

他国・他領域におけるデータアーカイブ利活用環境 米国におけるデータアーカイブの研究利用と 国内社会科学分野のデータアーカイブの現状 -

研究分担者	磯 博康	大阪大学大学院医学系研究科
研究分担者	大橋靖雄	東京大学大学院医学系研究科
研究分担者	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

研究要旨

疫学研究により得られたデータを広く共有化するためのシステムであるデータアーカイブ化に向けた課題を整理するために、先行している米国、国内社会科学分野の現状を把握し、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。

A. 研究目的

データアーカイブを二次利用するメリットは、

- 1) 既に行われている調査を繰り返さずに済み労力、資金とも無駄な投入を避けることができ、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こと、
- 2) 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、質の高い調査データにアクセスできること、
- 3) 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こと、である。

しかし、疫学研究のアーカイブデータの提供・利用環境はまだ整っていない。そこで、米国ならびに国内で先行している社会科学分野の事例と留意点を確認し、今後の方向性を考える一助とする。

B. 研究方法

実際に米国でデータベースを用いた研究に従事

されている Duke 大学瀬戸口聡子准教授より、情報を得た。

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. 研究結果

米国における二次データ利用の現状

米国では、電子カルテデータ、レセプトデータ、患者レジストリデータのみならず、コホート研究データについても公開が始まっている。これらのデータを用いることで、コホート研究を一から始めるのに比べ少ない労力でデータセットを作ることが可能である。また、NHLBI から研究費を受けた行われた RCT とコホート研究を含めた全てのデータを最大利用することを意図して、Biologic Specimen and Data Repository Information Coordinating Center を事務局としたデータデジタル化が行われている。データ提供にあたっては、インフォームド・コンセントに沿った形にするこ

と、個人同定情報は削除すること、地域情報は外すことなどがガイドラインで規定されている。このデポジトリでは、データあるいは生体試料二次利用希望者はHP上で、各研究の詳細を確認し、自身の用いたい研究に対しリクエストをすることができるが、その際、倫理審査を事前に受けておくなどの規則も定められている。このシステムにより、すでに他国からの申請も含め多くの研究が実施されている。

社会学分野のデータアーカイブの現状

社会科学系では、特に若手の研究者がデータアーカイブを利用して、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ことがより一般的になってきている。そのためのセンターのひとつであるSSJDA (Social Science Japan Data Archive)には現在、約1600件のデータが寄託されており、2013年度は2700件の利用があった。このように活用が進んでいる背景には、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。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は、データ解析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。

現在、SSJDAの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。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、多岐にわたっている。

D. 考察

国民の税金を投入し、多くの人手と長期の追跡を経て構築されたコホート研究データをアーカイ

ブ化、広く利用可能にすることは、研究の透明性確保、第三者による研究結果の検証、若手の育成に寄与するのみならず、研究の無駄・重複を減らし、必要な公費・労力を新しい有意義な研究に向けるという意義もある。そこで、日本よりデータベースの二次利用が進んでいる米国において、研究に用いられているデータベースの種類、活用事例とその成果、二次利用を進めるにあたっての留意点等について確認した。

日本では、二次データの公開システムとしてバイオサイエンスデータベースセンター(NBDC)が立ち上がったばかりであり、仕組みそのものに関する議論が十分に行われてはいない。今後、日本においても公的研究費を受けた研究を適切に二次利用することが求められているが、そのためには、個人情報保護と研究活用とのバランス、事務手続きの標準化・単純化と必要経費の設定、データの適正使用と質保障のためのサポート、共通化によるデータマニピュレーション、情報のロスとデータ容易使用のバランス、共通化プロセスの透明化と公正なシェアの仕組み等につき、議論を重ねていくことが必要と考えられた。

また、国内でも社会科学系分野のデータアーカイブは先行しており、そのためのセンターが設立され、現在では多くのデータが二次利用されている。しかし、今の形になるまでに、10~20年の年月を要しており、データ寄託がある一定数に達するまで、利用のメリットが十分に浸透するよう働きかけるとともに、利用のための環境整備も必要と考えられた。

E. 結論

疫学研究により得られたデータを広く共有化するためのシステムであるデータアーカイブ化に向けた課題を整理するために、先行している米国、国内社会科学分野の現状を把握し、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

特になし

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