

「ICPSR データの利用と課題」

西澤由隆（同志社大学）

1. はじめに

ICPSR(The Inter-University Consortium for Political and Social Research) は1962年に設立され、現在はミシガン大学のISR(The Institute of Social Research) の1部門として位置付けられている、世界でも最大の社会科学系のデータ・アーカイブである。⁽¹⁾ その運営は会員制をとっているが、1997年の時点で世界中で450を超える大学がメンバーとして登録している。⁽²⁾

また、年間にアクセスされるデータセットの本数が、延べ74,000本(2,220,000メガバイト) を超えるという数字からもその規模の大きさが分かる。⁽³⁾

ICPSRは、その使命を3つ掲げている。それは：

1. 社会科学データの収集とアーカイブの管理
2. 公開・公平を原則としたデータの運用
3. 効率のよいデータ利用環境の整備

である。⁽⁴⁾

コンピュータ技術の進歩と、データ収集が社会科学の道具として一般化する中で、社会科学者が年間に「生産」するデータの総数は飛躍的に膨らんでいる。そのような状況にあって、信頼できるデータを、迅速に、しかも公平に利用者に提供する営みは、きわめてチャレンジングなものである。その使命がどの程度達成できているかは、利用者一人ひとりが判断するところであるが、私の限られた経験から評価するとしても、相当程度の成果を上げているといえる。

さて、このICPSRについて、(1)その運営の現状と将来への課題と、(2)その利用方法について解説するのが、企画委員から私に課せられた課題である。以下、次のような構成で、議論を進めていく。⁽⁵⁾

2. ICPSRの組織とサービス

2-1. 組織構成

2-2. アーカイブの内容

2-3. 今後の課題

3. ICPSRの利用方法

3-1. Websiteからのダウンロード

3-2. 非加盟校のユーザーにも利用可能なリソース

2. 組織とサービス

2-1. 組織構成

ICPSRは、形式的にはISRの中のThe Center for Political Studies(CPS)の管理下に位置づけられているが、実質的には独立の組織として運営されているというのが私の印象であった。⁽⁶⁾ そしてそこでは、既に紹介した「ミッション」を達成するために、80人の専属のスタッフが、データの収集整理(Archive)・教育(Education)・コンピュータおよびネットワーク(Computing and Network)の3つの部門に配属されて活動している(1996-1997)。このように、量的にみるかぎりスタッフがずいぶん充実しているようではあるが、年間に200本以上のデータセット(延べファイル数にして約2,000本)が年間に新しくリリースされるという実績を考えると、けっして大きすぎる所帯ではなさそうである。

ICPSRの運営において、特に注目すべきことは、それが利用者コミュニティに開かれているという点である。ICPSRはミシガン大学にあるが、けっしてミシガン大学の学者のみがその運営方針を決めているわけではない。そのことを制度的に保証しているのが、Official RepresentativeとCouncilである。

ICPSRの加盟校には、それぞれOfficial Representative(OA)と呼ばれる代議員が置かれることになっている。OAは、それぞれの大学においてユーザー番号やパスワードを管理したり、データのアクセスに関する技術的な補助をするばかりか、加盟校の代表として隔年に召集される代表者会議に出席することができる。会議は、ICPSRのその時点での運営状況と将来的な活動方針をメンバー校に説明する機会であると同時に、各校からの要望を聞くなど、ユーザーからのフィードバックを得る機会としても位置づけられている。機関誌(Bulletin)には、そこでは「活発な議論(lively discussion)」が行われていると報告されている。「開かれた」組織としての重要な機能を果たしているようである。⁽⁷⁾

また、その運営方針・予算の作成・人事などについて実質的に決定するのが、Councilである。Councilは、現在は12名の社会科学の研究者で構成されているが、それぞれの任期は4年で、無報酬である。「開かれた

組織」の点から重要なことは、そのメンバーが、ミシガン大学やISRの研究者に限定されるのではなく、広く学会の各分野から選ばれていることと、それがOAによって選出されていることである。ここ数年の構成メンバーを見る限り、2年ごとにその半数が入れ替わっている。特定の研究者グループによる「寡占」ということを避ける努力の証であるといえるだろう。

2-2. アーカイブの内容

ICPSRが所蔵する膨大な量のデータセットの全容をこのような報告書で紹介するのは難しい。その全体像を一望するには、ICPSRが発行するカタログ (Guide to Resources and Services) を手にするのがよい。電話帳を思わせるような厚みのガイドブックには、所蔵する全データの study title・study number・principal investigator などの基礎的な情報に加えて、その調査の概要が紹介されている。また、以下に紹介するように、ICPSRのホームページにも、その電子版がある。こちらの方は、一般に公開させているので、非メンバー校の研究者にもアクセスが可能である。

現在のところ、全データセットを図1のとおり、18のカテゴリーに分類して整理している。選挙研究者にとって最も関心のあるカテゴリーはXIV. Mass Political Behavior, Attitudesであると考えられるが、そこはさらに図2のように細分されている。そして、たとえば、A-2-a: Election Studies Series-United Statesには American National Election Studiesシリーズの一連のデータセットが整理されている。また、A-2-b: Election Studies Series-Nations Other Than the United Statesにはオーストラリア・イギリス・カナダと各国の National Election Studyシリーズが並ぶ。日本の有権者を対象とし調査では、R. ワード・A. クボタの1967年のいわゆる「ミシガン調査」や三宅一郎の「宇治調査」もこの中に分類されている。

選挙関連以外の主なものを2・3列挙するとしたら、アメリカ連邦議会の法案に対する議員の投票状況データ・アメリカの国勢調査データ・Euro-Barometers and European Communities Studies・General Social Surveysなどがあげられる。

2-3. 今後の課題

これほどの規模の機関であるから、財政問題を初め一般の組織が共通に抱える多くの問題をICPSRも持っていることは想像に難しくない。ここでは、ICPSRの特殊性から生じる2つの問題(課題)を紹介しておこう。それは、(1) コンピュータ・ネットワー

クを中心とした情報通信技術の急速な変化にどう対応するかというデータの供給媒体にかかわる問題と、(2) データの受け入れ基準の確立の問題である。

供給媒体 コンピュータとそのネットワークに関連した技術が数ヶ月の単位で変化する今日、長期的な展望でデータ供給のシステムを構築するのは容易なことではない。実際、ICPSRのディレクターであるR. ロックウエルは、最新のAnnual Reportで、「技術的分野での選択において、ICPSRは幾度となく失敗をした ("ICPSR has made a series of mistakes in its technological decisions.")」と告白している。⁽⁸⁾

長い間、ICPSRでは、データの配給はマグネット・テープに依っていた。ところが、やがてフロッピーディスクへと媒体が変わり、さらに1990年前後からは、the Consortium Data Network (CDNet)を利用したftpによるネットワーク・サーバーからの直接のダウンロードが可能になった。そして、現在は、wwwのホームページからのダウンロードと、供給媒体が急激に変わっている。

データの配給についての問題が一段落した現在、ICPSRが進めている作業が、コードブックなどのデータに付随するドキュメントの標準化された電子化である。これまで、ICPSRでは、利用者の多様な使用環境に対応すべく、複数の書式の文書を用意してきた。具体的には、複数のワードプロセッサの書式・ASCII・formatted ASCIIなどである。ところが、このように複数の書式をいずれのデータセットについても用意するのは、膨大なコストがかかる。そこで、経費節約のために、書式の一本化が進められている。現在のところは、the Adobe Acrobat Readerに対応した、Portable Document Format (PDF) とすることが決まり、それにむけて作業が進んでいる。⁽⁹⁾

受け入れ基準 ICPSRが直面するもう1つの課題は、増加の一方をたどる所蔵データの管理をどのようにするかという問題である。ICPSRは、単にデータを収集することを目的としてはいない。その目的は、利用可能なデータの供給である。つまり、不特定多数の研究者が利用可能なように、データを整理(場合によっては編集・加工)する必要がある。つまり、送られてくるデータの質によって差はあるものの、大なり小なりICPSR側に負担がかかるわけである。ところが、リソースは有限であるため、無制限にデータを受け入れるわけにはいかない。したがって、何らかの「受け入れ基準」を設ける必要がでてきたのである。

重要なことは、これが、新規データ・セットの受け入れにかかわる問題だけではないという点である。というのは、多くのデータは、たとえばNational Election Studyや国勢調査データがそうであるように、継続的に増加する、いわゆる「シリーズ物」である。つまり、「現在受け入れているタイトルの全てを将来に渡って受け入れ続けることがもはやできない」というのである。

そこで、ICPSR Councilは、特別委員会(the Archival Development Committee)を設置し、1996年10月からその対応策を検討してきた。その報告書が、"New Wine in New Bottles: Moving to NAPA--Policy for ICPSR Archival Development Report of Archival Development Committee"として、1997年7月にまとまっている。⁽¹⁰⁾

そこでは、次のような4点の「基準」を提案している。

1. 教育および研究における実質的な価値・歴史的な価値・奇抜さの3つのいずれかの観点から、その重要性が示されるもの
2. 相当程度の需要が見込まれるもの
3. 妥当なコストで他の機関から入手することができないもの
4. データの整備にかかるコストが特に高いものについては、そのコストが正当化されるだけの重要性を持つもの

これら基準は、それなりに妥当なものであるが、「重要性」や「歴史的価値」をどのように測定するかは明言されていない。また、現時点での需要からは、10年(あるいは50年)先の需要は予測できない。つまり、これらの基準の実質的な運用は、なかなか困難であるといわざるをえない。とりあえずレポートが出たものの、この点についてどのように対処するかが、ICPSRが抱える当面(そして緊急)の課題であるという事態に変わりはない。

3. ICPSRの利用法

すでに紹介したとおり、ICPSRはメンバー制をとっている。加盟校に所属する教職員・学生は無制限のデータ利用が保証されている。そのほか、夏に開校されるSummer Seminarに参加する権利を有する。

以下では、まず、Websiteからのデータのダウンロードの手順を説明する。そして、最後のセクションでは、非加盟校の研究者にも利用可能なリソースを紹介する。

3-1. Website(ホームページ)からのダウンロード

(1) ICPSRのWebsiteへの接続

- ・ WWWのブラウザ(Netscape・MSエクスプローラなど)を起動する。
- ・ URLは、www.icpsr.umich.edu
- ・ 図3がICPSRの初期画面(ホームページ)

(2) ダウンロードするデータの特定

- ・ 初期画面の「archive」をクリックし、検索画面へ移動する。(図4)
- ・ データ名やstudy #が分かっている場合は、画面左半分の該当する箇所に情報を入力する。図4は、「World Values Survey」をダウンロードすることを想定した例。図5は、当該のタイトルで登録されているデータの一覧表。
- ・ テーマ別にデータを絞り込むのであれば、図4の画面右半分の該当するテーマをクリックする。図6は、「XIV. Mass Political Behavior and Attitudes」の詳細画面。さらに、その「2. Election Studies Series」の「b. Nations Other Than the United States」をクリックすることで現われるデータの一覧表が図7。

(3) コードブックとデータのダウンロード

[ここでは、「World Values Survey, 1981-1984 and 1990-1993」をダウンロードすることを想定して、作業をつづける。]

- ・ 図5の中の「World Values Survey, 1981-1984 and 1990-1993」に対応するDA(データ)をクリックして、「Download」画面に移動する。(図8)
- ・ コードブックについては「Freely available」とあり、メンバーシップに関係なくダウンロードが可能である。
- ・ 「Uncompressed」をクリックすると、数秒の内にコードブックが表示されるので、それを「名前を付けて保存する」。(図9)
- ・ データファイルについては、「authorized users」とあるように、加盟校ユーザーのみがアクセス可能である。
- ・ 「Compressed」・「Uncompressed」のいずれかを選択する。すると、「ユーザーID」と「パスワード」を入力する画面が現われるので、必要な情報を入力し、「enter」キーを押すとダウンロードが始まる。なお、「ユーザーID」と「パスワード」は、当該の大学のORが管理している。ftp用に用意され

ているCDNet Accountに対するIDとパスワードが基本的にはWebsiteにも使用できる設定となっている。

3-2. 非加盟校のユーザーにも利用可能なリソース

非加盟校のユーザーに対して、ICPSRのリソースがまったく閉ざされているかというところでもない。ICPSR BylawsのArticle I., Section 3に規定されているように、有償ではあるがデータを使用することができる。⁽¹¹⁾ また、すでに示したように、コードブックについては、会員・非会員に関係なく、Website(ホームページ)からダウンロードが自由にできる。また、American National Election StudiesのCD-ROM(The 1994 Post-Election Studyまでの全てのデータが集録されている)などは、個人でも入手可能な価格で、非会員にも頒布されている。⁽¹²⁾ また、ANESについては、The 1996 Pre-/Post-Election Study以来のデータは、通常のICPSRを通じての配布とは別に、NESのWebsiteから直接ダウンロードすることができるようになっている。こちらの方は、ICPSRの加盟・非加盟にかかわらず利用が可能である。

4. まとめ

現在のところ、日本ではICPSRのメンバー校となっている大学は限られている。年会費が高額であることがハードルとなっているのであろうけれども、全国規模の面接調査を1つ実施するときの費用に比較したとき、そのコストパフォーマンスはずいぶん高いといえる。これを機会に、日本でもより多くの研究者がICPSRを利用するようになることを願うものである。

注

1. それぞれが所蔵するデータの総数は、私が調べた限りでは、公式な文書には明記されていない。特定のデータセットが、複数のシリーズものとして重複して登録されていたりするために、延べ数としての「所蔵データ数」は、その規模の実質的な比較においてあまり意味を持たない。おそらく、そのような理由で「所蔵データ数」を明示的に示していないものと思われる。ただし、「コンピュータ処理のできる形態での研究・教育用社会科学データ・アーカイブ(computer-based research and instructional data for the social sciences)」としては世界最大であると謳っている。
2. 全世界の加盟数については、正確な数字は公表されていないようである。この数値は、報告者が加盟校リストをもとに数えたもの。

3. "Five-Year Data Ordering Summary" ICPSR Annual Report 1996-1997, p. 12.

4. "ICPSR Mission Statement" ICPSR Annual Report 1996-1997, p. (ii).

5. なお、ここで紹介する内容は、ICPSR Annual Report 1996-1997とGuide to Resources and Services 1996-1997: Inter-university Consortium for Political and Social Research(いずれもICPSR発行)および、wwwのWebsite: <http://www.icpsr.umich.edu> (1998年4月30日現在のもの)を基にしている。ICPSRの広報・出版物は電子化がきわめて進んでおり、前二者の内容は、そのほとんどがWebsiteに反映されている。本報告では、紙面の関係でそのWebsiteの内容のうち、限られた部分についての要点を紹介するにとどまっている。詳細は、上記のWebsiteをぜひ参照されたい(図3 ICPSR ホームページ)。

6. 私は、客員研究員として1993年4月から2年間、CPSで過ごした。

7. ICPSR Bulletin, December 1997, Vol. XVIII, No. 2.

8. もっとも、「そのことを誇りに思っている("and we are proud of them")」とロクツウエルは続けている。避けられない「失敗」であったと同時に、その教訓は今日じゅうぶんに活かされているという自負があるのだろう。"Report from the Executive Director" ICPSR Annual Report 1996-1997, p. 4.

9. "ICPSR's Documentation Initiatives" <http://www.icpsr.umich.edu/ICPSR/Developments/doc.html> 参照。

10. "New Wine in New Bottles: Moving to NAPA--Policy for ICPSR Archival Development Report of Archival Development Committee" <http://www.icpsr.umich.edu/ICPSR/Developments/arcdev.html>

11. ICPSR Bylaws Article I-Section 3: Services for Non-membersには次のように規定されている。"ICPSR data and documentation will be provided to individuals who are not affiliated with a member institution. Such individuals requesting data for their research or instructional use will sign a loan agreement, and submit payment for charges assessed, before data are provided." Guide to Resources and Services 1996-1997, p. xi.

12. CD-ROMの購入には、NESのホームページから、「Studies, Data, and Other Resources」へ移動し、「Ordering the NES CD-ROM」へ進む。NESのホームページのURLは、www.umich.edu/nas

図1 ICPSR Data Archive

図2 サブテーマ

図3 ICPSR ホームページ

図4 検索画面

図5 検索結果 データ一覧

図6 テーマ別詳細画面

図7 テーマ別データ一覧

図8 「ダウンロード」画面

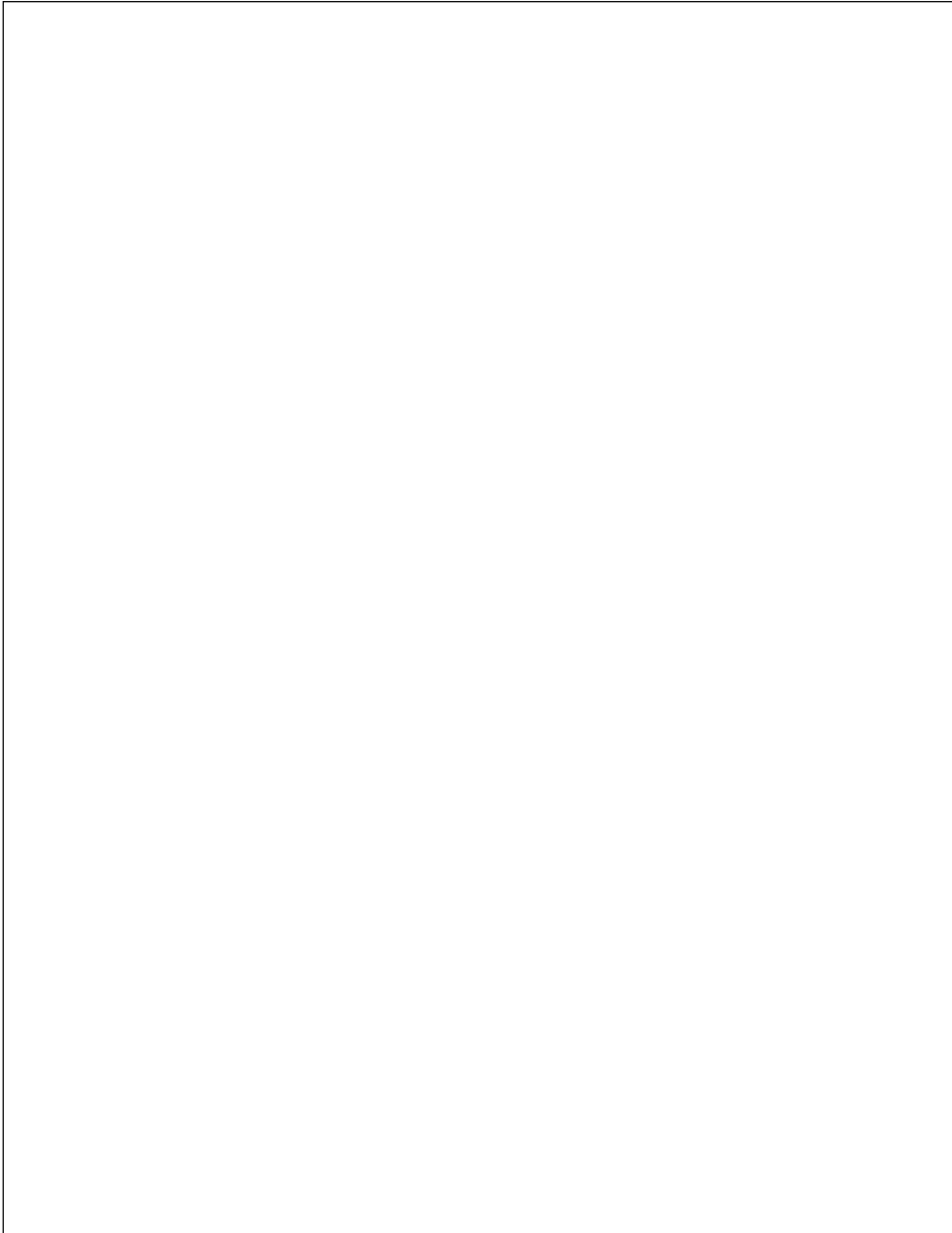


図9 ダウンロード後のコードブック