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

2025年1月31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

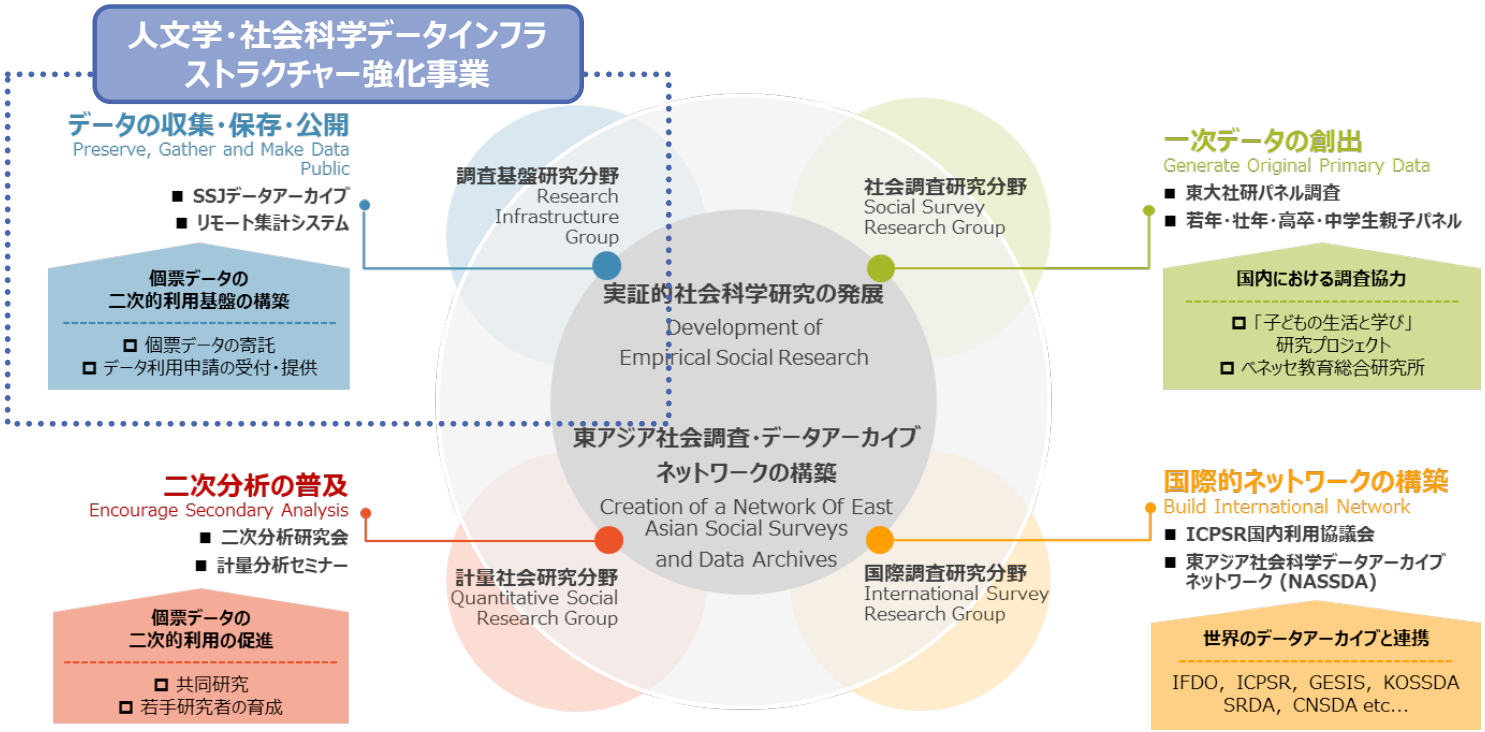
横内陳正

東京大学社会科学研究所

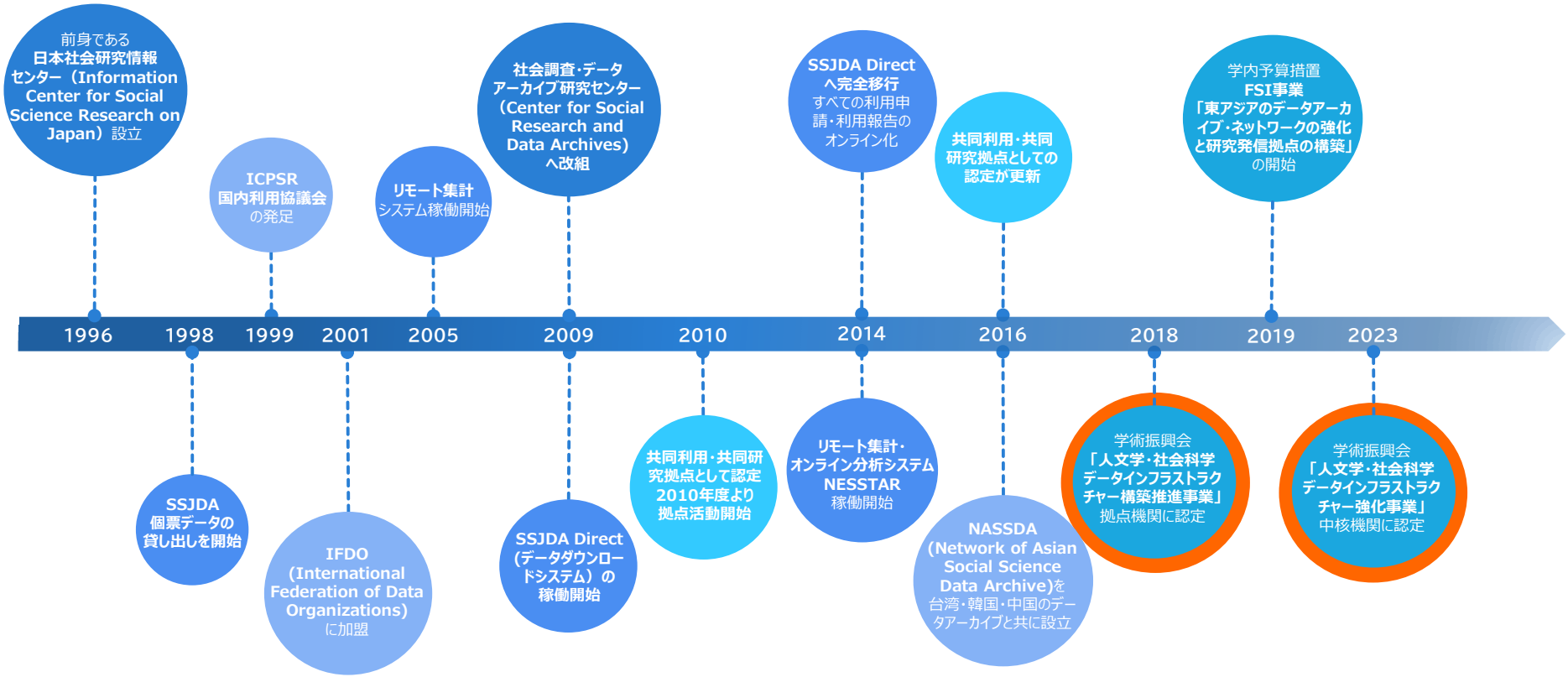
附属社会調査・データアーカイブ研究センター

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、次にあげる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	
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	5
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

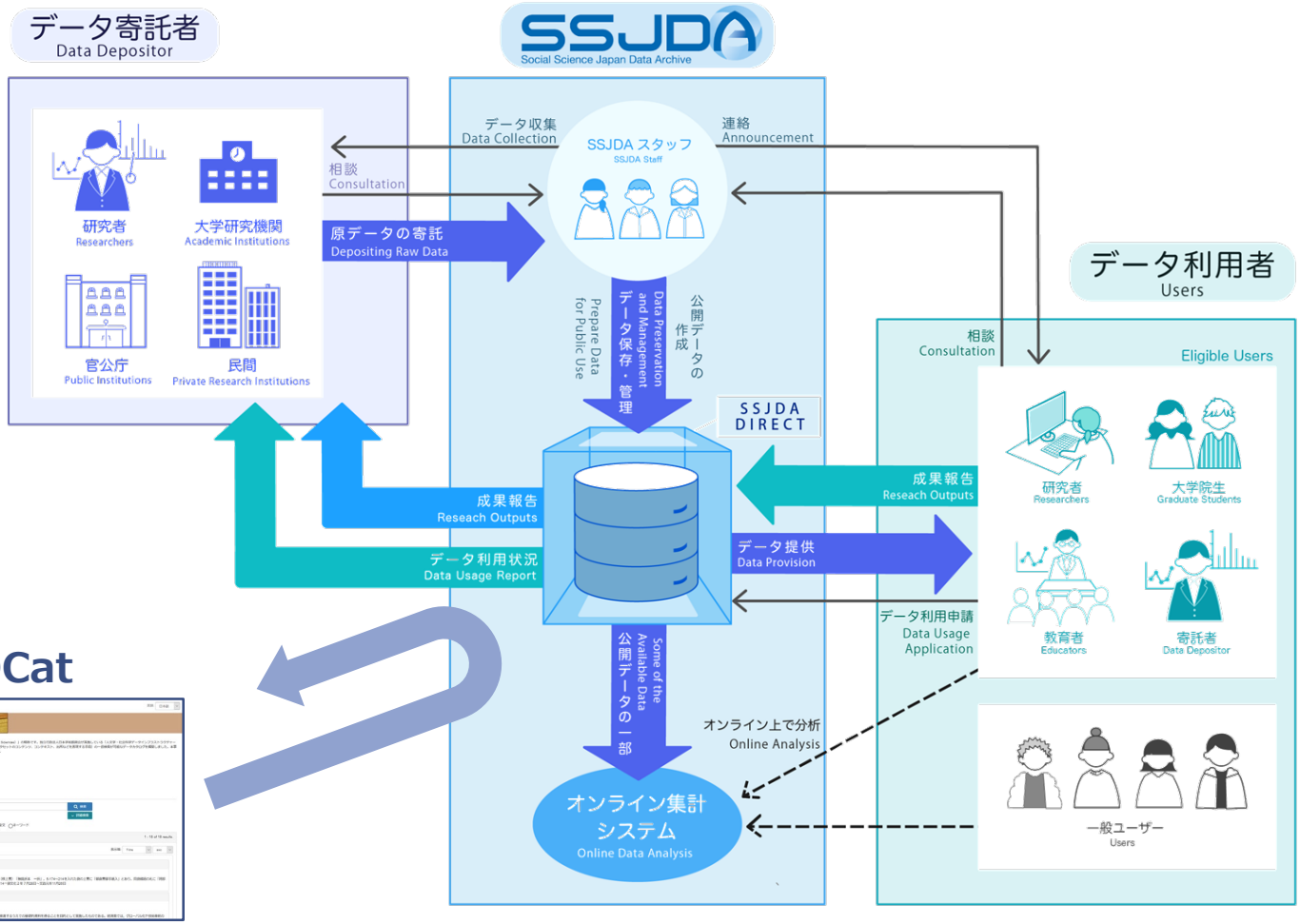
個票データ

	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



JDCat

本事業における主な取り組み状況

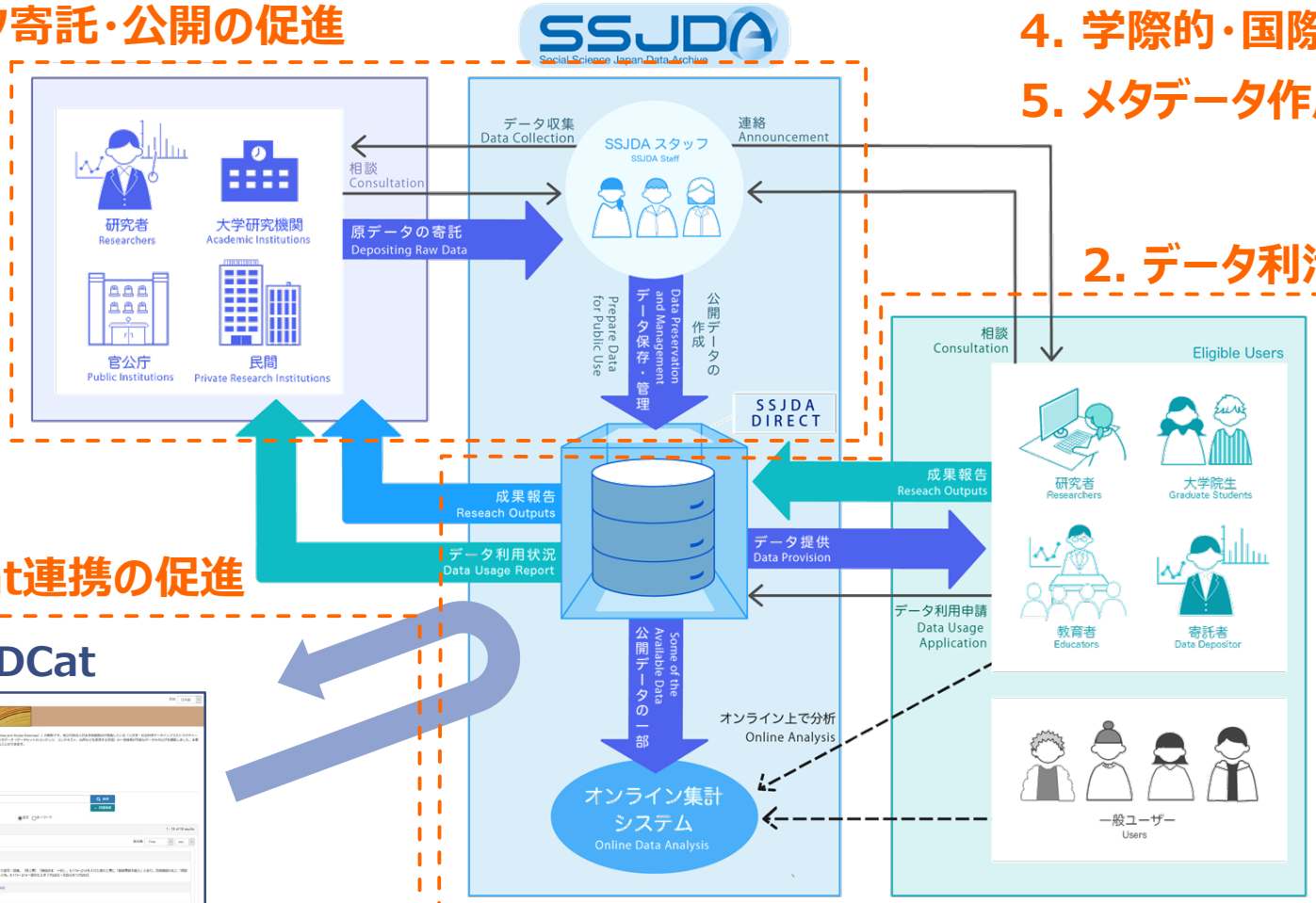
1. データ寄託・公開の促進

4. 学際的・国際的な活動

5. メタデータ作成へのAI導入

2. データ利活用の促進

3. JDCat連携の促進



JDCat

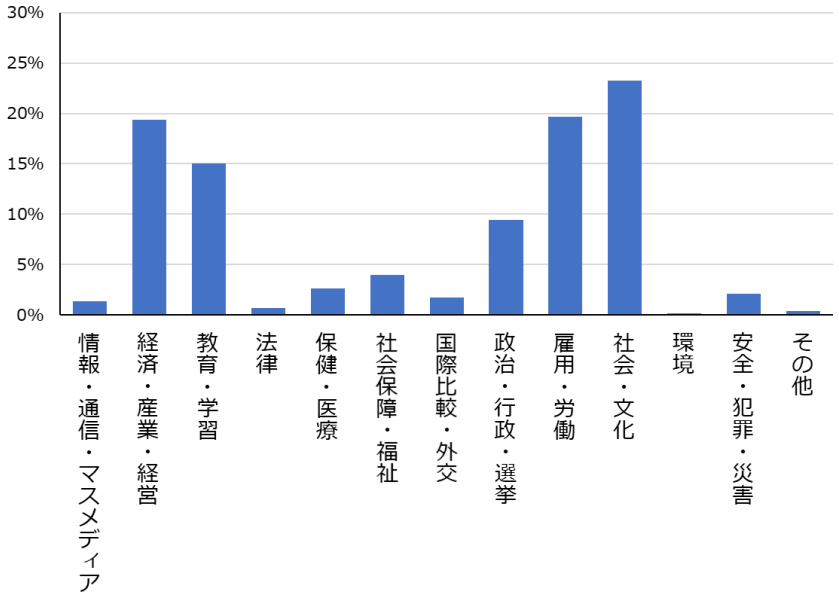
1. データ寄託・公開の促進

- 幅広い組織・研究者からのデータ寄託
  - これまでに200以上の組織・研究者から寄託
  - 約1,900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



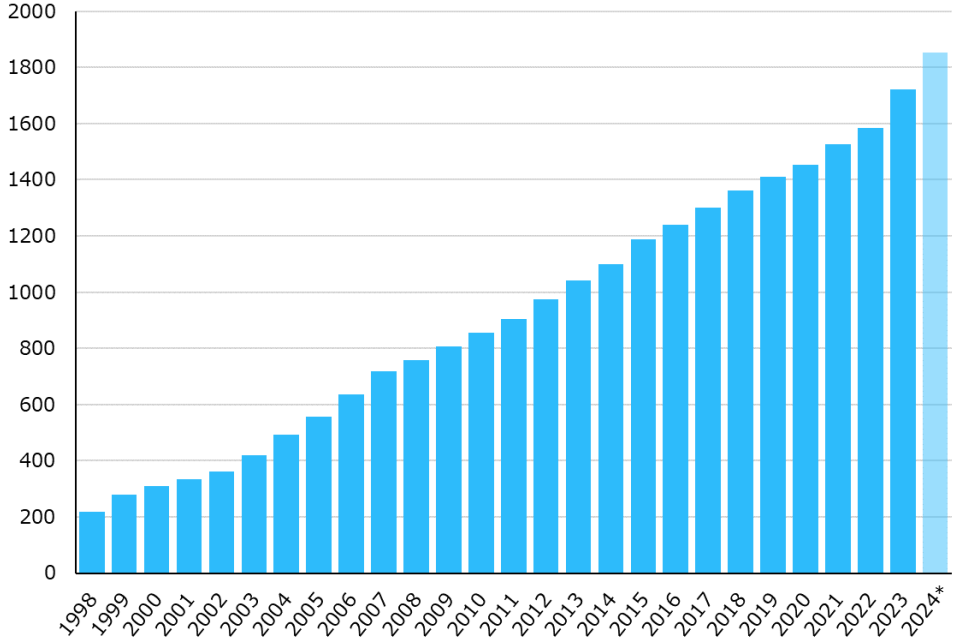
（2025年1月時点）

1. データ寄託・公開の促進

- データ寄託の促進
  - 個別相談会の実施
  - 寄託者への表彰事業
- データ公開の促進
  - AIを活用したメタデータ作成

寄託者表彰

公開調査データ数の推移



*2025年1月時点

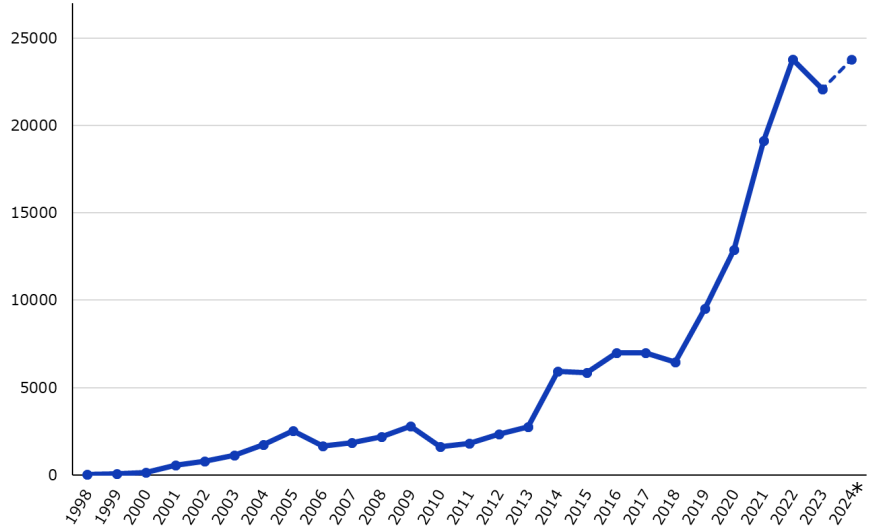
2. データ利活用の促進

- データの利用方法
  - 利用申請
  - オンライン集計システム (SSJDA Data Analysis)
  - 非制限公開疑似データ

非制限公開疑似データ

名前	ラベル
caseid	
sex	問1(1)_性別
ybirth	問1(2)_生年
mbirth	問1(2)_生月
ZQ03	問3_就業状況
JC_1	問4A(1)_現職・働き方
JC_41	問4A(4-1)_現職・従事時間（1日あたり）

提供データセット総数の推移



*2025年1月時点

オンライン集計システム

SSJDA Data Analysis

東大社研・若年パネル調査(JLPS-Y)wave1-6, 2007-2012(継続調査)

メタデータ

保存済みの出力のみ表示

クロス集計 > sex: 問1(1)_性別, ZQ03: 問3_就業状況

N 行% 列%

			問3_就業状況		
			している	していない	合計
問1(1)_性別	male	N	1514	179	1693
		行%	89.4%	10.6%	100.0%
	female	N	1248	426	1674
		行%	74.6%	25.4%	100.0%
	合計	N	2762	605	3367
		行%	82.0%	18.0%	100.0%

	値	自由度	漸近有意確率	正確有意確率
Pearson	126.3558	1	0.0000***	
連続修正	125.3486	1	0.0000***	
尤度比	129.3983	1	0.0000***	
Fisher				0.0000***
ファイ係数	0.1937			
クラメールのV	0.1937			

2. データ利活用の促進

- データ利活用の促進
  - 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および若手研究者の育成のための事業
    - 課題や参加者を公募する「二次分析研究会」
    - 誰でも参加可能な有料の「計量分析セミナー」
  - 社会科学データインフラ・ワークショップを開催

JDCatとの連携データを利用

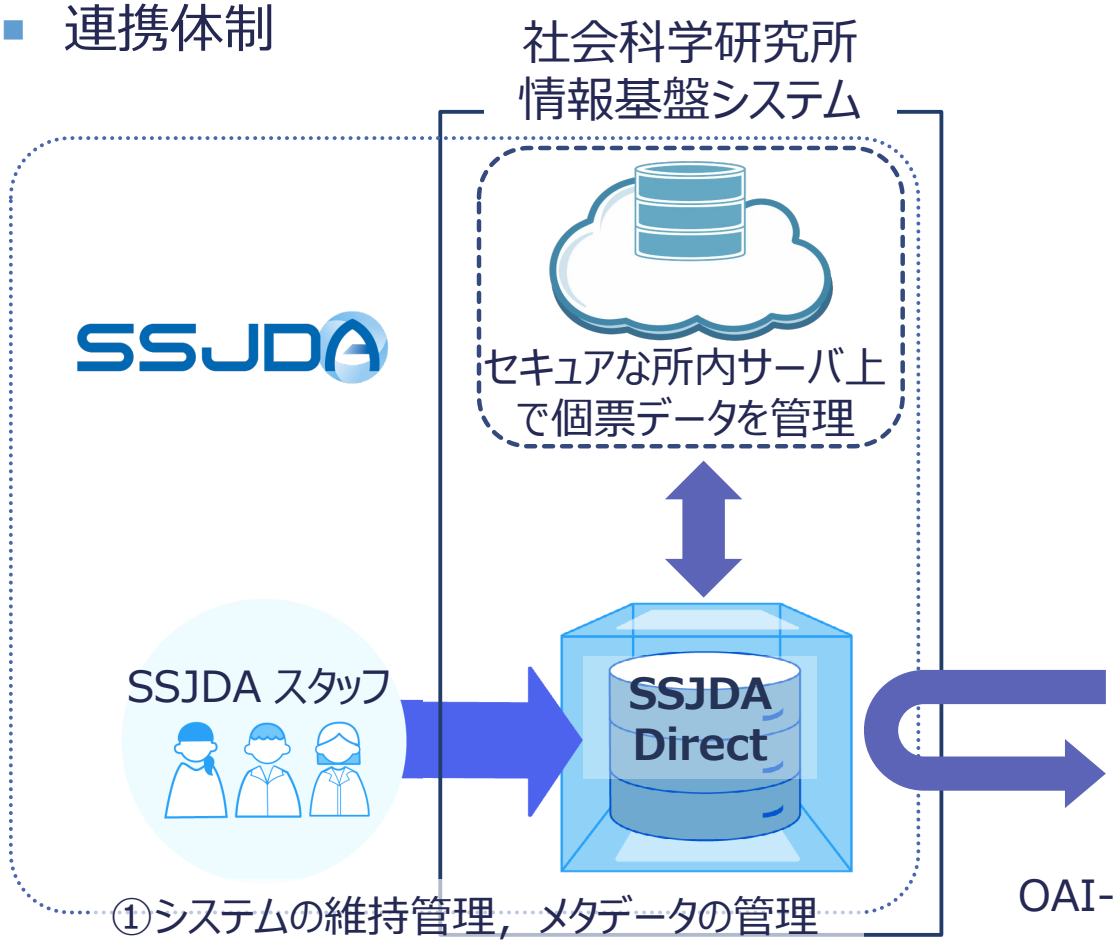
二次分析研究会

計量分析セミナー

データインフラ・ワークショップ

3. JDCat連携の促進

- 連携体制



①システムの維持管理，メタデータの管理
②メタデータの編集，JDCatへの連携作業

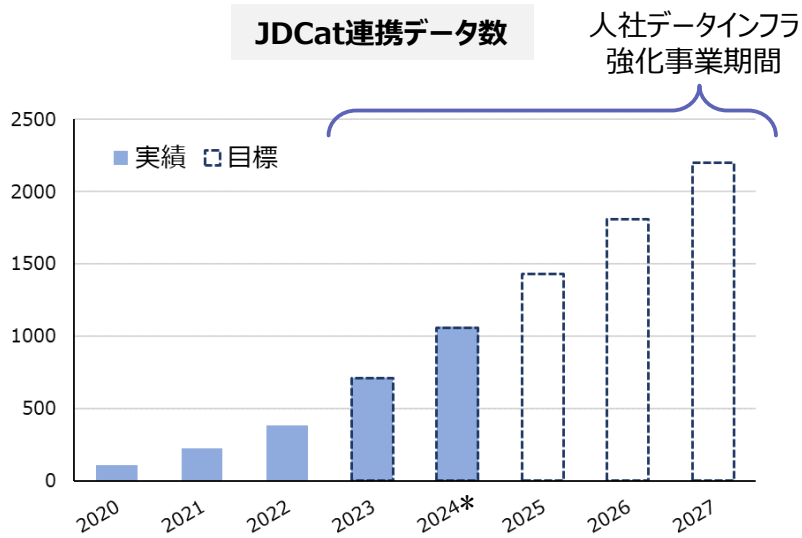
JDCat

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
- 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

JDCat連携データ数



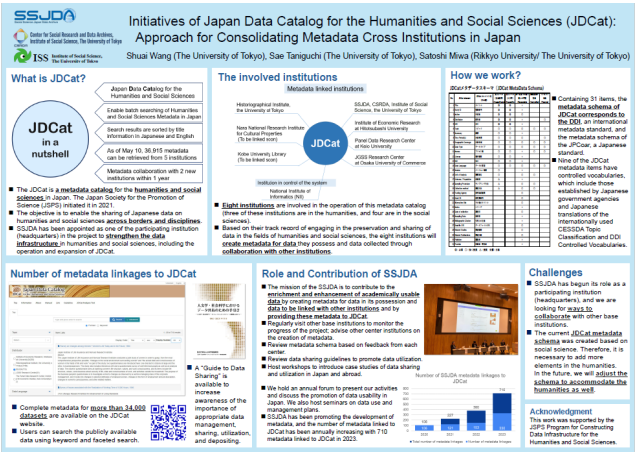
*2025年1月時点

メタデータスキーマ: DDI

4. 学際的・国際的な活動

- 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
  - IASSIST, JOSS, EDDI, WDS, etc.
- 東アジア社会科学データアーカイブネットワーク (NASSDA)
- 海外から専門家を招聘し国際ワークショップを開催

IASSIST (2024)



NASSDA (2025)

国際ワークショップ (2025)

5. メタデータ作成へのAI導入

- 従来のメタデータ作成方法
  - SSJDAのスタッフによる作業
  - 寄託者による作業

➡ **AI活用による省力化・効率化**

メタデータ (SSJDA Direct)

概要	
調査番号	PM050
調査名	東大社研・壮年パネル調査（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・卒論等）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ともない，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／会行報計）	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：1087　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: 多段抽出 層化2段無作為抽出 ・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歳階層）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 調査票による郵送配布，訪問回収法 ・対象者には追跡調査であることを事前に伝えた上で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 実査は，(社)中央調査社
DOI	10.34500/SSJDA.PM050
寄託者(経費)	（「調査の概要」を参照）

JDCat連携メタデータ

Item type	Harvesting DDI(1)
公開日	2024-01-29
タイトル	
タイトル	東大社研・壮年パネル調査（JLPS-M）Wave 1-5, 2007-2011 Japanese Life Course Panel Survey of the Middle-aged (JLPS-M) Wave 1-5, 2007-2011
作成者	東京大学社会科学研究所パネル調査プロジェクト 実査は，(社)中央調査社
配布者	SSJ データアーカイブ
配布者URI	https://csrda.iss.u-tokyo.ac.jp/ssjda/
配布者	SSJDA
配布者URI	https://csrda.iss.u-tokyo.ac.jp/en/ssjda/
編集者	SSJデータアーカイブ SSJDA
所蔵者・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 Japanese Life Course Panel Surveys (JLPS) project, Institute of Social Science, The University of Tokyo
URI	https://ssjda.iss.u-tokyo.ac.jp/Direct/gaiyo.php?eid=PM050
URI	https://ssjda.iss.u-tokyo.ac.jp/Direct/gaiyo.php?lang=eng&eid=PM050
アクセス権	制約付きアクセス restricted access
権利情報	https://csrda.iss.u-tokyo.ac.jp/access/condition/
権利情報	https://csrda.iss.u-tokyo.ac.jp/en/access/condition/
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ともない，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。
対象時期	2007 - 2011
調査日	2011/01 - 2011/03
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
観察単位	個人 Individual
サンプリング方法	確率: 層別抽出, 確率: 多段抽出 Probability: Stratified, Probability: Multistage
調査方法	自記式調査票：紙 Self-administered questionnaire: Paper

5. メタデータ作成へのAI導入

- 今年度の主な成果

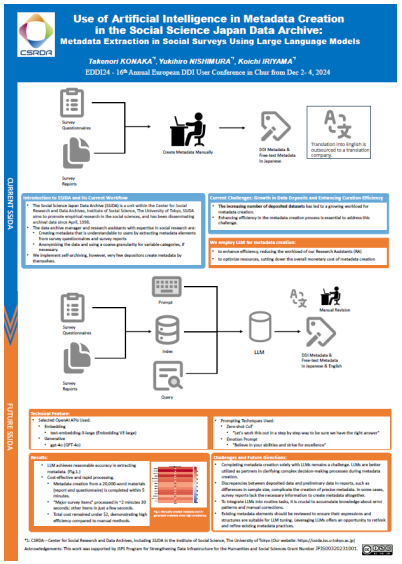
JOSS (2024)

https://joss.rcos.nii.ac.jp/session/overview/

データリポジトリとLLM勉強会 (2024)

https://dias.ex.nii.ac.jp/llm/

EDDI (2024)



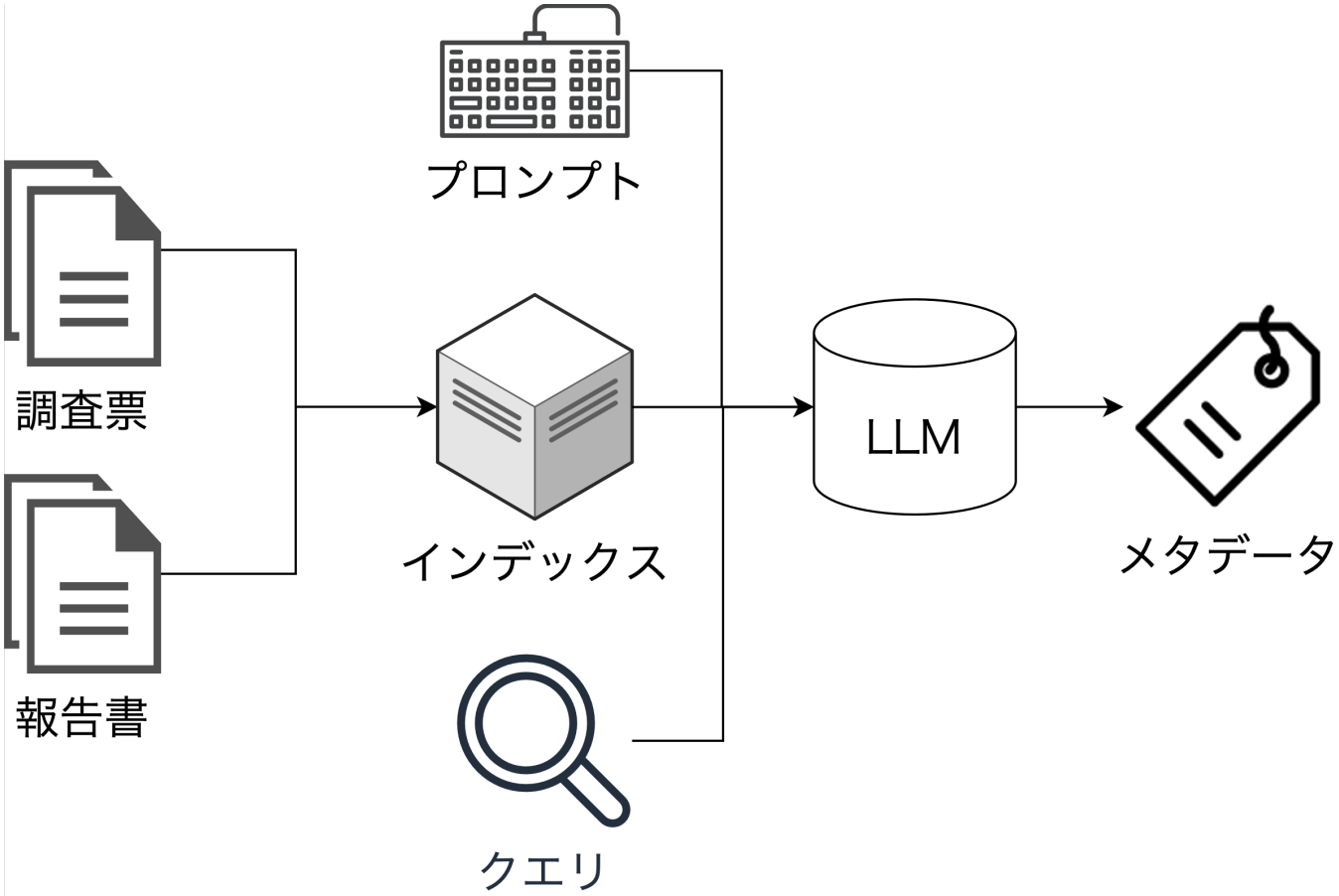
https://doi.org/10.5281/zenodo.14723230

社会科学分野におけるミクロデータ利用の現状と課題 (2025)

https://jmodc.org/event/sche_repo_moc_iss2025.php

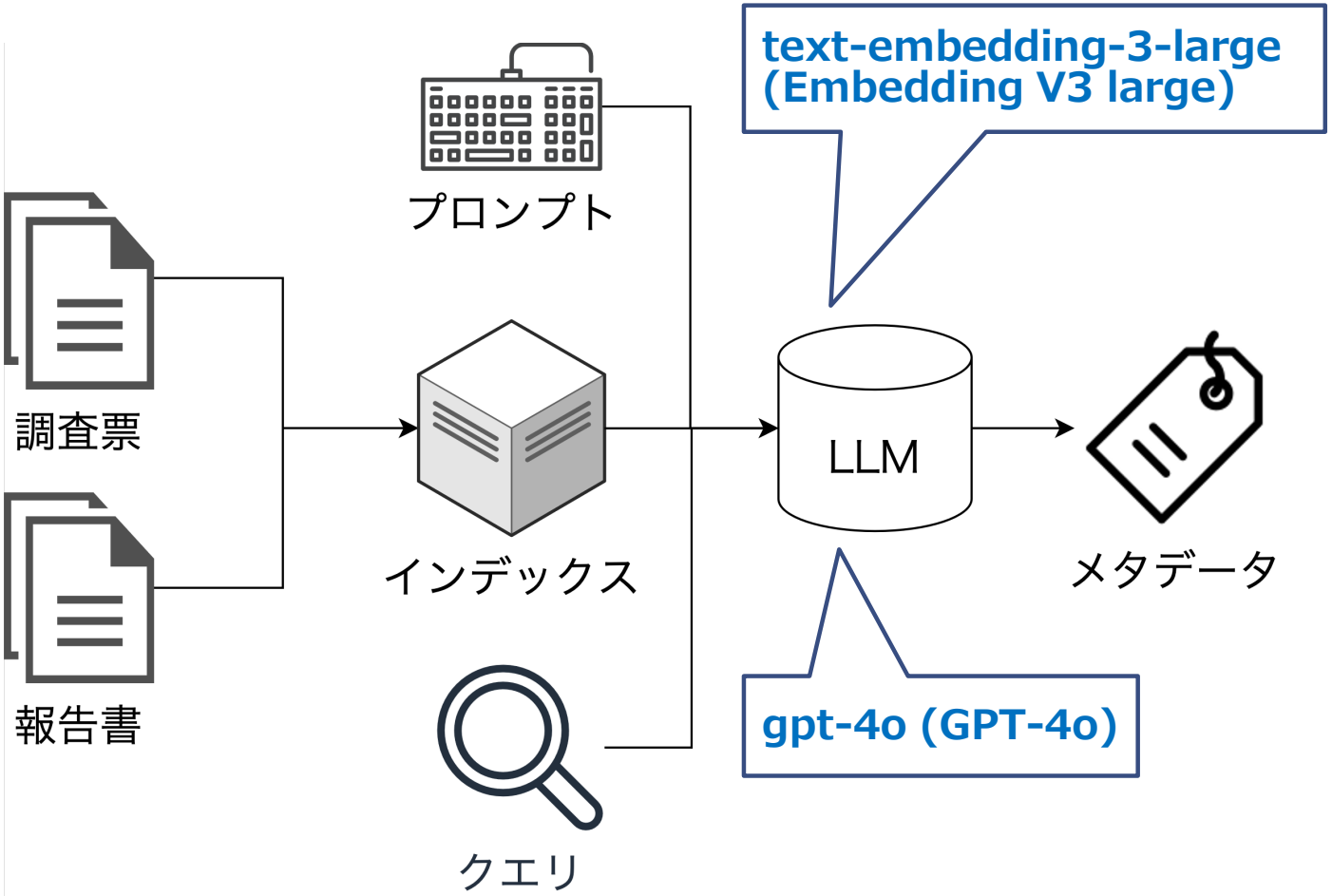
5. メタデータ作成へのAI導入

- 社会調査のメタデータ抽出フロー



5. メタデータ作成へのAI導入

- 社会調査のメタデータ抽出フロー



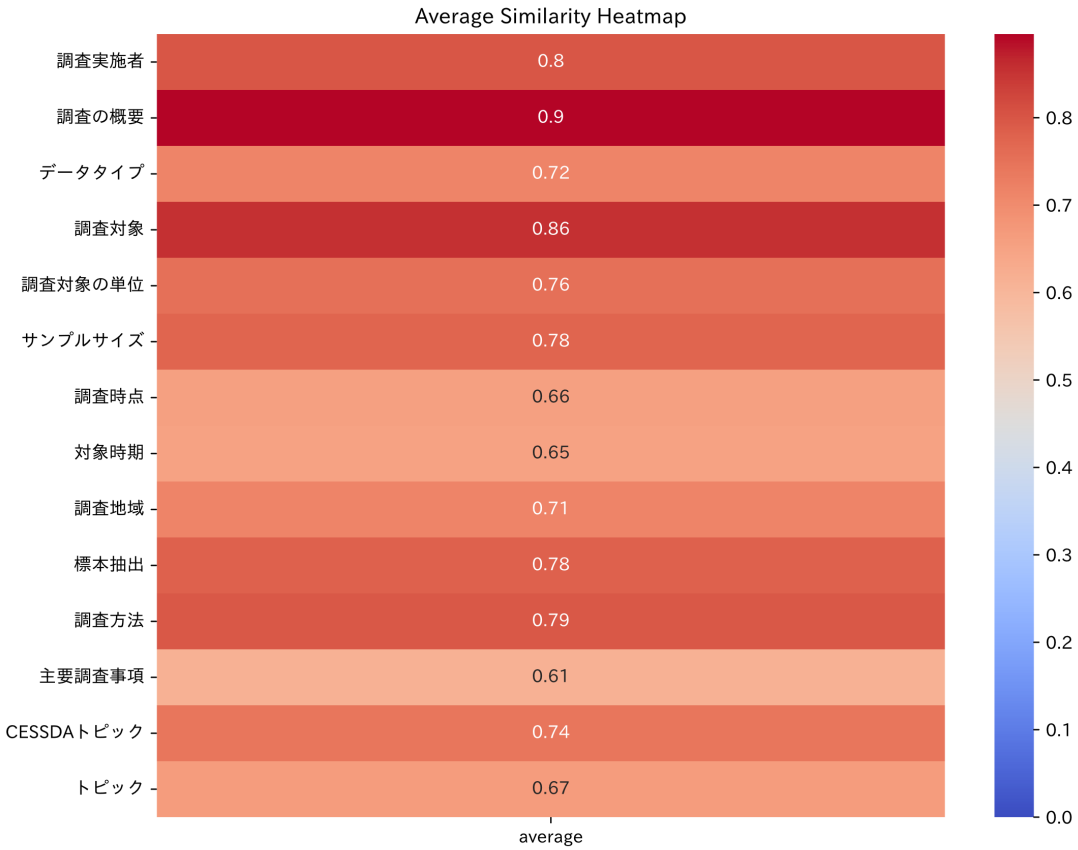
5. メタデータ作成へのAI導入

- メタデータ抽出結果のサンプル

	Webサイト (SSJDA Direct)	AI	秒数
主要調査事項	主要調査事項　　フェイス項目 性別，年齢，居住都道府県，居住地域，学生種別，本人職業，配偶者の有無，子どもの人数，同居家族，住居形態，世帯年収 (1)介護生活 介護の対象者 介護対象者の要介護度 実際の介護者 介護対象者の同居人 主な介護者 要介護になったきっかけ 介護対象者の住居 介護対象者の現在もしくは一番最近の住居 (2)サービスの利用状況と費用 ・利用状況 公的介護保険のサービスとして利用したもの 公的介護保険を要介護度ごとの限度額まで利用している（していた）か 限度額まで利用しない（しなかった）理由 限度額を超えるサービスとして利用したもの 日常的に利用する（した）有料の介護関連サービス等（公的介護保険の対象外） 介護にもっとお金をかけられるなら利用したいサービス ・費用 毎月の介護費用（自己負担分）をまかなう主な資金 ・リフォーム 介護のために実施した住居のリフォームの内容 リフォームにかかった合計費用 どのようなリフォームをしたいか（廊下やトイレなどに手すりを設置など）	1. 個人情報 (Personal Information) - 性別 (Gender) - 年齢 (Age) - 配偶者の有無 (Marital Status) - 職業 (Occupation) - 子供の人数 (Number of Children) - 同居している家族 (Family Members Living Together) - 住まいの種類 (Type of Residence) - 年収 (Annual Income) 2. 介護経験 (Caregiving Experience) - 介護経験の有無 (Experience with Caregiving) - 介護の相手 (Care Recipient) - 介護の期間 (Duration of Caregiving) - 介護の相手の要介護度 (Level of Care Needed by Care Recipient) - 介護の相手の年齢 (Age of Care Recipient) 3. サービスの利用状況と費用 (Service Usage and Costs) - 利用した公的介護保険のサービス (Used Public Long-term Care Insurance Services) - 自己負担額 (Out-of-pocket Expense) - 利用した上乗せサービス (Used Additional Services) - 上乗せサービスの利用料金 (Expense for Additional Services) 4. 介護に関する意識 (Attitudes towards Caregiving) - 介護前の準備 (Preparation before Caregiving) - 介護中の苦労 (Difficulties during Caregiving) - 公的介護保険制度への望み (Hopes for the Public Long-term Care Insurance System) 5. 介護に対する不安 (Concerns about Caregiving) - 身内の介護状態に対する不安 (Concerns about Family Member's Caregiving Situation)	155

5. メタデータ作成へのAI導入

- メタデータ抽出結果のコサイン類似度の平均



まとめ：今年度の主な取り組み状況

- **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
  - 連携データの利用拡大や共同研究の促進
  - 連携に必要となる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. メタデータ作成へのAI導入**
  - 今後の実装に向けた整備や試験的な運用