

解説



学術情報データベースの構成と利用

学術情報センターのデータベース†

根岸 正 光†

1. はじめに

文部省大学共同利用機関・学術情報センターでは、1987年から大学など研究者向けのオンライン学術情報データベース・サービス“NACSIS-IR (National Center for Science Information Systems-Information Retrieval Service)”を運用している¹⁾。1992年4月現在では、40種のデータベースが、このシステムに搭載され利用されているが、さらに増強を図っている。ところで、我が国における情報検索サービスは、1975年に東京大学大型計算機センターにおいて運用開始されたTOOL-IRが、その始めである。それ以降、大学関係、とくに7大学の大型計算機センターでは、多くの検索システムの開発が行われ、これに主に外国製のデータベースを搭載してサービスが行われるようになった。また同時に、データベース自体の作成も、教室などの研究者グループや学会などの手により、多く行われるようになった。1981年には、文部省科学研究費補助金の研究成果公開促進費の中に、データベース作成に対する補助率が設定され、研究者自身によるデータベース作成が、政策的にも支援されることになり、研究者のデータベース作成が盛んになったといえる。

1986年に発足した学術情報センターでは、その前身である東京大学情報図書館学研究センター(1976~1983年)および文献情報センター(1983~1986年)以来のデータベース関係活動を引き継ぎ、これを「学術情報システム」という構想の中でさらに発展させるべく努力してきた。本稿では、学術情報センターで経年的に実施している「学術情報データベース実態調査」の結果によって現状を鳥瞰した後、学術情報センターにおける

データベース活動を概括する。また、学術情報データベースの問題点と今後の展望も検討することにしたい。

2. 我が国の大学における学術情報データベース

2.1 学術情報データベース実態調査

学術情報センターでは、1988年以来、我が国の全大学などに対して、データベースの作成とサービスの状況に関する調査「学術データベース実態調査」を実施している²⁾。1991年度調査(1991年6月現在)での調査の概要は表-1に示すとおりである。国公立大学など、および短期大学・高等専門学校1093機関を対象として、調査票方式により、(1)計算機センター、図書館など機関内の部局におけるデータベース・サービスの実施状況と、(2)機関所属の研究者によるデータベースの作成状況を調査した。調査は学長など、機関の長宛に公文書で依頼し、また前年度回答のあったものについては、これを同封して、修正箇所を朱記してもらうようにしている。したがって、機関的な活動であるデータベース・サービスの状況については、網羅的な結果が得られている。一方、データベースの作成に関しては、研究者の個人的活動によるものも含めるという趣旨であるが、この調査票が学内において当該研究者個人にゆき渡らないという可能性も考えられるので、結果の網羅性は万全とはいえない。もっとも、この調査もすでに4年目であり、回答されたデータベースは学術情報センターの「データベース・ディレクトリ・データベース」(DBDR)において、公表されているという事情もあるので、データベース作成活動を積極的に公表したいとする研究者は、本調査に参加しているとみてよい。このことから、共同利用的性格をもつデータベースは、ほぼカバーされていると考えられる。

† Database Service at NACSIS: The National Center for Science Information Systems by Masamitsu NEGISHI (National Center for Science Information Systems).

† 学術情報センター

表-1 「学術情報データベース実態調査」の対象機関とデータベースの作成およびサービス実施状況
(平成3年度学術情報データベース実態調査による。以下、表-7 まで同様)

機 関 種 別	調 査 対 象 機 関 数	回 答			
		DB 作成状況 (作成 DB 数)	データベース・サービス実施状況		
			機関数	部局数	サービス DB 数
国立大学	95	656	50	78	276
公立大学	40	28	2	2	2
私立大学	374	292	46	49	82
大学共同利用機関	15	125	9	12	54
文部省・文化庁機関	19	55	5	5	20
短期大学・高等専門学校	550	17	6	6	6
合 計	1,093	1,173	118	152	440

2.2 学術情報データベースの現況

(1) データベースの作成状況

本調査によれば、我が国の大学などにおいて作成されているデータベースは、前表のとおり 1173 件である。1988年度からの推移をみると、この間データベース件数は 750 件から 1173 件に増加している。機関別では国立大学におけるものが全体の半数以上を占めている。以下、これらデータベースの分野別、データ種別、作成組織別、作成資金別、公開性別の比率を掲げると、表-2~6 のようになる。

まず分野別では、自然科学系統が全体の 41%

表-2 作成データベースの分野別件数比 (表-1 中の作成データベース 1,173 件における構成比。以下表-6 まで同様。分野区分は文部省科学研究費補助金の分類に準拠。複合領域の主なデータベースは情報学、地理学、体育学)

分 野 別	比率 (%)
多 領 域	24.9
人文・社会科学	22.9
文 学	14.5
法 学	1.7
経 済 学	4.2
そ の 他	2.5
自然科学	41.0
理 学	12.5
工 学	4.7
農 学	2.1
医 学	13.8
そ の 他	7.9
複 合 領 域	4.7
そ の 他	2.1
無 回 答	4.4

を占め、このほかに多領域と人文・社会科学がそれぞれ 23~25% という割り振りである。個別分野では、医学、理学がそれぞれ 13% 前後を占めるが、文学分野のデータベースが 170 点、14.5%

表-3 作成データベースのデータ種別件数比

デ ー タ 種 別	比率 (%)
文 字	46.2
書 誌	20.5
抄 録	7.4
全 文	7.8
辞 典	3.3
そ の 他	7.2
文 字 + 数 値	13.5
数 値	23.5
画 像	14.3
その他・無回答	2.5

表-4 作成データベースの作成組織別件数比

作 成 組 織 別	比率 (%)
個 人	29.0
研究会・グループ	25.6
機 関	41.1
そ の 他	4.1
無 回 答	0.2

表-5 作成データベースの作成資金別件数比 (重複回答)

作 成 資 金 別	比率 (%)
大学などの予算	73.4
文部省科学研究費補助金	30.9
他の補助金	12.2
そ の 他	18.8
無 回 答	1.9
合 計	137.2

で個別分野では第1位になっているのが注目される。次に、データ種別をみると、文字データによるものが半数近くであり、その半数近くは書誌情報のデータベースである。数値データベースは23.5%であるが、文字情報と組み合わせたものを合計すれば37%に達するし、画像などを含むデータベースは14.3%である。さきに掲げた分野別をこのデータ種別との関連でみると、多領域データベースは多くは書誌目録、医学では数値データ、理学では数値データと書誌目録、文学では書誌目録と文献全文データベースが大きな比重を占めている。文献全文データベースは全体で130点あるが、うち文学のものが68点を占めており、文学分野で全文データベースの作成が盛んであることが分かる。

大学などでデータベースを作成する場合、その作成組織や作成費用も大きな問題である。作成組織としては、機関的な事業としてデータベース作成が行われているものが41%であるが、研究者個人として取り組んでいるものも29%、またグループによる場合が25.6%という比率である。分野別との関連では、多領域データベースは、おおむね組織対応で作成されるし、医学でも半数近くが組織対応であるが、理学では、個人、グループ、機関がほぼ均衡し、文学では、機関と個人が相対的に多く、グループでの作成は少ない。作成資金の面では、大学などの予算による場合が73.4%であるほか、文部省科学研究費を得ているものが30.9%ある。多領域データベースがほとんど大学などの予算で作成されているほかは、分野間

に目だった差異はない。

作成されたデータベースの公開性については、全体の41%が公開中である反面、非公開とするものも35%ある。非公開の割合は医学分野で高く(52%)、また文学でも39%が非公開としている。前者はデータの性格上やむをえないものが多いとみられ、後者については、公開に至るほどデータ整備が進んでいないとの理由が主たるものと思われる。公開の方式では、公開データベースの57%がオンライン・システムによるものであるが、パソコン、ファイル提供、印刷物もかなりの比率を占める。なお、非公開データベースの59%はパソコン上のデータベースである。

(2) データベース・サービスの状況

大学などにおけるデータベース・サービスの状況は表-7に示すとおりで、現在440件のデータベースが、計算機センター(情報処理センター)と図書館を主体にサービスされている。大学におけるデータベース・サービスは、7大型計算機センターを中心に発展してきたという経緯もあり、現在でも117件のデータベースがこれら7センターで公開されている。大型計算機センターは、全国の国公立大学の大学院生以上の研究者のための全国共同利用施設であり、この7センターに学術情報センターを加えた8センターのデータベース案内として、「データベース利用ガイド」が毎年公開されている³⁾。

大型計算機センターではそれぞれ10件程度以上のデータベースが公開されているが、情報処理センター、図書館では1機関あたり、数点、多くは1~2点のデータベースが運用されている。図書館では、ほとんどが当然ながらOPAC(Open Public Access Catalog)といわれる蔵書目録データベースである。目録136点と大学で作成されたデータベース171点の合計307点は、原則的に先のデータベース作成調査における公開データベース479点中に含まれるので、作成データベースの64%が、ここに掲げたような組織的運用を通じて公開されていることになる。

サービス・データベースについては、国内外の他機関から導入したものの133点が含まれる。これは440点中の30%で、逆にいえば、件数からみた大学などでのデータベース自足率は70%ということになる。もっともOPACは他のデータベー

表-6 作成データベースの公開性と公開方式別件数比
(公開方式は重複回答)

公開性と公開方式	比率(%)	
公開中	40.8	
(公開方式)	オンライン	23.1
	バッチ	6.5
	パソコン	11.0
	ファイル	12.4
	印刷物	16.3
	その他	3.9
無回答	0.2	
応相談	22.9	
非公開	34.9	
無回答	1.4	

スとは性格を異にするから、これを除外すると56%という自足率が出る。つまり、大学環境で組織的に公開されているデータベースのうち、56%は我が国の大学環境内で独自に形成されたものである。外国製データベースは96点あるが、そのほとんどが大型計算機センターと国立大学の情報処理センターおよび大学共同利用機関という、国立の機関で運用されている。なお、大学共同利用機関の中では、学術情報センターの運用データベース30件(調査当時)が大きな比重をもっているが、これについては次に詳述する。

ところで、データベースは、同じものが複数の機関からサービスされる場合がある。表-7はサービス件数を統計したものであり、データベースの実体からみると延べ件数にあたる。実数ベースで集計するには、各データベースの内容的異同を判別して名寄せする必要があるが、サービス機関側で、データベースを分割/併合したり、また外部から導入したデータベースに独自のデータを付加してサービスする場合などがあるので、厳密な名寄せは実際上困難である。前出、学術情報センターの「データベース・ディレクトリ」を検

索・通覧して異同を判別するかぎりでの複数サービスの概況は次のとおりである。

全体に複数サービスの例は少なく、外国製データベースでは、遺伝子データのEMBL/GenBank/NBRF(欧州分子生物学研究所/Los Alamos研究所/Georgetown大学)、蛋白質構造データProtein Data Bank(Brookhaven研究所)、有機化合物の結晶構造に関するCambridge Crystallographic Data Centre Database(Cambridge大学)、英国電気学会(IEE)の文献抄録データベースINSPEC、米国議会図書館の書誌LC-MARCがそれぞれ5カ所ほどで運用されている。国内の大学外機関製のデータベースでは、日本経済新聞社の経済統計データNEEDSの各種系列が9カ所に、国会図書館のJapan MARCが5カ所に導入されている。大学などで作成されたものでは、量子化学データベースQCLDBが5カ所で運用されている。その他、2カ所でサービスされているものも数点みられるので、表-7の合計サービス件数440件は、データベースの内容からみた「実数ベース」では400種弱になると考えられる。なお、同じデータベースでも、各サービス機関の方針に即して、異なるシ

表-7 機関種別データベース・サービス実施部局数と作成・導入区分別データベース数
(作成・導入区分における「大学など」は、表側に掲げる国立大学から短大・高専までの各種機関を指す。「大学などの職員作成」には、サービスを実施している当該大学などの職員の作成したデータベースだけでなく、他の大学などの職員により作成されたものも含む。「国内外機関」とは、大学など以外の機関で、国会図書館や企業などである)。

機関・部局種別	サービス 実施 部局数	サービス・データベース数					計
		作成・導入区分					
		大学などの 職員作成	図書館	国内外部 機関作成	外国機関 作成		
国立大学	73	118	67	16	75	276	
大型計算機センター	7	74	3	4	36	117	
情報処理センターなど	12	11	6	6	30	53	
図書館	36	4	51	1	1	57	
その他	23	29	7	5	8	49	
公立大学	2		2			2	
図書館	2		2			2	
私立大学	49	13	48	20	1	82	
情報処理センターなど	10	1	8	19		28	
図書館	33	1	37	1	1	40	
その他	6	11	3			14	
大学共同利用機関	12	22	11	1	20	54	
文部省・文化庁機関	5	16	4			20	
短期大学・高等専門学校	6	2	4			6	
合 計	152	171	136	37	96	440	

システムと環境で提供されるから、サービスの単純な競合・重複が生じているわけではない。

3. 学術情報センターにおけるデータベース・サービス

学術情報センターのオンライン・データベース・サービス NACSIS-IR の現況は表-8~10 のとおりである。学術情報センターは、1980年の学術審議会の文部大臣への答申に盛られた「学術情報システム」構想に基づき、1986年に東京大学文献情報センターを改組して、大学共同利用機関として新たに発足したもので、(1)大学間を結ぶ「学術情報ネットワーク」、(2)大学図書館のためのオンライン目録システム NACSIS-CAT, (3)電子メールと(4)上記の NACSIS-IR を運用している。

表のとおり、NACSIS-IR のデータベースには、(1)センターで作成したもの<表-8>、(2)国内の他機関で作成されたもの<表-8>、(3)大学などの研究者が作成したもの<表-8>、(4)NACSIS-CAT により構築されるもの<表-9>、(5)外国機関製のデータベース<表-10>などがある。センター作成データベースでない場合、(5)は購入、(2)、(3)は無償で供与を受けているものである。

3.1 データベース・サービスの展開経過

(1) 前史

学術情報センターのデータベース・サービスは、その前史において、東京大学大型計算機センターの TOOL-IR と密接に関連している。すなわち、センターの前身である情報図書館学研究センターおよび文献情報センターにおいて、COMPENDEX (1978~) および洋雑誌総合目録データベース (現在の FSCAT, 1982~) のサービスが TOOL-IR の一環として実施されたからである⁴⁾。学術情報センターの発足後、これらを同センターの計算機に移行させると同時に、新たなデータベースをも導入して、1987年4月に NACSIS-IR との呼称のもとに、情報検索サービスを開始した。

(2) データベースの購入から形成へ

NACSIS-IR は、外国製文献データベース6点、MARC (Machine Readable Catalog) 3点、和・洋雑誌総合目録および科学研究費補助金研究成果概要と学位論文索引の合計13点のデータベースを

もって発足した。我が国の大学研究者の学術情報データベースに対する需要は、当初、TOOL-IR の CA Search をはじめとする、海外の有力データベースに向かっており、先の学術審議会の答申も、学術情報センターに外国製データベースのサービスを期待するものであった。そこで、NACSIS-IR では、それまで大型計算機センターなどに導入されていなかった EMBASE, SCI など大型の海外データベースを新たに導入して、研究者の需要にこたえることにした。

もっとも、このころすでに大方の関心は、我が国独自の学術情報データベースの形成という方向に動きつつあった。データベースの作成に膨大な労力・時間を要するのは周知のことながら、努力の結果、上記のとおり、総合目録のほか、科学研究費補助金研究成果概要と学位論文索引の2点の独自データベースを擁して開業することができた。その後、体制の整備を図りつつ、独自の学術情報データベースの形成活動を充実させて今日に至っており、この点についてはのちに整理する。

NACSIS-IR の検索システムは、TOOL-IR の経験を生かして、日立製の情報検索ソフトウェア ORION の上に、データベースそれぞれの特性に即したインタフェースを設計・開発して、付加したものである⁵⁾。なお、データベースあたりのレコード数の増加とこれにともなう索引数の増加に対処するため、現在、ORION をその後継ソフトウェアである MR 90 に順次置き換えている。

3.2 NACSIS-IR の利用状況

NACSIS-IR は、国公私立の大学・短期大学・高等専門学校などの研究者および図書館員を対象利用者としている。登録利用者は、1991年度では620機関の4,800名程度で、その内訳は、教員73.6%、院生6.1%、図書館員15.5%、その他4.8%であるが、国立大学の教官が全体の47%程度を占めている。データベースごとの利用量では、目録系の和図書総合目録 (JBCAT, 全表示件数の22%)、JPMARC (同20%)、洋図書総合目録 (FBCAT, 同14%) が断然多い。これには、この種の目録系データベースが、接続時間や表示量とは無関係の、セッションあたり30円という低額の定額料金であること (他は1分50円、表示1件13円の従量制) と、図書館員による頻繁な利用という要因があるとみられる。

表-8 NACSIS-IR 学術情報センター作成および国内機関作成データベース

(1992年4月現在. 公開予定分を含む)

No.	データベース名	データ件数	収録期間	運用開始	対象分野	内 容	作成機関		
1	科学研究費補助金研究成果概要データベース KAKEN	57,200	1985~	1987. 4	全分野	文部省科学研究費補助金による研究成果概要報告	学 術 情 報 中 心		
2	民間助成研究成果概要データベース JOSEI	700	1971~	1992. 3	全分野	民間助成財団などの助成研究成果概要報告(4財団)			
3	海外研究プロジェクトデータベース EXRP	81,000	最新版	1990. 5	科学技術	日米英仏独伊加スウェーデン8カ国での政府助成研究の概要			
4	経済学文献索引データベース KEIZAI	63,000	1983~	1991. 8	経済学	我が国の経済学文献の書誌、「経済学文献季報」に並行			
5	学位論文索引データベース GAKUI	54,600	1984~	1987. 4	全分野	我が国大学の博士学位論文の索引			
6	研究者ディレクトリ RES	13万	1988年5月現在	1991. 1	全分野	我が国大学など研究者の研究課題・論文など、「研究者・研究課題総覧1990」に並行			
7	データベース・ディレクトリ DBDR	1,026DB 114機関	1991年4月現在	1988. 4	全分野	我が国大学などで作成・提供されるデータベースの総覧			
8	雑誌記事索引データベース ZASSAKU, NDLJPI	805,600	1984~	1992. 1	全分野	日本の主要学術雑誌掲載論文の索引情報		国立国会図書館	
9	学会発表データベース	第一系: 電気・情報・制御 GAKKAI 1	58,400	1987~	1988. 4	電気・情報・制御	13学会	学 術 情 報 中 心	
10		第二系: 化学 GAKKAI 2	13,700	1988~	1989. 4	化 学	3学会		
11		第三系: 建築・土木・造園 GAKKAI 3	12,200	1990~	1991. 3	建築・土木・造園	3学会		
12		第四系: 生物学・農学 GAKKAI 4	730	1990~	1991. 3	生物学・農学	3学会		
13		第五系: 理学 GAKKAI 5	120	1991~	1992. 5	理 学	3学会		
14		第六系: 工学 GAKKAI 6	120	1991~	1992. 3	工 学	3学会		
15		第七系: 医学 GAKKAI 7	680	1990~	1991. 3	医 学	3学会		
16		第八系: 人文・社会科学 GAKKAI 8	130	1991~	1992. 3	人 文・社会科学	3学会		
17	全文データベース	学術論文データベース 第一系: 電子 PAPER 1	全文 460 抄録1980	1989~	1990. 4	電 子	学会誌掲載論文の全文 (図表は FAX で配信)	東京大学 史料編纂所	
18		学術論文データベース 第二系: 化学 PAPER 2	5,920	1983~	1989. 9	化 学			
19		学会予稿集 電子ファイル YOKO	35,000	1989~	1992. 3	全分野			学会の大会・研究会予稿集の版面を FAX で配信 (14学会)
20		現行法令データベース LAW	3,600	最新版	1990. 3	法 律			我が国現行法令の全文
21		臨床症例データベース	1,700	1987~	1992. X	医 学			臨床症例報告の全文 (9学会)
22		維新史料概要データベース ISHIN	18,300	1846~ 1871	1991. 11	歴 史 学			「維新史料概要」全10巻のうち1~2, 4~7巻の全文
23	木簡データベース MOKKAN	12,000	—	1992. 6	歴 史 学	我が国で発掘された木簡の釈文と報告書の索引情報	奈良国立文化財研究所		
24	研究者提供DB	家政学文献情報データベース KASEI	20,000	1979~	1992. 4	家 政 学	家政学関係の国内文献に関する索引情報	日本家政学会	
25		RAMBIOS RAMBIOS	5,000	1983~	1992. X	分 子 生 物 学	分子生物科学分野のレビュー文献情報	東京理科大学 RAMBIOS刊行会	
26		化学センサーデータベース	33,000	1988~	1992. X	化 学	各種化学センサに関する書誌情報と数値データ	化学データベース委員会	
27		脂質情報データベース	1,500	最新版	1992. X	生 化 学	脂質の名称, 化学構造, 生活活性などのデータと文献情報	日本脂質生化学研究会	

No.	データベース名	データ件数	収録期間	運用開始	対象分野	内 容	作成機関
28	学術関係会議開催情報	3,500	最新版	1992. X	全分野	学術会議登録学会の関係する研究会, 国際会議の開催予定	日本学術会議
29	学協会集会スケジュール	800	最新版	1992. X	工学など	日本工学会加盟学協会の研究集会の開催日程	日本工学会
—	練習データベース PRAC	検索練習用に各種データベースの一部を抜き出したもの					

表-9 NACSIS-IR 目録系データベース

(1992年4月現在)

No.	データベース名	データ件数	収録期間	運用開始	対象分野	内 容	作成機関	
30	目録所在情報DB	和図書総合目録 JBCAT 書誌 60万 所蔵 401万	最新版	1989. 4	全分野	我が国大学図書館など所蔵の図書雑誌の総合目録, NACSIS-CAT で蓄積されたデータを NACSIS-IR に移行したものの	学 術 情 報 セ ン タ ー	
31		洋図書総合目録 FBCAT 書誌 120万 所蔵 263万	最新版	1989. 4	全分野			
32		和雑誌総合目録 JSCAT 書誌 7.3万 所蔵 149万	最新版	1987. 4	全分野			
33		洋雑誌総合目録 FSCAT 書誌 11.1万 所蔵 90万	最新版	1987. 4 (1982)	全分野			
34	アメリカン・センター 図書館総合目録 ACCAT	6,300	最新版	1990. 7	全分野	アメリカン・センター6図書館 所蔵図書の総合目録	駐日米国大使館 アメリカン・ センター	
35	国会図書館科学技術 文会議録データベース NDLPROC	28,000	1985~	1992. 1	科学技術	国会図書館所蔵の科学技術 文会議録の目録情報	国立国会図書館	
36	M A R C	JP MARC JPM	107万	1969~	1987. 4	全分野	国内発行図書の書誌	国立国会図書館
37		LC MARC (Books) LCMB	320万	1968~	1987. 4	全分野	主に米国発行図書の書誌	米国議会図書館
38		LC MARC (Serials) LCMS	53万	1973~	1987. 4	全分野	欧文雑誌の書誌	

表-10 NACSIS-IR 海外機関作成データベース

(1992年4月現在)

No.	データベース名	データ件数	収録期間	運用開始	対象分野	内 容	作成機関
39	COMPENDEX PLUS COMPEN	234万	1976~	1990. 4 (1978)	工 学	"Engineering Index" 対応, 雑誌・会議録論文の抄録	米 Engineering Information
40	EMBASE EMB	232万	1984~	1987. 4	医学・薬学	"Excerpta Medica" 対応, 雑誌・会議録論文の抄録	蘭 Elsevier Science
41	Life Sciences Collection LIFE	102万	1982~	1987. 4	生命科学	CSA: Cambridge Scientific Abstracts 発行の各抄録誌対応	米 CSA
42	MathSci MATH	80万	1973~	1987. 4	数 学	"Mathematical Reviews", "Current Mathematical Publications" 対応, 論文抄録	米 American Mathematical Society
43	Harvard Business Review HBR	2,600	1927~	1987. 4	経 営 学	"Harvard Business Review" 掲載論文の全文	米 John Wiley & Sons
44	ISTP & B ISTP	176万	1982~	1987. 4	科学技術	"Index to Scientific & Technical Proceedings" 対応, 会 議録の目次的情報	米 Institute for Scientific Information
45	SciSearch SCI	336万	1987~	1988. 4	自然科学	"Science Citation Index" 対応, 論文引用索引	
46	Social SciSearch SSCI	61万	1987~	1988. 4	社会科学	"Social Sciences Citation Index" 対応, 論文引用索引	
47	A & H Search AHCI	56万	1987~	1988. 4	人文科学	"Arts & Humanities Cita- tion Index" 対応, 論文引用索引	

NACSIS-IR へのアクセス経路は、学術情報ネットワークを介在させることで、多様なルートが可能である。先ごろ実施した利用者へのアンケート調査によれば、学内の計算機センターなどを經由して接続するものと、これらを介さず電話回線で接続するもの（後述のアクセス・ポイント、DDX-T の利用を含む）の比率は 53 : 47 である。なお、通常の電話回線で直接接続するものが全体の 24% を占めるが、これは電話料金がさほど気にならない東京近辺の利用者も多いためであろう。

学術情報ネットワークの利用という観点からみると、上記の学内センター経由分は、そのセンターと学術情報センター間の接続に学術情報ネットワークを使っている。また、学術情報ネットワークには、「アクセス・ポイント」と称して、電話から直接ネットワークに乗り入れる口を全国 7カ所に設けており、これを利用するものが 9.5% ほどある。この両者の合計が学術情報ネットワークの経由分であり、非経由分との対比は 62 : 38 になるから、全国的には、学術情報ネットワークは有効に使われているとみることができる。

4. 学術情報センターにおける独自データベースの形成

これまで、我が国では独自のデータベースの形成が積極的に行われず、外国製のデータベースから情報を引き出すのみであり、これが国際的な情報摩擦を引き起こしているといった議論があった。その真偽をにわかに断定するのは困難であるが、近年における我が国科学技術の先鋭化ともなっており、日本の科学技術情報に対する海外の関心が高まっているのは事実であろう^{62,73}。データベースの作成には、何分、相当の労力と時間を要することは確かであるから、この際、我が国でも着実な努力をつみ重ねてゆく必要がある。このような情勢を受けて、学術情報センターでは、内外の研究者の需要にこたえるべく、学術情報データベースの形成を重点的に進めている。

4.1 データ収集における制度型、独自調査型、協働型

データベースの作成にとっては、誤りのないデータを漏れなく、定期的集めるということが最も重要かつ困難なことである。「科学研究費補

助金研究成果概要データベース」は、この補助金制度に基づき年度末に研究者から提出される報告をデータベース入力したものである。「学位論文索引データベース」は、学位授与大学で調製される授与簿のデータを入力したものである。これらは制度に依拠したデータ収集によって作成されるものといえ、収集面での安定性が高い。一方「データベース・ディレクトリ」は、前記のとおりセンター独自の「学術情報データベース実態調査」によるものである。独自調査によるデータ収集は一般に困難なことが多いが、年次的に継続実施することで安定状態を確保できる。センターでは、大規模な独自調査として、全国の大学図書館を対象に雑誌総合目録データ調査を 1976 年以来実施している。これはすでに大学図書館界で定着し、これを前提にオンライン目録システム NACSIS-CAT の普及が図られたという経緯がある。

学術情報データベースの形成にあたって、今後重要なのは学協会などとの協働である。学会は組織化された学術情報源であり、我が国独自の学術情報データベースの形成の多くは、学会との協力関係において可能になると考えられる。年次大会、研究会での発表要旨を収める「学会発表データベース」と「学会予稿集電子ファイル」（予稿集の版面を光ディスクに収録する）では、各分野の学会にひろく協力を呼びかけることで逐次参加学会が増加しており、現在 50 学会ほどになっている。「経済学文献索引データベース」は経済資料協議会との協力によるものであるし、各助成財団との協力で「民間助成研究成果概要データベース」を作成している。また 1992 年度に公開が予定される「臨床症例データベース」では、医学系 9 学会との協働で作業を進めている。

4.2 索引・抄録型から全文型データベースへ

学術文献に関するデータベースは、これまで、著者、標題など書誌事項だけを収録する索引型データベースや、これに抄録を加えた抄録型データベースが主流であったが、本文自体を収容するフル・テキスト（全文）データベースが順次盛んになりつつある⁸¹。これは、電算写植による雑誌印刷や電算化された編集システムの普及ともなうもので、コンピュータを利用した編集・印刷工程の途次に生ずる文書ファイルを、全文データベースに変換するという方式で作成される例が多

い。学術情報センターでは、この方法による学会誌の全文データベースの作成を学会との協力により進め、1989年から「学術論文データベース」として公開した。この場合、図表の扱いが問題で、外国のものは現状でも HBR (Harvard Business Review) データベースにみられるような、図表を含まないテキストだけの全文データベースである。「学術論文データベース」では、図表部分を光ディスクに収容して、利用者からの指示に応じて、これを FAX で送信するシステムを開発し、「完全な」全文データベースとした。

なお、今後は電算写植ファイルの変換、再利用のような後追いのデータベース作成ではなく、データベース化を前提にした一貫方式が望まれる。これに関して、欧米では、文書の構成を記述する言語 SGML (Standard Generalized Markup Language) の適用が進みつつある模様であり、センターでもこれに注目し、調査研究を行っている^{9),10)}。

5. 学術情報データベース発展のための今後の課題

5.1 研究者のデータベース活動の支援

学術情報データベースは、本来、研究者の研究活動と密接に関連して形成されるものである。前記のような学協会との協力に基づくデータベース作成はこの一環であるが、すでにみたとおり、データベースの作成は、大学や学協会の事業として組織化される場合だけではなく、研究者個人やグループの手による例も多い。こうした活動は、文部省科学研究費補助金研究成果公開促進費で支援されており、1990年度には69件に対して4億2000万円が助成された¹¹⁾。

データベースは、公開して、多くの人に利用されることで、その作成の本旨が発揮されるのが普通であろう。また公開利用を通じて、作成活動がさらに組織化される可能性も高い。学術情報センターでは、科学研究費によるデータベースをはじめとして、研究者により作成されたデータベースが、簡便に全国公開に移行されるよう、「大学等の研究者等提供データベースの受入事業」を1990年に制度化した。同時に、文献索引・抄録型データベースを対象に、常識的なフォーマットで作成されたデータベースは、パラメータを与えるだけ

で、NACSIS-IR のオンライン・データベースに変換されるシステムを開発し、受入の円滑化を図った。日本家政学会の「家政学文献データベース」は、この制度、システムによる受入データベースであり、このほか数点についても公開準備を行っている。

5.2 非文字データベースと公開方法の多様化

文献データベースは、データの基本構成が均一なので技術的に対処しやすく、検索方法も一様であるから利用者にも分かりやすい。そこで、上記の受入システムなど、体制の整備も進化した。一方、数値データベースに関しては、上記受入事業では、利用システム（アプリケーション・プログラム）付きでの提供を前提にしているが、これはシステムの移植作業を含むので実際上難しい面もある。数値データベースの場合、さらに、センター側で用意する一般的なプログラムだけでは不十分で、利用者側でも独自のプログラムを作ってデータ処理を行いたいという事情もある。この点は、テキスト・データベースと称される、多くは古典の全文データベースについても共通である。これは文字データベースではあるが、その利用法は通常の検索ではなく、たとえば用語法の統計分析など、利用者プログラムでの独自処理が望まれる。

この種のデータベースの公開利用方法としては、ファイルを交換する方法と、データベース（および利用ソフトウェア）の作成者の計算機システムに、利用者が直接乗り入れて使う方法のいずれかが当面考えられる。ファイルの交換については、大容量媒体である CD-ROM を用いるほか、高速回線でのダウンロードもあり得る。作成者システムへの乗り入れ利用には LAN 間接続が有効である。学術情報センターでは、1992年度から TCP/IP の高速通信網 SiNet を運用するが、この特性を活かして、多様な形態でのデータベース利用が今後考えられることになるであろう。

6. おわりに

学術情報センターのデータベースは、センターで開設した国際回線を通じて、米国科学財団 (NSF)、米国議会図書館および英国図書館からも利用できる。これは、我が国の学術情報の海外発信の一翼を担うという意図のもとに実施されてい

る。今後学術情報センターには、我が国の学術情報の組織化、データベース化をより一層促進する一方で、ゲートウェイの開設などを通じて、学術情報の国際的な相互流通にも積極的に貢献してゆくことが期待されよう。

参 考 文 献

- 1) 学術情報事務研究会編：NACSIS-IR 総合マニュアル（改訂版），p. 500, (財)電気・電子情報学術振興財団，東京（1992）。
- 2) 平成3年度学術情報データベース実態調査報告書，p. 45, 学術情報センター，東京（1991）。
- 3) 全国共同利用大型計算機センター，ライブラリ・データベース連絡会編：オンライン・データベース利用ガイド（第11版），p. 235（1991）。
- 4) 根岸正光，小澤 宏：データベース・サービスの10年，東京大学大型計算機センター編：最近10年のあゆみ，東京大学大型計算機センター，東京，pp. 140-154（1986）。
- 5) 根岸正光：情報検索システムの構築における利用者の特性とデータベースの特性—NACSIS-IRの構成を通して，大学図書館研究，第32号，pp. 1-11（1988）。
- 6) Negishi, M.: *Research Activities in Japan and Japanese Articles Registered in Western Databases*, Monch, D. et al. eds.: *Japanese Information in Science, Technology and Commerce: Proceedings of the 2nd International Conference*, IOS Press, Amsterdam, pp. 184-197（1990）. ISBN 90-5199-022-7.
- 7) Negishi, M. and Adachi, J.: *Overseas Acceptance of Japanese Scientific Papers as Seen*

in a Citation Analysis, Preprints of the 3rd International Conference on Japanese Information in Science, Technology and Commerce, INIST-CNRS, Nancy, France, pp. 243-259（1991）. ISBN 2-904975-72-1.

- 8) 根岸正光：フル・テキスト・データベースの応用動向，情報処理，Vol. 33, No. 4, pp. 413-420（1992）。
- 9) 田中洋一：文書記述言語 SGML とその動向，情報処理，Vol. 32, No. 10, pp. 1118-1125（1991）。
- 10) 根岸正光：電子原稿・電子出版・電子図書館—『SGML 実験誌』の作成実験を通して，情報処理学会研究報告，Vol. 91, No. 98（91-FI-24），pp. 8.1-8.8（1991）。
- 11) 文部省のデータベース振興策，(財)データベース振興センター編：データベース白書1991，(財)データベース振興センター，東京，pp. 238-244（1991）。（平成4年4月9日受付）



根岸 正光（正会員）

1945年生。1968年東京大学経済学部卒業。1976年同大学院経済学研究科博士課程修了。東京大学助手、講師、助教授（情報図書館学研究センター，文献情報センター）を経て，1986年より学術情報センター教授（データベース研究部門）。オンライン情報検索システム，学術雑誌総合目録データベースの開発に従事。最近は，論文引用度の分析による研究水準の国際比較，全文データベースの編集・検索システムなどを研究，ASIS会員。

