

第4章 既登録文献データ確認調査の実施について

周藤 真也

(早稲田大学社会科学総合学院)

1. 調査の目的

「社会学文献情報データベース」は、1995年以前のデータは、主として「社会学文献目録」(『社会学評論』の巻末に掲載)に収録された書誌情報に依っている。これら文献目録のデータベース化の作業は、概ね2000年には終了したが、文献目録の時代の情報には、基本的な書誌情報しか含まれておらず、データベース化を期に整備された英文の項目などの情報が欠けている。また、収録されているデータには、目録作成およびデータベース化の際に生じたと思われる入力の手違いや、そもそもの文献の基本情報に関する手違い、文献データの重複などが散見されていた。

このため、今回初めて、日本社会学会会員を対象として執筆者本人に文献情報を配布し、登録されているデータの確認をしてもらうことによって、データの修正、重複、および掲載されていない項目の情報の収集に努めるとともに、あわせて未登録文献の調査を行い、「社会学文献情報データベース」の充実を図ることになった。

2. 調査の概要

2.1. 実施時期

実施時期：2004年1月中旬～2004年3月末

当初、〆切を2月末としていたが、回収率を上げるため、いくつかの社会学関連のメーリングリスト等で呼びかけを行い、3月末まで受け付けることとした。また、4月以降に遅れて送られてきたものについてもできる限り対応した。

なお、例年同時期に行っている前年の新規分および過年の収録漏れ分を対象とした「社会学文献調査」と合わせて実施し、未登録の文献情報の収集に努めた。

また、修正情報の収集は、従来、「社会学文献情報データベース」の富山大学サイトで行ってきっていたが、今回の調査では、「既登録文献データ確認調査」の調査票に朱筆することによる用紙での提出に一本化して、会員向けのアナウンスを行った。

2.2.実施手順

1) 会員名簿との照合と、データ配布先の決定

本調査の実施にあたっては、まず既登録データの日本社会学会会員への配布のための準備作業が必要である。「社会学文献情報データベース」ならびにそのもととなった「社会学文献目録」は、日本社会学会の会員からの「自己申告」を基本情報として作成を行ってきたものであるが、共著文献など複数の著者がいるものも多く、どの会員からの申告に基づいたものであるのかの情報を保存していないため、各文献情報と会員個人とは必ずしも1対1で同定されるわけではない。また、「社会学文献目録」は、「自己申告」によるデータ収集をはじめから、50年近くが経過しており、物故者や既に現役を引退され退会された元会員の方も多数いると考えられた。このため、今回調査を実施するにあたって、まずはじめに「日本社会学会会員名簿」との照合を行い、各文献データをどの会員に配布し、確認してもらうかを決定する作業を必要とした。

データの配布先は、次のポリシーにより決定した。

- ① 「社会学文献情報データベース」の「著者名（漢字）」と「日本社会学会会員名簿」の氏名データを照合し、同じ氏名の会員に配布する。
- ② 「社会学文献情報データベース」の「著者名」データに、複数の著者が存在する場合は、先頭より順番に、「名簿」の氏名データとの照合を行い、最初に合致した会員に、データを配布する（これにより、筆頭著者が非会員あるいは現会員ではない場合や、訳者などもサポートされる）。
- ③ ①②によって配布先が決定しないもののみ「著者名（英字）」による照合を行い、データを配布する（一部の欧文論文は、「著者名（漢字）」の情報をもたないため）。

会員名簿のデータは、日本社会学会事務局を通し、事務委託をしていた財団法人日本学会事務センター（当時）から 2003年12月下旬現在のもの（3505名（機関会員を除く））を入手したのだが、この名簿データを確認したところ、次のような問題があった。

- ① 会員氏名の読みは、JIS X 0201 のカタカナ（いわゆる「半角カタカナ」）であった上、拗音促音との区別がなく、読みでの照合が難しく、「著者名（カナ）」との照合はあきらめざるを得なかった。
- ② 会員名簿データの会員氏名には、戸籍上の字体（旧字体や戸籍上のみで使用される特殊な文字などの異体字）を用いている会員が見受けられ、会員が通常、論文等を執筆する際に使用している字体と異なる会員が少なからず存在している。このため、特殊な文字については、一般的な文字に予め置き換える作業が必要となり、また異体字に

については、変換表を用い、自動的に変換して、データベースの著者名データと照合させるようにした。

なお、実際の照合は、調査実施当時の最新データ **Version 6.0 (2003年12月26日版、64,458件)** に基づき、**Perl5** による手製のスクリプトを用いて行った。

また、照合後も、次の点で手作業による調整作業を行った。

① 同字同名者への対応とデータの振り分け

会員内での同字同名は、執筆時期や発表媒体、題名や内容などから、基本的に執筆者を特定して振り分けを行った。会員外(元会員や物故者を含む)との同字同名についても、できるかぎりサポートするようにした。

② 筆名や旧姓で書かれた文献への対応

「著者名」の記述に違いがある場合でも、できるかぎりサポートするようにした。

これらの作業によって、最終的に、**64,458** 件中、**39,086** 件 (**60.6%**) が照合され、本調査における文献データの配布先として決定した。これを配布先会員で見ると、次のようになる。

2003年12月下旬現在の

日本社会学会会員 (機関会員を除く)	3,505 名 (平均 11.1 件、標準偏差 23.5)
既登録文献 1 件以上	2,244 名 (平均 17.4 件、標準偏差 27.5)
既登録文献 0 件	1,261 名

照合件数別を一覧表にすると次のようになる。

照合件数	会員数	%
1,2	493	22.0%
3,4	403	18.0%
5~9	364	16.2%
10~14	239	10.7%
15~19	172	7.7%
20~29	207	9.2%
30~39	107	4.8%
40~49	76	3.4%
50~69	82	3.7%
70~99	52	2.3%
100~149	34	1.5%
150~199	8	0.4%
200~244	5	0.2%
250~	2	0.1%
合計	2244	100.0%

2) 既登録文献データの配布

今回の調査にあたって、調査票の作成、配布、回収、入力作業は、日外アソシエーツ㈱に業務委託を行った。なお、業務委託にあたっては、個人情報保護の観点から、名簿情報のデータは、今回の調査にしか使用せず、使用後は廃棄するという契約を結び、なおかつ調査票の送付に最低限必要な項目の情報のみをお渡しした。

「既登録文献データ確認調査」の調査票にある案内文は日本社会学会データベース委員会が作成し、既登録データの記述形式は、同社と相談して決定した。調査票は、既登録文献が存在する会員と、存在しない会員によって文面が分かれ、それぞれ資料にある見本に準じたものを会員別に作成して配布した。調査票の配布は、「日本社会学会データベース委員会」名の封筒でヤマト運輸のメール便を使用して同社が発送業務を担当した。また、3月8日現在で回答のなかった会員 2,824 名については、督促状（はがき）を作成して3月12日に発送をした。

〈資料〉社会学文献情報データベース既登録文献データ確認調査票（登録文献あり）

〈資料〉社会学文献情報データベース既登録文献データ確認調査票（登録文献なし）

3. 調査結果

「既登録文献データ確認調査」調査票回収状況

宛先不明 203 名を除く会員 3,302 名中 858 名（26.0%、全会員に対しては 24.5%）

修正件数 レコード数：10,091 件 述べ項目数：29,340 項目

重複情報 500 組

その他データ削除要求 93 件

4. 修正作業等について

会員から返送されてきた調査票に基づいて、既登録データの修正作業を行った（実際の修正作業は、日外アソシエーツ㈱に業務委託し、確認・校正作業をデータベース委員会において行った）。

ただし、返送されてきた調査票には、会員がデータベースの内部構成を理解していないことからくる不必要、不適切な修正も見受けられ、実施方法については今後の問題を残した。

5. 調査の反省点、今後の継続的な修正情報、補充情報の収集に向けて

今回の「既登録文献確認調査」では、配布データの4分の1強についての修正データが集まり、多くの不足していた項目のデータ、データの誤りを直すことができた。この意味では、本調査は一応の成功を収めたといえよう。

しかしながら、準備期間が短く、またデータそのものも、配布する以前に行っておくべき修正や統一の点などについて、準備が多少不十分であったことは悔やまれる。また、会員にデータベースの構造や記述の形式を正確に理解してもらった上で、情報を収集することには限界があり、より完成度の高いデータベースを構築する上で問題を残した。これらことから、今後、継続的に修正情報の収集に努め、かつ速やかに公開データに反映されるようなシステム作りが必要になってくると思われる。