

統計調査データ・アーカイブの意義

前田幸男 東京大学大学院情報学環・准教授
東京大学社会科学研究所附属日本社会研究情報センター・准教授（兼務）

1. 歴史的な前提—アメリカ合衆国センサスを例に

○19世紀は手集計

—原票を保存するが統計情報として重要なのは集計表であった。

○製表機とパンチ・カード導入の意味（1890年米国センサス）

- 原票とは異なる媒体に情報が転記される。
- 原票以外の記録が再度分析可能な価値ある対象として認識される。
- 実際には数億枚の紙製カードを保存することは困難であり、廃棄された。

○大型コンピュータによる集計を開始（1960年米国センサス）

- 磁気テープに転記された匿名化マイクロ・データの保存と公開が現実的に。
- 匿名化された標本データの二次利用が始まる。
- 1960年代に、大学においても学術調査データを保存・公開するための組織が徐々に作られる（西独 ZA 1960, 米 ICPSR1962, 英 UKDA 1967）。

2. 政府統計を用いた再分析の意味

○重要な政策課題について研究するためには過去のマイクロ・データが必要

- 例1 少子化 最低でも二時点におけるデータが必要。
コーホート（生別年代）の分析を行うためには適切な再分類が必要。
集計表からは就業状態や学歴別の傾向を明らかにすることは困難。
- 例2 格差の拡大 最低でも二時点におけるデータが必要。
一時点のデータからは分かるのは資産・所得の配分状態のみ。
過去との比較を通じてのみ格差の拡大・縮小を研究できる。

○政策の対象となる集団を理解するためには標本規模の大きい調査が必要

- 学術調査の標本規模は1000～4000程度。
年齢、性別、職業などにより分類した上で詳細な分析を行うことは事実上困難。
- 政府統計であれば大規模な匿名標本データを作成できる。

- 例 若年無業者 通常の学術調査では分析に耐えうる数を確保することは極めて困難。
内閣府の研究会では「就業構造基本調査」の特別集計を利用。
就業構造基本調査は約45万世帯が調査対象

3. 政府統計データ・アーカイブが必要な理由

○調査の企画・立案・実施と統計データの保存・公開とを分離する

調査の実施部門は標本の設計、調査票の検討、実査等で多忙であり、統計調査の企画・立案・実施にあたっている部門が片手間に磁気データの保存と公開活動を行うことは非現実的。

○磁気データの最終的保存

紙媒体の資料はその紙質に適切な環境（温度・湿度・明暗）を保って保存すれば良いが、磁気データの保存はそれだけでは不十分。コンピュータおよび磁気媒体というハード面での技術革新、およびソフトウェアの変化の両面に注意しながら、適宜保存形式を変化させる必要がある。また、磁気媒体の経年劣化への対応も必要となる。

○政府統計の専門家集団と学術目的利用者の橋渡し

政府統計の作成にあたっては独自のプログラムを作成しデータの管理および分析が行われている。それに対して、学術目的利用者は SPSS、SAS、Stata 等の市販されている統計ソフトウェアを用いる場合が大半である。政府統計部門内部で利用しているデータをそのまま秘匿処理を施した後公開しても実際に分析する能力を持つ利用者は極めて限られる。そもそも通常の利用者が分析できる形式にデータを編集し、必要なメタデータを整備する必要がある。

○統計利用促進活動

講習会等を開催することで正しい利用が促進される。学術目的利用者の全てが社会調査や統計理論に詳しいわけではない。多くの社会学者にとって方法は必要に応じて身につけるものであり、適切なデータの利用を促す必要がある。

4. アメリカ合衆国のデータ保存・公開活動の実態－センサス・マイクロ・データを例に

○注意すべき前提条件

最初から政府統計を総合的に管理するアーカイブが構想されたわけではなく、歴史的に 1 つ 1 つの小さな決定の積み重ねとして、現在のように様々な組織が緩やかに連携しつつ、データの保存と公開とを行う態勢ができあがった。日本の政府統計を保存・公開する仕組みとしてそのまま模倣するには前提が異なるように思われる。

○連邦政府によるデータの公開と保存

① 商務省センサス局

- －業務に支障がない範囲内で公開している。
- －ウェブ上での集計に関しては様々な工夫が施されている（American Fact Finder）。
- －マイクロ・データを利用する場合は技術的要求水準が高い。

例えば 2000 年センサスの 1%抽出データを利用するためには 722 頁あるマニュアルから必要な情報を拾い出し、データを読み込むところからはじめなければならない。

- 積極的な利用者は学術目的のデータ・アーカイブからデータを入手できない組織
- データの永年保存についてセンサス局は責任を負っていない。
- ウェブ上で公開されているセンサスのマイクロ・データも直近のものだけ。

②国立公文書記録管理局 (National Archive)

- 磁気ファイルを保存する最終的な責任は国立公文書記録管理局が持っている。
- ただし、国立公文書記録管理局から磁気データを入手することは一般的・実践的ではない。
- 磁気データだけではなく個票についても最終的な保存の責任を持つ。

○大学組織によるデータの保存と公開

①ICPSR Inter-University Consortium for Political and Social Research,

Institute of Social Research, The University of Michigan

- データ・アーカイブの活動の一環。
- 加盟校研究者に対する研究支援としてセンサス・データの保存・公開を行っている。
 - 80年代までは磁気媒体に保存されたセンサス・データを研究者個人が予算を獲得して購入し、利用するのは非現実的であった。ICPSR がメンバー校の研究者のために一括して購入し、利便性を高めた上で、再頒布していた。
- センサス局と密接に連携し、利用を促進するための活動を行っている。
 - ①ファイルの編集と利便性の向上 SPSS、SAS を読み込むためのファイルを付加
 - ②センサス・データの利用を促進する講習会の開催
 - ③センサス・データを利用した学術刊行物の文献一覧を整備
- 公開用マイクロ・データは 1960 年センサス以降を所蔵・公開

②IPUMS-USA The Integrated Public Use Microdata Series (IPUMS),

Minnesota Population Center, The University of Minnesota

- センサスのマイクロ・データを用いた研究の促進に特化した活動
- 時系列比較が可能なデータセットの作成
 - 10 年ごとのセンサスは職業や人種の分類が時代によって変更される。極力比較可能なかたちに変数を統一する作業を行い、時系列の比較を促進している。
- 永年保存されている歴史的な個票から公開用マイクロ標本データを遡及的に作成し公開。
- センサス・データを利用した学術刊行物の文献一覧を整備

5. その他のアメリカ政府統計データの学術利用について

ICPSR は Current Population Survey, Consumer Expenditure Survey についてもデータを保存・公開している。ただし、センサス局や労働統計局と特に協定があるわけでない。ICPSR はデータを購入した上で、整理・編集し、加盟校の利用者に提供している。利用者も、複数の政府サイトを探し回るよりも、ICPSR のウェブサイトで一括して検索することを望む。

ICPSR

the best source for **CENSUS**
2000

Census 2000 data are free to all

The decennial Census of the United States is perhaps the most important research resource for social scientists. ICPSR is acquiring the entire 2000 Census collection and distributing it free of charge to all users. This project, supported by the National Science Foundation and the National Institutes of Health, guarantees that users will have unlimited, open access to these data in a variety of formats and that the data will be preserved for future generations.

A wealth of new Census products coming soon

- Reformatted summary file data, reorganized with a single file per state
- Condensed versions of select summary files containing a limited set of key variables for research and instructional use
- PUMS subsamples designed to facilitate exploratory research and instruction on special samples of the population
- A data finder that will point researchers to the data that best fit their specific research needs, and custom subsetting capabilities to permit users to create and download their own files

Training and outreach

ICPSR plans workshops in 2002 and 2003 to introduce researchers to Census 2000 data and to provide hands-on analytic experience. Representatives from the Census Bureau, along with professional demographers and sociologists, will provide instruction in effective use of the data and their application to different types of research projects.

Several collections now available

ICPSR is now distributing the Redistricting Data, Summary File 1, and Summary File 2 free of charge to all users. New 2000 Census data, including Summary Files 3 and 4 and Public User Microdata Samples, will be acquired and made available as they are released from the Census Bureau.

 Visit the Census 2000 Web site at:
www.icpsr.umich.edu/CENSUS2000

▼ DATA

[Create an Extract](#)
[Download or Revise Extracts](#)
[Analyze data online](#)
[Register as a New User](#)

▼ DOCUMENTATION

[What is IPUMS?](#)
[User's Guide](#)
[Variables](#)
[Samples](#)

▼ RESOURCES

[Enumeration Forms](#)
[Published Census Volumes](#)
[Revision History](#)

▼ RESEARCH

[Citation and Use](#)
[Bibliography](#)
[Related Sites](#)

▼ CONTACT US

[Feedback](#)
[IPUMS Staff](#)
[How to Help](#)



Integrated Public Use Microdata Series

census microdata for social and economic research

IPUMS-USA is a project dedicated to collecting and distributing United States census data. Its goals are to:

- Collect and preserve data and documentation
- Harmonize data
- Disseminate the data absolutely free!

Use it for GOOD -- never for EVIL

▼ IPUMS-USA NEWS

[PLEASE HELP: Save the ATUS!](#)
[Data release schedule](#)

▼ OTHER MPC PROJECTS

[IPUMS-International](#)
[IPUMS-CPS](#)
[NAPP](#)
[NHGIS](#)
[IHIS](#)

Funded by

