

SSJデータアーカイブにおける取り組み： これまでの成果と今後の展望

2024年3月11日

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フォーラム

横内陳正

東京大学社会科学研究所

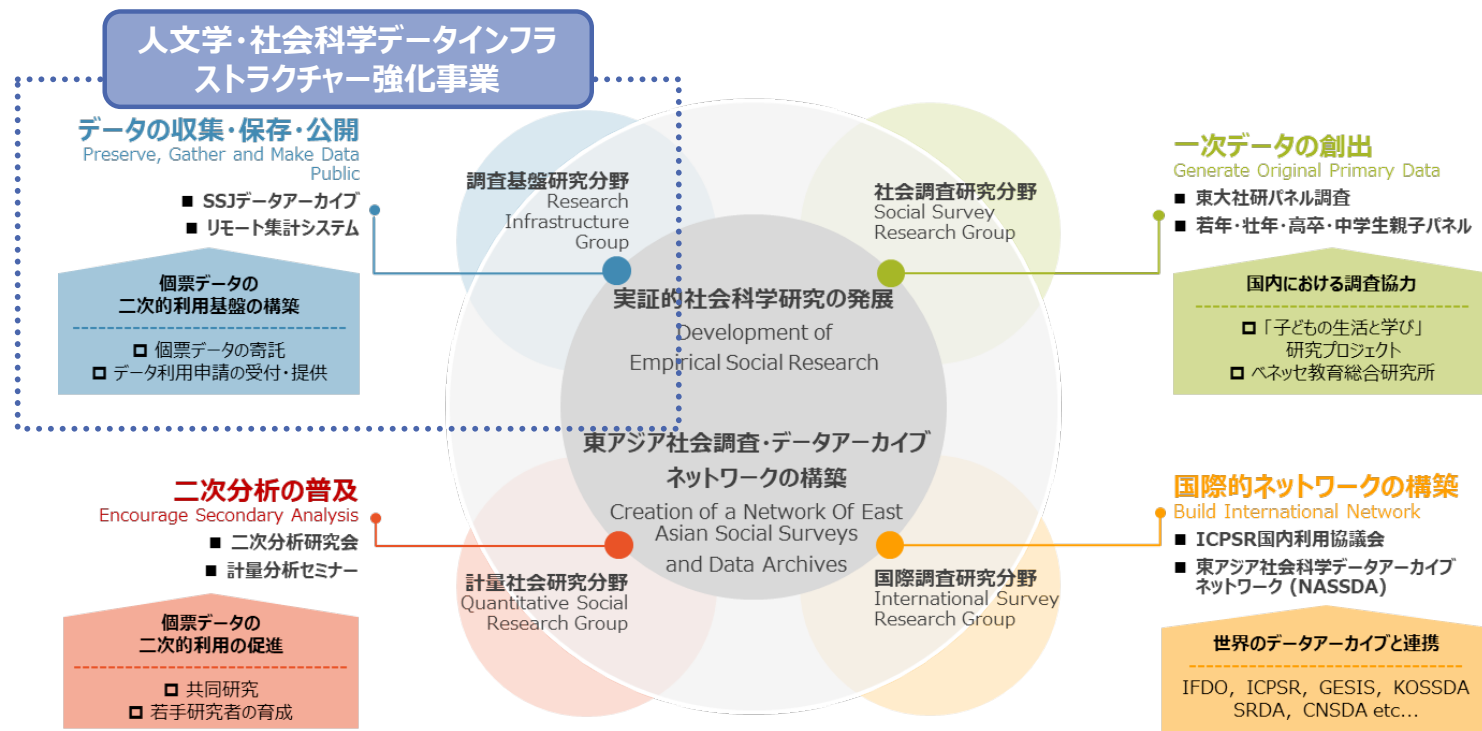
附属社会調査・データアーカイブ研究センター

社会科学とデータアーカイブ

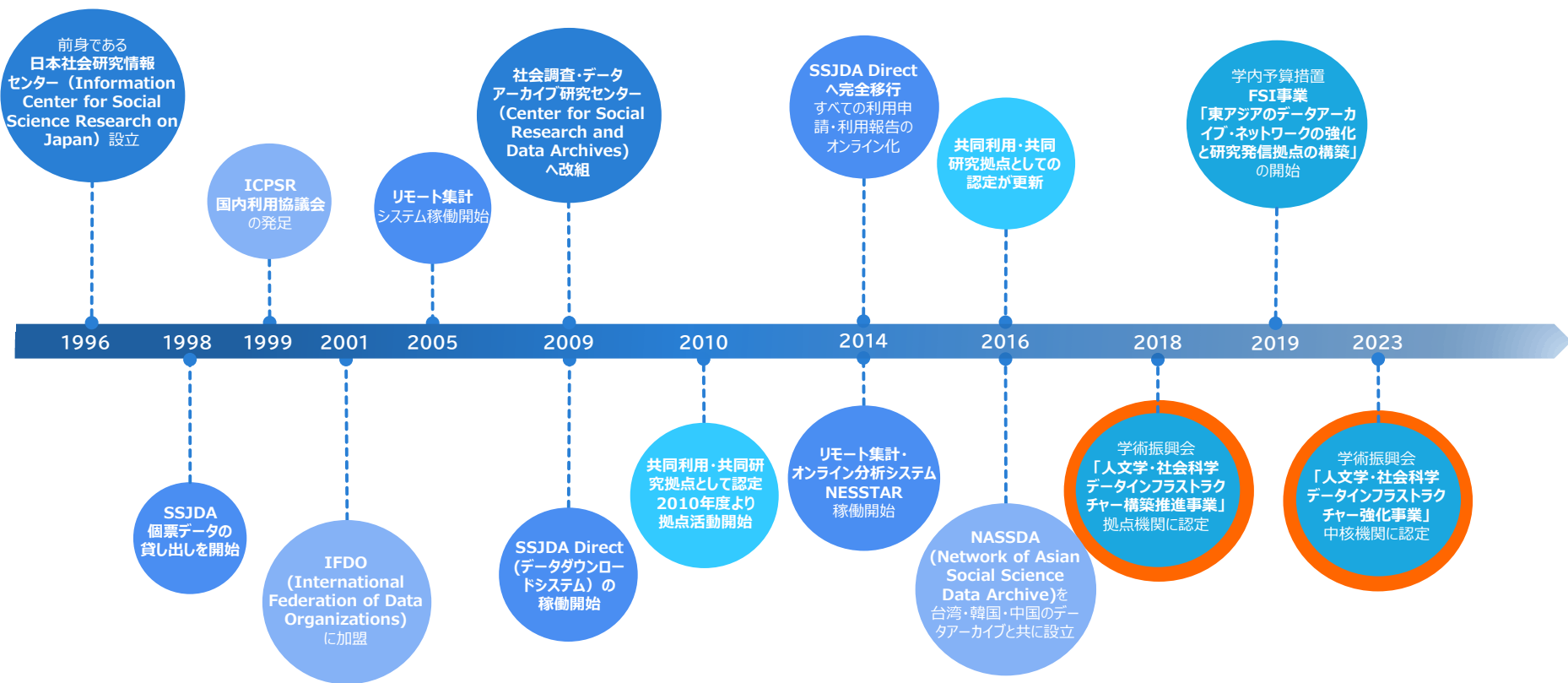
- 社会科学（Social Sciences）
 - 社会の中での人間行動や社会現象を対象とする科学の総称
 - 法学，政治学，経済学，経営学，社会学など
 - 研究データは主に社会調査によって収集される
 - 定量的データ：行政管理データ，集計データ，個票データなど
 - 定性的データ：インタビュー，フィールドワーク，文書記録など
- 「データアーカイブ」とは，こうした調査データを収集・保管，共有する機関
 - データ版の図書館のイメージ
- 「オープンサイエンス」や「データ駆動型科学」などの考えが重要視される時代
 - これらを支えるインフラとしてのデータアーカイブの重要性がより高くなっている

SSJデータアーカイブ

- 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) の事業のひとつ
 - SSJデータアーカイブ (Social Science Japan Data Archive: SSJDA)



社会調査・データアーカイブ研究センター(CSRDA)の沿革



SSJデータアーカイブの役割・意義・特長

■ 役割

■ データの収集・保管と共有

質問紙

問7. 現在（または直近）のお仕事に関して、次にあるAからGのそれぞれについて、あてはまる程度をお答えください。（○はそれぞれにつき1つ）

	かなりあてはまる	ある程度あてはまる	あまりあてはまらない	あてはまらない	部下は いない
A. 自分の仕事のペースを、自分で決めたり変えたりすることができる	1	2	3	4	X
B. 職場の仕事のやり方を、自分で決めたり変えたりすることができる	1	2	3	4	
C. 部下の仕事のやり方を、自分が決めている	1	2	3	4	
D. 教育訓練を受ける機会がある	1	2	3	4	
E. 仕事を通じて職業能力を高める機会がある	1	2	3	4	
F. 子育て・家事・勉強など自分の生活の必要にあわせて、時間を短くしたり休みを取るなど、仕事を調整しやすい職場である	1	2	3	4	
G. 今後1年間に失業（倒産を含む）をする可能性がある	1	2	3	4	

あなたのふだんの生活についておうかがいします。

問8. あなたはどのくらいの頻度で以下のことをしていますか。（○はそれぞれにつき1つ）

	毎日	週に5~6日	週に3~4日	週に1~2日	月に1~3日	ほとんどしない
A. 運動（ウォーキング・ジョギング・エアロビクス・水泳・テニスなど）	1	2	3	4	5	6
B. 1日に3食を食べる	1	2	3	4	5	6
C. 栄養バランスの取れた食事を取る	1	2	3	4	5	6
D. カップ麺やファーストフードを食べる	1	2	3	4	5	6
E. 食事の用意	1	2	3	4	5	6
F. 洗濯	1	2	3	4	5	6
G. 家の掃除	1	2	3	4	5	6
H. 日用品・食料品の買い物	1	2	3	4	5	6
I. 友人・恋人（配偶者は除く）と食事をする	1	2	3	4	5	6
J. 友人・恋人（配偶者は除く）と話をする	1	2	3	4	5	6
K. インターネットを利用する（仕事以外で）	1	2	3	4	5	6

個票データ

	ZQ07A	ZQ07B	ZQ07C	ZQ07D	ZQ07E	ZQ07F	ZQ07G
1	2	3	1	2	2	4	4
2	2	2	2	4	1	2	3
3	2	3	3	2	4	2	4
4	1	2	2	2	1	2	4
5	3	2	3	4	3	1	4
6	2	2	2	3	2	4	3
7	2	3	4	2	2	2	3
8	2	4	5	4	1	1	4
9	2	3	4	1	2	2	3
10	2	2	5	3	2	2	4
11	2	2	5	3	1	3	3
12	3	2	2	2	2	3	1
13	2	4	4	4	3	3	4
14	2	3	4	2	3	4	4
15	1	3	5	2	2	1	4

SSJデータアーカイブの役割・意義・特長

■ 役割

- データの収集・保管と共有

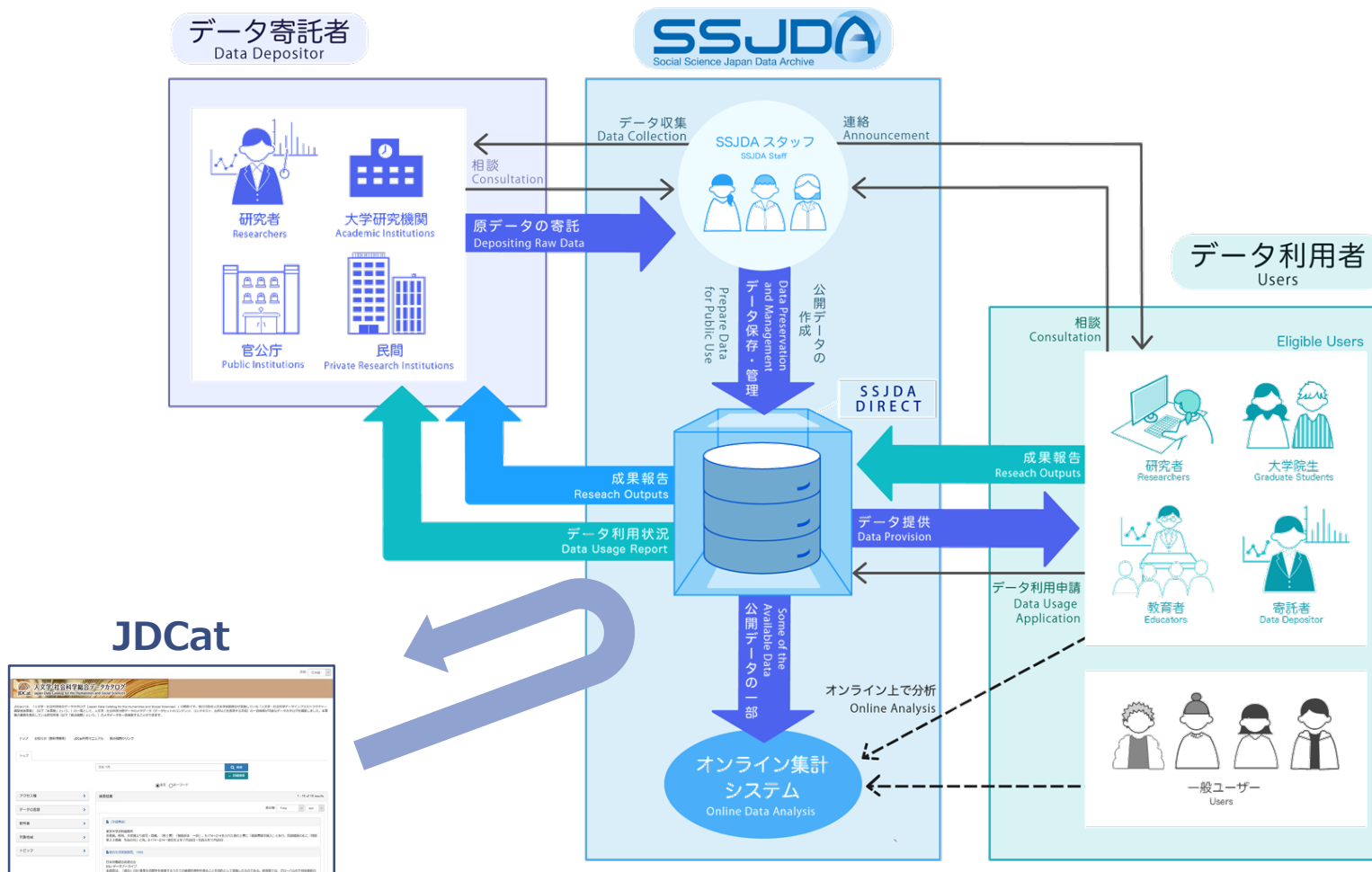
■ 意義

- データの散逸を防ぎ、安心して長期保存できる
- データが見つけやすく、アクセスしやすく、相互運用しやすく、再利用しやすくなる
- 新たな視角からの二次分析が可能になったり、分析の再現性を担保できる
- 大学院生や若手研究者にとっての研究資源であり、学部生に対する教育資源にもなる

■ 特長

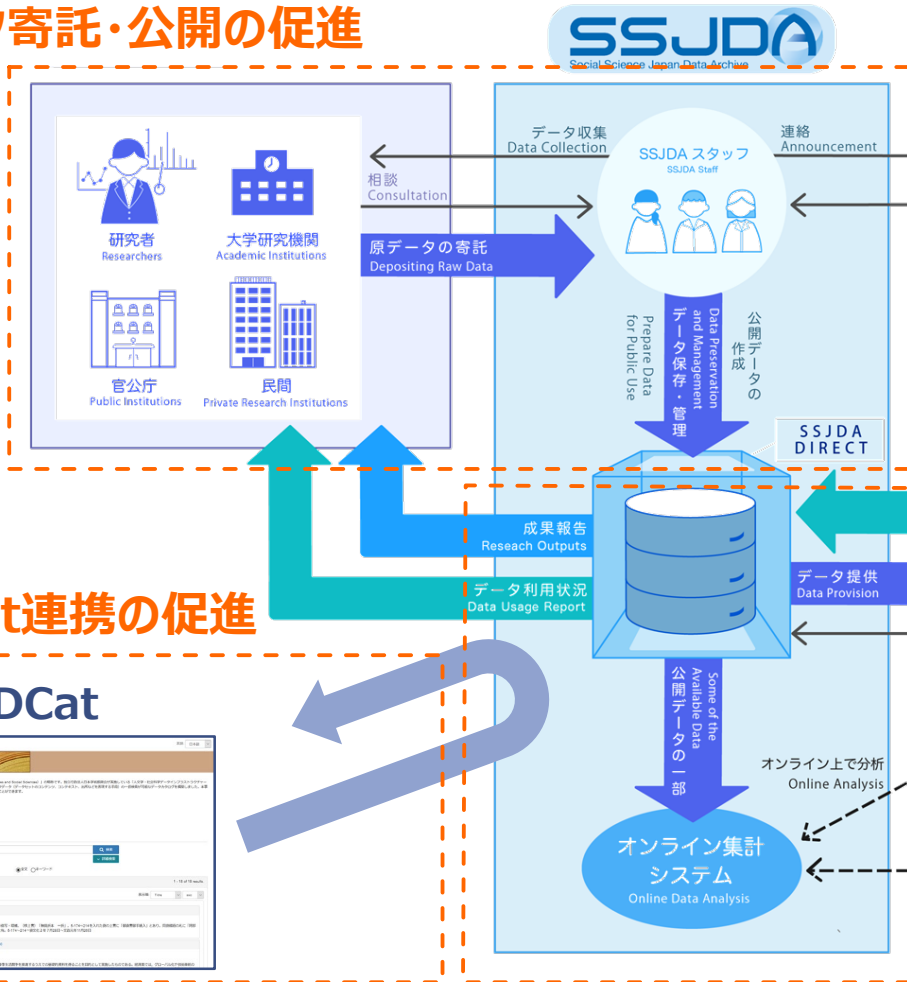
- 幅広い機関・研究者の方々から調査データが寄託されている
- 公開データを使った研究や教育を支援している
- データアーカイブ間の学際的、国際的な連携を行っている
- 寄託者や利用者への親身なサポートを行っている

SSJデータアーカイブの仕組み



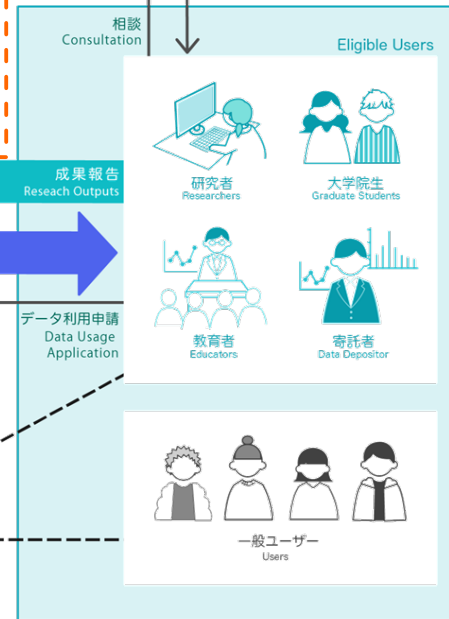
本事業における主な活動計画

1. データ寄託・公開の促進



- 4. 学際的・国際的な活動
- 5. その他の新たな活動

2. データ利活用の促進



3. JDCat連携の促進



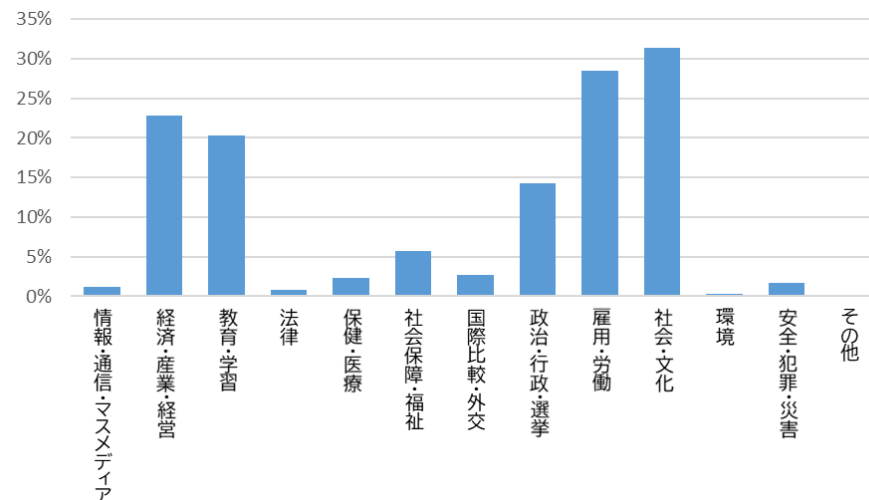
1. データ寄託・公開の促進

- 幅広い組織・研究者の方々からのデータ寄託
 - これまでに200以上の組織・研究者（代表者）から寄託
 - 約1,700の調査データを公開・提供

代表的な調査データと寄託者

JLPS（働き方とライフスタイルの変化に関する全国調査）
SSM（社会階層と社会移動全国調査）
JES（投票行動の全国的・時系列的調査研究）
全国就業実態パネル調査（リクルートワークス研究所）
学校教育に対する保護者の意識調査（ベネッセ教育総合研究所）
日本人の意識調査（NHK放送文化研究所）
食育に関する意識調査（農林水産省）
仕事と育児の両立に関する実態把握のための調査（厚生労働省）

調査データのトピック別割合



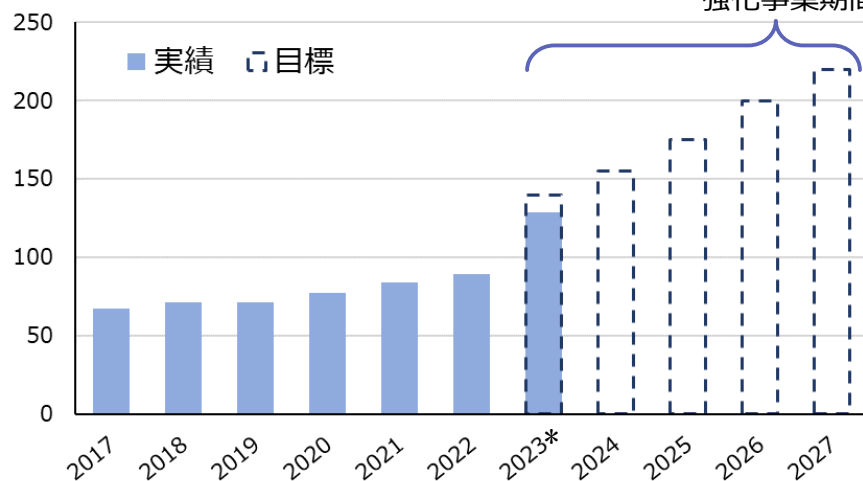
（2023年6月時点）

1. データ寄託・公開の促進

- 寄託を継続・拡大させるための取り組み
 - 学会メーリングリストを用いた発信；個別相談会の実施
 - 既存の寄託者への継続的なコンタクト
 - 寄託者用マイページの新設・改良
 - 寄託者への表彰事業

新規データ寄託件数の推移

人社データインフラ
強化事業期間



*2024年2月時点（差し替え分を除く）

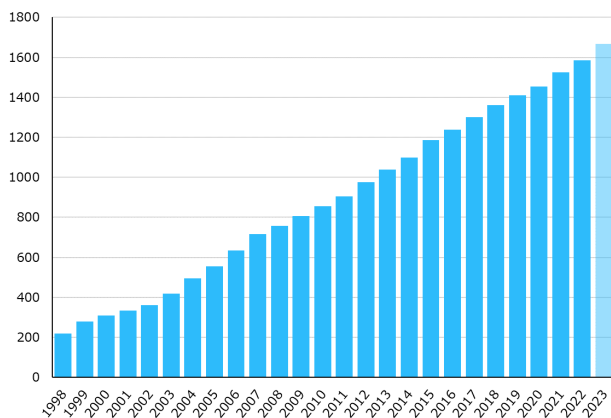
寄託者用マイページの画面

データ情報	
調査番号 (半角英数)	<input type="text"/>
* 調査名	<input type="text"/>
調査名(英語) (半角英数)	<input type="text"/>
* 目的	<input type="text"/>
* リモート集計 システム利用	<input type="text"/>
* 承認手続き	<input type="text"/>
* 報告書複写	<input type="text"/>
* データ登録	未登録 (0MB) ファイルを選択 選択されていません * 1つのZIPファイルに圧縮してからアップロードください。 * ZIPファイルのサイズが40MBを超える場合は、ssjda@iss.u-tokyo.ac.jpまでお問い合わせください。
公開状況ステータス	寄託者登録中

1. データ寄託・公開の促進

- 個票データのキュレーション
 - 公開の迅速化
- メタデータ作成
 - AIの活用
 - 国際規格（DDI）への対応と見直し
 - メタデータの英語化

公開調査データ数の推移



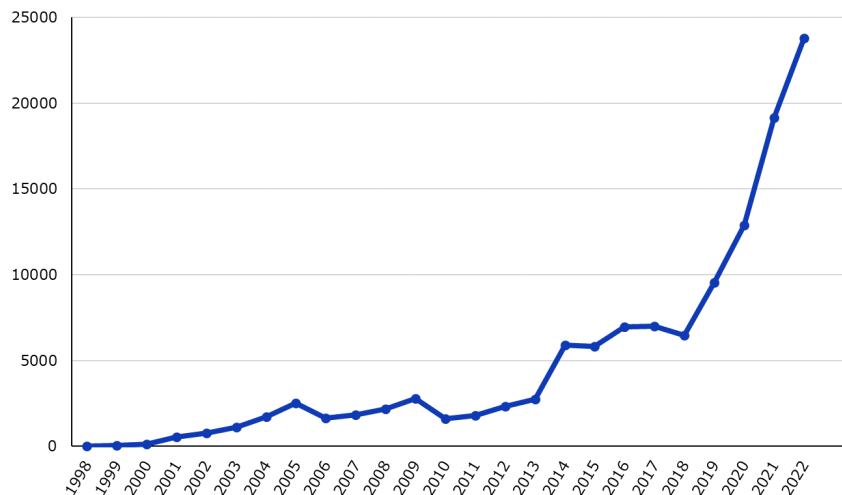
メタデータ

概要	
調査番号	PM050
調査名	東大社研・社年/パネル調査 (JLPS-M) Wave 1-5, 2007-2011
寄託者	東京大学社会科学研究所/パネル調査プロジェクト
利用申込先・承認手続き	利用方法の詳細は こちら SSJDAが利用申請を承認したときに利用できる
教育目的(授業などの利用)	教育(授業・卒業等)も可
利用期限	研究はなし教育は一年
データ提供方法	ダウンロード
メタデータ閲覧・オンライン分析システムNesstar	利用不可
調査の概要	労働市場の構造変動、急激な少子高齢化、グローバル化の進展などにもとまれ、日本社会における就業、結婚、家族、教育、意識、ライフスタイルのあり方は大きく変化を遂げようとしている。これからの日本社会がどのような方向に進むのかを考える上で、現在生じている変化がどのような原因によるものなのか、あるいはどこが変化してどこが変化していないのかを明確にすることはきわめて重要である。
データタイプ(量的調査/質的調査/混合調査)	量的調査 量的調査: ミクロデータ
調査対象	日本全国に居住する35~40歳の男女 (2006年12月時点)
調査対象の単位	個人
サンプルサイズ	継続調査 回収数: 1067 回収率: 85% 追加調査 回収数: 251 回収率: 31%
調査時点	2011-01 ~ 2011-03 2011年1~3月 (wave5)
対象時期	2007 ~ 2011
調査地域	日本 全国
標本抽出	確率: 層別抽出 確率: 多段階抽出 層化2段階無作為抽出 ・地域 (10地域) と都市規模 (4類型) の2層により層化 (271地点) ・さらに性別・年齢別 (5歳階級) に層化 ・抽出台帳は、住民基本台帳を基本とし、住民基本台帳の閲覧が不許可になった地点では、選挙人名簿を使用
調査方法	自記式調査票: 紙 調査員による郵送配布、訪談回収法 ・対象者には追跡調査であることを事前に伝え、調査に協力を要請し、郵送で調査票を配布。その後記入された調査票を、調査員が訪問し回収。
調査実施者	東京大学社会科学研究所/パネル調査プロジェクト 実施は、(社)中央調査社
DOI	10.34500/SSJDA.PM050
寄託者(経典)	(「調査の概要」を参照)

2. データ利活用の促進

- データの利用方法
 - 利用申請
 - オンライン分析 (SSJDA Data Analysis)
 - 非制限公開 (疑似) データ

提供データセット総数の推移



オンライン分析システムの画面



2. データ利活用の促進

■ 利活用の促進

- データを利用した若手研究者による論文を対象とした表彰事業
- 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の事業
 - 二次分析にかかわる共同研究を公募する「二次分析研究会」
 - SSJDAの公開データを利用した「計量分析セミナー」(毎年2回)
- 社会科学データインフラ・ワークショップを毎年開催

JDcatの
データを利用

優秀論文(寄託者)表彰



二次分析研究会



計量分析セミナー

2023年度
計量分析セミナー・春

2024年3月4日~8日 (週々3回) オンライン開催
10:30~17:00

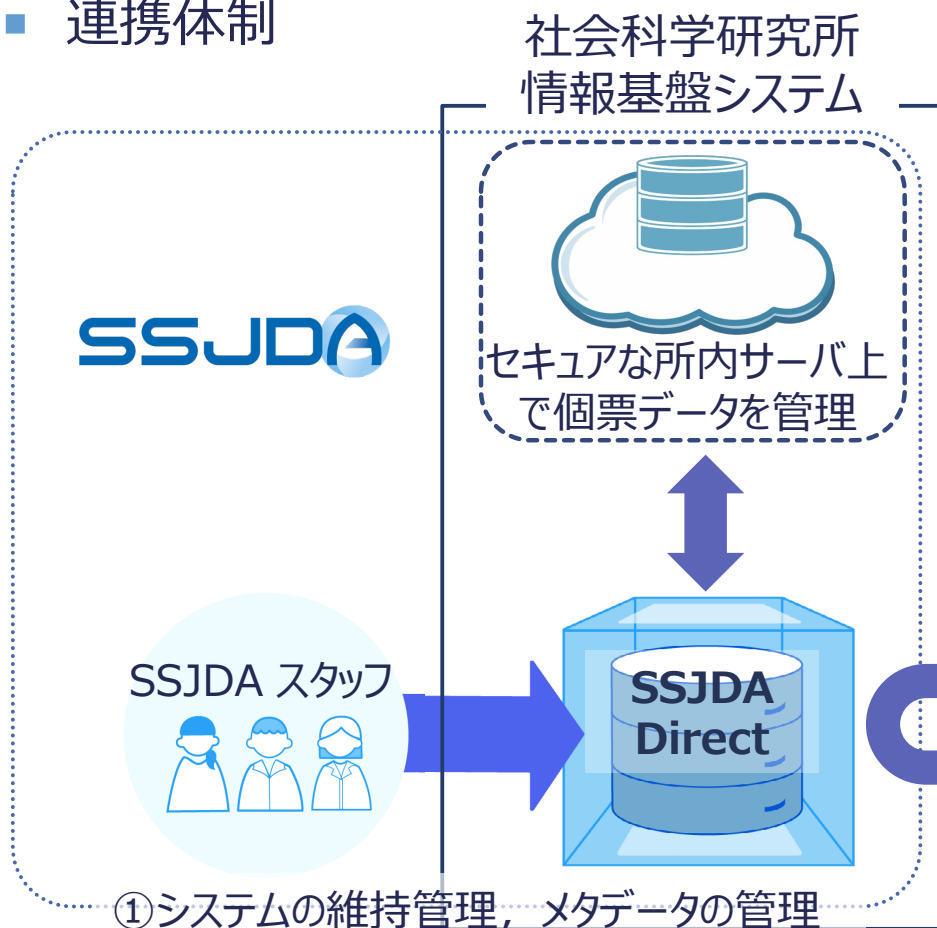
- March 4 Mon Statalによるパネル調査データ分析の実践
講師: 妻山 亮太 (学習院大学)
- March 5 Tue Rによる格差と因果効果の推定: 識別の対比と推定法の共有
講師: 川田 恵介 (東京大学)
- March 7 Thu イベント・ヒストリー分析/生存時間解析の理論
講師: 福田 百孝 (東北大学)
- March 8 Fri Rと機械学習による社会分析: 教師付き学習と因果推論/格差研究への応用
講師: 川田 恵介 (東京大学)

申込締切: 2024年2月2日(金)
定員(申込35名)になり次第締め切ります

講義詳細、お申込はウェブサイトをご覧ください
計量分析セミナー事務局: statom@iss.u-tokyo.ac.jp

3. JDCat連携の促進

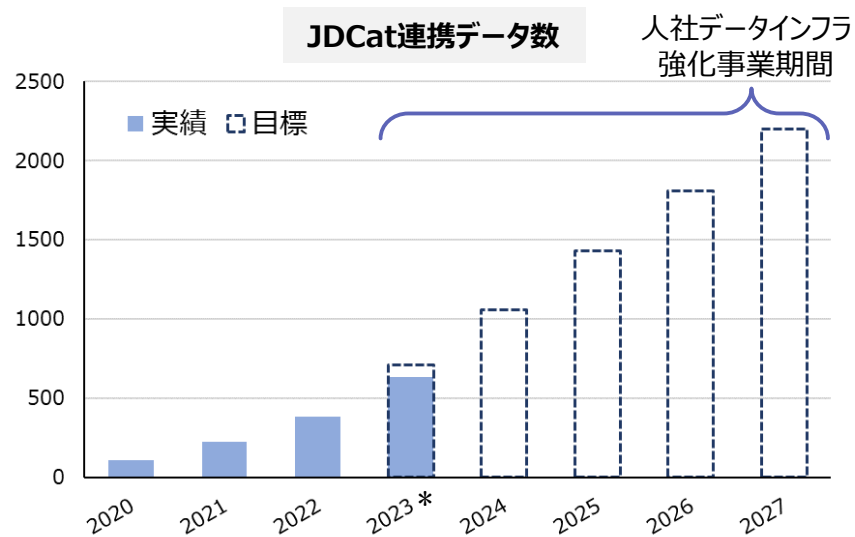
■ 連携体制



OAI-PMHプロトコルによる自動ハーベスティング

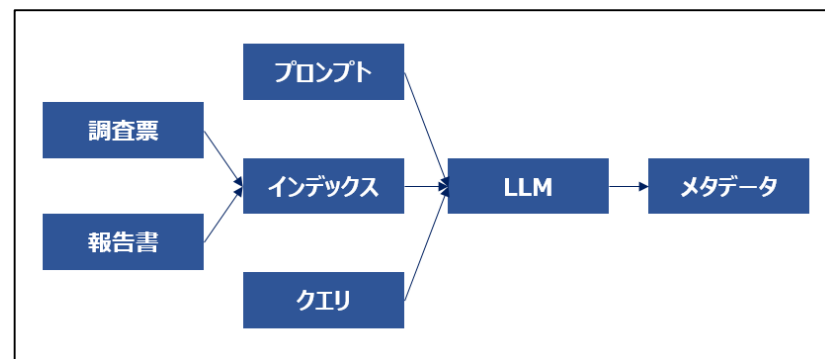
3. JDCat連携の促進

- 連携の継続と加速
 - メタデータ作成におけるAI活用
- 連携データの利用拡大や共同研究の促進
 - 二次分析研究会や計量分析セミナーなどを開催
- SNS等を通じた情報発信



*2024年2月時点

メタデータ作成AIのアーキテクチャ図




3. JDCat連携の促進

- メタデータスキーマや統制語彙に関する最新の動向を調査
 - 例) DDI-Lifecycle 4.0, DDI-CDI
- 収集した情報はJDCatの改修と運用のために関連機関にフィードバック
- CESSDA Vocabulary Serviceでの日本語版統制語彙のバージョン管理を行い、更新を検討

JDCat連携メタデータ

Item type	Harvesting DDI(1)
公開日	2024-01-29
タイトル	東大世研・世研(京大世研 (JLPS4)) Wave 1-5, 2007-2011 Japanese Life Course Panel Survey of the Middle-aged (JLPS4) Wave 1-5, 2007-2011
作成者	東京大学社会科学研究所(京大世研プロジェクト) 美野山, (社)中興調査社
記号	SSJデータアーカイブ
配布URL	https://icrda.iss.u.tokyo.ac.jp/ps4/
記号	SSJDA
配布URL	https://icrda.iss.u.tokyo.ac.jp/ps4/
記号	SSJデータアーカイブ
記号	SSJDA
所属性・言語	東京大学社会科学研究所(京大世研プロジェクト) Japanese Life Course Panel Surveys (JLPS) project, Institute of Social Science, The University of Tokyo
URI	https://icrda.iss.u.tokyo.ac.jp/Directories.php?web=PM050
URI	https://icrda.iss.u.tokyo.ac.jp/Directories.php?lang=english&web=PM050
アクセス権	制約付きアクセス
権利情報	https://icrda.iss.u.tokyo.ac.jp/ps4/access/
権利情報	https://icrda.iss.u.tokyo.ac.jp/ps4/access/condition/
概要	労働市場の構造変化、熟年世代の雇用、グローバル化の進展などにもとづいて、日本社会における就職、結婚、育児、養育、ライフスタイルのあり方は大きく変化を遂げつつあります。これらの日本社会がどのような方向に進むのかを解明する上で、現在進んでいる変化がどのような理由によるものなのか、あるいはどこが変化してどこが変化していないのかを明確にすることは極めて重要である。
対象時期	2007 - 2011
調査日	201101 - 201103
母集団	日本全国に居住する35～40歳の男女 (2006年12月時点) Male and female residents in Japan aged between 35 and 40 in December 2006
調査単位	個人 individual
サンプリング方法	確率 層別抽出 確率 多段階抽出 Probability Stratified, Probability Multistage
調査方法	自己応答調査書・紙 Self-administered questionnaire, Paper

DDI-Lifecycle 4.0



DDI-Lifecycle 4.0 Beta 1 Q&A

Technical Committee
28 February 2024

DOI: 10.5281/zenodo.10724547

15

4. 学際的・国際的な活動

■ 学際的な連携

- 国内外の会議でSSJデータアーカイブの取り組みや成果を発信
- 他のデータアーカイブやレポジトリ等に関する情報収集
- IASSIST, Japan Open Science Summit (JOSS), European DDI User Conference (EDDI), Research Data Alliance Plenary Meeting (RDA)

IASSIST (2023)

Development of the Self-Archiving System in the Social Science Japan Data Archive
Megumi Ikeda¹, Nobutaka Yokoyama² & Saeko Mura³
¹Institute for Social Sciences, The University of Tokyo; ikeda@iss.u-tokyo.ac.jp

Introduction

- The University of Tokyo, Institute of Social Science, Center for Social Research and Data Archives (CSRDA) created the Social Science Japan Data Archive (SSJDA) to support empirical research in social sciences in Japan and has been disseminating raw data since April 1998.
- The collection and storage have been smooth. Over 2,400 data sets have been put into the public domain.
- Our recent challenge is that we have been unable to keep up with the increasing volume of deposits. Therefore, SSJDA built a self-archiving system in February 2023.

Statistics of SSJDA

- Number of applications: 1,410 (75 applications from international researchers and graduate students around the world)
- The total number of data provisions is 23,800.

Self-Archiving System of the Social SSJDA

Overview of SSJDA

- What can depositors do with the Self-Archiving System?
 - Register raw data
 - Select conditions for data
 - Create metadata
 - Check the list of applications for use, etc.
- Status of Self-Archiving System Utilization
 - Of the 32 depositors, 21 have been self-archived (as of April 2023).
 - Depositors commented on the "convenience" of the system.

Challenges of our Self-Archiving Systems

- The lack of depositors creating metadata is a serious problem: many depositors need more experience creating FAIR metadata, and metadata creation is time-consuming, only one metadata currently created.
- The manual is already available. However, we need to show how to create metadata
 - To create benefits, such as shortening the time between deposit and release if they register metadata
 - Use AI to help create metadata
- Many people deposit data the previous way.
- More promotion needs to be done that the self-archiving system has been created.
- We need to inform depositors about the benefits of the self-archiving system (e.g., no need to contact CSRDA, no need to submit a deposit form).

JOSS (2023)

Data Documentation Initiative の紹介と活用事例

胡中孟徳 (konaka@iss.u-tokyo.ac.jp)
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
<https://csrda.iss.u-tokyo.ac.jp/>

SSJDA

4. 学際的・国際的な活動

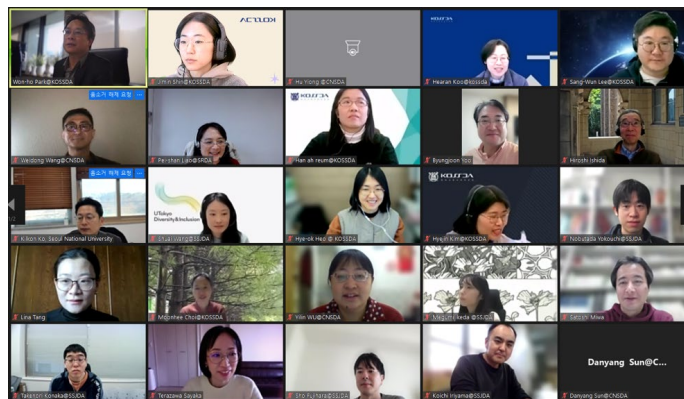
- 国際的な連携

- 海外から専門家を招聘し国際ワークショップを開催
- NASSDA (Network of Asian Social Science Data Archives)の創設
- ICPSR国内利用協議会のハブ機関
- IFDO (International Federation of Data Organizations)のBoard Member

国際ワークショップ (2024)



NASSDA (2024)



5. その他の新たな活動

- データリポジトリの国際認証CoreTrustSeal（CTS）の取得に向けた取り組み
 - データアーカイブのセキュリティ強化
 - 公開データの英語化
 - 研究成果の英語による発信を強化
- 質的調査データのメタデータ公開に関する検討
 - 海外機関へのヒアリングや情報収集

まとめ:本事業における主な活動計画

- **1. データ寄託・公開の促進**
 - 寄託を継続・拡大させるための各種取り組み
 - 個票データのキュレーションやメタデータ作成の効率化
- **2. データ利活用の促進**
 - 表彰事業, データの二次的利用に関する教育活動
 - 社会科学データインフラ・ワークショップ
- **3. JDCat連携の促進**
 - 連携の継続と加速のためのAI活用
 - 連携データの利用拡大や共同研究の促進
 - SNS等を通じた情報発信
 - 連携に必要なシステム改修の検討と実施
- **4. 学際的・国際的な活動**
 - 国内外の様々な分野における情報発信・共有
- **5. その他の新たな活動**
 - 質的調査データのメタデータ公開に関する検討