

ICPSR におけるデータの寄託から公開まで

前田幸男（東京大学社会科学研究所研究支援推進員）

SSJDA - 19

February 2002

本稿は筆者がデータベース室勤務中に従事した研究の成果である。このような機会を与えてくださった仁田道夫先生と佐藤博樹先生、そして論文構成上大変有益なコメントを草稿段階で寄せて頂いた清水誠先生に御礼申し上げる。草稿を読んで表現上のアドバイスを下さった今井亮祐氏（東京大学大学院）にも感謝したい。筆者がデータの保存・公開について興味を持ったのは、ミシガン大学大学院在学中に日本の選挙調査データを整理して以来である。その意味では、その機会を与えて下さったジョン・キャンベル教授、並びに様々なアドバイスを頂いた数名の ICPSR スタッフ・メンバーの皆さんにも御礼申し上げたい。

目 次

0 . はじめに	1
1 . ICPSR の組織	2
2 . データの寄託	3
3 . 被調査者のプライバシー保護	4
4 . アーカイヴィストによるデータの検査・整理	5
5 . データ利用に必要な文書（メタデータ）の整備	8
6 . データの公開 - ICPSR の側から	10
7 . ICPSR における分業	11
8 . データの提供 - 加盟大学の OR と利用者	12
9 . データの利用 - 教育の現場から	15
10 . プライバシー保護とデータ・ファイルの無断複製防止	15
11 . 結語	17
参考文献	21
有益なホームページへのリンク集	23
Appendix 1 . データ・アーカイヴィストの募集要項	24
Appendix 2 . 文書部門職員の募集要項	25
Appendix 3 . 文書部門管理職の募集要項	26
Appendix 4 . データ利用申請の電子メール	27
Appendix 5 . データ申請者に対して CAP から送られる電子メール	28

0. はじめに

アメリカ合衆国の社会科学研究においては調査データの統計的分析が大きな役割を占めている。それは、社会調査データが再分析のために共有されていること、並びに大学院で体系的な統計利用の教育が行われていることの二つによって支えられている。後者の教育に関しては、個々の教育機関でもある程度対応可能であろうが、前者の社会調査データの共有に関しては、何らかの組織的基盤が無ければ極めて困難である。アメリカでは、社会調査データを共有することを目的に ICPSR(Inter-University Consortium for Political and Social Research) が 1962 年に設立されている¹。

ICPSR はミシガン大学の政治学者 Warren E. Miller のリーダーシップのもと 1962 年に設立された(以下の記述は主に Converse, Kish, and Shanks 1999 による)。ミシガンの選挙研究プログラムに所属し調査を実施した研究者たちは、1950 年代から彼らが収集したデータを他の機関に所属する研究者たちと共有するべく努力してきた。当初は学外の研究者が求めたクロス表をパンチカードとソーターを使ってそのつど作成していたそうであるが、その作業に必要な職員の時間を提供するという意味において、ミシガンの選挙研究プログラムが実質的に費用を負担していた。しかし、そのような依頼が増加するにつれ、本体の選挙研究プログラムの作業の遅延が問題となった。そこで、調査の実施とデータの保存・公開を分離し、かつデータ保存用の編集およびクリーニングや公開にかかる費用を公平に負担するために、ICPSR の設立が企画されたそうである。計画の当初は、加盟各大学からミシガン大学へ料金を払う仕組みが受け入れられるかどうか疑問視する向きも多かった。各大学からの承認を得るにあたっては Miller の個人的かつ精力的努力に負うところが大きかったようである。

ICPSR は現在約 400 の加盟大学と 600 万ドルの財政規模(\$1 を 124 円で換算して約 7 億 5 千万円)を誇る巨大組織であるが、ICPSR が米国の社会科学研究に占める役割は極めて大きい。本稿では、ICPSR が受け入れた調査データを処理する過程を紹介することを通じて、社会調査データを公開する際の様々な問題、並びに社会調査データの二次分析に必要な研究・教育環境の整備について考察したい。

¹ 設立時の名称は ICPR (Inter-University Consortium for Political Research) であった。

1. ICPSR の組織

ICPSR の具体的な活動を紹介する前に、その組織構成を概観しておくことが有用であろう。ICPSR は組織全体を統括する部門（Administration）と具体的な業務を遂行する四つの下位部門から構成される。部門の構成と職員数は以下の表に掲げる通りである。

表 1-1 ICPSR 職員数

活動領域	職員数	備考
組織運営	9	3名兼任 内1名は他大学の教授
データ・アーカイブ	51	2名兼任 とともに他大学の教授
テクニカル・サポート	11	1名兼任
教育活動	6	3名兼任
コンピューターとネットワーク	8	2名兼任
総数	82	常勤職員数は50人前後

出典：Annual Report とミシガン大学ネットワーク上の職員録検索から著者が作成。職員録から雇用形態が判然としない場合もあったので、大まかな数字として理解していただきたい。

2000年度決算における人件費は総額4,112,200米ドル（\$1を124円で換算して約5億円）であり、その半分がデータ処理とアーカイブの運営に必要な職員の人件費である。運営費の総額は1,051,500米ドル（約1億3千万円）で、やはりその半分がデータ処理とアーカイブの運営に充てられている。詳しくは表1-2と表1-3をそれぞれ参照して頂きたい。なお、組織的には五部門構成であるが、財政上は機能別に六部門が立てられている。

表 1-2 ICPSR 人件費 - 2000年度決算

	ドル	円換算（\$1 = ¥124）
組織運営	588,200	72,936,800
データ処理・企画	1,911,700	237,050,800
アーカイブの運営	170,100	21,092,400
コンピューター部門	461,900	57,275,600
教育部門	646,200	80,128,800
利用者サポート	334,100	41,428,400
総計	4,112,200	509,912,800

出典：Annual Report p.32.

表 1-3 ICPSR の運営費 - 2000 年度決算

	ドル	円換算 (\$1 = ¥124)
組織運営	558,500	69,254,000
データ処理・企画	469,800	58,255,200
アーカイブの運営	14,800	1,835,200
コンピューター部門	371,400	46,053,600
教育部門	210,500	26,102,000
利用者サポート	169,300	20,993,200
総計	1,051,500	130,386,000

出典：Annual Report p.32. 六部門の運営費を合算しても総計と一致しないが、これは Annual Report においても一致しない。

2. データの寄託

ICPSR で公開している調査の大半が確率抽出された個人に対して行われたインタビューに基づく社会調査データである。他の種類の研究データとしては、利益集団・組織・企業を分析単位とするデータ、国家を分析単位とするデータ、国家間の紛争が分析単位となっているイベント・データ等が所蔵、公開されている。以下の記述は基本的に社会調査データの寄託・保存・公開を念頭に置く。

ICPSR に対して調査データを寄託するのは、大学に所属する研究者あるいは新聞社・世論調査会社等である。後者は私企業であるが故に、常に調査データを寄託する訳ではないが、大学に所属する研究者に関しては、一定の期間を経過した後には必ず調査データを公開するという規範が米国では確立しているといつてよい。事実アメリカ統計学会はその倫理規定の中で、データを共有することの重要性を明確にしている (American Statistical Association 1999, II-C-13 and II-F-1)。より具体的には、国立司法研究所 (司法省) の研究費を支給された場合は研究期間満了と同時にデータをアーカイブに提出する義務があり (National Institute of Justice, 2000)、保健・福祉省の研究費の場合は、研究満了 2 年以内にデータを提出しなければならない (National Data Archive on Child Abuse and Neglect, 1999)。全米科学協会 (National Science Foundation) の研究補助金を得て実施された研究のデータは研究期間終了 1 年以内に ICPSR 等のアーカイブに寄託する義務がある (National Science Foundation, n.d.)。

データ寄託者は ICPSR にデータを寄託する前に以下の事項を明確にし、かつ署名しなければならない²。

- 1) 寄託者は ICPSR がデータを標準的な利用条件のもとで（研究・教育のために）公開することに同意する。
- 2) 寄託者はデータに対する著作権を保持し、かつ ICPSR を通じて一般公開する権利を有する。
- 3) 寄託者は調査対象者を直接特定できる全ての情報を除去した。さらに調査対象者を間接的に特定する可能性がある情報を除去するために最大限の努力をした。
- 4) 寄託者はデータが公開されたことから、調査対象者の特定、彼らのプライバシーに関する情報の漏洩、またはプライバシーそのものの侵害が生じて、訴訟が発生した場合、ICPSR（組織並びに職員）及びミシガン大学を全ての法的責任から免除する。

寄託者が以上の事項に合意して初めて ICPSR はデータを受け入れ、アーカイヴィストの作業が始まる。

3. 被調査者のプライバシー保護

調査データを二次分析用に公開するためには、被調査者個人の割り出しを防ぐために細心の注意が必要であることは前項からも明らかであろう。ここでは、Dunn and Austin（Fall 1999）と ICPSR が発行している *Guide to Social Science Data Preparation and Archiving*（2000, pp.16-17）に依拠して、プライバシー保護とデータ公開との関係について少し詳しく見ておきたい。

被調査者を直接特定することができる情報には、氏名、住所（郵便番号を含む）、電話番号（内線番号を含む）、社会保険番号、免許証番号等がある。従来これらの情報は、ICPSR のアーカイヴィストによりデータ・ファイルから除去されてきた。しかし、近年は個人を直接特定できる情報を含むデータ・ファイルは ICPSR により処理されずに、一旦調査を行った研究者に対して返還されている。これは、個人を直接特定できるデータを ICPSR が（公開していなくとも）所蔵することから生ずる法的問題を回避するためである。実際、被調査者個人を特定できるデータの開示を求める訴訟が ICPSR に対して起こされる可能性が存在する。

² *Guide to Social Science Data Preparation and Archiving*(ICPSR 2000)に収録された ICPSR Data Deposit Form を参照。本文に掲載した 4 項目中、(3) と (4) は Data Deposit Form では一つの項目としてまとめられているが、内容を理解しやすくするためここでは別の項目を立てた。

被調査者の割り出しは、上記のような直接的情報がなくとも複数の間接的な情報を援用することからも可能になる。そのような間接的な情報としては、居住行政区画・選挙区、所属組織、卒業した学校の名前、勤務先、勤務先の地位等多数ある。このような情報は調査目的と密接に関連していることが多く、直接的な情報と比べて処理が難しい。具体的には除去、他の変数カテゴリーとの統合（例えば年収 15 万ドル以上を 1 つ下のカテゴリーと合併）あるいは要約変数の作成（例えば州を南部・東部等に変更）等の方法により処理されるのが通例である。

ICPSR は調査担当研究者に対して、以上のような変数を含むデータを公開することが被調査者のプライバシー保護に与える危険性について十分考慮し、必要な処置を施した後で、データを寄託することを要請している。その際、ICPSR のアーカイヴィストは、調査担当研究者が二次分析用のデータ・ファイルを作成するにあたって必要な技術的アドバイスを与える。

データが寄託された段階で ICPSR はプライバシー保護に対する配慮が十分になされているかどうかを再度吟味する。この段階で、まだプライバシー保護が十分でないと判断された場合、ICPSR と調査担当研究者は残っている問題を解決するために協議することになる。

4. アーカイヴィストによるデータの検査・整理

プライバシー保護について十分な処理が施されたデータ・ファイルおよび様々な文書が調査担当研究者から ICPSR に寄託された後に、アーカイヴィストによる検査と編集が行われる。そこではデータ・セットの利便性を向上させるクリーニングや再コーディングが行われると同時に、二次分析に必要なコードブックや調査概要等のメタデータが整備される。この項では主に Zelenock and Marz (1997) に依拠しながら、具体的にどのような検査と編集が行われるかを概観したい³。

寄託データが、公開までに受ける主な検査および編集は以下の通りである（順序不同）。

1) データの検査。具体的には以下の項目を検討する。

- イ) 寄託された調査データがどの程度クリーニングされているか
- ロ) 調査データが一般公開に適したものか
- ハ) データがどの程度使いやすいものか

³ National Data Archive on Child Abuse and Neglect (NDACAN) が発行する手引きは、調査を担当した研究者がアーカイヴに寄託するデータを作成する際に注意すべき事項を列挙しており、有益である (NDACAN 1999)。

2) 保存用ファイルの作成。調査者から寄託された原データの保存用ファイルを作り、ICPSR 本部から離れたデータ保管場所で保存する。後には ICPSR アーカイヴィストの手によって処理された公開用データも同様に保存される。

3) 被調査者のプライバシーが漏洩する危険性がないかどうかを検討する。社会保険番号、電話番号等の直接的な情報だけでなく、居住地、職業等の間接的な情報の組み合わせから生ずる危険性も検討される(「3. 被調査者のプライバシー保護」を参照されたい)。

4) 調査概要、コードブック等の文書が、どの程度詳細かつ正確かを検討する。質問表も機械可読ファイルへと変換され、公開される文書の一部となる。

5) 文書の記述と実際のデータが一致するかどうかを検討する。具体的には以下を参照。

イ) データ・ファイルに含まれる調査対象者数

ロ) 変数の数

ハ) データのコラム数 (Record Length)

ニ) ファイルの構造 (flat file or hierarchical file)⁴

ホ) 複数のファイルが1つのデータを構成する場合、それらのファイルを統合する方法

6) ウェブ上の検索に利用される文書を作成する。アーカイヴィストは提出された文書を利用しつつ、調査のデザイン、標本抽出法、調査の目的、データ分析上の注意、調査を使用した報告書・学術論文の文献目録等を含む調査概要を準備する。

7) 上記の項目を検討した結果、何らかの問題があることが分かった場合、調査を行った研究者、あるいは調査を実際に担当した調査機関と協議し、必要な措置をとる。

どの程度の時間と資金が特定のデータ・ファイルに注がれるかは一様ではない。近年、寄託されるデータ・ファイル数の増加に伴い、プライバシーの保護以外はアーカイヴィストによる検査を受けないまま一旦公開される場合も増えている (Archival Development Committee,

⁴ フラット・ファイルには2つの種類がある。LRECL and Card Image の場合は、一つの分析単位に関する情報が同数のカード・イメージ(コンピューター上は80桁1行が一つのカード)に保存されている。これはデータがパンチ・カードに保存されていた時代の名残である。単純な LRECL の場合は一つの分析対象につき1行が使われている。この場合桁数は変数の数に依存する。なお LRECL は Logical Record Length の略である。flat file は rectangular file とも言われる。調査が異なる分析単位、たとえば個人と世帯の両方を含む場合、二通りのデータ保存法がある。一つのファイルの中に、異なる分析単位の情報を違うフォーマットで保存する方法が一つ。もう一つは、調査データが複数のフラット・ファイルで構成されている場合である。これらの形式で保存されたファイルを階層ファイル hierarchical file と言う (ICPSR, November 2000、Glossary を参照)。

1997 を参照)。しかし、時間的あるいは資金的に可能な場合は更に以下の処理がデータに対して施される (順序不同)。

8) ハードコピーの文書を機械可読ファイルに変換する。コードブックはテキスト・ファイルとして提供されることも多い。詳細は「5 . データ利用に必要な文書 (メタデータ) の整備」を参照されたい。

9) SAS と SPSS のデータ定義ファイル (Data Definition Statement) を作成する。ICPSR が受け取ったデータの状態により、簡単な作業の場合もあれば、かなりの時間を要する作業となることもある。データ定義ファイルは、各変数の素データ上のコラム位置と作成されるシステム・ファイル上のフォーマット、変数ラベルと変数値ラベル、欠損値等を指定する。データ定義ファイルを編集することにより、利用者は素データから必要な変数だけを抜き出す、あるいは欠損値の指定を変更することができる。また、SAS や SPSS 以外の統計パッケージへのデータの移植もデータ定義ファイルを編集することで容易となる。なお ICPSR には、SAS と SPSS のシステム・ファイルから素データとデータ定義ファイルを作り出すプログラムが存在する。

10) 調査の内部で、変数のコーディングが統一されていない場合、統一的なコーディングに変更する。

11) 欠損値のコーディングを統一・標準化する⁵。欠損値がシステム欠損値 (コンピューター上はピリオド ". " やブランク " ") の場合、それが妥当かを検討する。具体的には、システム欠損値に「該当せず」(男性だけに聞かれた質問の場合女性はずべて「該当せず」になる) だけでなく、「回答無し」や「わからない」が混在する場合、両者を極力分離し、可能な限り「回答無し」や「わからない」に実際の数値を与える再コーディングを施す。

⁵ 例えば「わからない (DK)」に 8 または 88、「答えなし・回答拒否 (NA)」に 9 または 99、「対象外・非該当 (INAP)」に 0 が割り振られることが多い。パネル調査の場合は、特定の回で調査に漏れた対象者の全変数がシステム欠損値に、面接を受けた対象者で特定の質問に非該当の場合は 0 が割り振られることになる。ここで割り振られている数値は例であり、他の数値を用いる、あるいは、システム欠損値がコーディングの一部 (対象外・非該当等) に用いられる調査もある。あくまで「面接不能」、「非該当」、「わからない」、「回答無し」等、有効回答と扱われない場合が明瞭に分離されていることが重要なのである。但し、多くの統計プログラムはシステム欠損値に対して変数値ラベルを貼ることができないので、システム欠損値をコーディングに用いるのは避けるべきであろう (National Data Archive on Child Abuse and Neglect 1999, p9)。

12) ワイルド・コード⁶や矛盾した回答の有無、質問のフィルターが正しくデータに反映されているか等を精査する。ワイルド・コードについては単純な入力ミスを見つけるだけでなく、論理的に不可能なコード（たとえば20歳の年金生活者や、親より年上の子供等）が存在しないことの確認も行われる。

13) データと文書とを照合し、全ての変数の頻度、ワイルド・コード、欠損値コードが正しいかを確認する。

14) 保存とネットワーク上の移転を効率化するために、データの保存形式を変更し、ファイルのサイズを小さくする。例えばSPSSのデータ・エディターに1-0変数を直接入力すると、本来ならば1桁のスペースで済むはずだが、初期設定を使うと8桁分のスペースを占める1-0変数を作成してしまう。多くのデータ・ファイルは変数が初期設定で作成されたままになっているので、必要以上のファイル容量を費やしていることがある。

15) 以上の手続きで問題があることがわかった場合、調査を担当した研究者、あるいは調査実施機関と協議する。

なお、1996年だけで171の調査データ（約4000のデータ・ファイル）がICPSRに対して寄託された。

5. データ利用に必要な文書（メタデータ）の整備

前項の記述はデータ・ファイルの編集が中心であった。しかし、調査概要やコードブック等のメタデータが無ければ、調査データの二次分析を行うことは不可能である⁷（Green, Dionne, and Dennis, 1999）。近年のコンピューター技術の発展はメタデータのフォーマット並びに提供方法に大きな影響を与え、様々な改良が現在進行中であるが（Rasmussen 1995; Carrozza and Howe, 1998）。この項では主にAllen, Bitner, and Randle（1999）に依拠して、ICPSRが現在の方法で機械可読ファイルを作成するようになった経緯と、実際にコードブック等のメタデータが如何に処理あるいは編集されているかについて簡単に紹介したい。

⁶ ワイルド・コードとは端的に言えば、特定の変数にとって存在してはならない数字である。例えば、ある変数に対して、「1. はい、2. いいえ、8. わからない、9. 答えなし」というコードが与えられている場合3, 4, 5, 6, そして、7等の数字はワイルド・コードとなる。ワイルド・コードの多くはデータ入力時のミスにより生ずる。

⁷ データ寄託時に提出された文書の質があまりにも低い場合、調査内容を把握できない場合、ICPSRはデータの受け入れを拒否することがある（Archival Development Committee, 1997を参照）。

多くの調査データに関する文書（標本設計、調査の概要・目的、コードブック等）がワード・プロセッサで作成されているにも拘わらず、調査機関・研究者は ICPSR にハード・コピー文書を提出することが多い。1990 年代後半の段階ですら過半数の文書がハード・コピーで到着している。このような場合、ICPSR の職員は機械可読ファイルを手に入れるよう最善の努力をするが、大抵の場合この努力は失敗に終わるようである。

その一方、アメリカ合衆国においては、コードブックやデータ形式に関する文書は無料あるいは実費のみで提供されるべきだという規範が確立している(Green, Dionne, and Dennis, 1999, p.3)。さらに、インターネット（あるいは File Transfer Protocol）の発達により、機械可読ファイルに対する需要は増加する一方である。このような状況に対処するため ICPSR は、スキャナーを使って印刷された文書を機械可読ファイルに変換する方法を模索してきた。

ICPSR では 1980 年代後半に、OCR (Optical Character Recognition) を使った機械可読ファイルの作成に着手した。しかし、OCR プログラムの多くが数字の「1」(いち)とアルファベット小文字の「l」(エル)とを区別できない等、技術的問題があり、スキャナーで文書を読み込んでから最終稿を確定させるまでに相当のチェックと編集が必要であった。さらに、OCR プログラムに依存すると、文書内の表やグラフ、あるいは質問表の形式を保存できないという問題がある。そこで代替技術として bit-mapped image file を利用することが考慮された。しかし、この形式のファイルを保存、公開および利用することはそれ独自の技術的問題が存在した。例えば、ファイルのサイズは大きくなり、またテキスト検索は不可能になる。更に、かつては bit-mapped image file を直接印刷できるプリンターが存在しなかった。

ICPSR は以上の経験から bit-map の技術と OCR の技術の双方を利用したプログラムの利用を検討し、Adobe 社が開発した PDF (Portable Document Format) を機械可読ファイル作成時の標準フォーマットとして 1995 年に採用した。PDF の "Image + Hidden Text" フォーマットは、bit-map image により文書画像をそのまま保存するだけでなく、その背後にテキストを維持することができる。したがって、Adobe Acrobat により作成された PDF ファイルは文字検索が可能であり、かつ、ハードコピー文書の体裁をそのまま保存することが出来るのである (Green, Dionne, and Dennis 1999, p.20 も参照)。

以下では、実際の機械可読ファイル作成過程を簡単に紹介する。

1) データ処理を担当しているアーカイヴィストが EDC (Electronic Document Conversion) 部門に文書 (調査の概要、コードブック、頻度表等) を提出する

2) EDC スタッフによる文書の吟味、検査

3 a) 機械可読ファイルの場合

- イ) 提出されたファイルのフォーマットに適したプログラムの選定⁸
- ロ) 編集作業
- ハ) 文書のチェック
- ニ) PDF ファイルへの変換

3 b) ハードコピー文書の場合

- イ) 文書の状態に適したスキャナーとプログラムの選定
- ロ) イメージ・ファイル編集可能なファイルへと変換
- ハ) 編集作業 (読み取り過程で生じた誤りや原文書に存在した手書きのメモ等を消去)
- ニ) 文書のチェック
- ホ) PDF ファイルへの変換

Green, Dionne, and Dennis (1999, pp. 19-22) は印刷文書を PDF ファイルに変換する過程を詳細に記述しているが、読み込みから編集までの過程では Adobe 社から発売されている Adobe Acrobat Capture が使用されている。彼女たちによると Adobe Acrobat Capture は、画像を TIFF (Tagged Image File Format) ファイル、文書をテキストとして別々に保存し、後者はエディターによる編集が可能である。保存用の文書は、編集処理が全て終わった後 "Image + Text Format" の PDF ファイルを Acrobat Capture により出力して作成される。Adobe 社の日本語ホームページを閲覧したが、Acrobat Capture は英語版のみしか販売されていないようである。

4) PDF ファイルの編集

- イ) しおり、サムネイル、文書間のリンク、あるいは WWW へのリンクを付加する
- ロ) ICPSR で準備した表紙と冒頭の解説を挿入する

1997-1998 会計年度に EDC が PDF ファイルに変換した文書ページ数は 122,000。1 日あたり 470 ページの文書を処理したことになる。

6. データの公開 - ICPSR の側から

検査と編集が終わったデータ・ファイル並びにメタデータ・ファイル (文書ファイル) はアーカイヴィストからテクニカル・サービス部門の職員の手に渡り、バックアップ・ファイルが作成されると同時に各ファイルに関する情報がデータベースに登録される。公開用の

⁸ 一つの調査において複数のファイルが異なるフォーマットで存在することは稀ではない。

ファイルは ICPSR のサーバーにアップロードされると同時に、年四回発行される新規データ頒布用の CD-ROM に収録され加盟大学に送付される。

なお、ICPSR 加盟大学のみ頒布されるデータのほかに、分野別特別研究 (Special Topic Archives) のデータが無料で一般公開されている。公開のための費用は、外部の基金・研究費により負担されており、現在は以下の五分野のデータが公開されている。

Health and Medical Care Archive (HMCA)

International Archive of Education Data (IAED)

National Archive of Computerized Data on Aging (NACDA)

National Archive of Criminal Justice Data (NACJD)

Substance Abuse & Mental Health Data Archive (SAMHDA)

特別研究ごとにホームページが開設されているが、これらの研究分野に関するデータの検索、コードブックの閲覧、初歩的な分析 (頻度の表示、クロス表の作成、相関係数の計算、重回帰分析等) がウェブの上で実行できる⁹。

7. ICPSR における分業

以上、データの寄託から公開に至る過程を簡単に紹介したが、ここでは各項では触れることができなかった ICPSR 職員の専門化と分業について述べておきたい。

1) データ・アーカイヴィストの専門化

前掲の決算書からも分かるように、ICPSR の資源 (人員・時間) の大半はアーカイヴ部門に投入されている。膨大な量のデータが処理されているわけだが、それに付随して、データ・アーカイヴィストにもある程度の専門化が進行しているように思われる。何故なら特別研究のデータあるいはセンサス等のデータは領域毎の専門知識がなければ適切に処理することが難しいからである。実際、あるアーカイヴィストは ICPSR の夏期講習 (2001 年度) で「センサス・データの利用法」という短期集中講座を担当している (Granda, 2001)。これと関連するが、ICPSR 内での分業は、クリーニング、コーディング、文書作成等の機能別ではなく、刑事犯罪、公衆衛生、社会科学等といった具合に、調査が行われた分野別になっているように思われる。データを公開可能にするためには、調査あるいはデータについてある程度の理解を持つ必要がある。しかし、機能別の分業では特定分野のデータに関する知識が集積されない。おそらく、アーカイヴィストは、分野毎に公開可能なデータを作る作業に最初から最後まで一貫して携わり、機械可読ファイルへの変換や、保存用ファイ

⁹ ICPSR のホームページ上にこれらのウェブサイトへのリンクがある。

ルの作成等で必要になったときだけ、文書部門やコンピューター関係の職員から技術的な支援を受けているのではないと思われる¹⁰。

データ・アーカイヴィストを探す ICPSR の求人広告がミシガン大学のウェブ上で閲覧できたので、参考までにそれを Appendix 1 に掲載してある。

2) 文書部門の独立

近年のインターネットの発達により機械可読ファイルの重要性が増してきているのは先述の通りであるが、ICPSR 内部には機械可読ファイルの作成に特化した Electric Document Conversion (EDC) 部門が、アーカイブ部門の中に存在する。1999 年の時点で EDC 部門には常勤職員 5 名、非常勤職員 2 名が在籍していた (Allen, Bitner, and Randle, 1999)。

データそのものを担当するデータ・アーカイヴィストと EDC 所属の文書担当者との分業の境界ははっきりとしない。ただし、EDC 職員の一部は文書の編集を一切担当せず、文書の読み込みだけに専念しているという記述が前掲論文にあることを考えると、調査データの中身に関する記述作業はデータ担当のアーカイヴィストが行い、体裁の標準化や PDF に対するサムネイルの付加等の編集を EDC の職員が行っているように思われる¹¹。

EDC 部門の求人広告をミシガン大学のウェブの上で閲覧できたので、参考までにそれを Appendix 2 と Appendix 3 に掲載してある。前者は実際に作業に従事する職員、後者は部門統括者を求める広告である。

8. データの提供 - 加盟大学の OR と利用者

ICPSR 所蔵のデータを利用するためには、基本的に大学単位での加盟が必要だが、それは ICPSR が所属するミシガン大学も例外ではない。ICPSR はその名の通り、Inter-University Consortium であり特定の大学 (ミシガン大学を含む) を優遇することはできない。したがって、ミシガン大学に所属する研究者と学生が ICPSR 所蔵データを利用できるように、ミシガン大学は ICPSR に対して年会費を納めている。

加盟大学は OR (Official Representative) を指名し、大学は OR を通じて ICPSR からの連絡およびデータの提供をうける。末端の利用者が直接データをダウンロードする方式が導入される以前は、ICPSR 加盟大学の教員並びに学生は所属大学の OR を通じてデータを入手していた。正式の OR は大学の教員あるいは図書館・情報教育部門の職員であることが多いが (ICPSR 1997)、実際の業務を誰が請け負っているかは大学によって異なるように思われる。ミシガン大学においては、正式 (あるいは名義上の) OR は政治学部の教員が勤め、実際の業務は政治学部のコンピューター・ラボラトリー (CAP: Computer Assistance Program) に雇用された大学院生が行っていた。

¹⁰ この点は本稿の改訂のために直接インタビューを行って確認したい。

¹¹ この点は本稿の改訂のために直接インタビューを行って確認したい。

筆者は幸い 1999 年 9 月から 2000 年 12 月まで CAP に勤務したので、本稿では主に筆者の経験からミシガン大学における ICPSR データの提供について紹介したい。ただし、ミシガン大学で CAP が ICPSR 関連の業務を行ったのは 2000 年 2 月までであり、その後は ICPSR のウェブサイトから末端の利用者が直接ダウンロードできるようになった¹²。直接ダウンロードによるデータの提供は全米の大学に広がっており、ここで記述される方式によるデータの提供は急速に過去のものとなりつつある。

1) 利用手続きの流れ

イ) 利用者が ICPSR のウェブサイトを検索を行い、必要なデータ・セットの番号を特定する。

ロ) データ・セットの番号を電子メールで cap@umich.edu に送信する(実例を Appendix 4 に掲載)。

ハ) 電子メールを受け取った CAP のスタッフが ICPSR のウェブサイトアクセスしてデータ・セット及び必要なファイルを手元にあるコンピューターのハードディスクにダウンロードする。他のファイルは、素データを SAS や SPSS に読み込むのに必要なデータ定義ファイル (Data Definition Statement) であることが殆どであった。コードブックは ICPSR のウェブサイトから特定の手続きを踏むことなく誰でも無料でダウンロードできたので、利用者本人に直接入手してもらう。

ニ) ダウンロードされたファイルをサーバーにアップロードする。

ホ) 利用者に電子メール (実物は Appendix 5 に掲載) を送付する。

ヘ) 利用者がデータをダウンロードする。データのダウンロードが可能な期間は 2 週間で、それを超えた場合はもう一度同じ手続きを繰り返す必要がある。データがアップロードされたネットワーク上のディレクトリにはミシガン大学のアカウントとパスワードがなければアクセスできない。

2) 申請・利用の記録

CAP においては個別のデータ利用申請が処理された後、それを記録していた。記録事項は、申請者名、電子メールのアドレス、所属部局・学部のみで、大学に所属する限り特に身分を聞いた記憶はない。ただし、申請は電子メールを通じて行われたので、電子メールの最後の署名から所属、職位が判明することがほとんどであった。

私自身は、この記録を整理・利用した経験はなかったが、CAP を統括した責任者によると利用記録は大学本部に対して、ICPSR に加盟する (し続ける) ことの重要性を示すために利用されるとのことであった。

¹² 厳密に言うと、Harvard-MIT Data Center のウェブサイトからのダウンロード方式が半年ほど試行された後で、ICPSR のウェブサイトからの直接ダウンロードに切り替えられた。

短期滞在者（客員研究員・夏期統計講習の学生）の取り扱いに関してだが、大学に正式に所属している限り、滞在期限を問うことなくデータは渡された。ミシガンで毎夏に開催される ICPSR や SRC の講習に参加している短期滞在者も授業料を大学に払って参加している限り、大学が供与するサービスを受けられるのは当然と考えられていたように思われる。

3) データを統計プログラムへ読み込む

ICPSR から入手されたデータを利用するためには統計プログラムに読み込む必要があるが、この段階でつまづく利用者が少なからずいた。利用者に対する支援も OR の役割の一部であるが、もっとも頻繁に受けた質問は、素データを統計パッケージに読み込む方法と、SPSS あるいは SAS フォーマットのデータを他のパッケージに移植する方法の二つであった。他のパッケージは主に Excel, Stata や Limdep 等であった。

統計パッケージを学習する際に素データの読み込みを学習することは少ないようであり、初めてデータ利用申請をした学生の多くがデータ定義ファイルの利用法について質問をした。トラブルのほとんどのケースは SPSS のデータ定義ファイルの先頭にあるシンタックスで素データがハードディスクのどこに存在するかを特定することを忘れたか、あるいは SPSS シンタックス末尾の EXECUTE 文とピリオドを忘れたか、の何れかであった。また、変数がコラム上のどこに存在するかという情報は提供されているものの、データ定義ファイル自体が提供されていないデータ・セットが稀にある。この場合にはデータ定義ファイルの作成をゼロから手伝うことになる。データ定義ファイルが OSIRIS¹³でしか用意されていない場合も同様である。

SPSS と SAS 以外の統計プログラムへの移植は二通りの方法で解決された。Stat Transfer と呼ばれるデータ変換プログラムが市販されており、CAP のコンピューターにインストールされていた。このプログラムは主な統計パッケージおよび表計算プログラムのシステム・ファイルを相互に変換することができる大変便利なものである。但し、特定のプログラムが Stat Transfer によってサポートされていない場合や、何らかの理由で変換がうまく行かない場合は、システム・ファイルをタブ区切りテキスト・ファイルとして出力し、他のプログラムにそれを読み込ませるといった手法を使うことが多かった。但し、この場合全ての変数ラベルと変数値ラベルが失われる。なお最新版の Stat Transfer では OSIRIS フォーマットのデータを他のプログラムのシステムファイルに変換できるようになっている。

¹³ ミシガン大学の社会調査研究所（ISR）が開発したメインフレーム・コンピューターの上で走る統計プログラム。かつては広範に利用されたが、今は使われていない。OSIRIS フォーマットで保存された過去のデータの変換が ICPSR で現在進行中である。

9. データの利用 - 教育の現場から¹⁴

政治学部には大学院 1 年生のために情報教育のためのクラスが用意されている。前半ではミシガン大学でのコンピューターの利用法、後半では統計の授業で使う統計パッケージ（主に Stata）の利用法について週 1 回 1 単位（通常の授業は 3 単位）の講義が行われた。講義は博士候補資格を得た大学院生が担当するのが常であった。私が受講した 1996 年度には一通り ICPSR のデータの利用法が講義され、実際にデータの申請、ダウンロード、データの統計パッケージへの読み込みが課題として与えられた¹⁵。ここでは大学院における統計教育一般については触れないが、政治学部大学院の統計学中級の講義では、学生が公刊された論文を選んで結果の再現を試みる課題が与えられた。統計学上級においては各自の研究関心に沿ったデータを入手し分析する課題が与えられるのが常であった。学生が ICPSR から入手できる調査データに大幅に依存したのは言うまでもない。

10. プライバシー保護とデータ・ファイルの無断複製防止

ICPSR の方針では、データの利用者は被調査者のプライバシーを侵害しないこと、並びにデータ・ファイルを他の個人並びに組織に提供しないことに合意する必要がある。実際の規定は以下の通りである¹⁶。

- イ) データを統計分析のみに利用し、特定の個人および組織を特定するために用いない。ただし ICPSR から書面による合意を取りつけた場合はこの限りではない。
- ロ) 偶然にデータ・セットに含まれる個人あるいは組織を特定してしまった場合、その情報を利用しない。また ICPSR に対して個人あるいは組織を特定するように求めない。
- ハ) 複数のデータ・ファイルを統合して、個人あるいは組織の特定を可能にするデータ・セットを作成しない。
- ニ) ダウンロードしたデータ等を他の個人、組織などに提供あるいは売却しない。ただし、ICPSR から書面で合意を得た場合はこの限りではない。

¹⁴ ICPSR 所蔵データの利用にとどまらずに、教育現場における機械可読データ・ファイルの提供と利用者に対する支援に関心がある方は、Jacobs (1991) を参照されたい。また近年のインターネットの発達が発達研究・教育の現場における調査データの再分析に与えた影響と、今後の可能性については、Clark and Maynard (1998) を参照されたい。

¹⁵ Won-ho Park 氏による 2001 年度の講義のホームページを以下の URL で見ることが出来る。
<http://www-personal.umich.edu/~wpark/ps514/>

¹⁶ 原文は以下の URL を参照。<http://www.icpsr.umich.edu/ORG/Policies/respuse.html> .

以上の規定が掲載されているウェブページによれば、データの利用者は以上の規定を読むことになっている。しかし実際どれだけの利用者がこの規定を読んでいるかは疑問である。ミシガン大学政治学部のコンピューター利用の講義では、被調査者のプライバシーの問題については一切触れることがなかった。ただしデータの二次分析からプライバシーの侵害が生じた事例は私の知る限り皆無である。

次にデータの無断複製であるが、私が CAP に勤務した期間に他大学からの問い合わせは無かった。ICPSR を通すことなく大学「組織」間でファイルを融通することは皆無だったのではないと思われる。一方、個人間の場合には若干事情が異なるように思われる。例えば、ミシガン大学においてある個人が CAP を通じてデータを手に入れた場合、そのデータ・セットは最低 2 週間ミシガン大学のネットワーク内でアカウントを持つ者であれば誰でもダウンロードできる状態にあった。したがって、あるクラスで 1 人の学生がデータの利用申請をした場合、他の学生たちは特段の手続きを経ることなくデータを手に入れたことになる。裏を返すと 1 人の利用申請者のためにデータをアップロードした場合、その後 2 週間間にデータに何が起こったかは全く把握していなかった。

電子メールを通じた利用の申請は簡便ではあるが、それなりに時間がかかるものであった。素データ、コードブック、データ定義ファイルが一つずつしか存在しないのであれば簡単だが、古いデータ・セットの場合は素データが複数のファイルに分割されていることが多かった。それは更に複数のデータ定義ファイルが存在することを意味する。データの読み込みでつまづく利用者が多かったことを考えると、とりわけ SAS や SPSS のポータブル・ファイルが提供されていない調査に関しては、SAS や SPSS のシステム・ファイルを作成した友人から複製させてもらった方が簡便だったであろう。

但し、これは逆にミシガン大学内部のデータの融通や複製については特に問題視されることがなかったことの証左ではないかと思われる。また、データを提供する OR の側としても、全ての手続きを杓子定規に実行すると作業量が増加するので、CAP のスタッフから学生・教員に対して各自が手続きを行うよう励行することは有り得なかった。ちなみにカリフォルニア大学バークレイ校のデータセンターの記述によれば、バークレイ校に所属する学生・研究者の間でデータを複製することは問題視されていない¹⁷。

¹⁷ 原文は以下の URL を参照。 http://ucdata.berkeley.edu/new_web/icpsropen.html

11. 結語¹⁸

以上 ICPSR の活動並びに末端の教育現場におけるデータ頒布の実際を概観した。最後に調査データの寄託および公開を促進するために有用であると思われる幾つかの事項に触れておきたい。

データの寄託は調査を担当した研究者にとっては彼らの研究に「不必要」な時間を割き、かつ貴重な予算を減らす邪魔物と映るかもしれない。しかし、調査の初期から保存並びに二次分析用の公開を念頭におくことで、寄託・公開に必要な予算・時間面でのコストを大幅に削減できるのではないと思われる。人間の記憶は不正確であり、3年前の調査についての正確な概要を記述することは、様々な資料を再読した上でなければ難しいであろう。調査票の最終稿を機械可読ファイルで提出しようとしても、コンピューター・ハードウェアが2, 3年で時代後れになる現状を考えれば、保存場所が定かではないことも多いのではないか。また、類似した複数のファイルが存在し、最終稿がどれかを確定するのに時間を要する可能性も高い。実際、データを保存するにあたって最も困難な問題は数値情報であるデータ・セットを適切に理解するために必要な文書の質の低さ、あるいはその欠落である (Collins 1996)。調査の初期の段階から何が保存のために必要かという指針があり、かつある程度の時間を割くことができれば、文書の散逸や必要な情報が得られないといった事態は回避できるはずである。

コーディングについては、一定の標準的なコーディング方式が確立していることが、重要であろう。これには三つの利点がある。第一に、コーディングの指針が簡潔明瞭な形でデータ収集前に与えられていれば、調査担当研究者にとっても、コーディング作業やデータ・ファイルを作成する際の様々な技術的問題を回避するための手引きとなるであろう。第二に、プライバシー保護以外には編集・修正の必要が無いデータ・セットを初めから作成することにより、データの寄託・公開にかかる時間を大幅に削減できる。第三に、研究結果が追試により再現されることが社会科学において重要であるという立場に立てば、調査担当研究者が分析に使用するデータ・ファイルと二次分析用のデータ・ファイルができるだけ同じ内容であることが望ましい。何故なら、研究の本質とは直接関係のない技術的問題、たとえば欠損値の指定方法等、により分析結果が微妙に(場合によっては大幅に)変わりうるからである。その意味で、調査の初期の段階からデータの公開・保存を念頭に作業を進めることで、調査担当研究者は自己の研究の科学的価値を高めることができるのではないと思われる (UK Data Archive, n.d.)

しかしながら、データの公開は調査担当研究者に一定の負担を課すことも事実であり、費用や時間の問題を避けて通ることはできない。研究者の自発性にのみ依拠するのは残念ながら楽観的であり、アメリカの National Data Archive on Child Abuse and Neglect (NDACBN)

¹⁸ 以下でアーカイブという言葉を使う場合、社会調査データの保存・収集・公開を主な任務とする機関全般を指している。ここでの議論はあくまで筆者の個人的見解であり、東京大学社会科学研究所 SSJ データ・アーカイブの方針ではないことをお断りしておきたい。

の場合、研究費支給の要件としてデータの寄託が義務づけられて初めて多くの研究者がデータの保存・公開に協力的になったという (Collins 1996)。研究費の取得にデータの公開が要件とされれば、研究費の数パーセントをそれに充てるという義務づけも可能であろう。事実、全米科学協会は研究費応募計画の中にデータ共有のための費用並びにその計画を含めることを要求している (National Science Foundation, n.d.)。そのような仕組みがない場合、資料的価値が高い調査や、質の高い学術調査については、アーカイブの側で公開用データ・ファイル並びにコードブック作成に必要な時間および労働力を提供するのものの考え方はないであろうか¹⁹。

なおインターネットが普及した今日、個々の研究者が自己のホームページを通じて調査データを公開すれば良いと考える向きもあろうが、その方法は多くの問題点を孕んでいる。個々の研究者がコードブックの作成等に通暁していればよいかといえ、必ずしもそうではない。例えば調査概要等の文書に不備がある場合、あるいは利用者の側で調査の細部についての疑問が生じた場合、それはまず間違いなく利用者から調査担当者への問い合わせへとつながる。1回の短い電子メールのやりとりで疑問が解消するという保証はない。また、研究者がコードブックの作成に通じていない場合や、特殊なデータベース・プログラムなどを利用した場合事態は悪化する。筆者はミシガン大学大学院在籍中に短期間リサーチ・アシスタントとして国際政治関係のデータを整理した経験があるが、個人ベースで公開されているデータはデータ収集過程やデータ・フォーマットなどが不明瞭であり大変苦勞した。また繰り返し調査担当研究者に問い合わせの電子メールを送ったが、最後には相手が疲弊し質問に対して返答しなかった。研究者個人がデータの公開を研究・教育の片手間に続けることは、時間的に困難だと思われる。

データ・アーカイブには以上のような事態を防ぐためにデータの保存に必要な文書を保存する、あるいはその作成に必要な技術的なアドバイスを与えるという役割がある。これによってメタデータの質が向上するのみならず、個々の研究者が利用者からの問い合わせに煩わされないようにすることができる。したがって、アーカイブの側でも、調査概要などの文書の作成や調査データのコーディングについてのノウハウを蓄積していくことが求められると思われる。

最後に、社会調査データの安易な公開は、かえって新たなデータの収集を阻害するという意見についても一言述べておきたい (例えば、Herrnson 1995)。筆者の専門の政治学では、1995年にアメリカ政治学会の機関紙 *PS: Political Science and Politics* (September 1995) 上で、出版された結果の再現性 (replication) およびデータの公開をめぐる誌上論争が行われている²⁰。そこでの議論においては、大規模調査データに大幅に依存する社会学、社会心

¹⁹ 類似した意見について Beedham (1999) を参照されたい。

²⁰ そこでの論争は多岐にわたり、狭い意味での社会調査データの公開のみならず、エリートに対するインタビューに基づいた資料や、既存の公開データを組み合わせで作ったデー

理学、あるいは政治学のような分野で、追試による結果の確認と、違った問題関心からの二次分析が、学問の発展および教育上必要であることには異論が無いように思われる。むしろ抽象的な理念が実行に移される際に、如何なる方法が妥当かという点について対立が生じている。具体的には政治学が科学である以上、学術論文に利用されたデータは基本的に公開されるべきであるという立場 (King 1995a; Meier 1995) と、実際に調査を立案、企画、実施した研究者には一定期間 (あるいは永遠に) データの独占的使用権が有るという立場 (Herrnson 1995; Gibson 1995) がある。具体的に各人がおかれた境遇により態度が異なるのは自然であり、調査を実施する研究者が消極的であるのに対し (Sniderman 1995)、公開データに依存する若手研究者や方法論者が積極的だというのは驚くにあたらない (Box-Steffensmeier and Tate 1995)。

筆者は調査を実施した研究者に一定のデータ独占利用期間があるべきだという議論に異論はない。調査担当研究者がどれだけの独占利用期間を持つべきかについての議論に深入りする気はないが、その期間としては調査実施から2年から5年程度が適当ではないかと考えている。先述したように時間の経過と共に様々な情報が遺失する危険性を考えれば、寄託・保存の手続き自体は調査終了直後から開始されるべきである。その場合は2年程度で全ての手続きが完了するであろう (もっともこの場合、データの独占利用期間は実質的には皆無であろう)。一方、もし5年以上寄託の手続きが取られなかった場合は、データは散逸の危機に直面しているように思われる²¹。また、データを公開可能にしてから長期間公開しないのはアーカイブの運営上好ましくない。2年から5年というのはそういう観点から導きだされた数字である。この点に関しては、様々な方面からのご意見、ご批判を頂ければ幸いである。

本論に戻れば、二次分析用データの公開が新たなデータの収集を阻害するという意見の根拠は疑わしい。調査で自己の研究関心に沿った質問が聞かれていないことの落胆は誰しも経験したことがあると思われるが、自らの研究関心に沿って調査を設計する機会を、データの独占的利用権が無い、あるいはその期間が短いという理由で放棄する研究者がいるであろうか？さらには、二次分析用データの公開はむしろ一次データの収集を触発するのではないと思われる (King 1995 b)。何故ならば、二次分析が広範に行われることによって初めて、既存のデータを用いた分析の限界、並びにその限界を克服するために必要な調査の設計が明らかになるからである。事実、アメリカ政治学において研究者による一次データの収集が活発になったのは、ICPSR が設立された 1960 年代以降である (King 1991)。

夕等についても議論が戦わされている。また、統計分析の追試 (verification) も議論の焦点であるが、データの公開が無ければそもそも追試はできないわけであり (Gibson 1995)、便宜上ここでは社会調査データの寄託・公開に関する議論のみを紹介したい。

²¹ 筆者はさる高名な先生が「このデータはもうどこにあるか分からない」とおっしゃったことを鮮明に記憶している。先生は調査企画者のお一人であった。その段階で調査実施から15年が経過していた。

さらに、データが公開されることはそのデータの一次分析に基づく作品の価値をむしろ高めると言っても過言ではない。ミシガンの選挙研究者が刊行した『アメリカの投票者』(Campbell et al. 1960)は社会調査データに基づく選挙研究の古典中の古典であるが、彼らの集めたデータが速やかに公開されなければ、そして同一のデータに基づく反論と再反論という形での論争が無ければ、この本が現在学会において有しているだけの影響力を持ったとは思われないのである (King 1995b)。

参考文献

- American Statistical Association. *Ethical Guidelines for Statistical Practice*, 1999.
Available at <http://www.amstat.org/profession/ethicalstatistics.html>.
- Archival Development Committee. "New Wine in New Bottles: Moving to NAPA." July 1997.
Available at <http://www.icpsr.umich.edu/ORG/Policies/arcdev.html>.
- Allen, Zack W. Jr., Shelly Bitner, and Pat Randle. "Electronic Documentation at ICPSR: Migrating to a Standard Format." *ICPSR Bulletin*, Summer 1999.
- Beedham, Hilary. "The Royal Statistical Society Working Group on Archiving Data." *IASSIST Quarterly*, Vol. 23 No. 3 Fall 1999.
- Box-Steffensmeier, Janet M. and Katherine Tate. "Data Accessibility in Political Science: Putting the Principle into Practice" *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995.
- Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. *The American Voter*. John Wiley and Sons. 1960.
- Clark, Rich and Marc Maynard. "Using Online Technology for Secondary Analysis of Survey Research Data - "Act Globally, Think Locally." " *Social Science Computer Review*, Vol. 16 No. 1, 1998.
- Collins, Patrick T. "Establishing Data and Documentation Standards for Investigators who are Required to Archive Research Data." *IASSIST Quarterly*, Vol. 20 No. 3 Fall 1996.
- Converse, Philip E., Leslie Kish, and Merrill Shanks. "Warren Miller Remembered." *ICPSR Bulletin*, Spring 1999.
- Corrozza, Mark A. and Steven R. Howe. "Global Access to Data Resources: Where's the Metadata?" *IASSIST Quarterly*, Vol. 22 No. 2 Summer 1998.
- Dunn, Christopher S. and Erik W. Austin. "Protecting Confidentiality in Archival Data Resources." *ICPSR Bulletin*, Fall 1998.
- Gibson, James L. "Cautious Reflection on a Data-Archiving Policy for Political Science" *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995.
- Granda, Peter. *Census 2000: Accessing and Utilizing Data Resources*.
Its syllabus is available at <http://www.icpsr.umich.edu/TRAINING/Biblio01/toc.html>.
- Green, Ann, JoAnn Dionne, and Martin Dennis "Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata." Council on Library and Information Resources, June 1999. Available at <http://www.clir.org/pubs/abstract/pub83abst.html>.
- Herrnson, Paul S. "Replication, Verification, Secondary Analysis, and Data Collection in Political Science" *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995.
- ICPSR. "Survey of Official Representatives," 1997. Summary Report.
- ICPSR. *ICPSR 1999-2000 Annual Report*.

ICPSR. *Guide to Social Science Data Preparation and Archiving 2nd ed.*, November 2000.

ICPSR. *Handbook for Official Representatives*, October 2001.

Jacobs, Jim. "Providing Data Services for Machine-Readable Information in an Academic Library: Some Levels of Service." *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 1991.
Available at <http://info.lib.uh.edu/pr/v2/n1/jacobs.2n1>

King, Gary. "On Political Methodology." *Political Analysis*, Vol. 2, 1990 (published in 1991).

King, Gary. "Replication, Replication." *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995 a.

King, Gary. "A Revised Proposal, Proposal." *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995 b.

Meier, Kenneth J. "Replication: A View from the Streets" *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995.

National Data Archive on Child Abuse and Neglect. *Depositing Data with the National Data Archive on Child Abuse and Neglect: A Handbook for Investigators, revised ed.* 1999.
Available at <http://www.ndacan.cornell.edu/ndacan/Publications/handbklisting.html>

National Institute of Justice. *Depositing Data with the Data Resources Program of the National Institute of Justice: A Handbook.* 2000.
Available at <http://www.icpsr.umich.edu/NACJD/nijhandbook.pdf>

National Science Foundation. "Data Archiving Policy." n.d.
Available at <http://www.nsf.gov/sbe/ses/common/archive.htm>

Rasmussen, Karsten Boye. "Documentation - What We Have and What We Want: Report of an Enquete of Data Archives and their Staff." *IASSIST Quarterly*, Vol. 19 No. 1 Spring 1995.

Sniderman, Paul M. "Evaluation Standards for a Slow-Moving Science" *PS: Political Science and Politics*, Vol. 28, No. 3, September 1995.

UK Data Archive. *Good Practice in Data Documentation.* n.d.
Available at <http://www.data-archive.ac.uk/creatingData/goodPractice.doc>

Zelenock, Tom and Kaye Marz. "Archiving Social Science Data: A Collaborative Process." *ICPSR Bulletin*, May 1997.

URL が示してある場合を除き ICPSR 関係の文書・論文は次のページから入手可能
<http://www.icpsr.umich.edu/ORG/Publications/pubs-email.html>

IASSIST Quarterly は以下の URL から入手可能
<http://datalib.library.ualberta.ca/>

この論文に引用されていない文献に興味がある場合、ICPSR 夏期講習の Providing Social Science Data Service の文献リストを参照されたい。

有益なホームページへのリンク集

Council of European Social Science Data Archives
<http://www.nsd.uib.no/CESSDA/index.html>

Harvard-MIT Data Center
<http://data.fas.harvard.edu/>

International Association for Social Science Information Service & Technology
<http://datalib.library.ualberta.ca>

International Federation of Data Organizations
<http://www.ssd.gu.se/ifdotitel.html>

Inter-University Consortium for Political and Social Research
<http://www.icpsr.umich.edu/>

National Data Archive on Child Abuse and Neglect
<http://www.ndacan.cornell.edu/>

Roper Center for Public Opinion Research
<http://www.ropercenter.uconn.edu/>

UC DATA, Survey Research Center, University of California, Berkeley
<http://ucdata.berkeley.edu/>

UC San Diego, Social Sciences Data Collection
<http://ssdc.ucsd.edu/>

東京大学社会科学研究所附属日本社会研究情報センター S S J データ・アーカイブ
<http://ssjda.iss.u-tokyo.ac.jp/>

Appendix 1 . データ・アーカイヴィストの募集要項

Job Family: Professional/Administrative
Posting No: T-021628-PR
Job Title: DATA ARCHIVES SPECIALIST
Grade: 05
Department: ICPSR
Post Begin/End Date: 11/13/2001 // 12/07/2001
Job Code: 256500
Hours: 40.00

DUTIES:

Provide technical and substantive support services for the users of data holdings in the Inter-university Consortium of Political and Social Research (ICPSR) machine-readable data archive and for the professional and technical staff of ICPSR. Convert into standard formats appropriate for long-term archiving, preservation, and distribution, recent data holdings acquired as SAS/SPSS/Stata system files. Work with User Support staff to evaluate, develop, and utilize, as appropriate, procedures for updating data collections in a variety of legacy formats for use in current technological environments. Work with technical and supervisory staff to prepare reports and estimates for users and administrative staff as needed.

DEPARTMENT QUALIFICATIONS:

DESIRED QUALIFICATIONS:

Prior work experience with users of social science in a data library and/or a computing or research facility is very desirable. Knowledge and experience with Unix, Windows, or similar systems is extremely desirable. Experience with data base development and management. Ability to, with direction and guidance, develop and take full responsibility for a project or assignment.

MINIMUM QUALIFICATIONS:

College degree, preferably Masters, in a social science and/or information technology field plus approximately two years of related work experience. Work-related and diverse experience using SAS and/or SPSS, including data processing and analysis. Demonstrated interpersonal and communication skills plus the ability to work well with a variety of individuals in both individual and team settings. Ability to work in an environment that requires attention to details and working with procedures, while retaining flexibility to respond to changing demands and needs within a short time frame and for numerous clients. Ability to manage large and diverse information accurately and in a timely manner. Ability to work independently and in a team environment. Excellent oral and written communication skills in English.

Salary Grade	Minimum	Midpoint	Maximum
05	\$23,000.00	\$38,900.00	\$54,800.00

(<http://www.umich.edu/~jobs/current/bydepartment.html#ICPSR>)

Appendix 2 . 文書部門職員の募集要項

Job Family: Office
Posting No: T-021924-PR
Job Title: OFFICE ASSISTANT II
Grade: 05
Department: ICPSR
Post Begin/End Date: 11/20/2001 // 12/14/2001
Job Code: 511340
Hours: 40.00

Candidate in Mind

DUTIES:

To perform electronic document conversion of hardcopy and electronic versions of social science research documentation. Use various scanning techniques and software (Acrobat, Paperport, etc.) on a Pentium workstation in combination with a desktop scanner. Format and edit of electronic files (text and image) after scanning for conversion to portable document format (PDF). Assist in training & orientation of new temporary employee team members. Serve as a continuing mentor for new team member placed as office partner. Convert electronic documents to ASCII text, PDF, and/or image formats. Extract text from paper documents resulting from submission to optical character recognition (OCR). Assist in the creation of specialized CD-ROMs.

DEPARTMENT QUALIFICATIONS:

Desired: At least one year previous work experience scanning hard copy documents and converting those files to machine readable, computer format. Exceptional computer skills in regard to UNIX and Windows operating systems. Excellent word processing skills using Microsoft Word and Word Perfect.

Minimum: Two years of college or equivalent including course work utilizing skills; reliability in handling confidential files; one to two years communication skills. Ability to work as part of a team. At least one year of experience working with Adobe Acrobat and related scanning software.

MINIMUM QUALIFICATIONS:

High school diploma or equivalent combination of education and experience is necessary. Some secretarial or related experience is necessary. Ability to keyboard with speed and accuracy is necessary. Reasonable knowledge of and ability to use standard business software and computer applications is necessary. Some experience in spreadsheeting may be necessary.

Questions in regards to applying for this position can be E-mailed to employment.services@umich.edu

Salary Grade	Minimum	Midpoint	Maximum
05	\$23,000.00	\$38,900.00	\$54,800.00

(<http://www.umich.edu/~jobs/current/bydepartment.html#ICPSR>)

Appendix 3 . 文書部門管理職の募集要項

Job Family: Professional/Administrative
Posting No: T-020966-PR
Job Title: PRODUCTION MGR
Grade: 09
Department: ICPSR
Post Begin/End Date: 11/30/2001 // 12/07/2001
Job Code: 162320
Hours: 40.00

DUTIES:

Manage the administrative and operational activities of a ten-person group engaged in Electronic Document Conversion (EDC) at the Inter-university Consortium for Political and Social Research (ICPSR), a unit within the Institute for Social Research at the University of Michigan.

Plan, prioritize, and assign the work of the unit's personnel in converting paper and electronic documents to Acrobat PDF format. Coordinate the flow of services between EDC and other units of ICPSR. Hire, train, supervise, and evaluate personnel in the unit. Participate in ICPSR planning and policymaking committees, including at meetings of the elected ICPSR Council. Recommend acquisition of equipment and software for the EDC unit. Prepare cost and delivery schedule estimates for internal and external clients. Prepare reports on unit activities for management and the Council. Represent the EDC unit and ICPSR at professional meetings, conferences, workshops, etc. Monitor budget expenditures for the EDC unit. Plan, develop, and oversee the preparation of specialized CD-ROM products.

DEPARTMENT QUALIFICATIONS:

Necessary: Significant work experience in managing production activities. Demonstrated problem-solving and multi-tasking abilities. Thorough understanding of a wide range of computational hardware and software, including desktop computers, server networks, and peripheral devices. Significant work experience with word processing, database, and spreadsheet software. Excellent communications skills, both oral and written, in the English language. Demonstrated ability to work in a team environment with people from diverse backgrounds.

Desirable: Education beyond the Bachelor's degree, with specialization in information technology or the social sciences. Work experience with Adobe Acrobat suite of software for document conversion, and with document scanning equipment. Formal training in techniques of personnel management. Familiarity with social science data and system, desktop publishing software, and graphics software packages.

MINIMUM QUALIFICATIONS:

Bachelor's degree in an appropriate field or an equivalent combination of education and experience is necessary; in certain cases, an advanced degree is necessary. Some experience in the appropriate discipline is necessary. Some experience in public speaking is necessary. Some writing experience is desirable. Some teaching experience is desirable.

Questions in regards to applying for this position can be E-mailed to employment.services@umich.edu

Salary Grade	Minimum	Midpoint	Maximum
09	\$32,700.00	\$55,100.00	\$77,500.00

(<http://www.umich.edu/~jobs/current/bydepartment.html#ICPSR>)

Appendix 4 . データ利用申請の電子メール (例) 実物だが名前を変えてある

From: "George W. Bush" <gwbush@umich.edu>
Date: Tuesday, January 18, 2000 03:02 PM
To: cap@umich.edu
Subject: data

Hi -

I would like access to study 6438, Americans' Changing Lives.
I also need the codebook.

Thanks,
George

From: "Al Gore" <algore@umich.edu>
Date: Wednesday, November 17, 1999 12:28 PM
To: cap@umich.edu
Subject: data request

Hi Capsters,

Could you please post the General Social Survey cumulative file (#2685)
on the IFS space? Thanks very much.

Al Gore
Assistant Professor, Department of Communication

From: "John McCain" <mccain@umich.edu>
Date: Monday, November 15, 1999 07:21 PM
To: cap@umich.edu
Subject: request for data

Dear Sir/Madam,

I am a Finance Ph.D. student here at the University of Michigan Business School. I need some data for my research thesis. I talked with you today around 3pm and you asked me to send this e-mail. Basically, I need the International Financial Statistics Data (Study Title) provided by International Monetary Fund (Principal Investigator). The code for it is: ICPSR - 7629. I need monthly data from 1974 through 1998 (If you need more information about the specifics of the data that I need I can give you further information).

Thank you very much.
Best regards,
John

Appendix 5. データ申請者に対して CAP から送られる電子メール データ申請者の名前は変えてある。

From: Yukio Maeda <ymaeda@umich.edu>
Date: Tuesday, November 16, 1999 09:23 AM
To: "Donald Rumsfeld" <rumsfeld@umich.edu>
Cc: cap@umich.edu
Subject: Re: request for data # 7629

Hi, please see below for details.
The codebook is directly downloadable from the ICPSR website.

Thanks,
Yukio

The data set you ordered from ICPSR is available to you via File Transfer Protocol (FTP). This message includes instructions on how to obtain your data.

OBTAINING DATA SETS VIA FTP

=====

To download the data you will need an FTP program such as RapidFiler (for Windows) or Fetch (for the Mac). Contact 4-HELP if you need assistance with FTP. You should transfer all the files in binary mode.

The files you requested are in CAP's IFS space. To access this file space, use your FTP program to connect to login.itd.umich.edu (authenticate with your unickname and password). If you connected properly, you will initially be in *your* IFS space, which is:

/afs/umich.edu/user/firstletter/secondletter/yourunickname

Change directories to CAP's IFS space, which is:

/afs/umich.edu/group/lisa/compap

You may also refer to this directory as '~compap' (without the quotes). The data files you requested are in a subdirectory of compap defined by the data's ICPSR study number.

FILE NAMES

=====

All available files for your data set are in this directory. All or some of the following types of files may be available:

File Prefix Contents
DA Data
CB Codebook (text version)
CB Codebook in Adobe Acrobat format (if file has .pdf suffix)
DD Data Dictionary
DI Data Dictionary
SP SPSS Portable Data File (if file has EXPORT suffix)
SP SPSS Syntax file for reading data (may need editing)
SA SAS Syntax file for reading data (may need editing)
DOC Technical Documentation

ODICT Osiris format data dictionary
ODATA Osiris format data file

FILE COMPRESSION

=====

The files are compressed in gzip format, which can be decompressed with the gzip program.

The program is available on Unix -- type 'gzip -d [filename]' (without the quotes) at the %prompt to decompress a file.

The program is also available as a DOS program for IBM-compatible PCs. To obtain it, download the file gnu.exe from the gnu subdirectory of the compap directory. Gnu.exe is a self-extracting file that contains the gzip program and documentation for your PC. To extract the gzip files just type 'gnu' (without the quotes) at a DOS prompt, or double-click on the gnu.exe file in File Manager. Once you have the gzip program extracted, you can decompress a gzipped file by typing 'gzip -d [filename]' (without the quotes) at a DOS prompt. (Note that gzip should be in the same directory as your data file, or in a directory that is on your ms-dos path.)

TIME LIMITS

=====

Because we have limited IFS space, please download your files as soon as possible. We will leave them on IFS for as long as possible, but make no guarantees beyond two weeks.

CODE BOOKS

=====

If you need a copy of the codebook, and a machine-readable version is not available, you may purchase a copy from ICPSR Member Services. Or, if we have the codebook on the shelf, you may check it out from the CAP office for 3 weeks. (It may take us a week to get the code book on our shelf.) A personal check or U-M account number is required for a refundable deposit on the codebook.

Good luck, and let us know if you have any questions.

Yukio Maeda, Amaney Jamal, Won-Ho Park
Computer Assistance Program
5613 Haven Hall
615-0316

