

# 社会学的質的データのデジタル・ アーカイブ方法論・序説

— 北海道大学・布施鉄治グループによる炭都・夕張調査に即して —

How to Make Sociological Qualitative Data into Digital Archive? :  
based on investigation by Fuse group about the coal mine city, Yubari

齊藤 康則

SORD プロジェクトでは、1970 年代に「北海道大学・生活社会学研究会」によって行われた夕張調査を対象として、調査資料のデータベース化、調査票・調査データのデジタル化を進めている。本稿は、「社会調査史の博物館」の構築に向けた、約1年半にわたる夕張調査の再検討作業の中間報告である。以下では、「社会調査史の博物館」が有する機能として、〈資料室的機能〉と〈アーカイブ的機能〉の2つを取り上げ、前者について調査資料のデータベース化を、後者について〈調査票のアーカイブ化〉〈調査データのデジタル・アーカイブ化〉を論じる。画像化・テキスト化の2段階からなる〈調査データのデジタル・アーカイブ化〉に関しては、作業過程を通じた〈非定型の定型化〉の方法論を、具体的・実践的に述べる<sup>(1)</sup>。

## 1. はじめに

近年、デジタル技術の飛躍的進展とともに、「デジタル・アーカイブ」という言葉を目にする機会が多くなった。東京国立美術館・国立科学博物館をはじめとする先進的ミュージアムでは、収蔵物のデジタル・アーカイブ化が進められつつあり、今日ではその画像・映像がインターネット上で公開されている<sup>(2)</sup>。社会の情報化とともに、それまで一回性・固有性を特徴としてきた文化資産の公共財化・共有財化という、知の地殻変動が生じているのである（武邑 2003）。

一方、実証的・経験的な社会学の営為を振り返ってみると、そこでは社会的遺産ともなりうる調査資料が、「対象の発見—成果の提

示」という研究サイクルの中で、十分に整理・保管されることなく死蔵・廃棄されてきたことが分かる。SSJDA (Social Science Japan Data Archive) のような量的データのアーカイブが整備され、若手研究者・大学院生を中心に2次分析が活況を呈しはじめたのも、つい最近のことにすぎない（佐藤・石田・池田編 2000）。このように〈データの社会的性格〉がようやく認識されはじめた学界状況にあって、SORDは「北海道大学・生活社会学研究会」（代表 布施鉄治）および「東京大学・地域政策研究会」（代表 蓮見音彦・似田貝香門）による「構造分析」的な地域社会研究の調査資料を受け入れ<sup>(3)</sup>、それらを「社会調査史の博物館」として公開することを目指している。そこでは、量的データのアーカイブとは異な

SAITO Yasunori 東京大学・大学院人文社会系研究科・博士課程

表1 炭鉱労働者調査の質問項目（「夕張市民（の）生活実態調査」）

「夕張市民の生活実態調査（基礎表）」	「夕張市民生活実態調査（主人）」		「夕張市民の生活実態調査C（婦人対象調査票）」	
来夕の時期・理由	生活歴と職業歴	労働組合の参加経験	平日の生活時間の構成	就業の理由
生殖家族の構成	家族生活の困難	雇用主への要望	休日の生活時間の構成	就業に関する夫の意向
定位家族の構成	生活の転期	生活時間の構成	家事・育児の分担	収入の割り振り
両親の扶養状況	戦後日本と自分	生活上の問題と解決	洗濯・食事・掃除の状況	母性保護の認識と意向
家計の取支状況	両親・学校の教育	信頼できる人との関係	子供の養育の分担	健康状態
給与と保険・年金	労働時間の構成(1)	地域諸団体への加入	学校行事等への出席	労働上の問題と解決
インフレと生活	労働時間の構成(2)	夕張市外での諸行動	産児制限・妊娠中絶	労働組合の参加
耐久消費財の所有	労働時間の構成(3)	メディアとの接触	夫の健康状態	労働組合の参加経験
新聞・雑誌の購読	生産性向上の影響(1)	転職希望と定住志向	生活上の問題と解決	将来的な就業希望
現物給与の種類	生産性向上の影響(2)	子供の教育・職業の意向	地域諸団体への加入	転職希望と定住志向
住居の間取り	給与水準の評価	自治体・国への要望	夕張市外での諸行動	夫婦分業に関する意向
炭鉱災害と扶助	労働上の問題と解決	支持政党	生活歴と生活の転期	保育に関する意向
	労働組合の参加		戦後日本と自分	就業に伴う生活課題
			職業歴	子供の教育・職業の意向
			共働きに関する意向	両親・学校の教育
			就業希望の有無	自治体・国への要望
			現在の就業状況	支持政党

る、質的データの特殊性に根ざしたアーカイブ方法論の構築が必要となることは論を俟たない<sup>(4)</sup>。

以上のような問題意識にもとづき、本稿は、布施鉄治・北海道大学教授（当時）を研究リーダーとする「北大・生活社会学研究会」によって行われた、炭鉱都市・夕張をフィールドとする実証的社会調査（以下、「夕張調査」と略称）の調査資料に即して、質的データのデジタル・アーカイブ化の可能性について論じるものである。夕張調査に関連した調査票・フィールドノート・調査対象に関する刊行物などは、調査団の中心メンバーであった小林甫氏から寄贈された調査資料群の一部であるため、資料の網羅性が高いといえる。布施グループによる夕張調査資料を素材として、今後の一般公開に向けた資料整理・保管法（データベース化）、および質的データのデジタル・アーカイブ化のための、現時点での暫定的なノウハウを提示することが本稿の目的である<sup>(5)</sup>。

夕張調査の概略について簡潔に述べておこう<sup>(6)</sup>。布施グループは、炭鉱労働者を中心とした夕張地域の住民諸階級の「労働—生活過程」を、1973年から1983年まで断続的に調査・分

析し、その研究成果は『地域産業変動と階級・階層——炭都・夕張／労働者の生産・労働——生活史・誌』（布施編1982）にまとめられている。この夕張調査は、1973～1974年に実施された北炭平和鉱・三菱南大夕張鉱の鉱員・職員・組夫の面接調査に始まり、自営業者・誘致企業労働者・失業対策事業従事労働者・生活保護世帯・自治体労働者および小中学校・高等学校の生徒・保護者へと、調査対象を拡大させていった。調査者と被調査者とのインターフェイスとなる調査票についても、炭鉱労働者調査で使用された「夕張市民（の）生活実態調査」——A票（「基礎表」）・B票（「主人」）・C票（「婦人対象調査票」）という3種類の調査票によって構成される——を原型として（表1）、調査の経過とともに調査票自体のマイナーチェンジも認められる<sup>(7)</sup>。

以下、節をあらためて、質的データのデジタル・アーカイブ方法論について述べることにしよう。

## 2. 資料室的機能とアーカイブ的機能

### 2.1 調査資料のデータベース化

「社会調査史の博物館」には、それが保管する調査資料の特性に応じて、次のような2つ

1	分類	作成者	作成年月日	タイトル	概要	形態	注記
2	書籍	鎌田隆雄	1979.11.1	石炭の歴史	石炭の歴史(1)の序文、石炭の歴史(2)の序文、石炭の歴史(3)の序文	A6, 280頁	
3	書籍	北海道庁経済計画課	1979.11.1	石炭の歴史(1)の序文	石炭の歴史(1)の序文	A6, 12頁	
4	書籍	夕張市 企画部	1979.11.1	石炭の歴史(2)の序文	石炭の歴史(2)の序文	A6, 116頁	
5	書籍	夕張市 企画部	1979.11.1	石炭の歴史(3)の序文	石炭の歴史(3)の序文	A6, 195頁	
6	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(4)の序文	石炭の歴史(4)の序文	A6, 141頁	
7	書籍	夕張市	1979.11.1	石炭の歴史(5)の序文	石炭の歴史(5)の序文	A6, 124頁	
8	書籍	北海道庁経済計画課	1979.11.1	石炭の歴史(6)の序文	石炭の歴史(6)の序文	A6, 40頁	
9	書籍	夕張市	1979.11.1	石炭の歴史(7)の序文	石炭の歴史(7)の序文	A6, 76頁	
10	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(8)の序文	石炭の歴史(8)の序文	A6, 130頁	
11	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(9)の序文	石炭の歴史(9)の序文	A6, 189頁	
12	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(10)の序文	石炭の歴史(10)の序文	A6, 18頁	
13	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(11)の序文	石炭の歴史(11)の序文	A6, 28頁, 付5頁	
14	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(12)の序文	石炭の歴史(12)の序文	A6, 70頁	
15	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(13)の序文	石炭の歴史(13)の序文	A6, 69頁	
16	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(14)の序文	石炭の歴史(14)の序文	A6, 78頁	
17	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(15)の序文	石炭の歴史(15)の序文	A6, 28頁	
18	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(16)の序文	石炭の歴史(16)の序文	A6, 28頁	
19	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(17)の序文	石炭の歴史(17)の序文	A6, 28頁	
20	書籍	夕張市立旭丘小学校	1979.11.1	石炭の歴史(18)の序文	石炭の歴史(18)の序文	A6, 158頁	

図1 夕張調査資料のデータベース(「調査チーム参照資料」)

の基本的機能を指摘することができる<sup>(8)</sup>。第一は、「調査チーム参照資料」に関する〈資料室的機能〉である。調査チームは、研究プロジェクトの途上で、数多くの書籍(一般/専門)・雑誌(一般/雑誌)・学术论文・新聞記事・行政文書・団体文書を収集するが、調査チーム自身の手によって、それらが意識的に目録化されることは稀である。そこで、受け入れ機関は、DCMES (Dublin Core Metadata Element Set) などの標準規格を参考にしながら、収集した調査資料をデータベース化する必要がある。

布施グループの夕張調査についても、『鉦山保安テキスト 坑内/機械』『採炭学』『廃鉦譜』(書籍), 『経済評論』『地域開発』『月刊ガン』(雑誌), 『石炭業における技術革新と労務管理』『産炭地域の住民生活と社会保障』(学术论文), 「石炭関係新聞切り抜き」「北炭夕張新鉦事故関係」(新聞記事), 『石炭化学工業開発

調査報告書』『夕張市総合開発計画』(行政文書), 『石炭国有化共闘前進のために』『石炭政策に関する要望書』(団体文書)など、500点以上の「調査チーム参照資料」が存在する。そこには、当時の炭鉱労働者の「労働一生活過程」を知る上で貴重な一次資料である、作業日誌・操業証・給与明細書・家計簿といった生活記録も含まれている。こうした「調査チーム参照資料」について、SORDでは「作成者」「作成年月日」「タイトル」「概要」「形態」「注記」からなるデータベース化と(図1), フラットファイル・ボックスファイルを使用した大・中・小項目への体系的分類の作業を進めている(図2)。この作業が完了すれば、1960~80年代にかけての斜陽期の北海道石炭産業の、資料室・図書館としての役割を果たすことも可能となろう<sup>(9)</sup>。



図2 ボックスファイルに入れられた調査票・コーディングシート

## 2.2 調査票のアーカイブ化

第二は、「調査チーム作成資料」に関する〈アーカイブ的機能〉である。「調査チーム作成資料」には、調査票・コーディングシート以外にも、フィールドノート・学会報告・論文草稿・報告書などが含まれているが、それらについては、上記の「調査チーム参照資料」に準じたデータベース化と体系的分類を行うことが望ましい。むしろ、ここで問題となるのは〈調査票のアーカイブ化〉と〈調査データのデジタル・アーカイブ化〉のための方法論である。

一般的に言えば、量的調査の場合、SPSS・SAS等の統計パッケージに合わせたデータ処理によって、調査票はその役目を終え、廃棄されることがほとんどである。質的調査の場合も、この事情は似たり寄ったりで、研究プロジェクトの終了とともに調査票の死蔵・散逸が生じることが多い。その上、質的データは、データそのものの質に精粗があるため、一元的なデジタル化に適しているとは言いがたい。調査票が、そのデータが別の媒体へと置き換えられぬままに廃棄されることによって、ある時代・ある地域の貴重な歴史的証言は、半永久的に失われるかもしれないのである。

以上のようなデータの散逸・廃棄を防ぐた

めには、〈調査票のアーカイブ化〉と〈調査データのデジタル・アーカイブ化〉が必要不可欠である。それはいかにして可能か？ まず、〈調査票のアーカイブ化〉について、夕張調査で用いられた「夕張市民(の)生活実態調査」票の整理・保管に即して述べておきたい。

のべ115世帯、224人におよぶ炭鉱労働者とその家族の生活史は、実査から約30年の間、北大・生活教育研究室の中で、調査票は1組ずつ茶封筒に入れられた状態、コーディングシートは段ボール箱の中に積み重ねられた状態で保管されてきたようである。このように、必ずしも保存状態が良好だったとは言えないものの、一番上に重ねられたコーディングシートが色ヤケを起こしている以外に、目立った汚損・破損は見当たらない。しかし、ほとんどの調査票・コーディングシートは鉛筆書きで記されているため、仮にこれらの資料がそのまま2次分析に供された場合、原票を手にするたびに文字情報が不鮮明になることも予想される(この点に、〈調査票のアーカイブ化〉にとどまらない、〈調査データのデジタル・アーカイブ化〉の必要性がある)。

夕張調査の場合、〈調査票のアーカイブ化〉には(狭義の)「調査票のアーカイブ化」と「コーディングシートのアーカイブ化」という2つの段階が含まれる。調査票のアーカイブ化とは、もともとA票(世帯対象)・B票(主人対象)・C票(婦人対象)を一組にして、被調査者名を付した茶封筒に入れられていた調査票を、「炭鉱・階層」によって分類した上で(「北炭・鉱員」~「三菱・組夫」)、A票・B票・C票ごとにクリアファイルに入れ直す作業を中心とする(図3・図4)。上記のファイリングによって、調査票の整理・保管方針が「炭鉱・階層一被調査者単位」から「炭鉱・階層一調査票単位」へと変わった点に注意されたい。そして、調査票の混在・散乱を防ぐために、それぞれの調査票の1枚目の右肩には(同一の炭鉱・階層内で通しの)ナンバリングを

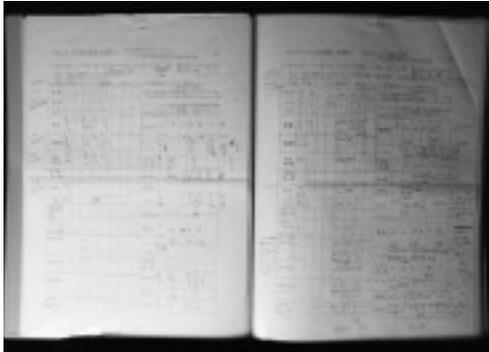


図3 クリアファイルに入れられた調査票

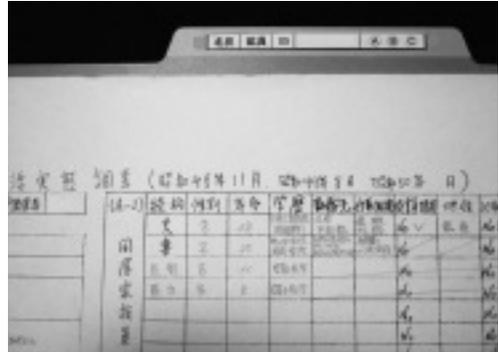


図5 個別フォルダに入れられたコーディングシート



図4 クリアファイルに入れられた調査票（北炭・鉱員）

施すとともに、「調査票番号」「被調査者名」「備考欄」からなる一覧表を作成している。同一の番号・被調査者名の調査票が複数部ある場合（下書きと清書など）、あるいは調査票とは別に調査メモ等の添付資料がある場合には、この一覧表の備考欄に記入する。以上の作業によって、調査票原票の可視性の確保と、その一元的な管理が可能になるといえよう。

一方、コーディングシートについては、「炭鉱・階層一調査票単位」の調査票とは対照的に、「炭鉱・階層一被調査者単位」の整理・保管を行っている。この作業の具体的手順としては、A票・B票・C票の調査データが書き写されたコーディングシートを、被調査者ごとに個別フォルダに挟んで綴じるとともに（5～6枚1組）、個別フォルダの見出しに、「炭鉱」「階層」「調査票番号」「被調査者名」から

なるメタ情報を記入していく（図5）。

以上が「調査票のアーカイブ化」の要諦であるが、この整理・保管法が2つの問題点を抱えていることを付言しておこう。第一は、調査票とコーディングシートの間の、整理・保管方針の違いである。現在のところ、調査票は「炭鉱・階層一調査票単位」、コーディングシートは「炭鉱・階層一被調査者単位」でファイリングされているが、今後の運用状況の如何によっては、両者の統一が必要となるかもしれない。その場合、調査票の整理・保管法を「炭鉱・階層一被調査者単位」へと変えることが、有力な代替案になるだろう。第二は、調査票とコーディングシートの双方に、被調査者名などの個人情報を書き残されている点である。原資料をマスキングすることで匿名性を確保することは可能であるが、それによって資料の歴史的価値が損なわれるとも考えられる。「社会調査史の博物館」としての一般公開に際しては、他のアーカイブの状況を参考にしながら、調査票原票の閲覧範囲、複写・貸借の可不可に関する慎重な議論が必要となる。

### 3. 調査データのデジタル・アーカイブ化

#### 3.1 コーディングシートの画像化

次に、「調査データのデジタル・アーカイブ化」についてである。調査データのデジタル

化の目的は、調査票に記された文字情報をデジタル形式へと移植することによって、原票の劣化・散逸の防止とデータの長期保存を可能とし、もって2次分析を含めた多様な利用に資することにある。質的データのデジタル化にあたっては、フォーマット化された量的データの場合にはみられないような、(調査票が)「意図せざる回答」のデジタル形式への組み込み・取り込みが課題となる。また、被調査者の回答は、セル内だけでなくセル外にまで及んでおり、図表・計算式・チャート等が調査票の欄外余白に記されていることも少なくない。こうした非定型なデータを、「定型性」と「再現性」のバランスを図りつつ、いかに媒体変換していくか——これが質的データをめぐる〈非定型の定型化〉という問題構制である。そして、この〈非定型の定型化〉のノウハウは、調査票の特性に応じて異なったものとならざるをえない。

さて、「夕張市民(の)生活実態調査」票の場合、次の点において〈非定型の定型化〉の作業がより重要性を帯びてくる。これは、調査票を介した「労働—生活過程」への接近という、布施グループの調査スタイルに大きく由来する問題でもある。のべ115世帯の炭鉱労働者とその家族の生活史は、構造化された調査票の枠組みによっては把握しきれない地平を持つものである。否、欄外余白の自由記述にこそ、追体験的・了解的に(再)構成すべき被調査者の生活の機微が記されているともいえる。「住居の間取り」設問に対する炭鉱住宅の間取りのスケッチ、「給与と保険・年金」設問に対する1ヶ月間の家計の収支一覧、「来夕の時期・理由」設問に対する調査時点までの職業経歴の年表などは、その一例であろう。このような外延を有するデータをデジタル化するには、複数のデジタル形式へと変換することによって、「定型性」と「再現性」の双方を担保することが求められる。

そこで、われわれは、〈調査データのデジタ

ル・アーカイブ化〉の作業を、次の2つの段階に分けて考えることとした。第一段階は「コーディングシートの画像化」、そして第二段階は「コーディングシートのテキスト化」である<sup>(10)</sup>。画像データおよびテキストデータという異なる水準のデジタル技術を用いることで、前者によって「再現性」の条件(=コーディングシートに視覚的に表された論理構造の担保)を、後者によって「定型性」の条件(=コーディングシートに記された文字群の規格化・一律化)を、それぞれ満たすことが可能となる。なお、画像化・テキスト化の対象物を(調査票ではなく)コーディングシートとした経緯については、布施グループの調査スタイルとも関連する、若干の説明が必要であろう。

「夕張市民(の)生活実態調査」票は、面接調査時に調査者自身の手で速記されたものであり、判読しがたい文字が散見される。また、記述が調査票の表裏両面にわたっている場合、調査者によって新たな用紙が追加されている場合もあり、調査票としての斉一性に乏しい。一方、コーディングシートは、恐らくは研究リーダーであった布施鉄治、あるいは岩城完之・小林甫など、調査団のコアメンバーによる未記入箇所のチェック等を経た上で、調査票の欄外余白の自由記述も含めて書き写されたものである。布施グループによる当時の分析においても、調査票ではなく(調査票に比べてハンディな)コーディングシートを用いていたことが推測される。そこで、われわれも、調査票ではなくコーディングシートに基づいて、画像化・テキスト化の作業を進めることにしたのである。

〈調査データのデジタル・アーカイブ化〉の第一段階は、「コーディングシートの画像化」である。ここでの画像化とは、コーディングシートを1枚ずつスキャンし、TIFF形式で保存し、識別可能なファイル名を施す作業を意味している。TIFF (Tagged Image

File Format)とは、アプリケーションに依存しない、汎用の画像交換フォーマットであり、保存を繰り返しても画質が劣化しない点に特徴がある。なお、スキャンされたコーディングシートの大きさは6400×4600ピクセル、サイズは28MBであることから(平均)、アーカイブとしての実際の運用にあたっては、JPEG等への圧縮が必要となるだろう。また、ファイル名には、「a\_hokutan\_koin\_01\_1」のように、「調査票種類(A票=a/B票=b/C票=c)」「炭鉱(北炭=hokutan/三菱=mit-subishi)」「階層(鉱員=koin/職員=shokuin/組夫=kumifu)」「調査票番号(01~)」「表裏(表=01/裏=02)」からなるメタデータを含めることで、保管・検索の便宜を図っている。以上が「コーディングシートの画像化」の具体的手順である(図6・図7)。

### 3.2 コーディングシートのテキスト化

そして、〈調査データのデジタル・アーカイブ化〉の第二段階が、「コーディングシートのテキスト化」となる。ここでのテキスト化とは、コーディングシートに記された文字情報を、データベースソフトに入力していく作業を意味する(なお、夕張調査データのテキスト化ではFile Maker Pro 8を使用している)。データベースソフトの利点として、リレーショナルデータの構築が可能であること、CSV・XML等の汎用形式への変換が可能であること、の2点が挙げられよう。

テキスト化のための最初の作業は、データベースソフト上での「入力フォーム」の構築である。一見したところ、コーディングシートのセルに(機械的に)対応させながら、「入力フォーム」を作成すれば良いと思われるかもしれないが、実際のところ、そうは行かない。それは、(調査票および)コーディングシートが、「セル数の限界」「意図せざる回答」という困難を抱えているためである。機種依存文字の禁止、半角・全角の区別といった基本

的な「入力ルール」であれば、あらかじめ定めることもできよう。しかし、「セル数の限界」「意図せざる回答」といった問題を解決する、〈非定型の定型化〉のための「入力ルール」は、以下のように、まさに遂行的に決めざるをえないのである。「入力フォーム」の生成と実際の入力作業において、これらの問題にどのように対処すれば良いのだろうか?<sup>(1)</sup>

「セル数の限界」問題は、「生殖家族の構成」「定位家族の構成」設問に多くみられるものである。同居家族員と他出者について尋ねた「生殖家族の構成」では7列(他出者については4列)、夫・妻の両親・兄弟について尋ねた「定位家族の構成」では8列のセルがコーディングシート上に用意されている。しかし、とりわけ「定位家族の構成」は当時の大規模な家族のありようを反映して、しばしば既定のセル数をオーバーしてしまう。その場合、コーディングシート上では1つのセルを2分割して記入したり、欄外余白に記入することによって、「セル数の限界」問題の解決が図られているのだが(図6-①)、こうした曖昧な記入法を「入力フォーム」上で再現することは、ほぼ不可能である。そこで、「入力フォーム」のセル数に関しては、全てのコーディングシートを精査した上で、回答の最大値に応じて決定する必要がある(最終的に、「生殖家族の構成」「定位家族の構成」のセル数は15列に設定。「炭鉱災害と扶助」についても5列に設定。図8-①)。

一方、「意図せざる回答」問題は、大きく3つに分けて考えることができる。第一は、番号・数字で回答するように指示されている設問において、それ以外の文字列が記入されている場合である。面接調査の際には、選択肢の書かれたカードが用いられたものの、コーディングシート上では、番号・数字が入るべきセルに単語・文書が記されているケースが数多くみられる。例えば、「現物給与の種類」設問は、住宅・その補修・石炭・水道・電気



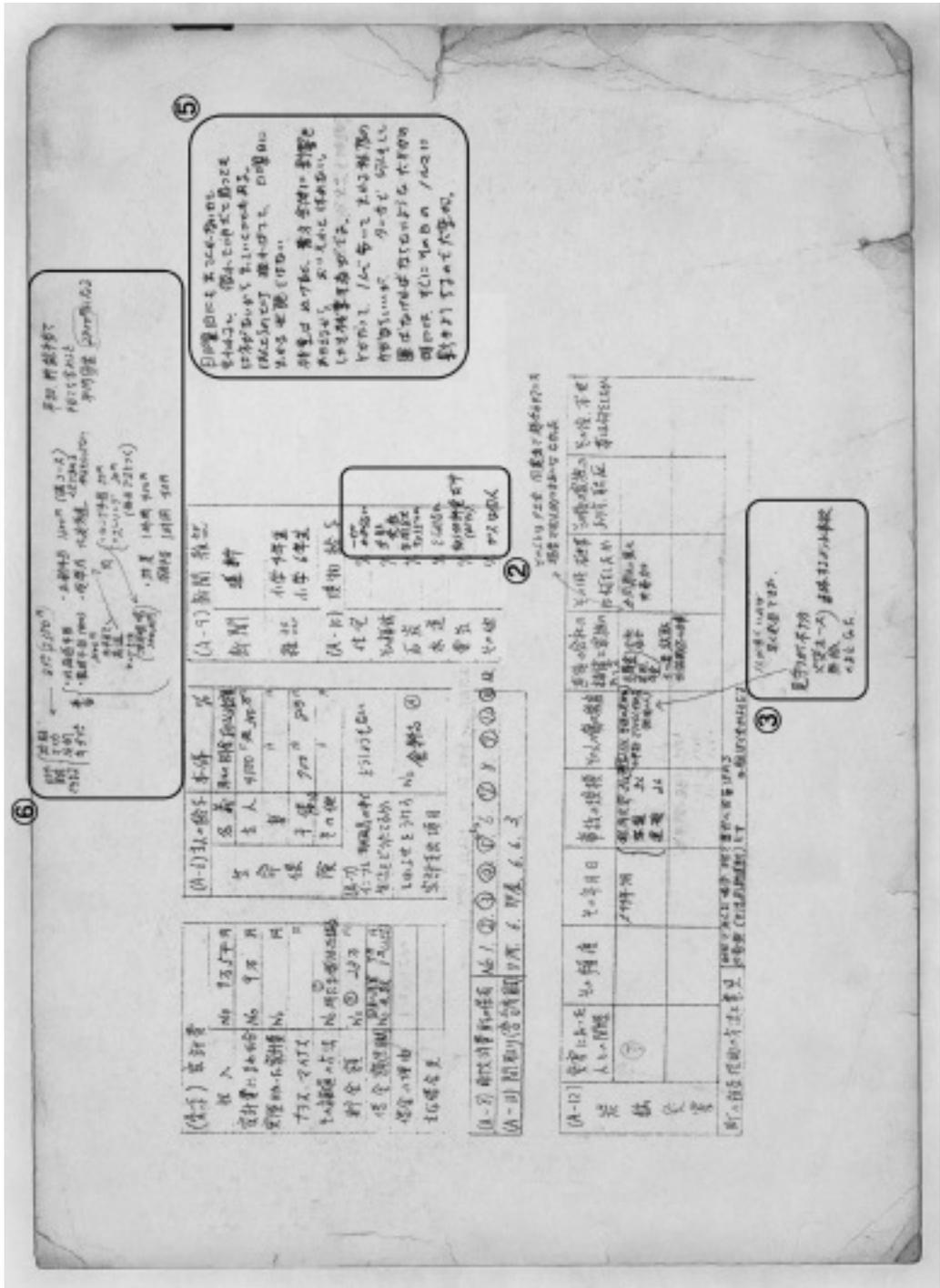


図7 「夕張市民生活実態調査」A票コーディングシート（裏面）





などに関する現物給与の割合を、パーセンテージで回答することが期待されているが、実際には、具体的な金額・分量と、それについての説明が書かれていることが多い(図7-②)。そうした単語・文書による回答を、ここでは、「小問に関連づけられた自由記述」と名づけておきたい。「夕張市民(の)生活実態調査」票のテキスト化にあたっては、この「小問に関連づけられた自由記述」を、そのまま「入力フォーム」の当該セルに打ち込むことにしている(図9-②)。それゆえ、2次分析にあたっては、分析者自身によるアフターコーディングが必要となる。

第二は、コーディングシートの欄外余白に書き残された自由記述である。自由記述には、設問(小問/大問)への(客観的ないし主観的)関連づけがみられるものと、そうではないものがある(ここで、客観的関連づけというのは、記入者自身が線・矢印を用いて設問との結びつきを示している場合であり、主観的関連づけというのは、入力者によって設問との意味上の結びつきが読みとれる場合である)。それぞれ、「小問に関連づけられた自由記述」「大問に関連づけられた自由記述」「設問に関連づけられない自由記述」と名づけておきたい。このうち、欄外余白に書かれた「小問に関連づけられた自由記述」(図7-③)；「炭鉱災害と扶助」設問)は、先述の、セル内に書かれた「小問に関連づけられた自由記述」の扱いに準じて、関連性の強いセルに打ち込むこととした(図9-③)。一方、「大問に関連づけられた自由記述」(図6-④)；「生殖家族の構成」設問)については、大問ごとに自由記述用のセルを作成し(図8-④)、他方、「設問に関連づけられない自由記述」(図7-⑤)についても、コーディングシートの表面・裏面ごとに自由記述セルを作成し(図9-⑤)、それぞれ、当該のセルに打ち込むことにした。

しかし、以上の〈非定型の定型化〉の入力ルールによっても、なお入力されずに残され

るものがある。それが、「図表」「計算式」「チャート」といった、第4の自由記述である。これらの文字・図形情報は、「入力フォーム」への“ベタ打ち”では意味内容が失われてしまうが、さりとて(テキスト化に準じた)何らかの入力作業をしないわけにはいかない。というのも、「給与と保険・年金」設問に対する給与明細表(図7-⑥)などは、当時の炭鉱労働者の「労働—生活過程」の重要な部分を表現しているからである。そこで、データベースソフトでは入力しづらい図表・計算式・チャートに関しては、コーディングシート上のオリジナルな記述を再現できるように、Microsoft Word(の罫線・図形機能)を用いて入力することとした(図10)。そして、「入力フォーム」側でもワード入力の有無が確認できるよう、コーディングシートの表面・裏面ごとに注釈チェック・注釈内容のセルを作成したのである(図9-⑥)。結果的に、ファイルメーカーとワードの二本立てにはなったものの、注釈セルの設定によって、ファイルメーカー上でのデータの一元的管理が可能になるといえよう。

以上が〈調査データのデジタル・アーカイブ化〉=〈非定型の定型化〉の諸階梯であるが、ここには2つの問題が残されている。第一は、画像データ・テキストデータの双方に、被調査者名をはじめとする個人情報が残されている点である。これは、〈調査票のアーカイブ化〉にみられるのと同様の問題であり、データの公開・提供に際しては、画像データ・テキストデータともに匿名性の確保が必要となる。第二は、自由記述のテキスト化が含まざるをえない「入力者の主観性」である。この「入力者の主観性」には、設問と欄外余白の自由記述との意味連関の読みとり、図表・計算式・チャート等のワード入力時の再現具合、などが含まれる。大学院生クラスによって構成された作業チームは、適宜、データ入力に関する共同討議を行ってはいるものの、それでも



図10 ワード上に入力された自由記述（「給与と保険・年金」設問）

テキスト化（のプロセス）には若干の曖昧性が残らざるをえない部分がある<sup>(12)</sup>。2次分析に際しては、分析者自身による画像データとテキストデータの照合作業が是非とも必要であることを、ここでも付言しておきたい。

### 3.3 〈読み取り装置〉の解釈＝再構築

〈調査データのデジタル・アーカイブ化〉の第二段階にあたる調査票（コーディングシート）のテキスト化は、これまで論じてきた「入力フォーム」の構築・「入力ルール」の生成という試行錯誤のプロセスにみられるように、「難航を約束された課題」（SORDメンバーの言葉）でもあった。この困難は、欄外余白に記された自由記述の膨大さに（直接的に）由来すると思われるのが普通であろう。「小問に関連づけられた自由記述」「大問に関連づけられた自由記述」「設問に関連づけられない自由記述」の入力にあたっては、入力者自身が意

味内容を読み取り、判断を加えなければならぬからである。しかし、問題を掘り下げて、社会調査という学問行為の中に位置づけ直してみると、以上の困難は、調査者と被調査者のインターフェイスとなる調査票をめぐる、〈読み取り装置〉のはたらきによる部分が大きいことが分かる。

コーディングシートの欄外余白に書かれた自由記述は、一瞥したところ、小問・大問の欄内に書かれた記述と同じように、自由記述として一律的に入力することが容易だと思える種類のものである。実際、われわれの入力作業においても、開始当初は別立てのテキストファイルにデータ化することにしていただが、作業チームにとって、このような機械的処理（＝欄内記述はデータベースソフトに入力し、欄外記述はテキストファイルに入力する）は非効率であると同時に、違和感を覚えるものであったという。それは、コーディ

ングシートに無秩序に書かれている（かのよ  
うに思われる）文字群が、少なくとも当時の  
調査集団にとっては、彼ら／彼女らの身体に  
埋め込まれた〈読み取り装置〉（＝マルクス主  
義的な社会変革論とアメリカ社会学の小集団  
論をベースとする）によって秩序立てて配置  
されていたことに由来する。それゆえ、欄内/  
欄外による入力方法の区別は、調査団に身体  
化されていた〈読み取り装置〉を無化しかね  
ないのである。

ここに、調査者の意図と自由記述の布置連  
関の探求を通してオリジナルな〈読み取り装  
置〉を解読し、それを踏まえてデータベース  
ソフト上に〈読み取り装置〉を再構築する必  
要が生じるといえる。実際の入力作業を介し  
た「入力フォーム」の構築・「入力ルール」の  
生成という、数次にわたる反照的なプロセス  
が示しているのは、〈読み取り装置〉の解読＝  
再構築の足跡なのである。

とはいえ、オリジナルな〈読み取り装置〉  
とリニューアルされた〈読み取り装置〉は、  
それぞれ性格を異にする。前者は、被調査者  
の「声」を社会（学）理論の鋳型に照らして  
「論理」化し、それを学会報告・発表論文とい  
ったアウトプットへと水路づけるためのもの  
である。一方、後者は、調査団のオリジナルな  
認識＝解釈枠組みを踏襲しつつも、そこから  
比較的自由的な観点において、当時の調査票と  
データ利用者を“つなぐ”役割を果たすもの  
である。

しかし、リニューアルされた〈読み取り装  
置〉をめぐる事情は単純ではない。ここでは  
2つの問題点を指摘しておこう。第一は、入  
力者が“つなぐ”相手であるデータ利用者が  
不確定な点である。量的データの場合、  
SPSS・SAS等の統計パッケージによるデー  
タ処理がほとんどであるのに対して、質的  
データの場合には「誰が、どのようにデー  
タを使うのか」にプラットフォームが依存す  
る度合いが高く、実際の用途が未決定な現時

で「入力フォーム」の妥当性を検証すること  
は不可能である。第二は、データ入力の作業  
がアウトプットと関連づけにくい点である。  
量的データの2次分析が有するある種の“取  
り組みやすさ”とは対照的に、質的データの  
場合には調査団の手になる記述の多元化・豊  
穡化をめざした〈調査票空間〉のテーマ化、  
調査史・実践史としての社会学史の描出にむ  
けた〈調査空間〉のテーマ化というように、  
2次分析の方向性は迂回的なものとならざる  
をえないのである（この点については後述す  
る）。

#### 4. むすび

本稿は、北大・布施グループによる夕張調  
査の資料群を対象として、調査資料のデー  
タベース化と調査票・調査データのデジタル化  
からなる〈質的データのデジタル・アーカイ  
ブ化〉の実践的方法論を述べたものである。  
小論を閉じるにあたって若干の内容的なまと  
めを行うとともに、〈読み取り装置〉をめぐる  
議論を踏まえて、ありうべき2次分析の方向  
性に関する示唆を与えたい。

これまで論じてきたように、「社会調査史の  
博物館」は〈資料室的機能〉と〈アーカイブ  
的機能〉の2つを有する。このうち〈資料室  
的機能〉にあたるのが「調査チーム参照資  
料」・一部の「調査チーム作成資料」につい  
てのデータベース作成と、その体系的分類で  
ある。一方、〈アーカイブ的機能〉にあたるのが  
「調査チーム作成資料」に関する〈調査票の  
アーカイブ化〉と〈調査データのデジタル・  
アーカイブ化〉である。〈調査票のアーカイ  
ブ化〉とは調査票のファイリング作業（による  
整理・保管）を意味し、〈調査データのデジ  
タル・アーカイブ化〉の作業は調査票の画像化・  
テキスト化から構成される。その結果、調査  
票は原票・画像データ・テキストデータとい  
う3つの水準（2つのデジタル形式）におい  
て保存されることになる。画像データとテキ

ストデータというバックアップの存在によって、調査票そのものの重要性は低くなるかもしれないが、保管場所のキャパシティが許すかぎり、調査票の“現物”が伝える息吹や肌理を大切にしていくことが「社会調査史の博物館」の(教育的)使命といえるだろう。

さて、以上の作業を通して、「夕張市民(の)生活実態調査」から、どのような2次分析が可能になるだろうか？ ここでは、2つの方向性を覚書風に記しておきたい。

第一は、調査票と調査データの批判的解読による記述の多元化・豊穡化である。画像データ・テキストデータと論文・報告書とのクリティカルな対話を通して、『地域産業変動と階級・階層』をはじめとする、布施グループのオリジナルなアウトプットとは異なった調査対象像が浮かび上がるだろう。やや卑近な例かもしれないが、この作業を介して、財政再建団体となった今日の夕張の「根っこ」のようなものを、炭鉱閉鎖の危機にあった約30年前の夕張に、あるいは見い出せるのかもしれない。

以上が〈調査票空間〉に内在した2次分析であるとすれば、第二のそれは〈調査空間〉そのものをテーマ化するものだといえる。調査票とそのデータにとどまらず、一連の「調査チーム参照資料」やフィールドノート、草稿を精査することによって、理論史・学説史としての社会学史ではなく、調査史・実践史としての社会学史を描き出すことができるだろう(佐藤2006)<sup>(13)</sup>。布施グループであれば、「労働—生活過程」分析と「社会機構—社会構造」分析の接続化というフレームワークが、対象と方法に即して、その妥当性を問われることになるといえる。

このような2次分析の方向性を示すことで、ひとまずの結びとしよう。

## 注

(1) SORD (Social and Opinion Research

Database) プロジェクトは札幌学院大学社会情報学部の研究プロジェクトであり、平成18年度から『社会調査史の博物館』としてのリジョン拠点データアーカイブの構築(平成18~21年度文部科学省科学研究費補助金基盤研究(B))をテーマとした研究活動を行っている。筆者は平成18年7月から本研究プロジェクトのリサーチアシスタントを務めており、本稿は上記の科研費による研究成果の一部である。このように貴重な機会を与えていただいた札幌学院大学社会情報学部に対して、心より感謝申し上げたい。

- (2) とりわけ全国各地の(中小の)美術館・博物館が収蔵物のデジタル・アーカイブ化を進めている背景として、文化遺産・歴史遺産をまちづくり・地域振興に結び付けようとする側面を指摘できる(笠羽2004)。このような動きについては、「集合的記憶」(Halbwachs 1950=1989)などの観点から、(文化資源学的に)なお詰めるべき論点があるように思われる。
- (3) 「生活社会学研究会」の調査資料は小林甫・前北海道大学教授から、一方、「地域政策研究会」の調査資料は似田貝香門・前東京大学教授から、それぞれ寄託を受けたものである。この場を借りて、両先生には心より感謝申し上げます。
- (4) 質的データのアーカイブ化の試みとして、富士常葉大学における「飯島伸子文庫」の開設、都市社会学会による「特殊飲食店女子組合員調査」データ(磯村英一による)の電子化を挙げることができる。前者については船橋(2007)、後者については玉野(2001)を、それぞれ参照のこと。
- (5) 本稿はSORDメンバーの共同討議・共同作業にもとづくものであり、とりわけ中澤秀雄(千葉大学)・西城戸誠(法政大学)・祐成保志(信州大学)の各氏との議論に負っている部分が少なくないことを、ここに記しておく。
- (6) 地域社会学の構造分析の系譜における布施グループの位置づけについては小内(1995)・

中澤(2007), 北海道の社会調査という文脈における夕張調査の知識社会学的検討については新藤(2005)を, それぞれ参照のこと。

- (7) この点については, 数十冊におよぶ小林甫氏の研究ノートを踏まえて, <夕張調査の展開過程図>を作成することを検討中である。
- (8) 以下の<資料室的機能>と<アーカイブ的機能>については, 木村・林(2006)を参照のこと。
- (9) この分野の先達として, 慶應義塾大学図書館の「日本石炭産業関連資料コレクション」を挙げる事ができる。これは, 昭和前期から平成の炭鉱閉山に至るまでの, 北海道の大手炭鉱(三菱・三井・住友・北炭・太平洋)の資料群を取蔵するものである。SORDのようなリージョンに特化した中小アーカイブにとっては, とりわけ大規模アーカイブとのネットワーク構築が重要なものとなる。http://www.mita.lib.keio.ac.jp/coal/を参照のこと。
- (10) 原純輔は, 非定型データの処理・分析の一方法として, 画像データへの変換・蓄積を挙げている。テキスト化されたデータとコード化されたデータは, 既存の統計パッケージでは関連づけにくいという技術的問題が存するため, 調査票の画像データ化には「解決の先延ばし」という意味もあったという(原2001)。夕張調査データのデジタル・アーカイブ化においても, 2次分析に際して(テキストデータの)アフターコーディングが必要とされるように, 上記の技術的問題が解決されたわけではない。しかし, NVivoのような定性データ分析ソフトの登場によって, こうした膠着状況に変化がもたらされる可能性も小さくはないだろう。
- (11) コーディングシートのテキスト化にあたっては, 庄司知恵子(北海道大学大学院博士課程)・吉野航一(同博士課程)・寺沢重法(同修士課程)の各氏に, 面倒な入力作業を引き受けていただいた。また, 久保ともえ氏(SORD事務局長)には, コーディングシートのスキャニング・コピーなど, テキスト化の下準備にあた

る煩雑な作業を手伝っていただいた。記して感謝申し上げたい。

- (12) ファイルメーカーとワードのいずれによっても, コーディングシートの記述内容を十分に表現しづらい場合, コーディングシート上に判読しがたい文字が書かれている場合には, 注釈セルを利用して入力者の「コメント」を添えることにしている。
- (13) アーカイブズ学の世界では, 公文書の所蔵場所が確保しづらくなるにつれて, アーキビストの役割が「記録の保存」から「保存すべき記録の選定」へと移っていった。それとともに, 「過程連係情報(process-bound information)」, すなわち情報を生み出し構造化するプロセスへの探求が始まったようである。以上のアーカイブズ学の展開は, SORDの「社会調査史の博物館」の構想に対する問題提起であると同時に(取蔵すべき調査資料の選択/排除), 調査資料の検討を通して調査史・実践史をテーマ化しようとする2次分析の方向性と, 興味深い符合を示している。Thomassen(2001=2006)およびCook(1997=2006)を参照のこと。

#### 参考文献

- Cook, T. (1997) What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift, *Archivaria*, Vol. 43: 17-63 = (2006) 塚田治郎(訳)「過去は物語の始まりである——1898年以降のアーカイブズ観の歴史と未来へのパラダイムシフト」記録管理学会・日本アーカイブズ学会編『入門・アーカイブズの世界——記憶と記録を未来に』日外アソシエーツ: 117-186
- 船橋晴俊(2007)「飯島伸子文庫の設置経過と内容」『日本及びアジア・太平洋地域における環境問題と環境問題の理論と調査史の総合的研究』平成15~18年度文部科学省科学研究費補助金研究成果報告書: 4-29
- 布施鉄治編(1982)『地域産業変動と階級・階層——炭都・夕張/労働者の生産・労働——生活

- 史・誌』御茶の水書房
- Halbwachs, M. (1950) *La mémoire collective*, Presses universitaires de France = (1989) 小関藤一郎 (訳) 『集合的記憶』行路社
- 原純輔 (2001) 「文化資源学の創設をめざして」『社会情報』11(1) : 9-26
- 笠羽晴夫 (2004) 『デジタルアーカイブの構築と運用 — ミュージアムから地域振興へ』水曜社
- 木村玲欧・林能成 (2006) 「先例調査にもとづく市民防災教育を視野にいたした災害アーカイブの立ち上げ」『アーカイブズ学研究』5 : 94-111
- 中澤秀雄 (2007) 「地方自治体『構造分析』の系譜と課題 — 『構造』のすき間から多様化する地域」蓮見音彦編『講座社会学 3 村落と地域』東京大学出版会 : 169-205
- 小内純子 (1995) 「地域社会学における総合的実証研究の意義と今日的課題」『北海道大学教育学部紀要』65 : 167-181
- 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
- 佐藤健二 (2006) 「地域社会のリテラシー — 調査史に学ぶ」似田貝香門監修『地域社会学講座 1 地域社会学の視座と方法』 : 213-242
- 新藤慶 (2005) 「産炭都市夕張の社会学的研究 — 布施鉄治編『地域産業変動と階級・階層』の知識社会学的検討を中心として」『社会情報』14(2) : 319-331
- 武邑光裕 (2003) 『記憶のゆくたて — デジタル・アーカイブの文化経済』東京大学出版会
- 玉野和志 (2001) 「『特殊飲食店女子組合員調査』のデータについて」『日本都市社会学会年報』19 : 219-229
- Thomassen, T. (2001) A first introduction to archival science, *Archival Science*, Vol. 1: 373-385 = (2006) 石原一則 (訳) 「アーカイブズ学入門」記録管理学会・日本アーカイブズ学会編『入門・アーカイブズの世界 — 記憶と記録を未来に』日外アソシエーツ : 47-64