

図1 NISLOにおける情報処理のフロー（参考文献13の第3図を改変）

<http://taxa.soken.ac.jp/JUMOKU/>で公開されている。

藻類分野では、東京大学応用微生物学研究所（当時）のご協力を得て、ミカツキモの顕微鏡画像を数日間にわたり16mm映画に記録して交配の様子を分析することを試み成功した。また藻類情報専門部会が総力をあげ、また、各機関のご理解も得て、6カ国39機関が維持していた3000種11,000株を収録した*World Catalogue of Algae*を出版した(5)。このカタログは、そのオンライン版を後述する微生物データセンター(WFCC - MIRCEN World Data Centre for Microorganisms, WDCM)のWebサイト(<http://www.wdcm.org/>)からデータベースALGAEとして公開している。データをとりまとめたから10年を過ぎているので更新が望ましいが、この間にインターネットによる各機関での情報発信が広がっているため、当然デジタルでのカタログ出版が考えられるであろう。

動物培養細胞分野では、雑誌「組織培養」に連載されていた株総説に対応した資料論文を収集したが、この株論文ファイルは現在では厚生省セルバンクで保管されている。植物培養細胞分野でも、当時4年間に発表された論文から、植物培養細胞に関する3000編を植物培養細胞情報専門部会において収集し文献検索システムも試作した(16)。

動物培養細胞分野と植物培養細胞分野で学んだことは、フルテキストの論文集が研究教育の貴重な情報基盤であること、同時に、専門家の使用に耐える品質の文献データベースの構築には知的集約が必須であること、一方で、知的集約を継続していく体制を整えることは容易ではないこと、であった。現在では、出版社が提供するフルテキストデータベース(例<http://www.scopus.com/scopus/home.url>)やPubMed

Central (<http://www.pubmedcentral.nih.gov/>)のような無料のフルテキストデータベースから図表を含む全文を簡単に入手できるようになったが、それでも、動物培養細胞分野と植物培養細胞分野の専門家によって管理された(*curated*)文献データベースを構築していく体制を整備することができていれば、国際的にも比類の無いデータベースに辿り着いていたのではないかと心残りである。

2. 情報技術の変貌

1970～80年代に開発したNISLOにおける情報処理のフローは図1のようにまとめることができる。この図の左下からの一連のボックスが、情報処理の対象の分析とデータと準備作業に対応している。左上から右へは、データの蓄積から利用までのワークフローに対応している。図1には明示的に書き込まれていないが、「誰が何のために必要としているか」を明確にしておくことが図1のフローの検討を始める前提条件である。

図1のフローは現在でも十分通用するが、フローを実現する情報技術はこの20～30年の間に変貌した。実データの入力やプログラムの作成の手段は、紙テープから始まりパンチカード、専用のビデオ端末そして汎用パソコンへと変遷したが、実験生物に関するデータを機械的に大量かつ網羅的にデジタルデータとして取得できるようになり、望めばデータベースに直接送り込むことも可能になった。データを扱うためのプログラミング言語は、筆者が触れた言語だけでもFORTRAN, PL1, BASIC, CそしてJava(14)やPerl(8, 17)へと主力が移行してきた。データファイルを管理するシステムとしては、コンピュータメーカー独自の

データベース管理システムからハードウェアに依存しない商用データベース管理システムそして高性能なパブリックドメインの(無料の)データベース管理システムが利用できるようになり(3, 15), 近年は, *eXtensible Markup Language (XML)*(2)で記述されたドキュメントファイルとしてデータファイルを保存する方式も普及してきている. データの出力方式は, B4サイズのプリントアウトからA4サイズのプリントアウト, 専用のキャラクター端末への表示, 専用のビデオ端末への表示, 汎用のパソコンへの表示と変遷してきた. 一連のシステムを動かす中核となるハードウェアは, メーカー独自のオペレーティング・システムを搭載したオフコン, ミニコン, 大型計算機を経て, WINDOWSパソコン, 汎用のオペレーティングシステムUNIXを搭載したサーバー, そして, パブリックドメインで成熟してきたLinuxを搭載したサーバーと広がってきた.

今ではデータの入りから出まですべてがデジタルの世界で処理できるようになったが, 紙テープを切り貼りしてプログラムに修正を加えていた風景, 数百枚のカードをカードリーダーに読み込ませていた風景, レーザープリンターから吐き出されたA4サイズの出力を始めて手にした感激は, 今でも, 手ごたえのある記憶としてよみがえってくる. こうした20~30年の情報技術の革新のなかでも, インターネットとWebが, データベースおよびネットワークの観点から特記すべき技術の双璧である. インターネットとWebの普及によって, 地理的な隔絶や個々のシステムの多様性を乗り越えて, 多数のデータベースや解析ソフトウェアを組合わせて利用することが, 画期的に容易になったのである.

3. WFCC - MIRCEN World Data Centre for Microorganisms (WDCM)

3.1 Queensland (オーストラリア) から和光 (日本) へ

1960年代にUNESCOは微生物学の振興に力を入れており, 微生物系統保存事業の振興についても専門家会合が開催された. その会合において議論が始まると参加者は直ちに, 議論のもとになる微生物系統保存事業の国際動向を示す基礎データが無いことに気づいた. この問題認識がWDCMの始まりである. 1964年には東京で初めて開催されたInternational Congress for Culture Collections (ICCC-1)の決議にデータセンターの設立が盛り込まれた. その後1972年にQueensland

大学(オーストラリア)の故V. B. D. Skerman教授が327機関の情報を登録した*the World Directory of Collections of Cultures of Microorganisms*第1版を出版したことによって, WDCMの活動が具体化した. Skerman教授はWorld Directoryの更新を継続していたがWorld Federation for Culture Collections (WFCC, 世界微生物株保存連盟)は個人的努力から組織的運営へと移行するために1996年にWDCMをSkerman教授から引き継ぐ機関を1985年に国際的に公募した.

一方, ライフサイエンス研究情報室はJCMを通して微生物分野のコミュニティーと連携を深めてきており, JCMを運営する理化学研究所ライフサイエンス培養生物部長兼東大教授であった駒形和男先生が室長を兼任されていた.

WDCMの運営機関の募集が開始された後, 理化学研究所内での議論を繰り返しながらも, 駒形和男先生のリーダーシップによって応募を決断した. 仄聞するところでは, 数カ国から応募があったが, WFCCの理事会に設けられた選定委員会で理化学研究所からの提案が採択され, 1986年春からライフサイエンス研究情報室はWDCMの運用を開始した. この引継ぎにあたっては, カードを含む一切の資料を引き取るために, 館野義男博士(現国立遺伝学研究所教授)がQueenslandに出向いた. 引継ぎ完了後, ライフサイエンス推進部の尽力によりWDCM事業の運営費を確保することができ, WDCMは理化学研究所において順調に滑り出した.

理化学研究所に移転した年の1986年直ちにWDCMはWorld Directoryを国際パケット交換網(International Packet Switching System (IPSS))を介して海外からもオンラインで利用可能なデータベースとして公開した. しかし, IPSSを直接利用するこの形態では, 当時はカップラーを介して電話回線に接続し, 特定の電話番号をダイヤルして接続するという煩雑な手順と, IPSSの利用料金に比較して, 利用可能なコンテンツの広がり小さかったことから, オンラインアクセスはほとんど行われなかった. この状況を一変させたのが, 1994年のWorld Wide Webの導入である. 宮崎智博士(現東京理科大学助教授)とシンガポールから客員研究員として来所したTan Tin Wee博士の尽力により, 国内のバイオ分野のWebとしてさきがけの一員として(初期の10サイト以内)Webサーバーの公開を開始した. 現在では, 図2に示すWDCMのトップページ(<http://www.wdcm.org/>)には月に数十万件のヒットがあり, Google(<http://www.google.co.jp/>)の

WFCC-MIRCEN World Data Centre for Microorganisms



当センター(培養生物世界データセンター)は、微生物や培養細胞を系統保存している系統保存機関の網羅的ディレクトリーと培養生物材料のデータベースならびに、生物多様性、分子生物学およびゲノムプロジェクトへのゲートウェイを提供しています。



<p>国際微生物株保存連盟(WFCC) 新着情報 from WFCC 総合的検索エンジン 学名確認 微生物・培養細胞資源センターのデータベース 微生物を学ぶ 配列・系統解析 ゲノムプロジェクト 生物多様性 生物学的研究における安全性 出版社 その他 Inter-Union Bioinformatics Group Report OECD BRC Report (PDF) Global Biodiversity Information Facility (GBIF) ダウンロードe-Workbench</p>	<p>SEARCH ENGINE</p> <ul style="list-style-type: none"> 全文キーワード検索 フィールド検索 AHMII Lineage search <i>NEW!!</i> <p>系統保存機関のリンクページ</p> <p>REGISTRATION/ UPDATE</p> <ul style="list-style-type: none"> 機関登録・更新 保有株リスト登録・更新 <p>WDCMを支援して下さっている機関と協力関係にある機関</p> <table border="0"> <tr> <td>■ WFCC</td> <td>■ NEDO</td> </tr> <tr> <td>■ MIRCEN</td> <td>■ CODATA</td> </tr> <tr> <td>■ CIB</td> <td>■ UNESCO</td> </tr> <tr> <td>■ JST/BRnet</td> <td>■ UNEP</td> </tr> </table>	■ WFCC	■ NEDO	■ MIRCEN	■ CODATA	■ CIB	■ UNESCO	■ JST/BRnet	■ UNEP	<p>サイト内の検索</p>  <p>Global search by <input type="text"/> Google <input type="button" value="search"/></p> <p>連絡先 </p>
■ WFCC	■ NEDO									
■ MIRCEN	■ CODATA									
■ CIB	■ UNESCO									
■ JST/BRnet	■ UNEP									

図2 WDCMのトップページ (<http://www.wdcm.org/>)

サイトの左端のカラムには WFCC の説明と各系統保存機関のサイト、微生物に関するデジタル教材のサイト、安全性のサイトなどへのリンクが用意されている。その最下段からは、微生物のデータベース構築・運用と分類同定の解析機能を備えた電子的ワークベンチ (*e-Workbench*) *InforBIO*、比較ゲノム解析用の *G-InforBIO* および複数の *InforBIO* ファイルの統合検索および公開可能とする *O-InforBIO* をダウンロードできる。サイトの中央のカラムには、WDCM が独自に用意した検索システムのメニューと、機関情報と保有株リストを登録・更新するメニューが用意されている。サイトの右端のカラムには、WDCM 内のメニューを検索するウインドウと、インターネット全体を検索するためのウインドウが用意されている。

「リンクしているサイトの数と重要度を加味した」ページ評価の指標も 6/10 と一定の水準に達している。ちなみに、塩基配列データベースを提供している *DNA Data Bank of Japan* (<http://www.ddbj.nig.ac.jp/>) の重要度は 7/10 という評価である。

3.2 WDCM データベースの利用

WDCM が提供しているデータベースを使って、維持菌株の学名、保有株のジャンル、サービスの種類、スタッフの氏名などから該当する機関を検索することができる。しかし、データベースはこうした検索以外にも使い道がある。すなわち、データベースの内容や利用状況を分析して、戦略を検討する基礎資料を手に入れることができる。表 1 に、WDCM の微生物系統保存機関データベース *CCINFO* の検索にどのような機

関略称 (*acronym*) が使われてきたか分析した結果をまとめた。年推移をみると、各略称の使用件数が 2001 年から 2003 年までが増えていることに加えて、*ATCC* の使用件数が圧倒的に多いことが明らかに読み取れる。筆者は、この統計を見るまで、*ATCC* のように良く知られている機関の情報をわざわざ *CCINFO* で検索しようとする利用者はそれほど多くないであろうと予想していたが、この予想ははずれた。おそらく、他のデータベースからリンクが辿られてくる例も含めて、*ATCC* への関心が非常に高いために *ATCC* の機関情報への要求件数も高くなっているのではなかろうか。表 1 でもうひとつ特記すべき点は、2001 年と 2002 年の 2 年間、2002 年の 1 月だけを除いて 23 ヶ月間 *IFO* が 2 位を占めていることである。*IFO* が国際的にも微生物系統保存機関のブランドであったことをう

表1 微生物系統保存機関ディレクトリーCCINFOに対する検索要求の分析例

	2001 01	2001 02	2001 03	2001 04	2001 05	2001 06	2001 07	2001 08	2001 09	2001 10	2001 11	2001 12
ATCC	334	ATCC 396	ATCC 637	ATCC 511	ATCC 669	ATCC 381	ATCC 378	ATCC 329	ATCC 338	ATCC 500	ATCC 439	ATCC 354
IFO	145	IFO 176	IFO 227	IFO 171	IFO 176	IFO 142	IFO 164	IFO 156	IFO 120	IFO 156	IFO 178	IFO 172
DSMZ	84	IAM 103	DSMZ 188	DSMZ 170	IAM 124	DSMZ 104	DSMZ 117	DSMZ 104	DSMZ 107	DSMZ 130	DSMZ 124	DSMZ 111
CBS	66	DSMZ 103	MAFF 139	NCTC 96	NCIMB 110	IAM 88	IAM 94	GTC GIFU 91	NCTC 89	GTC GIFU 120	IAM 97	CBS 97
IAM	53	JCM 99	IAM 122	NCIMB 91	DSMZ 101	CBS 77	NCTC 91	NCIMB 86	GTC GIFU 86	NCTC 111	NCTC 78	IAM 84
JCM	53	CBS 94	CBS 121	CBS 90	NCTC 82	NCTC 65	CBS 73	IAM 82	CBS 84	MAFF 106	JCM 73	JCM 74
GTC GIFU	49	MAFF 67	CCMP 101	IAM 73	CBS 81	NCIMB 55	NCIMB 71	CBS 76	IAM 71	IAM 105	NCIMB 71	NCTC 62
NCIMB	40	NCIMB 66	NCIMB 82	MAFF 70	JCM 69	JCM 48	MAFF 62	MAFF 63	NCIMB 63	CCM A 91	CBS 68	NCIMB 56
MAFF	39	NCTC 51	IMI 79	CCUG 58	CCUG 54	CCUG 45	GTC GIFU 56	JCM 57	JCM 60	BTCC 87	GTC GIFU 55	CIP 48
CCMP	30	GTC GIFU 48	JCM 68	NCIM 57	IMI 53	GTC GIFU 42	JCM 52	NCTC 54	MAFF 53	IID 81	MAFF 54	CRL 44
	2002 01	2002 02	2002 03	2002 04	2002 05	2002 06	2002 07	2002 08	2002 09	2002 10	2002 11	2002 12
ATCC	338	ATCC 329	ATCC 403	ATCC 613	ATCC 580	ATCC 398	ATCC 501	ATCC 755	ATCC 1174	ATCC 2005	ATCC 1758	ATCC 1280
DSMZ	138	IFO 179	IFO 206	IFO 133	IFO 196	DSMZ 155	IFO 170	IFO 253	IFO 248	IFO 293	IFO 288	IFO 157
IFO	135	DSMZ 140	VTB 128	VTB 122	DSMZ 147	IFO 143	DSMZ 114	DSMZ 187	DSMZ 191	NCTC 206	DSMZ 191	NCTC 144
IAM	102	IAM 103	DSMZ 119	DSMZ 103	CRL 124	NCTC 115	CBS 111	MAFF 160	NCTC 189	DSMZ 195	NCTC 159	DSMZ 113
CIP	98	VTB 79	IAM 93	IAM 78	IAM 114	CRL 112	IAM 105	IAM 154	IAM 170	IAM 150	IAM 135	CBS 87
NCTC	82	CIP 78	NCTC 88	NCTC 72	CBS 103	IAM 93	NCTC 96	NCTC 150	CBS 141	MAFF 146	MAFF 129	IAM 82
NCIMB	79	NCIMB 76	CBS 84	JCM 66	NCTC 102	NCIMB 85	NCIMB 92	NCIMB 139	NRRL 131	CBS 141	CBS 128	MTCC 76
JCM	79	NCTC 74	NCIMB 83	CBS 63	NCIMB 100	CBS 75	GTC GIFU 68	CBS 133	MAFF 118	JCM 108	JCM 96	NCIMB 67
CBS	73	CBS 74	CIP 77	GTC GIFU 62	VTB 92	JCM 68	JCM 66	JCM 119	VTB 108	NCIMB 99	MTCC 93	CIP 64
CCM	62	JCM 70	MAFF 77	CIP 58	JCM 83	CIP 67	CCRC 65	NRRL 113	JCM 107	VTB 90	VTB 91	KCTC 61

	2003 01	2003 02	2003 03	2003 04	2003 05	2003 06	2003 07	2003 08	2003 09	2003 10	2003 11	2003 12
ATCC	1845	ATCC 1702	ATCC 1864	ATCC 2218	ATCC 2192	ATCC 2572	ATCC 3158	ATCC 2870	ATCC 3291	ATCC 2831	ATCC 3883	ATCC 2660
NCTC	330	NCTC 262	NCTC 262	NCTC 242	DSMZ 211	NCTC 270	NCTC 239	NCTC 172	NCTC 268	NCTC 342	NCTC 309	NCTC 236
DSMZ	157	DSMZ 149	DSMZ 170	DSMZ 189	NCTC 210	IFO 182	DSMZ 171	KCTC 160	DSMZ 210	IFO 225	IFO 232	DSMZ 165
IFO	139	IFO 146	IFO 169	IFO 180	IFO 165	DSMZ 125	IFO 154	DSMZ 144	IFO 148	DSMZ 194	DSMZ 152	IFO 154
CBS	110	IAM 107	CBS 160	NRRL 174	CBS 125	MTCC 97	MTCC 114	IFO 106	CBS 146	CBS 169	CBS 126	CECT 141
IAM	104	NCIMB 102	VTB 147	NCIMB 133	IAM 98	CBS 95	CBS 106	MTCC 105	MTCC 134	MTCC 133	MTCC 114	NCIM 117
NCIMB	101	BCRC 95	NRRL 122	VTB 125	VTB 96	IAM 88	NCIMB 105	CBS 100	VTB 112	CIP 133	CECT 103	IAM 115
KCTC	80	VTB 92	IAM 107	CBS 102	RCB 95	JCM 86	KCTC 99	VTB 89	IAM 108	RCB 124	IAM 102	NCIMB 107
CIP	79	CBS 91	NCIMB 103	IAM 92	NRRL 92	CIP 80	IAM 99	NCIMB 83	CIP 106	IAM 122	NCIM 99	CBS 95
CNCTC	75	MTCC 89	GTC 85	RCB 83	MTCC 90	NRRL 76	CIP 90	NCIM 78	CECT 106	NCIMB 117	RCB 94	CCM 89
	2004 01	2004 02	2004 03	2004 04	2004 05	2004 06	2004 07	2004 08	2004 09	2004 10	2004 11	2004 12
ATCC	2223	ATCC 2709	ATCC 5097	ATCC 2172	ATCC 1676	ATCC 1609	ATCC 955	ATCC 1093	ATCC 1544	ATCC 1731	ATCC 1274	ATCC 1054
NCTC	269	NCTC 267	NCTC 570	NCTC 299	NCTC 310	NCTC 327	NCTC 361	NCTC 307	NCTC 331	NCTC 418	NCTC 384	NCTC 355
MTCC	180	DSMZ 182	VTB 343	MTCC 161	DSMZ 167	MTCC 241	MTCC 205	MTCC 188	DSMZ 188	NCIMB 167	DSMZ 200	MTCC 236
DSMZ	180	IFO 158	DSMZ 332	DSMZ 154	IAM 157	IFO 201	IFO 197	DSMZ 148	MTCC 179	DSMZ 166	MTCC 141	NCIMB 181
CPZ	139	MTCC 150	MTCC 306	NCIMB 141	MTCC 155	DSMZ 191	DSMZ 172	NCIMB 144	NCIMB 158	CBS 140	CBS 135	DSMZ 141
IFO	135	IAM 134	IFO 290	IFO 136	NCIMB 152	IAM 163	CBS 154	CBS 123	CBS 142	MTCC 136	BCRC 122	CBS 133
IAM	134	CBS 122	NCIMB 255	CBS 135	CBS 140	CBS 160	NCIMB 133	CCM A 121	CIP 133	IFO 136	NCIMB 120	IFO 130
CBS	128	VTB 115	CBS 237	VTB 116	IFO 139	NCIMB 137	CIP 128	IFO 121	NBRC 109	CIP 129	IFO 111	BCRC 122
ACM	96	NCIMB 97	IAM 229	IAM 113	CIP 118	CIP 130	IAM 124	NCIM 105	IFO 109	NRRL 111	IAM 108	IMI 120
NCIMB	95	NCIM 95	CIP 182	CIP 102	CCM A 109	JCM 123	JCM 122	CIP 99	JCM 103	BCRC 108	ACM 108	NCIM 100

2001年から2004年の間、検索に利用された機関略称(acronym)を集計結果を月ごとにとまとめ、利用件数の多い順に10位までを表にまとめた。なお、微生物系統保存機関の図別機関数やサービスの動向についてはWFCCのホームページ(<http://www.wfcc.info/>)の統計のメニューで閲覧できる。

Bacteria Search

Database:

[Select-All](#) [Clear-All](#)

- [ATCC](#) (ATCC Bacteria)
- [WDCM](#) (WDCM STRAINS Bacteria)
- [JCM](#) (Japan Collection of Microorganisms)
- [CBS](#) (CBS Bacteria Database)
- [TXSearch](#) (DDBJ TXSearch for Taxonomy Database)
- [BCCM](#) (BCCM Bacteria Database)
- [LBSN](#) (LBSN Bacteria NameDatabase)
- [ICB](#) (ICB database by MBI)
- [CABRI](#) (CABRI Bacteria Catalogues)
- [IFO](#) (IFO Bacteria database)
- [DSMZ](#) (DSMZ Bacteria)
- [CCUG](#) (CCUG Microorganisms databases)
- [CECT](#) (CECT Spanish Type Culture Collection)
- [HAMB1](#) (HAMB1 University of Helsinki, Finland)
- [KCTC](#) (KCTC Strain Database)
- [PHLS](#) (PHLS Public Health Laboratory Service)
- [ARS](#) (ARS-NRRL Culture Collection)
- [EGM](#) (Institute of Ecology and Genetics of Microorganisms)

Scientific name:

Genus : Lactobacillus	Species : plantarum	Search
-----------------------	---------------------	--------

図 3.1 AHMII (バクテリア) のトップページ

検索対象としたいサイトにチェックを入れて学名の全部あるいは一部を入力して検索すると、各サイトから返ってきた順にそれぞれの検索結果が一つながりの Web ページ (HTML ファイル) として順次表示する。

かがわせる統計結果である。IFO 以外にも毎月国内機関が表のベスト 10 に含まれている。仮想的に JSCC としてまとめて集計すると、興味深い結果が得られる可能性がある。このように、現状を蓄積したデータベースを分析することによって、系統保存事業の戦略を定量的な根拠に基づいて検討することが可能になる。

3.3 WDCM の機能拡張

1996 年には、宮崎智博士とともに複数の Web サイトを同時に検索して結果をまとめて表示する検索エンジン *Agent for Helping Microbial Information on the Internet* (AHMII) を WDCM から公開した (10)。なお当初この検索エンジンを *Agent for Hunting for Microbial Information on the Internet* と命名したが、いわゆる資源国の研究者からの示唆によってフルネームを修正したものである。現在は、コンピュータ言語の Java のスレッド機能をフルに活用した版を WDCM のサイトから公開している。図 3.1 ~ 図 3.3 にバクテリア、菌類・酵母および培養細胞用の AHMII トップページを掲載した。AHMII によってあらかじめ設定されている機関のデータを一括して横断的に検索でき

る。しかし、実際の検索は、各機関が Web から公開している検索機能に依存しているため、2つの問題を内在している。第 1 に、全機関が共通に提供しているデータ項目以外を使った横断的検索を実行できない点である。図 3.1 と図 3.2 では学名を共通項目として想定している。第 2 に、各機関からの検索結果の出力形式が共通化されていない点である。したがって、AHMII によって数百機関の Web サイトを同時平行検索することが技術的には可能であるが、より深い検索を実現するためには、検索に利用可能なデータ項目や出力形式の共通化が必須である。第 5 章では、病原細菌の場合のデータ項目共通化に触れる。

2003 年には *Simple Object Access Protocol* (SOAP) に基づいたサーバーを整備して (11)、Perl や Java といったコンピュータ言語から検索機能や解析機能を直接呼び出せる *Web services* の技術を導入した (4, 18)。この技術に注目したのは、AHMII の開発の際に、アクセス先の Web インターフェースを一つ一つ分析した結果に基づいてプログラムを作りこむ必要があったこと、加えて、サービスを継続していくためには、常にアクセス先の仕様変更を監視してしかるべくプログ

Fungi & Yeasts Search

Database:

Select-All Clear-All

- ATCC (ATCC Filamentous Fungi&Yeast)
- WDCM (WDCM STRAINS Fungi[yeast&mold])
- JCM (Japan Collection of Microorganisms)
- CBS-Fungi (CBS Filamentous Fungi Database)
- CBS-Yeast (CBS Yeasts Database)
- MMBB (Matsushima Mycological Memor. No.8)
- TXSearch (DDBJ TXSearch for Taxonomy Database)
- BCCMB (BCCM Biomedical Fungi&Yeasts Catalogue)
- BCCMI (BCCM Industrial Fungi&Yeasts Catalogue)
- CABRI (CABRI Filamentous Fungi&Yeasts Catalogues)
- IFO-Fungi (IFO Filamentous Fungi)
- IFO-Yeast (IFO Yeasts)
- DSMZ (DSMZ Fungi&Yeasts)
- CCUG (CCUG Microorganisms databases)
- CECT (CECT Spanish Type Culture Collection)
- HAMB1 (HAMB1 : University of Helsinki,Finland)
- KCTC (KCTC Strain Database)
- PHLS (PHLS Public Health Laboratory Service)
- NCYC (National Collection of Yeast Cultures,Norwich,UK)
- ARS-Fungi (ARS-NRRL Culture Collection - Fungi)
- ARS-Yeast (ARS-NRRL Culture Collection - Yeast)

Scientific name:

Genus : Saccharomyces Species : cerevisiae Search

図 3.2 AHMII (菌類) のトップページ

ラムに修正を加える必要があったことが動機となっている。Web servicesの世界では、どのようなデータベースやデータ解析機能をどのようにアクセスできるかを、コンピュータ言語が理解できる標準化された方式で記述することになる。したがって、図 4 に概略を示すように、個々の Web サイトを開いて人手で解析することなく、Java や Perl のプログラムが Web サイトを渡り歩いて、複数の情報資源をとりまとめて利用できる環境を自動生成する可能性が広がっている。

2004 年には WDCM は地球規模生物多様性情報機構 (Global Biodiversity Information Facility (GBIF), <http://www.gbif.org/>) の WFCC ノードとして、GBIF におけるデータ型式とデータ通信の標準仕様である DarwinCore と DiGIR の運用を開始した (図 5)。この際、国際塩基配列データベース (International Nucleotide Sequence Database (INSID)) を SOAP/Web services でアクセスし、GBIF のデータを DarwinCore / DiGIR でアクセスし、結果をとりまとめて表示する機能を試作した (図 6)。

3.4 和光から三島へ

さて、この間の 1995 年、筆者は国立遺伝学研究所 (以下、遺伝研) 生命情報研究センター長 (当時) の五條堀孝教授と館野義男教授と文京区本郷の喫茶店

Cell Lines Search

Database:

Select-All Clear-All

- ATCC (ATCC Cell Lines and Hybridomas)
- HyperCLDB (The Cell Line Data Base of European banks)
- JCRB (JCRB Cell Bank)
- RIKEN (RIKEN Cell Bank Database Search)
- CABRI (CABRI Cell Lines Catalogues)
- DSMZ (DSMZ Cell Lines)

Scientific name:

Keyword for search : Human Search

図 3.3 AHMII (培養細胞) のトップページ。

で、DNA Data Bank of Japan 事業へのお誘いを受けていた。その後、ほぼ 1 年を経過した 1996 年に筆者は遺伝研へ移籍した。移籍を決断した一番大きな要因は「生物情報の基本は塩基配列データにある」という認識であった。その翌年の 1997 年、理化学研究所から遺伝研への申し入れを受けて、日本微生物資源学会と世界微生物株保存連盟のご了解のもと、WDCM は理化学研究所から遺伝研へ移転した。この際、理化学研究所において計上されていた WDCM 運営費を理化学研究所から遺伝研へと引き継ぐことができずしたがって WDCM は長期的に安定した資金源を失った。しかし、遺伝研の一般研究費、不定期ではあるが UNESCO や American Society for Microbiology からの若干の資金提供に加えて、遺伝研の生命情報研究センター (現生命情報・DDBJ 研究センター) の支援を受けて、WDCM 事業を継続している。ここで、個人や組織の状況が変化したにもかかわらず、インターネットと Web のおかげで、WDCM の Web サイトは揺るぐことなく公開を継続できたことを付記しておきたい。したがって、筆者亡き後も志ある後継者とその支援者が存在すれば、大きな公的資金を確保できなくても、WDCM 事業の命脈を保っていくことが可能であろう。

なお、WDCM の微生物系統保存機関データベース CCINFO 登録機関数は、1989 年に 55 カ国 345 機関、

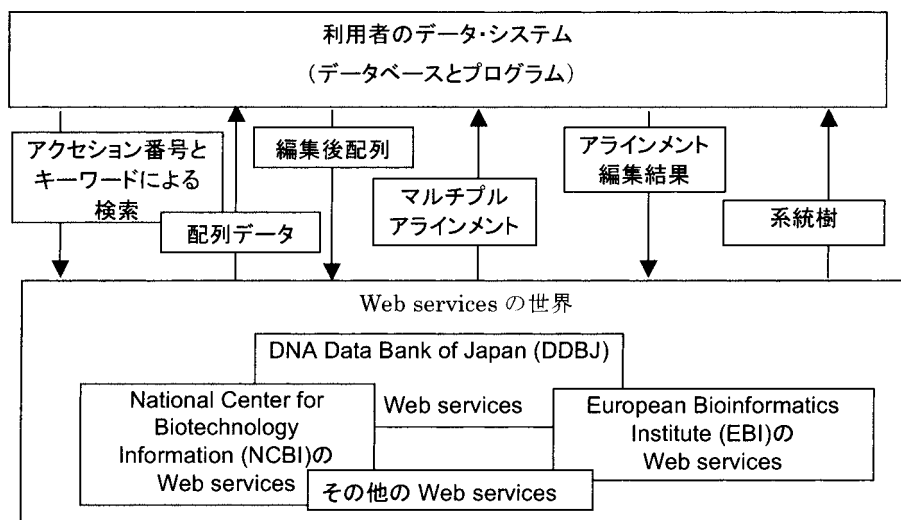


図4 Web サービス

DDBJ, EBI, NCBIなどが提供している Web サービスを、利用者はプログラムから直接呼び出してその処理結果を XML として受け取ることができる。XML にはデータの目印 (タグ) が埋め込まれていて構造化されているので、単なるテキストファイルや HTML ファイルと異なり、後処理するプログラムを容易に用意できる。さらに、プログラムで自動後処理したその結果を次の Web services の入力とすることができる。このようにして同一サイト及び分散したサイトが提供する Web services を組合わせて、例えば、配列のデータ入手から系統樹作成まで自動処理するシステムを効率的に構築できる。Web services を活用するとスクラッチからプログラミングする開発に比較して、およそ 10 分の 1 程度の入力で開発を完了できると思われる。

1993年に60カ国481機関、1999年に60カ国497機関、そして2004年末に65カ国489機関となっている (<http://www.wfcc.info/>)。

4. 電子的ワークベンチ (e - Workbench) InforBIO シリーズ

実験生物情報システム NISLO の基盤の上に WDCM を運営する機会を得ることができたが、それと平行して、筆者は微生物を対象とするデータベースとデータ解析システムの開発にも取り組んできた。1990年代前半から取り組んだ第1世代は、UNIXサーバーとサイベース社のデータベース管理システム SYBASE に基づいたパッケージであった。このパッケージは複雑でコストのかかるシステムとなり、実験のレベルを出ることはなかった。第2世代は、データベースファイルとして XML を採用し、プログラミング言語として Java を採用することによりハードウェアのモデルの如何にかかわらず WINDOWS, Linux および Macintosh で動作を可能とした。この版で、手元のパソコンでデ

ータの管理、数値解析 (9), 進化系統解析, 決定木同定, 確率的同定, クラスタ判別形質の推定, プライマーの設計などの一通りの機能を実行可能となったため、かねてより暖めていた電子的ワークベンチ (e - Workbench) のコンセプトが現実のものになった (12)。第1世代から第2世代にかけては宮崎智博士とともに試行錯誤してきたが、第2世代に入ってからしばらくたった2001年から微生物学者の田中尚人博士がチームに参加したことによって、急速に実用化が進み、事実、開発者グループ以外の研究機関でのご利用が広がり始めた。また、ご利用いただき始めた研究機関や DNA Data Bank of Japan (DDBJ) のスタッフとの議論から、e - Workbench として、伝統的な分類同定用の InforBIO から、微生物ゲノムの閲覧や比較が可能なる G - InforBIO および複数の InforBIO ファイルを統合して検索できる機能をインターネットに公開可能とする O - InforBIO へと発展させた。G - InforBIO の画面例を図7に紹介しておく。

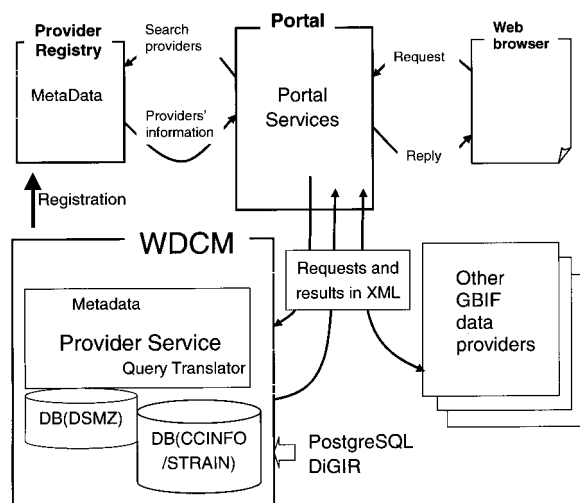


図5 地球規模生物多様性情報機構 (GBIF) における WDCM システム

WDCMがこれまで提供してきたデータベースを GBIF仕様でアクセス可能として (図の左下), その概要をメタデータとして GBIFに登録 (図の左上). その結果, 図の右上の GBIFへの検索要求をポータルサイト経由で受け取って結果を返す.

5. 病原微生物情報システム

5.1 はじめに

2001 ~ 2003年の3年間千葉大学真菌医学研究センターにおける振興調整費「病原真菌・放線菌の遺伝資源の国際的拠点形成のための基盤整備」事業に加えていただき, その後, ナショナルバイオリソースプロジェクトの病原微生物班にも加えていただき, 病原微生物を対象とするデータベースの開発とネットワーク構築に携わっている.

データベースは千葉大学真菌医学研究センター (以下, センター) における菌株の管理および研究を支援するシステムである. 始めに, センターのご関係者との全体会議や個別インタビューによって, 蓄積すべきデータ項目とデータ表記, 利用形態などを明らかにした. その結果に基づいて, データ解析機能を含むデータベースを設計したが, プロトタイプの作成・評価・改良を重ねて, センターにおける菌株受入業務, データ解析機能を含む研究業務, 保存業務および分譲業務に対応するデータベースを構築した. その後, 実データを投入した試験的運用を経て, 本格的運用を開始した. 本システムの開発にあたって, 筆者はシステム開発の場を提供しただけであり, 情報システムに形をつけ, 血を通わせたのは, センター長である西村和子先

生を始めとするセンターの先生方の知的作業であった.

5.2 ハードウェア環境

ハードウェア環境としては, 複数の利用者が複数の場所でデータベースを操作することを想定して, サーバに格納したデータベースを, ネットワークを介してパソコンから操作する環境を整備した. サーバとして, 1.3 GHzのCPUを2個, 4 GBのメモリおよび42 GBのディスクを備えたマシンを当初国立遺伝学研究所に2セット設置して, oliveとdaphneと命名した. オペレーティングシステムはコストとセキュリティーの観点からLinux (Red Hat Linux release 7.2) とした. サーバoliveには, これもコストの観点から選択したリレーショナルデータベース管理システム postgresql-7.3.3を導入し, サーバdaphneをWebサーバとしてApache2.0 (3) とTomcat4.1.30 (15) を導入した. センターからは, パソコンからインターネットを介して遺伝研のサーバに接続してデータベースを操作することを想定して試運転を行った. しかし, 菌株の取り扱いに日常的に使用するにはネットワークが不安定に過ぎたため, サーバをセンターに移行して, センター内のネットワークの中で閉じて利用する環境へ変更した.

5.3 データの利用権限

病原微生物のデータはひととき慎重に取り扱う必要があるため, データベースは, 利用者登録とパスワードならびにSSLの設定をして, 安全性を確保した. 利用者の権限も以下のようにきめ細かく設定した:

- ・システム管理者 (システム管理者用のアカウントで, システムを管理する上で必要な全ての権限を持つ)
- ・業務管理者 (IFM株管理者用のアカウントで, 菌株データ, 保存・分譲データの登録更新)
- ・IFM番号付与の権限を持つ業務従事者 (分譲受付, 入庫・出庫担当者用のアカウントで, 菌株データ)
- ・保存・分譲データの登録更新の権限を持つ研究者 (在庫管理業務などを行わない研究者用のアカウントで, 菌株データの登録更新などの権限を持つ)

対象とする菌株に, 各研究者が解析研究中の菌株である「Work株」と, IFM番号が振られてカタログから公開されている株である「IFM株」の2種類の状態があるものとした. 「Work株」に対しては, 菌株基本情報の登録者 (Registered by), 申請者 (Applicant),

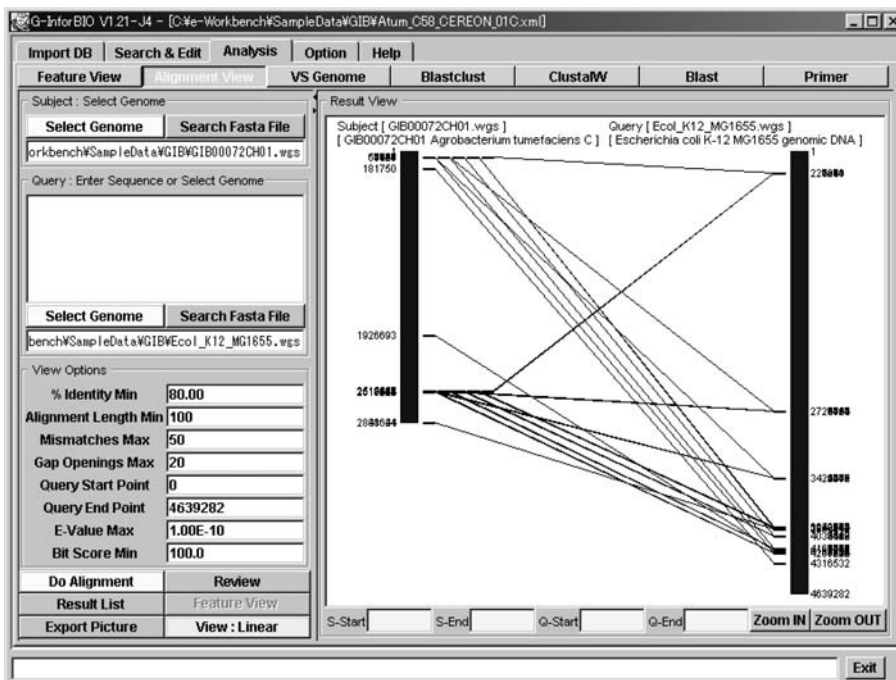


図7 G・InforBIOにおける微生物ゲノム比較の例

2種の微生物ゲノム配列において予測された *Open Reading Frames* (ORFs) の配列同一性をもとに、対応すると思われる ORFs を線で結んだ図。左上のウィンドウで解析対象の配列を選択し、左下のウィンドウで閾値などのパラメータを設定すると、右側のウィンドウに2本の配列を対照させた図が表示される。

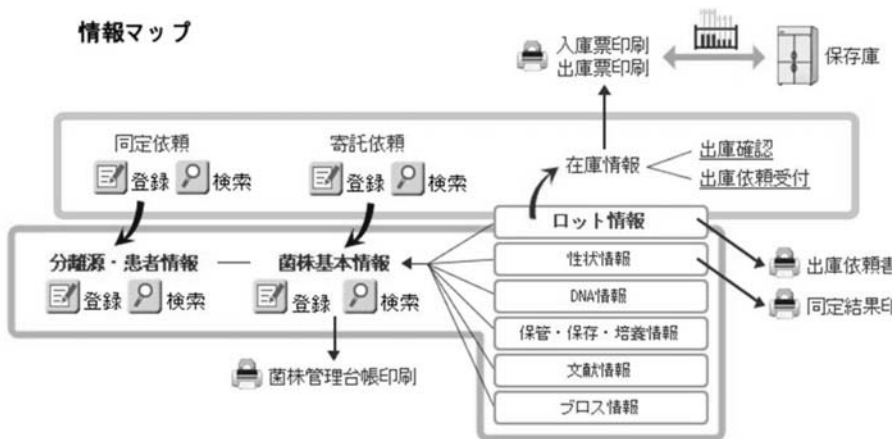


図8 千葉大学真菌医学研究センターで運用されているデータベースシステムの概要

たり、図 10 に表示されている機関が維持している病原微生物を検索することができる。特に病原細菌の場合は、大阪大学、岐阜大学、東京大学および理化学研究所が維持している病原細菌を横断的に一括検索することを可能としている。この検索システムは科学技術振興機構が実装したが、ナショナルバイオリソース

プロジェクトが始まる以前から、3大学の研究室の知的作業によってデータ項目の標準化が進められていたことが、一括検索を成功に導いた決定的要因である。

5.6 非専門家向けコンテンツ

病原性を示す菌の研究は、一般社会において病原微

【特徴の似た菌の仲間達】										
あなたの入力値は？	学名	サムネイル	スコア	中心色	周辺色	表面形状	しわ	集落形状	生育温度	生育速度
				緑	黄色	粉	有り	顆粒状	低温	速い
1	Mki Bio Science		75	緑	黄	羊毛	無し	顆粒状	低温	速い
2	Talaromyces trachyspermus		66	緑	黄色	フェルト	有り	広域	低温	速い
3	Alternaria alternata		58	褐色	黄	フェルト	有り	顆粒状	低温	速い
4	Trichoderma viride		58	緑	黄色	フェルト	無し	広域	常温	速い
5	Acronium falciforme		50	緑	白	羊毛	有り	広域	低温	速い
6	Aspergillus fumigatus		50	はじめ白色 黄緑色を経て 暗青緑色から 暗灰色	はじめ白色 黄緑色を経て 暗青緑色から 暗灰色	ピロード状か ら線毛状	有り	顆粒状	25	速い
7	Penicillium citreonigrum		50	菌糸は白色 であり、分生 子が形成さ れるに伴って 緑色に着色 される	菌糸は白色 であり、分生 子が形成さ れるに伴って 緑色に着色 される	羊毛	有り	広域	低温	速い
8	Penicillium citrinum		50	緑	白	粉	有り	広域	常温	速い
9	Phialophora verrucosa		50	緑	緑	羊毛	有り	広域	低温	速い

【再検索】	
	中心色: <input type="text" value="緑"/> 周辺色: <input type="text" value="黄色"/>
表面形状: <input type="text" value="粉"/> しわ: <input type="text" value="有り"/> 集落形状: <input type="text" value="顆粒状"/> 生育温度: <input type="text" value="低温"/> 生育の速さ: <input type="text" value="速い"/>	<input type="button" value="検索"/>

図9 千葉大学真菌医学研究センターにおいて試作した簡易同定システムによる同定結果例

利用者が菌株のデータをシステムに与えると、システムは判定用のスコアを計算し、図の左側から4列目に表示されているようにスコアが高い順に候補菌種の一覧表を提示する。

生物そのものに対する理解が深まっていてこそ健全に発展するものと考えられる。そこで、3世代の家族で構成される家庭とミクロンというキャラクターを設定して、電子紙芝居の形式をとってさまざまな病原微生物に関する科学的知見をわかりやすく提供しようと試みている。この電子紙芝居では、各ページを理解できたところで利用者が次の画面に進んだり、元の画面に戻って内容を確認できるように設定してある。また、やや専門的な情報は子画面で表示するように設定した。これまでに、水虫編、レジオネラ菌編、マラリア原虫編、および院内感染編の4編をポータルサイトの

「病原微生物を学ぼう」のメニューから公開している。この電子紙芝居作成は、シナリオの準備から挿入する微生物の顕微鏡写真まで、研究班に参画されている各研究機関(図10)の先生方のご協力があったからこそ実現できた成果である。

6. バタフライ効果の結末

筆者は、薄膜の物性研究から実験生物情報システム(NISLO)の開発へ不連続的に転進後、駒込分所から和光本所微生物系統保存施設への移転によって微生物情報との関わりが深まり、微生物データセンター

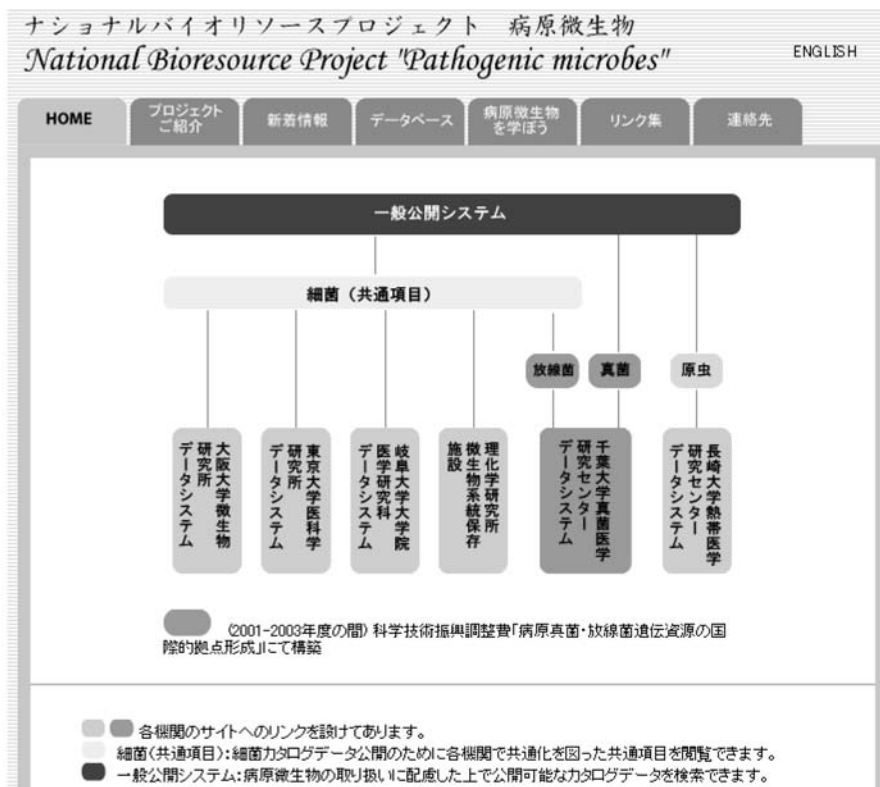


図 10 病原微生物ポータルサイトのトップページ

フォルダー形式のデザインのラベルが、選択可能なメニューに相当している。また、各機関のボックス、細菌（共通項目）のボックス、一般公開システムのボックスもすべてクリック可能であり選択可能である。

WDCMの運営にも携わるようになって、周辺分野も含めた各種国際事業の立上げや運用に関わるようになった。その後、理化学研究所内でのゲノムプロジェクトや電子計算機室の運営に短期間であるが携わった後、遺伝研への異動によって、配列データについては微生物ゲノムデータへの取組みへの展開がもたらされた。これらの研究事業を進めてきた過程で、WDCM事業の他にも多様な国際事業に関わる機会も得た。

NISLOの時代には、CODATA（後述）の中心人物のお一人であった故小谷正雄先生からのご依頼があり、CODATAを通してハイブリドーマデータバンクの構築にも協力した。ハイブリドーマデータバンクは、国際免疫学会などの学協会からの要望を受けて、国際学術連合会議（International Council for Science (ICSU), <http://www.icsu.org/index.php>)の科学技術データ委員会（Committee on Data for Science and Technology (CODATA), <http://www.codata.org/>)が企画したプロジェクトであり、米国のAmerican Type Culture Collection (ATCC, <http://www.atcc.org/>)とフランス

のCERDICおよびライフサイエンス研究情報室が協力して開発が始まった。筆者はこの事業において、米仏にCODATA事務局も加わったデータの取り扱いに関する方針の違いなどに起因する激しい国際的議論を初めて経験することになった。ライフサイエンス研究情報室は、分担した学術雑誌の論文からデータを収集し始めたが、このデータ収集体制を確立することが困難であったことと、米仏の対立からハイブリドーマデータバンクは結局、ATCCとCERDICの2機関それぞれにおいて命脈を保つことになった。CODATAにおいては、故大杉治郎京都大学教授が主導したアジアにおける科学技術データ資源のネットワークの立上げ、英国のFrank Bisbyが主導した生物学名の網羅的リストの整備を目指すSP2000プロジェクトの立上げに加わった。

一方、微生物を対象とするデータベースやネットワークの活動を背景として、OECD Mega-science ForumのBiological Informatics Task ForceにおけるGBIF構想のとりまとめ、とOECD Working Party for Biotechnology

の第1期 *Biological Resource Centers Task Force* (以下, *BRC TF*) の取りまとめに参画した。この他にも、三島に異動後は *DNA Data Bank of Japan* のスタッフとして欧州 *EMBL* ならびに米国 *GenBank* のスタッフと国際実務者会議や電子メールで日常的に議論をする状況になり、ヒトゲノム配列決定国際コンソーシウムにも加えていただいた。また、生物情報の標準化への関心をもっていったところ、多型情報の標準化の道具として XML に基づいた *Polymorphic Markup Language (PML)* を提言したが、社団法人バイオ産業情報化コンソーシウム (<http://www.jbic.or.jp/>) の標準化事業として取り上げられて *Object Management Group* (<http://www.omg.org/>) において国際標準になる見込みである。このような状況からここ数年は毎年 10 万マイル以上飛行している。いささか手を広げすぎた感を否認ない。また、*GBIF* 構想の実現時期には米国ワシントン DC での準備会合に出席中に母を失い時差もあって家族にも大きな負担をかけた。*BRC TF* 提言を議長としてとりまとめるに際しては、冷や汗だけでなく血を流すこともあった。これらは「自己過信もほどほどに」という黙示であったかもしれない。しかし、*GBIF* においては、文部科学省ライフサイエンス課と科学技術振興機構のご苦勞によって日本は *GBIF* 発足当初から投票権のあるメンバーとして参画し、岩槻邦男先生のリーダーシップのもと存在価値を示している。第1期 *BRC TF* では、当時の *OECD* 事務局の *Salomon Wald* 博士、*JBA* 理事の炭田精造博士ならびに米国微生物学会会長も努めた *Atlas* 教授によって国際会議の運営から英文報告書の作成まで鍛錬されるという望んでも得られない貴重な経験をすることができ、加えて第1期 *BRC TF* が打ち出した構想の実現がフランスをリード国とする第2期 *TF* において図られている。

そして 2004 年、生物学についても情報学についてもシステムティックな研究を達成しきっていないながらも日本微生物資源学会から学会賞を授与された。先達の時代認識と筆者の自己過信から始まったバタフライ効果の結末に、もって瞑すべし。

7. おわりに

データベースとネットワークの鍵を握っているのはハードウェアやソフトウェアではない。ヒューマン・ウェア (筆者の造語) である。人材である。例えば、メーリング・リスト、電子掲示板、ブログ (blog)、Wiki などの情報を共有する技術は利用者を選ばないが、利用するコミュニティと運用責任者によってそ

の活性に雲泥の差がある。

データベースとネットワークの目的を設定するのはコミュニティである。データとサービスの品質を維持するのもコミュニティである。データを活用するのもコミュニティであるが、それは中核から外延と広がっている。

データを操作する技術はこれからも進歩を続けていくが、技術に血を通わせて技術から最大限の恩恵を受けるのはコミュニティの専門家である。

文 献

1. Edwards, N. L. *Predictability: Does the Flap of a Butterfly's Wings in Brazil set off a Tornado in Texas?* . *The American Association for the Advancement of Science*, Washington DC (1972)
2. Erik, T. R. (宮下 尚, 牧野 聡, 立堀道昭訳). *入門 XML*, オライリー・ジャパン, 東京 (2004)
3. 藤田泰徳, 山崎文則. *オープンソースを使ったデータベースの構築 Apache サーバと PostgreSQL でデータベース*, セレンディップ (2000)
4. 丸山 宏, 浦本直彦, アンディ クラーク, 田村健人, 村田 真. *XML と Java による Web アプリケーション開発 (第2版)*, ピアソン・エデュケーション東京 (2002)
5. Miyachi, S., Nakayama, O., Yokohama, Y., Hara, Y., Ohmori, M., Komagata, K., Sugawara, H. and Ugawa, Y. *Japan Scientific Societies Press, Tokyo* (1989)
6. 日本実験動物学会調査ワーキンググループ. 過去 9 回実施された実験動物使用数調査結果の数値を通覧して. *実験動物* 42 : 675 663 (1993)
7. Rogosa, M., Krichevsky, M. I. and Colwell, R. R. *Method for coding data on microbial strains for computers. Int. J. Sys. Bacteriol.*, 21: 1A 184A (1971)
8. Schwartz Randal L. and Phoeni, T. (近藤 嘉雪訳) *初めての Perl (第3版)*, オライリー・ジャパン東京 (2003)
9. 菅原秀明. 「数値分類」鈴木健一朗, 平石 明, 横田 明 (編)「微生物の分類・同定実験法」, p. 231 248, シュプリンガー・フェアラーク東京, 東京 (2001)
10. Sugawara, H and Miyazaki, S. *AHMI: Agent to Help Microbial Information Integration. Nuc. Acids Res.* 31: 3727 3728 (2003)
11. Sugawara, H and Miyazaki S. *Biological SOAP servers and web services provided by the public*

- sequence data bank. *Nuc. Acids Res.* **31**: 3836-3839 (2003)
12. Sugawara, H, Tanaka, N and Miyazaki S. An e - Workbench for the Study of Microbial Diversity Design and Basic Functions. *Microbiolo. Cult. Coll.* **19**: 59-67 (2003)
13. 菅原秀明, 館野義男. 「理化学研究所における実験生物情報システム」NISLOの開発, 情報管理 **26**: 640-653 (1983)
14. 高橋麻奈. やさしいJava, ソフトバンクパブリッシング東京 (2002)
15. 竹形誠司. Java + My SQL + Tomcatで始めるサーバーアプリケーション開発入門, ラトルズ (2003)
16. Tateno, Y., Nemoto, U. and Sugawara, H. Bibliographical Data Base for Plant Tissue and Cell Cultures (IRIS). *Proc 5th Intl. Cong. Plant Tissue & Cell Culture* p. 787-788 (1982)
17. Tisdall, J. (水島 洋 監修・訳; 明石浩史, またぬき, 小林慎治 訳), バイオインフォマティクスのためのPerl入門, オライリー・ジャパン東京 (2002)
18. Wilkinson, Mark D. and Links, M. BioMoby: an open source biological web services proposal, *Brief. Bioinf.*, **3**: 331-341 (2002)