

〈研究ノート〉

社会調査データ・アーカイブの利用方法*

中 野 康 人**

1 データ・アーカイブ利用の意義と目的

本稿の目的は、社会調査データ・アーカイブを利用してデータ分析を行う方法を整理し、その実際的方法を紹介することにある。

データ・アーカイブとは、社会調査データおよび関連情報を蓄積・保管し、他者が利用可能な状態で公開しているものである。データ・アーカイブの歴史については真鍋（2003）などに詳しいが、その存在意義としては、調査コストの削減であるとか、公共財としての調査データを共有できること、などがあげられる。

一方利用者側の視点では、データ・アーカイブを利用すると次のようなことが可能になる。これらは、研究目的での利用だけでなく、教育目的での利用にも有益である。

二次分析 個票データが入手できれば、自ら実施した調査と同じような感覚で、データ分析ができる。もちろん、アーカイブ・データを利用するということは、データそのものは二次的な利用となる。二次分析とは、「社会調査によって収集された公開データを再分析し、既存の仮説や新しい仮説を検証したり、あるいは新しい分析手法を適用したりする」（佐藤ら 2000）方法である。

メタ分析 複数の調査データが蓄積されているので、同様のテーマに関する調査・変数を検索し、メタ分析の材料とすることができる。メタ分析とは、単体の調査にもとづいたデータを分析するのではなく、同じテーマ・変数について

横断的に複数の調査データを利用して分析をおこなう手法である。詳細は、Mullen（1989 [2000]）などを参照。

追試 先行研究で分析されたデータが公開されていれば、そのデータを入手して手で再分析・追試することができる。もちろん、新たなモデルや分析手法を適用して二次分析を行うことも可能である。

既存の調査データの単純集計 詳細な分析に供するだけではなく、単純集計レベルでも有意な情報を得られる。個票データが公開されていなくても、単純集計レベルの情報は公開されていることが多い。世論調査や意識調査のデータについて、「ある考えをもっている人はこのくらいいる」ということがわかる。

質問文の参考 データそのものではなく、質問文や変数のデータベースとしても利用可能である。同じ理論的概念でも、操作化されると異なる質問文になる可能性がある。『世論調査年鑑』で質問文の雛形をさがすのと同じ感覚で、データ・アーカイブに蓄積された調査票を利用できる。

以上のような目的をもってデータ・アーカイブを利用することを想定し、以下ではその具体的な方法を整理する。

2 作業の流れ

作業の流れは、おおよそ図1のようになる。以

*キーワード：データ・アーカイブ、二次分析、検索

**関西学院大学社会学部助教授

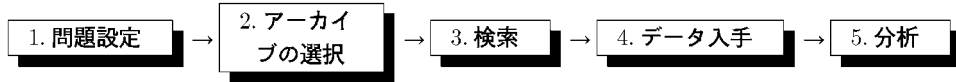


図1 作業の流れ

下に、各ステップの概要を示す。

2.1 問題設定

まずは、どのような目的でデータ・アーカイブを利用するにせよ、自らの問題設定をしておく必要がある。これは、研究のテーマもしくは分析しようとしている概念や変数を特定することを意味する。

2.2 データ・アーカイブの選択

次に、利用するデータ・アーカイブを選択する必要がある。後述するが、データ・アーカイブは複数あり、蓄積されている調査は必ずしも重複していない。特定の調査をターゲットにするのであれば、その調査が公開されているデータ・アーカイブを選択しなければならない。もし、探索的に検索を行おうとしているのであれば、複数のデータ・アーカイブで検索を試してみることも有用だろう。

世界中のデータ・アーカイブを完全に網羅したリストは今のところ存在しない。しかし、IFDO (International Federation of Data Organizations) のウェブ・ページ (http://www.ifdo.org/org_archives/arc_bfr.htm) などが参考になる。そのほかには、佐藤 (2000) や、CESSDA のページなども参考になる。CESSDA のウェブ・ページでは、CESSDA 加盟のデータ・アーカイブを一括して検索することができる (<http://www.nsd.uib.no/Cessda/>)。

2.3 検索

対象となるデータ・アーカイブが決まったら、設定した問題にかんする調査を検索する。詳細な検索の方法は、データ・アーカイブ毎に異なるが、ウェブ・ページに検索システムが統合されている場合が多く、基本的には「キーワード」で該当する調査を検索していく。

ただし、検索にヒットした結果が、自分の問題



図2 CESSDA : Archives in Europe

設定と概念的に一致するか否かはわからない。検索結果を注意深く吟味して、研究テーマに一致したデータを見付けなければならない。場合によっては、変数の再コードや新しい変数作成（計算）が必要になる。2次分析における再コードの方法などは、Elder et al.(1993)などに詳しい。

2.4 データおよび関連情報の入手

アーカイブの運営形態などによって異なるが、データ・アーカイブには個別の社会調査について、調査の概要（study description）、調査票・質問文、単純集計表（code book）、個票データなどが蓄積されている。調査毎に、実際に公開される情報にばらつきがあるし、利用に際しての手続きも異なり得る。最も開放度が高い場合は、個票データまでが自由に利用できる。一方、調査の概要はすぐに閲覧できるが、個票データなどは個別に調査主体に問い合わせないと入手できない場合もある。

また、情報入手の手段もウェブ・ページから検索・ダウンロードできる場合もあれば、アーカイブ事務局に申請してCD-ROMなどの媒体を郵送してもらう場合もある。

2.5 分析

分析対象にする調査が検索できたら、調査票やコードブック、個票データを入手して、目的にあった分析を行う。

3 具体的利用方法

次に、いくつかのデータ・アーカイブを実際に利用する方法について解説する。利用環境としては、2003年6月時点の関西学院大学社会学部での環境を前提とする。

3.1 ICPSR

ICPSR (Inter-university Consortium for Political and Social Research) は、アメリカのミシガン大学で運営されているデータ・アーカイブである。社会科学分野でもっとも大きなデータ・アーカイブの一つで、World Value SurveyやInternational Social Survey Programなど、日本のデータを含む

調査データが入手可能である。関西学院大学は、ICPSR 国内利用協議会 (<http://ssjda.iss.u-tokyo.ac.jp/icpsr-natl.html>) のメンバーであり、所定の手続きをとることにより、無料で個票データの利用が可能である。

3.1.1 検索

検索は、ICPSR のウェブページから可能である (<http://www.icpsr.umich.edu/access/>)。

図3が検索画面である。“Search”欄にあるテキストフィールドに、検索したいキーワードを入力する。右側のリストボックスで、検索するフィールド（title、study no、investigator、subject terms、all fields）を指定できる。オプションとして、“words”または“phrase”を選択できる。キーワードに複数の単語を入浴した場合、前者は各単語を個別に検索するのに対し、後者は一つの成句として検索する。“Web site”は、アーカイブデータだけでなく、ICPSR のウェブサイト全体を検索するオプションである。このオプションをチェックすると、調査票やコードブックだけでなくサイト内にある論文なども検索対象になる。“data holdings”は、調査票やコードブックなどデータ・アーカイブの本体を検索対象とする。

“Advanced Search”をクリックすると、より詳細なオプションを指定できるページに切り替わる。Advanced Searchでは、複数のキーワードをAND（must contain）、OR（should contain）、NOT（must not contain）で指定できる。また、検索フィールドも詳細に指定できる。データが更新された日付でも検索が可能である。

“Search”ボタンを押すと検索結果が出力される（図4）。

検索結果は、study no や title の一覧として出力される。各調査タイトルの下に、“description”、“downloads”、“related literature”の三つのリンクが張られる。“description”は、その調査の概要を出力する。“downloads”は、その調査に関連してダウンロードできるファイルのリストを出力する。“related literature”は、その調査をもとにした論文の一覧を出力する。ICPSR でデータを取得して論文を執筆した場合、その書誌情報をICPSR に報告する義務がある。ここで出力される

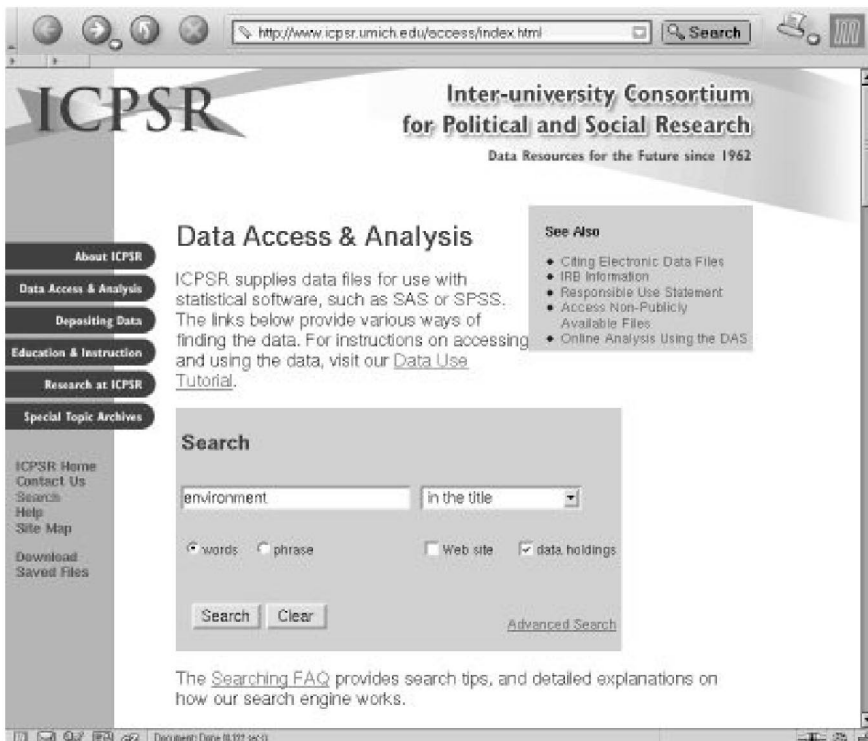


図3 ICPSR : Search

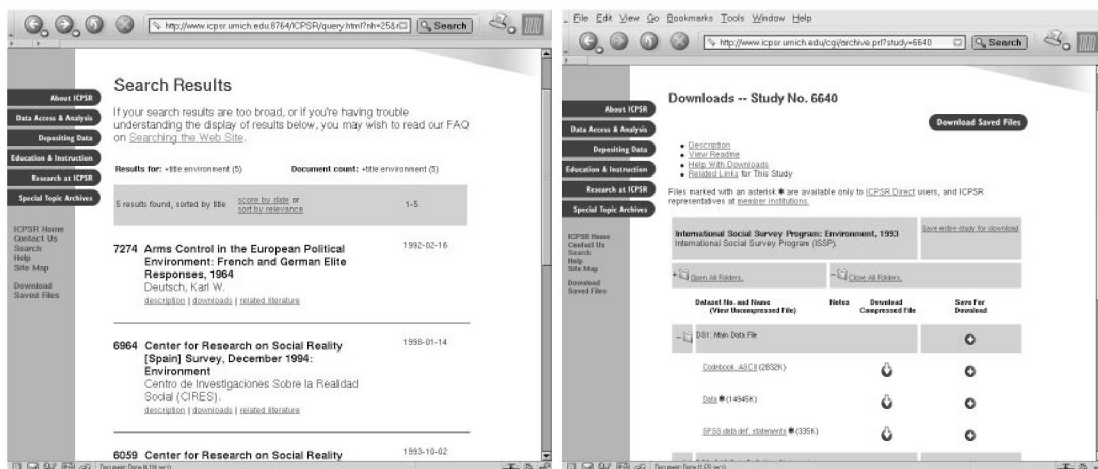


図4 ICPSR : Result & Download

書誌情報はそうした報告をもとにしている。自分の研究課題に関連する文献を発見するのにも役立つ。

3.1.2 データの取得

“downloads” に出力されるリストのうち、*

(アスタリスク) がついたものは、ICPSR のメニューにのみ閲覧が許可されたものである。それ以外のものは、そのままダウンロード・閲覧が可能である。特に個票データは*つきのものが多い。個票データをはじめとする*つきのファイルを手するには、ICPSR 国内利用協議会に利用を申請

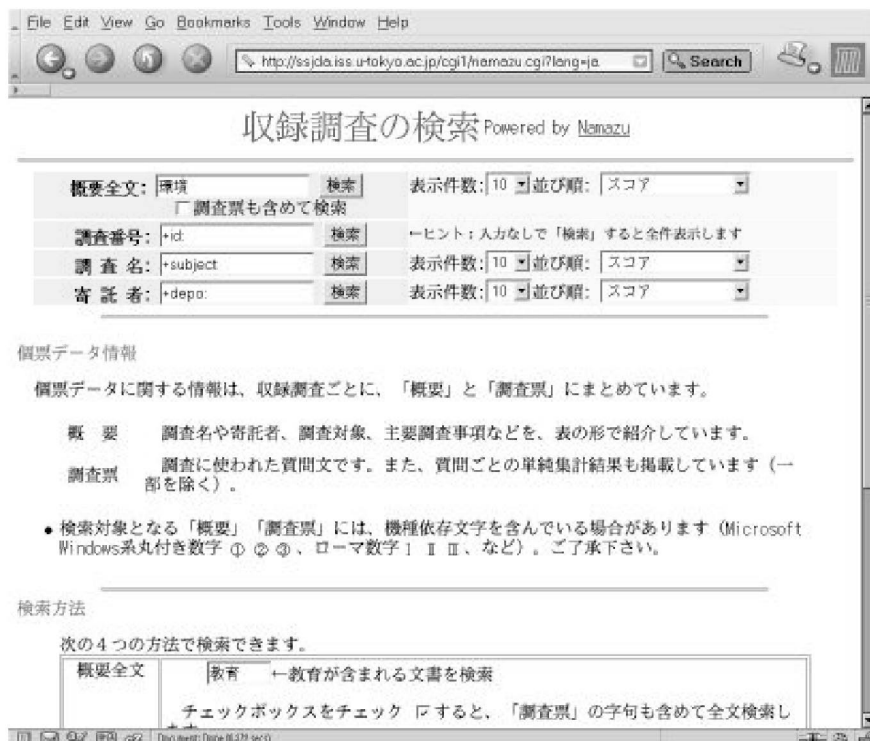


図5 SSJDA : Search

する必要がある。申請の窓口は、大学図書館のレファレンスカウンターである (<http://library.kwansei.ac.jp/reference.html#icpsr>)。

3.2 SSJDA

SSJDA (Social Science Japan Data Archive) は、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で運営されているデータ・アーカイブである。日本で唯一の IFDO メンバーでもある。

3.2.1 検索

検索は、SSJDA のウェブページから可能である (<http://ssjda.iss.u-tokyo.ac.jp/cgi1/namazu.cgi?lang=ja>)。

図5が検索画面である。基本的に、テキストフィールドに検索したいキーワードを入力して、「検索」ボタンを押すと検索結果が出力される。検索対象となるのは、「概要全文」「調査番号」「調査名」「寄託者」のいずれかで、それぞれ個別にテキストフィールドが用意されている。

検索結果は、調査番号や調査名の一覧として出

力される (図6)。各調査名をクリックすると、その調査の概要が表示される。概要のページから調査票および単純集計のページにもリンクが張られている。

3.2.2 データの取得

個票データの入手は、SSJDA に申請書などを郵送し、データが入った CD-ROM を返送してもらう形態をとっている。利用方法や制限などについては、SSJDA のウェブページを参照のこと (<http://ssjda.iss.u-tokyo.ac.jp/access.html>)。

3.3 コードブック

次に、データ・アーカイブなどから入手したコードブックなど、電子媒体の調査情報が手元にある場合に、そこから目的の情報を引き出す方法を紹介しておく。複数年次のデータが入った ISSP 調査の CD を入手して、そこから目的の調査内容を含む調査を捜し出すような状況を想定している。

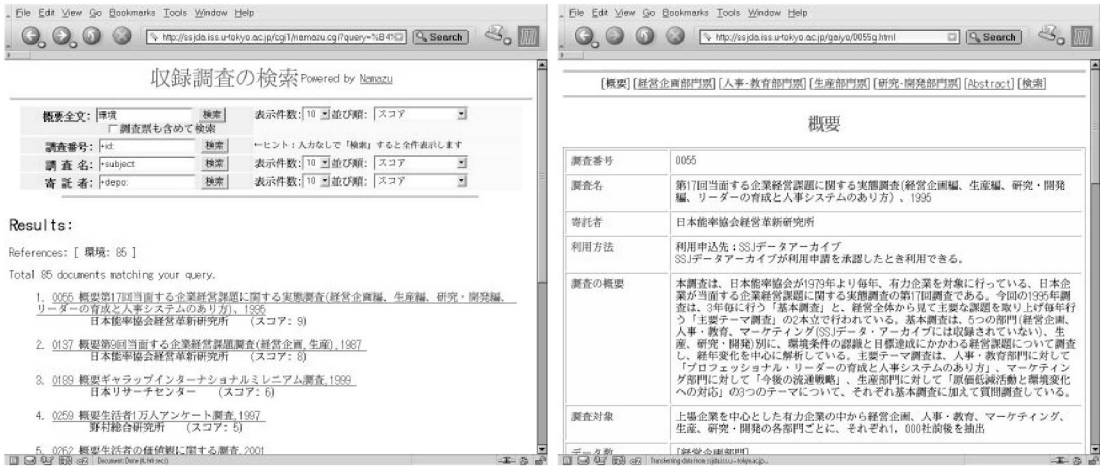


図6 SSJDA : Result

3.3.1 検索ツール

コードブックや調査票などのファイルがある場合、ファイルの内容を検索できるツールを使えばよい。特に、ファイルが複数あるような状況では、クロスファイルサーチ（複数のファイルの内容を検索できるツール）が必要となる。

Windows の「検索」 MS-Windows 系の最近の OS であれば、スタートメニューから「検索」を選んで、ファイル内容の検索ができる。ただし、出力結果がファイル名の一覧になるだけなのであまり便利とはいえない。

grep grep コマンドは、ファイル内容を検索する強力なツールである。検索結果も、ファイル名およびファイル内の行数まで細かく出力してくれる。ただし、基本的に CUI なプログラムなので、出力結果の読みやすさも含めて慣れが必要である。

Acrobat Reader の Index Search Adobe 社の Acrobat Reader は、現在インターネット上の文書配布で DeFact Standard になっている PDF 形式のファイルを表示するプログラムである。コードブックなどの調査関連文書も PDF ファイルで公開されていることが多い。また Acrobat Reader は無償で配布されている。

3.3.2 Acrobat Reader による検索

Acrobat Reader の Index Search 機能は、日本語版の Acrobat には含まれない。しかし、英語版の Acrobat などを使えば利用可能である。英語版は日本語版とは別にインストール可能であるし、また英語版のみでもフォント環境さえ整っていれば日本語ファイルの表示には問題はない。

Index Search は、メニューの file にある search から実行できる。検索を実行する前に、検索対象となるファイル群に対応した index file を指定する（図7）。Index file は検索情報を含んだファイルで、作成には Acrobat の Writer が必要となる。ただし、コードブックと同時に配布されていることが多い。

指定できたら、search の query で検索する。検

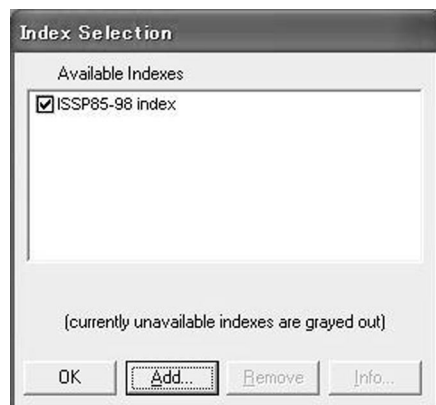


図7 ACROBAT : index file の指定

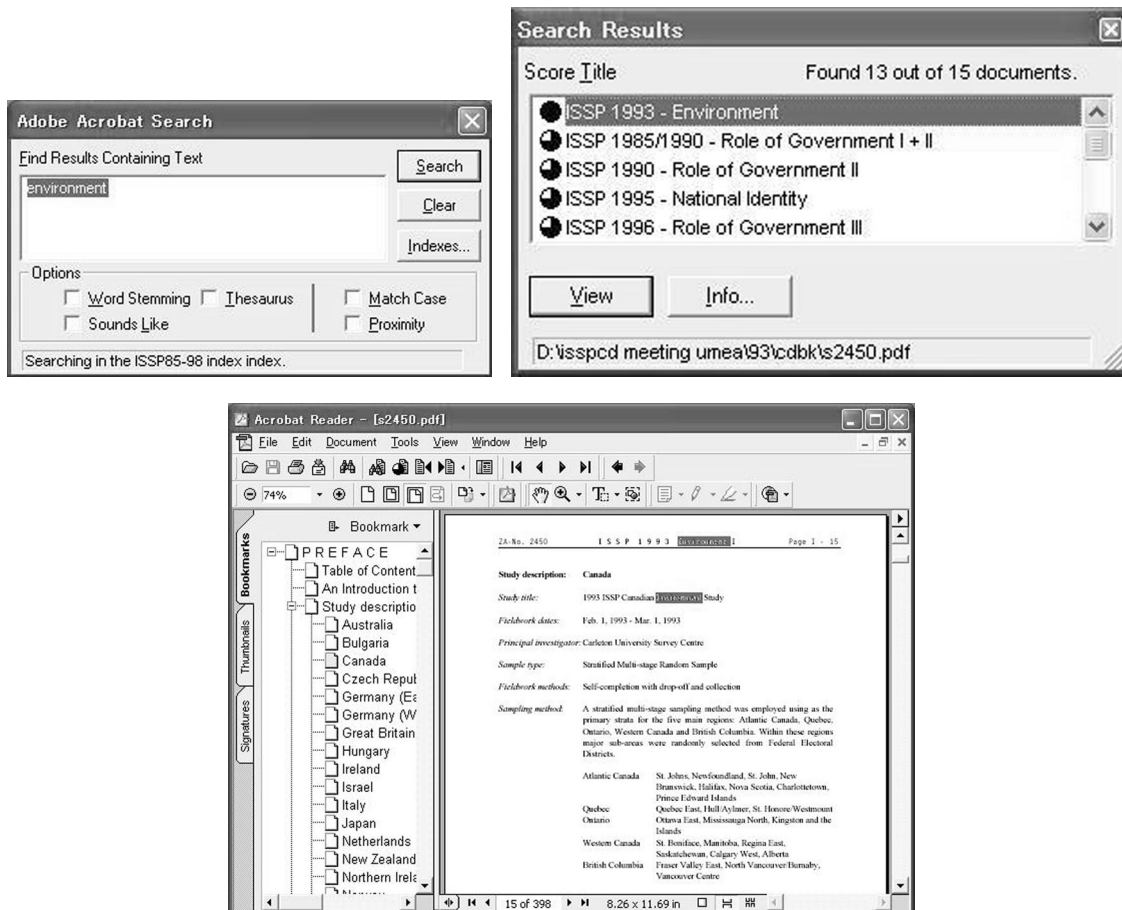


図8 ACROBAT：検索とその結果

索のオプションが細かく指定できるが、ここでは省略する。

検索結果は、まずキーワードを含むファイルの一覧の形であらわれる。そして、ファイルを指定すると内容が表示される。指定したキーワードがハイライト表示されるので、どの部分に目的の語があるのか分かりやすい（図8）。

以上の方法は単純にキーワードの検索をするため、一般的な単語を指定すると膨大なヒット数になってしまうという欠点がある。例えば、調査名に含まれる単語で検索した場合、コードブックのヘッダやフッタに調査名があると、毎ページにヒットする箇所が存在することになる。この問題を回避するためには、検索対象となるフィールドを指定する必要がある。例えば、質問文のみを対象とするとか、調査名のみを対象とするなどであ

る。こうした方法を実現するには、ファイルのデータベース化、またはファイルの統一した規格化が不可欠である。これまで紹介したデータ・アーカイブの検索システムは、コードブックなどをデータベース化することによって詳細な検索を可能にしている。

4 参考情報

以下に、データ・アーカイブの利用に際して参考になる情報源をいくつかあげておく。

IFDOnet [<http://www.ifdo.org/>]

(International Federation of Data Organizations)
データ・アーカイブに関する技術的な情報、アーカイブのリスト、二次分析のマニュアルなど

IASSIST [<http://datalib.library.ualberta.ca/>]
 (International Association of Social Science Information Service & Technology) リンク集が
 役立つ

CESSDA [<http://www.nsd.uib.no/Cessda/>]
 (COUNCIL OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES) アーカイブのリスト地図、
 CESSDA 内の総合検索機能など

ICPSR [<http://www.icpsr.umich.edu/>]
 (Inter-university Consortium for Political and Social Research) データ・アーカイブの他、
 DDI の開発など

ZA [<http://www.gesis.org/en/za/index.htm>]
 (Zentralarchiv für Empirische Sozialforschung, Universität zu Köln) データ・アーカイブの他、
 ISSP、Eurobarometer など

SSJDA [<http://ssjda.iss.u-tokyo.ac.jp/>]
 (Social Science Japan Data Archive) データ・
 アーカイブの他、二次分析研究会、リンクなど

SORD [<http://www.sgu.ac.jp/soc/sordhp/main.htm>]
 (Social and Opinion Research Database Project)
 日本社会学会会員によって実施された社会調査
 に関する情報を収集・蓄積

JPOLL, the Roper Center [<http://www.ropercenter.uconn.edu/jpoll/home.html>]
 日本の世論調査データのアーカイブ

GSS [<http://www.icpsr.umich.edu/GSS/>]
 (General Social Survey) GSS の調査票情報など

ISSP [<http://www.issp.org/>]
 (International Social Survey Programme) 日本
 を含む国際比較調査

社会調査の工具箱 [<http://www.nara-u.ac.jp/soc/staffs/mabuchi/toolbox.html>]

間淵領吾 (奈良大学社会学部) 編集のリンク集

文献

- 真鍋一史, 2003, 『国際比較調査の方法と解析』慶應義塾大学出版会。
 佐藤博樹・石田浩・池田謙一編, 2000, 『社会調査の公開データ 2次分析への招待』, 東京大学出版会。
 Brian Mullen, 1989, *Advanced BASIC Meta-Analysis* (『基礎から学ぶメタ分析』小野寺孝義訳, [2000], ナカニシヤ出版)。
 Glen H. elder, Jr., Eliza K. Pavalko, Elizabeth C. Clipp, 1993, *Working with archival data*, Sage.

謝辞

本稿を執筆するに先だって、データ・アーカイブに関する情報を得る機会が二つあった。一つは、2002年11月に関西学院大学社会学部で行われた Rolf Uher 氏 (ZA、ドイツ・ケルン大学) によるワークショップ。そしてもう一つは、社会学部から派遣され、筆者が2003年3月にドイツ・ケルン大学まで赴いた ZA 視察である。貴重な機会を与えて頂いたことを記して感謝したい。

Retrieving Information from Data Archives

ABSTRACT

The purpose of this paper is to review methods to retrieve information from social research data archives. The intended readership is researchers interested in secondary analysis of social research data and beginners in data archive utilization. First of all, The aims of using data archives are confirmed. That is to say, the aims are secondary analysis, meta-analysis, supplementary verification, descriptive analysis, and questionnaire references. To achieve these aims, the process needs to be taken in five steps; (1) defining a research purpose, (2) selecting an adequate data archive, (3) retrieving information from the data archive, (4) obtaining appropriate data, and (5) analyzing the data. Then we take examples of step (3) and (4) in some data archives. In addition, some useful URLs are listed.

Key Words: data archives, retrieving, secondary analysis