

社会調査個票デジタルデータの 収集、保存、二次分析について

前田幸男

東京大学大学院情報学環

東京大学社会科学研究所

附属社会調査データアーカイブ研究センター

本日のチュートリアルの概要

- 個人的回想
- 諸外国の社会科学調査データのアーカイブ（レポジトリ）
- データのライフサイクルと社会調査データの保存・共有
- 歴史的な経緯 - アメリカのセンサス・データを例に
- 数値データ、メタデータ、統計ソフトウェア
- 日本におけるデータ・ポリシー
- 東京大学社会科学研究所 **SSJDA** の活動紹介
- データの保存と公開を担う組織および専門家について

個人的な回想をお許し頂ければ...

- 大学院生時代（1993～2001）、データをインターネット上から直接ダウンロードしたことはなかった。
- 統計ソフトウェアで日本語を扱うことができなかった。
- メタデータという言葉は聞いたことがなかった。
- 印刷・製本されたコードブックを傍らにおいて、データの分析をしていた。
- 図書館情報学の専門家や古文書写真のデジタルファイルを保存する活動・研究をしている方々と会話をする機会はなかった。
 - そんな機会が訪れるとも考えなかった。
 - 突然、海外の社会科学データ・アーカイブの知人から電子資料の保存に関する国際学会の企画委員を頼まれる...なんで私に???

欧米の社会調査・統計調査データアーカイブ（例）

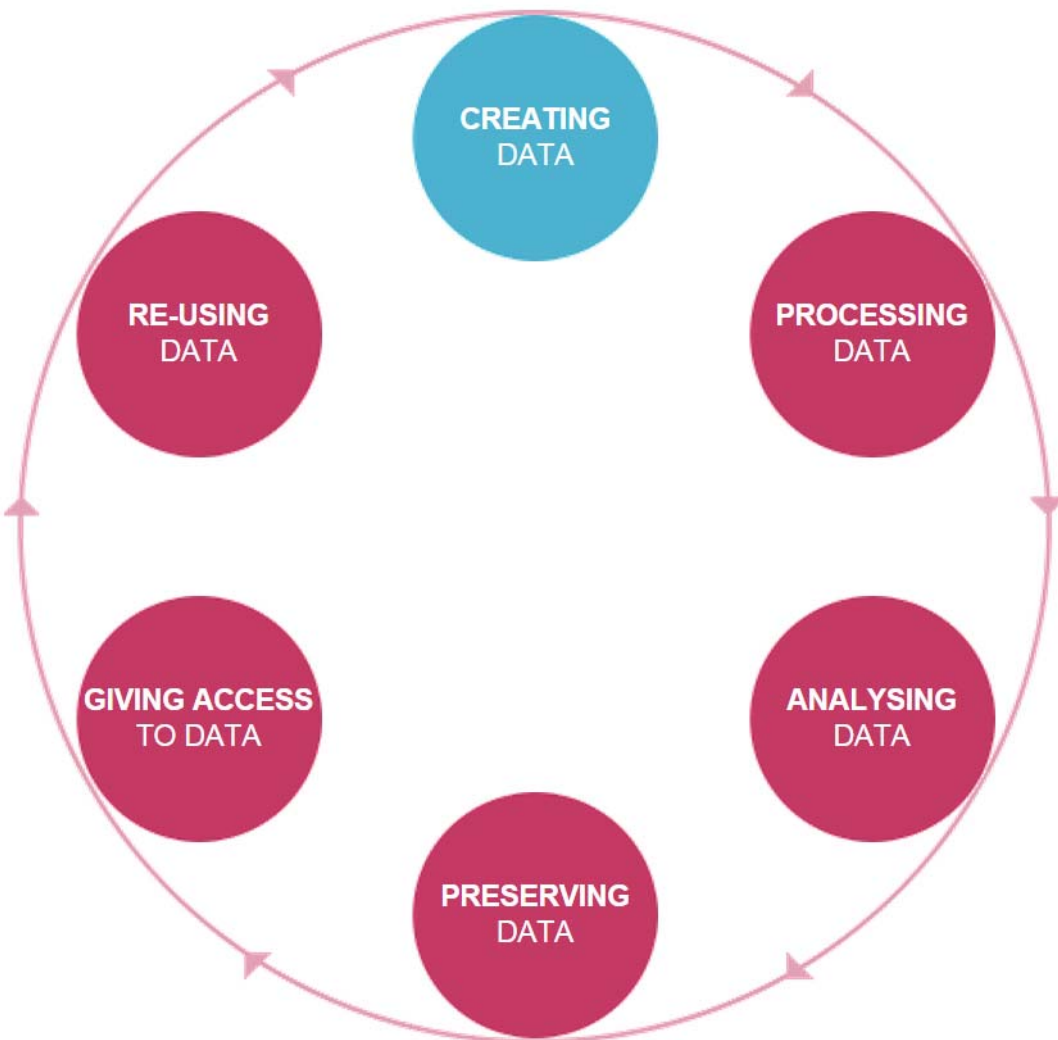
- アメリカ（1962） Inter-University Consortium of Political and Social Research (ICPSR) <https://www.icpsr.umich.edu>
- ドイツ（1960/1986） Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) <https://www.gesis.org/en/home/>
- オランダ（1964/2005） Data Archiving and Networked Services (DANS) <https://dans.knaw.nl/en>
- イギリス（1967） UK Data Archive (UKDA) <http://www.data-archive.ac.uk/>
- ノルウェー（1971） NSD - Norwegian Centre for Research Data <http://www.nsd.uib.no/nsd/english/>
- フィンランド（1999） Finnish Social Science Data Archive – FSD <http://www.fsd.uta.fi/en/>

東アジアのデータ・ アーカイブ（例）

- 台湾（1994）Survey Research Data Archive (SRDA) – 中央研究院
<https://srda.sinica.edu.tw/>
- 韓国（1983/2003/2015）Korea Social Science Data Archive (KOSSDA) ソウル大学
アジアセンター
http://snuac.snu.ac.kr/center_eng/
- 日本（1996/1998）Social Science Japan Data Archive (SSJDA)
東京大学社会科学研究所
<http://csrda.iss.u-tokyo.ac.jp/>

- 列挙した組織の形態、予算、規模は様々である。
- 立地は大学外か、大学隣接か、大学内の3パターン。
 - ドイツのGESISは連邦・州政府からの予算で運営されている。
 - 欧州諸国の場合、国からの直接的予算措置があるのが一般的。
 - アメリカのICPSRは主に大学を加盟組織とするコンソーシアムで、政府レベルの支援はない。
 - 日本のアーカイブは大学の一部門に過ぎず、脆弱である。
- 扱うデータの種類も違いがある。
 - 政府統計を扱う場合と学術データのみの場合
 - 質的データも扱う場合と、量的データのみの場合

データの ライフサイクル



- データの収集
- データの処理とクリーニング
- データ分析
- データの保存
- データの共有・公開
- データの二次分析

何故、社会調査データを保存するのか？（１）

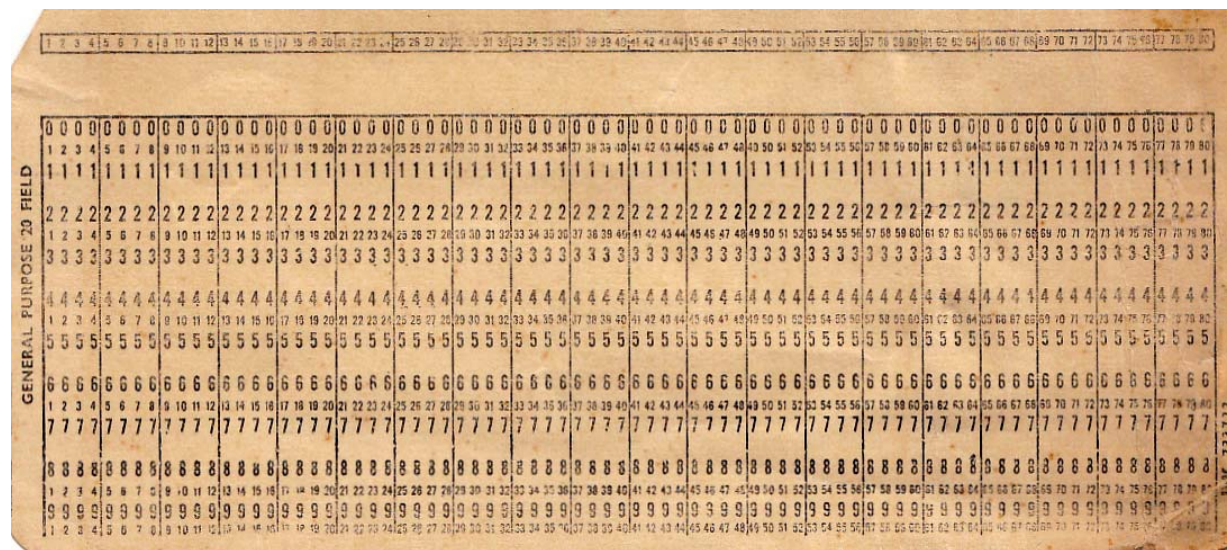
- 公文書や書簡が重要な歴史的な資料であるように、確率抽出された標本調査個票データは、その時代の社会の姿を記録する重要な資料である。
 - 行政記録の集計結果から得られる情報は限られている。
 - データが蓄積することで、可能になる分析がある。例えば、無党派層の増加や少子化・晩婚化というテーマは、複数時点のデータがなければ、検証不能である。
- 統計表・集計結果だけの保存では不十分。
 - 統計表や集計結果は、参照が容易であり、政策決定の材料、あるいは、国民一般への情報提供としては有益である。
 - しかし、研究上の関心、あるいは、行政上の目的から異なる集計・分析を試みるためには、個票データが利用可能な形態で保存されている必要がある。

何故、社会調査データを保存するのか？（2）

- 大規模標本調査には費用がかかる
 - 計画標本**3000**人、回収率**50%**、回答者数**1500**人程度の選挙調査で最低**2000**万円程度の予算が必要。これは一時点の横断面調査の金額。
 - 継続的に観察するパネル調査、親と子を調査する場合などは、更に費用が膨らむ。大規模調査の科研費の申請は研究期間を通算して億単位になることも珍しくない。
- 1つの社会調査における質問数（変数）は百問を超えることも稀ではなく、異なる観点からの分析（二次分析）に耐えうる。
 - 費用および回答者への負担を考えると、既存データで分析可能ならば、新しくデータを収集する必要はない。
 - 若手研究者が独力で全国大の標本調査を実施することは不可能であり、データの共有が大学院教育・研究上不可欠である。

歴史的な経緯

アメリカのセンサス・データを例に



By Ventriloquist - Own work, CC BY-SA 3.0,
<https://en.wikipedia.org/w/index.php?curid=32753387>

米国センサス個票データを例に

- センサスは憲法上の要請として**10年**に一度行われている。
- **19世紀**には様々な集計方法（計算の道具や機械）が考案された。
- 一つの画期が、**1890年**センサスから導入された **Herman Hollerith** の製表機（**Tabulating Machine**）であった。
 - センサスの調査票情報をパンチカードに（穴を穿つことで）転記し、機械的な集計を可能にした。



パンチカードは保存されたのか？

- パンチカードを保存するか否かは激論の対象となった。
- すでに、センサス・スケジュール（記入済み調査票）自体は永年保存の対象。
- 20世紀前半には研究者がデータを再分析できるとは思えなかったであろう。
- むしろ、集計結果・統計表の保存が重要だと考えられた。
 - パンチカードを保存するか否かは担当部局の判断に委ねられた。
 - パンチカードの劣化も早かった。
 - 結果として多くのパンチカードが廃棄された。
- ただし、パンチカードは、他の文書と同様の公文書として理解された点は重要。



センサスにおける大型計算機の利用

- UNIVACは1951年に導入され、1950年センサスの集計に部分的に利用された。磁気テープを利用したデータおよび計算結果の保存が可能に。
 - Hollerithの製表機は過去の遺物に。
- 1960年センサスからは完全にUNIVACによる集計が行われた。
- 磁気テープは保存されたのか？



磁気テープの扱い



By Daniel P. B. Smith. - Image by Daniel P. B. Smith., CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=7097264>

- パンチカードの保存は、現場の判断に委ねられていたが、その慣行が磁気テープにも引き継がれた。
- 磁気テープは高価であり、上書きされ再利用されていた。
 - NHKのテレビ放送番組のテープは放送開始から1980年頃まではほとんど保存されていない。
- ただし、1960年代以降は、民生用・学術用大型計算機の普及と相まって、統計データの二次分析が可能になった。
- 1960年代から70年代にかけて、政府統計データの磁気テープがNARAにおける保存の対象になった。
- また、学術調査の世界でも1960年代初頭にはICPSR (ICPR) のようなコンソーシアムが結成され、パンチカード形式でデータを共有するという慣行が徐々に普及していった。

NARAが電子データ（政府統計の個票データ）についても保存と提供の責任を負っている。



NATIONAL ARCHIVES

[Blogs](#) · [Bookmark](#)

Search Archives.gov

RESEARCH OUR RECORDS

VETERANS' SERVICE RECORDS

EDUCATOR RESOURCES

VISIT US

AMERICA'S FOUNDATIONS

Research Our Records

Home > Research Our Records > Census Records > Statistics from 1990 Census of Population and Housing

Census

[About Census Records](#)

[1940 Census](#)

[1930 Census](#)

[1850-1930 Census Clues](#)

[1790-1840 Census Clues](#)

[Nonpopulation Census](#)

[1935 Business Census](#)

[Indian Census Rolls](#)

Statistics from 1990 Census of Population and Housing

Electronic Records Reference Report

[Introduction](#)

[Summary Statistic Files](#)

[Public Use Microdata Samples \(PUMS\)](#)

[Geographic Reference Files](#)

[Special Tabulations with 1990 Census Data](#)



- NARA - The U.S. National Archives and Record Administration¹⁴

ICPSRからもセンサスデータは入手可能

 Find & Analyze Data

Log In/Create Account

FIND DATASEARCH/COMPARE VARIABLESDATA-RELATED PUBLICATIONSRESOURCES FOR STUDENTSHelp

Quick Download ▾
(5 datasets; 2,580,906 KB)

Table of Contents

- Access Notes
- Dataset(s)
- Study Description
 - Citation
 - Funding
 - Scope of Study
 - Methodology
 - Version(s)
- Related Publications
- Variables
- Utilities
 - Metadata Exports
 - Download Statistics

Census of Population and Housing 1970, 1980, 1990, and 2000 [United States]: Public Use Microdata Sample: Artist Extract Files (ICPSR 35534)

Principal Investigator(s): ⓘ United States Department of Commerce, Bureau of the Census; National Endowment for the Arts

Summary: ⓘ

The Research Division of the National Endowment for the Arts has prepared Artist Extracts from the Census Public Use Microdata Samples (PUMS) from 1970, 1980, 1990, and 2000. The extracts are intended to reduce the file size and make it easier to use census data to conduct research on artist occupations. The files contain records of all persons in households with one or more persons having a detailed occupation code considered "artist" by NEA. These occupations include: actors and directors; announcers; architects; authors; ...[more info](#)

NADAC

This study is maintained and distributed by the [National Archive of Data on Arts & Culture](#) (NADAC). NADAC is supported by the [National Endowment for the Arts](#) ↗.

Access Notes

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.

Dataset(s)

政府統計と学術社会調査データの相違

- 政府統計は、国によって法的な位置づけが異なる。
- アメリカの統計データは公文書一般と同じ枠組みで保存・公開される。
- 日本は統計は統計法、公文書は公文書管理法と別の体系
- 政府統計の個票データは、利用の手続きや制約は厳しい。
- 調査対象者数が多い。例えば、就業構造基本調査（2007）は約45万世帯に対して調査を実施。
- 学術社会調査はアカデミック・ルールに従う。
- 国によって状況は異なるが、資金提供機関（NSF等）が、データの共有・公開を義務づけることも増えている。
- 古いデータは散逸してしまったものも多い。保存は研究者個人とその後継者の努力に依存していた。
- 日本で本格的に学術社会調査データの保存と共有が行われるようになったのは1990年年代から。

日本における社会調査データ提供機関

- レヴィアアサン・データ・バンク
 - 1991年より活動 木鐸社の活動。
 - 選挙調査データを2～5万円で販売
- SSJDA (1996) 後の詳述
- RUDA (2011) 立教大学のデータ・アーカイブ 小規模だが、先駆的な取り組み。
- 一橋大学経済研究所 [社会科学統計情報研究センター](#)
- 独立行政法人 [統計センター](#)

会場では写真を投影します。

数値データ、メタデータ、統計ソフトウェア

メタデータとコードブック

- データを共有するためには、数値以外の付随情報（メタデータ）の共有が必要になる。
- メタデータとしては、例えば
 - 研究代表者
 - 調査実施機関
 - 母集団とサンプリングの方法
 - 調査実施時期
 - 調査の詳細（回収率）
 - 調査員の手引き書
 - 調査票/回答票
 - 自由回答の分類（コード）
- 以前は、印刷されたコードブックが必須であった。
- 21世紀に入ってから、多くの場合、PDF版のコードブックは作られている。
- 統計ソフトウェアの能力が向上することによって、以前と比べるとデータとメタデータの連係が強化された。
- メタデータをXMLの統一規格で記述することにより、高度な利用が可能になりつつある。
 - DDI – Data Documentation Initiative

データファイルの具体例 LRECL 1行1367列

[illegible]

S P S S によるデータ読み込みの例 単純なFlat Fileの場合

```
FILE HANDLE DATA /  
NAME="C:¥DATA¥07294-  
0001-Data.txt"  
LRECL=1367 .  
DATA LIST FILE=DATA /  
V1 1-4  
V2 5-8  
V3 9-10  
V4 11-12  
...  
V63 158-159  
V64 160-160  
V65 161-161  
...
```

VARIABLE LABELS

```
V1 "ICPSR調査番号 (7294)"  
V2 "面接番号"  
...  
V64 "16(a) 今の日本の憲法は変えた方が良く  
    と思いますか。それとも今のままで  
    良いと思いますか。"
```

```
V65 "16(b) (16(a)で「変えた方が良く  
    と思いますか。それとも多少変える程  
    度で良いと思いますか。"
```

```
...  
VALUE LABELS V64 1"変える" 5"今のま  
    ま" 7"その他" 8"わからない" 9"答  
    えなし" 0"非該当".  
VALUE LABELS V65 1"大幅" 5"多少" 7"そ  
    の他" 8"わからない" 9"答えなし" 0  
    "非該当".
```

メタデータの有無による情報量の違い

統計

数値データとメタデータが関連付けられた場合の出力

16(a) 今の日本の憲法は変えた方がよいと思いますか。それとも今のままでよいと思いますか。

度数	有効数	2051
	欠損値	72

16(a) 今の日本の憲法は変えた方がよいと思いますか。それとも今のままでよいと思いますか。

	度数	パーセント	有効パーセント	累積パーセント
有効数				
変える	345	16.3	16.8	16.8
今のまま	951	44.8	46.4	63.2
その他	82	3.8	4.0	67.2
わからない	655	30.9	32.0	99.1
答えなし	18	.9	.9	100.0
合計	2051	96.6	100.0	
欠損値	72	3.4		
合計	2122	100.0		

統計

数値データのみの場合の出力

V64

度数	有効数	2122
	欠損値	0

0を集計に含めて良いのかこれでは不明

V64

	度数	パーセント	有効パーセント	累積パーセント
有効数				
0	72	3.4	3.4	3.4
1	345	16.3	16.3	19.6
5	951	44.8	44.8	64.4
7	82	3.8	3.8	68.3
8	655	30.9	30.9	99.1
9	18	.9	.9	100.0
合計	2122	100.0	100.0	

ソフトウェア（SPSS）に読み込まれ、メタデータと関連付けられたデータ

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	7294	1	10	1	1	1	1	10	3	1	2	4	4	1
2	7294	2	10	1	1	2	1	10	3	1	2	4	4	1
3	7294	3	10	1	1	3	1	10	3	1	2	4	4	1
4	7294	4	10	1	1	4	1	10	3	1	2	4	4	1
5	7294	5	10	1	1	5	1	10	3	1	2	4	4	1
6	7294	6	10	1	1	6	1	10	3	1	2	4	4	1
7	7294	7	10	1	1	7	1	10	3	1	2	4	4	1
8	7294	9	10	1	1	9	1	10	3	1	2	4	4	1
9	7294	10	10	1	1	10	1	10	3	1	2	4	4	1
10	7294	11	10	1	1	11	1	10	3	1	2	4	4	1
11	7294	12	10	1	2	1	1	10	3	2	5	1	1	1
12	7294	13	10	1	2	2	1	10	3	2	5	1	1	1
13	7294	15	10	1	2	4	1	10	3	2	5	1	1	1
14	7294	16	10	1	2	5	1	10	3	2	5	1	1	1
15	7294	17	10	1	2	6	1	10	3	2	5	1	1	1
16	7294	18	10	1	2	7	1	10	3	2	5	1	1	1
17	7294	20	10	1	2	9	1	10	3	2	5	1	1	1
18	7294	21	10	1	2	10	1	10	3	2	5	1	1	1
19	7294	23	10	1	3	1	1	10	3	3	3	1	1	1
20	7294	25	10	1	3	3	1	10	3	3	3	1	1	1
21	7294	26	10	1	3	4	1	10	3	3	3	1	1	1
22	7294	28	10	1	3	6	1	10	3	3	3	1	1	1

	名前	型	幅	小数桁数	ラベル	値
67	V51	数値	2	0	11(a) 政治家...	{0, 非該当}...
68	V52	数値	2	0	11(b) (11(a)で...	{0, 非該当}...
69	V53	数値	2	0	11(b) 政治家...	{0, 非該当}...
70	V54a	数値	2	0	12(a) あなた...	{0, それ以上回...
71	V54b	数値	2	0	12(a) あなた...	{0, それ以上回...
72	V54c	数値	2	0	12(a) あなた...	{0, それ以上回...
73	V54d	数値	2	0	12(a) あなた...	{0, それ以上回...
74	V54e	数値	2	0	12(a) あなた...	{0, それ以上回...
75	V55	数値	1	0	12(c) (第一問...	{0, 非該当}...
76	V56	数値	1	0	12(e) (第二問...	{0, 非該当}...
77	V57	数値	1	0	12(g) (第三問...	{0, 非該当}...
78	V58	数値	1	0	これから最近...	{0, 非該当}...
79	V59	数値	1	0	14 税金を減...	{0, 非該当}...
80	V60	数値	1	0	15(a, b) ベト...	{0, 15(a)に関...
81	V61	数値	1	0	15(a) ベトナ...	{0, 非該当}...
82	V62	数値	1	0	15(b) (15(a)で...	{0, 非該当}...
83	V63	数値	2	0	16(a, b, c) 憲...	{0, 16(a)非該...
84	V64	数値	1	0	16(a) 今日の日...	{0, 非該当}...
85	V65	数値	1	0	16(b) (16(a)で...	{0, 非該当}...
86	V66	数値	1	0	16(c) (16(a)で...	{0, 非該当}...
87	V67	数値	2	0	17(a, b) 自衛...	{0, 17(a)非該...
88	V68	数値	1	0	17(a) 自衛隊...	{0, 非該当}...
89	V69	数値	1	0	17(b) (17(a)で...	{0, 非該当}...

ソフトウェアに依存することの長所と短所

- メタデータとデータが関連付けられることで、分析作業は格段に円滑に行うことができる。
 - いちいち、コードブックや調査票を確認せずにすむ。
 - 非該当をデータファイル上に維持したまま、集計からだけ除外できる。
- 集計結果（出力）を最低限の加工で、論文等に挿入できる。
- 例えば、複数のデータセットを一括して、変数ラベルを検索することも可能。

- データとメタデータの関連付けはソフトウェアに依存しており、かつ、ソフトウェア間の互換性は保証されていない。
- 同じソフトウェアでも古いバージョンのデータが必ず読めるという保証はない。
- 商用ソフトの場合は、中の仕組みは分からない。

Ex. “These data definition statements have been tested for compatibility with SPSS Release 6.13 for UNIX and/or SPSS Release 9.0 for Windows.”

統計ソフトウェアが数値以外の情報を扱う能力の変化

- かつては、アルファベット以外の文字は受け付けられなかった。
- かつては、変数ラベルの長さはSPSS v7.5 で、半角120文字以内限定されていた。
- SPSS v7.5 で、値ラベルの長さは半角60文字以内。
- 変数ラベルと変数値ラベル以外の情報を扱うことは少なかった。
- 現在は日本語等も受け付けられる。
 - UNICODE対応。
- SPSS v.22 変数ラベルの長さは、255バイト以内。
- SPSS v.22 値ラベルの長さは120バイト以内。
- 一般のソフトウェアでデータセット内の複雑な構造を記録・記述できる。
 - マルチレベルデータ、時系列データ、サンプリング・デザインなど

データの長期的保存と利用について (1)

Emulation

- 古いソフトウェアでデータを読み込み分析する環境を最新のコンピュータ上で再現する。
- データは古い形式のまま保存できる。
- データ・レスキューには有効な方法だが、利用上の制約が大きいのではないか。

Migration

- データの保存形式を常に最新の形式に保つことで（新しいバージョンに対応すること）、利用可能性を担保する。
- 利用者側から見るとありがたいが、全てのソフトウェアに対応できるわけではない。
 - 統計ソフトウェアの数も多い。
- 対応できない場合は、結局 Emulation に頼ることになる。

データの長期的保存と利用について (2)

ソフトウェアに依存しない形式で保存

- 統計ソフトウェア上のファイルを、**LRECL**形式に変換すると同時に、関連するメタデータを、**ASCII**形式で保存する。
- データとメタデータの関連付けは失われる。
- 保存にも、利用にも一定の作業、時間が必要となる。
- ただし、将来的に別のソフトウェアで読める確率はかなり高い。

磁気媒体自体の劣化防止・更新

- パンチカード
- 磁気テープ
- フロッピーディスク 5インチ
- フロッピーディスク 3.5インチ
- CD-ROM
- 内蔵／外付けハードディスク
- ZIPディスク
- MO
- USB
- DVD

データの長期的保存と利用について (3)

- 主要ソフトウェアについては、アップグレード時に互換性の確認が必要。
- 保存媒体の経年劣化を念頭に、定期的に新しい媒体に移すことが必要。
- 紙やモノと異なり、デジタルデータは瞬時に消滅することもありえる。
 - デジタルデータのレポジトリには、レポジトリの信頼性を担保するために一定の認証基準が作られている。
 - 社会調査データのアーカイブについてはData Seal of Approval (DSA) が有名。 <https://www.datasealofapproval.org/en/>
 - DSA は World Data System (WDS) と [認証基準の共通化](#)を行っている。

日本におけるデータ・ポリシー

統計調査・社会調査を巡る国内状況

- OECDは公的資金による収集されたデータは公開することを勧めている。

[“Principles and Guidelines for Access to Research Data from Public Funding”, April 2007.](#)

- 日本では文科省や学術会議がデータを二次利用のために公開することを義務づけてはいない。
- 日本の社会科学系の学会でデータの公開を義務づけたり、論文投稿時にデータの提出を求めている例はないと思われる。
- 研究者が個人的努力としてウェブ上でデータを公開することは増えている。ただし、長期的にデータが保存・提供されるか否かは不透明。

研究不正への対応としてのデータ保存

- 文部科学省は「研究活動における不正行為への対応等に関するガイドライン」を公表（2014年8月）。
- 「研究機関において、研究者に対して一定期間研究データを保存し、必要な場合に開示することを義務付ける旨の規程を設け、その適切かつ実効的な運用を行うことが必要である。」
 - これは、不正防止を念頭に置いており、データのライフサイクルを念頭に、研究データの共有・再分析を推奨するものではない。
- 「なお、保存又は開示するべき研究データの具体的な内容やその期間、方法、開示する相手先については、データの性質や研究分野の特性等を踏まえることが適切である。」
 - データ保存のフォーマットや必要なメタデータについての記述もない。

DMP Data Management Plan 国外

- DMPとは？

“A data management plan is a document that outlines a researcher’s management strategies and policies for data produced by his or her research project.” <http://guides.lib.umich.edu/socialsciences-dmp>

- 諸外国では資金提供機関がDMPを要求することが多い。

- アメリカの[National Science Foundation](#)

- ミネソタ大学図書館ウェブサイトに連邦機関の[一覧](#)が掲載してある。

- フィンランドの[Academy of Finland](#)

- イギリスの[Economic and Social Research Council](#)

- オーストラリアの[Australian Research Council](#)

- DMPの中にデータの保存や公開についての記述を求められる。

DMP Data Management Plan 国内状況

- 科学技術振興機構（JST）は2017年4月に「『オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針』運用ガイドライン」においてDMPを義務づけた。
- JSTによれば、DMPは、「研究者が自身で実施する研究プロジェクトにおいて研究成果として生じる研究データの取扱いを定めるものであり、具体的には、当該研究データの保存・管理、公開・非公開、公開範囲等に関する方針や計画について記載したものを指す」。
- 日本学術振興会は現在DPMの導入を検討中のようにあ

DMPの重要性

- 公的資金から得られたデータは公開するという世界的潮流。
- 研究者にとっては、データの保存・公開のための作業は直接研究には結びつかない「無駄に見える」作業。
- 必要な経費や人件費を研究計画の中に組み込む方が良い。
- データが公開されることを前提に、計画およびデータ収集の各段階で必要なメタデータを整理・保存するほうが格段に効率的。
- より多くの調査データが広く保存され、二次分析に供されるためにDMPは重要。

東京大学社会科学研究所 附属社会調査データアーカイブ研究センターの活動紹介



センター	SSJデータアーカイブ（調査基盤）			社会調査		計量社会		国際調査
センターについて	SSJDAについて	データの寄託	データの利用	社研パネル調査	共同調査と共同研究	二次分析研究会	計量分析セミナー	国際連携

検索・分析・ダウンロード

- ❖ データ検索システム
- ❖ データの利用申請(SSJDA Direct)
- ❖ リモート集計システム (Nesstar)

Center for Social Research and Data Archives



❖ 社会調査・データアーカイブ研究センターは、実証的社会科学研究の発展、社会調査・データアーカイブネットワークの構築に、主導的な役割を果たしていきます。

❖ 【SSJDAのデータを初めてご利用される方】は[こちら](#)。

センターからのお知らせ

一覧

- ・ [【9/20】リモート集計システム \(Nesstar\) で利用可能なデータが増えました。](#)
- ・ [国際ワークショップ\(2017年9月7日開催\):「データ・ライブラリアンシップ:アメリカ大学図書館におけるライブラリアンの役割変化について」のご案内を公開しました。](#)

SSJDAデータ公開情報

一覧

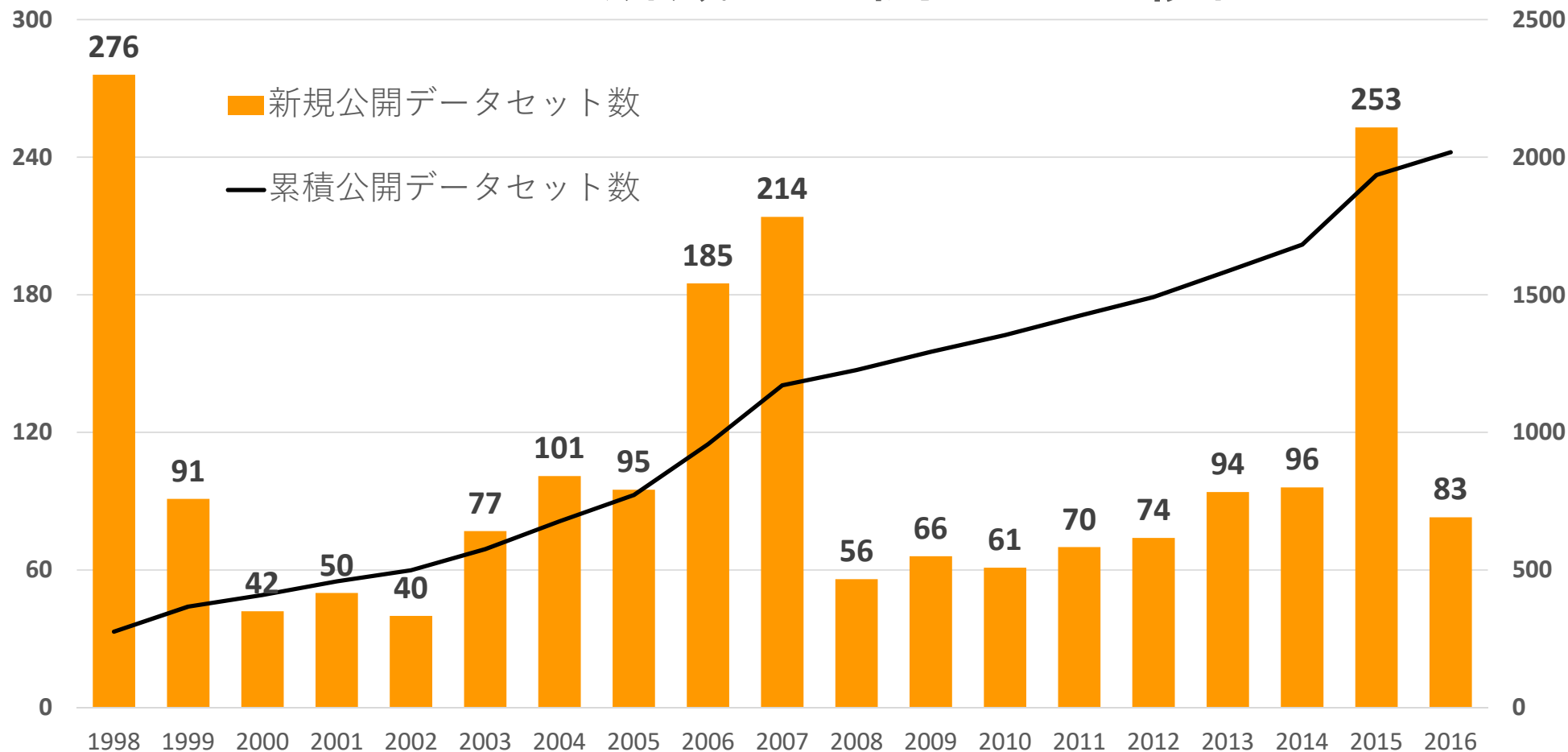
2017-08-23

- 0978「子ども・子育てに関するアンケート (SRC自主調査002) (サーベイリサーチセンター) を公開しました。
- 1084「介護の実態とニーズ調査, 2012」(ライフケアパートナーズ) を公開しました。

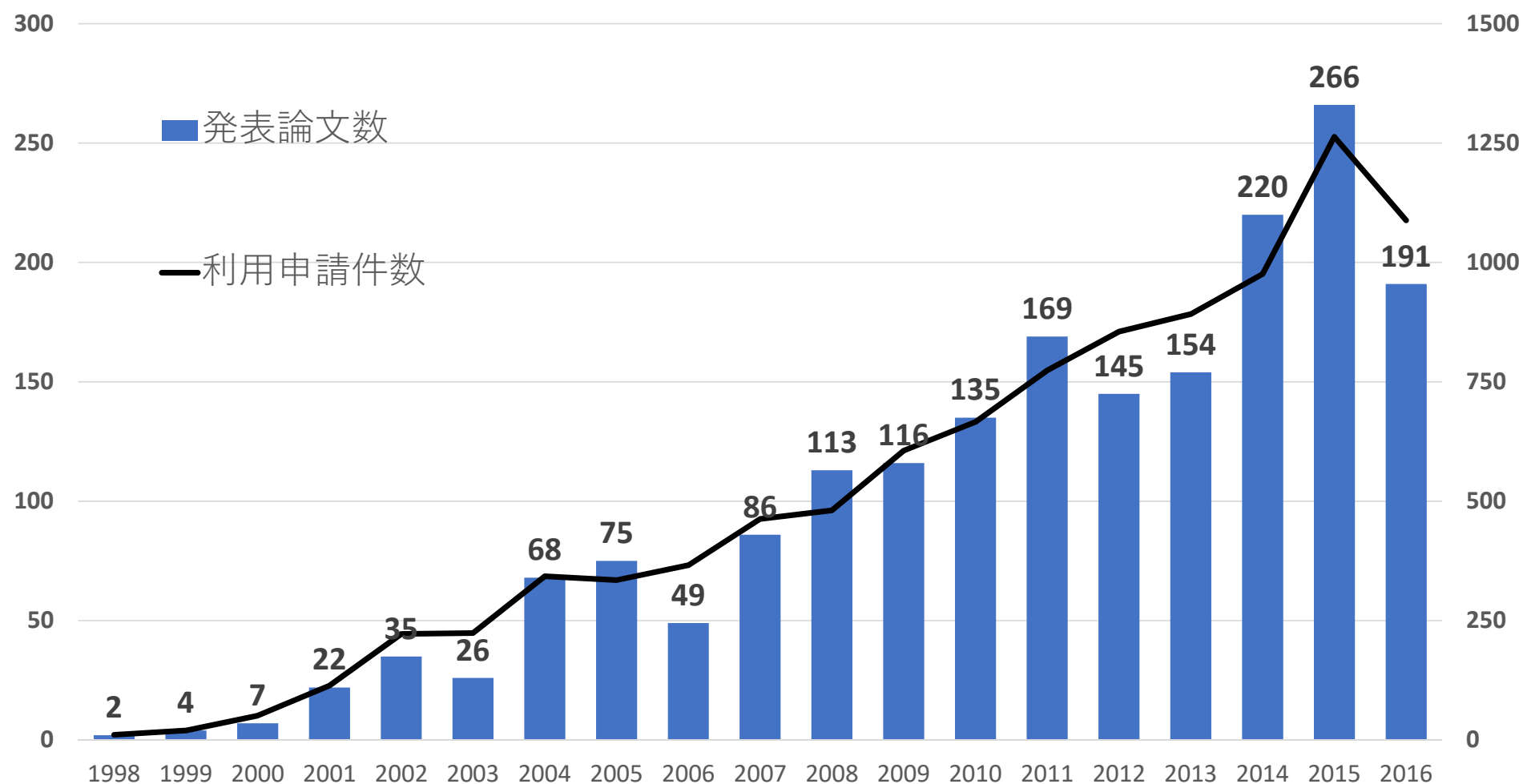
S S J D A で収集・保存・提供している データ

- 研究者が（科研費などで）実施した標本調査データ、各省庁が行った世論調査、労働組合が行った調査、民間企業の調査部門が行った調査など
- 研究者や調査機関にデータの寄託依頼をしている。
- 受け入れているデータのほとんどが、確率抽出された標本調査のデータである。
 - 研究上の価値があると判断すれば、インターネット調査等からのデータも受け入れは行っている。
- 集計データ（統計表）、質的データ等は対象外
- 政府統計を扱う予定はない。

SSJDAの所蔵・公開データ統計



利用申請件数と発表論文数（学位論文含む）



データの受け入れと確認

- 研究代表者はデータと関連文書を**SSJDA**に提出する。その際には、**PROSELF**というファイル転送システムを使う。
- 必要とされる情報（メタデータ）は、調査概要、集計結果（単純な頻度表）、標本抽出手続き、データ収集方法（面接、留め置き、郵送等）、データのフォーマット、コーディング等
- この段階で、データ利用の範囲と調査報告書の複写について確認する。
- 研究代表者（データ収集者）は、**SSJDA**とだけ意思疎通をすれば良い。個別の利用者とやりとりをする必要はない。
- 原データはそのままの形式で保存する。

データの整理

- データの単純集計が、報告書のものと合致するかを確認する。
- 個人識別子があれば、削除する。
 - 氏名、住所、電話番号等が残ったままのデータが寄託されることも稀ではない。
- 間接識別子については、必要な処理を行う。
 - 調査対象者が居住する市区町村が含まれる場合は、削除し、都道府県レベルの情報のみ提供する。現段階では衆院選小選挙区の情報提供している。
 - 従業員・被雇用者の調査の場合、企業規模や創業年の情報は削除する。
- データにエラーがあれば、必要な修正を行う。
 - 原データは、一部の人だけに聞かれる質問（サブ・クエスチョン）の処理が間違っていることが多い。
- データはSPSS形式およびタブ区切りテキストファイルで提供されている。

データ概要の作成

- 必要なデータかどうかを利用者が見極められるように、調査概要を作成して、検索可能にしている。
 - 寄託者から提供された文書から該当箇所を適宜抽出し、整理している。
 - 事前に調査内容の詳細と集計結果がわかるように調査票と単純集計はウェブ上から閲覧可能にしている。
- 提供されるデータファイルと、公開される調査票や集計表情報が合致するかは確認されている。
- データ公開前に、データおよびメタデータは寄託者・研究者によって確認・承認してもらう。
- データが公開されたら、ウェブ上で告知する。1カ月に1回送信される電子メールのニューズレターでも案内する。

データの提供方法：SSJDA Direct



[English](#)

東京大学社会科学研究所
附属社会調査・データアーカイブ研究センター

[マイページ](#) > [利用申請](#)

前田幸男さん

調査番号, データ名, 寄託者名で検索

詳細検索

－ 申請するデータを下記一覧より選択ください（調査名をクリックすると、データの概要が表示されます） －

<< 1 2 3 4 >> 全39件

－ データ一覧 －							
調査番号	調査名	寄託者名	データ提供方法	教育目的利用	利用期限	Nesstar	
0093	衆議院選挙に関する世論調査(1996年総選挙前後調査), 1996	選挙とデモクラシー研究会(JEDS)	ダウンロード	○	* 期限なし	--	選択
0247	日本人の民主主義観と社会資本に関する世論調査, 2000	選挙とデモクラシー研究会(JEDS)	ダウンロード	○	* 期限なし	--	選択
0751	若い有権者の意識調査(第3回), 2009	明るい選挙推進協会	ダウンロード	○	* 期限なし	--	選択
0752	第45回衆議院議員総選挙の実態, 2009	明るい選挙推進協会	ダウンロード	○	* 期限なし	--	選択
0773	日本人の社会的期待と総選挙に関する世論調査(Waseda-CASI&PAPI2009), 2009	W-CASI研究会2009	ダウンロード	○	一年間	--	選択
0774	日本人の社会的期待と選挙に関する調査(Waseda-CASI2010), 2010	W-CASI研究会2010	ダウンロード	○	一年間	--	選択
0904	第46回衆議院議員総選挙全国意識調査, 2013	明るい選挙推進協会	ダウンロード	○	* 期限なし	☒	選択
0905	第23回参議院議員通常選挙全国意識調査, 2013	明るい選挙推進協会	ダウンロード	○	* 期限なし	☒	選択
0992	第22回参議院議員通常選挙の実態, 2010	明るい選挙推進協会	ダウンロード	○	* 期限なし	☒	選択
0993	第17回統一地方選挙全国意識調査, 2011	明るい選挙推進協会	ダウンロード	○	* 期限なし	☒	選択

<< 1 2 3 4 >> 全39件

* ただし、教育目的の場合は1年

データ概要の具体例

概要	
調査番号	0905
調査名	第23回参議院議員通常選挙全国意識調査, 2013
寄託者	明るい選挙推進協会
利用申込先・承認手続き	利用方法の詳細は こちら SSJDAが利用申請を承認したときに利用できる
教育目的(授業など)の利用	教育(授業・卒論等)も可
利用期限	研究はなし教育は一年
データ提供方法	ダウンロード
メタデータ閲覧・オンライン分析システム Nesstar	メタデータ閲覧・オンライン分析システム「 Nesstar 」を用いた分析が可能。
調査の概要	この調査は、2013年7月21日執行の第23回参議院議員通常選挙における有権者の投票行動等の実態を調査し、今後の選挙啓発上の資料とすることを目的とする。 主な調査項目は次のとおり。 1)政治・選挙への関心、2)投票と棄権、3)投票選択の経緯、4)選挙媒体との接触、5)選挙制度について、6)選挙啓発への参加と接触 なお今回から、調査方法をこれまでの面接調査法から郵送調査法に改めて実施した。
調査対象	満20歳以上の男女個人
調査対象の単位	個人
サンプルサイズ	標本数: 3,000人 有効回収数: 2,019(67.3%)
調査時点	2013年9月18日～10月24日
調査地域	全国
標本抽出	標本抽出: 選挙人名簿を使用 抽出方法: 層化2段無作為抽出法 地点数: 210市区町村 210地点
調査方法	郵送配布郵送回収法
調査実施者	(公財)明るい選挙推進協会、実査は一般社団法人 新情報センター、埼玉大学社会調査研究センター
委託者(経費)	
寄託時の関連報告書・関連論文	「第23回参議院議員通常選挙全国意識調査－調査結果の概要－」、2014年5月、(公財)明るい選挙推進協会

データの提供

- 右記の誓約事項に同意して初めて、データは提供される。
 - 個々の調査対象を特定することはない。
 - データの出所を明記する。
 - 利用・実績を報告する。
- 実は、承認は自動的ではなく、担当者が実際に研究計画書などを読んだ上で承認する。
- 実績報告は、SSJDAの活動について報告し、翌年度の予算要求をする上では極めて重要。

誓約事項

1. 提供された個票データは、学術目的での二次分析にのみ利用します。個別データの秘密保護を図り、個々の調査対象を特定する分析は行いません。提供された個票データは、本利用申請をした者だけが利用し、第三者には再提供しません。

2. 調査に関する照会は、原則としてSSJデータアーカイブを通して行い、寄託者には直接行いません。

3. 二次分析の結果を発表する際には、個票データについて以下の文を付すことにより、個票データの出典を明記します。
二次分析に当たり、東京大学社会科学研究所附属社会調査・データアーカイブ研究センターSSJデータアーカイブから「〇〇〇調査」(寄託者名)の個票データの提供を受けました。

The data for this secondary analysis, name of the survey, name of the depositor, was provided by the Social Science Japan Data Archive, Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo

4. 登録したユーザー情報や申請内容について変更が生じた場合は、速やかに、SSJDA Directで変更した情報を修正します。

5. 利用期限終了後は、速やかに個票データを消去して、SSJDA Directで利用報告を行います。
その際、論文等(1)を発表していれば、利用した調査の寄託者数 + SSJデータアーカイブ分1の部数をSSJデータアーカイブ(〒113-0033 東京都文京区本郷7-3-1 東京大学社会科学研究所附属社会調査・データアーカイブ研究センター SSJデータアーカイブ)まで郵送にて送付します。

6. その他、SSJデータアーカイブの指示に従います。

7. SSJDA Directの利用により何らかの不利益を被ったとしても、寄託者及びSSJデータアーカイブの責任は一切問いません。

(1) 著書や論文、学会の大会報告書、授業等の実習報告書などの印刷されたものを指します。「論文等」が、印刷物の一部分であり全体の提出が困難な場合、該当する「論文等」を複写したもので可。
提出のあった論文等は、SSJデータアーカイブから各寄託者に1部ずつ転送します。

☐ 上記誓約事項を厳守することに同意します。

戻る

同意する

データの保存と公開を担う組織 および専門家について

- あくまで個人的感想
- データ・プロフェッション
 - 図書館にライブラリアンがおり、博物館に学芸員がいるように、データ・アーカイブにはデータ・アーキビストが必要である。
 - 諸外国のデータ・アーカイブの専門家には関連分野で博士号を持つ人も珍しくはない。
 - 専門家の学会もある。
 - [International Association for Social Science Information Services and Technology](#)
 - 学者が研究・教育の片手間で行うことではなくなっている。
- 社会科学（あるいはそれぞれの分野）のデータ収集・保存・共有を支えるインフラストラクチャの整備とそれを支える人材の育成が必要ではないか。デジタル・データの保存についての研究はその一環として非常に重要。

ご静聴ありがとうございました。

参考文献リストは末尾にあります。

前田個人への問い合わせは ymaeda@iss.u-tokyo.ac.jp

SSJDAへの問い合わせは ssjda@iss.u-tokyo.ac.jp

をお願いします。

参考文献リスト

- 前田幸男. 2007. 「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) 98巻9号: 112-118.
- ー. 2008. 「アメリカ合衆国におけるセンサス・マイクロ・データの保存と提供」 『統計』 59巻12号: 21-28.
- ー. 2008. 「もうひとつのアーカイブ」 『情報学研究: 学環: 東京大学大学院情報学環紀要』 73巻: 85-87.
- ー. 2008. 「海外におけるリモート集計システムをめぐる動向 (特集 リモート集計の現状)」 『エストレーラ』 175号: 18-27.
- ー. 2011. 「統計メタデータとデータ提供方法の変化--DDIについての概観」 『エストレーラ』 206号: 2-8.
- UK Data Archive 『データの管理と共有ー研究者向け最良事例』 2011年発行/翻訳2013年 (東京大学社会科学研究所附属社会調査・データアーカイブ研究センター)



研究者向け最良事例

2011年5月 (英語版)



社会調査・データアーカイブ共同利用・共同研究拠点
東京大学 社会科学研究所
附属社会調査・データアーカイブ研究センター

2013年2月 (日本語版)

- Adams, Margaret O'Neill. 1995. Punch Card Records: Precursors of Electronic Records. *American Archivist* 58: 182-201.
- ———. 2003. "Three Decades of Description and Reference Services for Electronic Records " In *Thirty Years of Electronic Records*, ed. Bruce I. Ambacher. Lanham, Maryland: Scarecrow Press, Inc.
- Adams, Margaret O'Neill, and Thomas E. Brown. 2000. Myths and Realities About the 1960 Census. *Prologue* 32 (4): 266-270.
- Ambacher, Bruce I., ed. 2003. *Thirty Years of Electronic Records*. Lanham, Maryland: Scarecrow Press, Inc.
- Anderson, Margo J. 1988. *The American Census: A Social History*. New Haven: Yale University Press.
- ———, ed. 2000. *Encyclopedia of the U.S. Census*. Washington, D.C.: CQ Press.
- Austrian, Geoffrey D. 1982. *Herman Hollerith: Forgotten Giant of Information Processing*. New York: Columbia University Press.
- Conrad, Mark. 1994. Reeling Them In: Accessioning the Electronic Records of the United States Government. *IASSIST QUARTERLY* 18 (1/2): 11-13.
- Fishbein, Meyer H. 2003. "Recollections of an Electronic Records Pioneer." In *Thirty Years of Electronic Records*, ed. Bruce I. Ambacher. Lanham, Maryland: Scarecrow Press, Inc.
- Henry, Linda J. 2003. "Appraisal of Electronic Records: Traditional Principles Endure." In *Thirty Years of Electronic Records*, ed. Bruce I. Ambacher. Lanham, Maryland: Scarecrow Press, Inc.
- Ruggles, Richard. 1973. "The Preservation and Use of Machine-Readable Records." In *The National Archives and Statistical Research*, ed. Meyer H. Fishbein. Athens, Ohio: Ohio University Press. 29-38.
- Blake, Kellee. 1996. 'First in the Path of the Firemen': The Fate of the 1890 Population Census. *Prologue* 28 (1).