

## もう一つのアーカイブ

前田幸男

私の所属する東京大学大学院情報学環・学際情報学府の文化・人間情報学コースではアーカイブ・歴史情報が4つの柱の一つとして掲げられ、研究・教育が行われている。片隅から眺めて理解する限り、古文書だけでなく、様々な種類の媒体に保存された多様な資料の保存と公開について活発な研究・教育が行われているように思う。このレビューでは、学環・学府の守備範囲から少し外れたところにある、重要資料の保存・公開活動について紹介したい。それは社会調査データのアーカイブである。

普通アーカイブというと文書が中心であり、その他は地図、音声記録、動画等が保存対象と考えられるが、統計資料もその対象となりうる。社会調査の機械可読ファイル（磁気データ）は、伝統的なアーカイブ理論から見ると異端であり、アーキビストの中には紙媒体資料と同じだけの価値を認めず、一度紙に転写した上で保存し、必要に応じて光学的に再読込をするべきと主張する人もいる。しかし、統計データ利用者の立場から言えば、社会調査データはコンピュータ上で再集計可能なところに意味があり、紙に転記された数字を再度読み込むことは、時間的・経済的資源の浪費と見なされる。実証的研究に従事する社会学者にとっては、機械可読ファイルを保存・公開する組織の存在は、歴史家にとっての公文書館や資料館と同様に、不可欠なのである。

確率的に抽出された標本に対する調査と、その集計・分析は1930年代に始まるが、現存する最古の標本調査データはアメリカ合衆国で1930年代に収集されたものではないかと思われる。ポール・ラザースフェルドたちがコロンビア大学の応用社会調査研究所で1940年代に行った調査のデータは現在でも利用可能な状態で保存・公開されている。20世紀中頃までは、調査票の情報はパンチ・カードに転記され分析されたが、1960年代以降は、徐々に磁気媒体へと移行した。社会調査データのアーカイブはこれらの磁気媒体を保存・公開する組織ということになる。

社会調査データの保存・公開の先鞭を付けたのは米国であり、世論調査・市場調査で有名なエルモ・ローパーが、Fortune誌等のために行った調査のパンチ・カードを1947年にウィリアムズ・カレッジに寄託したことが端緒になる。その後ローパー・センターが1957年に同カレッジに設立された（後にコネチカット大学に移転）。大学に所属する研究者側では、ミシガン大学社会調査

センターが行った全米選挙調査のデータを共有するために、同大学の政治学者が1962年にICPRを設立した。政治以外のデータをより広く扱うようになったことから後にICPSR (Inter-University Consortium for Political and Social Research) へと名称が変更されている。欧州においては1960年にドイツ連邦共和国でケルン大学にZA (Zentralarchiv) が設立され、イギリスでは1967年に現在のUKDA (U.K. Data Archive) の前身がエセックス大学に設立されている。

世界のデータ・アーカイブの中でも中心的な役割を果たしているICPSRは大学単位の会員制機関であり、加盟校に所属する研究者の研究・教育上の要望に応じてデータを提供している（日本でも東京大学を含む20以上の大学が共同でICPSRに加盟している）。設立の当時は依頼を受けるとパンチ・カードや磁気テープを複製して送付していたが、現在では加盟校の教員・学生であれば、自由にICPSRのデータをインターネットからダウンロードして利用できるようになっている。事情は基本的にヨーロッパのデータ・アーカイブでも同様だが、今日においても全てのデータ・アーカイブが直接ダウンロード方式で運営されている訳ではない。

さて、日本における状況だが、私が兼務している社会科学研究所附属日本社会研究情報センターでは1998年からSSJDA (Social Science Japan Data Archive) と呼ばれるデータ・アーカイブを運営している。現在では所蔵データセット数は1000を突破し、学術目的の利用者も年間1000人を超えている。ただし、日本における社会調査データのアーカイブの歴史は短い。蒲島郁夫教授（現在・東京大学法学部）と筑波大学多目的データバンク・プロジェクトの尽力で、木鐸社に商業ベースで（実態としては殆ど社会奉仕活動として）レヴィアアサン・データバンクが設立されたのは1991年のことである。この他、SSJDAよりも一足早く、SORD (Social and Opinion Research Database) が1998年に札幌学院大学に設立されている。いずれにしても、日本の社会調査データ・アーカイブの歴史はわずか15、6年に過ぎない。

では、何故社会調査データのアーカイブが必要なのであろうか。一つの理由は調査に巨額の費用を要する以上、その結果得られたデータは、保存して再利用すべきだからである。例えば、現在の日本で計画標本規模4000、回収率50%（回収標本規模2000）で全国調査を行おうとすると、軽く見積もっても2000万円程度の費用がかかる。到底個人でまかなえる金額ではない。また収集されたデータも、よほど単純な調査でない限りは、異なる理論的な関心からの分析に十分耐えうる。しかしながら、調査を実施した研究者と縁のない研究者、特に大学院生にとって、他大学の教員と直接やりとりするのは気の重い作業である。また、調査を実施した側でも、一つ一つの依頼に丁寧に応えるのは時間的にも大きな負担となる。そこで、調査実施者と二次利用を希望する研究者を媒介する組織としてデータ・アーカイブが必要になる。

データ・アーカイブが必要な理由の二つめは、過去のデータが蓄積されていることが、新しい現象を理解するために不可欠だからである。社会調査データは、調査実施時点における社会状態

の断面図ともいえるが、後の調査と質問項目および母集団が共通の場合は、現状を理解する上で非常に有益な情報を提供し得る。例えば価値観の変化を検討する場合、一定期間における変化を追跡する以上、最低二時点における調査データを用いた検討が不可欠となる。即ち、過去の調査データが利用可能な形式で保存されていないならば、新しい社会現象が現れても、その原因を十分に掘り下げて研究することはできないのである。

第三の理由としては、調査データの保存には個人の力を超えた組織的な努力が必要だという点が挙げられる。他の媒体と同様だと思われるが、機械可読形式で保存された調査データの永年保存にはかなりの手間を要する。マイクロフィルムが劣化するのと同様、磁気媒体も経年劣化するので、定期的に新しい媒体へと転写される必要がある。この他にも苦労は多い。例えば保存形式だが、一般的に使われているソフトウェアがあれば、そのシステム・ファイル形式で保存するのが簡便ではある。しかし、長期的な保存を念頭におくと、ソフトウェアの更新（生産中止）に伴い場合によって読めなくなるかもしれない形式で保存することは必ずしも得策とは言えない。従って、コンピュータ・ハードウェアおよびソフトウェアの技術革新に合わせて、保存の戦略を立てることが必要になる。

第四の理由は標準化されたメタデータの整備が必要だからである。機械可読ファイルも、その内容を説明するメタデータがなければ、無用の長物にすぎない。具体的には標本抽出の方法と調査票の情報が主だが、メタデータは当事者以外が数十年後に読んでも意味が通じるように準備する必要がある。例えば標本抽出も、「日本人を代表するように抽出した」という記述は失格で、母集団の範囲、台帳の種類、層化の方法、調査地点の抽出方法、情報収集の方法等を誤解の余地無く記述しなければならない。

このような地道な作業を経て、社会調査データの長期的保存と二次分析は初めて可能になる。日本ではデータ・アーカイブの歴史は浅いが、最近は徐々に国際的な認知も得てきている。社会調査データの分析に興味のある研究者、大学院生の方には積極的にSSJDAや他のアーカイブのデータを利用していただきたい。また、調査を実施された場合は、信頼のおけるアーカイブにデータとメタデータを寄託することをお勧めする。



前田幸男（まえだ ゆきお）

1969年10月30日

【専攻領域】政治学・社会調査

【著書・論文】

「大都市住民と選挙」（財）東京市政調査会編『大都市のあゆみ』東京市政調査会：249-276頁、2006年9月。

「性役割分業と政治参加」永井暁子・松田茂樹編『対等な夫婦は幸せか』勁草書房：97-118頁、2007年1月。

「自治体の統計・調査・政策立案体制を問う」『都市問題』・後藤新平生誕一五〇周年記念 八月号特別増刊、112-118頁、2007年8月。

【所属】東京大学大学院情報学環・準教授

東京大学社会科学研究所・準教授（兼務）

【所属学会】日本政治学会、日本選挙学会、日本統計学会、Midwest Political Science Association