

研究データ利活用協議会(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 活動成果報告書

1. 活動の目的

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して活動を行った。

2. 活動期間

2021年10月～2023年3月

3. 小委員会メンバー

- ・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）
- ・菊池 信彦（人間文化研究機構 国文学研究資料館）
- ・古藤 隆浩（東北福祉大学）
- ・谷口 沙恵（東京大学社会科学研究所）
- ・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）

4. 小委員会開催実績と活動の概要

開催期間：2021年10月27日～2022年10月12日

全7回 すべてオンライン開催

小委員会はおよそ隔月の頻度で開催した。委員はデータアーカイブ／リポジトリに所属する、あるいはかつて所属した人文学系5名（研究者4名、職員1名）、社会科学系2名（研究者1名、職員1名）、データアーカイブ／リポジトリに所属しない大学職員1名で構成され、多様な視点から議論を行った。

小委員会第1回から第4回にかけては、各委員が所属あるいは所属していた機関におけるデータ共有の取り組みについてそれぞれ報告がなされ、意見交換を行った。

小委員会第5回、第6回では、成果報告書についての議論がなされた。当初、2022年9月での活動終了予定であったが、成果報告書の作成にあたりさらに議論を深める必要性があったため、2023年3月までの半年間の延長を申請し、2022年9月の企画委員会にて承認された。

小委員会第7回では、作成した成果報告書ならびにデータ共有事例について意見交換を行い、最終確認がなされた。

5. シンポジウムでの報告

- ・2021年度 RDUF 公開シンポジウム（2021年11月22日）にて、小委員会の活動を報告
- ・2022年度 RDUF 公開シンポジウム（2022年11月11日）にて、小委員会の活動成果について報告

6. 成果報告書の概要

各委員の所属する、または過去に所属していたデータアーカイブ／リポジトリの活動つ

て下記 8.に記載する項目にわけて整理し、まとめたのが、本報告書 4 ページ以下の各機関事例報告である。

また、各データアーカイブ・データリポジトリの活動全般についてそれぞれスライドにまとめたのが、別紙の各機関のデータ共有事例集である。

7. 成果報告書 事例報告機関一覧

- (1) 人間文化研究機構 国立国語研究所
- (2) 人間文化研究機構 国文学研究資料館
- (3) 東京大学史料編纂所
- (4) 関西大学アジア・オープン・リサーチセンター
- (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 (CSRDA)
SSJ データアーカイブ (SSJDA)
- (6) 慶應義塾大学パネルデータ設計・解析センター

8. 各機関事例 記載項目一覧

| |
|--------------------------------|
| 機関概略 |
| 対象とするデータ |
| 上記 URL |
| 研究データとは (各機関で) 何を指すか |
| ポリシーの内容 |
| 個人情報 |
| ライセンス/著作権 |
| 寄託データの扱い |
| 二次利用の対応 |
| 成果物の提供の在り方 |
| 利用登録の有無 |
| 想定利用者 |
| データの質の担保と多様性の確保 |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 |
| データアーカイブ運営上の課題 |

人間文化研究機構 国立国語研究所

2022年9月30日現在

| | |
|--------------------------------|---|
| 機関概略 | 国語研は、日本語や社会における言語生活を科学的・総合的に研究している。日本語・言語に関する知を探究し、その成果を広く社会に提供することを目的としている。 |
| 対象とするデータ | 学問分野としては日本語学や言語学、日本語教育研究と呼ばれる分野が中心。 |
| 上記 URL | https://www.ninjal.ac.jp/ |
| 研究データとは（各機関で）何を指すか | 日本語・言語研究のための基礎資料やデータ（言語資源）。これらを整備し広く社会に提供を行っている。 |
| ポリシーの内容 | 研究成果・研究データの公開方針を決めている。 https://www.ninjal.ac.jp/resources/research-policy/ |
| 個人情報 | 個人情報保護ポリシーが決められている。 https://www.ninjal.ac.jp/utility/policy/#PRIVACYPOLICY 個人情報の利用目的は以下のとおり。 ・回答のため（お問い合わせの場合） ・メールマガジン配信のため（国語研メールマガジンにご登録の場合） ・研究所が実施する催し物の運営（参加者登録、開催案内送付）のため（催し物参加申込の場合） ・研究所が保有する資料・データベース等の利用登録及び本人確認のため（資料・データベース等の利用申請の場合） ・収集した情報を、統計及び分析のために使用する。 また、サイトの利用状況を統計的に把握するため、Google アナリティクスを利用している。Google アナリティクスでは個人情報は収集していない。 |
| ライセンス／著作権 | 国立国語研究所の Web サイトにあるコンテンツは断りのない限り国立国語研究所に帰属する。ライセンスは提供しているもの個別に設定されている。画像等の古典籍資料は CC ライセンスを使うことが多い。解析用 UniDic ではクリエイティブ・コモンズ・ライセンス（CC BY-NC-SA）とフリーライセンス（GPL/LGPL/BSD License のトリプルライセンス）の2種が設定されている。 |
| 寄託データの扱い | 名大会話コーパス（ https://mmsrv.ninjal.ac.jp/nucc/ ）のように受け入れ実績がある。日本語学会の論文データ等も受け入れている。具体的な寄託に関する規定はない。 |
| 二次利用の対応 | 利用する場合は事前連絡が必要。 |
| 成果物の提供の在り方 | 成果物を公表した場合、公表内容について連絡を求めている。 |
| 利用登録の有無 | 一部コンテンツ（コーパス等）は利用登録を求めている。 |
| 想定利用者 | 研究者・学生・一般 |
| データの質の担保と多様性の確保 | 公開物によっては記載がある。共通の規定はない。 |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | 中納言などコーパス関係ツールの使い方講習会をはじめ様々な催し物を実施している。 |

| | |
|----------------|--|
| データアーカイブ運営上の課題 | 国立国語研究所のデータを利用した成果物について、利用の報告が一部行われなことがしばしばあり、運用がうまくいっているか判断する情報の集約が十分に行えない場合がある。Web上に公開する場合、システムのセキュリティ対策が課題となっている。 |
|----------------|--|

(堤 智昭)

人間文化研究機構 国文学研究資料館

2021年12月17日現在

| | |
|--------------------------------|--|
| 機関概略 | 国内各地の日本文学とその関連領域の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 |
| 対象とするデータ | ・（他機関を含む）古典籍・古文書の書誌・撮影画像 ・古典籍・古文書の本文データ |
| 上記 URL | https://www.nijl.ac.jp/ |
| 研究データとは（各機関で）何を指すか | 国文学研究資料館として公開する画像データを指す。 |
| ポリシーの内容 | 「人間文化研究機構国文学研究資料館の画像データのオープン化に関する指針」および「人間文化研究機構国文学研究資料館データベース利用規程」に定められている。 |
| 個人情報 | 現在該当するものがない。 |
| ライセンス／著作権 | オープンデータについては、CC BY-SA を指定している。 |
| 寄託データの扱い | 人間文化研究機構国文学研究資料館データベース利用規程による。 |
| 二次利用の対応 | オープンデータについては所蔵先の明記を求めている。 |
| 成果物の提供の在り方 | 情報提供フォームでの通知を要請 |
| 利用登録の有無 | 利用登録はない |
| 想定利用者 | 研究者・一般利用者 |
| データの質の担保と多様性の確保 | とくになし |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | 研究会やシンポジウムの開催 |
| データアーカイブ運営上の課題 | 本資料作成者は現在当事者ではないので、現状の課題は不明。 |

（岡田一祐）

東京大学史料編纂所

2022年10月31日現在

| | |
|--------------------------------|--|
| 機関概略 | 前近代日本史関係の史料を対象とする研究所。国内外に所在する各種史料を収集・分析して、日本史の基幹史料集として編纂・刊行し、学界および社会にひろく公開し、貢献することを使命としている。 |
| 対象とするデータ | ・前近代日本史関係の史料 |
| 上記 URL | https://www.hi.u-tokyo.ac.jp/ |
| 研究データとは（各機関で）何を指すか | 前近代日本史研究のための基礎資料や関連データ。 |
| ポリシーの内容 | サービス利用ポリシー https://www.hi.u-tokyo.ac.jp/faq/kitei （東京大学史料編纂所公開用データベース利用規定） データ収集ポリシー：「利用条件確認書式」による |
| 個人情報 | プライバシーポリシー（東京大学の規定に準拠） https://www.u-tokyo.ac.jp/gen01/reiki_int/reiki_honbun/au07408321.html |
| ライセンス／著作権 | ・本所所蔵史料画像データは「CC BY」（クリエイティブ・コモンズ 表示 4.0 国際ライセンス）相当の条件で提供 ・本所出版物版面画像データは「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 ・その他の史料については申請などが必要 |
| 寄託データの扱い | 原史料所蔵者と個々にデータの扱いについて「利用条件確認書式」などにより規定。 |
| 二次利用の対応 | 利用条件あり https://www.hi.u-tokyo.ac.jp/faq/reuse （本所所蔵史料画像データ・本所出版物版面画像データ） https://www.hi.u-tokyo.ac.jp/library/use/ （その他の史料） |
| 成果物の提供の在り方 | データユーザーに対しては、成果物に関する情報提供を促している（任意）。 |
| 利用登録の有無 | 一般利用可能なデータについては、利用登録は不要。 |
| 想定利用者 | 研究者・学生・一般利用者 |
| データの質の担保と多様性の確保 | 【質の担保】データの公開にあたっては、担当教員がデータを精査する。その後、担当教員・職員がメタデータ作成や、確認作業を担当する。 【多様性の担保】データやメタデータの英語化に取り組んでいる。 |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | ・研究発表 ・研究会 ・史料編纂所ウェブサイト、Twitter を活用した情報発信。 ・史料所蔵者からの要望によるデータ公開 |
| データアーカイブ運営上の課題 | ・予算、人員確保。 ・業務の更なる効率化。 ・データの長期利用・長期保存。 |

（平澤 加奈子）

関西大学アジア・オープン・リサーチセンター

2022年2月14日現在

| | |
|--------------------|---|
| 機関概略 | 関西大学が所蔵する東アジア文化研究資料のデジタルアーカイブ化、および、それを利用したデジタルヒューマニティーズ研究推進拠点として設置された。直接的には、2017年度の「私立大学研究ブランディング事業」に「オープン・プラットフォームが開く関大の東アジア文化研究」が採択されたことで KU-ORCAS は設立されたが、その前身には 2007 年に採択された「グローバル COE プログラム」、2011 年に採択された「私立大学戦略的研究基盤形成支援事業」などのプロジェクトがあり、関西大学の東アジア研究およびデータベース構築は続けられてきた。 |
| 対象とするデータ | 関西大学デジタルアーカイブには、漢籍や中国語の語学資料等様々な東アジア関係資料のほか、大阪画壇（近世絵画）の資料、江戸末期から昭和初期にかけての漢学資料等および印象資料がある。 |
| 上記 URL | https://www.iif.ku-orcas.kansai-u.ac.jp/ |
| 研究データとは（各機関で）何を指すか | 研究データに関する具体的な定義は（少なくとも）外部に公開されている情報はない。 |
| ポリシーの内容 | 論文等の機関リポジトリ公開を定めたオープンアクセスポリシー（ https://www.ku-orcas.kansai-u.ac.jp/oa-policy/ ）を策定している。ただし、このポリシーは研究過程の中間データを対象にするものではない。 |
| 個人情報 | 個人情報の取り扱いを定めたプライバシーポリシー（ https://www.kansai-u.ac.jp/ja/info/privacy.html ）がある。KU-ORCAS では、実施していたコロナアーカイブ@関西大学（すでに閉鎖）で上記ポリシーに準拠した運営を行っていた。 |
| ライセンス／著作権 | 関西大学デジタルアーカイブは著作権保護期間満了を迎えた資料を公開している。そのため、公開している画像ファイルはパブリックドメイン、そのメタデータは CC0 ライセンスで提供している。 |
| 寄託データの扱い | 関西大学デジタルアーカイブは、関西大学総合図書館と関西大学所属研究者の個人蔵書を公開しており、寄託されたデジタル化データは提供していない。ただし、コロナアーカイブ@関西大学（すでに閉鎖）では、資料提供時に、クリエイティブコモンズライセンス CC-BY-NC の条件を承諾するよう求めていた。 |
| 二次利用の対応 | 関西大学デジタルアーカイブには「二次利用に際してのお願い」のページを公開しており、そこでは、「『関西大学デジタルアーカイブ』で提供する書誌情報や資料画像等に関しては、原則として、関西大学アジア・オープン・リサーチセンター（KU-ORCAS）の許諾なく、商用・非商用を問わず、無償で自由にご利用いただけます。複製、改変、出版物やウェブサイトへの掲載も自由です。」と記載している。ただし、「二次利用に際しては、関西大学デジタルアーカイブが提供する画像データであることを、著作権法に定める引用の方法に準じて明示していただく」こと、また、「画像データに対して加工や変形等の操作を行った場合は、その旨あわせて明示を」すること、そして「関西大学デジタルアーカイブを利用して得られた成果物（研究論文、専門書、復刻本等）を、KU-ORCAS 宛てに、1 部寄贈」を求めている。だが、それらはあくまで「お願い」ベースであり、必須のものではない。 |
| 成果物の提供の在り方 | 前項の「二次利用の対応」記載の通り、「関西大学デジタルアーカイブを利用して得られた成果物（研究論文、専門書、復刻本等）を、KU-ORCAS 宛てに、1 部寄贈」を求めている。 |
| 利用登録の有無 | 関西大学デジタルアーカイブでは利用登録することなく利用可能である。ただし、KU-ORCAS が提供している近代漢語語料データベースは利用に際し登録が必要となっている。 |
| 想定利用者 | 東アジア研究者を中心に、異分野の研究者、企業、自治体、市民といった人々を利 |

| | |
|--------------------------------|---|
| | 用対象としている。 |
| データの質の担保と多様性の確保 | 具体的な記載なし |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | 研究集会および国際シンポジウムの開催を通じた広報活動 |
| データアーカイブ運営上の課題 | 「デジタルアーカイブアセスメントツールに基づく関西大学デジタルアーカイブの自己評価結果の公表」（2021.02.01 付／URL: https://www.iiif.ku-orcas.kansai-u.ac.jp/news/20210201-0 ）によると、「関西大学デジタルアーカイブでは、『1. 組織的基盤の取組』に大きな課題を抱えていることがうかがえます。」とある。これは、「デジタルアーカイブアセスメントツール（ver.2.0）」に基づく自己評価であり、具体的には「組織的基盤の取組」に記載の「予算」、「人材確保」、「人材育成」、「長期保存（方針等）」、「長期保存（体制整備）」において課題があるとされている。なお、ここで指摘している課題は、あくまで2021年2月1日付の記事に基づくものであり、その後も継続していることを示すものではない。 |

（菊池 信彦）

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) SSJ データアーカイブ (SSJDA)

2022年10月31日現在

| | |
|--------------------|--|
| 機関概略 | 日本の労働調査、社会調査の個票データを収集・保管し、学術目的（一部データは教育目的）での二次的な利用のために提供する |
| 対象とするデータ | 日本社会を対象とした社会科学分野の質問紙調査による量的データ、個票データ |
| 上記 URL | https://csrda.iss.u-tokyo.ac.jp/ |
| 研究データとは（各機関で）何を指すか | メタデータを含む社会調査データ |
| ポリシーの内容 | <p>下記 URL「寄託の手引き」にて「SSJ データアーカイブ・データポリシー」を記載 https://csrda.iss.u-tokyo.ac.jp/96b3d9684193caebc46846519e477478d0f228c3.pdf</p> <ol style="list-style-type: none"> 1. データコレクションポリシー 寄託を受け付けるデータ分野、トピック、調査方法、データフォーマットなどを規定。 2. データ公開ポリシー データを公開するうえで必ず従うべきポリシーを規定。 3. 個人情報保護ポリシー SSJ データアーカイブに所蔵されている個票データは、匿名処理を施しており個人情報保護法の定める個人情報には該当しないが、これに準じた形で保護されることを明記。 4. データ管理ポリシー 個票データの管理体制について規定。 |
| 個人情報 | <p>上述のとおり「寄託の手引き」内に個人情報保護ポリシーを記載。また別途、CSRDA として個人情報保護方針を定めている。 https://csrda.iss.u-tokyo.ac.jp/csrda/privacy.html</p> |
| ライセンス／著作権 | <p>データの権利は寄託者が持つ。 SSJDA が作成したメタデータについては、ウェブサイト上での特段の明記はしていないが、著作権は放棄していない。</p> |
| 寄託データの扱い | 上述の「寄託の手引き」に定められている通り。 |
| 二次利用の対応 | <p>利用条件あり。 https://csrda.iss.u-tokyo.ac.jp/infrastructure/access/condition/</p> |
| 成果物の提供の在り方 | <p>成果物がある場合には、SSJDA データ提供システム（SSJDA Direct）での登録に加えて、紙またはデジタル形式で SSJDA まで提出してもらうよう利用者に依頼している。また、利用者が成果物を発表する際には、寄託者と SSJDA への謝辞を掲載する必要がある。</p> |
| 利用登録の有無 | 必要 |
| 想定利用者 | 研究者、学生。なおデータ寄託機関に所属する者であれば、研究者以外でも利用可能（ただし学術目的、教育目的での利用に限る）。 |
| データの質の担保と多様性の確保 | <p>（質の担保）データの公開にあたっては、寄託データ担当教員がデータを精査する。その後、データアーカイブマネージャーがデータクリーニング、およびクリーニング後のデータの確認作業を担当する。一定基準に達していない質のデータであれば公開しない場合もある。 （多様性）すでにデータを寄託している研究者だけでなく、新たな寄託者</p> |

| | |
|--------------------------------|---|
| | <p>候補である研究者・機関にデータ寄託を依頼するなどして、新規寄託者を増やすように努めている。</p> <p>メタデータと調査票の英語化を進めており、国際的な利用に対応している。各メタデータには DOI を付与しており、永続的なアクセス性を担保している。</p> |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | <ul style="list-style-type: none"> ・ SSJDA のデータを用いた公募型研究会の開催。 ・ 質問公募型の調査の実施と、質問を採択された研究者への調査データ提供。および SSJDA を通じた調査データの公開。 ・ SSJDA のデータを用いた計量分析セミナーの開催。 ・ SSJDA のデータを用いて執筆された論文に対する優秀論文表彰制度。 ・ ディスカッションペーパーの発行。 ・ CSRDA のウェブサイト、東京大学社会科学研究所メーリングリスト、CSRDA の公式 SNS (YouTube、Twitter) を活用した情報発信。 |
| データアーカイブ運営上の課題 | <ul style="list-style-type: none"> ・ 予算、人員確保は継続的な課題。 ・ 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、社会から公開の即時性が求められるようなデータは、優先的に迅速に公開できる体制を構築していく必要がある。 ・ 業務の更なる効率化のために DX 化を進める必要がある。 |

(谷口 沙恵)

慶應義塾大学パネルデータ設計・解析センター

2022年10月31日現在

| | |
|--------------------------------|---|
| 機関概略 | 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提供する |
| 対象とするデータ | 個人・企業を対象としたパネル調査データ |
| 上記 URL | https://www.pdrc.keio.ac.jp/ |
| 研究データとは（各機関で）何を指すか | パネル調査により得られた情報をデータ化したもの |
| ポリシーの内容 | サービス利用ポリシー https://www.pdrc.keio.ac.jp/service-policy/ サイトポリシー https://www.pdrc.keio.ac.jp/site-policy/ |
| 個人情報 | プライバシーポリシー https://www.pdrc.keio.ac.jp/privacy-policy/ |
| ライセンス／著作権 | データ収集ポリシーについては作成中 |
| 寄託データの扱い | データ収集ポリシーについては作成中 |
| 二次利用の対応 | 利用資格・利用要件あり https://www.pdrc.keio.ac.jp/paneldata/howto/ |
| 成果物の提供の在り方 | データユーザーに対しては、成果物に関する情報登録を促しているが、任意であるため、未登録のままのユーザーもいる。 |
| 利用登録の有無 | データ利用の際には、データ申請システムのアカウント登録が必要 |
| 想定利用者 | 国内外の大学または国公立・民間研究機関に所属する研究者・大学院生・学部生（研究生・研究員等も含む） |
| データの質の担保と多様性の確保 | 【質の担保】 データクリーニング、合成変数の提供、サンプルバイアスを補正するウエイトの提供、データのパネル化プログラムの提供など 【多様性の担保】 データやメタデータの英語化、DOIの付与 |
| 公開・共有データを活用した研究を増やすための取組み事例・提案 | <ul style="list-style-type: none"> ・ 学生向けの論文コンテスト（JHPS Award） ・ ディスカッション・ペーパーの発行 ・ 海外のデータアーカイブへの提供（LIS, CNEF） ・ 海外の研究機関との共同研究の実施 |
| データアーカイブ運営上の課題 | — |

（石井 加代子）