

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 活動報告

菊池 信彦（国文学研究資料館）
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）
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）

2022/11/11



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。

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）

菊池 信彦（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）

これまでの活動状況

- 小委員会組織後、1か月半～隔月の頻度で、委員所属先の大学・研究機関におけるデータ管理の状況や実態、課題を報告し、共有を行った。
 - 現在は、各機関における状況を事例集としてまとめる作業を続けている。当初予定では、2022年9月で活動終了予定だったが、半年間の延長中
- * 委員会期間中に委員の異動等の理由で、報告機関の活動にすでにかかわっていないケースがある。そのため、掲載できないものや、「表に出ている」情報を基に記述する方針としているものもある。

成果報告書 各機関事例 記載項目

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

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 利用登録の有無	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。 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庫データベースは利用に際し登録が必要となっている。

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。
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。
個人情報	<ul style="list-style-type: none"> ・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。

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
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の分野、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>
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>
利用登録の有無	<p>必要</p> <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。</p> <p>（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、継続的なアクセス性を担保している。</p>
データの質の担保と多様性の確保	

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
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
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

成果物の意義

- 「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。

2022年度末には小委員会活動の成果物として公表予定。