



第10回学術コミュニケーションセミナー

事例報告：研究データの利活用

Social Science Japan Data Archive(SSJDA)における
データ利活用の取り組み

2022年7月25日(月)

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
谷口沙恵（特任専門職員 / URA）

目次

1. Social Science Japan Data Archive (SSJDA) の概要
2. SSJDAにおけるデータ収集・保存・提供のしくみ
3. SSJDAデータ利活用の状況
4. 二次分析研究推進のための支援
5. 国内外データアーカイブ関連機関とのネットワーク
6. 課題と展望

1. Social Science Japan Data Archive (SSJDA, SSJデータアーカイブ)の概要

- 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（当時は日本社会研究情報センター）に、1996年5月SSJデータアーカイブを設置し、個票データ*1の収集と保存を開始。1998年より個票データの提供を開始。
- 社会科学全般（社会学、法学、政治学、経済学、経営学、心理学、教育学、社会疫学など）の調査の個票データ（量的データ*2）を収集し、提供
- データ寄託は無償。寄託者（機関）は、研究者、大学、公的研究機関、民間研究機関、一般財団法人、公益法人、官公庁、マスコミなど。
- データ利用は無償。ただし登録制。
- 利用対象者は、大学または公的研究機関所属の研究者、教員の指導を受けた大学院生、SSJDAへデータを寄託している機関所属の方に限る。
- データ利用目的は、学術目的の二次分析*2に限る。
ただし一部データについては、講義などの教育目的の利用も可能。



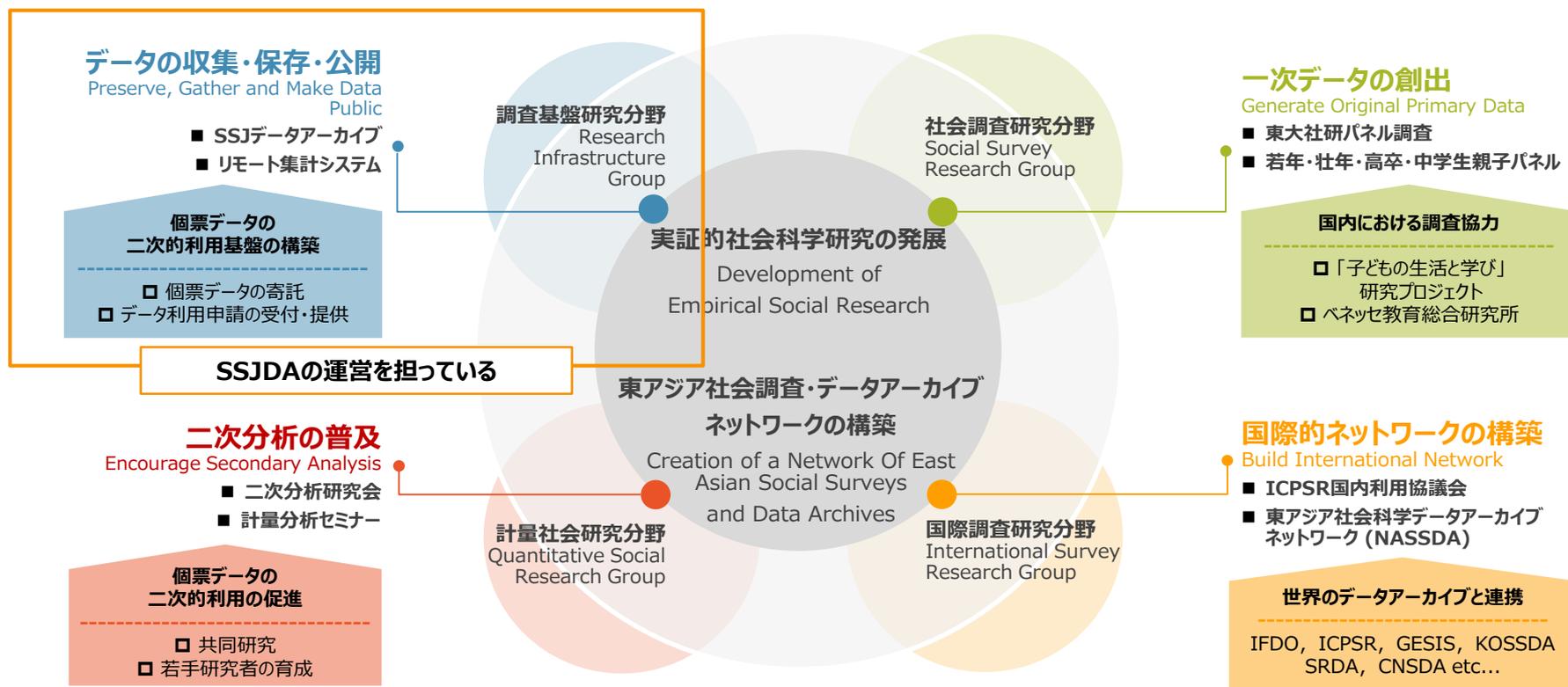
*1 個票データ…アンケートなどの質問票の回答で、個人情報の秘匿処理がなされたもの。

*2 量的データ…年齢、身長、年収など、数量として測ることができるデータ

*3 二次分析…ここでは、調査実施者たちが分析を行い論文等を執筆した後、第三者がそのデータを用いて分析することを指す。

1. SSJDAの概要

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 (CSRDA) 組織図



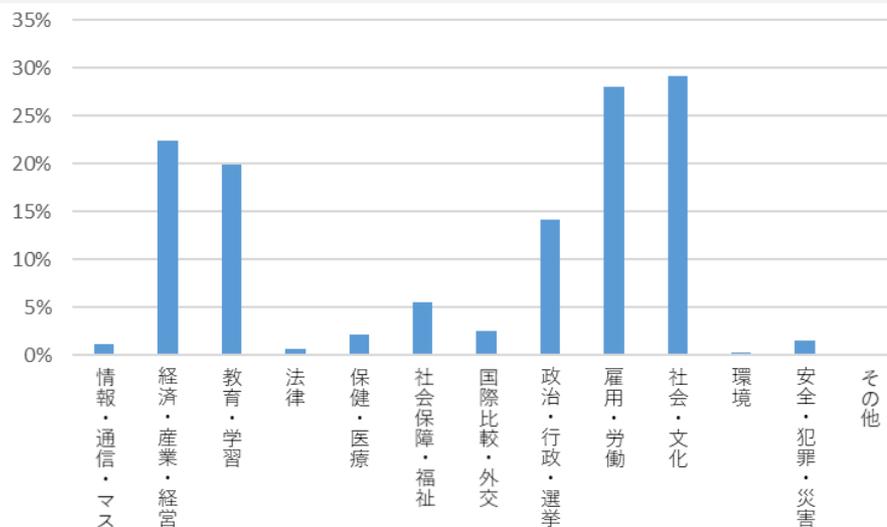
データアーカイブの意義と役割

- 社会調査の個票データの散逸を防ぐ。検索可能であり、データにアクセスでき、相互運用できて、再利用が可能であるというFAIR原則に沿った役割を果たす。
- 研究の再現性の担保
- 調査実施予算がない学生や若手研究者も、データアーカイブから提供される個票データを利用した二次分析による研究が可能となる
- 個票データの二次分析により、多くの新しい研究が可能となる
- 今後実施される社会調査の質の維持・向上
- 同様の調査の重複を防ぐことで、調査費用と手間を抑え、研究者と調査対象者双方の負担を軽減する

SSJDAはこれらの意義と役割に基づき活動している。

SSJDA 提供データについて

- SSJDAから公開しているデータ（データセット）は、主に社会・文化，雇用・労働，経済・産業・経営，教育・学習，政治・行政・選挙などのトピックが多くを占める。（全1542件／2022年7月現在）
- データセットは、これまでに200を超える機関・組織及び研究者(代表)から寄託されている。主な寄託機関は下記①の通り。
- よく利用されている主なデータセットには、例として②などがある。



SSJDAデータセットのトピック別割合

① 主な寄託機関

□ 公的研究機関

日本政策金融公庫総合研究所，労働政策研究・研修機構，2015SSM研究会データ管理委員会，東京大学社会科学研究所，その他個人研究者など

□ 民間研究機関

リクルートワークス研究所，第一生命経済研究所，ヘネッセ教育総合研究所，電通，NHK放送文化研究所など

□ 官公庁

内閣府，厚生労働省，経済産業省，農林水産省など

② 主なデータセット

□ JLPS(働き方とライフスタイルの変化に関する全国調査)

社会科学研究所が実施する若年，壮年，高卒，中学生親子を対象とした3つのパネル調査

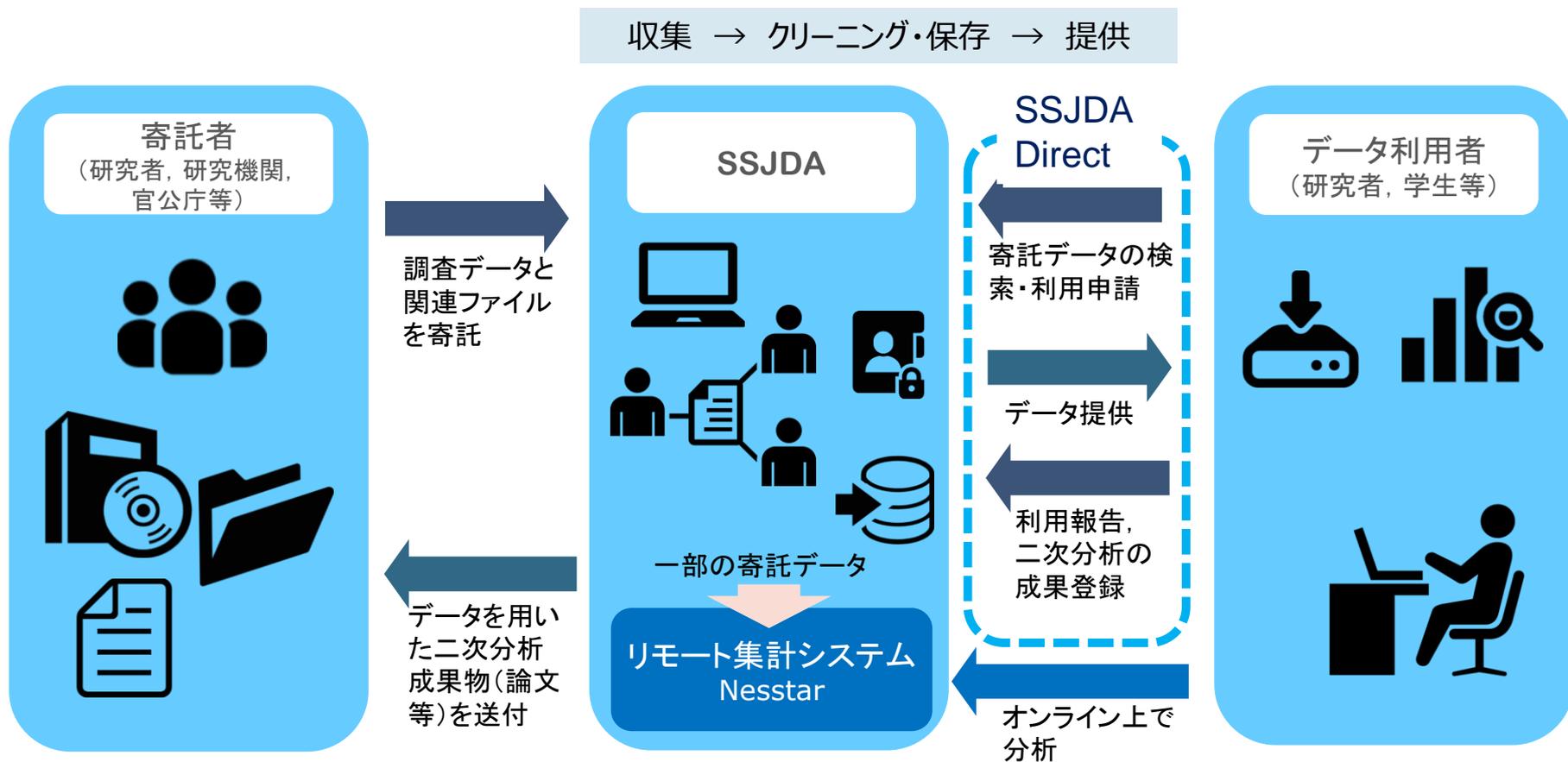
□ JES(投票行動の全国的・時系列的調査研究)

現代日本の政治意識や投票行動に関するパネル調査

□ SSM(社会階層と社会移動調査)

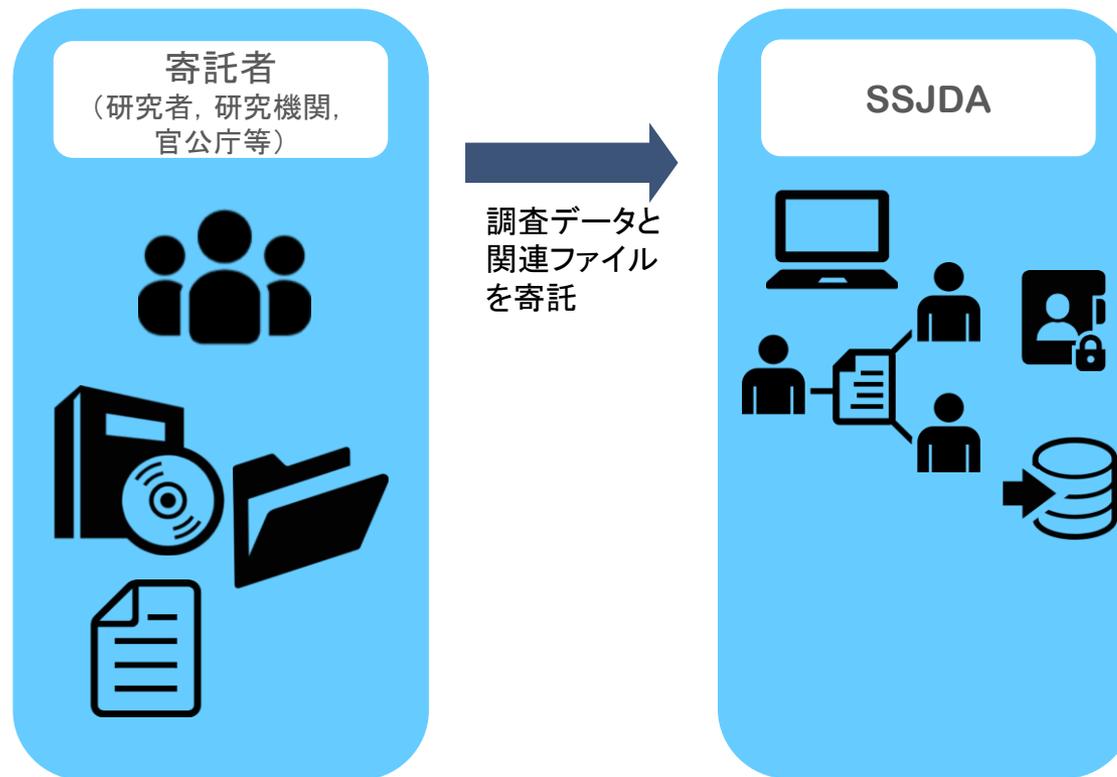
日本における不平等や格差，職業経歴に関する情報を多く含んだ調査

2. SSJDAにおけるデータの収集と提供のしくみ



2. SSJDAにおけるデータの収集と提供のしくみ

①データを収集する



また、新たなデータを収集し続けるために、以下のような研究者・研究機関へ積極的に寄託を働きかけている。

- これまでにデータを寄託した寄託者
- KAKEN（科研費データベース）を通じてピックアップした、社会調査を実施した研究者
- 社会調査を実施している機関
- 社研教員のつてを通じて など…

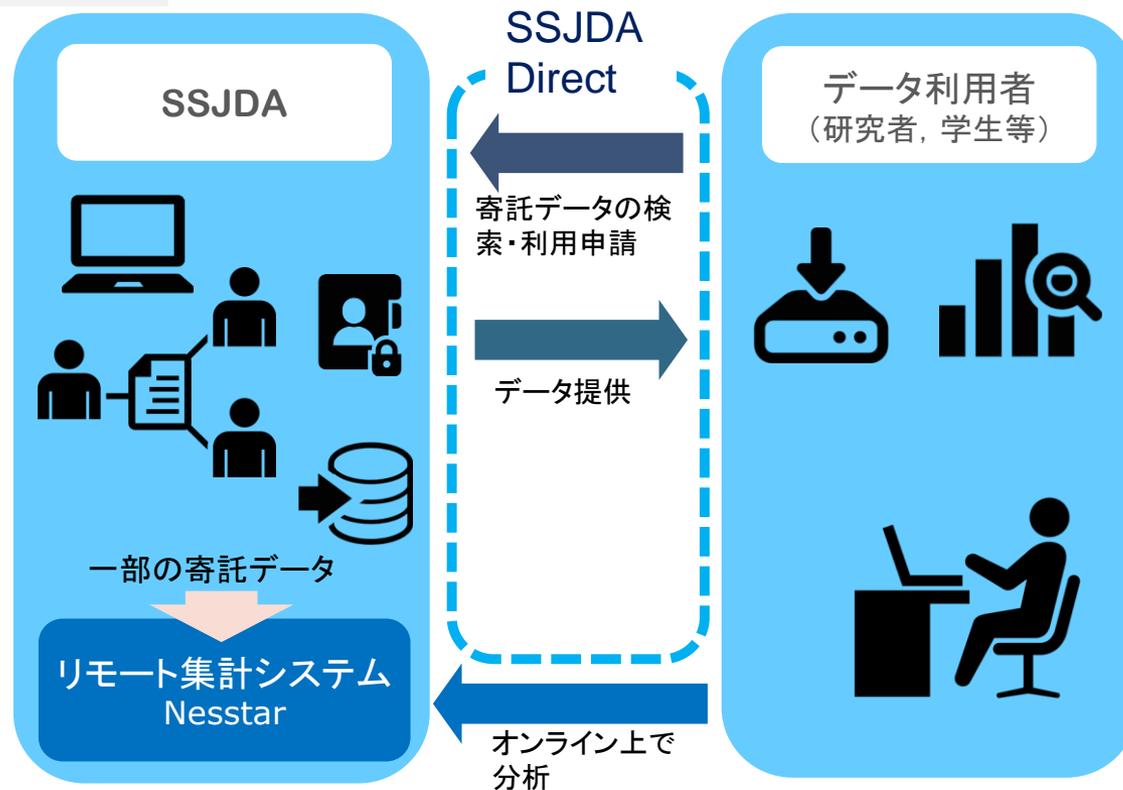
②データのクリーニング作業と保存



- 寄託者から寄託されたデータのクリーニング（変数の命名，無回答・非該当の識別，単純集計の検証など），プライバシーや企業秘密にかかわる情報を保護するための秘匿処理を行い，DAT，SPSS，Stata形式に変換する。
 - 寄託データについて，公開へ向けてメタデータ作成を行う。メタデータは社会調査メタデータの国際規格であるData Documentation Initiative(DDI)に準拠し，人文学・社会科学総合データカタログ（Japan Data Catalog for the Humanities and Social Sciences，JDCat）のメタデータスキーマに沿った独自のメタデータ項目を用いている。
 - 寄託者から許諾を得た一部のデータは，オンライン上で利用可能なりモート集計システムにデータを搭載する。
- すべての工程はデータアーカイブ運営担当である教授2名の指導助言と助教2名，データマネージャー1名のチェックのもとで専門スタッフにより行われる。

2. SSJDAにおけるデータの収集と提供のしくみ

③データの公開と提供



- オンラインデータ検索・提供システム SSJDA Directを通じてメタデータを公開。検索は誰でも操作可能
- アカウントを持っていれば, SSJDA Directを通じて利用申請できる。申請承認後にデータダウンロードが可能 (一部例外として郵送対応の特別データあり)
- 年間70件~80件程度のペースで新規データを公開
- 一部のデータについてはSSJDA Directに加えて, リモート集計システム Nesstarを通じて提供

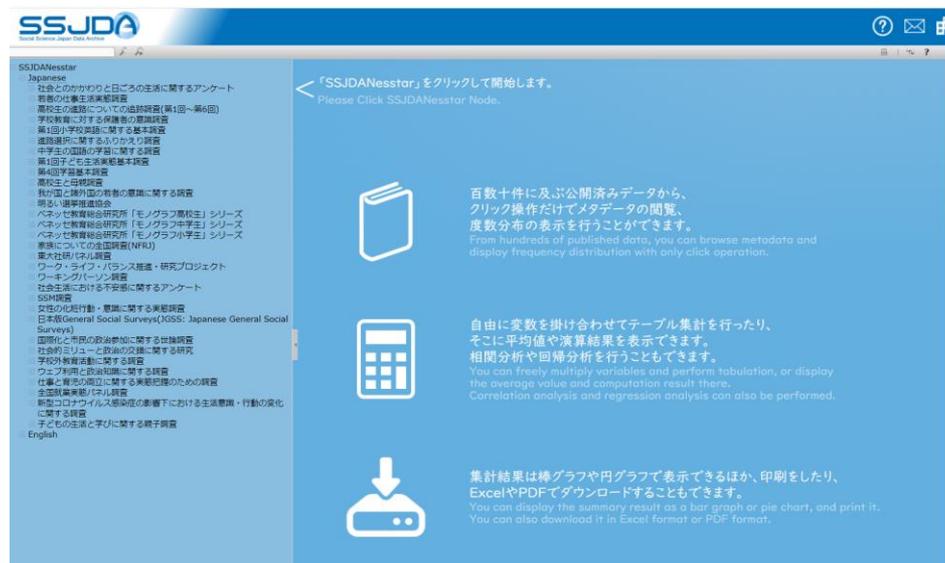
2. SSJDAにおけるデータの収集と提供のしくみ

④ より開かれたデータ提供方法：リモート集計システム

リモート集計システムNesstarは、社会調査系メタデータの国際標準であるData Documentation Initiative(DDI)に対応したシステム。CSRDAのウェブサイトにて公開。

Nesstarはオンライン上で誰でも単純集計や簡単な分析を行うことが可能となっている。

- 利用登録不要
- 利用資格なし
- 利用料なし
- 集計結果や分析結果は棒グラフや円グラフで表示できるほか、印刷やExcel形式、PDFでのダウンロードが可能

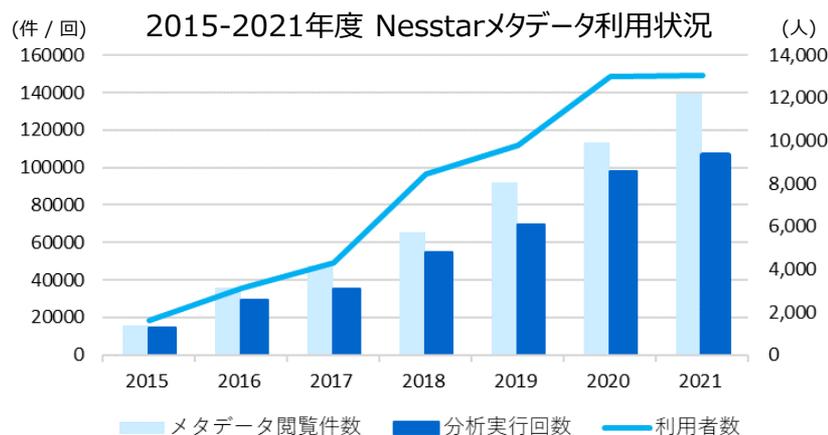


Nesstarの利用開始画面

ただし、

- 利用できるデータはSSJDA Directから公開されているうちの、一部のものに限られる
- 簡単な分析に限られる
- 個票データ自体はダウンロードできない。分析はNesstarのシステム上でのみ実行可能
- 利用誓約事項あり (<https://csrda.iss.u-tokyo.ac.jp/infrastructure/remote-analysis/>)

右表の通り、利用者は年々増加している。2023年度には新たなリモート集計システムの開始を予定。



3. SSJDAデータ利活用の状況

- 収録データは、CSRDAウェブサイト上で公開されているオンラインシステムSSJDA Direct (図1) を通じて、検索・利用申請が可能.
- 利用者数は年々増加している(表1).
- 近年では教育利用目的での利用申請も増加している(表2).



図1 SSJDA Direct データ検索画面

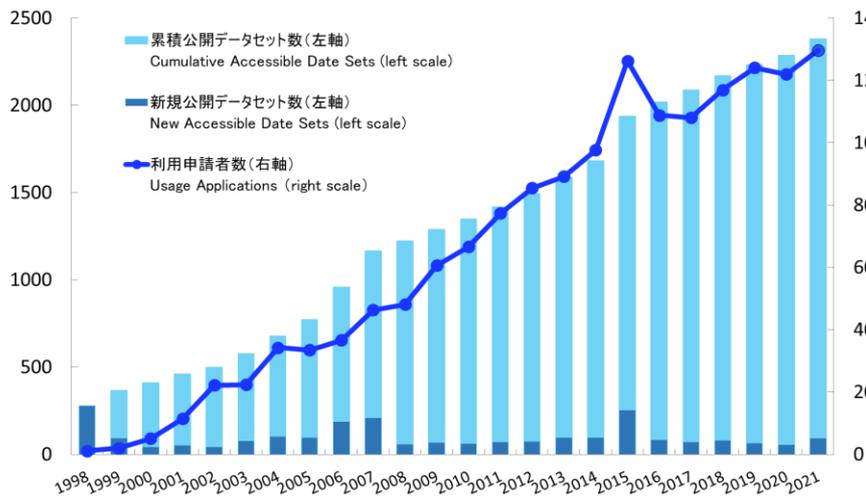


表1 公開データセット数の推移と利用状況 (2022年3月末時点)

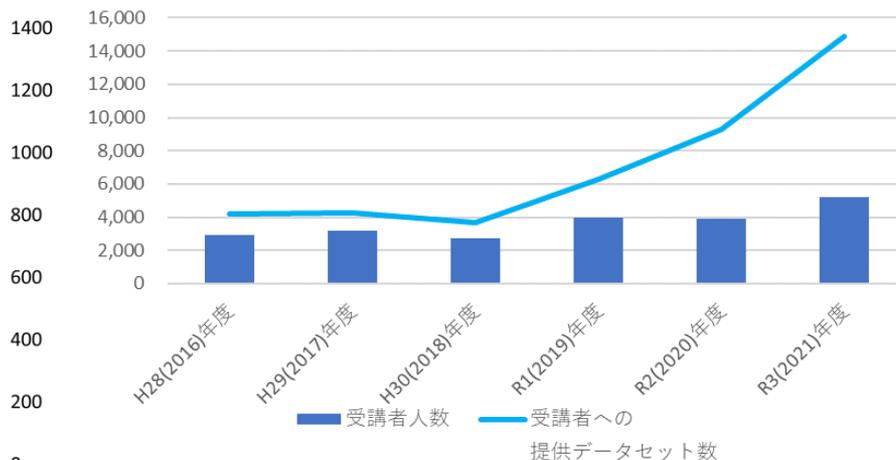
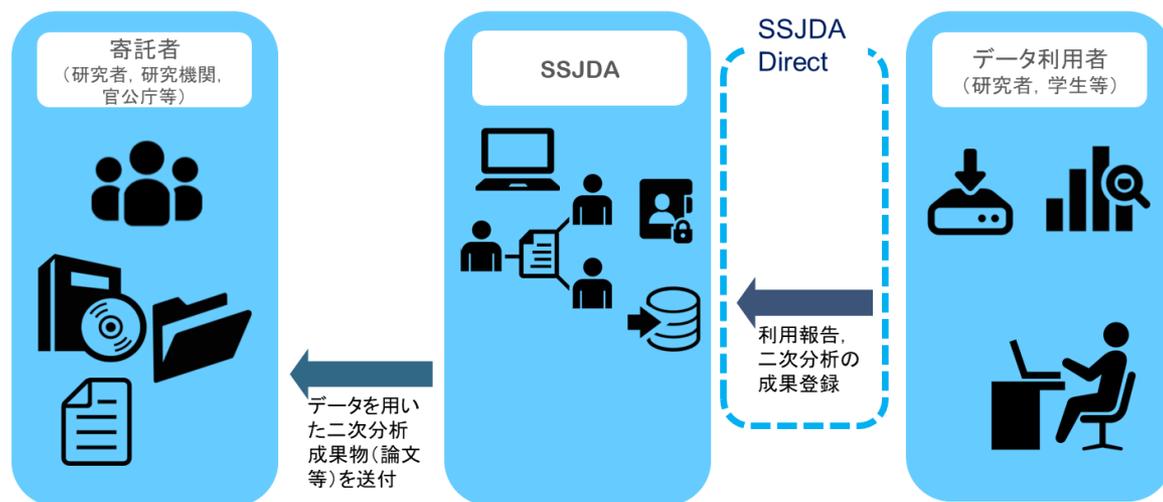


表2 提供データセットを用いた講義等の受講者人数と受講者への提供データセット数

3. SSJDAデータ利活用の状況



データの利用報告と成果登録の流れ

- 利用者はデータを用いて執筆された論文・著書などの研究成果物を、SSJDA Directを通じてSSJDAに登録
- SSJDAは利用者から報告された成果物をデータ寄託者に報告
- 成果物には学位論文の報告も多い（下表）
- 良質で利用者数の多いデータの寄託者と、優れた論文を毎年表彰する制度がある（下写真）
- インパクトファクターの高い国際誌に掲載される論文や、受賞論文も多く報告されている。

学部教育への貢献

公開・提供している個票データの一部は学部教育にも利用可能。
SSJDAから提供されたデータを用いて、毎年度卒業論文などの学位論文が多数執筆されている。

SSJDAデータ利用者より報告された学位論文数		
学位論文数	(うち博士論文)	
H28(2016)年度	128	3
H29(2017)年度	165	2
H30(2018)年度	161	6
R1(2019)年度	212	6
R2(2020)年度	216	4
R3(2021)年度	193	2

寄託者・優秀論文表彰 -独自の表彰制度-

毎年度末に、年間を通じて利用申請が多く質の良いデータを寄託した寄託者/機関を1,2名(機関)、そしてSSJDAのデータを利用した特に優れた論文を1,2本選出して表彰している。



寄託者・優秀論文表彰式

国内外に評価の高い優れた研究成果の例

西村幸浩「死刑制度に対する態度の規定要因の検討—生活不満と信頼に着目して—」 2021年4月

日本において死刑を支持する態度が、どのような要因によって規定されているかを明らかにすることを目的に、SSJDAの寄託データ「日本版General Social Surveys <JGSS-2010>」を用いて分析。その結果、家庭外の生活不満から政治的信頼が低下し、厳罰と死刑を支持するに至るという構造モデルが見いだされ、国内の学術誌『法社会学』に掲載された。
2022年5月、同論文は日本法社会学会第23回日本法社会学会奨励賞（論文部門）を受賞。

Sho Fujihara, Takahiro Tabuchi “The impact of COVID-19 on the psychological distress of youths in Japan: A latent growth curve analysis” Feb, 2022.

若者とその親のパネル調査データを用いて、COVID-19によって若者の心理的ディストレスがどのように変化したのか、また特にディストレスの変化の大きかった若者はどのような人なのかを明らかにした、社会科学・医学の領域横断的研究論文。国際学術誌Journal of Affective Disorders (IF4.839) に掲載。

4. 二次分析研究推進のための支援活動

公募型共同研究 二次分析研究会

二次分析研究の普及・促進および若手研究者の育成のために、二次分析に関わる共同研究を毎年度公募して開催している。年度末には成果報告会を開催し、成果報告書を刊行。

2022年度研究グループの一例

戦後福祉国家成立期の福祉・教育・生活をめぐる調査データの二次分析

テーマ

戦後福祉国家成立期の福祉・教育・生活をめぐる調査データの二次分析

使用データ

0001	新規学卒者(中卒)労働市場調査, 1953
1207	京浜工業地帯調査(従業員個人調査)
1213	貧困層の形成(静岡)調査
1311	団地居住者生活調査
1331	「ボーダー・ライン層」調査
労働調査24	国鉄女子労働者調査
労働調査61	福祉資金行政実態調査
労働調査62	高齢者生活実態調査
労働調査63	ソーシャル・ニーズ調査



成果報告会の様子



これまでの成果報告書
※ただし2020年度からオンラインでのPDF刊行のみに変更

計量分析セミナー

SSJデータアーカイブで公開されている調査データを利用した計量分析のセミナーを、毎年夏と春の2回開催している。講師はCSRDAのスタッフを中心に、学外からも講師を招聘。

※コロナ禍のため、現在はオンライン会議システムを用いたオンライン開催のみとなっている。

2022年度計量分析セミナー・夏 開講予定一覧(9月6-9日開催)

二次分析入門!

2022年9月6日(火) 10:30~17:00

担当講師:三輪 哲・池田 めぐみ・横内 陳正(東京大学)

Stataを用いた計量分析入門

2022年9月7日(水) 10:30~17:00

担当講師:麦山 亮太(学習院大学)

Rを用いた主成分分析と因子分析

2022年9月8日(木) 10:30~17:00

担当講師:森 知晴(立命館大学)

Rによる因果推論入門:再現可能性改善に向けたquarto, matching, multiple-testingの活用

2022年9月9日(金) 10:30~17:00

担当講師:川田 恵介(東京大学)



過去のセミナー開催の様子

5. 国内外データアーカイブ関連機関とのネットワーク

CSRDA (SSJDA)では他機関とのデータ相互利用推進のため、国内・海外のデータアーカイブとの連携強化を進めている。

国内

ICPSR国内利用協議会 ハブ機関として運営を担当
<http://jna-icpsr.jp/>

人文学・社会科学総合データカタログ JDCatとの連携
<https://jdcatalog.jp/>

*CSRDAは2018年度よりJSPSデータインフラ構築推進事業 拠点機関として認定を受け活動している。

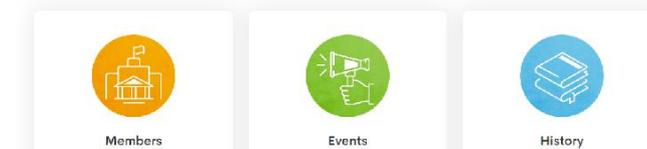
CiNii Research との連携
<https://cir.nii.ac.jp/>

海外

Network of Asian Social Science Data Archive (NASSDA) <https://nassda.org/>
2016年に日本 (SSJDA) , 韓国 (KOSSDA) , 台湾 (SRDA) , 中国 (CNSDA) と共同で結成。

International Federation of Data Organization (IFDO) <http://ifdo.org/>

Social Science One <https://socialscience.one/>



6. 課題と展望

- 予算，人員確保は継続的な課題。
- 人員面での都合から，従来はデータの寄託から公開までに一定の時間を要する，という課題がある。今後は，コロナ関連などのように社会情勢から公開の即時性が求められるようなデータは，優先的に迅速に公開できる体制を構築していく。
- 2023年度より，新たなリモート集計システムを公開予定。新システムでは新たにデータの変換機能を追加予定。マウスを使ったクリックによる操作（GUI）を中心にした，初学者向けのインターフェイス。スマートフォンでも閲覧可能な仕様であり，大学生，さらに高校生や中学生にも利用しやすいシステムを予定している。
- 2022年度内に，疑似データ（登録・利用資格不要，利用条件なし。非制限での公開データ）の提供を開始予定。こちらはSSJDA Directシステムを通すことなく直接ユーザーがデータをダウンロード可能なデータとなる。（学生や一般の方の分析練習用を想定）
- 今後も新たな寄託データを積極的に受け入れつつ，利用者の裾野を広げつつ，できる限り永続的なデータアーカイブ運営の維持を目指す。
- 国内，国外のデータアーカイブとの連携を深めてデータの利用促進を進め，データ利用の利便性向上と利用者の拡大を目指す。

ご清聴ありがとうございました

関連URL

東大社研CSRDA <https://csrda.iss.u-tokyo.ac.jp/>

SSJDA Research Paper Series (日本語) <https://csrda.iss.u-tokyo.ac.jp/infrastructure/rps/>

CSRDA Discussion Paper Series (英語) <https://csrdadps.com/>

CSRDA公式YouTube



CSRDA公式Twitter

