

研究資源共有化システム ニュースレター



大学共同利用機関法人 人間文化研究機構 高度連携情報技術委員会

Inter-University Research Institute Corporation National Institutes for the Humanities <http://www.nihu.jp>

【巻頭言】 人間文化研究機構の情報関連事業について将来構想の検討

人間文化研究機構 永村 眞

人間文化研究機構では、6機関（国立歴史民俗博物館、国文学研究資料館、国立国語研究所、国際日本文化研究センター、総合地球環境学研究所、国立民族学博物館）によって長年にわたり蓄積されたデータ群を、より有効に活用しようとの目的で、様々な取組を重ねてきました。その事業の中核は高度連携システム事業の一環としての nihuINT 事業であり、その事業を支える処理システムとして nihuINT を開発し、機能の高度化を図ってきました。

この nihuINT のもとで機構に蓄積されたデータ群により、ジャパンサーチへのデータ提供や国立国会図書館との連携、さらには海外ユーザーによるデータ活用の便宜のための国際リンク集の充実など、多様なデータ利用を実現しています。

しかし大学共同利用機関法人としての本機構が、特に情報関係の事業分野で、その役割を十分に果たすためには、6機関が蓄積したデータ群のより高度な利用を実現することにとどまらず、業務の間口を広げ、より多くの大学による情報事業と連携する事業に着手する必要があります。

そこで、総合情報発信センターのもとに「高度連携システム等将来構想ワーキンググループ」を設置し、その検討結果をうけ、その第三期後半から第四期にわたる情報関連事業の将来構想として、今年7月に「人間文化研究機構の情報関連事業第三期後半における基本方針」を作成しました。この基本方針のもとで、「第三期後半における人間文化研究機構の情報関連事業について、第四期におけるあり方を見据え」て、以下の基本目標を定めました。

1. 機構を構成する各機関が果たすべき大学共同利用の拠点としての役割を促進する存在として、機構の事業を位置付ける。
 - これにより、各機関のミッションに沿った情報の提供を実現し、機関間の連携の促進など、こ

れまで以上に実質的な事業展開を図る。

2. 各機関が持つ研究資源や知識の連携により、新たな研究データの構築、それらの効果的な運用や長期保存のための環境整備を進める。

- これにより、人文学の研究に欠かすことのできない情報基盤を国内外に提供し、日本における、もしくは日本を対象とする人文学のより一層の高度化を推し進める。

3. 現行の高度連携事業（nihuINT 事業）がになうシステム・事業を展開しつつ、各機関が持つ研究資源の利活用をさらに促進するための役割・機能を検討する。

- これにより、より効果的で筋肉質なシステムとデータの提供が可能となるとともに、シンプルでユーザーの要望に即したデータインフラストラクチャーを構築する。

この基本方針では、6機関の研究活動のなかで生まれた貴重なデータ群の有効活用を図り、機構は調整機能を果たすため、nihuINT を中核とするデータインフラの構築の進展を図るとともに、他の機関・大学・個人とのデータ連携の推進を掲げています。そして基本方針のもとで、将来的には人間文化研究機構が、広く日本の人文学全般にわたるデータインフラの構築など、情報処理事業の中で大きな役割を果たすことが求められるといえましょう。



ジャパンサーチとの連携を開始

人間文化研究機構 大内 英範

2019年2月、「ジャパンサーチ」(*1)が一般向けに公開されました。ジャパンサーチとは、「書籍等分野、文化財分野、メディア芸術分野など、さまざまな分野のデジタルアーカイブと連携して、我が国が保有する多様なコンテンツのメタデータをまとめて検索できる『国の分野横断統合ポータル』」(*2)のことです。正確には「ジャパンサーチ (BETA)」であり、一般向け試験公開というステータスではありますが、21機関から51データベース、1,800万件以上のデータを収集し(*3)、検索に供している、巨大なポータルサイトです。その運用は、内閣府に置かれた「デジタルアーカイブジャパン推進委員会及び実務者検討委員会」(*4)の方針の下で、国立国会図書館が担っています。



図1 ジャパンサーチ

ジャパンサーチは、原則として実際にデータを所有するさまざまなアーカイブ機関から、分野別の「つなぎ役」を介してデータを収集することになっています。人間文化研究機構は、その人文学分野での「つなぎ役」となり、まずは nihuINT に搭載されている、機構内機関のDBの中から、機関において提供可と判断したものについて、提供を始めました。

2019年11月現在、nihuINTからジャパンサーチに提供し、公開されているDBは以下の通りです。なお一部のDBについては、nihuINTのメタデータではなく、nihuINTに登録された原データでの提供である場合もあります。また、国立民族学博物館は nihuINT からのデータ提供ではなく、同館から直接ジャパンサーチにデータを登録しています。

■国際日本文化研究センター

宗田文庫図版資料／ちりめん本／平安京都名所図会／季語検索／在外日本美術／日本関係欧文図書目録／日文研オープンアクセス

■国文学研究資料館

吾妻鏡／絵入源氏物語／二十一代集／コーニツキー版欧州所在日本古書総合目録／古筆切所取情報／新奈良絵本／歴史人物画像／近代文献情報(明治期出版広告)／連歌・演能・雅楽／歴史物語／日本実業史博物館コレクション／蔵書印

■国立国語研究所

『方言文法全国地図』地図画像／『日本言語地図』地図画像／米国議会図書館本源氏物語翻字本文／国立国語研究所学術情報リポジトリ

■国立民族学博物館

標本資料目録／標本資料記事索引／映像資料目録／ビデオテーク／音響資料目録／みんぱくりポジトリ／梅棹忠夫著作目録(1934～)

nihuINTのメタデータを提供したDBに関しては、原則として検索結果から nihuINT のデータへのリンクが表示されます。nihuINT から原データへのリンクをクリックすれば、当然、原データを見ることができます。

ジャパンサーチでは、Web上での検索のほか、APIによってさまざまな形での利活用を可能としています。中心的に利用されると想定されるのが SPARQL API ですが、簡易的な Web API 機能によるデータ取得もできるようになっています(*5)。

ジャパンサーチとの連携によって、nihuINT が持つ日本の人文学に関する多種多様なデータを、他のさまざまなアーカイブ機関のデータと連携させて利用できるようになりました。もちろん、ジャパンサーチでデータを見つけた利用者がリンクを辿ってきたり、関連情報を直接探しにくることで、nihuINT や各機関の利用増も見込まれます。

なお、上記DBのほか、2019年度中に提供・公開DBの追加を行なうための準備を進めており、連携はさらに広がる予定です。

*1 <https://jpsearch.go.jp/>

*2 「ジャパンサーチとは？」(<https://jpsearch.go.jp/about>)による。2019年11月16日参照。

*3 「現在のデータ」(<https://jpsearch.go.jp/stats>)による。2019年12月4日参照。

*4 https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/index.htm

*5 詳細は「開発者向け情報」(<https://jpsearch.go.jp/api>)からのリンクを参照のこと。



図2 ジャパンサーチから nihuINT、nihuINT から機関の原DBへの遷移例

国立国語研究所のデータベース・データセット

国立国語研究所 高田 智和

国立国語研究所は、コーパスを始めとして、言語研究資料のデータベース・データセットを公開しています。国立国語研究所 web サイトの「データベース」(https://www.ninjal.ac.jp/database/) (図1) では、一覧、研究目的別 (現代日本語/日本語教育・日本語学習/方言と言語の多様性/日本語史/研究文献/その他)、データ種別 (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/その他データ) の3種分類によって、全96種のデータベース・データセットへアクセスすることができます (2019年11月現在)。

国立国語研究所は、70年前の研究所創設から実施してきた調査研究の成果を、オンライン検索が可能なデータベース、あるいは、ダウンロード型のデータセットとして提供しています。特に後者の提供形態がこれまでは主流で、データ利用者が自身のパソコンでデータを扱えることを重視してきました。現代風に言うと、検証可能なデータ利用環境の実現を実践してきました。



図1

国立国語研究所は、データベース・データセットの基になった調査研究の資料も保存しています。資料群の内容は「国立国語研究所研究資料室収蔵資料」(https://rnr.ninjal.ac.jp/) (図2) で公開しています。方言を始めとする話し言葉や音声の調査であれば、話者にインタビューをして調査票 (図3) や書き起こしを作り、分析のための情報カードを作成します。近年はインタビューの録音や録画を収録します。書き言葉や文字も同様で、新聞や雑誌の用語用字調査では、語彙カードや集計表を作成します。これらが、データベース・データセットの基になった調査研究の資料です。

録音や録画は、かつてはオープンリール、カセットテープ、8mmフィルム、ビデオテープなど、さまざまな記憶媒体で保存していました (図4)。記憶媒体は経年劣化を起し、再生用機材も生産中止により再生困難になったため、現在はデジタル化を進めてい

ます。保存継承と再利用のためのデジタル化です。デジタル化音声・映像は、研究教育目的に限り、来館利用の形で提供しています。50年前の話し言葉は、今現在の採取が不可能なので、貴重な言語研究資料となっています。

電子化されたデータベース・データセットと、その基である紙資料・メディア資料を包括した保存と提供を、国立国語研究所は今後も続けて参ります。

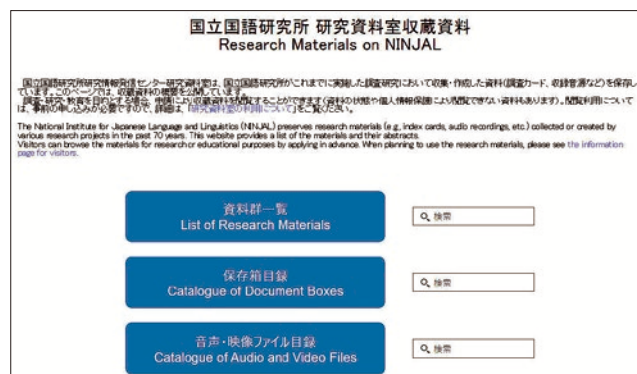


図2

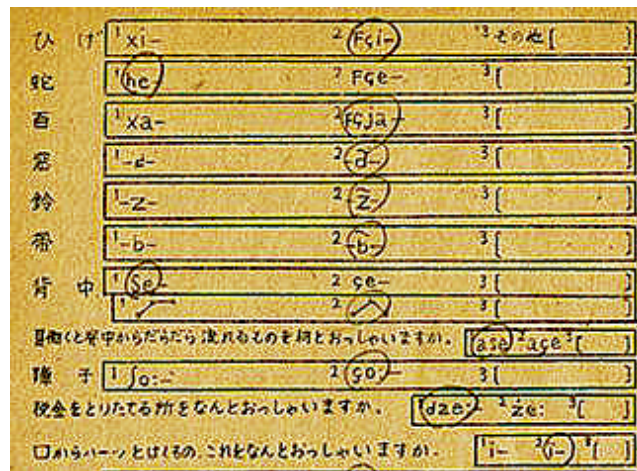


図3



図4

第14回 人間文化研究情報資源共有化研究会
人文科学におけるオープンサイエンスの課題

日時 2019年2月15日(金) 13時30分
会場 大阪大学中之島センター (507 講義室)

〈開催要旨〉

人間文化研究機構において第3期の nihuINT が公開されて、2年が経過しました。人文科学の研究データを取り巻く状況は、データ基盤の整備状況や、オープンデータの動向、人文科学の「危機」といった議論を含め、大きく変化しつつあります。そのような状況の中で、人間文化研究機構が持つべきデータ、持つべきシステムとはどのようなものであるべきか、次のステップのための議論をさらに深めなければなりません。本研究会では、nihuINT のありようそのものから含め、「ゼロベース」での議論を行い、人間文化研究機構のみならず、人文科学が持つデータの形式・保持方法・利用方法などについて、様々な見地から検討しました。

〈プログラム〉

- 趣旨説明 人間文化研究機構 永村 眞
- 「nihuINT の「後」とは何か」 人間文化研究機構 大内 英範
- 「関西大学 KU-ORCAS のデジタルアーカイブ構築とそのデータ利活用に向けて」
関西大学アジアオープンリサーチセンター 菊池 信彦
- 「歴史資料の共有化とデータ連携・利活用」
東京大学史料編纂所 榎原 雅治
- 「商用による情報提供と公的な情報提供の関係、未来」
株式会社ネットアドバンス 田中 政司
- パネルディスカッション
「人文学データの今後と人間文化研究機構が目指す情報基盤」 司会 国際日本文化研究センター 関野 樹



CONTENTS

- 人間文化研究機構の情報関連事業について将来構想の検討 永村 眞・1
ジャパンサーチとの連携を開始 大内 英範・2
国立国語研究所のデータベース・データセット 高田 智和・3
第14回、第15回人間文化研究情報資源共有化研究会プログラム・4

第15回 人間文化研究情報資源共有化研究会
人文系研究データの生成と管理
—「可逆性」の実現のために—

日時 2020年1月25日(土) 10時10分
会場 日比谷図書文化館 4階スタジオプラス

(小ホール)

〈開催要旨〉

研究不正防止を主目的としたガイドラインにより、研究者は研究ノート等の研究データを作成して適切に保管・開示すること、研究機関はそれを研究者にさせることが求められています。しかし、これからは不正を防ぐために保管するだけでなく、研究データを公開・共有して新たな研究資源とすることで、次の研究を生み出すことが期待されています。ただ、人文系の中には、きわめて限られた研究者だけが閲覧を許される写本や、保管自体が難しい音声や一次資料とするものなどがあります。そうした研究では、論文から一次資料に戻ることができる「可逆性」の実現が容易ではないことが考えられます。人文系研究データ生成の際の問題点とはどのようなもので、どのように解決されるべきか、研究データによる「可逆性」の実現をテーマに議論します。

〈プログラム〉

- 趣旨説明 人間文化研究機構 大内 英範
- 「人文系研究データのライフサイクルを考える」
京都大学学術研究支援室 天野絵里子
- 「日本古典文学研究における「可逆性」とは」
慶應義塾大学附属研究所道文庫 佐々木孝浩
- 「話し言葉調査の「可逆性」」 国立国語研究所 高田 智和
- パネルディスカッション
「人文系研究データの生成と管理」
司会 国際日本文化研究センター 関野 樹

〈申込み先〉(当日受付あり)

大学共同利用機関法人人間文化研究機構
本部事務局センター事務室情報発信係
Tel. 03-6402-9234
E-mail. cip-office@nihu.jp



研究資源共有化システム ニュースレター 第13号

発行日 2020(令和2)年1月10日
発行・編集 大学共同利用機関法人人間文化研究機構
高度連携情報技術委員会

〒105-0001 東京都港区虎ノ門4-3-13 ヒューリック神谷町ビル2階
TEL. 03-6402-9200 (代表)