

## SORD プロジェクトの現状と課題

The Present Situation and Problems of  
Social and Opinion Research Database (SORD) Project

### 新國三千代

札幌学院大学社会情報学部の新國三千代と申します。

SORDプロジェクトの現状と課題と題しまして報告させていただきます。

最初にプロジェクトの現状をご紹介いたしまして、今回のシンポジウムのテーマでございます、データアーカイブ時代における社会調査の課題、すなわち、調査情報の生成・蓄積・活用、これに関するSORDプロジェクトの課題ということにつきましてお話ししたいと思っております。

SORDと申しますのは、もう皆さんご存じかと思っておりますけれども、Social and Opinion Research Database (社会調査データベース)の頭文字をとった略称でございます。これは先程、秋山学部長から説明がございましたが、1991年4月に本学に社会情報学部が設立されたときに、当時の田中一学部長が「社会調査と情報系の研究者が集まって全国的な共同研究を進めよう」と発案されて開始されたものでございます。学部事業として位置付けておりました。

本プロジェクトの当初の目的は、実際に社会調査のデータベースを構築しながら、データの作成・蓄積・管理・提供についての基礎的な研究を行うというものでございました。10年が経ったわけですが、このプロジェクトに参加したメンバは、本学部教員が延べ7名、学外からは全国的に著名な社会調



新國 三千代 氏

査分野の研究者延べ9名の合計16名で今に至っております。本日講演される原純輔先生もメンバとしてご協力下さいました。

さて、そもそもSORDプロジェクトとは、いったいどんなことをやってきたのかということなのですが、この10年間の成果を挙げると言われますと、社会学分野における社会調査に関するデータベースを構築したこと、その結果、社会調査の実態を知ることができるようになったこと、さらに社会調査分野のデータベースやデータアーカイブに関する議論の活発化に多少なりとも貢献してきたことではないかと思っております。

少し詳しく述べますと、1994年から3回ほど、日本で実施された社会調査に関する全

国調査をいたしました。これは主に日本社会学会会員を対象に行った調査でございますが、その社会調査の概要をデータベースに蓄積して、それを1,000ページ程の報告書にして情報提供者全員に配布しました。1990年から1997年ぐらいまでの調査がここに入っております。協力いただいた回答者は400名強で、現在はさらに増えておりますし、他の学会の方からも提供をいただいております。現在データベースに収録されている件数は1,144件です。このうち約300名の方からWeb上で公開してもよろしいという理解を得ておまして、現在SORDホームページから730件の調査概要情報が検索、閲覧できるようになっております。

SORDホームページは1998年の8月から運営しております。それから現在に至るまで約2万件を越えるアクセスがございます。このデータベースの閲覧につきましても、閲覧の際に登録をいただいているのですが、延べ800名ほどの登録者がおります。昨年本学部の中澤秀雄さんが中心になってホームページを一新いたしました。しゃれたページになっております（URL⇒<http://www.sgu.ac.jp/soc/sordhp/>）。英語版も作成して下さいまして、それ以降、アクセスが飛躍的に伸びているというところでございます。このように、調査概要の情報に関しましてはかなり公開しているのですが、SORDプロジェクトでは調査データすなわち個票データそのものも蓄積して、提供しています。現在12件あります。実際にこれらの個票データの利用も5件ほどありました。

SORDホームページで社会調査概要情報を閲覧する場合にはメールアドレスを入れて、初回の場合はお名前とか、所属等を入力していただきますけれども、ここで登録をして利用するというかたちになっております。データベースの閲覧登録者は延べ800名ほどおりますが、内訳は、大学や研究所に所属する方

が76パーセントを占めております。その他は公的機関、調査機関、民間の方です。特徴的なのが大学院生、学生の利用が57パーセントで約6割を占めていることです。教員、研究者が約20パーセント。最近では海外からのアクセスが増えておまして3.5パーセントいます。ほとんどが大学関係者で、特に日本の社会について研究している方、それから外国にいる日本人で日本について調べている方、そういう方がアクセスしているようです。以上がSORDの今までの現状でございます。詳細につきましては、本学部の紀要に報告や論文を掲載しておりますのでそちらをご覧ください。

さて、この10年を振り返ってみますと、SORDが設立された当初は、まだ日本の社会学分野では、どんな調査を行っているのかという実態も明らかではなかったという状況でした。それから、もちろんデータベースはなかったのですが、データアーカイブについてもいろいろな企画がなされて、提案もされていたのですが、マスコミ機関を除いた大学関係機関では存在していなかったという状況でした。

それがSORDが立ち上がって、いろいろな活動をしていく中で、1997年ですが、東大の社会科学研究所に日本社会研究情報センターというのができまして、ここでSSJDAという、Social Science Japan Data Archiveの略なのですが、この様なデータアーカイブが設立されました。ここに専任スタッフと運営費がつきまして、いわゆる公的な組織としてデータアーカイブが発足したというわけです。そこでは、今までに組織で実施したデータの保存と提供ということを積極的に進めておられまして、今年の12月に、IFDO (International Federation of Data Organizations for Social Science) という、これは資料収集組織の国際連盟みたいなものですが、そこにここが加盟することが日本で

初めて認められまして、日本の窓口的役割を現在担っています。

では、今 SORD は何をやっているのかと申しますと、今までとあまり変わらないのですが、個人が実施した調査データの保存と提供ということを、その後も続けております。

このような中で最近、調査データを用いた二次分析、あるいは調査データの公開に関する議論が急速に進展してきております。特に去年の12月に、東大出版会から「社会調査の公開データ」という本が出版されまして、これは午後にお話しされる佐藤博樹先生が編集されたものなのですが、これが出ましてから版を重ねて、かなりこのことに関する議論が活発になってきていると思います。このことにつきましては佐藤博樹先生の方からお話しがございます。

このように10年間で随分状況は変わってまいりました。そこで今までの SORD が課題として抱えてきた問題を改めて問い直してみようと考えまして、SORD プロジェクトで認識されてきた課題をまとめてみました。

主に調査データの生成・提供・利用に関する課題でございます。これにつきましては、先程言いました SSJDA や海外のデータアーカイブ [1] などでは、それぞれ組織独自の方法あるいは方式で対応しているわけでございます。海外では、もう既に40年以上も前からデータアーカイブ活動というのが活発に行われておりまして、既にそれぞれのところで解決をみている課題だと思っておりますが、日本におきましては、まだ議論と時間が必要な問題のように私は感じております。

生成と言いますのは、データを提供するための生成という意味でございますが、これに関しましてどのような議論があったかと申しますと、調査データを整理して、編集をして実際に提供をするのですが、編集のときに、たとえば調査に答えた人が特定できないようなかたち、たとえば通し番号にするとか、

空欄にするとか、そういうことをやるわけですけれども、それでも特定される危険性はないのかとか、しかしそれをあまりやってしまうと、実は本当に知りたいことを明らかにすることができなくなるのではないかといったことです。前者については、実際に経済学分野の研究者で生のデータにあたって、どれだけ特定できるかということ进行分析されている方々もおられます。

それから、データを提供するというのだけれども、データの評価はどうなっているのか。これは前から議論がされておりまして、たとえばデータのクリーニングはどの程度なされているのか、これは基本的なことなので、研究者はよく心がけていると思うのですが……。これに対して、データの質、本当に提供するような質をもっているのか、という議論があります。これは非常に難しいわけです。さらにデータのランク付けということも海外なんかではやっております。こういったことにつきましては、誰がどのような方法でやるのかといった結構難しい議論がございませぬ。

SORD としましては、これにつきましては、データクリーニングがしっかりしていて、ご本人が提供するというものはすべて受け入れるというかたちで、今まで取り組んできております。データの評価につきましては、実は、提供するデータそのものの質というのもあるのですが、それがどう利用されるか。そして、それがどのような成果を挙げていくのかということでも評価がまた変わってきます。従って、SORD としては、質につきまして最初の受け入れ段階で厳しい制限を加えていないというのが現状です。

それからデータの提供に関しましては、これは個票データの公開なのですが、これについてずっと以前から問題になっているのは、社会学分野の調査データについて、「学会等に倫理綱領、倫理規約のようなものがないの

に、そういうことでやっていいのだろうか」という議論があります。海外では、基本的にこういう倫理規約というものをつくって、データを公開するというかたちになっておりますが、まだ公開できるデータがなかったという経過もあるわけですが、これはやはりきちんと議論をして綱領作りをする必要がある問題です。

もう1つ、これからの調査データについてはいいとしても、過去の調査データを公開するとすると、今まで調査をする人は、調査に応じる人に対し、“この研究のみでデータを使用するのであって、他の用途には使いません”という約束をして、そういう信頼関係のもとに調査を実施してきたわけですが、こういうことをやると調査を受ける人との信頼関係が損なわれてしまうのではないかと、そういう議論もかなりございました。特にデータベース、昔のものはデータベース化されるなんていうことは全然想定されておりましたので、確認することは今や不可能ということでございます。これについてどうなのだろうかと、そういうような問題を投げ掛ける方もいらっしゃいます。

それから運営組織のあり方、これにつきましてはSSJDAが出現して、ある程度の方向が出たと思っております。これは組織としてきちんとやっていくという意味では、資金と人材をそこに注いで、それなりのかたちをつくっていくということが実現したのだと思います。

SORDの方はどちらかというと、教員のボランティア的な要素が強くて、今までデータを皆さんと協力して収集して蓄積して利用しよう、というようなかたちで進展してきたと思います。こういったあり方というのは、今後どう発展していくかというあたりは、私たちが今後を考える上で非常に興味を持っている課題ではございます。

それから利用につきましては、そもそも当

初から他人がつくったデータを利用できるのかという議論がかなりございました。つまり、調査をする人はそれなりの自分の問題意識や明らかにしたいことに基づいて調査を企画し、調査項目を考えて設定しているのだけれども、問題意識が違う他の人が作成した調査を使うことは考えられない、また、調査の企画についてしっかりと把握していないと利用するなんていうことはできないなど、さまざまな議論がございました。

特に日本の場合、二次分析の教育というのはほとんどなされておりましたので、間違った使われ方をして、こうではないか、ああではないかと議論されてはたまったものではないという、そういった議論も結構ございました。二次分析の教育については最近、この辺のことも佐藤博樹先生のお話であるかと思うのですが、かなり議論が進んで教育がなされているという状況になっていきます。

これも教育に関することなのですが、利用のモラル形成というのがなされていないのではないかと、つまりデータをもらって、適当に改ざんする、こんなことは普通、研究者の場合には考えられないのですけれども、改ざんしたり、あるいはそれを誰かにコピーをして流してしまう。そういうことに対するモラルの形成がなされていない状況で、こういったデータの提供・利用なんていうことを云々されることは問題なのではないかと。こういったことが、従来からずっと議論されてまいりました。それなりの時間は掛かるかとは思いますが、海外では解決されている問題のようですので、いずれは日本においても解決されていく課題であると思っております。

実は、ずっと以前の資料を見ていましたら、ドイツのケルン大学にZAという中央データアーカイブというのがございまして、大きなデータアーカイブなのですが、ここの以前のScheuch 所長さんがこんなことを言ってい

るです。1960年代に、当初はデータライブラリーという言い方をしていたそうなのですが、それを目指して活動をしていたら、「なぜ原データを収集して提供する必要があるのか。分析した報告書で十分ではないか」と、そのような議論が当時あったそうです。やはり賛否両論があったと思うのですけれども、それを経て30年後、今振り返ってみると、データライブラリーの存在が、実はデータを整備して質を高めてきた。それから原データを多くの研究者が利用することで、多様な観点からの分析が可能になってきた。たとえば、新たな指標を発見することも可能になった。データライブラリーがあることによって複数の調査で、同じ質問の回答分析が可能になって、たとえば、複数の時系列調査や国際比較調査などで同一の質問に対する回答データを分析することができる。そういうことが可能になって、狭量の経験主義を越えた社会科学が可能になった。そのようなことをおっしゃっています。報告書というのは特定の人の主観的な要素が入ってくるので、やはりデータライブラリーの存在というのは意味があったと、このように30年前を振り返っておっしゃっています（参考文献〔2〕から新國が要約）。

ということは、このような活動というのは、成果が出るまでには、実はある程度の時間と議論が必要なのではないかとそのように今は振り返って思っています。

そう考えますと、時間がかかるのだったらSORDも少し腰を据えて将来を展望した方がよいのではないかと改めて考えています。特に公的な、ある程度の運営費が維持できるような公的な機関はでき上がっておりますので、お金と人をできるだけかけないでこういった活動を進めていくことも考える必要があります。SORDはお金がそんなにあるわけはありませんから。例えば、今までの、社会調査はどんな調査が行われているかという情報は非常に貴重で、これは今後数年おきにで

すが調査をして追加蓄積をしていくと、結構充実したものになるのではないかと思います。こういったこと、あるいはホームページの更新などをお金と人をかけないでやっていく。そして特徴のある取り組みを考えていったらどうかと思っております。更に今検討していることは、教育研究における公開データの利用方法に関するワークショップのようなものを開いて、たとえばSSMのデータなんかは非常におもしろいと思うのですけれども、こういうのを積み重ねていく。あるいは特色あるデータを収集する。たとえば教育用サンプルデータ、あるテーマに焦点を当てたデータを収集して、そういった特徴のあるものを出しながら、ボランティアによる情報の収集、提供、活用といったものを追求していくというのもあり得るのではないかと思っております。

ただ、これだけだとおもしろくないので、やはりその成果を何らかのかたちで、かかわっている人にとっての何らかの業績になるかたちにもっていきたいと考えています。たとえば、今、考えている研究課題としては、公開データの教育利用の事例に関すること、それから社会調査でどんな調査が行われているかという情報を集めておりますので、社会調査をメタ的に分析するという、それから、現在ホームページ上で中澤さんがBBS、掲示板というのを主催していますが、十数件の書き込みがあります。大学院生、学生がほとんどなのですが、私が横で見ていると非常におもしろいなと思ったのが、学生が社会調査に関するいろいろな質問を聞いてくるわけなのですが、それに対する中澤さんの答え方が、動機づけとか教育的示唆に富んだおもしろいものになっているのです。これを見ていて、これがたとえば今後どういうふうに進展していくのかというのをちょっと注目しているのですけれども、たとえば社会教育学的な視点、コミュニケーション学的な視点から、何か課

題としておもしろいものがあるのではないかというふうに考えております。

それから、次は SORD のマルチメディアデータベース化の基礎研究でございますが、本学部の佐藤和洋さんが代表として取り組んでいる課題でございます。調査データというのは数字のデータだけではなくて、たとえば質的なデータ、聞き取り調査などの音声データ、それから私どものところでは調査票というのを現在 250 件ほど収集しておりますが、調査票というのはいろいろレイアウトがデザインされておまして、そういった情報も重要なので、調査票そのものを画像として、マルチメディアとして扱う、そういったことも研究課題として行っております。

これにつきましては原純輔先生の方から、さらにおもしろいお話が伺えるかと思っております。実は今情報処理学会に「人文科学とコンピューター」という研究会がございまして、ここで“人間の歴史文化遺産のデジタルアーカイブ”ということが非常に活発に議論されております。10月にシンポジウムが開かれるようなのですけれども、最新のデジタル技術を使って、人間の歴史文化遺産を蓄積

して行って、それを次の世代に継承していくということが、非常にホットな話題、最新の研究課題として注目されております。

そういう意味では、本日、ご講演をいただく方々のお話は、まさに議論が盛んなおもしろい研究課題のところでございますので、私どもの SORD プロジェクトも、今後を見据えながら、皆さんのお話しを伺って、これから次のステップを検討する糧にしていきたいと思っております。

御傾聴どうもありがとうございました。

#### 注釈および参考文献

[1] たとえば、下記を参照のこと

ICPSR : Inter-university Consortium for Political and Social research (大学協会)  
<http://www.icpsr.umich.edu/>

CESSDA : Council of European Social Science Data Archive (欧州社会科学調査資料委員会)  
<http://www.nsd.uib.no/Cessda/europe.html>

[2] データライブラリーのありかたに関する研究—その機能と特徴を中心として—、データライブラリーに関する研究会(代表 磯村英一)、調査研究報告書、p.80 から引用、(1995)。