## RDUF小委員会 「人文学・社会科学のデータ共有における課題検討」 データ共有の取り組みに関する事例集

#### 収録事例一覧

1. 国立国語研究所 (堤智昭・筑波大学人文社会系) 3ページ

2. 国文学研究資料館 (岡田一祐・北海学園大学人文学部) … 14ページ

3. 東京大学史料編纂所 (平澤加奈子・東京大学史料編纂所) … 27ページ

5. 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・東京大学社会科学研究所) … 41ページ

6. 慶應義塾大学パネルデータ設計・解析センター (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) … 48ページ

\*なお、本事例集に収録した内容は、各担当によるスライド作成時点のものである。

# 国立国語研究所のデータ公開事例

筑波大学 堤智昭

2022/09/30

# 国立国語研究所について

大学共同利用機関法人人間文化研究機構の機関

共同利用機関法人

現代語・古典語、標準語・方言、書き言葉・話し言葉、日本語の非母語話者による日本語習得過程など、日本語研究の基礎データとなる大量の言語資源を整備し、大学・研究者コミュニティ・一般社会に提供する

- 。提供するデータ(種別ごと)
  - 。コーパス、オンライン辞書、言語地図、画像・PDF、ツール、カタログ、その他

https://www.ninjal.ac.jp/info/aboutus/

https://www.ninjal.ac.jp/resources/

# 研究所のデータ公開方針

- ●国立国語研究所オープンアクセス方針
- ●国立国語研究所研究データ管理・保存・利活用ポリシー
- ●国立国語研究所学術情報リポジトリ運用指針
- ●共同利用推進センターが管理・配布するデータセットに対するライセンス付与の基本方針

が公開され、これを基準に細かな部分は各サービスごとに決めている

ポリシー公開URL: https://www.ninjal.ac.jp/resources/research-policy/

# コーパス開発センター

## 言語コーパスの開発・提供

- 。現代日本語書き言葉均衡コーパス(BCCWJ)
- 。日本語話し言葉コーパス(CSJ)
- 。近代語コーパス(CMJ)
- 。日本語歴史コーパス(CHJ)

他機関からのデータ受け入れ(移管)

。名大会話コーパス

https://ccd.ninjal.ac.jp/index.html https://mmsrv.ninjal.ac.jp/nucc/



# コーパスの利用について

コーパスを公開して、どなたにでも利用していただけるようにするため、著作権処理を実施

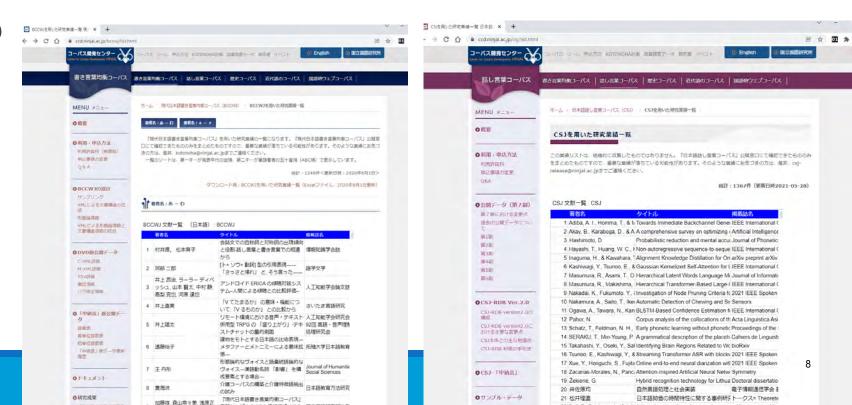
。現代語等の著作権があるものは許諾をとっている

個人情報保護法に抵触する情報があればテキストの該当部分を伏せ字

- 。場合によってはサンプル全体を除外
- 。個人情報保護法には抵触しなくても、著作権者からの希望があれば、情報 の一部を伏せ字
  - 。例えば新聞社からの要請で、新聞記事中の「~容疑者」「~被害者」等の実名部分を伏 せ字にしている

# コーパスを用いた研究成果

- コーパスを利用した研究成果には「利用による成果」である旨の明示を依頼
- 。窓口への申し出、確認できたものを成果一覧にまとめている
- 。任意・手動なところが
- 。一部抜けがある場合も



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

# Web茶まめ

コーパスの研究成果であるUniDicを使った形態素解析ツール

Web上で動く形態素解析アプリ

- 。研究成果・公開データの利用促進を目的
- 。研究・授業で使われている
- 。案外利用者がいる?
- 。利用登録や制限は特になし
  - 。2015年公開以降利用上の問題はない

高負荷になって他サーバに影響を及ぼしたりなど 利用が増えたら可能性はあるが今のところは問題なし

https://chamame.ninjal.ac.jp/



# 画像・データの公開

## 国立国語研究所が所蔵するもの

- 。国語研図書館所蔵の貴重書
  - C•Cライセンス⑥ ① で公開
  - 。 少し前に(2021年10月22日) NCがとれて商業利用も可能に・・・・
  - 。原本の資料名と「人間文化研究機構国立国語研究所所蔵」であることを表示すれば良い

## 他機関が所蔵するもの

- 。米国議会図書館アジア部日本課が所蔵する『源氏物語』
- 大英図書館提供の天草版『平家物語』『伊曽保物語』『金句集』「言葉の和らげ」「難語句解」
  - 。各機関から公開許可を得て公開している
  - 。多くの場合は公開元もC・Cライセンスで公開している

# データ公開の問題点

## 成果物の窓口への申し出

- 多くは申し出をして頂いているが一部抜ける場合もある
- 。全数把握はやはり困難

## とにかく維持が大変

- 。他機関からの移管されたもの
  - 科研費の期限が来て管理しきれなくなったもの等もある
  - 。名大会話コーパス(基盤B 平成13年度~15年度)
  - ・雑誌『国語学』全文データベース(日本語学会からの移管)
    - 。PHP3時代に作られた検索プログラム
    - ・謎のデータ形式
    - 。時々出てくるセキュリティ問題

# Webセキュリティに関して

公開している世界中からWebページにガンガン攻撃が来る

。よくあるフレームワークのセキュリティホール探し、よく使う名前のファイル探 し等など

## 所内の公開要件も年々厳しく

- 。PHP等使っているプログラムのバージョンを次々上げていかないといけない
  - 。PHP7.4を使っていると2022年3月で公開停止、7.3以下はその前に公開停止済み
- 。使っていると思っていなかったJqueryのバージョンで引っかかったり
- 。Log4j脆弱性のようなものが出ると対応を迫られる
- 各プロジェクトで対応しているとなかなか大変
  - 。動的コンテンツは極力減らす方向になりつつある

# おわり

# 国文学研究資料館と人文学データ共有歴史的典籍ネットワーク事業を中心に

# 本発表について

- 本発表では、国文学研究資料館の人文学データ共有について紹介 します
- ・発表者は現在、国文学研究資料館に所属しておらず、また、国文 学研究資料館から本発表の内容を承認されたものではありません
  - あくまで発表者の私見を述べるものです。

# 本発表の構成

- ・ 国文学研究資料館の紹介
  - ・日本文学と日本史にかんするデータ
- 歴史的典籍NW事業の紹介

# 国文学研究資料館の紹介

## 略史

- ・1966年、日本学術会議が国語・国文学研究資料センターの設置を提言
  - ・1970年に学術審議会が国文学研究資料センターの緊急設置を報告
- ・1972年、旧三井文庫の敷地にあった文部省史料館(1951年)を包含するかたちで**国文学研究資料館**が設置
- ・2004年、大学共同利用機関法人化、08年現在地に移転
- ・2013年、古典籍データベース研究事業センターを設置、14年に古典籍共同研究事業センターに改称
  - ・2021年、人文学データ駆動研究センター・学術大型研究計画設置準備 室が設置

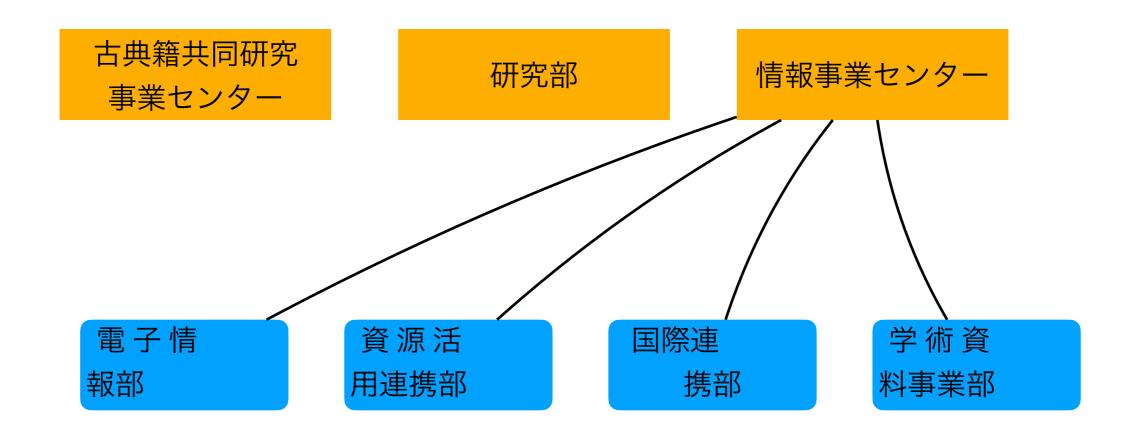
# 国文学研究資料館の紹介

## ミッション

- ・「国内各地の日本文学とその関連資料を大規模に集積し、日本文学をはじめとするさまざまな分野の研究者の利用に供するとともに、それらに基づく先進的な共同研究を推進する」
  - 各地の古典籍調査を継続して実施(書誌、必要に応じてマイクロフィルム撮影)
- ・設置経緯から、史資料分野の研究者が所属

# 国文学研究資料館の紹介

## 組織構成



# 国文研のデータ形成

## 国文研本体

- ・1974年に文学電子テキストの検索実験(東大大型コンピュータ?)
- 1978年から国文研内で大型コンピュータ稼働、各種検索システムの実験がスタート
  - ・古典文学のテキストを独自フォーマット、SGML、XMLなどで作成試 行
  - ・岩波書店刊行の国書総目録を引継ぎ、日本古典籍総合目録データベース として維持
    - ・明治なども(近代書誌・近代画像データベース)
  - ・国文学年鑑(~2007年)→国文学論文目録データベース

# 国文研のデータ形成

## 史料館

- ・史料情報共有化データベース
  - ・「全国の歴史資料保存利用機関が閲覧公開する史料情報を、学術研究・史料保存活動に関係する人々をはじめ、歴史資料に興味をもつ全ての人びとに、インターネットを通じて広く提供することを目的に作成」
- ・館蔵史料のデータベースなども

# 歴史的典籍NW事業の紹介

## 概要

- ・人文社会科学系初の大規模学術フロンティア促進事業
- ・日本語の歴史的典籍(≒古典籍)の画像データベースの構築
  - ・大学等所蔵機関と連携し、30万点の古典籍画像を作成・公開
  - ・ 国内外の拠点大学・機関を中心に、撮影を進めている
- ・2013年を準備年度とし、2014年から10か年計画

#### 唯一の日本古典籍ポータルサイト

キーワードを入れて検索 Q 検索 詳細検索

● 書誌から探す ○ 画像タグから探す(全てのタグ) ○ 全文から探す(収録書誌)

#### C

#### お知らせ

2021.12.8 システム保守点検のため令和3年12月23日(木) 9:30~15:00 (予定)の間、断続的にサービス停止が発生いたします。新日本古典籍総合データベースはご利用いただけません。 ご迷惑をおかけいたしますが、何卒ご理解いただきますようお願い申し上げます。

2021.11.19 新日本古典籍総合データベースで、下記のとおり画像を公開いたしました。 ・筑波大学附属図書館所蔵資料400点

2021.6.3 新日本古典籍総合データベースの公開画像件数が15万点を突破しました。

2020.12.7 現在、タグを付与した画像だけを検索するシステムはメンテナンス中です。

2020.7.28 「国書総目録」のページへは、右上の三点メニューをクリックしてください。

https://kotenseki.nijl.ac.jp/



#### おすすめキーワード

本サービスでよく利用されるキーワードから検索することができます。

源氏物語 伊勢物語 犬 徒然草 源氏 万葉集 医学 古今和歌集 枕草子 物語



#### ピックアップコンテンツ

# 歴史的典籍NW事業とデータ

- ・ 書誌データ
  - ・国文研所蔵のものを中心に書誌データ、画像データ(本文データ)を揃えて国文研オープンデータセットとして提供
- 画像データ
  - ・国文研で撮影済みのマイクロ資料のものをデジタル化
  - あらたに事業でカラー撮影
- 本文データ
  - くずし字データセットとして派生

# 歴史的典籍NW事業とデータ

- ・書誌データにはDOIを附与
- ・画像データは、所蔵先の意向に応じて、従来通りの制限的利用・再利用可ライセンス(CC、非商用制約も容認)・利用制限無しでの提供
  - ・いずれについても、画像ウェブサービスの一般的形式であるIIIFで 公開されているので、利用はしやすい
- ・本文データは作成が進んでおらず、まとまった形では提供していない
  - ・副産物としてのくずし字画像については、オープンデータセット として公開済み

# 現状と課題

- ・ 古典籍は著作権が基本的には消尽している
  - ・所有者の意向という「疑似著作権」問題
    - ・ほんとうはCCも……
  - ・差別問題などの課題もある
    - ・古典籍の内容に馴染みすぎて、われわれが差別をしてしまうなどのことは避けなければならない
- ・本文提供が進んでいないのが難



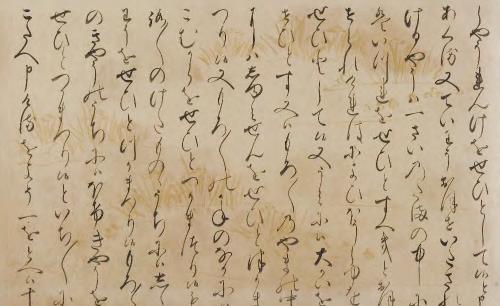
# 東京大学史料編纂所における データ利活用の状況

東京大家史料編集所

HISTORIOGRAPHICAL INSTITUTE

The University of Tokyo

平澤 加奈子(学術専門職員・シニアURA)





## 東京大学史料編纂所の紹介

- ・日本の古代・中世・近世・明治維新期の史料を調査・研究。
- →『大日本史料』『大日本古文書』などの史料集を編纂・出版。
- →広く市民・研究者に提供。日本史研究の基盤を支える。

#### 〇史料収集事業

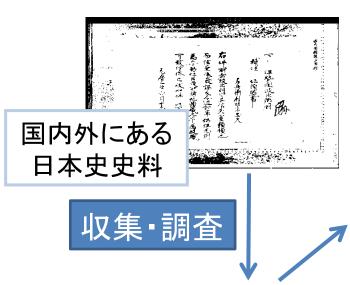
- ・明治以来、国内外に所在する歴史史料の調査・撮影を継続。
- ・近年では、収集したデータのデジタル化を進めるとともに、 デジタル撮影からデータベース搭載にいたる進捗を総合的に管 理する仕組みを整備。
- ・技術の進化にあわせてシステムを更新し、100年以上にわたる史料調査の成果を、常に最新の状態で蓄積。



## 史料編纂所の研究フロー

対象:前近代日本史史料(奈良~明治5年)

例:『大日本古文書』



メタデータ

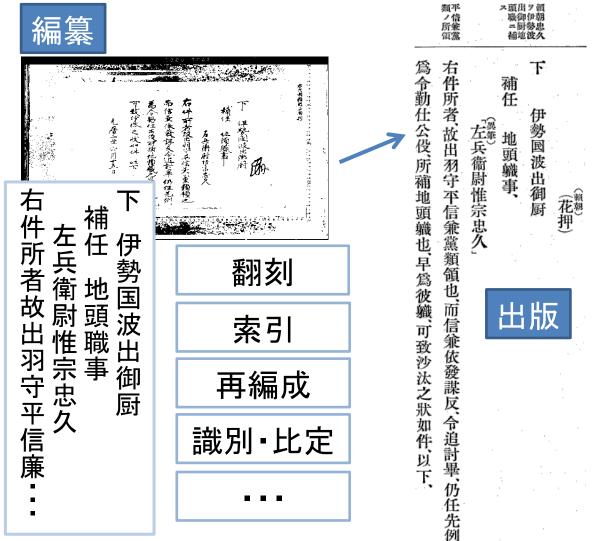
•名称:源頼朝下文

●日付:元暦2年6月15日

史料画像

•Tiff, 400dpi, 256 Gray

scales



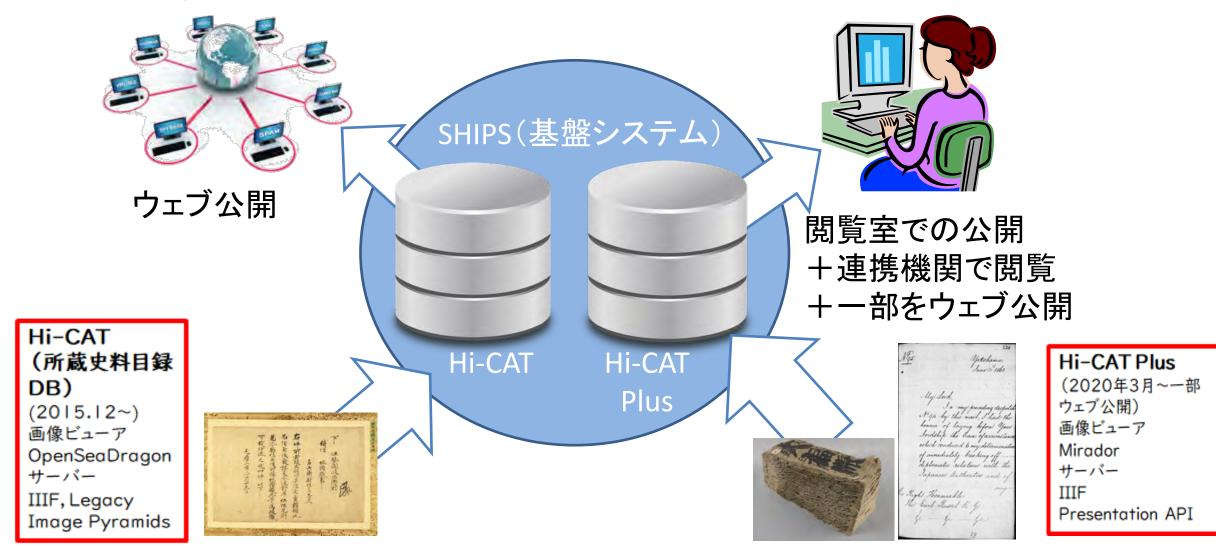
## 東京大学史料編纂所のデータベース

種類 目録,本文,画像・ 図像・写真、知識 (人名・地名・和 暦・文字など) 範囲 日本前近代 (奈良~維新) DB数 40 うち所外公開:30



1984年から作成開始。史料編纂所の編纂・研究活動から発展。

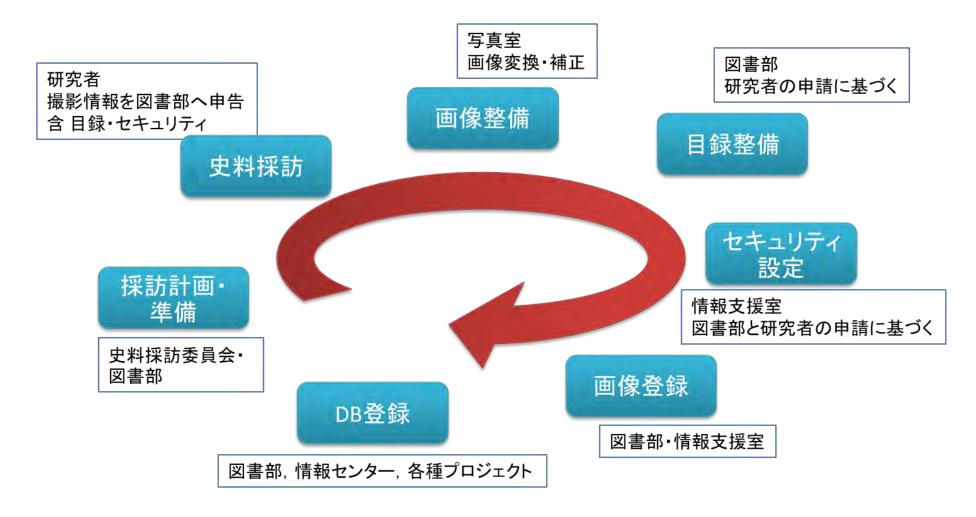
## 史料編纂所所蔵史料と所外史料のデータ管理



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

# 史料採訪の流れ:機関としての規約化



史料編纂所内で規約化(規約化は2008年開始) 試行は2010年,実施は2011年から

## 利用条件の明確化

#### 画像データの利用条件設定(2019年4月公開)

### 東京大学史料編纂所所蔵史料の掲載・放映手続きについて



貴重書(原本・ 古写本類)

貴重書 (特殊蒐書)

**写本** (本所での請求記号が **4**から始まる史料 例:4101-1) **その他**(左記の種別に 当てはまらないもの)

※「貴重書(模写)」「影写本」など



※「貴重書(模写)」 は含まれません。 お問い合わせくだ さい。



**2** 

画像公開の確認

本所データベースで 画像が公開されている\*1



図書室にお問い合わせください

https://www.hi.u-tokyo.ac.jp/library/contact/index.html



\*1「大日本史料稿本」については、 掲載・放映等への利用はお断りします。 \*2 「貴重書(原本・古写本類)」、「特殊蒐書」で公開画像がなく、既存の刊行物等(許可済)から画像を転載する場合は、本所への問い合わせは不要です。

利用の際は東京大学史料編纂所の所蔵史料であることを明示してください。例:『史料名』(東京大学史料編纂所所蔵)

#### CC BY 4.0 相当の条件でご利用いただけます

利用の際は東京大学史料編纂所の所蔵史料であることを明示してください。

例:『史料名』(東京大学史料編纂所所蔵)

・改変を行った際はそれについて明示してください。

例:『史料名』(東京大学史料編纂所所蔵)を改変

改変した史料画像データを本所が作成・公開したと捉えられるような公開・公表は行わないでください。

・「大日本維新史料稿本」の場合は下記の例のような表示をお願いします。

例:東京大学史料編纂所所蔵「大日本維新史料稿本」所引「三田藩家老日記」

#### 注

- ・CC BY 4.0 相当の条件に該当する史料で、本所の公開画像を使用せず、既に掲載使用されている画像(既刊書に掲載など)を使用する場合や、複製申請により入手した画像を使用する場合にも、問い合わせ手続き不要です。所蔵表示は左と同じです。
- ・史料種別(区分)が「本所出版物」の画像データ(「史料集版面ギャラリー」掲載)のご利用に関しては、本所webサイト(<a href="https://www.hi.u-tokyo.ac.jp/fag/reuse/index.html">https://www.hi.u-tokyo.ac.jp/fag/reuse/index.html</a>)をご確認ください。

## 利用条件の明確化②

#### 本所出版物 (版面画像データ)

- ・史料集版面ギャラリーにて閲覧可能 https://www.hi.u-tokyo.ac.jp/dip/
- ・利用条件を設定

https://www.hi.u-tokyo.ac.jp/faq/reuse/

史料集版面ギャラリー

東京大学史料理算所により開算、出版した史映真の版画画像ギャラリーです。福算・出版の詳細は工芸をでご確認ください。 グリエイティブ・コモンズ・ライセンスの「CCBYNICSA」(グリエイティブ・コモンズ 表示・非算利・証440国際ライセンス)相当の条件で提供しています。詳細は工芸会でご確認ください。

- 『大日本史料』[+] ※
- ・『大日本古文書』(+)
- · 『大日本近世史料』[+] 🏎
- INDAMEDIA [4] ~
- 『大日本維新史料 編年之部』(+) 🛥
- 『大日本維新史料 類類之部』[+] —
- 『復古記』[+] ==
- 『明治史要』[+] \*\*
- · 『維新史』[+] 😕
- 『明冶維新史料選集』[+]
- 『保古飛呂比』[+] ※
- 『日本関係海外史料』[+]
- 『日本関係海外史料目録』[+] ≥

東京大学史料掘簋所

#### 利用条件:史料画像データの利用

東京大学史料編纂所(以下、「本所」)が公開している画像データ(関連するメタデータを含む)の利 用条件は下記の通りです。

#### ■ 【1】本所が所蔵する原本等の史料画像データ

史料種別(区分)が「貴重書(原本・古写本類)」・「特殊蒐書」(一部を除く)・「写本」のものが 該当します。

※史料種別は、所蔵史料目録データペース検索結果画面では下図の例のように表示されています。



#### 【利用の際の条件】

- 利用の際は東京大学史料編纂所の所蔵史料であることを明示してください。例:『史料名』(東京 大学史料編纂所所蔵)
- 改変を行った際はそれについて明示してください。
   例:『史料名』(東京大学史料編線所所蔵)を改変
   改変した史料画像データを本所が作成・公開したと捉えられるような公開・公表は行わないでください。

該当のデータはクリエイティブ・コモンズ・ライセンスの「CC BY」(クリエイティブ・コモンズ <u>表示 4.0 国際ライセンス</u>)相当の条件で提供しています。上記の「利用の際の条件」に従う限り、手続きなく 自由に利用できます。利用料は発生しません。

本件に該当するもののうち、「特殊蒐書」のリストは<u>こちら</u>をご覧ください。 画像ファイルのダウンロード方法は、<u>こちら</u>をご覧ください。

#### ■ 【2】本所出版物の版面画像データ

史料種別(区分)が「本所出版物」の画像データすべてが該当します。該当のデータは<u>史料集版面ギャラリー</u>にて閲覧することができます。

#### 【利用の際の条件】

• 利用の際は東京大学史料編纂所の編纂・出版物であることを明示してください。

例: 『書名』 (東京大学史料編纂所編)

改変を行った際はそれについて明示してください。

例:『書名』(東京大学史料編纂所編)を改変

改変した史料画像データを本所が作成・公開したと捉えられるような公開・公表は行わないでください。

該当のデータはクリエイティブ・コモンズ・ライセンスの「CC BY-NC-SA」 (クリエイティブ・コモンズ 表示 非営利 - 継承 4.0 国際ライセンス) 相当の条件で提供しています。

非営利目的に限り、上記の「利用の際の条件」に従う限り、手続きなく自由に利用できます。利用料は 発生しません。

## 利用条件の明確化3

#### 所外の画像史料データ

- ・現在、一部の画像データを 本所DB(Hi-CAT Plus)より所外 公開
- ・利用条件については、各史料 群により異なる(史料所蔵機関 との取り決めによる)

\*例:都城島津邸所蔵史料

https://www.hi.u-

tokyo.ac.jp/faq/reuse\_miya
konojo/



Hi-CAT Plusは、史料編纂所がこれまで撮影・収集した国内および海外に所在する史料の画像データを閲覧するためのデータベースです。 このうち所蔵者の許可を得られたものは、検索結果にイメージボタンが表示され、画像データを閲覧できます。

【史料群】	說明
禁裏公家文庫	<u>天皇家・公家関係史料を中心とした研究プロジェクト</u> (禁裏公家文庫研究)によるモノクロマイクロフィルムスキャン・カラーデジタル撮影データ <u>■公開史料と利用の案内</u>
都城島津邸所蔵史料	部城島津邸所蔵史料(宮崎県都城市)のカラーデジタル撮影データ <u>■公開史料と利用の案内</u>
東寺文書(琵琶湖博蔵)	滋賀県立琵琶湖博物館(滋賀県草津市)所蔵「東寺文書」のカラーデジタル撮影データ <u>■公開史料と利用の案内</u>
岩倉具視関係史料	海の見える杜美術館(広島県廿日市市)所蔵「岩倉具視関係史料」のモノクロマイクロフィルムスキャン画像データ (*目録データの注記欄に思文閣出版刊『岩倉具視関係史料』の巻・文書番号を記載) <u>■公開史料と利用の案内</u>

その他の区分は 史料編纂所図書室端末 で利用できます。

データベースについて 利用規定 データベース更新スケジュール よくあるご質問

©2022 東京大学史料編纂所

# KU-ORCASの データ運用管 理体制

#### 菊池信彦

(関西大学アジア・オープン・リサーチセンター特任准教授\*)

\*本資料作成当時/2022年度から国文学研究資料館



# はじめに

- 自己紹介
  - 西洋史、NDL、DH研究から関大でDA企画運用管理 ≠開発
- 報告内容
  - KU-ORCASの関西大学デジタルアーカイブの運用管理体制
- KU-ORCASとは
  - 東アジア文化研究(中国語、中国史研究が中心)のDH研究とその成果を発信するPH研究のプロジェクト
  - 2017年度~2021年度の5年間/次年度も継続予定

# 関西大学デジタルアー カイブ

- 関西大学デジタルアーカイブ
   (https://www.iiif.ku-orcas.kansaiu.ac.jp/)
- ■関大図書館所蔵資料が大半、教員個人蔵書が一部
- 構成データベース4つ
  - 東アジア(なんでもあり、中国語教育、映画パンフ等)
  - 大阪画壇(絵画)
  - 泊園文庫(漢籍メイン)
  - 泊園印章 (ハンコ)
- ●他に、コロナアーカイブ、近代漢語語料 庫、VR2つ

# メタデータと画像サーバの管理

- メタデータ: データベースでそれぞれ異なり、図書館書誌をベースにしつつも、独自項目が立てられていて、把握が難しい
- 画像サーバ:
  - サーバ保守は外注だが、一次対応は任期付き教員が対応
  - サーバ内データは長年の研究の蓄積から整理が困難
- →維持管理の難しさ

# 課題とその解決に向けて

- デジタルアーカイブという実務を多く含む領域のため、長期的な維持管理にはデジタル な研究活動を行う「アーキビスト」が必要
- だが、新しいポストは難しいので、
- ①任期付きの研究者が進めやすいように、協同を確立させる
- →そのためには、人文系研究者の研究スタイルの転換に向けた意識改革と事務と研究者の協同体制が必要
- ②評価体系の構築し、研究者のインセンティブを高める必要もある。
- →「組織的営み」としての研究に必要な、サーバやデータの維持管理という業務を研究評価に加え、それを組織内外で共有させていくこと



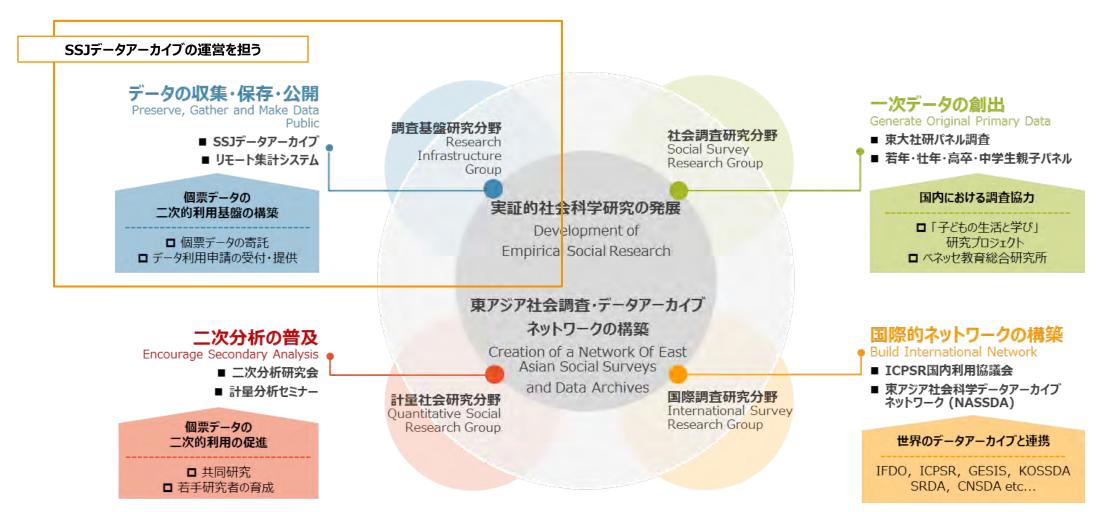


## SSJデータアーカイブにおけるデータ利活用の状況

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)

## 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)

大きく4つの分野で構成されており、そのうちの1分野がSSJデータアーカイブの運営を担っている。





## Social Science Japan Data Archive (SSJデータアーカイブ、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(当時は日本社会研究情報センター)に、1996年5月 SSJDA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
- データの受け入れは無償。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。
- データの利用は無償。ただし利用登録制。
- 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。





### 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。

#### ①主なデータ寄託機関

□公的研究機関

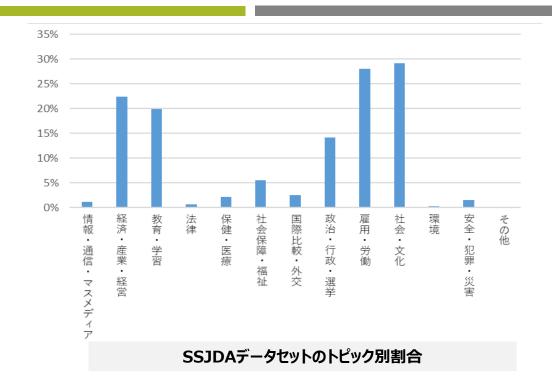
日本政策金融公庫総合研究所、労働政策研究・研修機構、 2015SSM研究会データ管理委員会、東京大学社会科学研究所、その 他個人研究者など

□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、ベ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



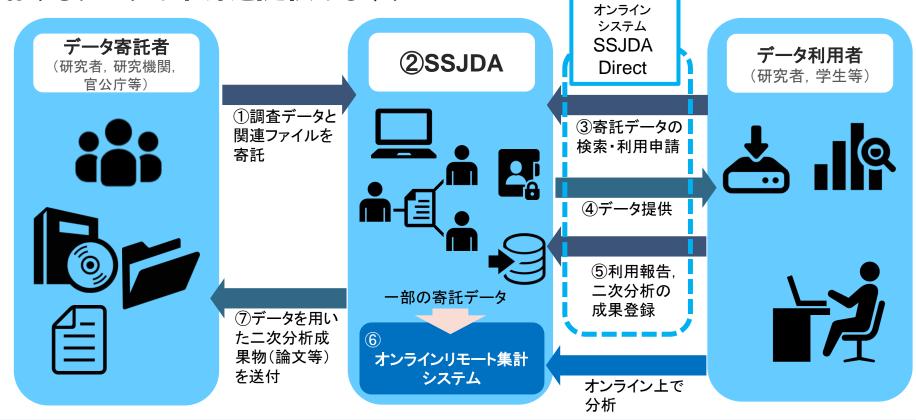
#### ②主なデータセット

- □ 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
- □ 全国就業実態パネル調査 日本における就業状態の変化、所得の変化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





## SSJDAにおけるデータの収集と提供のしくみ



- ① 寄託者よりデータの寄託を受託。受託の際にはSSJDAスタッフがデータ概要に関する聞き取り等を行う。
- ② SSJDAスタッフによる寄託データのキュレーションを実施し、統計ソフトSPSS、Stataの各形式で提供用データを用意。メタデータを作成し、オンラインデータ検索・申請システム「SSJDA Direct」より公開する。
- ③ 利用者はSSJDA Directを通じてデータを検索し、利用申請を行う。
- ④ SSJDAは、利用者からのデータ利用申請をチェックし、認められた申請についてデータを提供する。
- ⑤ データ利用期間終了時、利用者はSSJDA Directを通じてデータの利用報告ならびに成果物の登録を行う。
- ⑥ 一部のデータについて、利用者はデータ利用登録を行うことなく、リモート集計システムを通じてオンライン上でデータ分析を行うことが可能。
- ⑦ 利用者のデータ利用状況、および提出された成果物は、SSJDAから寄託者へ報告・送付する。





## 提供データの利活用について



- データ利用者数は年々増加している(表1)。
- データ寄託者とデータを利用して執筆された論文に対して、独自の表彰制度を設けている(写真1)。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)選出して表彰している(寄託者表彰)。 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(優秀論文表彰)。
- 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、次の2つの事業を 実施している。
  - ① 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(写真2)。年度末には成果報告会を開催し、成果報告書を刊行している。
  - ② SSJDA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(写真3)。主に学生や若手研究者を対象とした 内容となっている。

#### 写真1 寄託者·優秀論文表彰



寄託者 · 優秀論文表彰式

#### 写真2 公募型共同研究 二次分析研究会



研究成果報告会の様子

#### 写真3 計量分析セミナー



過去のセミナー開催の様子



## 今後の課題と展望

- 予算の確保、人員の確保並びに育成は継続的な課題。
- 人員面での都合から、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、コロナ関連のデータのような、社会情勢から公開の即時性が求められるデータは、優先的に迅速に公開できる体制を構築していく。
- 今後も新規データの受け入れや公開データ数の増加させ、新たなコンテンツやセミナー等のイベントを充実させるなど、利用 者の利便性向上に努めるとともに、利用者数の更なる拡大を目指す。
- 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、日本の実証的社会科学研究の発展に寄与する。





## 慶應義塾大学 パネルデータ設計・解析センター

データアーカイブ機能の紹介

慶應義塾大学 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

#### 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一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

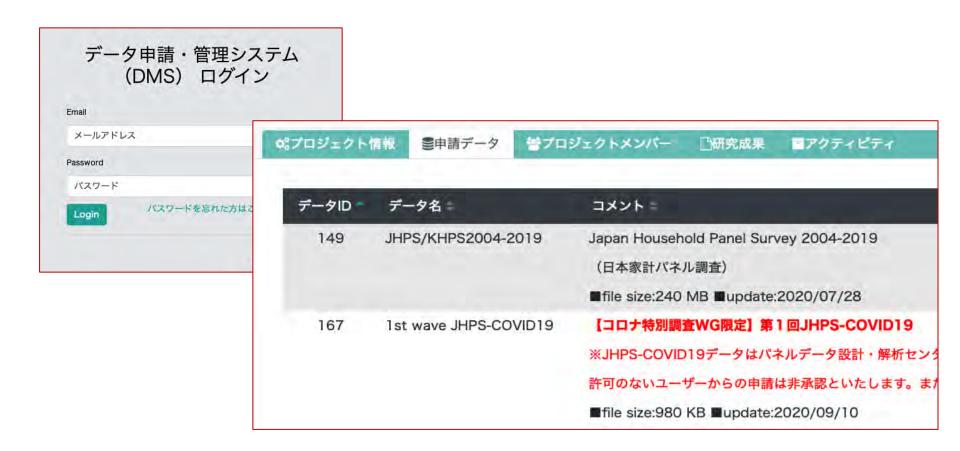
• パネルデータに関するデータアーカイブの中核的な機能を担う機関を目指して、当センターが実施するパネル調査および、他機関から寄託されたパネルデータを国内外の研究者へ公開

Research Center at

## アーカイブ機能:パネルデータの整備・管理・提供

- ✓ データ利用の利便性とデータの質の向上
  - ~ オンライン申請・提供システムの運用
  - ~ 母集団推計ウエイトやデータ構築プログラムの提供
  - ~ データ・変数カタログの整備
  - ~パネルデータの研究業績データベースの整備
  - ~ 外部機関のパネルデータの受け入れ態勢の整備
- ✓ 研究資源の共有と国内外の研究者の参画
  - ~ 6年間で1000件以上のデータ提供・利用
- ✓ 国際的データアーカイブへのデータ提供
  - ~ LIS: 国際比較可能な所得の個票データベース
  - ~ CNEF: 国際比較可能な家計パネルデータベース

#### データ申請・管理システムの画面】



#### データ変数カタログ】

- □ データ・変数カタログの整備
  - JHPS/KHPSの<u>膨大な変数情報をデータベース化</u>し、変数の内容・該当年・ 回答状況などが把握できるシステム。
  - データ管理システム内に設置。アカウント作成で誰でも閲覧可能。

