

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

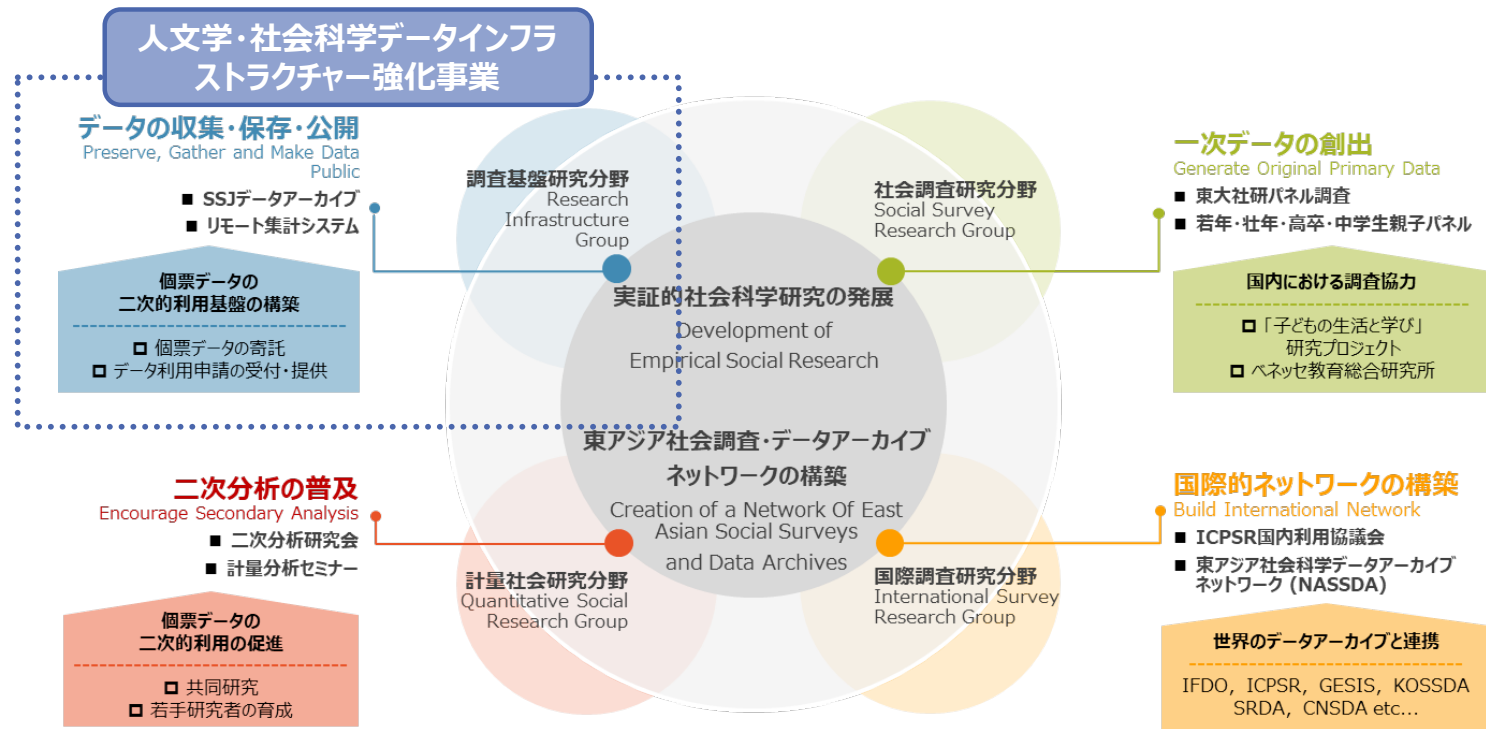
横内陳正

東京大学社会科学研究所

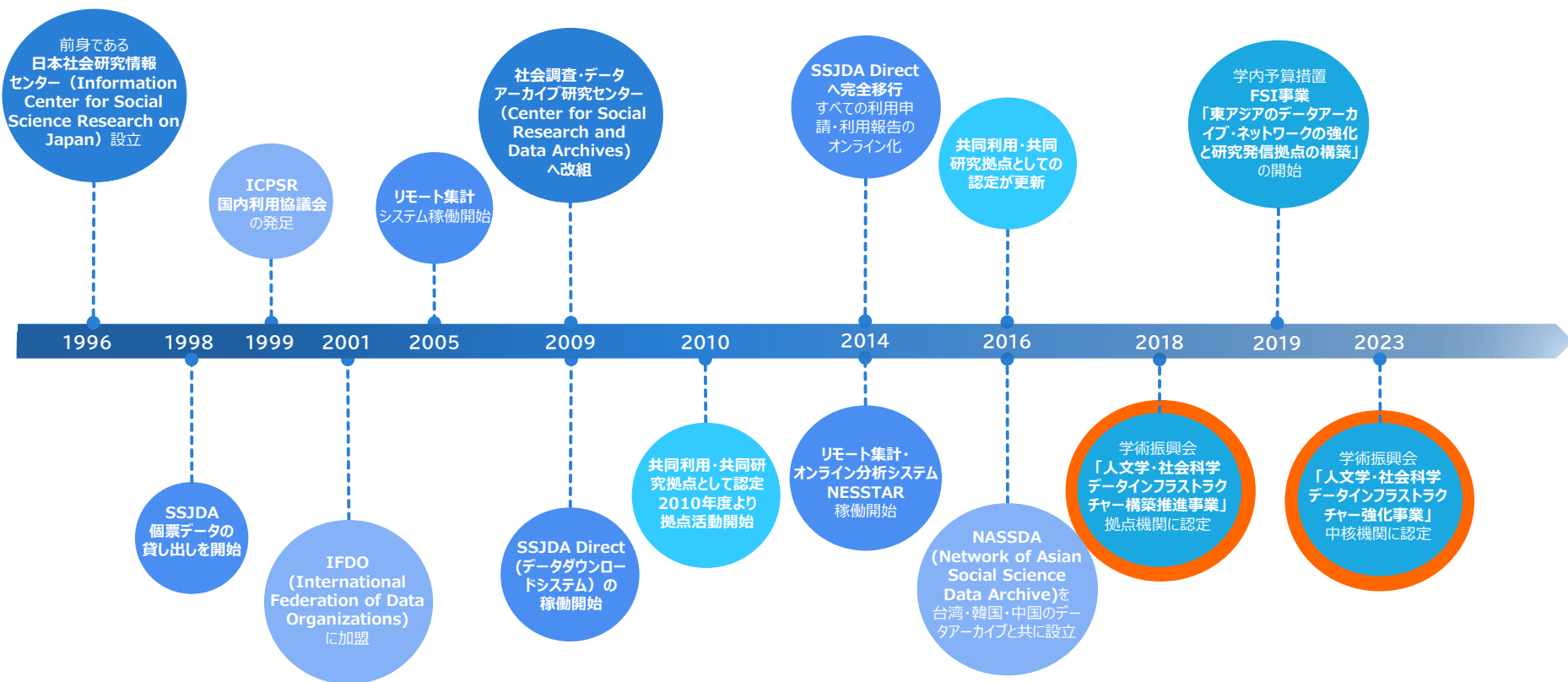
附属社会調査・データアーカイブ研究センター

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E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 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出来る	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出来る	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

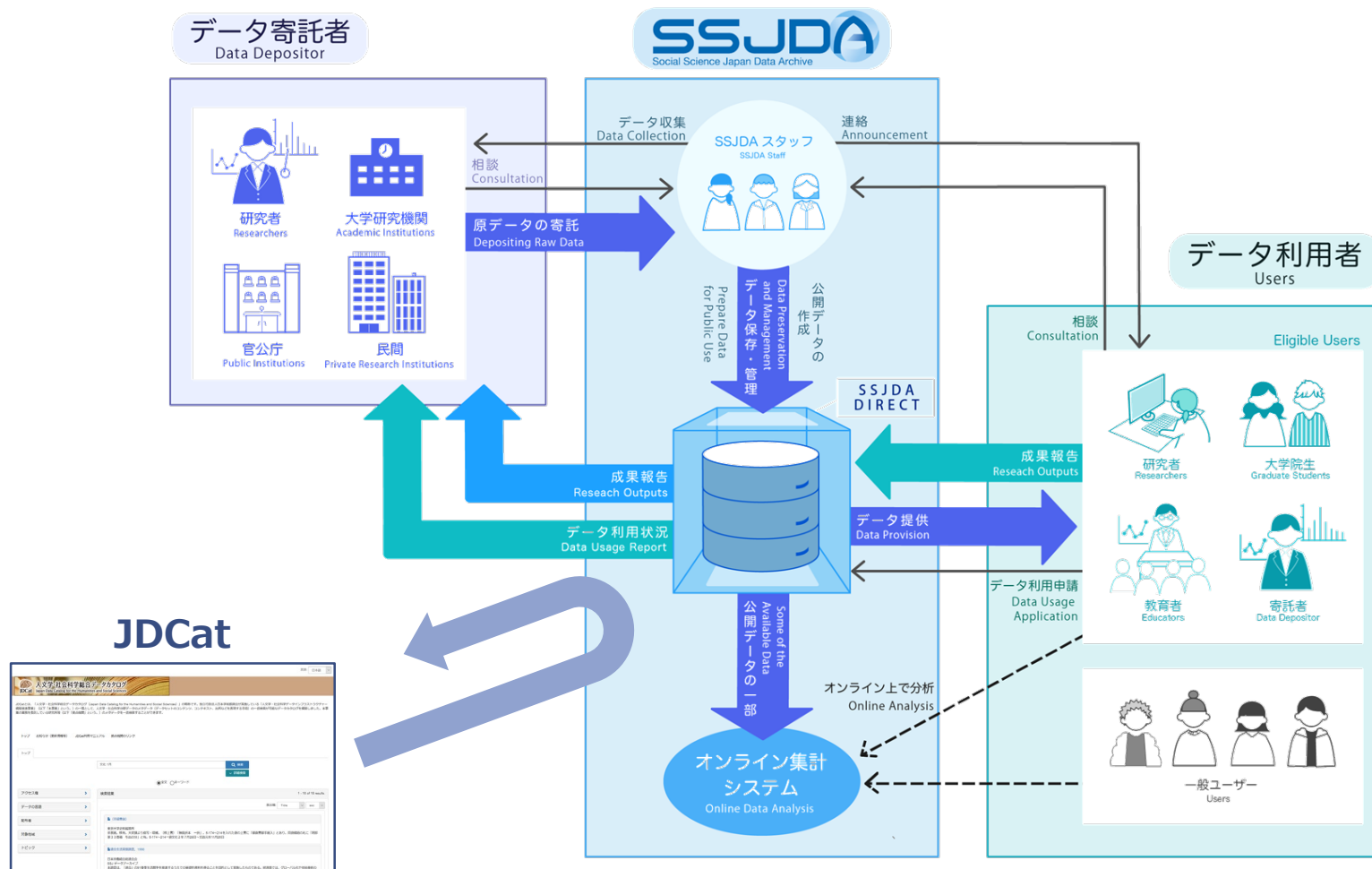
- 意義

- データの散逸を防ぎ、安心して長期保存できる
- データが見つかりやすく、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

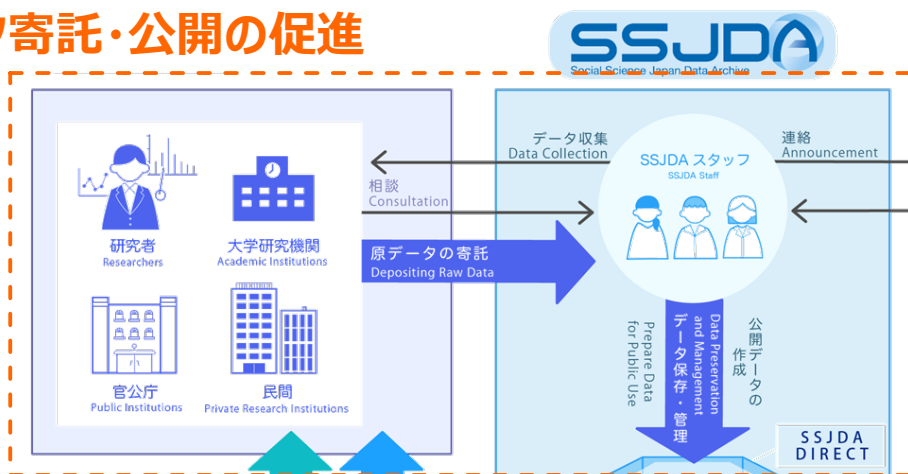
- 幅広い機関・研究者の方々から調査データが寄託されている
- 公開データを使った研究や教育を支援している
- データアーカイブ間の学際的、国際的な連携を行っている
- 寄託者や利用者への親身なサポートを行っている

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

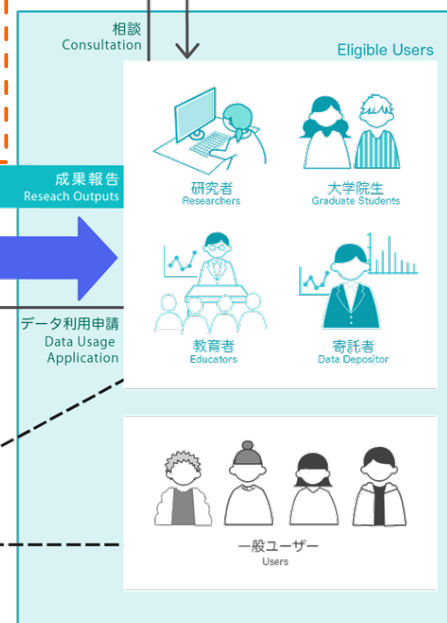
1. データ寄託・公開の促進



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



オンライン集計
システム
Online Data Analysis

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）

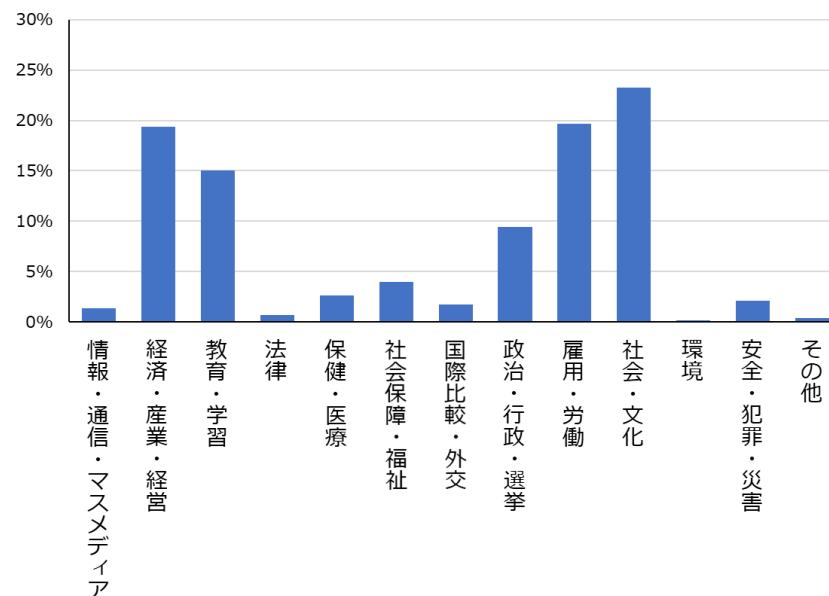
学校教育に対する保護者の意識調査（ベネッセ教育総合研究所）

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）

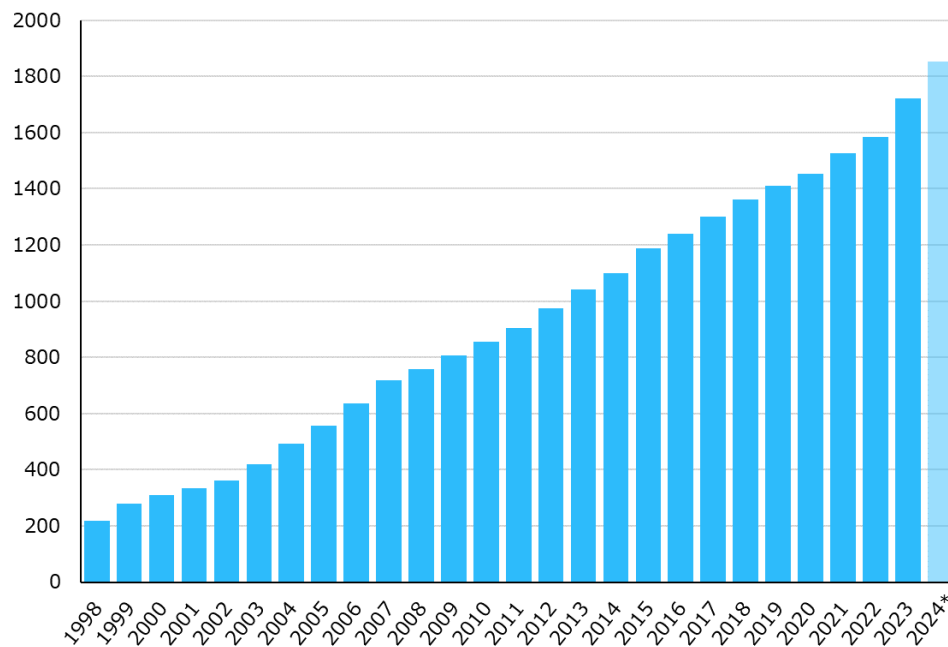
1. データ寄託・公開の促進

- 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

2024年度
計量分析セミナー・春

2025年2月17日～21日
(2月18日) オンライン開催
10:30～17:00

17 Mon. 機械学習を用いた格差と因果効果の推定:
Balancing Weightsを軸とした手法整理
講師:川田 恵介 (東京大学)

19 Wed. Stataによるパネル調査データ分析の実践
講師:渡山 亮太 (学習院大学)

20-21 Thu. Fri. ベイズ統計モデリング:
マルチレベル分析を中心に
講師:清水 裕士 (関西学院大学)

申込締切:2025年1月15日(火)
定員(各区分)になり次第 締め切ります

詳細・申込はウェブサイトをご確認ください

SSJDA
Social Science Japan Data Archives

データインフラ・ワークショップ

社会科学データインフラストラクチャー
ワークショップ

社会科学分野におけるDMP
:他分野との比較からみえる課題と展望

2024年10月1日(火) 13:00-14:00
オンライン開催 (Zoom)

オープンサイエンスの考え方が広まりつつあるなか、昨今では様々な研究助成においてデータマネジメントプラン (DMP) の作成が必要とされるなど、研究者の意識だけでなく実践においても変化が求められています。

そこで本セミナーでは、まず、社会科学分野とその比較対象として人文科学分野に賛同し、それぞれの分野におけるDMPの現状について概観します。つぎに、両分野の比較を通して、社会科学分野におけるDMPを巡る課題と展望について、参加者とともに考えます。

講師:石田賢示 (東京大学社会科学研究所・准教授)
北岡タマ子 (人間文化研究機構・特任准教授)

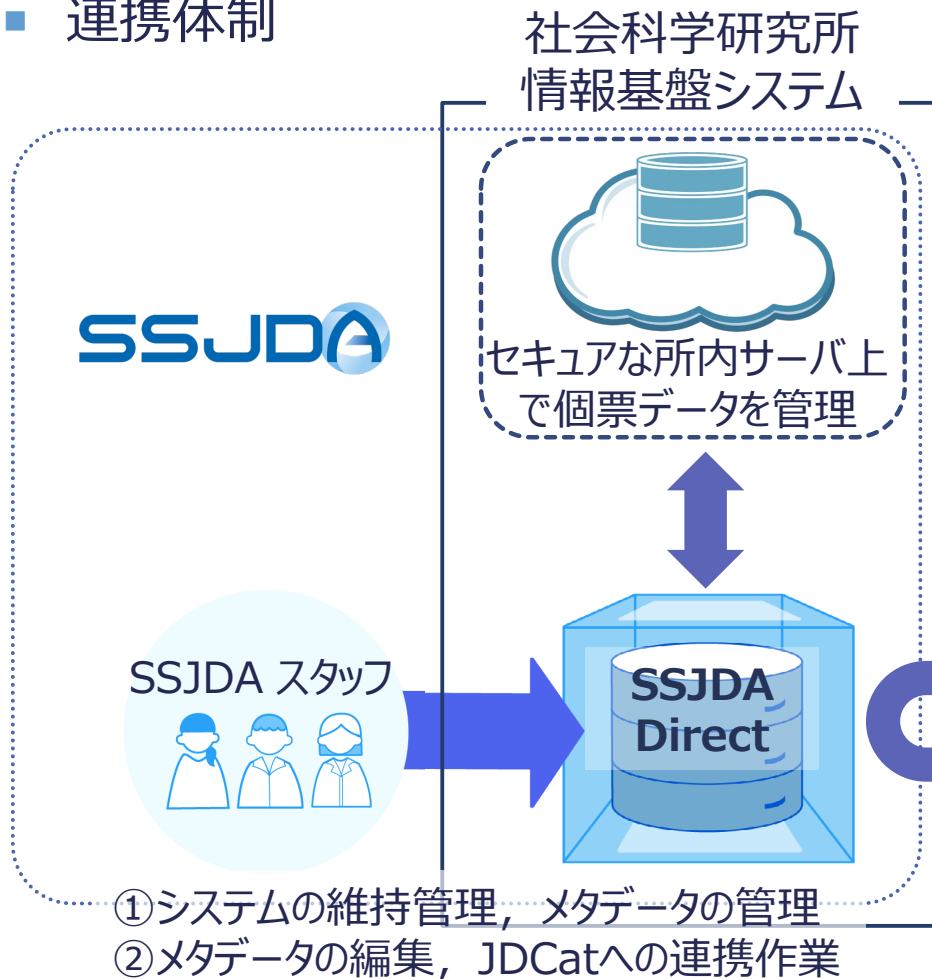
申込登録はこちら

本イベントは、2024年人文・社会科学データインフラストラクチャー推進事業
(共同実施:SSJDAと東京大学社会科学研究所)の一環として開催されます。

主催: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.u-tokyo.ac.jp/>

3. JDCat連携の促進

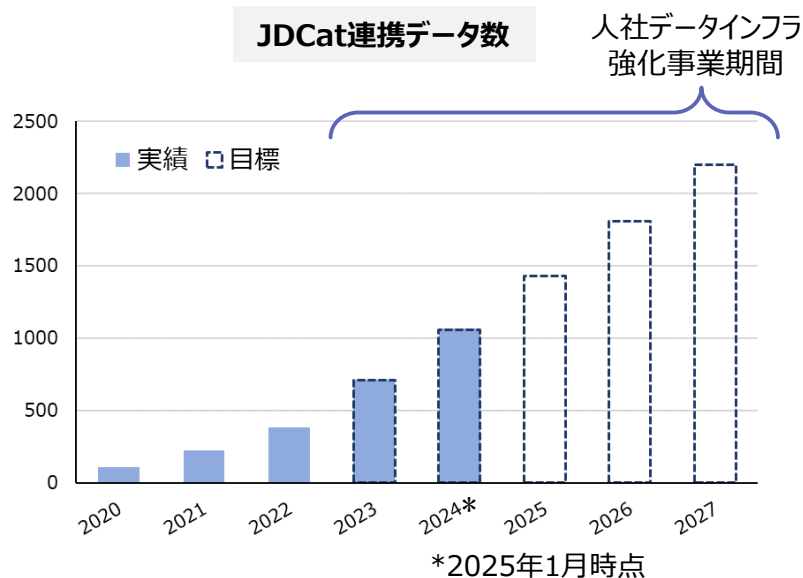
■ 連携体制



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
- 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



メタデータスキーマ: DDI

Center for Social Research and Data Archives,
Institute of Social Science, The University of Tokyo


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

- 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


NASSDA (2025)



SSUDA
Society for Social-Usability Data Analysis
The University of Tokyo, The University of Tsukuba
ISS
Institute for Information Science and Society
University of Tsukuba

Initiatives of Japan Data Catalog for the Humanities and Social Sciences (JDCat): Approach for Consolidating Metadata Cross Institutions in Japan

Shuai Wang (The University of Tokyo), Sae Taniguchi (The University of Tokyo), Satoshi Maeda (Rikkyo University) The University of Tokyo



JDCat
Japan Data Catalog for the Humanities and Social Sciences

What is JDCat?

Japan Data Catalog for the Humanities and Social Sciences

Crossed meta-repository for the Humanities and Social Sciences Metadata in Japan

Search results are sorted by the information in Japanese metadata

As of May 10, 36,915 metadata can be retrieved from 11 institutions

Metadata consolidated with 1-year institutions in Japan

- The JDCat is a metadata platform for the humanities and social sciences in Japan. The Japan Society for the Promotion of Crossed Meta-Repositories in 2021
- The objective is to enable the sharing of Japanese data on social issues and humanities between libraries and universities.
- SSUDA has been appointed as one of the participating institution data in the future. We will strive to strengthen the data infrastructure in humanities and social sciences, including the operation and expansion of JDCat.

The involved institutions

Metadata linked institutions

Interdisciplinary Institute,
The University of Tokyo

Library of Science
The University of Tokyo

Yokohama City University
Library

Rikkyo University
(To be linked)

SSUDA, Catalog of Social Science
Metadata

Library of Economics Research
at Hitotsubashi University

University of
Tokyo Library


University of
Osaka University of Commerce

Institution in control of the system

- **Explicit institutions** are those in the operation of the metadata catalog (three of these institutions are in the humanities, and two are in the social sciences)
- **Based on the track record of engagement in the preservation and sharing of data in the future**, we will strive to strengthen the data infrastructure across the entire field of data processing and data collected through collaboration with other institutions


How we work?

JDCat → JDCat → JDCat (Metadata)



- Containing 31 items, the metadata scheme of JDCat corresponds to the ISS, an international metadata standard, and metadata scheme of the SSUDA, a Japanese standard
- **None of the JDCat metadata control vocabularies**, which include those from the Japanese government agency, translators of the CEESDA Tools
- **Controlled Vocabularies**

Number of metadata linkages to JDCat



Number of metadata linkages to JDCat

■ Complete metadata for **more than 34,000** datasets are available on the JDCat

Users can search the publicly available metadata and download the data

Role and Contribution of SSJDA

■ **SSJDA is a public research center for the environment and enhancement of academically usable data** that will contribute to the development and use of data to be linked with other institutions and by providing these metadata to JDCat

- Regulatory role and state institutions to monitor the project, adjust other institutions on the creation of metadata
- **Provide metadata schemes based on feedback from each institution**
- **Review data sharing guidelines to promote data utilization**
- **Provide metadata schemes and data sharing and utilization in Japan and abroad.**

■ The lead in our current plan to present our activities and discuss the position of data usability in Japan, we also host seminars on data and management plans

■ SSJDA has been promoting the development of metadata, and the number of datasets with 710 JDCat linkages increased in 2023


Challenges

■ SSJDA has begun its role as a participating institution (development), and we are looking for **partners** to collaborate with other base institutions

■ The current JDCat metadata schemes were created based on social science. It is necessary to add more information for the humanities. In the future, we will adjust the schemes to accommodate the humanities as well

Acknowledgement

This work was supported by the JSPS Program for Crossed Data Infrastructure for the Humanities and Social Sciences





国際ワークショップ^o (2025)

NASDAQ ASO / NASDA (Network of Asian Social Science Data Archives) Meeting
CERDA国際フォーラムワークショップ / CERDA International Workshop

CERDAはデータアーカイブス・ネットワーク強化プロジェクト / Strengthening Data Infrastructure for Social Sciences

 Network of Asian Social Science Data Archives

社会科学データアーカイブ連携の国際動向 International Trends in Social Science Data Archive Collaboration

Register here until Jan. 7, 2025

2025年1月9日(木)
9 Jan. 2025 9:30~11:50 (JST)

Hybrid／ハイブリッド開催

For more details
https://ccvwa.hse.ie/topics/anInfoPanel/public_investures/

社会科学データの取扱いが難しくなって、多くの国内、地産、民間等のグループが組まれている。データタイプ別の協力が必要である一方で、その成功や否かは一時的であり、必ずしも一定ではあるとせん。

基調講演者は、IASIST（International Society for Social Science Information Services and Technology）の副会長であるBrobay Joseph Berdenfeldtを用いて、地域および世界における社会科学データアーカイブの現状に、東アジア圏での研究上の課題について概観しに紐結るものとす。

予言書的な含意のある建設的立場にて紹介いたします。

ワークショップの後半では、東アジアデータアーカイフオーストワーク(NASDAQ)、メンバー機関間の活動への参加が報告されます。

As the role of social science data archives has grown over many national, regional, and international groups have formed. While collaboration among data archives is increasingly important, the approaches and contents of these cooperative ways are not always uniform.

For the keynote speech, Mr. Brobay Joseph Berdenfeldt, Vice President of ASIISAT (International Association for Statistical Information Systems and Technology), will provide an overview of the current status of social science data archiving worldwide. He will also address potential opportunities for NASDAQ and its members to form alliances with other data organizations that could be mutually beneficial. In the second part of the workshop, the East Asia Data Archive Network (NASDAQ) and its collaborative activities will be introduced, followed by reports on the initiatives and efforts of its member institutions.

主催： 東京大学社会科学研究所所属会合企画、データタイプ研究会センター

幹事：日本学術会議事務局、NAC、日本人文学者・研究者データアーカイブ強化推進委員会、CCVWA/JASSOCIATION

Funded by Nippon Foundation & University of Tokyo, Japan Academy, Japan Society for the Promotion of Sciences (JSPS) Grant-in-Aid for Scientific Research (KAKENHI).

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/ナール調査 (JLPS-M) Wave 1-5, 2007-2011
寄託者	東京大学社会科学研究所/ナール調査プロジェクト
利用申込先・承認手続	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
メタデータ提供形式・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極めて重要である。
データタイプ (量的調査/質的調査/混合/先行研究)	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 層化は総無作為抽出 ・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層等) に層化 ・抽出方法は、住民基本台帳を基本とし、住民基本台帳の閲覧が不可能になった地点では、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、訪問調査 ・訪問調査は追加調査であることを事前に伝え、同意の上で調査票を配布し、郵送で調査票を配布。その後に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/ナール調査プロジェクト 実質は、(社)中央調査社
DOI	10.34500/SSJDA.PM050
寄託者(館蔵)	(「調査の概要」を参照)

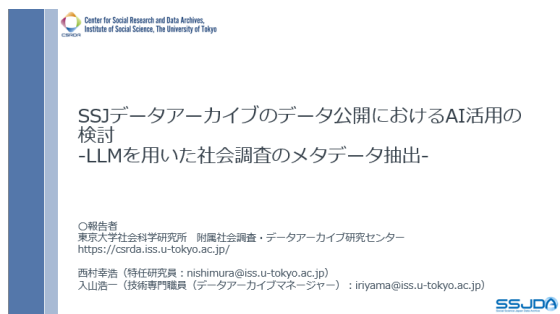
JDCat連携メタデータ

Item type	Harvesting DOI(1)
公開日	2024-01-29
タイトル	東大社研・壮年/ナール調査 (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/ナール調査プロジェクト 実質は、(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/ナー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極めて重要である。
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

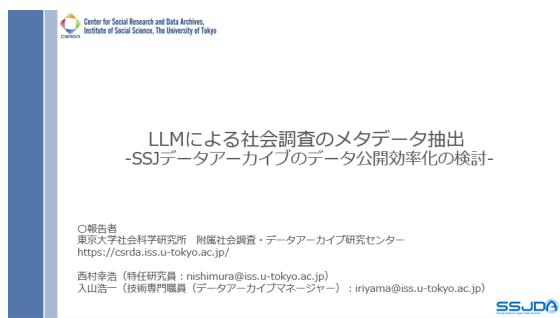
■ 今年度の主な成果

JOSS (2024)



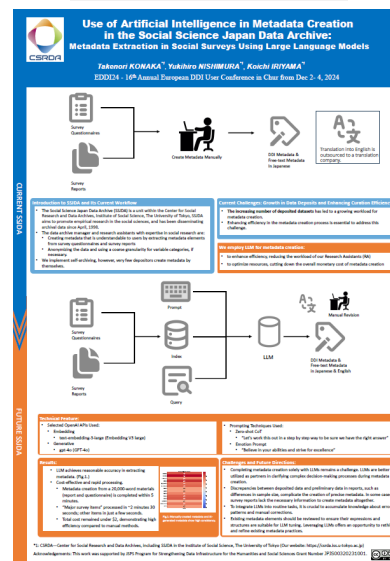
<https://joss.rcos.nii.ac.jp/session/overview/>

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>

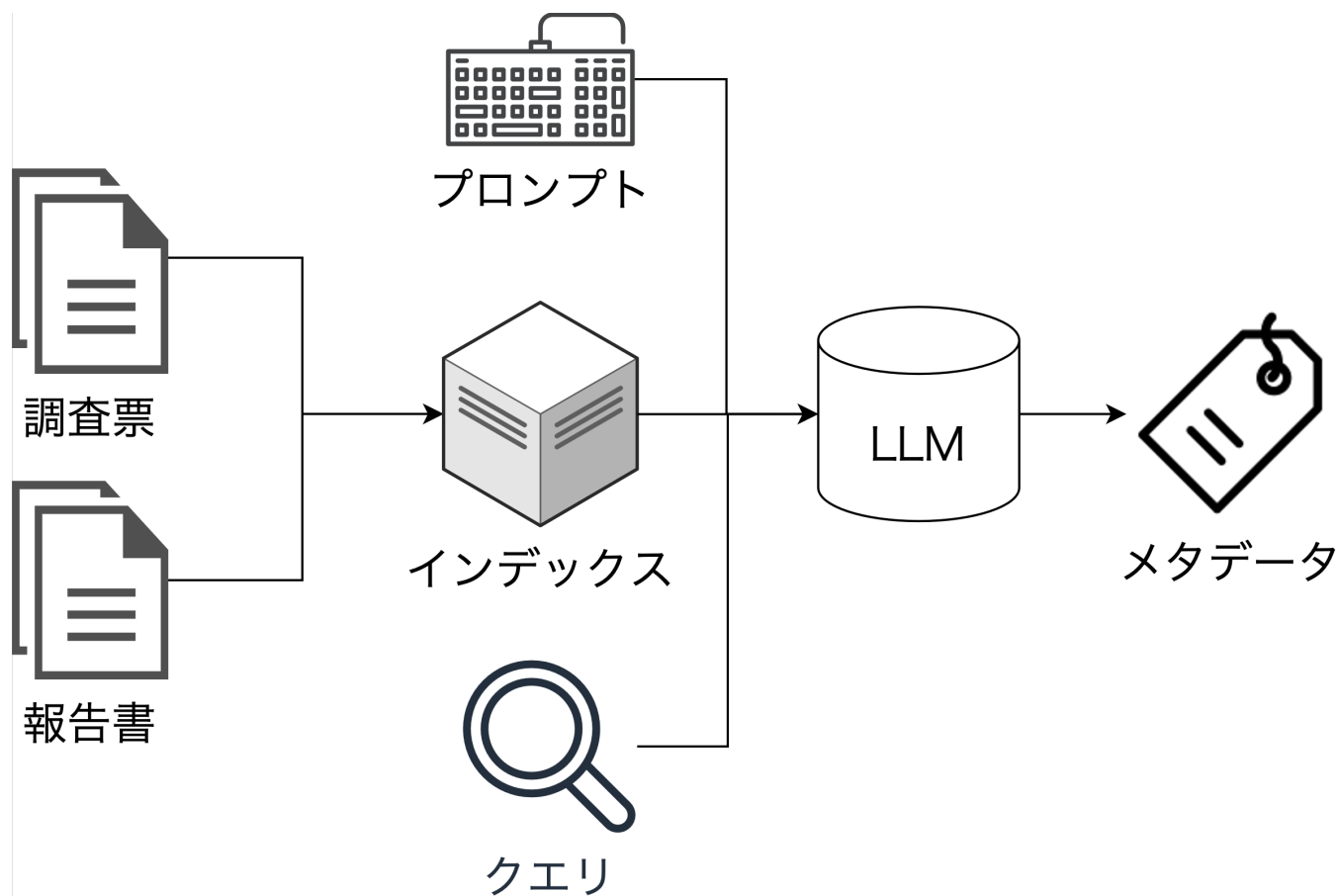
社会科学分野におけるマイクロデータ利用の現状と課題 (2025)



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

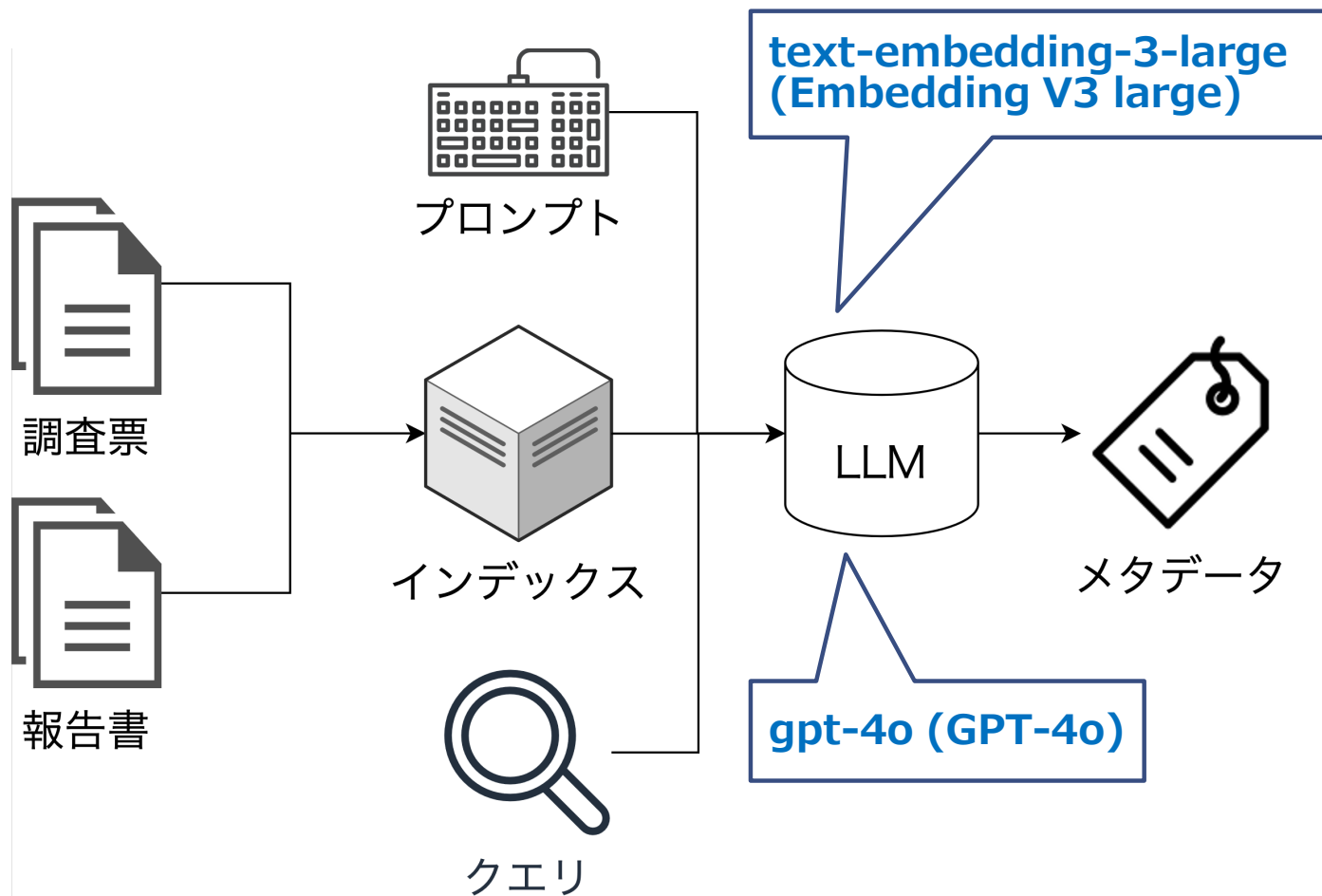
5. メタデータ作成へのAI導入

- 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)

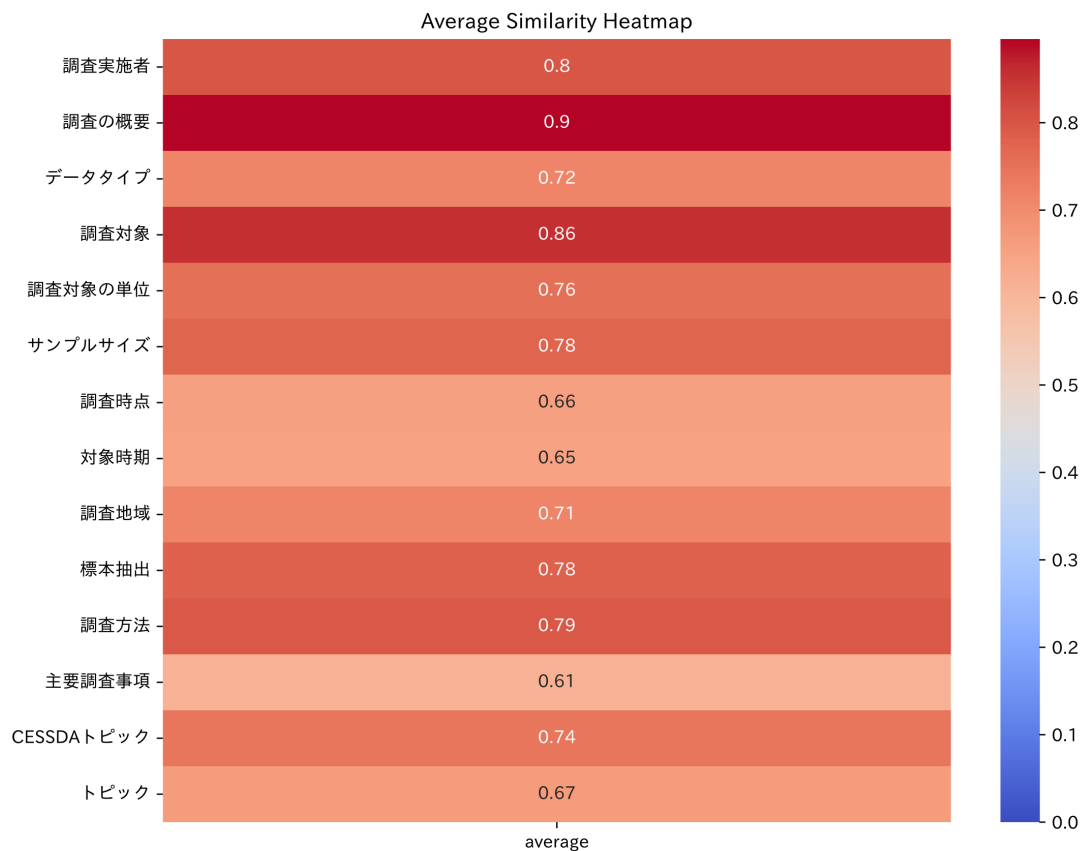
AI

秒数

主要調査事項	<p>主要調査事項 フェイス項目</p> <p>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</p> <p>(1)介護生活</p> <p>介護の対象者</p> <p>介護対象者の要介護度</p> <p>実際の介護者</p> <p>介護対象者の同居人</p> <p>主な介護者</p> <p>要介護になったきっかけ</p> <p>介護対象者の住居</p> <p>介護対象者の現在もしくは一番最近の住居</p> <p>(2)サービスの利用状況と費用</p> <p>・利用状況</p> <p>公的介護保険のサービスとして利用したもの</p> <p>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</p> <p>限度額まで利用しない（しなかった）理由</p> <p>限度額を超えるサービスとして利用したもの</p> <p>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）</p> <p>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</p> <p>・費用</p> <p>毎月の介護費用（自己負担分）をまかなう主な資金</p> <p>・リフォーム</p> <p>介護のために実施した住居のリフォームの内容</p> <p>リフォームにかかった合計費用</p> <p>どのようなリフォームをしたか（廊下やトイレなどに手すりを設置など）</p>	<p>1. 個人情報 (Personal Information)</p> <ul style="list-style-type: none"> - 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) <p>2. 介護経験 (Caregiving Experience)</p> <ul style="list-style-type: none"> - 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) <p>3. サービスの利用状況と費用 (Service Usage and Costs)</p> <ul style="list-style-type: none"> - 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) <p>4. 介護に関する意識 (Attitudes towards Caregiving)</p> <ul style="list-style-type: none"> - 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) <p>5. 介護に対する不安 (Concerns about Caregiving)</p> <ul style="list-style-type: none"> - 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. 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