

## 大規模社会調査データの保存・公開・活用

——東京大学社会科学研究所社会調査・データアーカイブ研究センターと Social Science Japan データアーカイブ(SSJDA)の試み——

東京大学 石田 浩

社会調査データを収集・保存・公開するデータアーカイブやデータライブラリーは、実証的な社会科学的研究にとって、もはやなくてはならない重要な資源となっている。世界的にもデータアーカイブの普及と発展は目覚ましく、世界規模のデータアーカイブの国際組織（International Federation of Data Organizations）やヨーロッパ地域の連合体（Council of European Social Science Data Archives）などをはじめとして、国を超えた協力・連携が模索されている。

社会調査の個票データを保存・公開するデータアーカイブの最も大きな学術的意義は、既存データの二次分析を可能とし、すでに収集されたデータから新たな知見を導き出すことが可能となることである。このことは実証研究の「再現性」を担保することにも通じる。調査データの再利用は、類似の調査を実施する多額の経費と時間を節約するだけでなく、調査対象者への回答負担の軽減を生み出す。さらに新たに調査を企画するときには、調査票項目や調査のやり方について既存調査を参照することができ、調査の質が向上する可能性がある。類似の調査項目をいれることにより、過去の調査を用いた時系列分析や国際比較研究が容易になるという利点もある。このように電子化されたデータを保存・公開することは、社会調査を貴重な「公共財」あるいは「リサーチ・ヘリテージ」として認識し学術の発展のために共有していくことにつながる。

本報告では、東京大学社会科学研究所に設置された社会調査・データアーカイブ研究センターと Social Science Japan データアーカイブ(SSJDA)の活動が、大規模な社会調査の個票データを「リサーチ・ヘリテージ」として継承していく試みのひとつであることを紹介する。社会調査・データアーカイブ研究センターは、我が国における社会科学的な実証研究を支援することを目的に、1996年4月に日本社会研究情報センターとして設置され、2009年4月に現在の形に改組された。2010年からは文科省の共同利用・共同研究拠点のひとつとして認定され活動を継続している。組織的には、センター長を含む教授4名、准教授4名、助教4名（以上教員12名の内3名が兼任）、技術職員（データアーキビスト）1名、（非常勤を含む）事務職員7名で構成されている。4つの研究分野、(1)調査基盤研究分野（SSJDAの運営）、(2)社会調査研究分野（独自の社会調査の実施）、(3)計量社会研究分野（二次分析の普及と計量分析の教育）、(4)国際調査研究分野（国際連携の推進）により成り立っている。

社会調査・データアーカイブ研究センターは、SSJDAを設置し1998年4月から個票データの提供を行っており、現在おおよそ1500ほどのデータセットを保存・公開している。SSJDAは、調査を実施した機関や研究者から個票データの寄託を受け付け、寄託されたデータは、まず調査票・コードブック・一次分析結果との照合を行い、公開用の個票データの整備を行う。その過程で必要に応じて調査対象者や事業所が特定される恐れのあるときには、特定される危険性の高い識別情報の削除やトップコーディング・グルーピングなどの方法により秘匿処理を行っている。さらに個票データとは別に、データの概要や調査項目などのメタデータを検索できる形で提供することにより、利用者が調査データの内容を事前に把握し、最も分析に相応しい調査データを選ぶことができる工

夫がされている。利用については、研究利用と教育利用の2種類があり、学術目的の二次分析研究のために、研究者・教員の指導を受けた大学院生、学部生などに公開されている。データの提供については、利用頻度の高い調査データは、ウェブ上から直接ダウンロードできるSSJDirectのシステムを開発し、迅速な提供に努めている。データ利用後には、執筆した論文などを報告する利用報告書の提出が義務づけられている。

社会調査データの保存・公開を行うアーカイブ事業は、社会調査の個票データを「リサーチ・ヘリテージ」として継承していく重要な第一歩であるが、それだけでは十分とはいえない。社会調査・データアーカイブ研究センターでは、社会科学的な実証研究の発展に貢献するために、4つの役割を果たしている。その4つとは4研究分野と対応した（1）データの保存・公開（データアーカイブ）（2）社会調査データの創出、（3）二次分析の普及・促進、（4）国際的ネットワークの構築である。社会調査・データアーカイブ研究センターでは、SSJDAに寄託する新たな独自データの創出にも取り組んでおり、若年・壮年者を対象として毎年追跡する「働き方とライフスタイルの変化に関する全国調査」（Japanese Life Course Panel Survey）を実施している。データクリーニング・コーディングを丁寧に行う質の高い調査データを自ら創出し、公開する作業に取り組んでいる。

SSJDAから利用可能な調査データを用いた二次分析の普及・促進を図るための「二次分析研究会」を2000年度から毎年開催している。センターがテーマと使用データを決めて参加者を募る「参加者公募型二次分析研究会」と、応募者がテーマ、使用データ、そして参加者を決める「課題公募型二次分析」の2つを毎年公募し運営している。二次分析の分析手法を講義する「計量分析セミナー」も2007年度から毎年開催している。優秀な二次分析の論文を執筆した若手研究者とデータの寄託者を表彰する論文表彰・寄託者表彰事業にも取り組んでいる。さらに、データアーカイブの国際的な連携を推進するために、世界のデータアーカイブ機関と情報交換やワークショップの開催などを実施している。

散逸してしまう貴重なデータを収集・保存し、公開するデータアーカイブの活動は、社会調査を貴重な「リサーチ・ヘリテージ」として認識し学術の発展のために共有していくために極めて重要である。しかしデータアーカイブに加えて、公開されているデータをよりよく活用するための二次分析の普及・促進の活動、独自の質の高い調査データの創出、そして国際的なネットワークの構築といった活動が同時に行われることで、「リサーチ・ヘリテージ」としての調査データが有効に活用され、学術の発展に貢献する優れた研究が生み出されてくる基盤をはじめて形成することができる。