

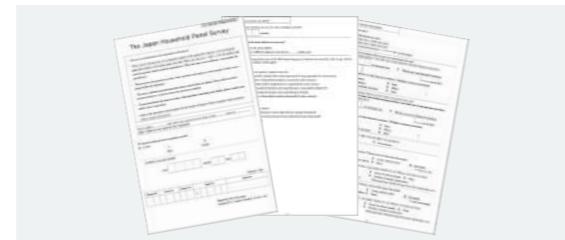
3. オンライン分析ツール（BI ツール）の作成

膨大な情報を含むパネルデータは、専門の統計解析ソフトを利用しないと十分な分析ができません。データを整理して統計解析ソフトでの分析に取り掛かる前段階として、データの確認や分析のヒントを得るために、オンライン上で簡単な集計を実現できるシステムを構築しました。



4. グローバル化への対応

皆様のご協力のもと作成されたパネルデータは、調査票や各種手引き、変数カタログに至るまですべて英語化し、国を超えて世界の学術研究で活用されるよう努めています。経済協力開発機構（OECD）の国際比較分析などにおいても、日本を代表するデータとしてたびたび活用されています。



5. さまざまなパネルデータの提供

慶應義塾大学パネルデータ設計・解析センターでは、皆様にご協力いただいている「日本家計パネル調査」のみならず、国内の研究機関からパネルデータの寄託を受け入れており、さまざまなパネルデータの共有化を進めています。「パネルデータのことなら慶應義塾大学パネルデータ設計・解析センター」といったように、パネルデータに関する国内屈指のデータ機関を目指しています。

Our Activity

活動報告

ニュースレターをご覧くださいどうもありがとうございました。

新型コロナウイルス感染症の流行から早くも3年が経とうしておりますが、わが国でも今年の夏にこれまでにない感染者数の増加を経験しました。この感染症に罹患された方、治療・療養を余儀なくされた方、生活・仕事に支障が生じている方、並びにご家族の皆様に対し、心よりお見舞い申し上げます。また不幸にお亡くなりになられた方のご冥福を心よりお祈り申し上げます。

慶應義塾大学パネルデータ設計・解析センターでは、新型コロナウイルス感染症の流行以降、皆様のご協力のもと『新型コロナウイルス感染症が社会に与えた影響に関するJHPS特別調査』（「JHPSコロナ特別調査」）を継続し、未曾有の世界的危機下における人々の暮らしへの影響を捉える数々の研究を行い、必要な政策提言をしてまいりました。こうした取り組みが評価され、国からの大型研究助成である「科学研究費助成事業（特別推進

研究）」を新たに獲得することができました。皆様方による調査へのご協力なしには成しえなかったこと、深く御礼申し上げます。

また、今年度は「JHPSコロナ特別調査」による研究成果を1冊の本にまとめることができましたこと、謹んでご報告申し上げます。現在、出版の最終調整を行っており、慶應義塾大学出版会より『コロナ禍の家計のレジリエンス格差』というタイトルで2023年3月ごろの出版を予定しております。

センターでは国際共同研究も引き続き進めております。昨年度までは難しかった対面での交流も徐々に再開し始め、今夏には日仏共同研究プロジェクトの一環として、フランスの社会科学高等研究院より研究者を招聘し、慶應義塾大学三田キャンパスにて研究セミナーを実現することができました。また、来春には日英共同研究プロジェクトの一環として、ロンドン大学の研究者らと英国にてセミナーの開催を予定しております。

私共の調査・研究は回答者の皆様や国からの研究助成から支えられていることを肝に銘じて、今後も社会に役立つ研究ができるよう邁進していく所存でございます。これまでの皆様のご支援に厚く感謝申し上げますとともに、今後とも引き続きご協力のほど、よろしくお願ひ申し上げます。



Interview 学術分野や国を超えた 研究データの共有化・利活用の必要性と取り組み

慶應義塾大学パネルデータ設計・解析センターでは、皆様からご協力いただき収集した貴重なパネルデータについて、センター内部の研究者が活用するだけではなく、学術分野や国を超えて、貴重な研究資源を多くの研究者に共有・利活用する取り組みを積極的に進めています。こうした研究データの共有化や利活用については、日本学術振興会による「人文学・社会科学データインフラストラクチャー構築推進事業」として、現在、国をあげて積極的に進められています。慶應義塾大学パネルデータ設計・解析センターも当事業の拠点機関として、パネルデータに関する日本屈指のデータ機関として活動を進めています。今回は「人文学・社会科学データインフラストラクチャー構築推進事業」の主要メンバーである東京大学教授・前田幸男先生に研究データの共有化・利活用の必要性と取り組みについてお話をうかがいました。

データの共有化・利活用と専門知識を持つ人材の育成

前田先生の研究分野と研究データの共有化・利活用とのかかわりについて、お話しください。

私の専門はもともと政治学です。大学院生の頃、選挙や世論調査データを使った政治学研究がしたかったのですが、私の進んだ大学院には調査をしている教員はいませんでした。その頃の日本には、ほかの研究者が収集したデータを共有するような基盤がなかったのでアメリカに渡った経緯があります。アメリカではすでにデータの共有化に関する取り組みが進んでいて、政治学の研究のかたわら、データの共有化や利活用に関する基盤整備に関して学び、その分野にも詳しくなりました。

データの共有化・利活用とは何でしょうか？

データの共有化とは、調査を実施した研究者だけでなく、それ以外の研究者にもデータを提供することです。以前は、データは調査をした人で利用が止まっていましたが、一回取ったデータを長期で蓄積していろいろな人に使ってもらうことには大きなメリットがあります。調査実施者は自身の研究のために調査設計していますが、別の研究者がそのデータを見たときに新しい研究の材料やヒント、分析のきっかけになります。また国の研究助成を受けて調査を実施している以上、自分たちの利用にとどめておくだけではなく、ほかの研究者にも利用してもらい、公共的な研究資源になることが大切です。

「データインフラストラクチャー」というのは聞きなれない言葉ですが、わかりやすく教えてください。

「データインフラストラクチャー」とは、データの共有化・利活用を促進するために、個々のデータの整備やメタデータ（データの説明が記載されたデータ）の作成、さまざまなデータの情報を一括管理して対応できるようにする制度や組織を指します。「データアーカイブ」という言葉が一番わかりやすいかもしれません。

図書館に司書がいて図書資料の情報整理をしているように、研究データにもいろいろなデータを整備してみんなに情報を開示するようなデータの専門職員が必要だと考えています。信頼できるデータを適切なメタデータとともにすぐにアクセスできる状態にして開示することは、学生の教育や研究者の研究を支えるオープンサイエンスには欠かせないことだと思います。

データを安心して使えるルール作りの重要性

オープンサイエンスとはなんですか？

わかりやすく説明しているのは2014年に作成された「FAIR」原則ですが、研究者やそうでない人も、学術的研究や調査の成果などの論文やデータが簡単に見つかり、アクセスでき、再利用できるようにする活動のことを指します。個々の研究者が自身の利用したデータについて整備・公開するのは大変な時間と労力を要するので、オープンサイエンスを実現するために、個々の研究者に代わって、安心して使ってもらえるデータを共有するための専門的知識をもつ人材を育成し、安定した組織を作る必要があります。そうすることで、研究者は自分で収集したデータをデータ整備の専門家に託すこともでき、また、必要なデータを簡単に見つけ利用することもできるようになります。研究者は自分の研究に専念することができるというメリットも生まれます。

(参照：<https://rcos.nii.ac.jp/document/openscience/>)

データ整備の専門知識を持つ人材の育成で、いうと学術界ではどのような位置づけになるのでしょうか？

おもに図書館情報学や情報技術の知識が必要になってしまいます。専門分野のデータを扱うためのキャリアパスとして、例えば「社会学」や「情報学」の修士号が必要だと思いますが、どの分野でなければいけないという明確な決まりはありません。日本の大学ではデータ共有に必要な活動は研究者の業務の一貫としてやっていますが、海外では専門職としての待遇が用意されていることが多い、そうしたトレーニングを積んだ人材がいることは、データを利用する研究や教育の支えにもなっています。

データを共有する上での留意点は何ですか？

まずは個人が特定されることがないようにデータを編集することです。データの中にある地理情報・性別・年齢・職業などを掛け合わせるだけでも、個人が特定されてしまう可能性が高まります。ですので、広く利用されるデータについては、地理情報を大きな単位にまとめたり、年齢を10歳刻みにしたりするなど、データをみんなが安心して使える一定の指針が必要です。また、調査を実施する側では、メタデータといった調査設計や回収率などデータに関する説明文書をあらかじめ作成しておくことも大切です。調査を実施した本人はそのデータがどのように収集されたのかわかりますが、ほかの研究者は文書なしにわかりませんし、本人であっても時間が経つ忘れてしまうことがあるため、早めに説明文書を作成することは重要です。

海外と日本の違いと今後の取り組みについて

海外ではデータ共有化・利活用に向けてどんな取り組みをしていますか？

国によってかなり体制が異なります。ヨーロッパでは国の組織が一元管理する体制をとっている場合もあり、例えば、ノルウェーやオランダでは国の予算で、データ共有化のための組織が整備され、専門のスタッフが常駐しています。イギリスやアメリカでは、大学主導で分権的に行っているように見えます。日本も国立や私立の大学が多様に存在し、それぞれ政府の予算も違うため、一元管理するのが難しい構造になっています。現在の日本の制度を考えると、既存の構造の中でゆるやかに協力しながらデータインフラの制度設計をしていくのが現実的だと思っています。私は、そのための環境を整えていくことが将来的に大事であると考えています。



東京大学教授 前田 幸男先生

学術研究目的でのデータ共有化・利活用に向けた取り組み

“ビックデータ”や“データ活用”という言葉をよく耳にする今日ですが、経済学や社会学といった社会科学の研究分野においても、データに基づく実証研究がますます重要視されてきています。社会問題の原因究明や経済法則の解明、政策の効果測定などにデータは不可欠であり、貴重な研究資源として共有化・利活用への取り組みが積極的に進められています。

慶應義塾大学パネルデータ設計・解析センターでは、毎年皆様にご協力いただき調査を実施し、パネルデータを収集しておりますが、このデータについても貴重な研究資源として、学術研究目的で共有化・利活用を進めております。データの共有化を進め、様々な視点から活用されることで、データは想定されていた以上の学術的価値を生み出すことができます。

本センターにおける学術研究目的での共有化・利活用に向けた取り組みは、日本学術振興会「人文学・社会科学データインフラストラクチャー構築推進事業」の一環として、国から研究助成を受けて実施しております。ここでは、皆様にご協力いただき収集した貴重なパネルデータを研究者間でどのように共有化し利活用を進めているのか、センターの取り組みをご紹介いたします。学術研究目的でのデータの共有化・利活用に際しては、回答者個人が特定されることのないよう細心の注意を払っていることなどをご理解いただき、今後も安心して調査にご協力いただければ幸甚です。

皆様にご協力いただき収集した貴重なパネルデータは
こうして共有化されています。

データ共有化までのプロセス

1. データの生成

データクリーニング

皆様がご回答くださった調査票について、前年の回答と著しい矛盾がないかなど1件1件確認したうえで、エクセルなどの表計算ソフトに数値として入力します。

データの匿名化

回答者個人が特定されることのないよう、氏名や住所、生年月日などの個人情報は入力するデータに決して含めません。また、お住いの都道府県の情報についても、職業や性別、年齢といった情報と組み合わせることによって個人が特定される可能性があるため、提供するデータに含めることはしません。ご回答くださった調査票に基づきデータ入力をする段階で、こうしたデータの匿名化処理をいたします。

データの精査と微修正

データの匿名化処理が終わると、最初の一年間は、調査を実施した本センターの研究者が優先的にデータを利用して分析を行います。センターの研究者は、研究計画に沿って分析を進めながら、データの精査を行い、正しいデータが共有化されるように必要に応じてデータの微修正を行います。

2. 学術研究目的でのデータの共有化

センターの研究者の優先利用期間が終わると、匿名化されたデータは、本センターのホームページにある「データ申請・管理システム」にて、一般の研究者に向けた貸し出しが開始されます。

データの利用希望者は、ウェブの申請画面に、所属や職位、連絡先、データを使ってどのような研究をどのように行うのかなどについて入力して、データの利用申請を行います。申請された内容に基づき、利用資格・要件が満たされているか、研究目的や研究計画が適切であるかなどについてセンターの職員が厳正に審査を行い、利用の可否を決定します。データは非営利・学術目的での利用に限り、国内外の大学または国公立・民間研究機関に所属する研究者・大学院生・学部生に提供されます。貸し出しに際しては、誓約書に記載された事項に従うことを義務付け、誓約事項に違反した場合には、今後データの利用を禁止するなどの厳しい処置をとっています。

以前は紙の申請書とCD-ROMに保存したデータを郵送でやり取りしていましたが、データの利活用を促進する観点から、2017年よりオンライン上のデータ申請・提供システムに変更しました。これにより、国内はもとより海外からも申請しやすくなりました。



学術研究目的でのデータ共有化・利活用を 促進する取り組み

1. パネルデータ作成プログラムの提供

ご回答いただいた各年のデータは、同じ回答者の他の年の回答と接続することで“変化”を捉えることができるパネルデータに生まれ変わります。質問内容は多岐にわたり、毎年新しい質問が追加されたりしているため、複数年のデータを接続する作業には細心の注意と膨大な時間を要します。各年のデータを接続してパネルデータにする作業は、データを利用する研究者にとって分析前に取り掛かるべき共通の作業であるため、センターではこの作業を自動化したプログラムを作成し、利用者に無償で提供することで、利用者のデータ整備にかかる負担を軽減し、データの利用促進を図っています。

A	B	C	D	E	F	G
1	Id	year	sale	age	work	health
2	10001	2019	0	34	0	1
3	10001	2020	0	35	0	1
4	10001	2021	0	36	1	2
5	10001	2022	0	37	1	2
6	10002	2018	1	57	1	3
7	10002	2020	1	58	1	3
8	10002	2021	1	59	1	3
9	10002	2022	1	60	0	4
10	10003	2012	0	29	1	1
11	10003	2013	0	30	1	1
12	10003	2014	0	31	1	2
13	10003	2015	0	32	1	1
14	10003	2016	0	33	1	1
15	10003	2017	0	34	1	1
16	10003	2018	0	35	1	1
17	10003	2019	0	36	1	1
18	10003	2020	0	37	1	1
19	10003	2021	0	38	1	1
20	10003	2022	0	39	1	1

2. 变数カタログの作成

調査票のボリュームと調査回数からもお察しいただけるように、ご協力いただいている調査から生成されるパネルデータの情報量は膨大で、データを利用する研究者にとっては、何年の調査にどういった情報が入っているのかを判断することは容易ではありません。当センターでは、キーワードによる検索エンジンを活用し、必要とする情報がどの調査年にどのような形で質問されているのか、素早く調べができるシステム「变数カタログ」を構築し、「データ申請・管理システム」上で公開しています。

