

磁気媒体・パソコン通信による全国的規模の教育情報収集の試み

井上 常茂

坂谷内 勝 吉岡 亮衛

国立教育研究所研究協力者
(佐賀県立鹿島実業高等学校)

国立教育研究所

従来、コンピュータ処理を前提としたデータ収集は、調査票の送付に始まり、調査票への記入、回収、機械可読のデータへの変換という作業手順で行われてきた。

本報告書は、昨今のパソコンやワープロ専用機の普及を考慮し、これまでの調査票によるデータ収集に代えて、全国のデータ提供者が、それぞれ保有する多種多様な情報処理機器や、文書編集用ソフトウェアを使って作成した機械可読のデータを、磁気媒体やパソコン通信で収集を試みた結果、および、その実用化に向けての諸問題について取りまとめたものである。

On nationwide collecting
of Educational Information
using Floppy Disk or PC Network

Tsuneshige Inoue

Masaru Sakayauchi Ryohei Yoshioka

A research collaborator of National
Institute for Educational Research
(Saga Prefectural Kashima
Vocational High School)

National Institute for Educational Research

539 Takatsuhara, Ooaza, Kashima,
Saga 849-13, Japan

6-5-22 Shimomeguro, Meguro-ku, Tokyo 153 Japan

So far, the procedure for collecting data to be computerized is sending out questionnaires, having them filled in, receiving them back and changing them into computer readable data.

In this report, in consideration of the recent spread of personal computers and word processors, we try to collect data by personal computer networks or floppy disks, asking informants to supply their information in computer readable form using their own various information processing machines or word processor software instead of collecting data by questionnaires as at present. We also summarize the diverse practical problems raised.

1. はじめに

全国の教育センターや教育研究所等（以下、「教育センター等」と呼ぶ。）のうち、教育情報システムを整備し、教育情報データベースを作成して、各地域の教育関係機関へ教育情報を提供している機関のほとんどは、データベースの作成・更新用のデータを収集するために、調査票を利用している。

国立教育研究所でも、従来、データベースの作成・更新をするために、全国の教育センター等に調査票を送付して、教育情報システム等の調査を行い、調査票を回収し、機械可読なデータに変換する、という作業手順で行ってきた。

しかし、今年度より、データ提供者側で機械可読のデータを入力すれば、①データ収集者側の作業が簡素化できる、②データ提供者側に機械可読なデータが残り、情報処理機器での管理が可能になる、③ペーパレスになる、などの効果を期待して、磁気媒体やパソコン通信によるデータ収集を試みた。

本研究は、全国の教育センター等から教育情報の収集を行うに当たり、①多種多様な情報処理機器や文書処理用ソフトウェア（以下、「ソフト」と呼ぶ。）を利用して入力されたファイルの読み込みができるか、②人名や地名など JISコードにない文字をどのように取り扱うか、③今後における教育センター等の情報収集の作業量の軽減を図れるか、をねらいとしている。

なお、磁気媒体やパソコン通信による調査を実施する前に、この方法による調査が可能であるか否かを確認するため、アンケート方式により、全国教育研究所連盟への加盟機関および国立教育研究所で把握している、教育センター等 643 機関を対象として事前調査を行った。その結果、確実にテキストファイルに変換して提供できるのは、306 機関（全調査機関のうち 47.6%）であり、このうち 67 機関は、パソコン通信でも可能であるとの回答があった。

この 306 機関からどのくらいのデータ件数を収集できるかを、平成 3 年度のデータ提供実績に基づいて集計したところ、8,356 件（総レコード数 10,636 件のうち 78.6%）であることが分かった。よって、この磁気媒体やパソコン通信による収集法で、全データ件数の 80% 近くが収集でき、データの提供者側が直接、機械可読のデータを作成して提供することが可能であると判断し、実施することとした。

2. 磁気媒体・パソコン通信による教育情報収集のための方法

データ収集のために利用した磁気媒体は、いろいろな磁気媒体がある中で、調査回答機関が取扱に慣れていること、調査内容を入力するための十分な記

憶容量があること、郵送に便利で、価格が安いことを考慮し、フロッピーディスク（以下、「FD」と呼ぶ。）を利用することとした。

2. 1 収集データの種類

当研究所の教育情報データベースを更新するために、毎年、全国の教育センター等に対し研究所で行っているもので、教育センター研究成果、教育ソフト、教職員研修講座、教育情報システム、汎用コンピュータ設置状況、パーソナルコンピュータ設置状況、保有データベース等に関する情報について調査を行った。

2. 2 FDを送送するための事前の作業

FDを送送するための事前作業は、次の(1)から(5)の順に行った。

(1)FDの初期化とラベルの貼付

FDを初期化して、各機関番号のラベルを貼った。

(2)教育情報調査票および各調査の実施要領の印刷

今年度からFDやパソコン通信を使って、教育情報データの収集を実施したため、従来の調査表や実施要領に若干の修正を加えた。①調査票については、図1のように機械処理を容易にするために項目番号「#01#」「#02#」を加えた。②実施要領には、「フロッピーでご回答いただく場合」と「フロッピーによる情報提供の場合の留意事項」および「フロッピー入力例」をつけ加えた。

図1 パーソナルコンピュータ設置状況調査票
(平成4年8月1日現在)

#01 回答公表の可否	1. 公表可		2. 公表不可	
#02 通 番				
#03 メーカー名				
#04 機種・モデル名				
#05 セット数	セット			
#06 R A M 容量	KB			
#07 ハードディスク容量	MB			
#08 フロッピーディスクドライブ	フロッピーサイズ	台	台	台
	ドライブ数			
	記憶密度			
#09 プリンタ名				
#10 プリンタ	1. dotプリンタ	2. 熱転写プリンタ	3. レーザプリンタ	4. その他
#11 PC用ソフトウェア端末	1. 端末として利用 2. 端末として利用していない			
#12 その他の周辺機器				

(3) サンプルファイルの作成とFDへのコピー

図2のように、図1の調査票の各項目をFDへどのように入力したらよいのか、レイアウトが判るように各FDにサンプルファイルを入力した。

なお、ファイル名はデータの提供者側からデータを収集した後、調査の種類ごとにファイルを結合する際、機械的に処理できるように、ファイル名と拡張子を指定した。また、通番はデータ入力が複数の担当者で行われたり、1つの文書作成容量に制限があるとき、複数のファイルに分ける必要があるの付けた。

例 C 9 9 9 9 0 0 1 . P S N (パソコン設置状況調査)
 機関番号 通番 内容

図2 サンプルファイル

- # 0 1 # 公表可↓
- # 0 2 # 0 0 1 ↓
- # 0 3 # 富士通↓
- # 0 4 # FMR-50HV↓
- # 0 5 # 1 1 ↓
- # 0 6 # 2 0 4 0 ↓
- # 0 7 # 4 0 ↓
- # 0 8 # 5インチ(2)2HD↓
- # 0 9 # MS-DOS↓
- # 1 0 # リーザープリンター↓
- # 1 1 # 端末として利用↓
- # 1 2 # ↓
- # 1 3 # 教育諸調査データ、教育情報の検索サービス、教材研究(CAI/CMI)、教職員研修パソコン通信↓

(↓は改行「CR+LFコード」)

(4) 昨年度提供されたデータの各機関ごとのFDへのコピー

各データ提供者側の入力作業負担を軽減するため、昨年度収集したデータで今年度もあまり大きな変更がなく、かつ提供件数が多いもの(教職員研修講座調査)について、ファイルを当該機関ごとにコピーした。

(5) 調査票・実施要領の配付

FDやパソコン通信での提供可能な機関へは、FDと一緒に同封して送付した。

3. 磁気媒体・パソコン通信による教育情報の収集

3. 1. 収集状況

(1) FDでの収集状況

今年度のデータの収集状況は、データ提供機関の総数が342機関で、レコード総数11,184件であった。このうちFDによる提供は、156(全データ提供機関のうち45.6%)機関で5,978件(総レコード数の53.5%)であった。

読み込みのできないFDが3機関分あったが、2機関はワープロ専用機で入力し、1機関はパソコンのソフトを利用していたが、当研究所と異なる機種やソフトであったので、テキストファイルに変換できなかった。これらの読み込みのできないFDのファイルについては、調査票に記入した回答が同封されていたので、データ入力会社へ外注して入力データを作成した。残りの153機関から提供されたFDは、ハード的に破損したり、中身のファイルが消失することもなく、100%読み込み処理することができた。

(2) パソコン通信による収集状況

パソコン通信による収集は、5機関(全データ提供機関のうち1.5%)から451件(総レコード数の4.0%)のデータを収集することができた。

パソコン通信によって収集したデータは、ファイルに異常がなく100%読み込むことができた。

(3) 入力のためのソフト分類

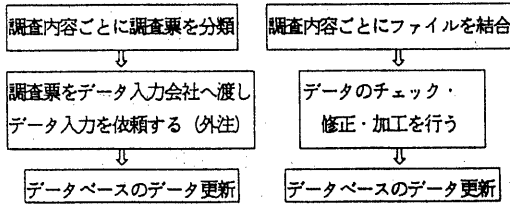
テキストファイルの形式でデータの提供を依頼したので、各機関がどのようなソフトを利用して、データ入力を行ったのか、FDやパソコン通信で提供した161機関の集計は表1のとおりである。ワープロソフトで入力した機関のうち、一太郎を92機関(89.3%)が利用していた。その次に多かったのがテキストエディターで33機関(機械可読化して提供した機関のうち20.5%)が利用していた。また、ワープロ専用機は10機関(機械可読化して提供した機関のうち0.6%)が利用していた。

なお、3.1(1)で述べた3機関を除いて、当研究所でテキストファイルに変換したのが3機関分あったが、残りはすべてテキストファイルに変換されていた。

表1 入カソフト別収集状況

入カソフト名	テキストファイルとして読み込み可能					テキストファイルとして読込不可		合計
	ワープロソフト	テキストエディター	ワープロ専用機	その他不明	パソコン通信	ワープロソフト	ワープロ専用機	
提供機関数	102 (63.4%)	33 (20.5%)	8 (5.0%)	10 (6.2%)	5 (3.1%)	1 (0.6%)	2 (1.2%)	161 (100%)
レコード数	3,923 (59.7%)	1,518 (23.1%)	111 (1.7%)	26 (6.5%)	451 (6.9%)	103 (1.6%)	34 (0.5%)	6,566 (100%)

図3 データ収集後の作業手順
【従来の収集後の作業】 【今年度の収集後の作業】



3. 2 収集後の作業

昨年度までの調査票での収集後の作業手順と、今年度実施した磁気媒体・パソコン通信による収集後の作業手順が、どのように変化したかを比較してみると、図3のように、従来、データを収集者側が行っていた、収集したデータの入力作業を、データの提供者側で行うようになったために、データの収集者側においてデータ入力作業が省略され、そのための時間・費用・労力が軽減され、収集したデータを数枚のFDに保存して管理できるようになった。

しかし、機械可読の状態を提供されたファイルを、そのまま利用できるか否かを確認するために次の作業を行った。①ワープロ文書のままの場合はテキストファイルに変換する。②各調査の種類ごとにファイルを結合する。③レコードの中にデータ提供機関を識別するための項目のないものについては、新たに機関番号を入力する。④一項目中の複数生起のデータの区切り記号を入力・修正する、⑤同じ内容の

レコードが繰り返し入力されている場合は、そのうちの1レコードを残して削除する。⑥空白や改行のための属性記号を入力する。⑦データベースの文字属性に一致させるように入力された文字を全角、又は、半角に修正する。⑧データベースの項目の順番に並べ換える。

4. 今後の課題

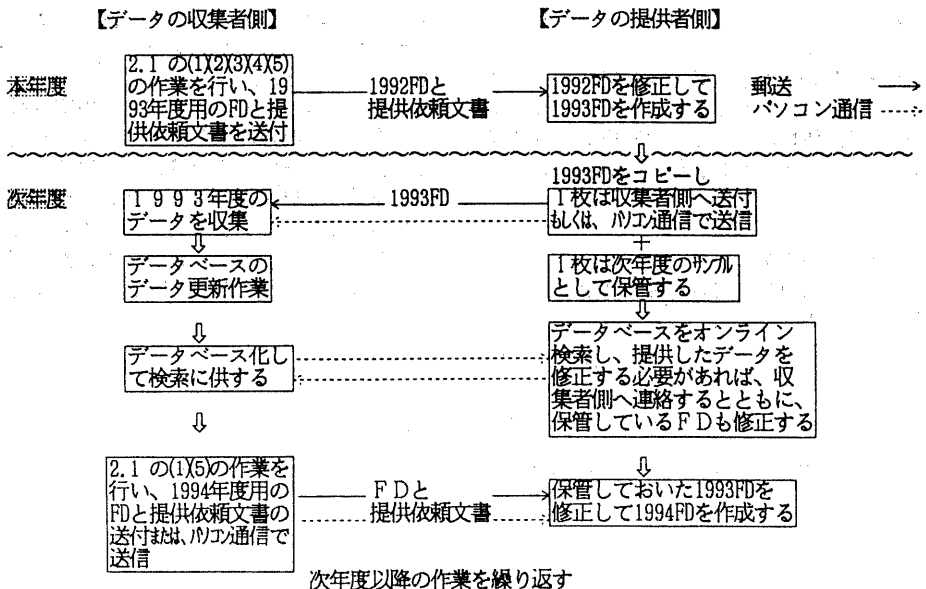
4. 1 事前の作業

FDの購入は、初期化する時間を軽減するためには、費用は少し高くなるが、初期化されたFDを購入することが望ましい。

FDの中のファイル名には、どの機関のファイルなのか分かるように、事前にファイル名が決められていたが、各レコードにも、どの機関のデータなのかを識別する項目を設ける必要がある。

また、2. 2の(2)から(4)で述べた事前の作業を軽減するために、図4のようにデータの提供者側で、データを入力した今年度のFDのコピーを保管し、次年度そのコピーしておいたFDのファイルを修正して、データの収集者側へ送付もしくはパソコン通信で送信すれば、初年度に必要とした事前の作業が軽減できるようになると考える。さらに、パソコン通信による調査回答が可能な機関へは、電子掲示板を利用して教育情報提供依頼文書を送信することで、ペーパレスが可能となり、提供依頼文書の印刷や、郵送に要する費用や労力が軽減され、2. 1の(1)から(5)の作業を軽減できるようになると考えられる。

図4 教育情報提供依頼から収集までの流れ



4. 2 磁気媒体での収集

今後は多種多様な機種あるいはソフトで入力されたファイルを、テキストファイルに変換できるソフトを当研究所が準備するか、もしくはデータ提供者側のワープロやソフトに、テキストファイルに変換する機能がないのであれば、入力したファイルを印刷して送付してもらい、OCRを利用して機械可読な状態に変換することなどが考えられる。

また、FDやパソコン通信で収集できたのは、収集した総レコード件数のうち6,429件(57.5%)であった。アンケート調査の結果をもとに、1.で述べたFDやパソコン通信で80%近くを収集できると予測した数字をかなり下回った原因は、磁気媒体での提供が可能であると回答した機関の中に、調査票で提供した機関が73機関(このうち13機関はFDで1部を提供)レコード件数が2,216件であったことにある。これらの機関に事後調査を行ったところ、主な理由は次のとおりであった。①アンケートに回答した担当者、実際のデータ提供の担当者が違っていたこと、②データ提供を担当する部課が異なるために、機械可読の状態に入力するための機種が違い、一つのファイルにまとめることができなかったこと、③各担当者にFDへ入力してもらう許可を得ていなかったこと、④各担当者が入力するための技能を、必ずしも持っていないかったこと、⑤機器が十分揃っていないかったこと、などである。

これは、各機関で教育情報を収集するための組織的な取り組みや、機械可読な状態にして、それをデータベース化して管理する、といった取り組みが行われていないためであろうと考えられる。

4. 3 パソコン通信での収集

パソコン通信で送信されたファイルの中に、圧縮ファイルにしたものがあつた、このファイルの中に解凍するためのアーカイバが組み込まれており、解凍することができたが、今後は送信するデータ量が多い場合、通信費の軽減を考えて、ファイルを圧縮して送信されることが多くなると考えられるので、圧縮に利用したアーカイバ名を記入する項目を設ける必要があるとともに、この圧縮されたファイルを解凍するアーカイバを準備する必要がある。

パソコン通信で送信されたファイルは、どのような機種やソフトで入力されても、関係なくテキストファイルの状態では収集できるし、ペーパレスにも繋がるので、今後さらにパソコン通信でのデータ提供が望ましいと考えるが、当研究所のIDの取得状況は、パソコン通信によるデータ提供が可能であると回答した機関のうち、IDを取得しているのは40機関で、残りの27機関は取得しておらず、パソコン通信が可能な全機関のID取得が必要である。

4. 4 データの入力

収集したファイルには、空白を使い、見やすく入力されたレコードがかなりあつた。しかし、データベースの検索結果の出力形式は、入力レコードの形式と異なる場合があり、必ずしも入力されたイメージで出力されない。そこで、余計な空白は入れないようにし、英文の単語と単語の間の空白のような必須の空白には、空白にするための属性記号を予め用意し、この属性記号を使用してもらおう、実施要領に明記しておく必要があると考える。

また、人名・地名・旧字体など第1水準・第2水準JIS漢字コードにない文字が外字で入力されているものがある。JIS漢字コードにないため文字化けし、検索語を利用して検索できない原因となる。したがって、人名や地名でその字形に意味を持つのであれば、別途方法を考慮する必要があるが、意味がなければJIS漢字コードにある異体字(新字体)を使用すればよい²⁾と考えられる。

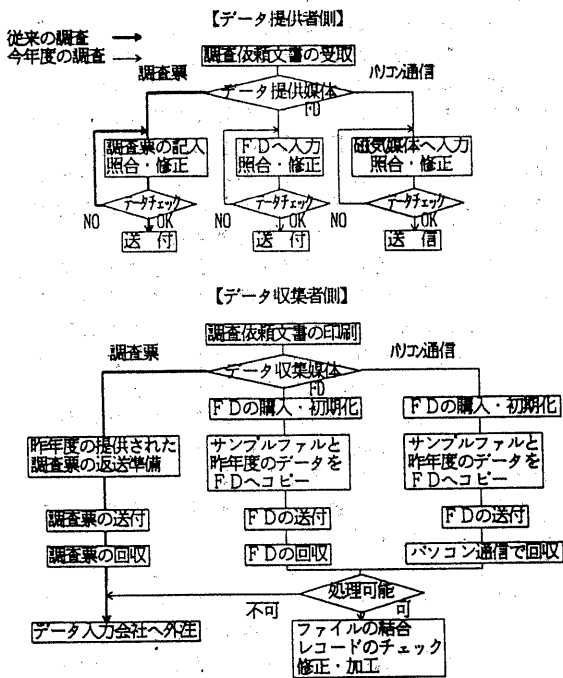
文献の要約等の文書を入力する際に、図や丸付き数字・ローマ数字等を使う場合があるが、この図の罫線や丸付き数字・ローマ数字もメーカー独自のキャラクターコードであり、規格化されたコードでないので、テキストファイルへ変換した場合に文字化けしたりする原因になる。図表などを用いて記述する必要がある場合は、それを調査票に記入して送付してもらい、画像データベース化して画像表示する方法もあると考えられる。

さらに、パソコンや汎用コンピュータの機種名やソフト名などの固有名詞が、通称や俗称を用いて入力されている場合も、データベースを検索する際に、条件が一致しなくて検索されない原因になる。すでに収集されたデータに関しては、当研究所で統一して修正し各データ提供者側が当研究所のデータベースをオンライン検索して確認し、違っている場合は連絡するようにすれば、解決できると考えられる。今後は、修正されたデータを参考にして記入すれば良いと考える。

4. 5 新しい作業の発生

図5のようにデータ収集者側と提供者側別に、従来の調査票によるデータ収集作業と今年度のFDやパソコン通信による収集作業に分けてみると、データの収集者側では、データを機械可読化するための入力作業が軽減され、収集したデータ管理がやりやすいといったメリットが得られた反面、FDの購入・初期化、FDへのサンプルファイルや昨年度のデータのコピー、FDのファイルの結合、レコードのチェック・修正・加工といった新しい作業が生じた。しかし、FD購入の費用は、FDを繰り返し利用することで減少するし、FDへのコピーなどの事前作業も、4.1で述べたように軽減することが可能

図5 今年度のデータの提供者側と収集者側の作業



である。収集したFDのファイルの結合、レコードのチェック・修正・加工の作業についても、データ提供者側が入力作業に慣れ、データの入力の仕方を標準化することによって、機械処理による修正・加工が可能となり、作業を軽減し簡素化することができる。

データ提供者側における、機械可読の入力作業は、調査票に記入する手間と同じであり、調査票を送付する手間と、FDを送付あるいはパソコン通信での送信する手間が同じであると考えれば、新たな作業が発生したとは考えられない。一方、そのデータでデータベースを作り管理すれば、検索資料として活用できる。このように今回の調査方法がもたらすメリットは大きい。

5. おわりに

FDやパソコン通信による収集法には、4. で述べたような課題があるが、この方法によると、データ収集者側には、データを磁気可読化する作業の軽減や、収集したデータの管理がしやすくなるなどのメリットがあった。

データ提供者側でも、これまで行ってきた調査票への記入の手間とデータ入力の手間が同じだと考えれば、機械可読なデータが残り、そのデータでデータベースを作って管理すれば検索資料として活用で

き、ペーパレスにも繋がるといったメリットがあり、データの入力作業が、データ収集者側から提供者側へ移ったことによって、トータルの作業量が減少したと考えられる。

また、多種多様な情報処理機器やソフトを利用して入力されたデータの読み込みは、FDやパソコン通信で収集した101機関8,566件のうち、3機関(1.9%)137件(2.1%)が不可能であった。今後は、①どのような機種やソフトで入力されたファイルでも、テキストファイルに変換できるソフトを準備する、②データ提供者側の機種やソフトにテキストファイルに変換する機能がないのであれば、データ入力形式で印刷したものを送付してもらい、OCRで読み取ってテキストファイルに変換する、③パソコン通信で送信することにより、テキストファイルとして読み込む、ことで磁気媒体やパソコン通信によるデータ収集が可能であり、実用化できると確認した。

今回の調査で、調査内容と無関係のファイルが入力されたFDがあった。これは磁気媒体全体にいえることだが、ファイルの中身が直接目で確認出来ないことや、印刷された内容とFDの中身との照合が不十分であったことによるものであると考えられる。今後は、データ提供者側において、機械可読のデータをチェックするための内部承認手続きを、明確にすることが必要である。

また、はがきアンケートでFDによる提供が可能と回答した担当者と、実際に磁気媒体へ入力して提供する担当者が違っていたため、相互理解をもってデータ提供されなかった、といったこともあった。今後は、情報提供の内部組織の整備が必要であり、また、FDなどの磁気媒体を公文書として取り扱うことの可能性について検討する必要がある。

今回、磁気媒体やパソコン通信での提供を行わなかった機関の中には、すでに教育情報データベースを作成している機関や、当研究所のオンラインのIDを取得している機関がある。これらの機関については、今後、機械可読のデータを提供することにもなう長所・短所を吟味し、データ収集者と提供者の両者が総合して、効率的に機械可読データが作成できるように検討する必要がある。

【参考文献】

- 1) 吉岡亮衛：「国立教育研究所の教育情報研究と地域教育情報システムの現状」, 科教研報, Vol. 7, No. 3, pp67-71, 1992
- 2) 星野 聡：「日本史データベース」情報処理学会, Vol33, No. 10, pp1109-1115, 1992