研究データ利活用協議会　公開シンポジウム

「人文学・社会科学のデータ共有における課題検討」小委員会　活動紹介

谷口 沙恵（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）
石井 加代子（慶應義塾大学パネルデータ設計・解析センター）
岡田 一祐（北海学園大学人文学部）
堤 智昭（筑波大学人文社会系）
平澤 加奈子（東京大学史料編纂所）
渡邊 要一郎（東京大学史料編纂所）

小委員会の設立目的

人文学・社会科学のデータ共有における課題を解決するために国内のデータアーカイブ等の事例を共有し、提言を行うことを目的とする。

初期委員

石井 加代子（慶應義塾大学パネルデータ設計・解析センター）

岡田 一祐（北海学園大学人文学部）

谷口 沙恵（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）

堤 智昭（筑波大学人文社会系）

平澤 加奈子（東京大学史料編纂所）

渡邊 要一郎（東京大学史料編纂所）

成果物

- 各委員が所属、あるいはこれまで関わってきたデータアーカイブ、データベース等の事例集

  事例1. 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

  事例2. 慶應義塾大学パネルデータ設計・解析センター

  事例3. SAT大正新脩大藏經テキストデータベース

  事例4. 国立国語研究所

  事例5. 新日本古典籍総合データベース

  事例6. 東京大学史料編纂所

- データ共有の際の指針や教材の作成。

成果物から得られるインパクトや効果

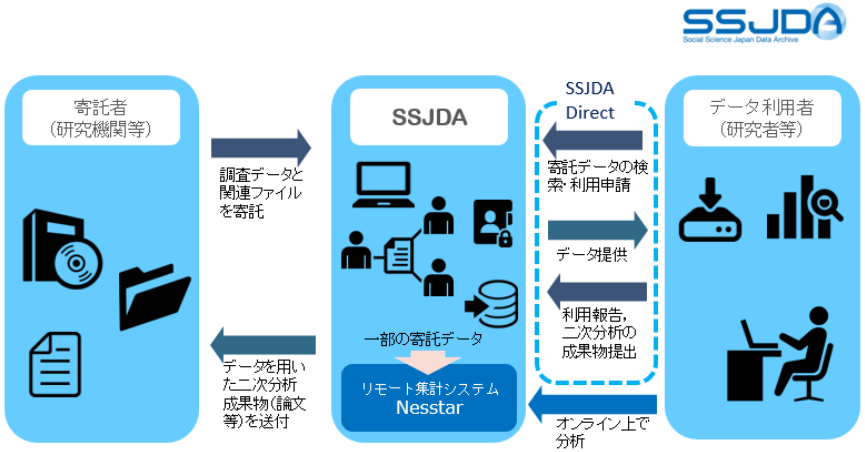
- 「研究データ」の定義が大きく異なる人文系と社会科学系の、研究データに関する用語や概念の整理を示すことができる。
- 複数の機関におけるデータ共有・公開の問題点、困難を共有・整理し、問題解決へつなげる知見を提示できる。
- 成果物を公開することによって、これから新たに共有・公開を開始する機関や研究者、支援人材の参考となる。

事例１：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 Social Science Japan(SSJ)データアーカイブ

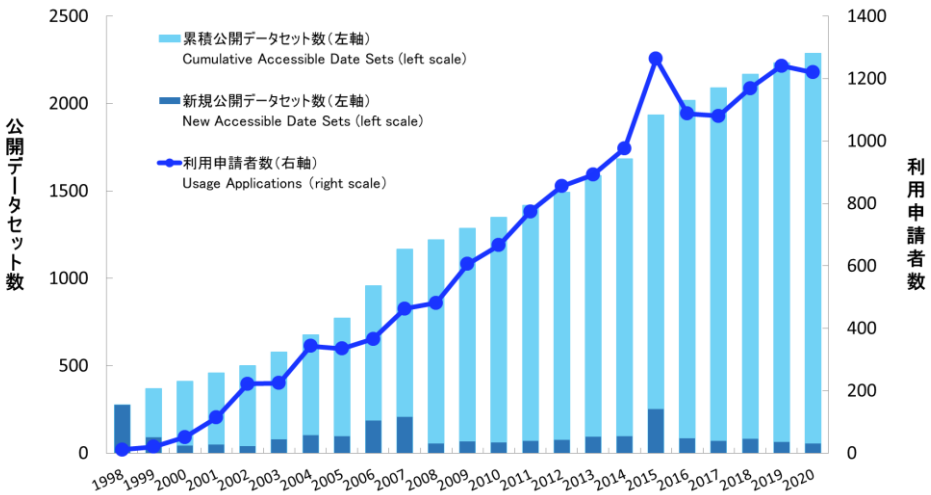
社会科学の実証研究を支援することを目的として、学術利用（一部データは教育利用も可）のための個票データの提供を行っている。

取扱う研究データ：
　社会調査の個票データ（量的データ）
利用条件：あり、登録制
利用料：無料
利用方法：
ウェブサイトからオンラインシステムを用いてデータ検索・利用申請が可能（ただし利用申請は利用登録者のみ可）

SSJDA データの寄託と提供のしくみ



SSJデータアーカイブのデータ公開・利用実績（1998-2020）



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

事例2：慶應義塾大学パネルデータ設計・解析センター

パネルデータとは？

同一主体を長期に渡り追跡調査することで得られたデータ。
制度変更の効果の測定や、個人や企業の行動変化、因果関係の把握が可能。

- 組織体制
  - わが国におけるパネルデータに関する研究分野の中心的な役割を担っていくことを目的に設立された研究教育組織。
  - 文部科学省の共同利用・共同研究拠点の認定。
- 使命
  - 「日本家計パネル調査(JHPS)」などパネル調査の継続的実施
    - 「消費生活に関するパネル調査（JPSC）」「日本子どもパネル調査（JCPS）」 etc。
    - パネルデータを活用した研究—多岐にわたる質問項目を活かして、所得格差、資産格差、教育、ワークライフバランス、健康など、幅広いテーマの研究を実施。
  - **パネルデータの公開・共有**
    - **パネルデータに関するデータアーカイブの中核的な機能を担う機関を目指して、当センターが実施するパネル調査および、他機関から寄託されたパネルデータを国内外の研究者へ公開**

事例3: SAT大正新脩大藏經テキストデータベース

大正新脩大藏經 テキストデータベース
संगणिकीकृतं तैशोत्रिपिटकं
Saṃgaṇikīkṛtaṃ Taiśotripiṭakaṃ
The SAT Daizōkyō Text Database

・仏典2920件、漢字にして一億字超の『大正新修大蔵経』の漢文テキストデータベースを電子検索可能な形でオンラインで公開

・電子テキストと元の大正蔵刊本テキスト画像の対照する機能を IIIF (International Image Interoperability Framework) によって提供

・電子テキストデータを人文資料電子データの国際標準である TEI (Text Encoding Initiative) 化するプロジェクトを進行中

・対象：『大正新修大蔵経』
・利用条件：個人、団体による学術目的、あるいはそれ以外の非営利目的
・利用方法：https://21dzk.l.u-tokyo.ac.jp/SAT/ からデータ検索可能

事例4：国立国語研究所の研究データ公開事例

- ✦ 現代日本語書き言葉均衡コーパス（BCCWJ）を始めとした言語コーパスの公開
- ✦ 研究資料として利用できる所蔵物の公開
  - ▪ データを作ることが研究成果
  - ▪ 利用されることで成果の価値も高まる
  - ➢ データ作成の業績価値の担保
  - ▪ BCCWJでは利用した研究業績を収集、公開
    - ✦ 窓口で確認できたもののみー抜けが存在
    - ➢ 公開されたデータを利用したときの対応方法の一般化
    - ➢ 活用事例収集の指針や標準の作成の重要性

これら課題の本協議会での検討

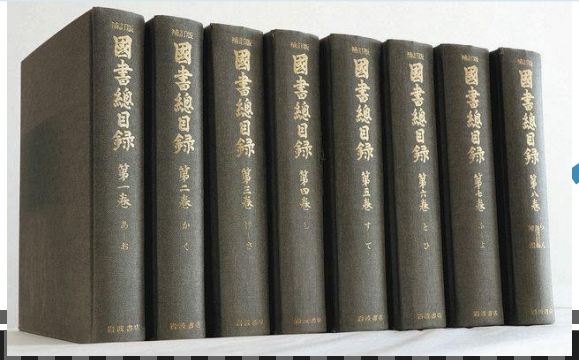
https://ccd.ninjal.ac.jp/bccwj/list.html

事例5：新日本古典籍総合データベース（日本語の歴史的典籍の国際共同研究ネットワーク構築計画）

新日本古典籍総合データベース

- 日本語の歴史的典籍（古典籍）資料を研究データに
- 画像デジタルウェブアーカイブのデファクトスタンダードであるIIIF形式で画像を公開
- Japan Link Center正会員としてDOIを各古典籍に付与
- 所蔵者の許諾によりCCライセンスやPD資料として提供

『国書総目録』以来蓄積された書誌データ（『日本古典籍総合目録データベース』）



十数万点の資料を撮影しIIIF形式で公開
©東京書籍株式会社付設教科書図書館東書文庫
DOI: 10.20730/100266840

事例6：東京大学史料編纂所のデータベース

東京大学史料編纂所
HISTORIOGRAPHICAL INSTITUTE, The University of Tokyo

- **史料編纂所におけるデータベース数** 40（うち、所外公開31）
  https://wwwap.hi.u-tokyo.ac.jp/ships/shipscontroller
- **種類**　目録, 本文, 画像・図像・写真, 知識　（人名・地名・和暦・文字など）
- **件数**　画像：2000万件　データ：約700万件

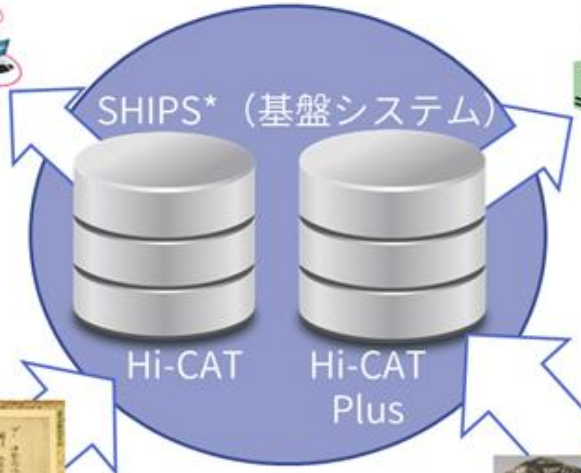
Hi-CAT（所蔵史料目録DB）
(2015.12～)
画像ビューア
OpenSeaDragon
サーバー
IIIF, Legacy Image Pyramids

Hi-CAT Plus
（2020年3月～一部ウェブ公開）
画像ビューア
Mirador
サーバー
IIIF
Presentation API

所蔵史料と所外の史料

ウェブ公開

閲覧室での公開
＋連携機関で閲覧
＋一部をウェブ公開



史料編纂所所蔵史料の画像

国内・海外から収集した史料の画像

* Shiryohensanjo Historical Information Processing System

本所所蔵史料と他機関所蔵史料について基盤システム（SHIPS）上でメタデータと画像を管理
データ利用条件設定 https://www.hi.u-tokyo.ac.jp/faq/reuse （2019.4～）

今後のスケジュール

- 1か月半に一度、zoomを利用した定期的なオンライン会議の開催する。そのほかは適宜メールでのやり取りを中心に活動を行う。
- 成果物がまとまった段階で、イベント等でのアウトリーチ活動を行う。

2021年10月～　　～2022年9月

事例収集
委員の所属機関、または関係する機関のデータ共有・公開の事例をまとめ、委員の間で共有する。

事例のとりまとめ
委員から報告された事例をまとめ、データを共有・公開する際に共通する問題意識や課題などを整理し、論点をピックアップする。

成果物の作成
委員が分担し、成果物（文書、動画、スライドなど）を作成する。

成果物公開
アウトリーチ
成果物の公開と、イベント等でのアウトリーチを実施予定。

ご清聴ありがとうございました