情　報　学　基　礎　41－1
（1996．4．18）

国立国会図書館のデータベースと情報検索

杉山　時之

国立国会図書館
総務部情報処理課

国立国会図書館（ＮＤＬ）が四半世紀にわたり開発してきた書誌情報と，これをデータベースとして提供する検索サービスについて言及する．情報サービス環境が激変している現在ＮＤＬの場合でもサービス形態が多様化してきている．館内の来館利用者への資料の所在やステータス情報のサービスから，インターネットによる全世界の情報利用者への出版情報提供サービスまで，巾ひろく要望に応えるべくシステムの構築に取り組んでいる．また図書館が所有する出版物に基づく書誌情報や一次情報のデータベース化を積極的に行い，電子図書館の実現に向けて歩んでいる現状を紹介する．

The databases and the information retreival systems in the National Diet Library

Tokiyuki Sugiyama

Information Processing Division
National Diet Library

I reffer to the information retrieval services which use the bibliographic information developed during a quater of a century in the National Diet Library. As we are exposed in rapid changes of information service field, our library is also making effort to develop the systems which is able to respond to requirments by users. I will introduce the present situation of the NDL producing bibliographic data and degitalized primary information based on the possessed materials and heading for an electronic library.

1．はじめに

国立国会図書館は納本制度のもとにわが国の出版物をもうら的に収集し，そのデータベース化を進めてきている．創立以来半世紀にわたる資料の蓄積と，四半世紀におよぶデータベース化活動は多様化する21世紀の情報社会を前に，情報の電子化と提供という大きなテーマに直面している．それは技術的な課題のみならず，出版と流通，利用と保存といった情報のサイクルに関わる経済的な課題と政策や法制面に関わる課題も含んでいる．こうした課題を解決することによって，需要と供給のバランスを見いだし，きたるべき本格的な電子情報社会が実現されていくものと思われる．

国立国会図書館は2002年に国立国会図書館関西館（仮称）を開館すべく準備を進めている．これはまさしく21世紀の幕開けともいえるわけで，電子化された情報を提供する次世代図書館として一般に理解され期待されており，これに応えるべく計画が進められている．

本稿は現在の国立国会図書館（以降ＮＤＬと記す）が抱えるデータベースとそのサービスの形態について紹介するものであり，関西館の計画については別の機会をとらえていただきたい．

2．情報資源とその利用

情報資源としての主要な資料の所蔵量および年間増加量は次の通りである．（平成7年12月現在）

① 和図書　4,362,000 冊，年 120,000 冊増
② 洋図書　2,000,000 冊，年 38,000 冊増
③ 和雑誌　93,500 種，年 4,400 種増
④ 和新聞　6,450　種，年 180 種増
⑤ 洋雑誌　48,100 種，年 1,250 種増
⑥ 洋新聞　1,410　種，年 12 種増
⑦ マイクロフィシュ
⑧ レコード（含ＣＤ）

その他，地図，楽譜，博士論文など．

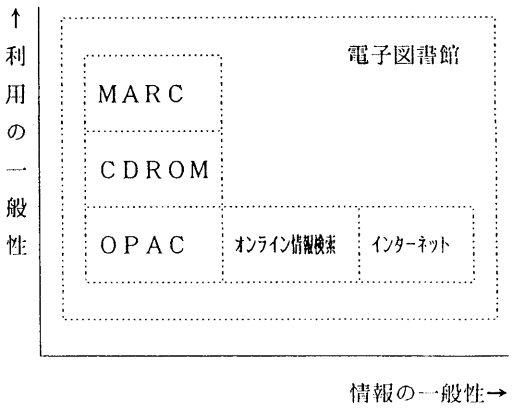
利用実態を表すいくつかの統計的数値は次の通りである．（平成6年度 実績）

① 閲覧者数
　年間： 46.6 万人，1日平均： 2,000 人
　開館日数： 231 日
② 館内貸出および出納
　図書：年間 42 万点，1日平均：1,820 点
　雑誌：年間 89 万点，1日平均：3,850 点
③ 複写
　年間：26 万件，564 万枚

3．データベースのサービス形態

国立国会図書館が行っているあるいは行おうとしているデータベースの提供サービスの構成および位置づけは次の通りである．

データベース
- 書誌情報データベース
  - →ＭＡＲＣ
  - →ＣＤＲＯＭ
  - →オンライン情報検索
  - →ＯＰＡＣ（開発中）
  - →インターネット（開発中）
- 一次情報（イメージ）データベース
  - →電子図書館パイロットプロジェクト



3．1　MARC

MARCは図書館業務のコンピュータ化を支える最も古典的な基盤のひとつである．1960年代末に米国議会図書館が手がけてから四半世紀を越えて現在もその役割は変わりない．各国の国立図書館がその国の出版物をもうら的に収集し，その書誌情報をコンピュータ可読の形すなわちデータベースとして提供する義務を負っている．

わが国でも1981年から JAPAN MARC の名で刊行しており，日本の出版物の約８割がたがこれに収録されていると考えられている．国立国会図書館の所蔵資料では，毎週受け入れられる約1,600冊年間で８万冊の図書について日本全国書誌としてデータベース化されており，過去に刊行された図書についても遡及版と称して，平成10年にはほぼ全数のデータベース化を完了する予定である．

雑誌および新聞についても逐次刊行物の MARC として完成しており，こちらも日本で出版されるものはほぼ100％収録されている．MARC はデータベースとして書誌情報の素材をもうら的にかつ標準化して提供するものあり，図書館活動の基盤として今後も重要な役割を果たしていくことになろう．

3．2　CD－ROM

MARC が書誌情報の素材提供であるのに対してこれを検索システムとしてパッケージ化したものが JAPAN MARC の CD-ROM 版，すなわち J-BISC である．データの収録範囲や内容もほぼ JAPAN MARC と同じであり，ローカルユースとしてスタンドアロンあるいはネットワーク形態で利用されている．MARC の利用には汎用コンピュータとソフトウェア開発等の手当てを必要とするのに比べて，パソコンで簡単に全国書誌の検索システムが実現できる点で，需要は増加の一途をたどっている．

CD-ROM 化されているのは図書だけではなく，日本で出版された雑誌の中の論文や記事の書誌情報，すなわち雑誌記事索引についても既に約120万件が検索出来るようになっている．1985年以降に出版された論文と見ていい．実は1975年以降の分約 110 万件についてもデジタル化がなされており，数年後には CD-ROM 化される予定である．また雑誌の書誌情報も CD-ROM 化を計画している．従来のカード形態や冊子形態の目録類もこれを電子化する方向であり，ローカルユースとしての利用が適切と思われるものは今後順次 CD-ROM 化されることになろう．

3．3　オンライン情報検索

現在約 150 ほどのユーザパスワードを発行してサービスされているオンライン情報検索システム（NOREN）では，10種類のデータベースが公開されている．ユーザは国会，各省庁の図書館，県立図書館，指定都市立図書館といった範囲に限定されているが，システム環境の整備が進展すればこの限定も次第に外していく予定であり，それには有料化の課題もある．このシステムで検索できるデータベースは20数種類あるが，公開の数も順次増やしていく予定である．

後述する OPAC やインターネットでのサービスと競合することになるが，データベースの種類の豊富さと検索機能の充実さという点で他のサービスとは質的に異なり，専門検索サービスという位置付けになる．

3．4　OPAC

図書館の利用者が直接端末を操作してアクセスする電子目録を OPAC と称している．CD-ROM で代用する場合もあるが，一般には常に最新情報を提供できるように，更新系を含めたオンラインシステムの構成にする．従来のカード目録を電子化したものという流れもあり，図書館の所蔵する図書や雑誌等の資料の目録情報がその対象となり，記事論文レベルやフルテキストの範囲までは含めることはせず，所蔵する資料の所在およびそれが利用可能なのかどうか，さらに貸出処理と連結してシステム化する必要がある．

ＮＤＬでは現在 OPAC として CDROM を代用するかたわら，本格的な OPAC 開発を進めている．

図書や雑誌のほか，会議録，論文，記事等，あらゆる種類の索引あるいは目録情報を視野に入れており，来館者はもとより非来館者に対してもサービスする計画である．資料の所在と利用を前提とする点でオンライン情報検索サービス等とその主たる目的が異なり，より所蔵資料に密着したいわば閲覧情報サービスとでも表現できる位置付けになろう．

３．５　インターネット

インターネットへの情報提供は二つの条件が考えられている．ひとつは無料，もうひとつは不特定多数すなわちインターネットをアクセスする人は誰もサービスを受けられるということである．こうした条件から実はＮＤＬでは量，質ともに限定した形で提供するすべく，現在開発を進めている．性格的には所蔵情報提供というより，日本全国書誌すなわち日本の出版情報提供サービスという位置付けと考えられる．当面の計画では最新５年間ほどの和図書と，和雑誌については過去の刊行物も含めて全数の書誌情報を検索できるようにWWWサーバシステムを構築する予定である．出版情報として最低限必要な項目に絞り，予想される大量のアクセスに対処したいと考えており，サービス開始は今年の夏には実現したいと思っている．

４．　データベースの種類と内容

４．１　書誌情報データベース

ＮＤＬ書誌データベースは，1970年代に開発された目録や索引編集システムの副産物としてのＣＴＳ用データを検索用に流用したものからスタートしている．現在オンライン情報検索システムで検索できるデータベースは次の表の通りである．なおこれらの中には米国議会図書館や英国図書館等の外部機関が作成したものが７種類ほど含まれているが，他はすべてＮＤＬが作成したものである．データ件数は平成８年３月末現在の数値である．

項番	データベース名	データ件数
1	和図書目録	1,897,000
2	和逐次刊行物目録	98,000
3	洋図書目録	161,000
4	洋逐次刊行物目録	47,000
5	点字・録音全国総合目録	154,000
6	英国図書館所蔵会議録	357,000
7	支部図書館逐次刊行物総合目録	4,300
8	統計情報総索引	13,000
9	和逐次刊行物総目次総索引	7,500
10	国際逐次刊行物目録	756,000
11	国際逐次刊行物目録（日本）	25,500
12	国会会議録索引	414,500
13	科学技術関係欧文会議録	47,500
14	日本科学技術関係逐次刊行物総覧	18,500
15	音楽資料目録（ＣＤ）	20,000
16	NHK-音楽ＣＤ目録	75,500
17	博士論文目録（日本）	115,500
18	文部省科学研究費補助金研究成果目録	28,500
19	雑誌記事索引	1,267,500
20	和図書著者名典拠	235,000
21	和図書出版者典拠	61,000
22	和図書件名典拠	23,500
23	科学技術関係逐次刊行物目録機関名典拠	12,000
24	国際交換機関・寄贈者典拠	970
25	US/MARC(BOOKS)	3,929,000
26	US/MARC(SERIALS)	710,500
27	US/MARC(GPO)	342,500

以下３つの主要なデータベースについてその概要を紹介する．

１）和図書目録データベース

ＮＤＬはわが国の出版物を網羅的に収集し，その目録または索引を作成する使命を担っている．その出版物の中核をなすのが和図書すなわち日本で出版された図書である．現在和図書のデータベ

ースは全体の80%が完成しており，その出版年の状況は次の通りである．このデータが図書館や情報サービスの世界で広く利用されいる JAPAN/MARC（図書編）であり，また JBISC の名前で知れ渡っている JAPAN/MARC CDROM版である．

その書誌的記録フォーマットはわが国の事実上の標準となっているが，世界標準である UNIMARC と比較すると，日本語という特殊性が故に多少ズレがある．フルテキストの時代に入りつつある現在，MARC の影も薄くなりつつあるが，印刷媒体が今後も存続し得るであろうし，電子出版が主流になってもその検索索引情報として重要な役割を果たすことになるであろう．

対象図書の出版年区分			件数	MARC/CDROM
遡及版	江戸以前	-1967	6万件	（平成10の見込）
	明治	1868-1911	12万件	完成済
	大正	1912-1925	10万件	（平成9年の見込）
	昭和	1926-1947	20万件	（平成8年の見込）
		1948-1955	11万件	完成済
		1956-1968	17万件	
		1969-1976	20万件	
最新版		1977-1988	62万件	
	平成	1989- 現在	49万件	

2）和逐次刊行物目録データベース

和図書と並んでＮＤＬの基本的な書誌データベースのひとつである．日本で出版された逐次刊行物と新聞の書誌情報を含み，ＮＤＬでの所蔵状況が記録されている．ＮＤＬは UNESCO の主催する UNISIST（世界科学情報システム）のプロジェクトの中でわが国のナショナルセンターに指定されており，わが国で発行される逐次刊行物に ISSN（国際標準逐次刊行物番号）の付与の任務を果たしている．そのために網羅的に収集し，データベース化を進めている．現在，和逐次刊行物目録データベースとして約 9.8 万タイトルを擁し，JAPAN/MARC（逐次刊行物編）を刊行している．逐次刊行物は図書と比較して情報の速報性に優れていることからその利用が多く，その中に含まれる記事や論文のデータベースとリンクして機能する重要な書誌情報である．ＮＤＬでは和洋の逐次刊行物の受入記録システムによって，巻号レベルの受入記録をコンピュータ処理している．いわゆるチェックインシステムである．このシステムと次に紹介する雑誌記事索引データベースを連結することにより，逐次刊行物の利用システムさらにドキュメントサプライシステムの実現を考えている．

3）雑誌記事索引データベース

わが国で刊行されている約3,000の主要雑誌に収録されている論文記事を検索するための索引データベースである．昭和23年から継続して刊行している同名の冊子体索引は昭和50年からコンピュータによる編集に切り替えたため，このデータに検索用のキイワードの付与を行い，オンライン検索データベースとして活用してきている．また平成７年から CDROM 版も刊行するようになった．

実は平成８年からは冊子体の索引誌を中止し，情報検索データベースを志向する採録方針に方向を転換することになり，採録誌数を今までの 3,000 誌から 5,500 誌に増加し，年間約 23 万件の採録を目指している．さらにこれを磁気テープの形で，いわゆる MARC として，販売する予定である．現在までのデータベースとしてのデータ量と CDROM 化の状況は次の通りである．

① 1975年から1984年までの採録分

約 110 万件で，CDROM 化は数年後の予定 ．

現在索引誌編集用データからデータベース用データへの再生のため調整を行っている.

② 1985年から現在までの採録分
約 120 万件で，CDROM 化済み.

このデータベースは人文社会分野および科学技術分野の両方から採録しており，とくに人文社会分野の論文記事データベースがわが国では少ない状況からその利用価値が評価されており，現在は学術情報センターの NACSIS-IR でも提供されている．さらに民間のデータベースサービス機関からの引き合いもあり，図書および雑誌の目録データベースと並ぶ重要な位置を占めつつある.

4．2 一次情報データベース

関西地区に2002年に開館を予定している第二国立国会図書館（仮称）の設立は，電子図書館というキイワードと密着した概念で広く認識されている．実現する図書館が電子図書館と呼ばれるにふさわしい形態で開館できるためには多くの技術的非技術的課題をクリアしなければならないわけだが，データベースについては1994年から通産省の特別認可法人である情報処理振興事業協会との共同による「パイロット電子図書館プロジェクト」の中でその具体的なイメージと実現の可能性が実証されようとしている.

このプロジェクトで作成されたデータベースは文字，絵，表など印刷物に表現された情報をそのままイメージでデジタル化したものであり，全文データベースではないが一次情報の一形態として紹介しておく．現在は実証実験システムのレベルであり，限定されたユーザが評価実験のために利用している.

① 国立国会図書館所蔵貴重書　約7,100 枚
ＮＤＬおよびＮＤＬ支部東洋文庫所蔵の貴重書東洋文庫所蔵の国宝および重要文化財11タイトル（1,236枚）が含まれる．内容は，江戸期の浮世絵，錦絵，掛け軸，古地図，奈良絵本など．高品質の写真撮影によるフィルムから，カラーイメージ（約5,000X4,000画素）で入力.

② 明治期刊行図書（ＮＤＬ所蔵）
約21,000冊，約600万 頁
明治時代に刊行された教育，経済，産業，政治社会，統計，教育等の社会科学分野の図書.

③ 第２次世界大戦前後の刊行図書
（ＮＤＬ所蔵）　2,486冊，約75万 頁
主に仙花紙（粗末な再生用紙）に印刷された図書で，この時期の図書としては数が少なく資料的価値が高い.

④ 国内刊行雑誌（ＮＤＬ所蔵）
24タイトル，約100万 頁
わが国の代表的な総合誌，政治・経済誌．1980年１月以降1994年12月までに刊行された分．中央公論，文芸春秋，潮，エコノミスト，国立国会図書館月報など.

⑤ 国会審議用調査資料　260冊，約6,000 頁
国会で審議が予期されるテーマに関する調査レポート．国会議員に提供している国政のさまざまなトピックについての報告書.

⑥ 憲政資料　約7,000点，約34,000 頁
三島通庸関係文書，三島通庸(1835-1888)は明治時代の政治家．倒幕運動に従い，明治時代は各県県令，警視総監等を歴任．保安条例の執行等に辣腕をふるった人物.

⑦ 出版者から原紙料の提供を受けた資料
18タイトル，約160万 頁

帝国議会議事速記録（東京大学会）	約 8万 頁
幕末明治日本国勢地図集成（柏書房）	約200 頁
朝日ジャーナル（朝日新聞社）	約27万 頁
少年サンデー（小学館）	約50万 頁
キネマ旬報（キネマ旬報社）	約24万 頁

など.

5．おわりに

1990年代，われわれはかつてない情報化の激しい波を受けている．電子化と通信ネットワークとくにインターネットというふたつの，しかもこれらが共鳴する形で社会を襲っているといっても過言ではない．図書館の世界でも例外ではなく，衆目の焦点は電子図書館に結ばれているといえる．しかしそれは波の表層部分での話であって，図書館の底辺では相変わらず印刷物資料をベースとした地味な図書館活動が行われており，これからも行われ続けることに変わりはない．

本稿で紹介した四半世紀にわたり構築してきている基本的な書誌情報や索引情報は，電子図書館時代が到来しようともこれを支える柱のひとつとして生きつづけるに違いない．電子化された書誌情報と一次情報は車の両輪として情報社会の道を走りぬけていくのではないだろうか．