

社会科学系データアーカイブの現状 SSJDA から

研究分担者	磯 博康	大阪大学大学院医学系研究科
	大橋康雄	中央大学理工学部人間総合理工学科
	祖父江友孝	大阪大学大学院医学系研究科
研究代表者	玉腰暁子	北海道大学大学院医学研究科
研究協力者	藤原 翔	東京大学社会科学研究所

研究要旨

データアーカイブ化に関し、社会科学系分野の現状と疫学研究に応用する場合の課題を把握・整理した。社会科学系分野では、既に多くの調査データがアーカイブ化され、利用されている。その促進のために二次利用教育も行われており、疫学研究のデータアーカイブを構築していくためには、社会科学系データアーカイブの運営システムから学ぶと同時に、利用のための環境整備も必要である。さらに、多くの社会科学系の調査データは、調査時点で個人情報を収集しないことも多く、対象者を追跡することが前提であるコホート研究とは、そもそもの配慮事項が異なる部分もある。特に、研究開始時点での対象者への説明のあり方、完全に連結不可能匿名化にするタイミングなど、今後の検討課題と考えられる。

A. 目的

社会科学系分野で進んでいるデータアーカイブ化の現状を把握し、疫学研究、とくに追跡が終了したコホート研究のデータアーカイブ化を進めるための参考とする。

B. 方法

社会科学系分野のデータアーカイブセンターの一つ、東京大学社会科学研究所の附属社会調査・データアーカイブ研究センターの藤原翔氏より、情報提供をいただいた。

C. 結果

[社会科学系データ]

社会調査とは、社会現象について知るために、一定

の科学的方法を用いて、現地においてデータの収集・分析をする過程と定義される。政治学、経済学、経営学、社会学、教育学、法律、法学調査等がその範囲となり、調査目的により、家庭、職業、教育、収入、疾病等、様々な変数が含まれる。

社会科学系調査では、調査時点で個人情報を収集しないことも多い。そのため、厳密には個人情報には当たらないと考えられるが、特殊な状況下では地域・属性などが分かる事で対象者が絞り込まれる(例えば、ある高校のある年時の卒業生で身長が 190cm、ある地域で子どもが 6 人、など)こともあり、情報の取り扱いには注意が求められる。

[社会科学系データアーカイブの意義]

社会科学系では量的、質的データのいずれもデー

データアーカイブに寄託されており、二次利用者がオリジナルな枠組みで分析を行い、新たな知見を出していくということがより一般的になってきている。その背景には、データアーカイブセンターが設立されたこと、ならびに二次分析のメリットが広く研究者に認識されたことがある。したがって、データアーカイブセンターの意義は、統計調査、社会調査の個票データを収集・保管し、その散逸を防ぐとともに、学術目的での二次的な利用のために提供することにある。このようにデータが収集・公開され、第三者が分析することにより、データの再現性を確認することにつながる。また、特に公的資金が投入され実施された調査データに関しては、調査者個人のものではないという認識も広まりつつある。

一方で、データアーカイブを二次利用するメリットはいくつか挙げられる。第一に、既に行われている調査を繰り返さずに済むため、労力、資金とも無駄な投入を避けることができる。第二に、特に多くの変数を得るような調査では得られたすべての情報を調査者が解析することはないため、利用されていない変数について独自のアイデアで解析することで、新たな知見を得ることができる。第三に、若手研究者にとっては、自身で小規模な回収率の低い調査を行うことに比べ、質のよい調査データにアクセスできる。さらに、学生教育の際にも、実データを用いた教育を行うことができる。

[データアーカイブセンターの沿革と運営]

1996年5月、東京大学社会科学研究所の附属施設として、日本社会研究情報センターが設立された。1998年4月から、センターではSSJデータアーカイブ(Social Science Japan Data Archive)を運営し、統計調査、社会調査の調査個票データと調査方法等に関する情報を収集・保管し、学術目的での二次分析のために、学内外の教員、大学院生等に提供を開始した。同時に、二次分析普及のために研究会や計量的研究法に関するセミナーも実施された。2009年4月に社会調査・データアーカイブ研究センターと改組

され、現在にいたっている。

データアーカイブ二次利用促進と円滑化のために、さらに次のような活動も順次行われている。二次分析優秀論文表彰(2005年6月開始)、リモート集計システムの稼働(2005年10月開始)、SSJDA Direct(データダウンロードシステム)の稼働(2009年4月開始)、寄託者表彰(2010年2月開始)、NESSTARシステム(メタデータ閲覧・オンライン分析システム)の稼働(2014年1月開始)、SSJDA Direct(データダウンロードシステム)への完全移行(2014年2月)。

現在、データアーカイブセンターの活動として行われている業務の主なものは、データ寄託の依頼・受付、データ整理、データ秘匿処理、メタデータの作成、データ利用の受付・提供、リモート集計の提供、二次利用成果の公開、データ寄託者の表彰、二次利用促進と適切な解析のための研究会・セミナーの開催等多岐にわたる。

運営費用は文部科学省(2010年度より国立大学法人共同利用・共同研究拠点)、東京大学社会科学研究所から運営費、データアーカイブに関わる科学研究費で賄われている。

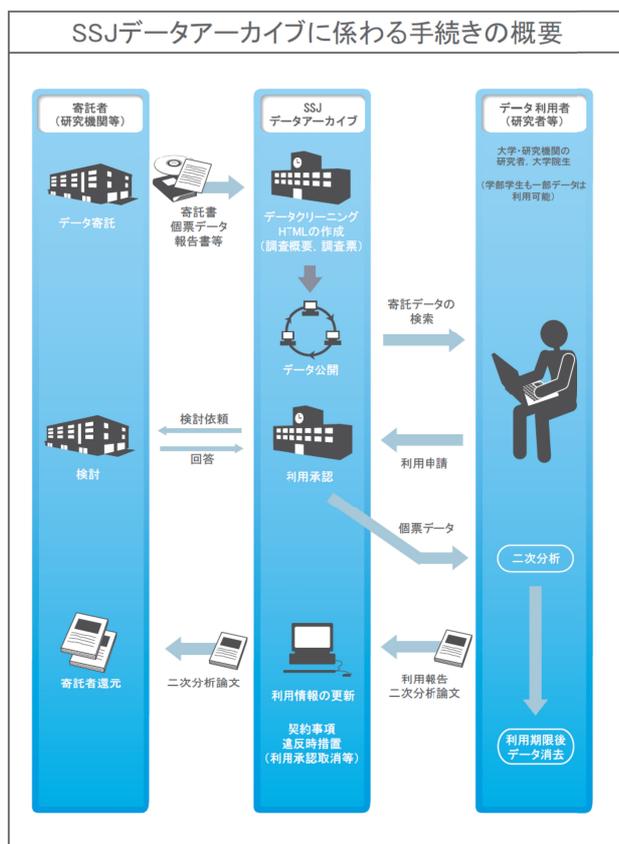
[他の社会科学系データアーカイブ]

国際的には、Roper Center(アメリカ)、ICPSR(アメリカミシガン大学)、GESIS(ドイツ)、UKDA(イギリス)等があげられる。国内では、RUDA(Rikkyo University Data Archive)、SORD(Social and Opinion Research Database、札幌学院大学、社会調査)、KUMA(神戸大学マイクロデータアーカイブ、公的統計)、Hi-Stat Social Science Database Network(一橋大学経済研究所、公的統計)、レヴァイアサン・データバンク(政治学系データ)等がある。

[データアーカイブ利活用の現状]

現在、SSJDAでは次の図のような仕組みで手続きが行われている

(<http://ssjda.iss.u-tokyo.ac.jp/2008image.pdf>)。



SSJDAには現在、約1600のデータセットが公開されており、2013年度は2700件の利用があった。

提供にあたっては、調査対象者個人が特定できるような情報は公開されないようセンターが秘匿処理を行っている。また、利用は学術目的のみで、利用者も研究者、大学院生、教員の指導を受けた大学生に限定される。そのうえで、利用者には、個々の回答者等が識別できる形式では発表しないことの誓約が義務付けられている。多くのデータの利用期間は1年(延長申請可能)であり、利用期限後は、個票データ消去が求められている。

D. 考察

社会科学系分野では、既に調査データのアーカイブ化とその利用の仕組みが整っており、多くの人々の協力により、うまく活用されているようである。しかし、データ寄託がある一定数に達し、利用のメリットが十分に浸透しなくては、労力のみが大きくなる懸念がある。したがって、疫学研究のデータアーカイブを構築していくためには、社会科学系データアーカイブの運

営システムから学ぶと同時に、利用のための環境整備も必要である。さらに、多くの社会科学系の調査データは、調査時点で個人情報収集しないことも多く、対象者を追跡することが前提であるコホート研究とは、そもそもの配慮事項が異なる部分もある。特に、研究開始時点で対象者への説明のあり方、完全に連結不可能匿名化するタイミングなど、今後の検討課題と考えられる。

E. 結論

東京大学社会科学研究所の附属社会調査・データアーカイブ研究センターにより運営されているSSJDAを例に、社会科学系分野のデータアーカイブ化の現状を把握し、疫学研究、とくに追跡が終了したコホート研究のデータアーカイブ化を進めるための知見を得た。

F. 健康機器情報

なし

G. 研究発表

1. 論文発表
なし
2. 学会発表
なし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし