

データベース作成の一事例 ～電気化学データベース～

横浜国立大学工学部 有澤 博
仁木克己

はじめに 近年の科学技術の発達には、その成果として膨大な量の学術情報の生産とつながっている。これに対して、集大成した学術情報とコンピュータの大容量記憶装置とにあき、データベース管理システム (DBMS) とデータ通信システム (DCS) を用いて、データの迅速な流通をはかろうという計画が、所々で出はじめている。このようなものを一般には学術情報流通システムとよぶ。

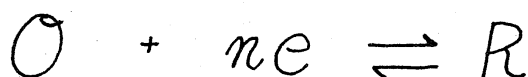
著者らは現在、文部省科学研究費「データベース作成」の助成のもとに、電気化学における電極反応パラメータのデータベース化を進めている。しかし、このような活動は、我国では先例が少なく、実際に行なってみると、多くの問題点や、新たな課題にぶつかることも多かった。本稿ではこれら諸問題を整理して報告したい。

電気化学データベースの内容

電気化学データベースは次の4つの構成要素から成る。

1. 電気化学の電極反応速度パラメータ
2. 電解質溶液の電気伝導度（伝導率）データ
3. 電解質溶液の活性度係数（活動度）データ
4. 1〜3に含まれるデータのソースレファレンス情報
（文献データファイル）

金属の電解製錬，精製，ソーダ電解など，産業の基幹となる電気化学工業は，電力消費型の産業で，これらの反応装置と設計する場合，最適条件（省エネルギーを含めて）を求めらるために，上記のものは基礎的な必須データである。本稿でとり上げるのは，このうち1の電極反応速度パラメータであり，電気化学データベースの中でも，最も重要なデータである。反応速度パラメータとは電極反応



（活性物質）

において，活性物質の種類はもちろん，電極の種類，支持電解質（反応物質ではないが，電子移動に寄与する物質），溶媒，添加物，温度等，様々の条件によって測定された平衡点に関するパラメータで，転移係数（Transfer Coefficient）と反応速度定数（Rate Constant）が主要なものである。⁽¹⁾

電気化学データベースのオリジナル・ソースは、世界各国で
 発行される学術雑誌に掲載された論文である。これらの文献
 点数は膨大な数にのぼり、前記の本質的な諸条件以外にも、
 測定方法に関するアイデアや、表現、評価方法などによって
 非常にバリエーションに富んだものになっている。

現在、国際化学会連合(IUPAC)の正規の依頼を受け
 て、電気化学反応データ収集の国際センターとしての役割を
 担うと同時に、文献や単純な数値情報でない、複雑な学術情報
 のデータベース作成活動のあり方について、模索を続けてい
 る段階である。

データベースの作成手順

本稿では、学術情報の収集と提供に関する一般的な議論は
 他の文献⁽²⁾にゆき、電気化学データの収集とデータベース
 化の具体的な手順に焦点を絞って述べる。現在までに行なっ
 てきた、あるいは今後次第に実行に移されることになる
 手順は次の通りである。

- (1) データの解析。電極反応データのデータ構造の解析
 およびこれを利用する側の要求の解析。これには、言
 うまでもなく化学の専門家による分析が必要である。
 しかしそこから得られるものは、一般には不十分、あ

いまいなもので、データベースに熟知した者による再分析、再構成が必須である。

- (2) データ収集 国内外の各文献ごとに情報の提供者を定め、それぞれの文献の別刷（またはコピー）を送ってもらうと同時に、文献中の必要なデータもサマリーしたもの（データシートとよぶ）を送ってもらう。データシートの作成は原論文から主観を交えずに情報抽出する必要があり、化学の専門的知識をもつ多くの協力者が必要である。データシート作成者を コンパイラとよんでいる。

- (3) 入力変換 集められたデータシートをいわゆる機械可読ファイル（一定の書式をもった MT ファイルなど）に変換する。データシートは化学の専門家ではあるがコンピュータの専門家では必ずしもるい人によって作成されるため、そのまま機械可読な形に変換できるには限らない。そこで、対象分野（化学）とコンピュータの双方に、ある程度の知識を有する人による作業を伴う。このようにしてできた機械可読ファイルも データファイル、データファイル上の書式も 内部形式とよんでいる。

- (4) Check and Authorize データファイルに作成され

に全データについて、化学の立場から見て明らかに誤りや、入力時にあけるミス等を修正、除去し、データファイルとして完成されたものにする。

- (5) データブック作成 データファイルも編集し、重要な項目についてはインデックスを作成して、プリントアウトも本の形にする。これをデータブックとよぶ。データブックは図書館等で総目録として利用される他、データの提供者（外国）に配布することも考えられる。（磁気テープ媒体そのものによる提供も考えられるが、外国の中には、必ずしもコンピュータが手近で利用できるとは限らない所もあり、書籍の形の提供も重要である）

- (6) DBMSによる運用 データファイルをもとにデータベースとDBMS（データベース管理システム）により運用する。現在商用化されているDBMSは、もともと事務計算を主体として考えられており、数値データや化学式などをもとり扱うようにはできていない。また検索論理も十分な能力をもっていないことが多い。したがって、いくらとも学術データベースのあり方を展望するという観点からは、既存のDBMSで満足できるものはなく、システムの設計、試作も必要である。

データベース作成の現状と問題点

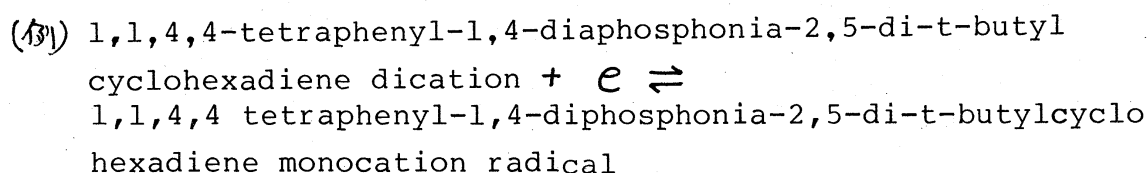
以上の考察のもとに、ここでは前項の手順の具体的な実施方法と、問題点について述べる。

(1) 入力変換に関する問題

図1にデータ・シートの例を示す。ここでも明らかなように、データ形式は、項目は分けているものの、かなり自由に記入している。本プロジェクトにおいても、データの収集段階において、各項目の記述法や欄指定を厳密に規定するべきだとの意見が初期には見られた。しかし実際にはコンパイラに必要以上の負担をかけ^{現在では}ため、むしろ原論文の意図を反映すべく、できるだけ自由形式にデータシートを作成してもらい、これをもう一度コーディングシートとよばれるシートに書き写している。コーディングシートも原論文の意図を正めることのりよう注意が払われているが、一方コンピュータ入力かきやすいように工夫されている。たとえば原論文で複数の測定（条件を変えたもの）が行なわれている場合、データシートでは1枚になるが、コーディングシートでは測定ごとに独立したシートになる。また文字種やデリミタ等もコンピュータ側で認識しやすいような形になっている。コーディングシートの例を図2に示す。これらの作成は現在は化学系学科を卒業した女性数名にお願いしている。

(2) データファイル形式の問題

電気化学データベースには有機物を含む物質名を記憶する必要があり、非常に長いものもあるが、短かくて済むものもあり、その差が激しい。長い反応式の例を下に示す。



したがって可変長のフィールドをもつことが本質的に必要である。また電極の表面処理法など、特殊な測定法のみ記入され、コメント的な役割をもつものもある。以上を勘案してデータファイルの形式(内部形式)としては、(イ)フィールド長を10バイト単位に任意に設定できること、(ロ)記述のする項目は、フィールド単位にも、大項目(論理レコードタイプ)単位にも省略できること、(ハ)相対形式の記述(直前に定義されたデータに対して変更点だけを再定義するもの^{*})と許すこととを条件として次のように設計した(表1参照)。

*注.

本データベースの特徴として、他の条件とすべて同じにして1つか2つのパラメータだけを動かした測定(たとえば温度変化など)が多く見られる。このような場合と冗長性を少なくするため、変更したところのみデータとしてもつのである。

- ・ 1レコード 80バイト, 1フィールド 10バイト,
- ・ 大項目 (「反応式」, 「電極」, 「速度定数」, など) ごとに論理レコードを形成する。論理レコードはレコードをまたいで定義される。論理レコードの1部は省略可。
- ・ 1論理レコード中に, 各項目をフィールド値として記述する。フィールドは識別子と伴って与えられ, 省略可能である。また "—" 識別子による継続も可能。

レコード例を下に示す。これは図2のEAS (活性物質名とその濃度) 項目に対応している。(識別子については図3を参照)

ES	:MTCr(CN)6(3-)	:V12.0E-3	:**	:**	: 00000000
	10	10	10	10	10

(3) 表現の多様性

上例にもあるように, 測定値などのデータでは有効数字や書き方, 精度によって, いろいろの表現方法があり得る。

例. 1.2 ± 0.05 , $1.15 \sim 1.24$, $\sim 1.23 \approx 1.2$

データ作成者は測定法, 機器精度などから, とも適切な表現方法を選んでいると考えられるので, どれかに統一することは好ましくない。本データベースは上の4通りを認めることとしてある。また濃度値などでは 0.5 mol/l のような表現の他に「Saturated」(溶けるだけ溶かした) のような例もあり, 将来いっしょに INVERTED INDEX をつける場合に困難を伴うものと思われる。

(4) データ構造とデータモデルの問題

電気化学データと DBMS によって運用しようとする場合、通常のデータベース (formatted) に比べて問題点が多い。まずデータ構造として、項目が多くかつ分割 (階層化) の度合いが高い。たとえば

- ・ 1 文獻は複数測定を含む
- ・ 1 測定とは同一物質 (化学種) に対して、いろいろなパラメータ (支持電解質, 溶液, 温度など) と変化したものの 1 つである。
- ・ 測定に用いられた測定法についても、その精度や、変更点、対象パラメータ範囲などの属性がある。

容易に分かるように、これを関係データベースのように、フラットな複数ファイルへの分散で構成しようとした場合、ファイル数が増え、しかもファイル間にもまたぐ関数従属性が頻発する。

例. [測定 文獻], [測定 支持電解質], [測定 溶媒], ...
 [測定 測定法 変更点]

また検索する場合を考えると、そのキーとなる検索項目は非常に多様である。したがって、Join にあたる演算 (集合の結合) とハッシュに行なうことになり、データベースとしての検索効率が低くなるおそれがある。またユーザに Join 操作

を実行させるのは、誤りをおかしやすい。

さらに、このようなデータの場合、多様な属性 (attribute) が存在するが、各測定についてみると、実際に値が与えられる属性値はごくわずかである。たとえばある測定に限ったときのみ得られる測定値や、またその補正が必要な場合などである。以上から考えると、通常のデータベースのように、「ぎっしりつまったレコードの集まり」としてのファイル構造では間に合わず、新しいモデルにもとづくデータベースを考へなければならぬと思われる。我々は、この問題については、新しいモデル AIS によるアプローチを考えている。⁽³⁾

現状と今後の課題

昭和57年 2月までに、7000データシート程度の MT 化がすすみ、今後データの収集は順調に伸びつづけてゆくものと思われる。現在、これらのデータもとろあえずデータブックという形で公開することも進めており、これは、内部形式に近い形式も、できるだけユーザに分かりやすい形でフォーマット化したものであり、いくつかの項目についてはインデックスも設ける。その形式例を図3に示す。白スキの部分に値が入れられる。またこの枠は、実際には、データ長に合わせて上下方向に伸び縮みする。この形式のデータは、ラインアリメント

カを製本したものを公開すると同時に、最新の資料を、N-1ネットワーク（大学間学術情報ネットワーク）も含めて、オンライン的にアクセスできるようにする予定である。

前に述べたように、この種のデータベースの作成は我国では先例が少なく、また情報の内容が高度に専門的で、利用者のレベルも高いだけに、多くの未解決の問題がある。最終的には、DBMSまでを含めて、まったく新しい思想でシステムを設計しなければならなくなると思われる。今後、学術情報システムに広く利用できるDBMSの試作を進めるとともに、電気化学データベースの高度な運用を試行してゆきたいと思っている。

参考文献

- (1) 玉虫・伊豆津他：電極反応の基礎，共立出版
- (2) 有澤：大学図書館におけるデータベース・システムの役割，大学図書館研究（掲載予定），（1982）
- (3) 有澤：Entity-Associationモデルによるデータベース設計
情報処理学会「アドバンスド・データベース・システム」シンポジウム，（1981）

謝辞 データベース設計、データ作成にご協力いただいている 三國房子氏
他データベース作成室の方々，プリントアウト・処理を設計・製作していただいている
森雅一氏 他の方々に 謝意を表します。

表1. 内部データ形式

項番	先頭からの バイト数	大きさ	タイプ	内 容
1.	0	2	A	論理レコードタイプ名
2.	2	3	A	補助記述子 相対/絶対形式 種別
3.	5	1		"：" (デリミタ)
4.	6	2	A	フィールド記述子
5.	8	10	A	フィールド値 (データ)
6.	18	1		"：" (デリミタ)
7. ? 9.	19			4~6 のくり返し
10. ? 12.	32			4~6 のくり返し
13. ? 15.	45			4~6 のくり返し
16. ? 18.	58			4~6 のくり返し
19.	71	1		(空白)
20.	72	8		シーケンスフィールド

注意 ・ フィールド記述子 "****" は Filler 用として 予約されている。

・ 論理レコードタイプ名およびフィールド記述子における"ー"は継続指定。

図1. データシート例

Type of Electrode Reaction		$\text{Cr}(\text{CN})_6^{3-} + e \rightleftharpoons \text{Cr}(\text{CN})_6^{4-}$	
Electrode System	Working electrode	Hg (DME)	
	Electroactive species	$\text{Cr}(\text{CN})_6^{3-}; 2.0 \times 10^{-3} \text{ M}$	
Medium (supporting electrolyte, solvent, surface-active substance, etc.)		1.0 M KCN	
Temperature		25°C	
Electrode potential	$E^\circ, E^\circ', E'_{1/2}$ etc.		
	Range of the measurement		
Other conditions			
Transfer Coefficient	anod.	meas.	
		cor.	
	cathod.	meas.	$\alpha = 0.59$
		cor.	
Rate Constant	meas.	$k_p = 0.24 \text{ cm sec}^{-1}$	
	cor.		
Activation energy (including Gibbs energy, enthalpy, and entropy of activation)			
Diffusion Coefficient			
Experimental Conditions	Method	AC polarographic measurements	
	Time-domain of measurem.	d.c. scan rate 25 mV/min.	
	WE (geometry, etc.)	Hg (DME)	
	Counter electrode		
	Reference electrode		
	Hydrodynamic condition		
Remarks			
Reference	Authors	Eric R. Brown, Hsueh L. Hung, Thomas G. McLeod,	
	Journal	Donald E. Smith and Glenn L. Boorman Analytical Chemistry 40, (1968), 1424	

COMPILER: F. Mikuni

図2. コーディングシート例

ECDATA Coding Sheet

03302

YNU Electrochemical Data Center

SNO			177 — $\phi\phi$ — $\phi\phi$		
OXST			$\text{Cr}(\text{CN})_6(3-)$		
ORXN			$\text{Cr}(\text{CN})_6(3-) + e \rightleftharpoons \text{Cr}(\text{CN})_6(4-)$		
NOSN			1		
WE			; Hg ; DME		
EAS			2.0×10^{-3} ; $\text{Cr}(\text{CN})_6(3-)$		
MEDS	A		H ₂ O		
MEDE			1.0 ; KCN		
MEDA					
PHIS					
TEMP			25		
ELPO					
OCN			dc scan rate = 25 mV/min.		
ATCM					
ATCF					
CTCM	A		$\phi, 59$		
CTCF					
ATSL					
CTSL					
RCME	ϕA		$\phi, 24$		
RCFR					
ACEN					
DOX					
DRED					
METD			ACP } fundamental harmonic analysis		
RANG					
STRT					
REFE			SCE		
HYCD					
RXMN					
AUTR			Brown, E.R.; Hung, H.L.; McCord, T.G.; Smith, D.E.; Braman, G.L.		
JOUR			Anal. Chem., 1968, 40, 1424 (Eng)		

図3. 出力形式設計図

1. P. FORMAT																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260
261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280
281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340
341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380
381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400
401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440
441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460
461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480
481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500
501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520
521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540
541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560
561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580
581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600
601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620
621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640
641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660
661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680
681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700
701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720
721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740
741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760
761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780
781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800
801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820
821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840
841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860
861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880
881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900
901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920
921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940
941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960
961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980
981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000