

報 告

データ・コードの標準化の体系調査とその背景*

上 田 陸 奥 夫**

1. ま え が き

文明が進歩し、高度の経済成長を背景に社会が複雑化してくると、社会の要求としての情報の需要と供給が増大してきた。これは、グーテンベルヒの活字の発明、あるいは産業革命のときにもあったが、近代社会においてはとくにいちじるしい。

情報の需要と供給の多くなった理由は、第1に工業化社会から情報をさらに必要とする脱工業化社会へ、すなわち情報化社会への推移と、第2にコンピュータの急進歩が情報の処理能力の拡大を可能にしたためである。

情報処理技術の進歩により、将来における情報は自然文の形をとることも多くなると考えられるが、データはその処理の過程においては、処理効率の点からコード化される。すなわち、コードによってシステム化される。

一方では、情報化時代の到来とともに、情報の交換、または共同利用の需給が増大してくると、情報、すなわちデータはコード化された形において授受されることが多くなる。このことから、データをコード化して処理効率をあげるだけでなく、データ交換または共同利用のためにも、コードの統一が必要となってくる。とくに、利用頻度の多いコードについては、標準化の必要が生じてくる。

ここで、システム技術として、データとそのコード化に対するコンピュータ利用者の考え方をもとに、現存するコード体系の内容の把握と、標準コードの必要度を調べて、標準コード設定の資料を作ることにした。また、標準コードの需要に関するテーマとして、データ・バンク作成過程において、標準コードがどれだけ必要かを示した実例をかかげることにした。

* Survey and its Consideration for the Standization of Data Coded Representation by Mutsuo Ueda (Social Insurance Agency)

** 社会保険庁

前者は、日本情報処理開発センターの実施したデータ・コード標準化体系調査報告書であり、後者は、工業技術院の実施した事業所共用ファイルの作成および処理実験である。いずれも筆者自身の体験をもとに説明しよう。

2. 情報処理システムにおけるデータ・コードの役割

情報処理における情報とは、広義の知識の所産であり、さらに狭義の知識とデータとに分かれる。ここでいうデータは、ある定められた方式にしたがって、事象とか思想の表現されたものであり、何らかの方法で伝達・処理できるように記号的表現をとるものである。その記号的表現がデータのコード化であり、これはデータをシステム化したことになる。

情報処理システムとは、具体的に、原書類の記入、整理およびコード化、カードへのパンチ、オンライン処理によるデータの伝達、データのコンピュータ処理、および処理結果であるアウトプットの利用などの各段階を含んでいる。したがって、これらのシステムのなかで、データの伝達、処理、分類、記録および検索などの効率、すなわち、全情報処理システムにわたってシステム効率をあげるために、データのコード化という問題がおきてくる。情報処理システムが複雑になってくると、とくに、その効果の評価がさらに困難になってくる。

情報処理システムにおけるコードの考え方は、つぎのように拡張されてきている。

- すなわち情報処理システムの範囲の拡大
これによって、データの質・量ともに多くなり、システム効率の向上はコード化しかない。
- アプリケーションの複雑化
データの処理アプリケーションが多くなり、データの伝達、処理、分類、記録および検索が高度化するにつれて、標準コードの必要が多くなる。

○ データの交換および共同利用の需要増
経営情報管理システムのような高度の情報管理
には、データの交換および共同利用が絶対必要で
あり、そのコードは標準化しなければならない。

このように、情報処理システムの効率向上のために
データのコード化が、適用範囲の拡大に従って、単
なるコード化でなく、標準コード化となり、システム相
互間の総合システム効率をとりあげるまでに変化をき
たしているのである。

3. データ・コードの標準化の必要性

物品とか物体、物質のような有形的なもの、あるい
は形状的なものについては、標準化はつきものである。
たとえば、事務用品の形状、あるいは機械部品の寸法
のようなものがこれにあたる。これは対象も明確であ
り、標準化の目的もはっきりしており、製造者と利用
者の利害が一致する場合が多いためである。

このことは、電子計算機関係についても全く同じで
ある。電子計算機と情報処理に関しては、ご承知のと
おり、国際標準化機構 (ISO) のうち、TC 97 が国際
標準化の中心となっている。

ISO/TC 97 の標準化活動の経過を見ても、物質的
な面、つまりハードウェア関係の標準化が先行し、論
理的、無形的あるいは抽象的な面、つまりソフトウェ
ア関係の標準化は後になっている場合が多い。

第1図の ISO/TC 97 の機構図のうち、前者に属す
るものとして、SC 2, SC 3, SC 4, SC 6 および SC
8 がある。これらの標準化活動は、国際的にも、国内

的にもはやかである。これに対して後者に属する
SC 1, SC 5, SC 7 および WGK は、何となく弱い
感じである。

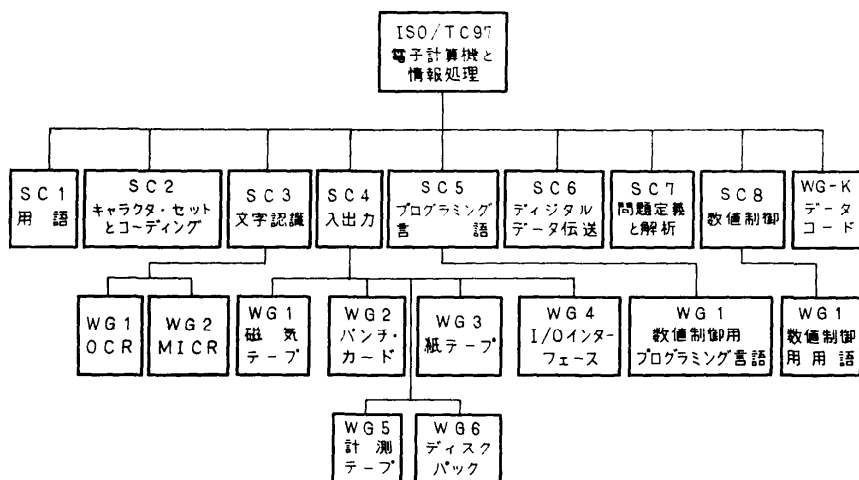
ただし、WGK, すなわちデータ・コードの標準化
は、ソフトウェアの固有なものと異なり、情報処理の
システムの思考の問題としてとりあげられている。

前章に述べたようなデータ・コードの重要性、すな
わち、情報処理システムの範囲の拡大、アプリケーション
の複雑化およびデータの交換・共同利用の需要増
という新しい要素群が、従来の標準化の様相を変えよ
うとしている。要するに、データ・コードの標準化の
必要性は、ISO/TC 97 の他の分科会のそれとは異な
ったこれらの要素群によって急激に増加している。

データ・コードの標準化の必要性をいまずこし具体
的な経営情報管理システム (MIS) におけるデータの
問題として扱ってみよう。

企業の MIS における、企業が外部から収集しうる
データには、経費と機密の関係で限度がある。したが
って、このような外部データは情報サービス業者か、
他の企業または官庁のデータ・バンクからの提供を待
つしかない。とくに、官庁からデータは統計資料が多
く、これには多額の調査費と処理費がかかっている
(例: 45 年国勢調査の費用総額、約 55 億円)。

この外部データの提供は、媒体は磁気テープが多い
と思われるが、問題はその内容である。データ・コー
ドが自企業のそれと異なるときは、コードの変換が必
要となり、データの交換・共同利用の効率が悪い。し
かも、このようなコードは、いわゆるキー・コード



第1図 ISO/TC 97 機構図

(Key Code)が多いので、データのファイル構成上まで影響を与える。たとえば、逐次的なファイル構成(Sequential File)のキー・ワードのコードが異なる外部データのファイルでは、コード変換のほかに、ファイル構成の再編成の処理を行わないと利用できない(とくに大量ファイルにおいて)。

外部データと企業内のデータとのコードが統一されていれば、上記の損失はなく、オンライン的な処理が可能となり、外部データの利用価値もあがってくる。このような需要が多くなると、コードは標準化へと進まなければならない。多数の企業、官公庁で共通に使用されるデータ・コード(キー・コードが多い)の標準化は、このようにして生まれる。

MISなどの高度なデータの取扱いだけでなく、経理業務の一元化(給料振込、公共料金の自動払込など)のために、磁気テープをデータ交換の媒体として、データ交換が行なわれるようになってきている。この場合、それらのデータに含まれるコードの標準化は当然であり、理屈ではなくなっている。

4. データ・コード標準化の背景

——体系調査を中心に——

標準コードの必要性は十分に認識できたが、その標準をどこで決めるかは、利用者ならずとも大いに気になるところである。設定の機関としては、国際的な問題もあるので不本意の向きもあるが、政府関係機関にならざるを得ない。しかし、標準コードの内容は、使用者協議のうえ決定すべきものであり、またそうでなければ使用者の同意は得られず、せっかくの標準化も利用されなくなる。

企業にとっても、官公庁、その他の団体にとっても、標準コードの実施となると、現状コードの移行において多少とも利害得失があるので、その設定は大いに問題がある。各使用者が標準コード設定のためには損失が目立つようだと、理想的な標準コードにしてもいささか心もとない。ここにハードウェアの標準化と異なったむずかしさがあると思う。

使用者がどのような標準コードの設定を望んでいるか、増大するデータの伝達、処理、分類、記録および検索などをいかにこなすか、またはデータの交換・共同利用に対応する標準コードはどうあるべきかは、各企業の性質、規模、経営状態などによっても異なるが、要するに情報処理システムに適合したものが望ましい。ここで、日本情報処理開発センターの事業として、

データ・コード標準化のための体系調査を行なうことにした。

このようなデータ・コードの傾向を把握するため、広く使用者側の意見を調べる必要がある。全使用者(中規模以上のコンピュータ利用者約2,000)の調査が理想的であるが、費用と時間の関係で、約700の社または団体に調査票を送り、320通ほどの回答を得た。

調査の内容は、つぎのとおりである。

(1) 現在使用中のコードについて

コードの名称、作成年、目的用途、作成機関、コードの構造、桁数、適用範囲、コード表の提示の可否

(2) 標準化を希望するコードについて

コードの名称(40ほど指定)

作成の緊急度

(3) データ・コードの標準化に対するご意見

この調査をもとに、調査の集計と分析を行ない、データ・コードの標準化体系をまとめてみた。

この調査の概要は、つぎのとおりである。

各コンピュータ使用者は、自社向きではあるが、システムの開始時から多くのデータ・コードを使用している。しかも、将来の予測に対する考えも述べられており、現状コードに不満足ながらも、現システムのための標準コード設定の意図を持つ企業が多い。

第1表はユーザの業種ごとに、各企業または団体における現在使用中のコード(質問1)の回答数である。これによって、標準コードの需要の傾向がわかる。

第2表は、各企業または団体が標準化を希望するデータ・コード(質問2)の回答数である。第1表とともに、データ・コードの標準化の貴重な資料である。

第3表は、ユーザの業種ごとに、各企業または団体におけるデータ・コードの標準化のシステムに関するさまざまな意見をとりまとめたものである。標準化について大きく5つの意見にまとめてみた。

情報処理機械の昔と今、すなわち、PCSのシステムとコンピュータのシステムとは、データの伝達、処理、分類、記録および検索などは異質のものであるので、データ・コードも異なった設計を持つのが当然である。しかし、皮肉なことに、現在大型機のコンピュータを持つ使用者はほとんど長期間にわたる(20年以上はザラである)PCSの使用経験を持つ。このために、データ・コードはPCS時代のものをそのままコンピュータで利用されている例が非常に多い。これらの企業および団体では、MISなどのような高度のシ

第2表 標準化希望順位

コード名	回答数			コード名	回答数		
	A	B	C		A	B	C
地域名・国名・国際機関名	71	24	15	産業 (業種)	133	11	5
外国都市名	34	21	27	銀行・金融機関	117	30	8
都道府県	176	12	4	勘定科目・会計科目	148	23	4
市区町村	157	17	8	固定資産目 (償却分類)	126	22	6
住居表	85	31	11	株式銘柄	34	25	17
年号	142	10	5	公社債銘柄	30	23	18
日付	125	7	7	国際収支	8	22	22
時刻	85	11	13	建築物 (構造など)	42	18	22
性別	129	12	8	鉄道 (駅名)	40	23	18
職業	94	35	14	港湾 (空港)	22	24	14
職業種	102	34	12	計量単位	85	22	7
職能・技能・資格	85	39	13	貨幣	57	20	11
学歴	108	40	11	函書	23	35	23
個人番号	61	34	29	新業所	18	19	22
企業 (会社名)	99	44	13	工業所	13	17	19
業所	61	28	17	力文字・漢	89	20	11
学校・研究機関	57	33	16	学部・学	3		
官庁	65	34	13	決済条	2		
材料	55	24	13	所得税	1		
用途	28	26	14	自動車	1		
加工法	20	27	17	病名	2		
商品	76	30	9	英数字	1		
商品取引 (商品流通)	32	14	15	外国	1		
薬品	21	22	18	業種			

(注) A: 2年以内, または早急に。
 B: 3~5年, またはとくにいそがない。
 C: いまのところ必要ない。

第3表 産業別意見分析

意見	合計	産業																					
		鉄業	建設	食品	繊維	紙パルプ	出版	石油・化学	窯業	鉄鋼	機械	電気	電子計算機	輸送	その他	商業	金融	運輸	電力	学計	組合・研究	中央官庁	地方公共団体
標準化が必要	12								1		3					3		2		2		1	
早期に決定	28	4	1	1	1		1	1	1		3				5	5		1	1	1	1	2	1
目標と時期	2																	1					
短期間で修正しない	1																						
全国的な統一	16						2			5	1		1	1	3			1				2	
制定機関をはっきりと公表の方法	22		1				2		1	1	2				4	4	1	1	1		6	1	1
メンテナンス	17						1		1	1	1				2	3	2	1		2	1	2	1
各界の意見をきく	11						1		1	1	1				1	2	2	1		2	1	2	1
方法	5						1		1	1	1										2	1	1
JISより業界で	6	2	1		1						1					1							
企業間の統一 (大ワクだけJIS)	14						2		1	2	1	1			1	1				1	1	3	
具体的なコードのみ	19						4		1		1	1			1	1			5	1	2	1	1
桁数は少なく (基本的なもの)	18	1	1	1	1				1	2	1	1		1	2	1	1		1	2	1	2	1
実用性のあるもの	11		1	1	1					2					1					2	1	3	
融通性 (共通性) のあるもの	11						1								2	1	1	2		1	2	1	2
記憶しやすいこと (理解)	5				1					1										1			
数字だけに限る	3											1										1	
体系的に考える (分類など効率化)	32	3	1		1				1	3	2	2		1	1	2		3	3	3	3	6	
データ・フォーマットの標準化	2		1																		1		
OCRとの関連	1				1																		
関係法規と秘密保護	6		1		1			2							1	1							
移行の問題	10									1	1				2	1		1			2	2	
経費、強制度	12									1	1	1				3	1		1	1	3		

システムの需要がとくに強いのである。

このような状態にありながらも、現状コードを批判し、社内のデータ・コードの統一化を目標として、まず社内の合理化をはかり、将来はデータ・コードの標準化を期待し、よりよい情報処理システムの希望を持っている企業が多い。これは第2表の標準化希望のコードの多いのをもてみてもわかる。

しかも、その希望コードは、かつて当学会が43年に出版した“コードの手引”作成にあたって調査した事例に明らかなように、自社使用中のコードと標準化希望の同名コードとの内容は一致の要請はなく、標準化コードが適切であれば、それにしたがうという空気が強い。自社だけ別のコードは使用できない、それは企業にとってマイナスであると割り切ったみかたが多い。

5. データ・コードの標準化の現状

現在、データ・コードの標準化はどのように進んでいるであろうか。標準化制定の方法として、法規などによる強制力を持つものか、企業団体あるいは使用者団体の自発的協定によるものかのいずれかが考えられるが、強制力は言論の自由に反し、自発的協定などは守られたためしがないという欠点を有している。この両者の欠点を補って、制定を国家機関(日本規格協会)で行ない、標準コードの使用については強制はしないという立場で、日本標準規格(JIS)において定めるのが順当であろう。

しかし、制定までの原案作成、専門委員会などは完全に全国民に開放されていると断言できる。この点については、諸外国も全く同様である。したがって、ISO/TC 97 という国際機構も、このような各国の自由な意見を反映して、国際標準化の活動も円滑に行なわれている。

全国民に開放すべきであるという理由は、一般物質の標準化(たとえば、机、家庭用品など)と同様にデータの伝達、処理、分類、記録および検索の過程が家庭に、個人に及ぶからである。これは情報化時代では当然であって、実際にキャッシュレス方式、CAI方式などの対個人情報サービスが発生しており、標準コードの需要の範囲が大きく拡張されようとしている。

現在までに、JISとして標準コードに制定したものをあげると

都道府県コード	JIS C 6260
市区町村コード	JIS C 6261

日付の表示	JIS C 6262
時刻の表示	JIS C 6263
性別コード	JIS C 6264

これらのコードあるいは表示方法は、データ処理の基本的項目であり、データ・フォーマット上でもキー・コードになることが多い。このうち、都道府県コード、市区町村コードは国内統計用コードをもとに、日付の表示、時刻の表示についてはISOの推薦案をもとにしている。

このほか、JIS制定予定のものに、つぎのものがある。それぞれ右記のものを参考資料としている。

産業コード	日本標準××分類(統計用)として5年ご と改訂
職業コード	
商品コード	
計量コード	

6. 事業所コードの現状

—データ・バンク作成実験より—

これは、工業技術院における“電子計算機利用に関する技術研究会”のなかのプロジェクトとして、“事業所共用ファイル作成WG”の報告の抜粋である。

(1) 共用ファイルの作成の目的

政府各省庁における電子計算機の利用は急速に進んでおり、それに伴って、磁気テープ、磁気ディスクなどにファイルされる情報の量は年々大幅に増加している。

一方、これらの情報の利用状況をみると、その多くは情報を収集した省庁でのみ利用され、技術面、その他いろいろ問題があるにせよ、他省庁の利用にはあまり供されていないように思われる。

収集された情報の有効利用をはかるためには、別々に収集した情報を関連づけさせ、それらの情報の分析を行なう必要がある。

以上の観点から共用ファイルの試作および利用実験を行ない、情報の相互利用の技術的基盤を固めるとともに、将来の行政情報ネットワーク・システムにおけるデータ・バンクの形成に備えて、その問題点の解決をはかることを目的とした。

(2) 調査対象と使用電子計算機

共用ファイルの試作は事業所ファイルを作成することとしたが、その調査対象、使用電子計算機等は第4表のとおりである。

対象事業所は東京都の立川市、府中市、三鷹市など

第4表 調査対象範囲と電子計算機

項目	事業所統計	工業統計	商業統計	社会保険	労災保険	失業保険
調査対象範囲	甲 民 営 乙 サービス 丙 国	甲 20人以上 乙 19人以下 丙 本店	甲 法人 乙 個人 丙 飲食店	5人以上	5人以上	5人以上
調査年月またはデータ収集年月	41年7月	42年12月	41年7月	44年8月現在	44年11月現在	44年11月現在
周期	3年毎	毎年	2年毎	月1回更新	月1回更新	週1回更新
電子計算機	IBM 360	NEAC 2200	NEAC 2200	FACOM 230-50	UNIVAC III	UNIVAC 1108
磁気テープ	9トラック 800 BPI	7 556	7 800	7 800	7 556	7 800
レコード長	77(B)	396(C)	260(C)	444(C)	139(C)	36(W)
ブロック長	77×20(B)	336×8(C)	260×10(C)	444×20(C)	139(C)	900(W)
対象事業所数	28,133	不明	不明	3,752	4,907	5,032

第5表 事業所に関するコード

ファイル	コ	ー	ド
事業所統計	府 県, 市区町村,	調 査 区,	事業所番号
工業統計	" "	" "	事業所番号
商業統計	" "	括 番, 票 番	番 号
社会保険	事業所, 局 署,	事業所番号	
労災保険	局 , 署 ,	基幹番号, 技 番号	
失業保険			

13市および北多摩郡の一部で、従業員5人以上の事業所を対象とした。

(3) 作成の手順

事業所に関する各ファイルの結びつけを行なう場合、各調査とも同一の事業所コードがついているのが最も望ましいが、今回の調査においては、第5表のように、事業所コードのつけ方がそれぞれ異なっていた。

第5表のように事業所コードのつけ方が異なる場合のマッチングの方法には、つぎの3つの方法が考えられる。

- (i) 調査票または原書類にさかのぼり、目視によりマッチング、転記を行なう。
- (ii) 事業所の名称、所在地が磁気テープなどにファイルされている場合は、それらを電子計算機によりマッチングさせる。
- (iii) (i), (ii)の併用、すなわち事業所コード(キー・コード)を目視によりマッチングさせ、他のデータは電子計算機を利用し、結合させる。

今回はすでに磁気テープにファイルされている情報をなるべく活用するとともに、(i)の目視によるマッチング、転記のほう大な労力を回避する立場から(ii)と(iii)の方法を採用した。

すなわち(ii)の方法により、事業所の名称、所在地を記録してある社会保険データと失業保険データのマ

ッチングを行ない、その他のデータについては、磁気テープに事業所の名称、所在地が記録されていなかったため(iii)の方法で社会保険データの名称、所在地のリストを各データの調査票または原書類とのマッチングを目視により行ない、その各事業所コードの対応表を基に、電子計算機により事業所に関するデータの結合を行なった。この手順を図示すればつぎのようになる。

(4) 事業所ファイル試作の結果

今回のファイルにはできる限り、原データの内容を盛り込むこととし、1事業所あたり約900桁となった。

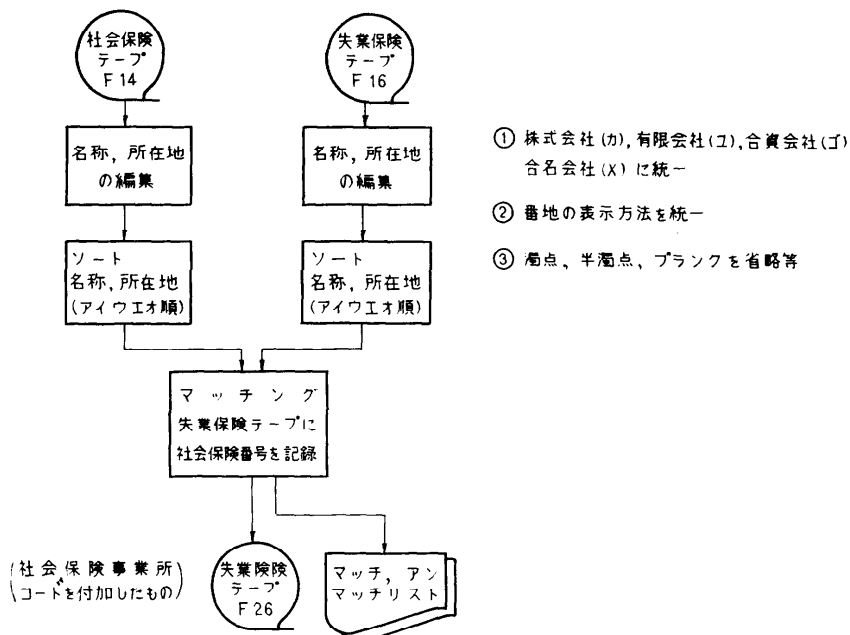
また、マッチングした結果はつぎのとおりであり、マッチング件数は必ずしもよいとはいえないが、今後データ・バンク、情報ネットワークシステムを考える際には、さらに検討しなければならない問題が多いといえる。

事業所ファイルの試作結果

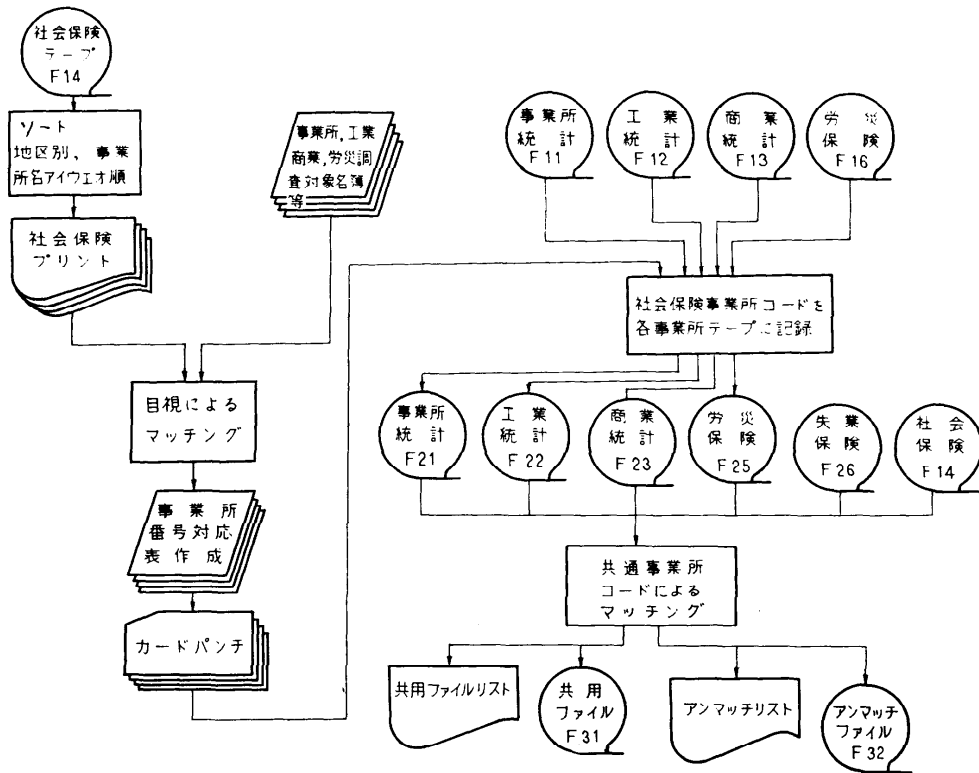
ファイル名	マ	ッ	チ
事業所統計	1,991		
工業統計	922		
商業統計	472		
社会保険	—		
労災保険	1,738		
失業保険	2,302		

以上の共用ファイル作成の過程をみると、作成を困難にし、アンマッチの件数が多いのは事業所コードの不統一が最大の原因である。もし、事業所コードが統一していれば、単なるマッチング作業として、目視作業、転記、事業所の名称、所在地による照合などのわずらわしさから開放されていたであろう。

単なる1コードの不統一が、情報処理システムに与える影響のいかに大きいかを身をもって体験した好例である。



第 2 図 電子計算機によるマッチング処理フロー ((ii) の方法)



第 3 図 目視および電子計算機によるマッチング処理フロー ((iii) の方法)

7. データ・コードの標準化の将来

現在はシステムの時代であるという。コンピュータを中核とする情報処理システムは、まさにその好例である。このシステムにおけるデータの処理効率を向上するものが、データ・コードの標準化である。しかし、システム全体として考えるとき、システムの各要素の標準化、たとえば、磁気テープ、I/O 装置あるいはプログラム言語がある。もちろん、データ・コードも多少内容は異なるが（システムの思考には他のものよりは近い）その一員である。現在のシステムはこれら各要素によって成り立っている。

ここで考えられるのが、各要素の標準化のほかに、システムとしての標準化、別の表現をすると総合的な包括的なもの、あるいは組織的なものの標準化かも知れない。システム標準化というものの中味はよくわからないが、情報処理システムもその一部なのか、または全部かも知れない。これが将来の標準化の姿とするならば、逆にシステム標準化が各要素の標準化と矛盾

が生じてくるおそれがある。すなわち、システムの標準化が他を圧倒するかどうかということである。

たとえば、オペレーティング・システムを例にとってみよう。OS の考えかたはシステムの的であり、その手法の結果が判明し、さらに高度化されようとしているが、ここで OS の標準化の期待も声も多い。もし OS が標準化への道をとるとすれば（実際はまだだが）、システム標準化の典型となるであろう。このシステム標準化の前には、ハードウェアの標準化は非常に弱いものとなるのではないか。

さて、データ・コードがシステム標準化の一部として考えられたとき、データ・コードがそのなかにどのような形で他の要素と関連を持つかは不明である。その理由として、各要素は情報処理技術の限らない進歩によってどんどん変化するからである。しかし、データの伝達、処理、分類、記録および検索などの方法に画期的な変化がない限り、データ・コードのシステム的な変化も遠い将来のことであろう。

(昭和 45 年 8 月 5 日受付)