SSJDA－094

社会調査・データアーカイブ共同利用・共同研究プロジェクト

古い社会調査データの復元・アーカイブ化における諸問題

「児童養育費調査 北会津 農村調査」の作業事例

那須蘭太郎（東京大学）[1]
入山浩一（東京大学）[2]

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

2025 年（令和 7 年）5 月

[1] 東京大学社会科学研究所学術専門職員，nasu-ran@iss.u-tokyo.ac.jp
[2] 東京大学社会科学研究所技術専門職員，iriyama@iss.u-tokyo.ac.jp

古い社会調査データの復元・アーカイブ化における諸問題
「児童養育費調査 北会津 農村調査」の作業事例

那須蘭太郎（東京大学）・入山浩一（東京大学）

概要

本稿は，「児童養育費調査 北会津 農村調査，1966」を事例として，実施から時間が経過した古い社会調査データの復元作業とアーカイブ化作業の記録を残すことを目的としている。古い社会調査データを復元し，研究者間で適切な形で公開，アーカイブ化することは，「歴史」となった過去の事象，人々の生活，社会についての新たな知見の創出をもたらす。また，古い社会調査データに対し，社会学が発展させてきた方法を導入することで，歴史研究に新たな視点を導入することを可能とする。一方で，古い調査データのアーカイブ化は，資料の散逸を防ぎ，過去の研究成果の再現と，それを深化させることも可能にする。このような重要性をふまえ，本稿では，膨大に残されている森岡清美調査資料群の調査票原本の復元・アーカイブ化の端緒として児童養育費調査データを対象としたデータアーカイブ化作業を行った。その作業過程や生じた困難，過去の報告書との整合性の確認などを記録している。本稿の記録は，今後行われるであろうデータ復元作業とアーカイブ化作業を結び付け，古い調査データの適切な管理・共有を促すことで，広く社会科学におけるデータ蓄積の促進に寄与する。また，質的データが多く含まれる本調査資料群は，質的データのアーカイブ化の重要性を示しており，今後の議論の必要性を強調している。

1. 導入

本稿は，1966 年に実施された「児童養育費調査 北会津 農村調査」のデータ整備過程を事例に，古い社会調査データの復元・アーカイブ化・公開作業をする際の困難・課題やその意義をまとめることを目的とする[3]。

古い社会調査データは，過去の意識や行動に関連した問いに答える歴史的資料としての価値を持つ（佐藤 2009）。また一方で，歴史的問いに対して，社会学が蓄積してきた社会調査データを活用することは，社会学的な論理の投入と社会学におけるデータ処理の投入をもたらし，歴史学・社会学の双方の発展に寄与する（佐藤 1998）。特に，多様な資料・史料を利用する歴史学，または歴史社会学におけるミクロデータの活用が，当該学問分野の発展に資するものとして理解されてきた（相澤ほか 2013）。

このように，実施から長い時間が経った過去の社会調査データの学術的価値が指摘されてきた一方で，そのデータの管理・保管にも重要性がある。社会調査データの保存は，実証

[3] 本稿で事例として取り上げた児童養育費調査 北会津調査は，すでに SSJ データアーカイブから，「児童養育費調査 北会津 農村調査（世帯調査票），1966」（SSJDA 調査番号：1754）として公開されている。

研究における再現性の担保という点で価値があるとともに，データの提供は二次利用を促進し，研究資源の蓄積に貢献する（佐藤 2010）。一方で，特にコンピュータが普及する以前に実施された調査データは，紙の調査票が残るのみであり，これらの資料をデジタル化し，二次利用可能な形でアーカイブ化することは，データ利用の促進による新たな研究成果の創出に寄与する[4]。また，過去に実施されてきた社会調査データのアーカイブ化は，貴重な資料・史料の散逸を防ぐ役割を持つ。

つまり，古い社会調査データを復元し，アーカイブ化することは，第 1 に歴史的問いに対する社会学的な発想の投入を促し，社会学における「複数の変数間のより複雑な関係」（佐藤 1998: 2）への注目が，新たな知の生産を助ける可能性がある。また第 2 に，より資料保存の観点に拠るならば，調査データの散逸・消失を防ぎ，適切な管理と共有を行うことは，過去の論文の再検討や新たな観点の検証を可能とするという利点を持つ。

本稿では，「児童養育費調査 北会津 農村調査」（以降，北会津調査）を事例とする。北会津調査は，児童養育費調査シリーズのうち，「児童養育費調査 川崎 労働者調査」（以降，川崎調査）に続くものとして，農村家族における児童養育費の実態を，家族周期（Family life cycle）の観点から明らかにすることを目的とした調査である（森岡 1967）[5]。これらの児童養育費調査シリーズは，のちに詳しく述べるが，森岡清美氏の残した膨大な社会調査データ「森岡清美調査資料群」（以降，森岡資料群）の一部である。

森岡資料群には，農村や家族，宗教に関する社会学的研究に資する膨大な社会調査が含まれる。とはいえ，その一部は調査情報が十分に残っていない，あるいは可読性の低い記入済み調査票しか現存していないなど，アーカイブ化する際に解決すべき課題が多いものも含まれる。そこでわれわれは，調査情報や調査票が比較的しっかりと残っており取り組みやすい，児童養育費調査のデータアーカイブから着手し，現存する膨大なデータを順次アーカイブ化していくことを試みた。森岡資料群が保有する膨大な調査データをアーカイブ化し，適切な方法と範囲で公開することで，資料散逸を防ぎ，新たな知見の創出を促進させることを狙いとした。

以上をふまえ，本稿では 1960 年代に行われた社会調査データのアーカイブ化を事例とし，その作業過程の詳細を示すとともに，作業において生じた困難・課題やその意義をまとめることで，実施から時間が経過した社会調査データのアーカイブ化を促進することを目的とする。

以降では，第 2 節で社会調査データの歴史的価値について，主に計量歴史社会学に関する研究群のレビューを行い，第 3 節で社会調査データのアーカイブが持つ役割とその重要

---

[4] 本稿における「アーカイブ化」とは，実施された社会調査データを適切な方法で収集・保管・保存するとともに，学術目的の二次利用のためのデータ提供にあたって必要となる一連の作業を指す。

[5] 「児童養育費調査 川崎 労働者調査（世帯調査票）, 1965」についても，同様のアーカイブ化作業を行い，すでに SSJ データアーカイブで公開されている（SSJDA 調査番号：1630）。

性について指摘する。第 4 節では，事例となる調査について紹介し，第 5 節で具体的なアーカイブ化作業の流れを説明する。第 6 節にて，生じた課題や困難についてまとめ，第 7 節では，調査が行われた当時に作成された報告書と復元データの分布の整合性を検証する。最後に，第 8 節でまとめと議論を行う。

2. 社会調査データの歴史的価値

社会学，あるいは広く社会科学の分野では，社会調査データの二次分析が従来活発に行われてきたが，その際の問いの対象は，現代あるいは調査時点現在に限定されたものではなかった。「計量歴史社会学」と題されるディシプリンにおいては，社会調査データの二次分析をもとに，過去の現象を分析・理解することが目指された。『理論と方法』で組まれた計量歴史社会学の特集における佐藤（1998）の導入は，計量歴史社会学の役割と試みを整理した嚆矢のひとつである。歴史というフィールドに対して計量社会学を投入することは，社会学が発展させてきた社会学理論とその高度な方法的フォーマライゼーション，そして計量的方法の導入という，いわば歴史と数理を出会わせる試みである（佐藤 1998）。これ以降，計量歴史社会学の進展はめざましい。2005 年には『社会科学研究』で歴史社会学の特集が組まれ，2024 年の『理論と方法』には会長講演として，渡邊勉による論考が掲載された（渡邊 2024）。歴史学研究の方法論を糸口とし，計量社会学や数理社会学と歴史研究との接点が見出されている（渡邊 2024）。

計量歴史社会学において，分析可能な社会調査データは必要不可欠である。特に実施から時間が経過した古い社会調査データを再分析することが，「歴史」となった過去の事象に対する方法的に適切な探究を可能とする。一方で，古い社会調査の多くは，コンピュータの普及以前に実施されたこともあり，紙の回答済み調査票の状態で現存しているものが多く，それらに対して二次分析を行うには，統計処理可能なデジタルデータに「復元」する必要がある。

計量歴史社会学を標榜する近年の研究プロジェクトは，そのような復元作業を広く行ってきた。最も中心的なものは，「東京大学社会科学研究所所蔵労働調査資料」（以降，労働調査資料）の復元であろう。労働調査資料は，東京大学社会科学研究所図書室に所蔵されていた資料群を指し，65 の労働調査が含まれている（労働調査論研究会編 [1970] 2001）。いずれの調査も，調査票の原票のみが保管されており，計量分析が可能なデータセットに復元する必要があった（佐藤 2025）。この労働調査資料の復元作業の詳細は，主に『社会科学研究』第 76 巻（2025 年）の「特集 社研労働調査資料のデジタル復元」に記されている。

古い社会調査データは，「歴史」となった過去の事象や人々の生活，社会を理解する際に極めて貴重な資料としてその重要性を保持している。さらに，計量歴史社会学が標榜するような歴史的資料への計量分析の導入は，社会学が発展させてきた理論・方法を歴史研究に持ち込むことを可能とし，新たな知見の創出に寄与するといえよう。こうした古い社会調査データの多くが，紙の調査票原票の状態で残されていることからも，復元作業は極めて大きな

価値を持った作業である。

3. 社会調査データのアーカイブ

一方で，社会調査データの復元作業は，その後のデータ管理の検討を要する。その際，データアーカイブは重要な役割を果たす。社会調査データのアーカイブ化，二次分析のためのデータ提供の流れは，1990 年代の終わりに端を発している。SSJ データアーカイブ（以降，SSJDA）は，社会調査の個票データを収集・保管するとともに，社会科学の実証研究を支援することを目的として，個票データの提供を行っている。

データをアーカイブ化し，それを広く利用可能にすることにはいくつかの利点がある。第 1 に，データの散逸を防ぐことにある。社会調査の個票データは，研究代表者の所属機関異動や退職，不適切なメディア管理，時代変化によりメディアの読み出しができなくなることなどのさまざまな背景から，散逸する危険に常に曝されている（三輪・佐藤 2018）[6]。これは，古い社会調査データに関して特に問題である。まず，実施から時間が経過した多くの社会調査データは紙媒体であり，さらに特に 1960 年代までの調査では，調査票に酸性紙が使用されていることが多く，経年劣化が激しい。調査票に記載された回答情報が読めなくなることによるデータの散逸の危険性がある。さらに，調査データを所有している研究者が，大学間の異動を経験したり，退職をしたりすることも，古い社会調査データにとっては，散逸をもたらしかねない，特に深刻な問題である。近年では，Data Management Plan（以降，DMP）の策定が要求され，調査データの管理・公開に関する事前の計画が求められているものの，戦後まもなく実施された古い社会調査においては，まだ DMP の発想が浸透していなかったこともあり，研究者の異動・退職などの契機によってデータが散逸してしまう危険性が十分にある。これに対し，デジタル処理されたデータをアーカイブ化し，専門の組織が適切に管理することは，散逸を防ぎ，貴重な情報が失われることを防ぐ。

第 2 に，新たな観点の導入による知見の創出にある。データアーカイブは，研究データを適切に公開することによって，二次分析を格段に容易にしてきた。調査当初に想定された仮説とは別の観点からデータが吟味されることは，新しい理論を構築する機会を提供する（佐藤 2012）。二次分析者は当然，調査設計の背景やねらいを理解している必要があるが，そのうえで導入される新たな視点は，既存のデータから読み取れる考察の幅を格段に増加させ，社会科学全体の知の生産に寄与する。実施から時間が経過した古い社会調査データに対する二次分析においては，特に研究者と調査実施者との時代的な距離が遠いことからも，過去の事象に対する現代からの問いの投影を可能とする。古い社会調査データを現代的な視点でみることは，このような重要性を持っており，それを可能とする二次分析の促進において，データアーカイブは重要な役割を担ってきた。

第 3 に，研究成果の再現・再検討にある。2024 年の『理論と方法』第 39 巻では，「社会

[6] 研究代表者の退職による資料の保存の危機に関しては，仁田（2025）が，労働調査資料の動向を極めて詳細に，リアリティを持った描写を行っている。

学において再現性とどう向き合うか」と題した特集が組まれ，その中で朝岡（2024）は，社会科学研究データの公開の潮流とその方法に関して議論を行っている。研究データの公開において，データアーカイブは大きく寄与してきたといえよう（三輪・佐藤 2018）。適切なデータ管理と公開は，研究の透明性を上げ，結果の再現性を向上させる。これは，古い社会調査に対する二次分析にも関連する。過去に実施された調査から得られた知見を，さらに深掘りし，社会現象に対する理解を深めるためには，第 1 に，当時導出された知見が再現可能であるかどうかが重要な問題となる。データアーカイブへのデータの保管と公開は，研究データの再現性を高め，議論の進展に寄与する。

このように，様々な調査実施者によって行われた社会調査データを収集し，広く適切に公開することは，学術的，社会的重要性を持つといえよう。以降では，古い社会調査データの復元作業・アーカイブ化作業を紹介することを通して，データ復元とアーカイブ化という作業を結び付け，今後復元される様々な社会調査データが適切に公開され，社会科学全体におけるデータ資源が豊富なものとなることを目指す。

4. 事例となる調査資料について

4.1. 児童養育費 北会津 農村調査について

本稿では事例として，児童養育費調査のうち，北会津調査を使用する。児童養育費調査は，3 地域で異なる時期に行われた調査によって構成されている。1965 年に神奈川県川崎市で行われた調査（川崎調査），その川崎調査に続くものとして，1966 年に福島県北会津村で行われた調査（北会津調査），1968 年に静岡県掛川市で行われた調査（以降，掛川調査）がある。

以下，中鉢編（1970）を参考に，児童養育費調査の目的や意義を説明する。児童養育に対する手当・扶助制度が少しずつ確立されていった 1950 年代から 1960 年代にかけて，異なる賃金体系に置かれる労働者が，どのような，またどの程度，児童養育による経済的圧迫を受けているのか解明することは急務であった。児童養育費調査は，年功的賃金体系に支えられる基幹産業の大企業常勤労働者（川崎調査），「かつての給源でありまたしばしばその経営家族主義の原型として指摘された農家複合家族」（中鉢編 1970: 226）（北会津調査），その中間に位置づけられる地方中小都市の常用勤労者家族（掛川調査）の異なった世帯の比較を通してこれを明らかにする目的で実施された。また，本調査の特徴の 1 つに，世帯の生活構造や児童養育の構造を，家族周期の観点から理解しようとする点がある。「およそ個人の生涯生活は，その出生，養育，就業，結婚，育児，退職，死亡のそれぞれの段階において，他の個人と不可避的な人格的関係をとり結び，その行動と意識の基本構造をつくりあげる」（中鉢編 1970: 3-4）という想定のもと，特定集団の生活構造が多様なコーホートの世帯主を持つ世帯の重層によって形成されているととらえている。本調査は，多様な生活構造を，異なる家族周期によって分解し，理解していくことを目的としているといえよう。

児童養育と生活構造の把握のために，児童養育費調査では世帯調査，家計調査（家計簿），

栄養調査の異なる 3 つの調査票を使用した調査を行っている[7]。各調査はそれぞれ家族の生活史と生活状況，消費構造，栄養状態に関する項目で構成され，それぞれで社会学，家政学，栄養学の専門家による検討が行われた。これら 3 つの調査のうち，森岡資料群としてアーカイブ化可能な資料は世帯調査に限定される[8]。したがって，以降のアーカイブ化作業はすべて世帯調査票で聴取された項目に限られる。世帯調査票では，主に子どもの状況，同居家族の状況，他出中の家族の状況，仕事の状況，世帯の生活状況，夫婦の状況などが聴取されている。

異なる賃金体系や生活を送る世帯間の比較を行うという研究目的上，異なる 3 地域で実施された調査結果の比較が必要とされる。初めに実施された川崎調査は，すでに SSJDA においてアーカイブ化作業が完了している（SSJDA 調査番号：1630）。掛川調査についても，今後復元作業が行われ，アーカイブ化される予定である。

北会津調査は，1966 年 8 月 18 日から 20 日にかけて，福島県北会津郡北会津村で行われた調査である。調査員の訪問による個別面接法に基づき，他記式調査票によって情報が聴取された。福島県北会津郡北会津村は，水田稲作農村であり，調査が実施された 1960 年代当時，水田率，農業専業率は全国的にみても高水準であった（中鉢編 1970）。したがって，「複合家族を維持しつつ安定した自作経営をつづけている農家世帯」（中鉢編 1970: 6）を標榜する北会津調査の対象としては適切であると考えられた。標本の抽出は，福島県北会津郡北会津村の水稲単作専業農家のうち，経営面積や家族構成などに基づいて対象が選定され，30 世帯の不備票を除く 124 世帯が対象となった。調査設計や調査過程などの詳細は，中鉢編（1970）に詳しく記載されている。

調査の有効回収数は 124 世帯と少ないが，聴取事項は詳細である。表 1 は，北会津調査（世帯調査）の主要調査項目である。主な調査事項は，児童・生徒の学校段階や発育状況，家族の構成や職業，別居家族の状況，世帯の所有材や所得，家族内での役割，農業収入や農地面積などの状況，その他家族に関する意識などが含まれている。特に，農業に関する聴取事項は豊富であり，農業所得の内訳や，所有する農業機器，雇っている労働者など，世帯の社会経済的背景を把握するのに十分な情報が得られている。

本データは，1960 年代の農村社会における児童養育費や生活の状況を把握することのできる貴重なデータである。また，他の児童養育費調査シリーズである川崎調査は大都市における大企業組織労働者家族，掛川調査は地方小都市における常用勤労者家族の状況をとらえている（中鉢編 1970）。したがって，これらの調査データと比較することで，1960 年代の児童養育費や生活の状況に関する異なった背景を持つ集団の比較が可能となる。

表 1　北会津調査の主要調査事項

[7] 川崎調査には，栄養調査は含まれていない。
[8] SSJDA に寄託された資料は，この世帯調査のみであった。

（1）児童の状況
性別,生年月,年齢・月齢,発達段階,学校種別,幼稚園・保育園種別,収入を伴う仕事の有無,収入の使い方,子供部屋の有無,自分専用の勉強机の有無,身長,体重,死亡した子どもの有無,死亡した子どもの性別
（2）家族の状況
続柄,出生年月,年齢・月齢,農業経営上の地位,家内労働・日雇・内職の有無,農閑期のみにする仕事の内容,学歴,卒業・在学の別,年間総所得,病気がちの家族の有無,病気がちの家族の続柄,現在病気の家族の有無,現在病気の家族の続柄
（3）別居家族の状況
続柄,現在の所在地（都道府県）,別居中の事由,仕送りの状況,帰省頻度,仕事の内容,職種
（4）世帯の状況
農地面積,農業所得合計,年払いにしている品物の有無,年払いにしている品物,年ぷ払いにしている品物の有無,年ぷ払いにしている品物,風呂場の場所,風呂の利用方法,便所の場所,台所の流しの蛇口の有無,炊事に使う燃料,冬暖房に使うもの,所有器具,もち家かどうか,部屋数,合計畳数,月ぎめ新聞の種類,日用品の買い物のお金を出し入れする家族の続柄,大きい買い物のお金を出し入れする家族の続柄,重大な出来事を中心的に決める家族の続柄,家の創立時代
（5）農業の状況
農業所得の内容,作る米の量,野菜を買うことの有無,買う野菜の内容,雇う人数,払う日当,住み込みでない年雇いの有無,住み込みでない年雇いの月平均賃金,自宅で作るか（味噌・醤油・つけもの）,現在使っている農業器機,借りて使っている農業器機
（6）夫婦の状況
結婚年月,結婚時年齢,結婚後の姓,夫婦の出生順位,夫婦のきょうだいの生存・死亡の別,夫きょうだいの性別,夫婦の出身地,夫婦の両親の存否,夫婦の親との同居経験,子どもの進学見通し,長男に希望する職業,追加出生意欲,理想だと思う子供数,妊娠中の家族の有無,妊娠中の家族の続柄
（7）その他
臨時収入があった場合の使用用途,最も近い生活の仕方
（8）フェイス項目
世帯類型,主な回答者

4.2. 森岡清美調査資料群について

以上のようにして実施された児童養育費調査は，森岡資料群のひとつとして保管されてきた。森岡資料群とは，宗教社会学・家族社会学分野において大きく貢献された森岡清美氏の残した一連の調査資料群を指す[9]。森岡清美氏は，実施した膨大な調査資料を，自宅裏の

[9] 森岡資料群を構成する調査の一覧は，小林（2019）を参照のこと。

物置に保存していた。これらの調査のうち，最も古いものは 1947 年，最も新しいものは 1985 年の調査であり，1950 年代から 1960 年代頃に実施された調査資料が最も多い（小林 2019）[10]。森岡氏の実証研究は，実際に現地に足を運び，数日から一週間程度滞在する中で，丹念に行ったフィールドワーク，調査票調査，インタビュー，各種資料収集に基づいており，その調査データは豊富な情報を保持している（小林 2019）。その膨大なデータは，戦後の混乱期から高度経済成長期にいたるまでの家族・農村・宗教の実態を知るうえで，極めて大きな貢献を果たす。また，社会学における当時の知見が，どのような方法論と検討のうえで創出されたのか，当時の研究論文における議論がどのような調査に依拠してなされたのかに関する貴重な情報を提供する（小林 2019）。

森岡宅の物置に保管されていた膨大な調査資料は，「質的データとしてのライフストーリーのアーカイヴ化と＜調査遺産＞継承の経験的研究」（2011-2013 年度，研究代表 小林多寿子）のプロジェクトの一環として整理された。森岡資料群は，森岡氏の物置から，一度小林多寿子研究室に持ち運ばれ，そこで資料内容のリスト化，資料のスキャンがなされた。紙の調査資料は，保管されていた「箱」，各箱の中において紐で結ばれていた「束」，束の中に主に封筒で整理されていた「袋」をもとに分類され，おおよそ調査ごとにきれいに整理されている（小林 2019）。SSJDA は，2022 年 1 月に，研究代表者である小林多寿子氏から，一連の森岡資料群を受領した。上述した児童養育費調査は，この寄託された森岡資料群の一部である。

5. 作業の過程

以降では，北会津調査の復元・アーカイブ化作業の詳細を示す。作業は大きく，(1) 作業準備，(2) データ入力作業，(3) クロスチェック作業，(4) データクリーニング作業，(5) 最終チェックに分かれる（図 1）。作業は，監督者 1 名と，作業者 3 名によって行われた。以下に，作業段階別に作業記録を示す。

5.1. 作業準備

作業に先立ち，作業準備を行った。この段階では，以降のデータアーカイブ化作業が円滑に進むよう，いくつかの準備を行った。第 1 に作業環境の構築，第 2 にラベルファイルの作成，第 3 に不明点メモシートの作成，第 4 に作業説明書の作成である。

5.1.1. 作業環境の構築

データ復元作業に先立ち，表計算ソフトの共有設定を行い，作業者全員が同時アクセス・

[10] 多くは森岡氏によって実施された調査であるが，一部森岡氏以外によって実施された調査も含まれている。

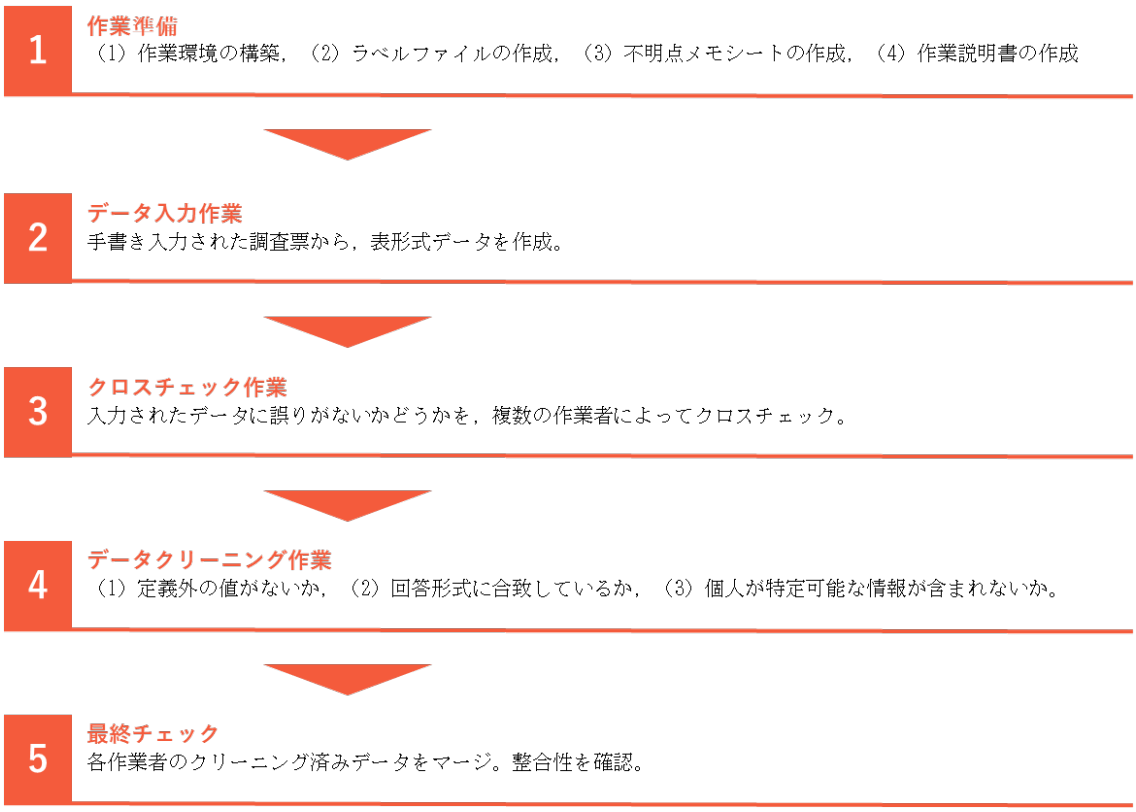


図 1　作業の流れ

同時編集作業が可能なファイルを作成し，作業環境を構築した。図 2 は，作業フォルダの構造を示したイメージ図である[11]。作業フォルダは，（1）入力作業，（2）クリーニング作業，（3）データバックアップによって構成されている。（1）入力作業フォルダには，データ入力作業時に使用されるファイルを格納している。このフォルダには，データ本体など，作業過程で編集されるデータを含む data.xlsx ファイル，作業時のマニュアルである作業説明書，元のデータである記入済み調査票が含まれている。データファイル（data.xlsx）には，データを入力していく data シート，変数の内容とその説明が記載されている変数ラベルシート，各変数の値の意味と注意事項が記載されている値ラベルシート，作業時に生じた不明点や相談事項をまとめる不明点メモシートが含まれている。（2）クリーニング作業用フォルダには，データ入力後のデータクリーニング作業に係るファイル，主にスクリプトが含まれる。入力作業が完了したのち，各作業者はアーカイブ化に向けたデータクリーニング作業を行い，作成したスクリプトはこのフォルダに保存した。（3）データバックアップは，入力された表形式データを定期的にコピーし，バックアップしたファイルを格納するフォルダである。data シートに入力されたデータを定期的にバックアップし，このフォルダに格納した。

各フォルダは東京大学社会科学研究所内のセキュアなネットワークに保存されており，監督者と作業者のみがアクセスできるよう厳格に管理した。また，作業履歴を保存するとと

[11] 実際のフォルダ・ファイル名とは異なる。

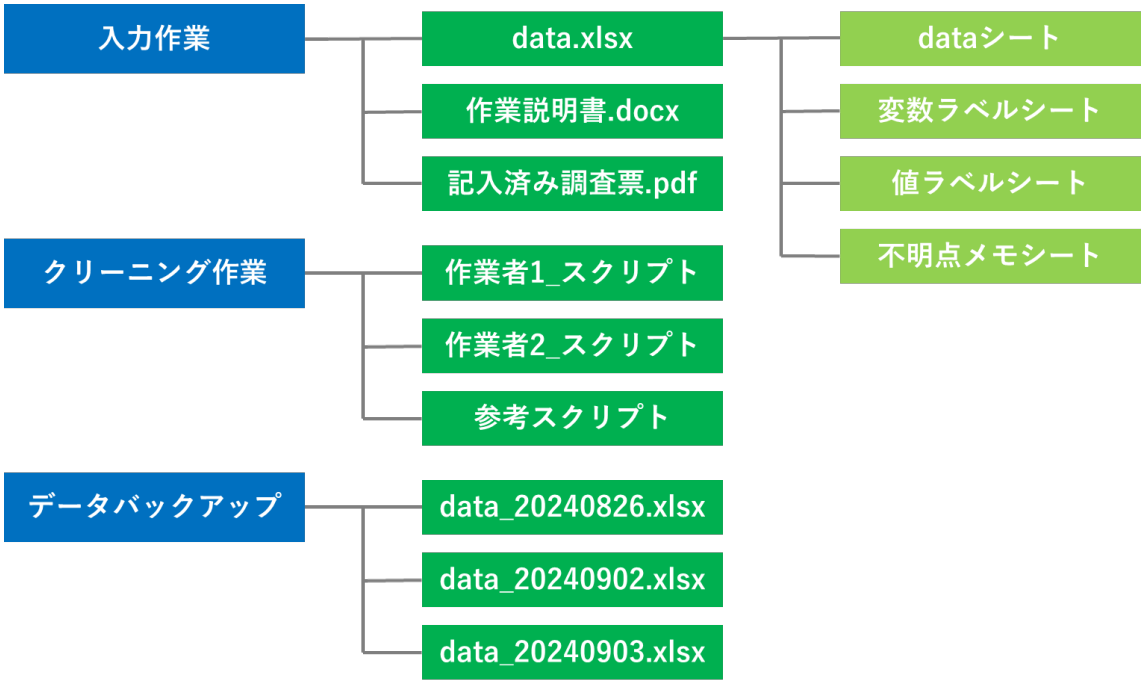


図 2　作業フォルダの階層構造

もに，データの消失を防ぐために，入力データである data シートのバックアップは定期的に行われた。古い社会調査の復元作業においては，複数の担当者が分担することが一般的であると考えられるため，作業フォルダが誰にでもわかりやすく，整理されていること，ファイルとその目的が一意に識別可能であることが重要である。

5.1.2.　ラベルファイルの作成

北会津調査の調査票データ（つまり，面接調査の際に調査員が記入を進めたシート）には，値ラベルが明示されている変数と明示されていない変数が混在している。また，変数ラベルは直接記載されていない。それらの情報は，中鉢編（1970）の末尾に別途記載されているため，まずは変数ラベルと値ラベルを表形式にまとめ，入力作業が効率的に進むよう準備を行った。変数ラベル・値ラベルの一覧表は，メインのデータを含む「data.xlsx」（図 2 を参照）の内部のシートにまとめられ，データ入力者が適宜参照できるようにした。さらに，各変数の入力にあたっての注意点は，ラベルシートに併記し，入力作業を行いながら，一瞥して注意点を確認することができるように工夫した。

5.1.3.　不明点メモシートの作成

古い社会調査の復元作業は，データ入力者が調査に直接参加していないため，聴取の意図や細かな情報の意味を汲み取り切れず，入力作業において不明点が生じる可能性が高い。そこで，入力作業中に悩んだことやわからなかったことは，表形式のシートにまとめ，すべての作業者間で共有した。監督者は，この不明点メモシートを定期的に確認し，解決策を検

記入者	id	変数	不明・相談点	記入日時	対応	対応日	対応者	対応完了フラグ	要確認フラグ（1となっている黄色の箇所はご確認いただけますと幸いです）
	2	q16_3	解読が難しい。見たままだと「エエ」となるが、意味が通らない		後ほど議論する。**以降，解読困難な文字は，まとめて「解読困難」列に解読困難な変数名を入力。**今回のケースは，すでに記入済み。		那須	1	1
	2	q53_6	字が不明瞭で読めない		「解読困難」列に解読困難な変数名を入力した。後ほど議論。		那須	1	0
	2	q71a,b,c	複数の情報が書いてあり、入力の仕方が分からない。		那須がざっと確認したところ，今後もこのようなMA回答が続出する。そのため，**q71aはMA形式にする。**		那須	1	1
	63	q49b,q50b	q49bに数値ではなく「共同」と記載されており、q50bも斜線となっている。		9999無回答と入力。		那須	1	0
	63	q53_6	一部の字が不明瞭で読めない。〇で記入。		「解読困難」列に解読困難な変数名を入力した。後ほど議論。		那須	1	0
	65	q50	回答欄に「／1反」の記載あり。		q49,q50は，この手の記載がかなり多い可能性がある。**このような記載は，数値データの妥当性にかかわる問題であるため，dataシートの「備考（欄外記載など）」に記載しておく。**このケースとmemo_id=23のケースは記載済み。		那須	1	1
	67	q35a	1,4両方に二重丸がついている。		主に使う燃料が特定できないため，そちらは9999，時々使う燃料のMAの方に，1,4の両方を1と入力。対応済み。		那須	1	0

図 3　不明点メモシートの記入例

討するとともに，共有が必要であると判断した不明点についてはフラグを立て，すべての入力者が確認することを促した。

図 3 は，実際に使用された不明点メモシートの例である。データ入力を行う作業者は，不明点が生じた場合，このシートに記入を行い，監督者は対応を示す。作業者全員に確認を促したい不明点については，要確認フラグ列に 1 のフラグを立てたうえで，黄色くセルを塗りつぶした。作業者に対しては，細かいことであっても迷ったことがあれば，この不明点メモシートに記載するように促した。

5.1.4.　作業説明書の作成

データ入力を行う分担者間で認識の齟齬が生じないよう，入力作業に関して詳細な指針を記載した作業説明書を作成した。調査の意図や背景についても簡単に説明を付し，入力作業の助けとなるように工夫した。

以上の作業準備が完了した段階で，データ入力作業に関する作業説明会を実施した。作業説明書，記入済み調査票，入力する表形式ファイル，ラベルシートを用意し，実際の作業の流れを説明した。

5.2.　データ入力作業

以上の作業工程を経て，いよいよデータ入力作業に入った。データ入力作業は，2024 年 8 月から同年 10 月にかけて行われた。北会津調査は，調査当時に記入された手書きの調査票データの形で保存されている。われわれの目標は，調査票に記載された手書きの回答情報を，二次分析しやすい表形式データに整理することにある。そうすることで，データを保存するだけでなく，研究利用を促すことで新たな知見の創出につながる。したがって，まず初めに，手書きの調査票データから，電子データ（図 2 における「data.xlsx」うち data シート）

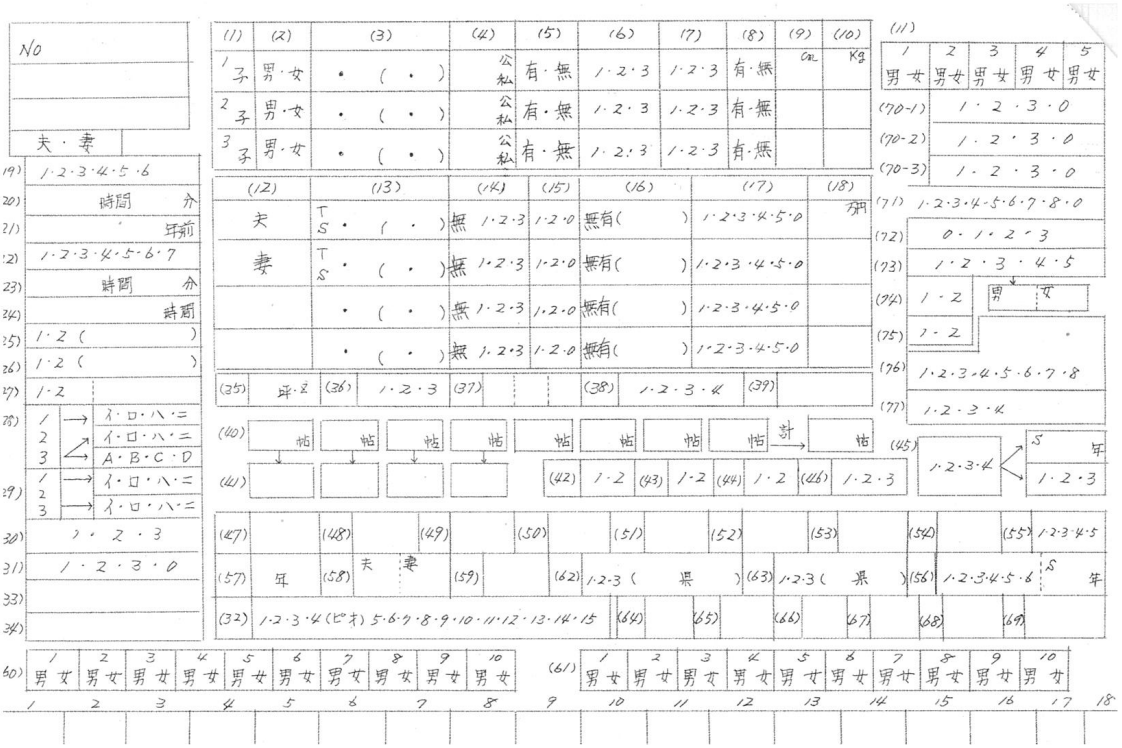


図 4　調査票の例（未記入のもの）

に入力し直す必要がある。図 4 は，データ入力作業において参照される手書きの調査票（未記入）である。このような調査票から，回答を確認し，それを表形式のデータに変換してゆく。このデータ入力作業に当たっては，3 名の入力者が分担して入力を行い，ヒューマンエラーの影響を抑えるために，入力していない他 1 名の入力者によってクロスチェックが行われた。各入力者は，寄託された手書きの調査票から，表形式データの各セルにデータを入力してゆく。

手書きの調査データから表形式の電子データを作成する際には，まず初めに「見たままに」入力する作業を行い，その後次第に数値データとしてまとめ上げていく作業手順とした。つまり，最初のデータ入力作業では，手書きの調査票からそのまま表形式の電子データに入力した。これには大きく 2 つの目的がある。第 1 に，歴史資料に残された情報を可能な限り残すこと，第 2 に，入力者間の入力ルールの不整合を防止する目的である。

第 1 の目的については，前田ほか（2024）が指摘するような，「『計量分析可能なデータセットの整備』と『過去の歴史資料の保存』という，両立可能に見えて相当なジレンマを発生させる 2 つの要求」（前田ほか 2024: 8）と関連している。北会津調査には，かなり多くの不規則回答が含まれる。例えば，本来は数値で回答すべき設問が，文字で回答されている場合や，設問の意図から一部外れるような回答が散見される。このような調査票への記載は，統計分析を行うための表形式データに反映させる際には困難を伴う。さらに，二次分析を行う際にも，こうした不規則な回答は基本的には利用できない。したがって，通常であれば，

情報を縮約し，調査票の設問に対する回答のみを表形式データにまとめるであろうが，今回のデータ入力作業では，過去の歴史資料を可能な限り原状のまま残すという観点から，解読可能なすべての情報をデジタル化した。そのうえで，その後のデータ公開の段階では，調査設問外の記載事項などは削除して非公開とした[12]。

第 2 の理由としては，入力者間の入力ルールの不整合を防止する目的である。かりに，データ入力の段階で，「見たままに」入力しない方針をとった場合，異なる入力者間で入力の傾向に体系的な違いが生じてしまう可能性がある。3 名のデータ入力者が同時に作業を行うことは，入力の際のルールや規則が一貫しない危険性をもたらしかねない。古い調査をデジタル化する場合，調査当時の狙いや意図を十分に理解しない入力者が作業を行うことになるため，そのリスクは通常の社会調査データの入力作業と比べても大きくなる可能性がある。このようなリスクを抑えるためにも，入力作業の際には，手書き調査票から「見たままに」入力する方針とした。

5.3. クロスチェック作業

作業者によるデータ入力作業が完了したのち，入力作業における入力違いがないかどうかを，1 名がクロスチェックをした[13]。データ入力者とは異なる作業者が，入力されたデータを確認し，齟齬が確認されればセルを黄色く塗りつぶす方法をとった。クロスチェック前のデータは，別途バックアップを取り保存した。黄色く塗りつぶされたセルを再確認し，最終的な値を確定した。このクロスチェックにより，2,521 件のセルが修正された。これは，全セル（50,220 件）のうち，約 5.02%（=2,521/50,220）にあたる。ただし，そのうちのほとんどは非該当条件による非該当・無回答の修正によるものであった。

5.4. データクリーニング作業

データ入力が完了したのち，データクリーニング作業を行った。データクリーニング作業は，統計分析ソフト（SPSS と一部 R）を使用して行われた。2 名の作業者によりデータクリーニングが行われ，1 名は調査の前半の設問，もう 1 名は調査の後半の設問といったように役割が分担された。この作業では，(1) 定義された値以外の値が記入されていないかどうか，(2) 調査で想定されている回答形式（シングルアンサー/マルチアンサー）と，実際に回答された回答形式に齟齬がないかどうか，(3) 自由記述項目に個人を特定できる回答がないかどうかの確認を行った上で，変数ラベル，値ラベルの付与を行った。

第 1 に，定義された値以外の値が入力されているかどうかを確認し，エラーがあれば修

---

[12] これは，調査設問外の情報には，個人が特定可能な情報などが含まれる可能性があるためである。

[13] 本来であれば，異なる入力者 2 名が独立に入力作業を行い，完了後にそれらを照合し，齟齬があるセルを確認・修正するという方法をとるべきであるが，様々なリソースの制約から，今回は，1 名の入力者が入力した内容を，他の入力者が目視で確認し，誤りがあれば修正をするという方法を採用した。

正を行った。ほとんどすべてのセルでこのエラーは確認されなかったが，一部の変数ではこれが確認された。例えば，問 4（発達段階）に関する聴取事項では，小学生から大学生まで，児童の発達段階を聴取しているが，いくつかのケースで「農業従事者」「家事」など，調査票では定義されない回答が含まれていた。これに関して，既存の情報を極力活用するとの立場から，調査に先がけ定義されていた回答カテゴリを拡張し，調査票に記載された情報を反映する修正を行った。

第 2 に，調査で想定されている回答形式と，実際に回答された回答形式との照合を行い，必要に応じて修正した。北会津調査には，しばしば調査設計時にはシングルアンサーとして設計されているが，実際の多くの回答にマルチアンサーが含まれる設問がいくつかある。入力時に十分に精査しなかった回答形式について，クリーニング段階では精査した。多くの設問では，情報の縮約をせず，たとえ調査票上ではシングルアンサーを標榜した設問であっても，1 つ以上のケースでマルチアンサー形式の回答がなされていた場合，デジタル化されたデータはマルチアンサー形式をとるように修正した。

第 3 に，自由記述データにおいて，個人が特定可能な情報が含まれていないかどうかを確認し，そのような情報はマスクした。これは，データの公開にあたって重要となる。個人が特定可能な情報も含めた表形式データも，非公開ではあるが，SSJDA にて別途保存されている。

5.5. 最終チェック

データクリーニングが完了した段階で，2 名の作業者によって作成されたクリーニング済み表形式データを，監督者がマージし，最終チェックが行われた。全体を通して齟齬がないか，入力漏れや変数漏れがないかどうかを確認した。

6. 作業時に生じた困難・課題

作業を進行する中で，（1）欄外記載事項をどのように処理するか，（2）解読困難文字の存在，が問題となった。本節では，これらの課題や困難にどのように対処したのかを説明する。

6.1. 欄外記載事項

このデータ入力の段階では，欄外に記載された情報をいかに表形式データに反映させるかが問題となる。今回は，手書きの調査票に記載された情報は基本的には残す方針で入力を進めた。欄外記載事項に関しては，特に注意深く検討を行った。まず，記入済み調査票から，欄外記載事項を別列にまとめた。これらの欄外記載事項のうち，70%は，農業従事者の日当（約 27%）・年間総所得（約 16%）・農業従事者数（約 15%）・使用している冬暖房（約 13%）で構成されており，これらの項目に欄外記載事項が集中している。特に，農業従事者の日当に関しては，欄外記載が多く発生していた。これは，調査では日当を尋ねている一方で，実

表 2　報告書との整合性（夫婦出身地）

報告書		実際のデータ		
	割合		度数	割合
		［夫］		
同村内	91%	この家	104	0.839
		同じ部落	4	0.032
		北会津村	5	0.040
		小計	113	0.911
		［妻］		
同村内	29%	この家	17	0.137
		同じ部落	5	0.040
		北会津村	14	0.113
		小計	36	0.290
隣接市町村	62%			
		N	124	1.000

際の回答は「1 反あたり」の給与を答えるなど，回答の仕方にブレがあるケースが多いことが原因である。年間総所得にみられる欄外記載事項に関しては，その多くが「年金」など，年間総所得に関する付加的な情報を記載しているケースが多くみられた。欄外記載事項が多いこれら 4 変数に関しては，特に別列を作成し，メインデータに反映しきれないものを備考変数としてまとめた。それ以外の変数に関連する備考や欄外記載事項についても，メインデータへの反映を検討しつつ，備考変数としてまとめて残した。

6.2.　解読困難文字

第 2 に，手書き調査票の可読性が問題となった。本調査は 1960 年代に実施されたため，入力作業を行った 2020 年代とは異なった表現や書き方がされる場合が散見され，また，筆圧の薄さや記入者の癖なども相まって，解読できない文字が多く存在した。こうした解読困難文字に関しては，作業者間で議論を行うとともに，調査項目から推察を行った。この議論の際には，インターネット上の情報源も参照し，例えば当時使用されていた農機具の名称や，地名などから推測し，情報を復元することができた。それでもなお，解読ができない場合には，「〇〇」と代替した。データセット全体にわたり，最終的に「〇〇」と代替されたセルは 2 件にとどまった。

7.　過去の報告書とデジタル化データとの整合性

本節では，「家族周期と児童養育費」（中鉢編 1970）で報告された北会津調査の基礎集計をもとに，デジタル化された本データとの整合性を確認する。とはいえ，報告書には膨大なクロス集計表が含まれるため，ここでは，主要な項目の単純集計に限定して確認を行った。

表 3　報告書との整合性（夫婦の学歴）

報告書		実際のデータ			
	%		度数	%	累積%
		［夫］			
9 年以下	58%	小学校	5	4.032	4.032
		旧制高等小学校・新制中学	67	54.032	58.065
		旧制中学・高女・実業学校・新制高校	45	36.290	94.355
		旧制高専・新制高専・短大	5	4.032	98.387
		大学	1	0.806	99.194
		無就学	0	0.000	99.194
		無回答	1	0.806	
		［妻］			
9 年以下	70%	小学校	3	2.419	2.419
		旧制高等小学校・新制中学	84	67.742	70.161
		旧制中学・高女・実業学校・新制高校	34	27.419	97.581
		旧制高専・新制高専・短大	2	1.613	99.194
		大学	0	0.000	99.194
		無就学	0	0.000	99.194
		無回答	1	0.806	
		N	124	100.000	

第 1 に，夫婦の出身地についての，中鉢編（1970）の記述との整合性を確かめる。中鉢編（1970）では，「夫の出身地は 91%が同村内で，残りは隣接市町村に散在している。妻の出身地はそれほど集中せず，同村内 29%，隣接市町村 62%，残りはその他の会津地方となっている」（中鉢編 1970: 116）と記載されている。これに関して，デジタル化されたデータにおける同様の設問の分布を確認したものが，表 2 である。報告書に記載されている「同村内」とは，調査設問の問 60（夫の出身地）と問 61（妻の出身地）における「1. この家」「2. 同じ部落」および，市町村自由記述において北会津村を記載したケースであると定義し，実際のデータと報告書との整合性を確認した[14]。いずれについても整合したと考えられる。ただし，報告書に記載されている，妻の出身地が近隣市町村であったケースについては，定義が不明であり，カウントすることがかなわなかったため，確認できなかった。

表 3 は，夫婦の学歴に関する報告書の記載と，デジタル化されたデータの集計結果との整合性を比較したものである。報告書には，「夫の 58%，妻の 70%が 9 年以下の教育を受けた」（中鉢編 1970: 116）と記載されている。この記述をもとに，デジタル化したデータとの整合性を確認する。表 3 によれば，旧制高等小学校・新制中学までの累積割合と，報告書の

[14] 市町村自由記述は，個人の特定可能性の観点から，非公開としている。

表 4　報告書との整合性（結婚年齢・年数，出生）（平均）

	報告書	実際のデータ
結婚年齢		
夫	23.2	23.331
妻	21.3	21.274
結婚年数	13.9	13.952
理想的な子ども数	3.3	3.274
追加出生意欲	0.3	0.266

記述とが整合していることがわかる。戦前・大正期の旧制高等小学校は，一般に旧制尋常小学校入学から 9 年程度で卒業するため，報告書の記述と整合的である。

次に，表 4 は，報告書に記載された数値と，実際のデータの整合性について，結婚年齢，結婚年数，理想的な子ども数，追加出生意欲を比較したものである。結婚年齢については，「夫 23.2 歳，妻 21.3 歳で結婚し,」（中鉢編 1970: 116）という記述をもとにしている。結婚年数に関しては，表 L-2（中鉢編 1970: 117）の周辺セルを，理想的な子ども数および追加出生意欲については，表 L-3（中鉢編 1970: 118）の周辺セルを参照している。大きく異なる集計結果とはなっていないものの，四捨五入を前提とした場合，夫の平均結婚年齢，結婚年数でわずかな差がみられる。

最後に，表 5 は，所有する耐久消費財に関する報告書との比較結果を示している。報告書の数値は，表 L-9（中鉢編 1970: 123）の集計結果をもとにしている。多くの項目で，報告書に掲載されている数値を整合している。ミシン，洗濯機，オートバイ・スクーター，自動車の所有割合は，実際のデータと若干の齟齬があるものの，その大きさも顕著ではない。多くの項目では，報告書とデジタル化されたデータとの間に整合性がある。

以上の比較より，完全な一致はかなわないものの，多くの項目で報告書の集計結果と整合的であると結論付けられる。

8. 結論

本稿では，森岡資料群のうち，特に「児童養育費調査 北会津 農村調査」を事例として，実施から時間が経過した古い社会調査データの復元とアーカイブ化の作業過程を記録した。

実際の調査データのアーカイブ化過程は，本稿に述べたとおりであるが，その中でもいくつかの困難にあたった。調査設計上の設問事項からはみ出した欄外記載事項の情報をどのように処理するか，字の薄れや劣化による可読性の低下に起因する解読困難文字をどう対処するかといったことが生じた困難の代表的なものである。また，コーディング・データクリーニングの段階においては，異なる記入者間でルールを十分に明確化し，統一させることが，特に古い社会調査データの復元作業には重要である。

森岡資料群のアーカイブ化は端緒についたばかりである。森岡資料群には，児童養育費

表 5　報告書との整合性（所有耐久財）

	報告書	実際のデータ	
	%	度数	%
テレビ	100	124	100.000
ミシン	88	108	87.097
洗濯機	80	98	79.032
オートバイ・スクーター	52	63	50.806
電気冷蔵庫	41	51	41.129
カメラ	35	43	34.677
電気・ガス釜	25	31	25.000
扇風機	20	25	20.161
電話	11	14	11.290
ピアノ・オルガン	5	6	4.839
ステレオ	2	3	2.419
ミキサー・ジューサー	2	3	2.419
トースター	2	2	1.613
応接セット	2	2	1.613
自動車	34	41	33.065
所有点数平均	4.6	4.621	

調査以外にも膨大な資料が含まれており，その多くは質的なデータである。これらを，いかにしてデジタル化し，アーカイブ化するか，さらに適切な方法と範囲において公開するかといった課題が山積している。質的調査研究の文脈では，フィールドノート，パーソナル・ドキュメント，オーラル・データなどの調査資料をどのように扱うか，具体的に検討しなくてはならない事態に直面している（小林 2019）。さらに，質的調査データの再分析・再解釈は，新しい知見や観点を生み出すうえで有用であるとされ（Kuula 2011），いくつかの国外のデータアーカイブではすでに質的調査データのアーカイブが行われている（Bishop and Kuula-Luumi 2017）。日本では，質的データのアーカイブが十分に行われていないが，国外の事例をもとに，質的データアーカイブの方法を模索する動きもある（横内ほか 2025）。本稿で行った児童養育費調査は，主に統計分析による二次分析を前提としたデータアーカイブであるが，今後の森岡資料群のアーカイブ化作業においては，量的調査データのアーカイブ化を行ってきた SSJDA においても，質的調査データのアーカイブ化の検討を要する。

謝辞

本調査データのアーカイブ化作業に当たっては，JSPS 人文学・社会科学データインフラストラクチャー強化事業 JPJS00320231001 の支援を受けた。

本調査データのアーカイブ化作業に当たっては，多くの方々のご協力を賜った。まず，大変貴重な調査データを膨大に残してくださった森岡清美先生に感謝申し上げる。森岡先生の残してくださった調査データは，当時の人々の様子を知る貴重な証言となっている。また，森岡資料群をご寄託くださった小林多寿子先生とそのご関係者様に深く御礼を申し上げる。森岡資料群を適切にスキャンし，保存する作業は，アーカイブ化のために必要不可欠であった。その多大なご尽力に心より御礼申し上げる。

データの復元・アーカイブ化作業にあたっては，東京大学大学院教育学研究科の石橋大樹さん，金井明日香さん，東京大学教育学部の鍋田萌子さんにご協力いただいた。記して感謝申し上げる。

本稿の執筆にあたっては，東京大学の鎌田健太郎さん，山口ゆり乃さんから大変有益なフィードバックをいただいた。記して感謝申し上げる。

付記

本プロジェクトのうち，アーカイブ化作業の監督・管理・記録と，本稿の執筆を那須が担当し，プロジェクトの企画・作業者間の調整を入山が担当した。

本稿で表明されている見解は，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公式見解ではないことをお断りしておく。

［文献］

相澤真一・小山裕・鄭佳月，2013，「社会調査データの復元と計量歴史社会学の可能性――労働調査資料（1945-1961）の復元を事例として」『ソシオロゴス』37: 65-89.

朝岡誠，2024，「社会科学研究データの公開について」『理論と方法』39(2): 220-32.

Bishop, Libby, and Kuula-Luumi, Arja. 2017. Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open*, *7*(1). https://doi.org/10.1177/2158244016685136.

小林多寿子，2019，「リサーチ・ヘリテージとしての森岡清美調査資料群――森岡調査資料との出会い」『一橋社会科学』11(別冊): 3-27.

Kuula, Arja. 2011. Methodological and Ethical Dilemmas of Archiving Qualitative Data. *IASSIST Quarterly*, 34(3–4), 12–17. https://doi.org/10.29173/iq455.

前田一歩・堀江和正・瀬戸健太郎，2024，「『国鉄労働組合婦人部実態調査』（1952 年）の復元過程――『データセットの整備』と『資料保存』は両立しうるか」SSJ データアーカイブリサーチペーパーシリーズ，91.

三輪哲・佐藤香，2018，「オープンサイエンス時代における社会科学データアーカイブの意義――SSJ データアーカイブのこれまでとこれから」『情報の科学と技術』68(10): 489-

94.
森岡清美，1967,「農村家族における児童養育費の研究(I)――家族周期からみた生活構造の分析を中心に」『季刊社会保障研究』3(4): 57-70.
中鉢正美編，1970,『家族周期と児童養育費――児童養育費調査報告書』至誠堂.
仁田道夫，2025,「戦後労働調査資料について――1980 年代の保存所蔵事情を中心に」『社会科学研究』76: 159-74.
労働調査論研究会編，[1970] 2001,『戦後日本の労働調査』東京大学出版会.
佐藤香，2009,「歴史的資料としての社会調査データ」『年報社会学論集』22: 22-31.
――――，2025,「社研労働調査資料のデジタル復元」『社会科学研究』76: 1-6.
佐藤俊樹，1998,「特集　計量歴史社会学」『理論と方法』13(1): 1-4.
――――，2010,「SSJ データアーカイブの現状と課題――社会調査・データアーカイブ研究センターの活動」『アーカイブズ学研究』12: 3-10.
――――，2012,「実証研究におけるデータアーカイブの役割と課題――SSJ データアーカイブの活動実績を踏まえて」『フォーラム現代社会』11: 103-12.
渡邊勉，2024,「計量社会学による歴史研究の可能性」『理論と方法』2024, 39(1): 2-18.
横内陳正・谷口沙恵・石田賢示，2025,「質的データの保管・共有に関する海外データアーカイブの実践――FSD と DANS の事例」『SSJ Data Archive Research Paper Series』No.93.