図書館、図書館関係その他の国際機関と密接に連携して、情報の共有・交換に努めるとともに、電子情報時代における海外との協力事業を推進します](#)」という目標を設定し、国際協力事業をさらに推進しています。この目標に関連して、書誌データや典拠データを、日本だけでなく世界中の人々に活用してもらえよう、私たちが世界の関係機関にデータを提供しています。

本誌前号の「[ひろがる、つながる書誌情報 —「NDL 書誌情報ニューズレター」から見たこれまでとこれから](#)」では、当館作成の書誌データ・典拠データが世界とつながる様子を、本誌広報犬 CANE(カーネ)君が説明してくれました。

本稿では、「国際的なデータ提供」と題し、日本の書誌データと典拠データがどのように世界に提供されているかをご紹介します。

1. 書誌データ・典拠データの国際的な提供の変遷

現在当館が書誌データ・典拠データを提供している海外の機関は、次のとおりです。

- 国際連合教育科学文化機関(UNESCO) (以下、ユネスコといいます)
- ISSN 国際センター[1]
- OCLC(OCLC Online Computer Library Center)

表1は、当館の書誌データや典拠データの国際的な提供の変遷をまとめたものです。

表1 書誌データ・典拠データの国際的な提供の変遷

1950年	ユネスコの「 Index Translationum 」(世界翻訳書目録)へ、外国語で書かれ日本語に翻訳された資料の書誌データを提供開始。
1976年	ISSN 国際センター 開設(当時の名称は ISDS 日本センター[2])。2002年に現在の名称に改称)。ISSNを登録した書誌データを ISSN 国際センター(当時の名称は ISDS 国際センター。1993年に現在の名称に改称)へ提供。
2010年	JAPAN/MARC(M)を OCLC に提供することを、OCLC と合意。OCLC が維持する書誌データベース WorldCat を通じて、世界中に公開開始。
2012年	VIAF(バーチャル国際典拠ファイル)への参加について、OCLC と協定を締結。JAPAN/MARC(A)を VIAF にて公開開始。
2013年	OCLC に提供する書誌データの範囲を拡大。JAPAN/MARC(M)に加えて、JAPAN/MARC(S)、雑誌記事索引データを WorldCat にて公開開始。
2014年	JAPAN/MARC(M)と JAPAN/MARC(S)を JAPAN/MARC(M/S)として一本化し、OCLC に提供。引き続き WorldCat にて公開。

当館では、1950年から、ユネスコが運営している「[Index Translationum](#)」(世界翻訳書目録)にデータを提供し、国際貢献に努めてきました。この「Index Translationum」は、1932年に始まった、国際連盟の図書普及事業の一つです。このデータベースは世界中の国立図書館が提供したデータによって構成され、世界各国の翻訳書を検索することができるものです[3]。

1976年には当館に[ISSN 日本センター](#)(当時の名称はISDS 日本センター)が開設されました。ISSNとは、逐次刊行物を識別するための国際的なコード番号です。ISSN 日本センターでは、発行者からの申請に基づいて、日本国内で発行される出版物にISSNを付与します。そして、ISSNを付与した資料の書誌データをISSN国際センターに送信します。送信された書誌データは、ISSN国際センターのデータベース(ISSN Register)に登録されます[4]。

また、世界各国の図書館が参加する組織として、[OCLC](#)があります[5]。当館は2010年にOCLCと覚書を交わし、当館の書誌データと典拠データを世界に公開してきました[6]。

2012年にJAPAN/MARC(A)の提供を開始しましたが、それに先立ち2008年に一度、参加申請とサンプルデータの送付を行っています。その時は、文字コードがVIAFのシステムに対応しておらず、参加できませんでした。2012年になって、MARC21フォーマット、Unicodeに対応した当館の典拠データをOCLCに改めて送付し、協定を締結して10月からVIAFに参加することとなりました。

2. 書誌データ・典拠データ提供の現状

表2は書誌データ・典拠データ提供の現状をまとめたものです。

表2 書誌データ・典拠データ提供の現状

提供先	提供データ	提供開始年	総提供件数	提供頻度	提供データの種類・収録期間など
WorldCat	JAPAN/MARC (M)	2010	約 450 万	週 1 回	【データの種類】 図書、視覚障害者用資料、電子出版物、地図および音楽録音・映像資料の書誌データ。 【収録期間】 明治～現在。 【備考】 2014年度から JAPAN/MARC (S)と一本化して、JAPAN/MARC (M/S)として提供。
	JAPAN/MARC (S)	2013	約 15 万	週 1 回	【データの種類】 逐次刊行物(雑誌・新聞・通信、年鑑・年報など)の書誌データ。 【収録期間】 明治～現在。 【備考】 2014年度から JAPAN/MARC (M)と一本化して、JAPAN/MARC (M/S)として提供。
	雑誌記事索引	2013	約 1,100 万	週 1 回	【データの種類】 国内刊行和文雑誌の記事索引。 (一部外国刊行和文雑誌・国内刊行欧文雑誌を含む。)
VIAF	JAPAN/MARC (A)	2012	約 100 万	週 1 回	【データの種類】 名称典拠(「個人名」「家族名」「団体名」「地名」「統一タイトル」)
ISSN 国際センター	ISSN の書誌データ	1976	約 4 万	月 1 回	【データの種類】 ISSN 日本センターで ISSN を付与した資料の書誌データ。
ユネスコ Index Translationum (世界翻訳書目録)	翻訳書の書誌データ	1950	約 13 万	年 1 回	【データの種類】 前年に発行された図書のうち、外国語で書かれ日本語に翻訳された資料の書誌データ。 【備考】 2014 年は送付休止。

WorldCat と VIAF は、ともに OCLC が運営しています。

[WorldCat](#) は世界各国の図書館の書誌データを取り込んだ、世界最大の書誌データベースです。誰でも WorldCat 上で当館書誌データを検索することができます。海外の図書館等で日本語資料を整理するときに、日本語に通じたカタログがいなくても、OCLC の参加館であれば、当館が作成した書誌データをコピーして自館のデータを作ること(コピーカタログ)ができます[7]。

[VIAF](#) は、各国の図書館が作成した典拠データをもとにした国際的な典拠ファイルです。VIAF では、各言語による典拠がひとかたまりで表示され、世界各国の誰もが使いやすい形で、典拠レコードを共有することが可能になっています。

当館が参加することで、世界中の書誌作成機関において、日本人の著者や日本の団体著者の同定識別が容易になりました[8]。

3. 書誌データ・典拠データ提供の実務

こうした国際的なデータ提供を、当館がどのように行っているのかをご紹介します。

OCLC に提供する JAPAN/MARC(M/S) と JAPAN/MARC(A) を例として実際の作業を見てみましょう。

OCLC、VIAF へは、毎週 1 回月曜日に、1 週間分のデータを送っています。前々週の金曜日から前週の木曜日にかけて作成・更新したデータを前週金曜日夜に抽出します。

作成ルールに則っているか、項目の漏れがないか、正しい記号やコード値を使用しているか等、当館担当者がデータの品質を確認した後、OCLC および VIAF の FTP サーバにデータを置き、先方の担当者に、抽出期間、総件数、内訳件数を英文メールで連絡します。

先方の担当者はそのメールを見て、FTP サーバからデータを取得し、JAPAN/MARC(M/S) を WorldCat へ、JAPAN/MARC(A) を VIAF へ反映する仕組みとなっています。

OCLC へのデータ提供は、当初は年数回程度でしたが、徐々に回数を増やし、今では週 1 回行うようになりました。

先方の担当者とのやり取りは英語で行うので、英語の苦手な筆者は四苦八苦しながら作業していますが、提供したデータが先方のデータベースに反映されたのを確認すると、海外旅行に出たわが子が旅先に無事ついてくれた時のような安心した気持ちになります。

【おわりに】

インターネットの普及、電子出版のひろがり、あるいはクールジャパン戦略の推進と日本のコンテンツの海外紹介などにより、海外から日本の出版物の情報を知りたいというニーズは、一段と高まってきています。

当館で現在行っている国際的なデータ提供は、NDL-OPAC や国立国会図書館サーチなどとともに、まさしく海外から日本の情報にアクセスしたいというニーズに応えるものです。当館では、このほど[全国書誌データをご紹介するパンフレット](#)を作成し、ホームページに掲載しました。このパンフレット『[全国書誌データをご利用ください](#)』(PDF: 408KB)の表紙には、日本の書誌データ・典拠データ(六角形のデザインが各種データを示しています)が日本から世界へとつながっていく図を配しました(図 1)。

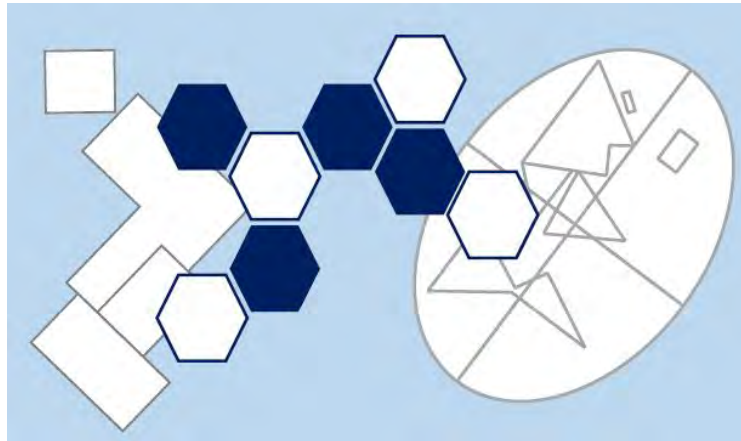


図1 パンフレット『全国書誌データをご利用ください』に使用した図

この図が示すように、今後も、書誌データと典拠データが日本と世界のさらなる架け橋になるよう、努力していきたいと考えています。

吉村 風

(よしむら かぜ 収集・書誌調整課)

[1] ISSN(International Standard Serial Number: 国際標準逐次刊行物番号)

[2] ISDS(International Serials Data System: 国際逐次刊行物データシステム) は、ISSN を国際的に管理する ISSN ネットワークの旧称。

[3] ユネスコ「[Index Translationum](#)」([世界翻訳書目録](#))については、本誌2012年1号(通号20号)のコラムでもご紹介しています。

http://dl.ndl.go.jp/view/download/digidepo_3487126_po_2012_1.pdf?contentNo=1#page=11, (参照 2014-11-7).

[4] 国内出版物の ISSN は、オンラインジャーナルを除き、[NDL-OPAC\(国立国会図書館蔵書検索・申込システム\)](#)で確認することができます。

国内発行オンラインジャーナルについては、[国立国会図書館サーチ](#)で確認できます。詳しくは、本誌今号のコラム「[国内発行オンラインジャーナルの ISSN 書誌データが国立国会図書館サーチで検索できます](#)」をご参照ください。

[5] 現在、170 개국・地域の 72,000 以上の機関が参加しています。

<https://oclc.org/news/media-kit/boilerplate.en.html>, (参照 2014-11-7).

[6] OCLC を通じた国立国会図書館作成書誌データ(JAPAN/MARC)の国際的提供について

http://www.ndl.go.jp/jp/library/data/oclc_agreement.html, (参照 2014-11-7).

[7] 日本語資料の書誌データを OCLC へ提供している機関は複数あります。そのため、WorldCat でヒットした日本語資料の書誌データでも、当館が作成したものではない場合があります。

[8] VIAF の概要、当館が参加した経緯、VIAF の使い方については、本誌2012年4号(通号23号)から2013年2号(通号25号)までの連載記事「[典拠の国際流通—バーチャル国際典拠ファイル\(VIAF\)への参加](#)」でご紹介しています。

http://dl.ndl.go.jp/view/download/digidepo_4059584_po_2012_4.pdf?contentNo=1#page=12, (参照 2014-11-7).

http://dl.ndl.go.jp/view/download/digidepo_8103221_po_2013_1.pdf?contentNo=1#page=22, (参照 2014-11-7).

http://dl.ndl.go.jp/view/download/digidepo_8226998_po_2013_2.pdf?contentNo=1#page=17, (参照 2014-11-7).

OCLC の多言語書誌構造化の取組み

—利用者にとって最適な表示を目指して

[本誌前号](#)で予告しました、世界最大の図書館サービス組織 [OCLC\(OCLC Online Computer Library Center, Inc.\)](#) による多言語書誌レコードの構造化の取組みについてご紹介します[1]。

OCLC の研究部門である OCLC Research では、[WorldCat](#) のデータの品質向上や、OCLC 各種サービスにおける表示画面の改善を目指し、多言語書誌レコードを活用した取組みを行っています。WorldCat には、世界中の参加館から提供された 3 億件以上の書誌レコードが収録されています。そのうち、英語以外の言語で作成された書誌レコードが半数以上を占めています(図 1)。

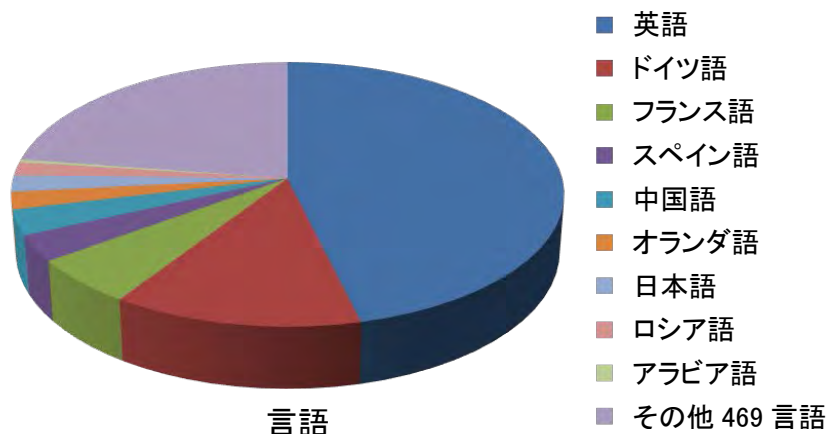


図 1 WorldCat 収録レコードの言語の割合[2]

書誌レコードを作成する際のルールとして、「資料にあるがままに記録する」(転記の原則)というものがあります。たとえば、タイトルは原則として資料にあるがままに記録されるため、同じ資料であれば、どの機関が作成した書誌レコードでも、同じ言語で記録されます。一方で、資料に表示されていない概要や件名(主題)、注記等の場合、書誌レコードの作成に使われる言語(目録用言語)はさまざまです。

また、一つの著作に対してさまざまな言語による翻訳書があります。たとえば、夏目漱石の『吾輩は猫である』の場合、原作と翻訳書では本文の言語が異なるだけでなく、タイトル(『I am a cat』『Je suis un chat』等)の言語も異なります。

このように、ひとことで多言語書誌といっても、目録用言語がさまざまな場合や、資料本文の言語がさまざまな場合があります。WorldCat に収録されているそれらの多言語書誌は、同じ資料に対する書誌レコードであったり、同じ原作の翻訳書の書誌レコードであったりと、データ間に何らかの関連性をもっています。

そこで、OCLC Research では、この大量の多言語の書誌レコードを解析し、[FRBR](#) でいう同じ「著作」ごとにグルーピング[3]したレコードの中から、利用者にとって最適な言語や文字の書誌情報を WorldCat の画面[4]で表示できるような「多言語書誌構造化」プロジェクトに取り組んでいます。

【利用者にとって最適な表示を目指して】

WorldCat には各国語版の画面があり、各項目の見出しの表示言語が選べます。たとえば、日本語の表示を選ぶと、イギリスの物理学者スティーヴン・ホーキング博士の著書“The Grand Design”の書誌情報は、図2のように表示されます。

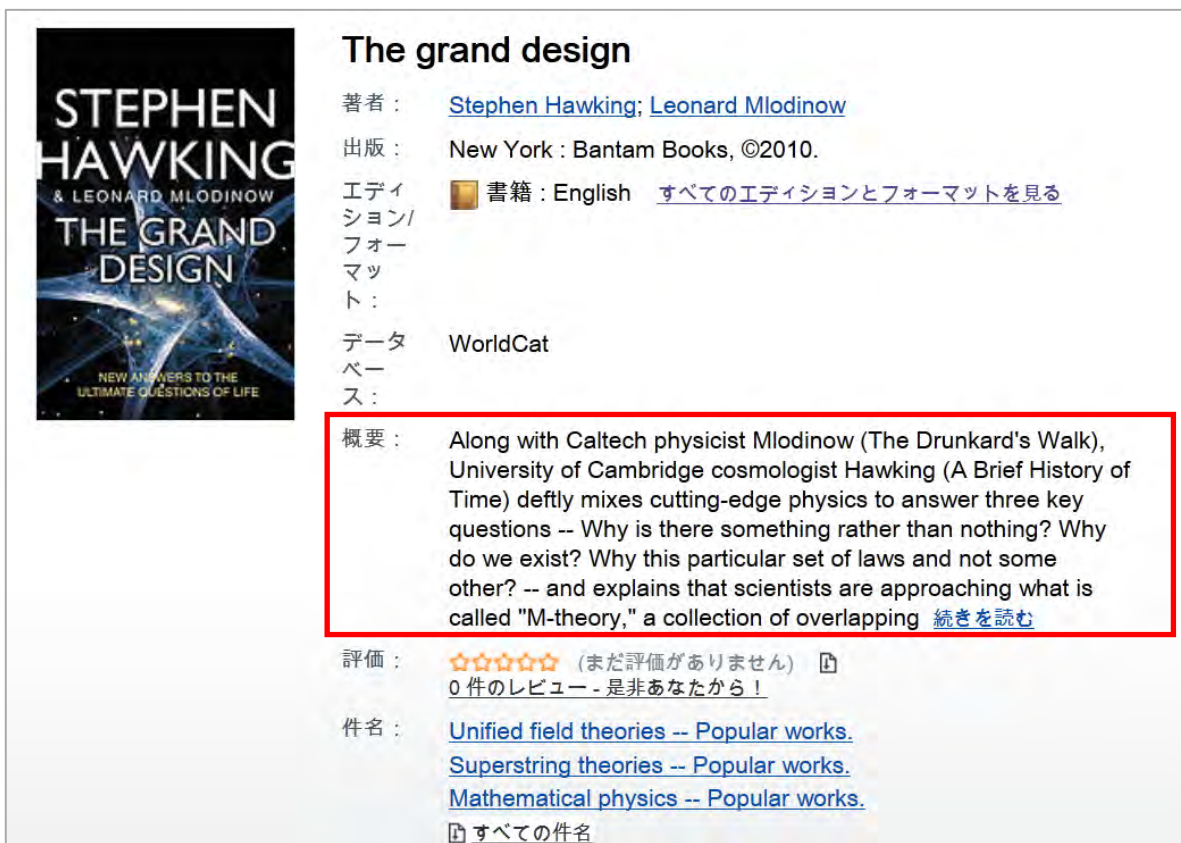


図 2 現在の WorldCat の書誌情報詳細表示画面(日本語) [5]

多言語書誌構造化の取組みが進むと、WorldCat 内のある「著作」に関する多言語書誌レコードのかたまりの中から、選択した表示言語と同じ目録用言語で作成されたデータを集めて、表示できるようになります。図3は、OCLC が考える利用者にとっての最適な表示を実現したイメージ画面例です。たとえば、日本語の画面では目録用言語の英語で表示されていた概要(図2赤枠参照)が、図3では、選択した表示言語と同じイタリア語で表示されるようになります。件名(主題)も同様です(図3赤枠①②参照)。これは、さまざまな目録用言語による“The Grand

Design”の書誌レコードや、各言語による翻訳書の書誌レコードから、それぞれの言語に関するデータ等をもとに、イタリア語の画面を見ている利用者にとって最適な表示を構成するデータが自動的に選択されています。実験段階のため詳細な情報が公開されていませんが、今後の動向を注目していきたい取組みです。



図 3 WorldCat の書誌情報表示画面イメージ(イタリア語)[6]

【翻訳書のグループ化[7]】

多言語書誌構造化プロジェクトでは、さまざまな言語で翻訳された「著作」に重点が置かれています。その一環として、WorldCat の多言語書誌レコードの解析・グルーピングの成果を活用し、[バーチャル国際典拠ファイル\(VIAF\)](#)で翻訳書のグルーピングの精度を向上させる取組みが行われています。

VIAF には、原作のタイトルとその著者から成る「著作の統一タイトル」レコードと、その翻訳書のタイトルと翻訳者からなる「表現形」レコード(それぞれ [FRBR](#) という「著作」と「表現形」)が登録されています。これは、WorldCat の各参加館から提供された書誌レコードをもとに生成されたものです。これらのレコードには“xR”という目印(フラグ)が付けられており、VIAF の参加館が提供している典拠レコードと識別することができます。

この“xR”付きの典拠レコードがVIAFでどのように用いられているか、村上春樹の『色彩を持たない多崎つくると、彼の巡礼の年』を例に見てみましょう[8]。まず、VIAFの[村上春樹の典拠レコード](#)を確認します(図4)。「Works」欄に表示されている村上春樹の著作の一覧から、『色彩を持たない多崎つくると、彼の巡礼の年』のリンクをクリックすると、「[著作の統一タイトルレコード](#)」が表示されます(図5)。画面の「表現形」欄を見ると、オランダ語、ドイツ語、イタリア語…と、この作品がさまざまな言語で翻訳されていることがわかります。さらに、たとえば、イタリア語のタイトル『L'incolore Tazaki Tsukuru e i suoi anni di pellegrinaggio』をクリックすると、翻訳言語、原作のタイトルと著者、そして翻訳者等がひとまとまりになった表現形のレコード「[村上, 春樹, \(1949-\). | 色彩を持たない多崎つくると、彼の巡礼の年 Italian | \(Pastore : 2014\)](#)」を確認することもできます。

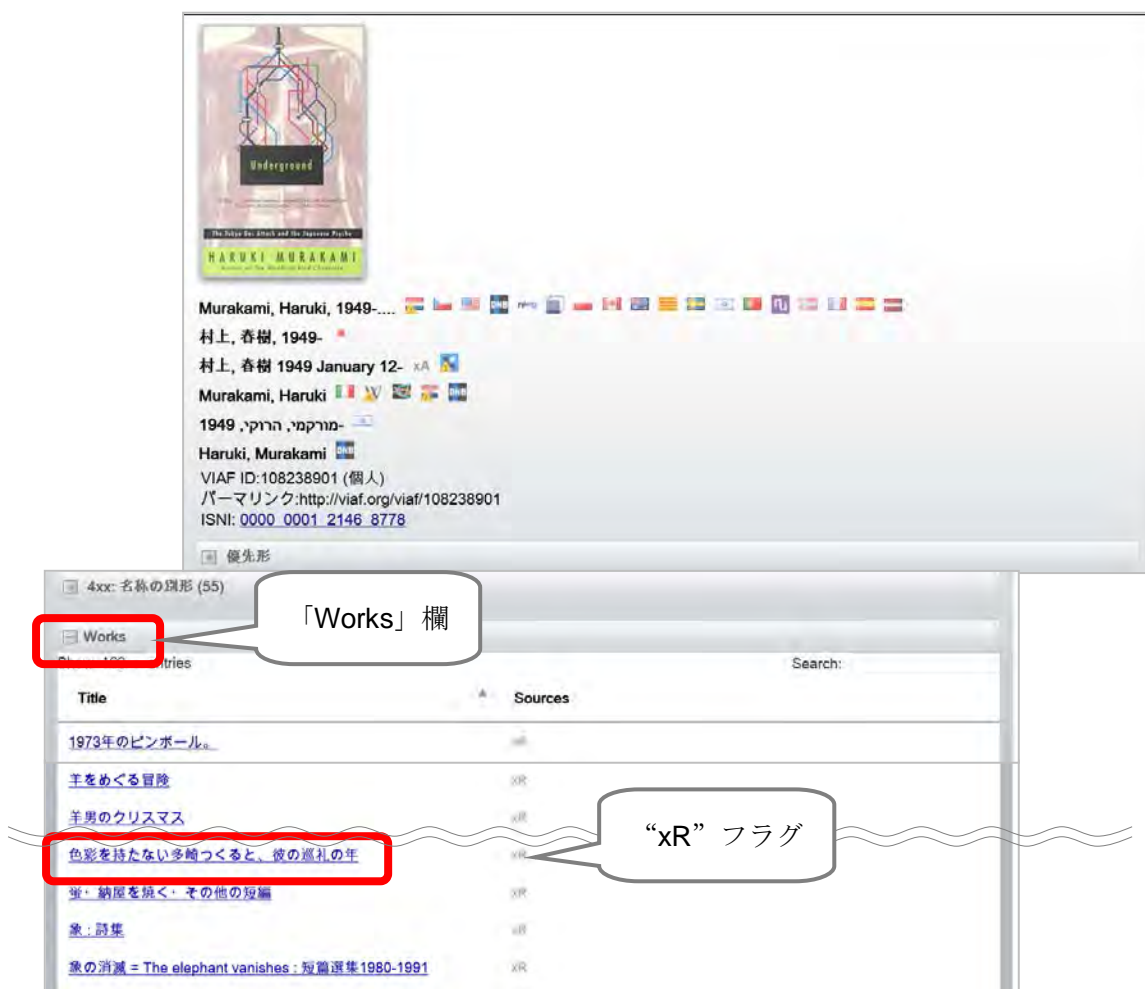


図4 VIAFの詳細表示画面(個人の典拠レコード)

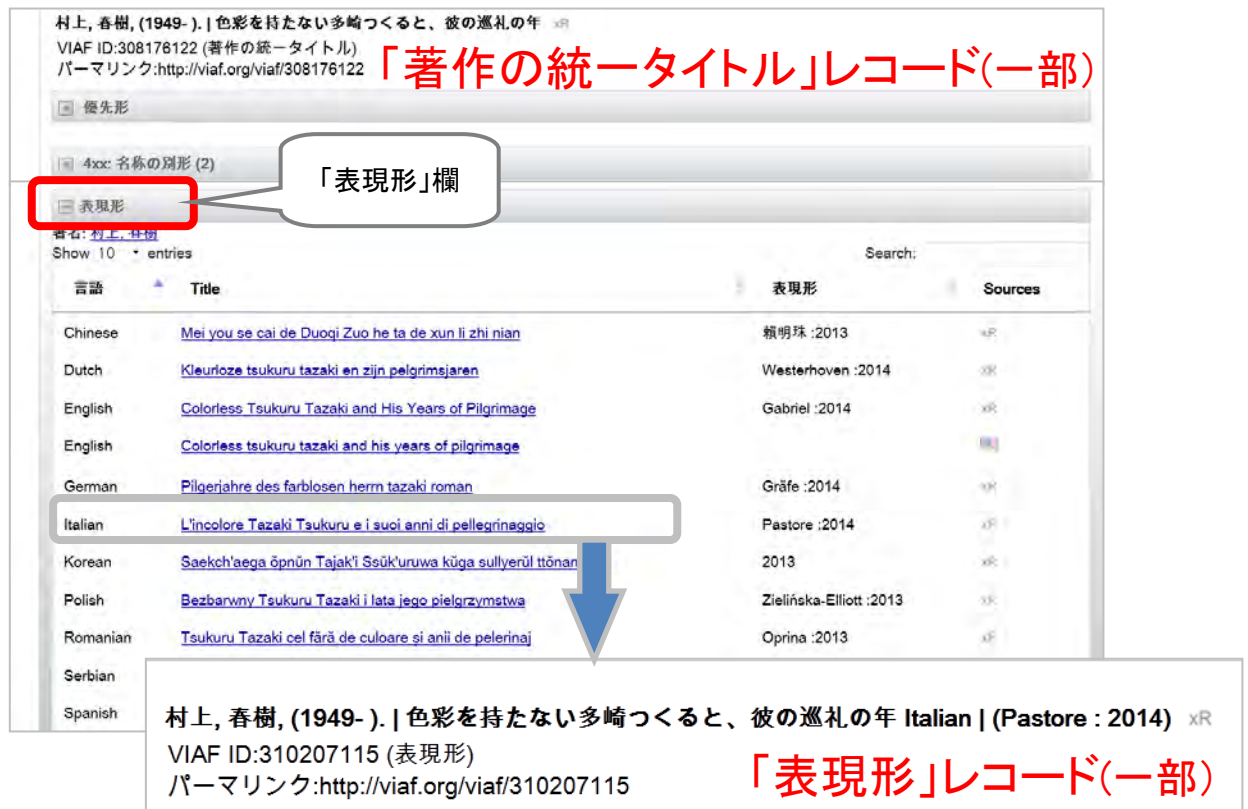


図 5 VIAF の詳細表示画面(著作の統一タイトルから表現形のレコードへ)

また、海外の著作について、日本語の翻訳書があるかどうか確認することもできます。2014年5月までに、原作と翻訳書をあわせて約100万件の“xR”付きの典拠レコードがVIAFに追加されています。この“xR”付き典拠レコードが追加されることで、原作とリンクされた翻訳書のレコードの数が増え、より多くの「著作」について、さまざまな言語で書かれた翻訳書が一覧できるようになります。また、WorldCatに収録されている3億件以上のレコードのうち、たとえば、英語以外の作品が英語に翻訳されている割合はどのくらいか？最も多くの言語に翻訳されている作品は何か？著者は誰か？といったことが調べられるようなサービスの実現に向けた開発も進められています。現在、「著作」に関する情報(「著作の統一タイトル」レコード)がさらに見やすくなるように、表示画面を改善中です。

【多言語書誌構造化を支える取組み】

多言語書誌構造化の取組みは、OCLC Researchによる他のプロジェクト「[著作レコード](#)」や“GLIMIR”(Global Library Manifestation Identifier)の成果を反映しながら進められています。そこで、最後にこの二つのプロジェクトを簡単にご紹介します。

・著作レコード:「著作」レベルのリッチなレコードの作成

OCLCが考える利用者に最適な表示の実現に向け、FRBRでいう「著作」に関する情報を充実させるためには、より網羅的かつ正確なグルーピングを「著作」のレコードごとに行う必要があります。鍵となるグルーピングは、OCLC Researchが開発したFRBR Work-Setアルゴリズムに基づいています[9]。このアルゴリズムは、WorldCatの書誌レコードのタイトルと著者の情報を基に、同一の「著作」に関連付けられる書誌レコード(FRBRの「体現

形)を機械的に判断するものです。この結果、workset[10]と呼ばれる「著作」ごとのグループができます。workset内のすべてのデータを解析し、ある「著作」に関するデータを集約することで、個々の書誌レコードよりもさらにリッチなレコードとして利用者に提供することができるようになります。このアルゴリズムを使ったWorldCatにおける「著作」のレコードの充実化を図る実験プロジェクト[11]の成果が、2011年にプロトタイプシステムとして公開されました。これらを基に、さらに精度の向上を目指しながら、WorldCatの表示画面の改善等に活用される予定です。

・GLIMIR：体現形レベルの表示の改善

GLIMIRは、WorldCatのデータの品質改善に向けた試みの一つで、グルーピングした書誌レコード(FRBRの「体現形」)のかたまりごとに識別子を与えるものです[12]。同じ資料に対して異なる言語で作成されたレコードだけでなく、同じ内容でも冊子体やオーディオブック、電子書籍のように形態が異なる資料のレコードについてもグループ化しています。プロジェクトは2009年に開始され、その成果は、重複している書誌レコードの検出・解消や、FRBR Work-Set アルゴリズムの改善に活用されています。また、WorldCatや、WorldCat参加館の目録作成サービス“Connexion”に順次取り入れられており、一件の書誌レコードから、さまざまな言語や形態の資料にたどりつける表示画面が実現されています。今後は、VIAFの表示画面の改善にもその成果が活用される予定です。

OCLCでは、これらのさまざまな取組みを通し、多言語書誌レコードの活用可能性を模索しています。取組みの効果をあげるには、活用する書誌レコードの量的な拡充も必要ですが、レコードそのものの品質の向上も重要です。レコードの品質は世界中に存在する各提供機関に拠るところが大きいため、この多言語書誌構造化の取組みは、国際的な連携プロジェクトといえるかもしれません。現在のWorldCatでは、書誌レコードを中心に構造化されたデータが表示されていますが、今後は、目録用言語や本文の言語の情報によって関連付けられる「著作」中心のデータの構造化が目指されています。一機関が作成する書誌レコード、典拠レコードの利用だけでは限界がありますが、OCLCを通じて各機関のレコードが共有され、言語の情報によってデータ間に新たな関連性が生まれることで、さらなるサービスの可能性が期待できます。

柴田 洋子

(しばた ようこ 収集・書誌調整課)

[1] 下記のページと講演スライドを参考にしています。

OCLC Research. Multilingual Bibliographic Structure,

<http://oclc.org/research/activities/multilingual-bib-structure.html>, (参照 2014-11-10).

Janifer Gatenby. Multilingual WorldCat,

<http://www.slideshare.net/JaniferGatenby/multilingual-presentation-ifla-2013-0819>, (参照 2014-11-10).

Janifer Gatenby. Multilingualism in WorldCat and VIAF,

<http://www.slideshare.net/JaniferGatenby/multilingualism-ifla-2014-08>, (参照 2014-11-10).

[2] 下記スライド2枚目から筆者が和訳しました。

Janifer Gatenby. Multilingualism in WorldCat and VIAF,

<http://www.slideshare.net/JaniferGatenby/multilingualism-ifla-2014-08>, (参照 2014-11-10).

[3] この記事では「クラスタリング」と同様の意味で使っています。クラスタリングとは、簡単にいえば、類似しているデータごとにかたまり(クラスタ)に分けること、つまりグループ化することです。

[4] この記事でご紹介する WorldCat の画面は、厳密には、データベースである WorldCat をウェブ上で公開している WorldCat.org ですが、区別せずに表記しています。

[5] 出典は下記の資料ですが、赤枠を追記するなど筆者が加工しています。

Janifer Gatenby. Multilingualism in WorldCat and VIAF,

<http://www.slideshare.net/JaniferGatenby/multilingualism-ifla-2014-08>, (参照 2014-11-10).

[6] [5]と同じ。

[7] VIAF におけるグルーピングの詳細については、下記をご覧ください。

Thomas B. Hickey and Jenny A. Toves, Managing Ambiguity In VIAF. D-Lib Magazine. 2014, 20(7/8),

<http://www.dlib.org/dlib/july14/hickey/07hickey.html>, (参照 2014-11-10)

[8] VIAF の画面や検索方法の詳細な説明は、本誌の連載記事「典拠の国際流通—バーチャル国際典拠ファイル(VIAF)への参加」の1回目(2012年4号(通号23号))および第2回目(2013年1号(通号24号))をご覧ください。また、この記事では日本語版の画面をもとに説明していますが、各言語版についてはブラウザの設定で変更することができます。

http://dl.ndl.go.jp/view/download/digidepo_4059584_po_2012_4.pdf?contentNo=1#page=12, (参照 2014-11-7).

http://dl.ndl.go.jp/view/download/digidepo_8103221_po_2013_1.pdf?contentNo=1#page=22, (参照 2014-11-7).

[9] OCLC Research. FRBR Work-Set Algorithm,

<http://www.oclc.org/research/activities/frbralgorithm.html?urlm=159780>, (参照 2014-11-10).

また、このアルゴリズムの研究開発については、下記の中で日本語でわかりやすく解説されています。

橋詰秋子. OCLC の FRBR 化の取組み : xISBN サービスを中心に. カレントアウェアネス. 2008, (296), p.10-11,

<http://current.ndl.go.jp/cal665>, (参照 2014-11-10).

[10] Work-Set、work set 等さまざまな表記がありますが、この記事では、下記のページにあわせて「workset」と表記しています。

OCLC Research. Multilingual Bibliographic Structure,

<http://oclc.org/research/activities/multilingual-bib-structure.html>, (参照 2014-11-10).

[11] OCLC Research. Work Records in WorldCat,

<http://www.oclc.org/research/activities/workrecs.html>, (参照 2014-11-10).

[12] Janifer Gatenby et al. GLIMIR: Manifestation and Content Clustering within WorldCat. Code4Lib Journal.

2012. 17, <http://journal.code4lib.org/articles/6812>, (参照 2014-11-10).

コラム:国内発行オンラインジャーナルの ISSN 書誌データが 国立国会図書館サーチで検索できます

国立国会図書館は [ISSN 日本センター](#)として活動し、日本国内で刊行される逐次刊行物に ISSN(International Standard Serial Number: 国際標準逐次刊行物番号)を付与しています。逐次刊行物には、雑誌、新聞、紀要などの冊子体や、CD-ROM、DVD-ROM などのパッケージ系電子出版物だけでなく、オンラインジャーナルなどのオンライン出版物も含まれます。

2014年からは[国立国会図書館サーチ](#)で、ISSN登録手続きが完了した国内発行オンラインジャーナルの書誌データを検索できるようになりました[1][2]。

[国立国会図書館サーチ](#)の検索画面から、タイトルや出版者名、ISSN等で検索を行ってください。



図1 国立国会図書館サーチ簡易検索画面

検索結果数が多い場合は、「検索結果一覧」画面左側の「検索結果の絞り込み」で、資料種別＝「デジタル資料」、データベース＝「国立国会図書館蔵書」、所蔵館＝「国立国会図書館」を選択してください(図2参照)[3]。



図2 検索結果一覧画面

詳細画面右側の「見る・借りる」欄に「請求記号：ISSN」と表示されているものが、国内発行オンラインジャーナルの ISSN 書誌データです(図3参照)[4]。



図3 検索結果詳細画面

詳細画面にはオンラインジャーナルの URL が表示されていませんが、詳細画面右下の「検索結果を出力」の「書誌情報を DC-NDL (RDF) で出力」を選択すると、「<owl:sameAs rdf:resource=」に続いて表示されます(図4参照)。

```

-<rdf:RDF>
-<dcndl:BibAdminResource rdf:about="http://iss.ndl.go.jp/books/R100000002-1024844618-00">
  <dcndl:catalogingStatus>C7</dcndl:catalogingStatus>
  <dcndl:bibRecordCategory>R100000002</dcndl:bibRecordCategory>
  <dcndl:record rdf:resource="http://iss.ndl.go.jp/books/R100000002-1024844618-00#material"/>
</dcndl:BibAdminResource>
-<dcndl:BibResource rdf:about="http://iss.ndl.go.jp/books/R100000002-1024844618-00#material">
  <dcterms:identifier rdf:datatype="http://ndl.go.jp/dcndl/terms/ISSN">2188-0972</dcterms:identifier>
  <dcterms:identifier rdf:datatype="http://ndl.go.jp/dcndl/terms/ISSNL">2188-0972</dcterms:identifier>
  <rdfs:seeAlso rdf:resource="http://iss.ndl.go.jp/issn/21880972"/>
  <rdfs:seeAlso rdf:resource="http://iss.ndl.go.jp/issn/21880972"/>
  <owl:sameAs rdf:resource="http://niche-life.com/issue.html"/>
  <dcterms:title>ニッチェ・ライフ = Niche life</dcterms:title>
-<dc:title>
-<rdf:Description>
  <rdf:value>ニッチェ・ライフ = Niche life</rdf:value>
  <dcndl:transcription>ニッチェ ライフ = Niche life</dcndl:transcription>
</rdf:Description>
</dc:title>
-<dcndl:alternative>
  <rdf:Description>

```

図4 DC-NDL (RDF) で出力した書誌情報

国立国会図書館サーチでの提供は、国内発行オンラインジャーナルの ISSN 書誌データ登録後、おおむね二日後となります。ぜひご活用ください。

(逐次刊行物・特別資料課 整理係)

- [1] オンラインジャーナル ISSN 書誌データは、当館所蔵資料を示すものではありませんので、NDL-OPAC では検索できません。
- [2] 国内発行オンラインジャーナルの ISSN 書誌データは、当館ホームページでリストでの提供も行っています。
<http://www.ndl.go.jp/jp/aboutus/issn/index.html#anchor12>, (参照 2014-11-10).
- [3] 詳細検索画面から検索する場合、データベース = 「国立国会図書館蔵書」、資料種別 = 「デジタル資料」、所蔵館 = 「国立国会図書館」を最初から選択して検索することができます。
- [4] 「見る・借りる」欄の「国立国会図書館蔵書(NDL-OPAC)」をクリックしても、当館所蔵資料ではないためNDL-OPACの書誌詳細画面は表示されません。

コラム:書誌データ利活用(5)

— テキストエディタを使った NDL-OPAC ダウンロードファイルの加工

今回のコラムでは、[NDL-OPAC](#) からダウンロードしたファイルを、テキストエディタで加工する方法について、いくつかご紹介いたします。

【はじめに】

NDL-OPAC からダウンロードしたファイルは、すべてテキスト形式です。テキストエディタで開いて編集等を行うことができます。テキストエディタにはさまざまな種類があります。今回のコラムでは、[Mery](#) というテキストエディタ[1]を例に、以下の三つのケースでの加工方法をご紹介します。

1. 記号区切り形式(\$区切り) : 途中で改行されてしまうレコードへの対処
2. 標準形式 : 途中で改行されてしまうデータ項目(タグ)への対処
3. ALEPH シーケンシャル形式(ASF) : UTF-8 から SHIFT-JIS への文字コード変換

1. 記号区切り形式(\$区切り) : 途中で改行されてしまうレコードへの対処

記号区切り形式のファイルをダウンロードしたときに、1 レコードが途中で改行されているケースに遭遇したことはないでしょうか。これは、1 レコード中の文字数が多いと、一定の長さ(決まったバイト数)で改行されてしまうからです。これでは、Excel にうまく取り込むことができません。

この問題を解決するために、正規表現[2]を用いた文字列の置換を行います。NDL-OPAC からダウンロードしたファイルを Mery で開き、検索メニューから置換を選んでください(図1参照)。

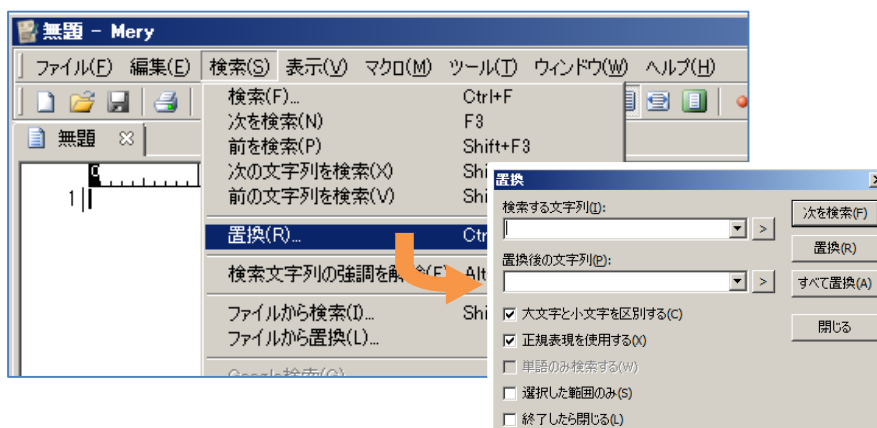


図1 置換画面

<加工方法>

①図1-1をご覧ください。通常、レコードNo. の下に1レコードが1行で表示されますが、No. 56とNo. 59のレコードは2行にわたって記載されており、途中で改行されていることがわかります。また、Excel に取り込む場合、レコードNo. の行も、行の先頭にある半角スペース5文字分も、必要ありません。

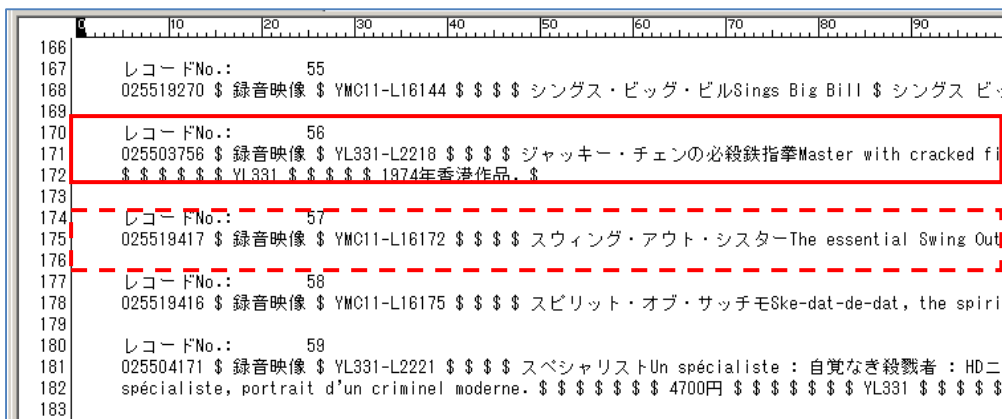


図 1-1 ダウンロードファイル(記号区切り形式)

(表示メニューから「折り返さない」を選択した画面。以下の図も同様。)

②レコード No. の行から 3 行分をまとめて置換します(図 1-1 の囲み部分を参照)。以下の表のとおり置換画面に入力し、「すべて置換(A)」ボタンを押してください。図 1-2 のようになります。入力する文字列の意味は、コラムの最後で説明します[3]。

検索する文字列(I) :	^¥s¥s¥s¥s¥s レコード.+¥n¥s¥s¥s¥s¥s ([0-9]{9}).+¥n *(.*)¥n
置換後の文字列(P) :	\$1\$2¥n
チェックボックス	「正規表現を使用する(X)」にチェック

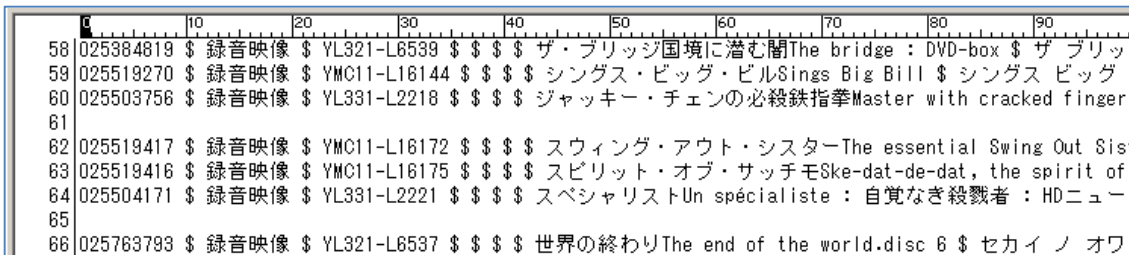
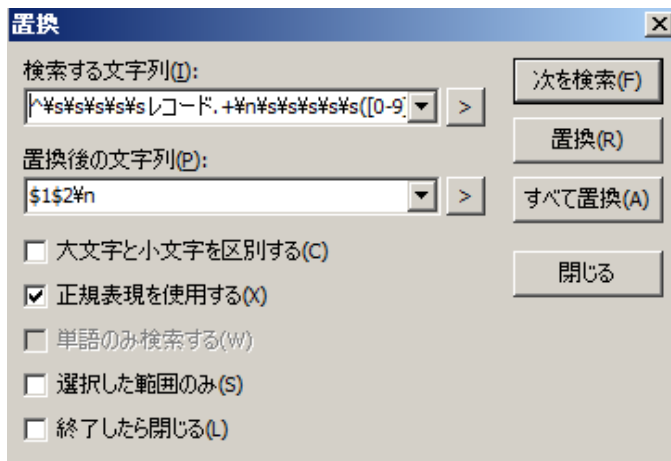
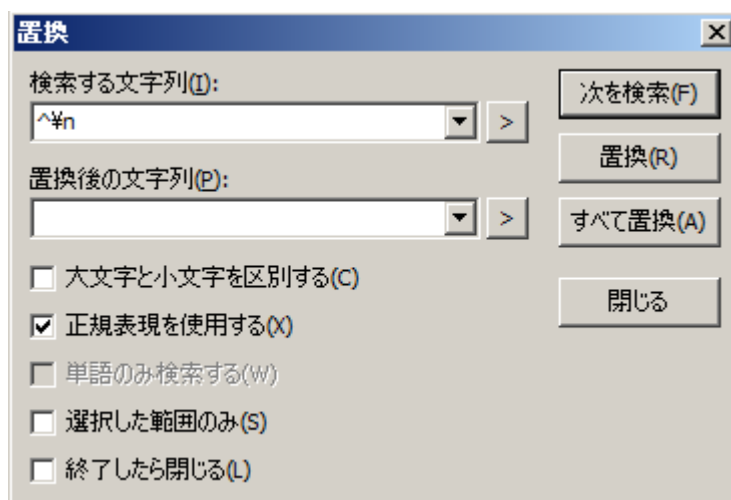


図 1-2 置換処理終了

③図 1-2 の状態でも Excel に取り込むことは可能ですが、空白行が必要ないので削除します。以下の表のとおり置換すると、図 1-3 のように空白行がなくなります。これなら Excel に取り込んでからの手間が減ります。

検索する文字列(I) :	^¥n
置換後の文字列(P) :	(何も入力しない)
チェックボックス	「正規表現を使用する(X)」にチェック



	10	20	30	40	50	60	70	80	90
54	025384819	\$ 録音映像 \$	YL321-L6539	\$ \$ \$ \$	ザ・ブリッジ	国境に潜む闇	The bridge : DVD-box	\$	ザ・ブリッジ
55	025519270	\$ 録音映像 \$	YMC11-L16144	\$ \$ \$ \$	シングス・ビッグ・ビル	Sings Big Bill	\$	シングス	ビッグ・ビル
56	025503756	\$ 録音映像 \$	YL331-L2218	\$ \$ \$ \$	ジャッキー・チェン	の必殺鉄指拳	Master with cracked finger	\$	
57	025519417	\$ 録音映像 \$	YMC11-L16172	\$ \$ \$ \$	スウィング・アウト・シスター	The essential	Swing Out Siste		
58	025519416	\$ 録音映像 \$	YMC11-L16175	\$ \$ \$ \$	スピリット・オブ・サッチモ	Ske-dat-de-dat, the spirit of S			
59	025504171	\$ 録音映像 \$	YL331-L2221	\$ \$ \$ \$	スペシャリスト	Un spécialiste : 自覚なき殺戮者 :	HDニューマ		
60	025763793	\$ 録音映像 \$	YL321-L6537	\$ \$ \$ \$	世界の終わり	The end of the world.disc 6	\$	セカイノ	オワリ

図 1-3 不要な空白行の削除(テキストエディタでの加工完了)

2. 標準形式：途中で改行されてしまうデータ項目(タグ)への対処

標準形式のファイルでは、1つの項目の中の文字数が多い場合に、図 2 の注記のように途中で改行されることがあります。これも正規表現を使った文字列の置換で対処できます。

	10	20	30	40	50	60	70	80	90
17									
18	レコードNo.:		2						
19	資料種別		録音映像						
20	請求記号		YL321-L6535						
21	タイトル		A memory of K						
22	責任表示		[GoRA 原作 ; GoHands 原作・アニメーション制作].						
23	出版事項		[東京] : King Record, [2014.8]						
24	形態/付属資料		ビデオディスク 4枚 (318分) : DVD + 1冊.						
25	注記		期間限定版.						
26	注記		原作・シリーズ構成:GoRA 原作・アニメーション制作:GoHands 監督・キャラクター						
27			CGIディレクター:長瀬義則 編集:丹彩子 音響監督:高橋秀雄/田中亮 音響効果:田						
28			声の出演:浪川大輔/小野大輔/小松未可子/津田健次郎/櫻井孝宏/堀江由衣/福山潤/						
29	内容細目		(episode01) Knight (episode02) Kitten (episode03) KIICHEN (episode04) Knock-or						
30	発売番号		KIBA-92123 (---)						
31	発売番号		KIBA-92124						

図 2 ダウンロードファイル(標準形式)

図 2 の 27 行目、28 行目の先頭にある半角スペース 26 文字分と前行の改行を削除すると、図 2-1 のように途中で改行されなくなります。

検索する文字列(I) :	%n%s%s%s%s%s%s%s%s%s%s%s%s%s%s%s%s%s%s
置換後の文字列(P) :	(何も入力しない)
チェックボックス	「正規表現を使用する(X)」にチェック

置換

検索する文字列(I): 次を検索(F)

置換後の文字列(P): 置換(R)

大文字と小文字を区別する(C) すべて置換(A)

正規表現を使用する(X) 閉じる

単語のみ検索する(W)

選択した範囲のみ(S)

終了したら閉じる(L)

↓

17	
18	レコードNo.: 2
19	資料種別 録音映像
20	請求記号 YL321-L6535
21	タイトル A memory of K
22	責任表示 [GoRA 原作 ; GoHands 原作・アニメーション制作].
23	出版事項 [東京] : King Record, [2014.8]
24	形態/付属資料 ビデオディスク 4枚 (318分) : DVD + 1冊.
25	注記 期間限定版
26	注記 原作・シリーズ構成:GoRA 原作・アニメーション制作:GoHands 監督・キャラクターデザ
27	内容細目 <episode01> Knight <episode02> kitten <episode03> KITCHEN <episode04> Knock-or
28	発売番号 KIBA-92123 (---)
29	発売番号 KIBA-92124
30	発売番号 KIBA-92125

図 2-1 データ項目途中の改行削除

3. ALEPH シーケンシャル形式 (ASF) : UTF-8 から SHIFT-JIS への文字コード変換

ASF 形式は、UTF-8 の文字コードでしかダウンロードできません。SHIFT-JIS への文字コード変換も、Unicode に対応するテキストエディタでできます。ASF 形式のファイルを開いたときに文字化けするテキストエディタは、おそらく Unicode に対応していないものです。

今回使用している Mery では、名前を付けて保存するとき(図 3 参照)に、エンコードの選択項目で「日本語(シフト JIS)」を選択して保存すれば、文字コードが変換されますので、とても簡単です。



図 3 文字コードの変換

【おわりに】

テキストエディタを使うと、今回ご紹介した加工方法のほかにも、さまざまな加工を行うことができます。テキストエディタをうまく使って、全国書誌データを便利にご活用いただければと思います。

また、今後のこのコラムで取り上げてほしいテーマ・内容などがありましたら、メール(bib-news@ndl.go.jp)でお寄せください。

(収集・書誌調整課 書誌サービス係)

[1] Unicode、サロゲートペアなどに対応し、正規表現で検索・置換が可能なテキストエディタ。

<http://www.haijin-boys.com/wiki/>, (参照 2014-11-10).

本稿でご紹介する三つのケースのうち、1と2については、正規表現およびNDL-OPACからのダウンロードの際に指定した文字コード(UnicodeまたはSHIFT-JIS)に対応しているテキストエディタであれば、加工できます。3については、Unicodeに対応していれば加工できます。

[2] 端的に言えば、いくつかの文字列を一つの形式で表現するための表現方法。たとえば、「あいうえお」と「かきくけこ」は「.+」(半角ピリオドと半角プラス)で、電話番号「03-3581-2331」は「[0-9¥-]+」で表現できます。正規表現の意味は、[3]をご覧ください。

[3] 正規表現の意味は、次のとおりです。また、図 1-2 で使用した文字列を説明します。

正規表現	説明
^	行の先頭。半角アクセントで表現。
¥s	半角スペース。半角円記号と小文字の s で表現。
¥n	改行。半角円と小文字の n で表現。
.	任意の文字。半角ピリオドで表現。
.+	1 つ以上の任意の文字の連続。半角ピリオドと半角プラスで表現。
.*	ゼロまたは 1 つ以上の任意の文字の連続。半角ピリオドと半角星印 (半角アスタリスク) で表現。
[0-9]{9}	半角数字が 9 文字分。半角数字は、0-9 を半角大カッコで囲んで表現。個数は、半角中カッコで数字を囲んで表現。
[0-9]+	半角数字の連続。
\$1、\$2	変数。
(●)	半角小かっこで囲まれた部分●は、変数に代入できます。変数は、\$1、\$2、・・・と表現。
図 1-2 で使用した文字列の意味	
^¥s¥s¥s¥s¥s レコー ド.+¥n¥s¥s¥s¥s¥s([0-9]{9}).+)¥n *(.*)¥n	<p>¥n が三つあるので、3 行分を表現。</p> <p>1 行目：^¥s¥s¥s¥s¥s レコード.+¥n 行の冒頭に半角スペースが 5 個あり、次に「レコード」の文字列があつて、その後には何かの文字列があつて、最後に改行。</p> <p>2 行目：¥s¥s¥s¥s¥s([0-9]{9}).+)¥n 半角スペースが 5 個あり、半角数字が 9 個で、次に何かの文字列があつて、最後に改行。</p> <p>3 行目：■*(.*)¥n ここの部分は、途中改行された行と空白行の両方にヒットするような文字列になっています。そのため、半角スペースを¥s で表現せず、そのまま表現する方法(網掛け部分)を用いています。</p>
\$1\$2¥n	<p>置換後の文字列で、変数(\$1、\$2)を含みます。置換前の文字列(検索する文字列)で表現すると以下のとおり。</p> <p>([0-9]{9}).+)(.*)¥n</p>

掲載情報紹介

2014年9月26日～2014年12月24日に、国立国会図書館ホームページに掲載した書誌情報に関するコンテンツをご紹介します。

- ・[「国立国会図書館典拠データ検索・提供サービス \(Web NDL Authorities\)」についてのページを新設しました。](#)
(掲載日：12月11日)
- ・[「書誌情報提供サービス」のページを更新しました。](#)
(掲載日：12月11日)
- ・[Unicode 外の文字リストを更新しました。](#)
(掲載日：11月14日)
- ・[国立国会図書館分類表 \(NDLC\) を更新しました。](#)
(掲載日：11月14日)
- ・[Unicode 外の文字リストを更新しました。](#)
(掲載日：10月7日)

NDL 書誌情報ニュースレター(年4回刊)

2014年4号(通号31号) 2014年12月25日発行

編集・発行 国立国会図書館収集書誌部

〒100-8924 東京都千代田区永田町1-10-1

E-mail: bib-news@ndl.go.jp(ニュースレター編集担当)