

国立国語研究所学術情報リポジトリ

日本語研究資料の整備と公開： 国立国語研究所研究資料室の取組み

メタデータ	言語: Japanese 出版者: 公開日: 2016-01-29 キーワード (Ja): キーワード (En): archives, EAD, RDBMS, audio-visual data, digital data 作成者: 寺島, 宏貴, TERASHIMA, Hirotaka メールアドレス: 所属:
URL	https://doi.org/10.15084/00000817

日本語研究資料の整備と公開 ——国立国語研究所研究資料室の取組み——

寺島宏貴

国立国語研究所 研究情報資料センター 非常勤研究員 [–2015.08]

要旨

国立国語研究所は戦後の開所以来、日本語研究に関する様々な資料を生成・蓄積してきた。本稿は、ことば資料のアーカイブズ構築に向けた、国語研・研究情報資料センター研究資料室の取組みについて述べる。最初に、所内の各部署から移管された紙資料、また視聴覚・電子メディア資料がそれぞれ専用の保管庫に（紙資料は資料群ごとに、メディアは媒体ごとに）収蔵されていることを紹介する。その上で、資料群の階層構造を記述すべく以前採用されていたデータ形式であるEAD（資料群階層をマークアップ言語によって表現するための国際規格）の運用に際して、記述項目・記述量の多さや、既存の紙・メディア目録との関連付けが行なわれていなかったことを示す。次いで、EADによる記述項目を基に資料群全体を把握しやすくすべく再構成された「研究資料室収蔵データベース」によって、概要情報と既存目録とがリンクした点を説明する。最後に、新しい移管体制に向けた取組みとして移管票の改訂（申込書形式による組織間での資料受渡し—受入れの明示）、また検討課題として電子データの媒体変換が必要である点について述べる*。

キーワード：アーカイブズ、EAD、リレーショナルデータベース、視聴覚データ、電子データ

1. はじめに

1948（昭和23）年12月に開所した国立国語研究所（以下国語研）は70年近くにわたり方言や語彙、音声、日本語教育、言語コーパスといった多様な研究プロジェクトを積み重ねてきた（国立国語研究所 1949–1994, 2000–2006, 2014）。過去に実施された調査研究では、視聴覚・電子メディアを含む膨大な研究資料が生み出された。しかし、プロジェクトの成果である学術論文や刊行物は別として、国語研の所蔵資料が研究所の内外で利活用される機会は少ない。

本稿では言語学ないしは日本語学という、ことばを扱う領域にとっての資料所蔵施設＝アーカイブズはいかなる形とすべきかとの課題を念頭に、国語研・研究情報資料センター研究資料室でこれまで実施してきた取組みについて述べる。次節で国語研所蔵資料の概略を紹介したのち、国語研において過去に実施された研究資料の移管、および資料目録整備とともに、目録のデータ形式が抱えた問題点に言及したい。その際、資料移管時の過誤や、後述のEAD化に伴う資料群階層のXML記述における問題を扱う。

最後に、資料受入れのための新たな体制に向けた取組みについて説明する。まず研究資料のデータ

* 本稿は、人文科学とコンピュータシンポジウム「じんもんこん2014」（2014年12月13・14日、於国立情報学研究所／一橋講堂）でのポスター発表である寺島宏貴・星野雅英・高田智和「ことばの研究資料アーカイブ（ママ）—国立国語研究所所蔵資料の利活用に向けて」、ならびに第121回NINJALサロンにおいて発表した寺島・星野（2015）を改訂したものである。

タ管理について、上記の EAD を登載したデータベースに代わり新たに所内公開された「研究資料室収蔵データベース」について解説する。さらに移管体制の見直しや閲覧環境の整備によって外部（ないしは他の研究領域）への利用提供を図るべき点、さらに電子媒体を含む各種メディアの管理に関して述べることとしたい。

2. 中央資料庫・中央メディア保管庫所蔵資料について

2.1 紙資料

国語研の所蔵資料は、主として紙資料と各種のメディア資料とに大別される。収蔵庫には過去の研究プロジェクトにおいて使用してきた機器類も保管されているが、本節では紙とメディアのみについて述べる（図1）。

まず紙資料の一例として、社会調査において作成された調査票の原本やその集計・分析資料、語彙調査の対象となった学校教科書や雑誌といった資料現物が保管されている。

また、各研究プロジェクトの遂行にあたって作成された公文書等の事務書類も散見される。これらは文書箱に収納し、研究プロジェクトごとにグルーピングした上で、紙資料専用の中央資料庫に保管している。同資料庫の気温は、資料保存の観点から 23°C に保たれている。

現時点での文書箱の総数は約 4,000 箱程度であるが、今後も所内各部署からの資料受入れによって収蔵量の増加が見込まれる。これに加え、方言や語彙の調査で作成された推計約 1,000 万枚のカード類が木製・ステンレス製のカードケースに、同じく地図資料も専用ケースに保管されている。以上に述べた紙資料の中には、昭和 20 年代の社会調査・語彙調査に関する貴重書類もみられる。

2.2 各種メディア資料

次にメディア資料については、各種媒体に記録された調査・分析データが多くを占める。中でも、過去の言語調査において録音・採集された音声が膨大にあり、特にカセットテープ（以下 ct）、DAT テープ（以下 dt）に記録されたものが圧倒的多数にのぼる。

約 50,000 本の ct、dt には、社会調査での外国人を含むインフォーマントによる会話・談話を始めとして、語彙調査等のデータ採集用に作成されたテレビ・ラジオ録音、また国語研に来所した著名人の講演録音がみられる。なお国語研が各地で開催した研究会や、国際シンポジウムの録音記録も残っている。

この他にも CD、DVD、MD、オープンリールテープ（以下 os）、オープンリールビデオ、8mm フィルム、β・VHS ビデオカセットテープと、媒体は多様である。また音声・映像ばかりではなく、調査研究の過程で作成された磁気テープ、フロッピー・MO ディスクといったものも含まれる。

以上に紹介した各種メディアは中央メディア保管庫に収納し、媒体ごとに専用のメディアケースに収納されている。紙資料と同様、室内温度は 23°C である。



図1 紙およびメディア資料

3. 研究資料データベースについて

3.1 これまでのデータ形式

次に、過去に行なわれてきた研究資料の目録化と、その問題点を説明する。ここでいう目録化とは資料群概要を示すための XML 記述の作成すなわち EAD 化を指し、EAD とは別に存在する Excel 形式の紙・メディア資料目録である「配架リスト」・「メディア ID 一覧」とは異なる（これら Excel 表については 4.2 で詳述する）。

国語研では下記の流れで、資料の目録化を図ってきた。

- ① 移管者の側で資料の移管票を作成し、資料受取者へ提出。
- ② 資料の受入れおよび燻蒸。
- ③ 燻蒸後の紙資料・各種メディア資料をそれぞれの保管庫に移管。
- ④ 移管票の情報や資料現物をもとに、EAD によって、各資料群概要の XML 記述を作成。

このうち④の EAD (Encoded Archival Description アーカイブ符号化記述または符号化永久保存記録記述) について青山 (2002: 270) ならびに五島 (2003: 268) によって確認しておくと、EAD は XML 一般ではなく、電子的検索手段の特性に配慮された文書記述言語 SGML (Standard Generalized Markup Language) サブセットの国際規格であり、アメリカのアーキビストによって 1998 年 9 月に公開された。欧米のアーカイブズにおけるデファクトの検索手段であり、SGML によって記録史料の全体像を表現するためのものである。

EAD は、1994 年に国際アーカイブズ評議会が制定した国際標準である ISAD(G) 等の記述要素を取り入れ、また英米図書マーク規則 (USMARC)、ダブリンコア (Dublin Core) との整合性を持たせて機能する¹。それらをコンピュータ上の検索手段として表現するのに必要な各種タグを規定した EAD のメリットとして、アーカイブズの群や小群からなる階層的な秩序を反映できることが挙げられる (五島 2003: 269, 坂口 2010: 387)²。

日本のアーカイブズへの導入は、国文学研究資料館の附置組織であった史料館において 2001 年後半から 2002 年 4 月にかけて実施された。XML で書かれた EAD データ全文を表示する XSLT スタイルシートの表示が成功した 2002 年 5 月に、アーカイブズ記述ないし検索手段の基本形として、日本初の EAD 適用事例が公表された (五島 (2007: 221) による)。日本におけるアーカイブズ学ともども歴史研究者 (例えば山崎 (2003) が挙げられる) を中心として実施された EAD の導入について、五島 (2007: 219) は、史料の各所蔵機関において資料情報が抱える問題—資料群全体を示す概要情報が皆無であり、例えば古文書 1 点ごとの書誌的情報、画像情報、

¹ USMARC について青山 (2002: 270) は「もっとも古い記録史料記述に関する電算化のルールである英米目録規則のなかの」規則であると紹介している。ダブリンコアは永田 (2003: 222) によると、1997 年のネットワーク情報資源の記述に関するワークショップにおいて「ネットワーク情報資源の発見のために基本的なメタデータ」〔筆者注:「データの体系づけられたデータ」のこと (永田 2003: 222)〕要素からなるコア・セットが、開催地の名をとって」呼ばれたものであるといい、また USMARC は現在 MARC21 (MARC:Machine Readable Catalog 機械可読目録) と称する。ダブリンコアの要素リストについては永田 (2003: 223) を参照のこと。

² ISAD(G) (第 1 版) と EAD タグとの対応リストは森本 (2003: 251) が紹介している。

本文情報に偏る一が背景にあったと指摘する。

国語研では、2003年度にEADを採用した資料記述・検索システムの構築に着手し、2005年9月までに一定程度の完成をみている（この点の詳細は森本（2006）を参照のこと）。国語研でのEAD運用に際しては、資料の移管後、配架された資料（紙資料・メディア）に基づいて階層構造の把握を行なった。EAD化にあたり原則一つの研究プロジェクトを1資料群として、プロジェクトの課題・研究内容や組織歴に基づいた階層編成を行ない、XML記述を作成した。

以下、資料群の編成について、各階層の記述項目および記述内容を挙げる（次頁図2も参照のこと）。

(1) フォンド (fond) レベル

共通の出所を持つ資料群の全体（レコードグループ）³。原則、一つの調査・研究プロジェクトに相当する。下記項目の記述を行なった。

- | | |
|--------------------------|-------------------------|
| ・版（第n版） | ・物理的特徴 |
| ・記述名称 | ・編成（階層分け） |
| ・発行者 | ・資料内容 |
| ・記述作成者 | ・評価 |
| ・発行年月日 | ・追加受入れ |
| ・大きさ | ・関連資料（別フォンド等） |
| ・形態／媒体 | ・分離資料 |
| ・表題（調査名、研究プロジェクト名） | ・公開条件 |
| ・要約（資料群の要約） | ・利用／複写条件 |
| ・資料作成年月日（現状では年代のみ） | ・必要機器（記録媒体に関する事項） |
| ・配架位置（中央資料庫、中央メディア保管庫など） | ・他の検索手段 |
| ・管理歴（移管元と移管先） | ・オリジナル資料 |
| ・資料作成者（担当部署と担当者） | ・複製 |
| ・組織歴、履歴 | ・関連出版物（成果刊行物およびメディア等） |
| ・入手情報 | ・資料管理概要（現在の管理状況等） |
| ・数量、形態（紙資料とメディアの数） | ・備考（資料の状態やメディアID等、留意事項） |

(2) シリーズ (series) レベル

シリーズは、内部組織（研究ユニット、もしくは研究プロジェクトのサブテーマ）単位の階層である。シリーズ単位での記述項目は次の通りである（一部）。

³ 全国歴史資料保存利用機関連絡協議会（1997: 93）はフォンドを「同出所史料群 fonds」の見出として「形態や記録媒体にかかわらず同じ出所をもつ資料の総体をいう。出所は特定の個人、家族、組織体などであり、その活動や業務の過程で、有機的に作成、利用、保管されてきたもの。[fondsはフランス語、国際的に広く使われつつある用語であるため、カタカナでフォン、あるいは英語読みしてフォンド、フォンズと表現することもある]」（原文ママ）と説明している。

- ・表題
- ・資料作成年月日（年代のみ）
- ・配架位置
- ・資料作成者（特記すべき場合のみ）
- ・編成（サブシリーズレベル）
- ・備考（メディア ID 等）

(3) アイテム (item) レベル

アイテムは、資料 1 点ごとの最小分割単位である。項目はシリーズレベルより簡略化された。やはり上位レベルとの重複項目が多かったが、アイテムレベルのみ「収納容器」の項目が追加されている。

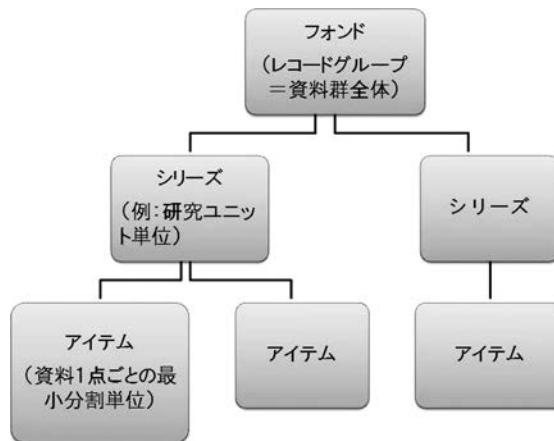


図 2 EAD のツリー構造 (国語研の場合)

3.2 XML 階層記述のデメリット

次に、EAD システムの問題点について述べる。

EAD の導入時には、おそらく多様な資料群に対応すべく記述項目が数多く設けられた。しかし、上記に挙げたように各階層の項目が多すぎ、上位・下位レベルでの重複も多くみられた。どの項目に、どのような情報を記述するかが不明になりがちであり、同じ記述を何度も繰り返すリスクが常につきまとった。また、アイテムについては文書 1 件や各種メディア 1 本ずつというように、大量に登録されるケースもみられた。

これらの問題のため、記述者によって情報量にバラつきが出ることになった。さらに、記述者が下位のシリーズ・アイテムの階層フォルダを自由に増やしたためツリーが細かく枝分かれした (図 3)。場合によってはサブフォンド (subfond) レベル、サブシリーズ (subseries) レベルの記述も行なわれた。各資料群の XML 記述はフォンド—サブフォンド—シリーズ—アイテムのうち 2 ~ 4 階層 (一部 6 階層もあり) から編成され、資料群によってはサブフォンドなし・フォンドのみ、フォンド—シリーズからなる記述もみられた。

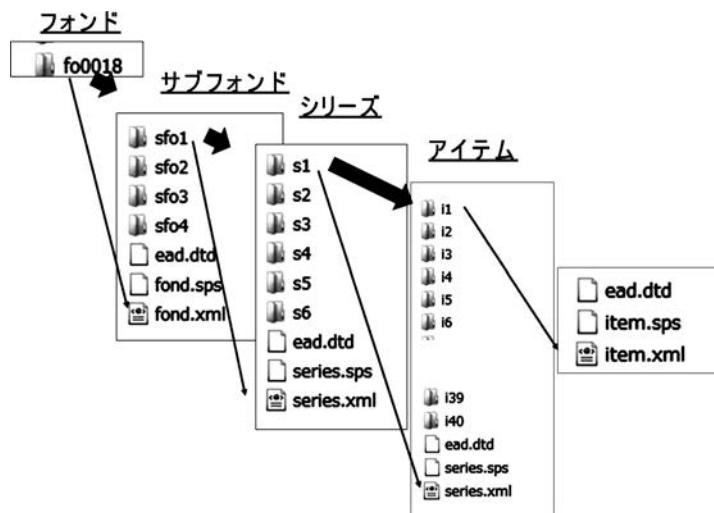


図3 資料群階層 (フォンド～アイテム)

EADシステムの運用では上記のような各レベルの細分化のほか、大量のアイテム（各種メディア）が歴史的古文書の目録を探るように1点ずつ登録されることもあった。ベースとなっている移管票へ過剰に情報を付加したために、移管票よりもはるかに長大な記述が生まれた（次頁図4・5）。各資料群の記述は、概要を示すにもかかわらず概要把握には向かないという問題に陥った⁴。

また紙・メディア資料については先に触れたように、移管票のほか、それぞれ表形式の目録が存在する（4.2参照のこと）。しかし、EAD記述はこれらの情報とは全く別個に作成され、既存の紙・メディア目録と直接には結びつかないデータとして運用された。あるフォンドに移管資料が追加された場合、EAD、紙・メディア目録の両方へ情報を記述する必要があった。紙・メディア目録上の棚位置を示す番号やメディアIDについても、逐一XML形式に書き直し手間が生じたのである。

以前、EAD記述へのアクセスのため国語研ウェブサイトにおいてEAD検索システムが公開されていた。同システムの利用者は、キーワード検索等からEAD公開画面を表示させ、各フォンドのシリーズ・アイテム記述から請求資料を特定し、閲覧請求を行なう。請求を受けた資料管理者は、紙・メディア目録をもとに出納を行なった。とはいっても、すでに述べたようにEADの記述量はフォンドによってまちまちであり、資料請求のための目録としては心許ないものであった。

このような事情から、XML形式によるEADに代わり、概要情報と紙・メディア目録とがリンクしたデータベースを新たに構築し、これに伴う資料移管体制の見直しが大きな課題となつたのである。新データベースの話に移る前にまず、新たな移管体制に向け所内で実施した基礎作業について以下で述べておく。

⁴ のちに記述の簡略化を目的として、サブフォンド、サブシリーズ、アイテムの階層は原則使用しないこととなった。国語研におけるEADの運用は総じて実験性の高いものであったが、しかし記述量が膨らむほど（=何でも書いてある）多角的な検索に耐えうる、との見方もできよう。

図4 EAD公開画面 (f00020: フォンドーシリーズーアイテム)

図5 EAD公開画面 (fc00020:サブサブシリーズ-アイテム)

4. 新たな資料受入れ体制に向けて

4.1 資料庫総点検の実施

これまで、各資料群の EAD 記述は国語研ウェブサイトで公開されていた。しかし積極的な発信に繋がらず、国語研の所蔵資料にどんなものがあるかは資料管理者のみぞ知る、という状態となっていたことは否めない。実際のところ、所内の職員や所外の OB・OG を中心に、EAD 公開画面を通じた閲覧請求は年に数件であった。

同時に、利用者のための閲覧室も完備しておらず、所内の研究図書室と比較しても研究資料の利用に適した環境の充実を図っていなかった。中央資料庫・中央メディア保管庫や、配架リスト・メディア一覧の存在すら所内で周知されているとは言い難い。

2013（平成 25）年 10 月以降、新たな資料管理者に交替してから、利用者が研究資料にアクセスしやすい環境の整備を行なっている。まず、中央資料庫・中央メディア保管庫の総点検（棚卸作業）を実施した。ただし後者に関しては特に劣化が懸念される、貴重な音声を含んだ ct・dt・os を当面の対象として所在確認を行なった。紙・メディア資料の棚卸を通じて、配架済みであっても紙資料の目録に未入力であった資料の抽出や、移管票提出済み・燻蒸済みにもかかわらず実際は未移管となっていた資料が特定された。フォンド ID が付与されていない資料群も多く、これらに ID を発行した結果、fo0181 でストップしたままであった ID は fo0221 まで增加了。

資料の未移管や未入力、またフォンド ID 未付与の問題は資料管理者の側で発生した、アーカイブズにおけるヒューマンエラーである。ここ数年にわたって、専従の資料管理者を置かなかつた点も要因の一つに数えられよう。その一方、上記に触れたように移管資料はこれまで全点を燻蒸処理していた。しかし、取り扱いに注意を要する貴重書類は今後の移管資料に含まれない点を踏まえ、燻蒸は移管プロセスから外した。

4.2 「配架リスト」の所内公開と「メディア ID 一覧」の整備

資料群に関するデータには、EAD とは別に紙資料の目録である「配架リスト」、メディア資料のそれである「各種メディア ID 一覧」が備わる（次頁図 6・7）。研究情報資料センターは 4.1 で述べた資料庫総点検を経て 2014 年 6 月、所内職員を対象に配架リストを「移管資料群一覧」（次々頁図 8。フォンド ID の一覧）とともに公開した（その際、目録上の個人情報等は除外している）。

配架リストは Excel 形式による箱単位の中央資料庫所蔵物一覧であり、「シリーズ—アイテム」のアイテムに対応し、各箱にフォンド番号・箱番号が付されている。箱内容の記述が詳細すぎるものも少なくない。各種メディア ID 一覧も同じく Excel 形式によって、媒体ごとにリストアップされた中央メディア保管庫所蔵一覧である。これもアイテムに対応するリストであるが、しかしフォンド番号に関連付けられていないメディアが多数みられた。中には、フォンドが明らかであるにもかかわらずフォンド ID が付与されていないものも大量に確認された。フォンド不明または未付与のものも一定数存在し、これらに今後 ID 付けを行なう必要がある。

また、各種メディア ID 一覧に関してはメディア 1 本ごと、すなわちアイテムレベルで ID が振られている（一例として、カセットテープならば ct0001 である）。この一覧ではメディアの種

機器目 録	棚位置	分類 配 架	作成部署	調査名	内容	旧箱ID	新箱ID	箱サイ ズ	移管者	備考
53	○ 045-E-1R	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 富良野 継続 面接調査票 2	北調12 (行動2 研)	fo0001-02				
54	○ 045-E-2L	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 富良野 言語生活調査票 調査員記録簿, アクセント聞き取り票	北調13 (行動2 研)	fo0001-03				
55	○ 045-E-2R	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 富良野 パネル 面接調査票	北調14 (行動2 研)	fo0001-04				パネルの調査票 (北調1回目 富 良野 面接調査 票, 社会生活調査 票)はfo09~
56	○ 045-E-3L	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 札幌 面接調査票 1	北調16 (行動2 研)	fo0001-05				
57	○ 045-E-3R	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 札幌 面接調査票 2	北調17 (行動2 研)	fo0001-06				
58	○ 045-E-4L	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 札幌 面接調査票 3	北調18 (行動2 研)	fo0001-07				
59	○ 045-E-4R	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 札幌 言語生活調査票	北調19 (行動2 研)	fo0001-08				
60	○ 045-E-5L	行動1研, 行動2研	北海道における共通 語化および言語生活 の実態 2回目	北調2回目 札幌 調査員記録簿, アクセント聞き取り票	北調20 (行動2 研)	fo0001-09				
	○ 045-E-5R	行動1研, 行動2研	北海道における共通 語化および言語生活	北調2回目 高校生調査 調査票1 (1~6)	北調21 (行動2	fo0001-10				

図 6 配架リスト

1	棚位置	Reel ID	調査名	タイトル	複製ID	その他ID	移管時箱 ID	備考
2	10-A-1L	os00001	話研録音資料(一般)	ラジオことばの研究室(1) 「易者の ことば」「デパートめぐり(ほか)	dt00391 dt00765	1	os-B-001	
3	10-A-1L	os00002	話研録音資料(一般)	ラジオことばの研究室(2) 「就職試 験」「女の社交場」ほか	dt00392 dt00393 dt00766 dt00767	2	os-B-001	
4	10-A-1L	os00003	話研録音資料(一般)	3人の女性	dt00394 dt00395 dt00768 dt00769	3	os-B-001	
5	10-A-1L	os00004	話研録音資料(一般)	九段高校生	dt00396 dt00770	4	os-B-001	
6	10-A-1L	os00005	話研録音資料(一般)	ラジオ新聞人20の扉	dt00397 dt00771	5	os-B-001	
7	10-A-1L	os00006	話研録音資料(一般)	ラジオ歌手20の扉	dt00398 dt00772	6	os-B-001	
8	10-A-1L	os00007	話研録音資料(一般)	ラジオ私は誰でしょう	dt00399 dt00773	7	os-B-001	
9	10-A-1L	os00008	話研録音資料(一般)	研究室の電話(1)	dt00400 dt00774	8	os-B-001	
10	10-A-1L	os00009	話研録音資料(一般)	石野家雑談	dt00401 dt00775	9	os-B-001	
11	10-A-1L	os00010	話研録音資料(一般)	3人の青年	dt00402 dt00776	10	os-B-001	
12	10-A-1L	os00011	話研録音資料(一般)	ラジオ街頭録音	dt00403 dt00777	11	os-B-001	
13	10-A-1L	os00012	話研録音資料(一般)	職安女子部	dt00404 dt00778	12	os-B-001	
14	10-A-1L	os00013	話研録音資料(一般)	職安男子部	dt00405 dt00779	13	os-B-001	
15	10-A-1L	os00014	話研録音資料(一般)	絵画館のおばさん	dt00406 dt00780	14	os-B-001	
16	10-A-1L	os00015	話研録音資料(一般)	芸能界	dt00407	15	os-B-001	
	openreel-sound	cassette DAT openreel-video	16mm 8mm 8mm VHS β Umatic	DV DV-CAM MD CD				

図 7 各種メディア ID 一覧 (シート分割前の os リスト)

別ごとにシートが分かれており、シートそれぞれの入力項目が異なる。

しかし、このメディア ID 一覧はフォンド ID と厳密な対応づけがなされていたわけではなかった。例えば音声資料に関して、フォンド ID と未対照のまま放置されていたものが多数みられた。

分類	資料群No.	研究課題等	移管件数
語彙表・ 語彙調査	fo0060	『分類語彙表: 増補改訂版』の作成	17
文法	fo0061	話しことばの文法の調査研究	7
文法	fo0062	話しことば研究室の資料	5
文法	fo0063	マス・コミュニケーションの研究: 委託研究の資料	1
計量	fo0064	大量語彙調査機械化のための準備的研究	1
文字・表記	fo0065	漢字の字体に関する基礎的研究: 中国の文字改革	38
語彙研究	fo0066	動詞・形容詞等の意味・用法の記述的研究	2
文字・表記	fo0067	現代語の表記法に関する研究: 送りがな・漢字	15
近代語	fo0068	明治初期の漢語の研究: 『明治初期漢語辞書』	2
計量	fo0069	情報化社会における言語の標準化: 日本語の正書法及び造語法とのあり方	2
計量	fo0070	言語処理の高度化(複合語・日英)	4
計量	fo0071	学術用語の語構成の研究	4
語彙研究	fo0072	雑誌用語の変遷に関する研究	6
日本語教育	fo0073	日本語教育研修会の実施	13
日本語教育	fo0074	日本語教育長期専門研修の実施	72
日本語教育	fo0075	日本語教育夏季研修の実施	34

図 8 移管資料群一覧 (fo0060 ~ fo0075)

これについて、まず各種メディア ID 一覧の図にあるような媒体ごとのシートを 1 ファイルずつ分割した上で、特に ct・dt、また os の一覧にデータクリーニングを施し、リスト内の複製情報 (ct または os から dt への) の抽出を行なった。その上で対照作業を行なった結果、フォンド不明分の音声を除くとおよそ 8 割方の ct・dt・os にフォンド ID が付与された。今後、他メディアについても順次データクリーニング及びフォンド ID 付けの処置を実施していく。

4.3 移管票（旧）の記載項目

次に移管票について説明する（次頁図 9・10）。移管票では、移管者・資料受取者が次の項目に必要事項を記入する。ただし、4.6 に述べるように研究情報資料センターでは今後この移管票は使用しないこととし、新たな資料受入れ申請書のフォーマットを検討中である。

(1) 移管票の記述項目

A. 移管資料の概要

- ① 移管年月日
- ② 移管量（保存箱・メディア箱・カードケース箱ごとに数量を記載）
- ③ 移管者名
- ④ 受取者名
- ⑤ 調査・研究名称
- ⑥ 担当部署・担当者名

資料庫移管資料の概要			
調査ID	移管年月日		
	保存箱	メディア箱	カードケース等
移管量			
移管者名			
受取者名			
調査・研究名称(一般的に研究所で使われている呼称)	実施期間		
正式名称および実施期間			
正式名称	期間		
担当部署および担当期間			
担当部署名	期間		
担当者名			
研究・調査の概要(成果刊行物・関連調査等があれば記載)			
保存・利用にあたっての留意事項(資料の媒体・個人情報の扱いなど)			

図 9 移管票（旧）：移管資料概要

<資料移管リスト記入例>					
1. 資料の場合					
保管番号	内容			所在	
1	中学生漢字検査調査 漢字物語テスト(S40.12) 答案(問題1~3)				
2	中学生漢字検査調査 漢字物語テスト(S40.12) 答案(問題1~6)				
2. 提供資料メディア・電子メディアの場合					
メディア番号	メディア種別	内容	数量	備考	所在
1	オーブンリール	「語ごとにば研究室」(音源資料(一般の部) No.1~10)	10本		
1	DAT	「語ごとにば研究室」(音源資料(一般の部) 話題1~話題10)	10本	同様のオーブンリール をDAT化したが	
1	CD	「語ごとにば研究室」(音源資料(一般の部) 内容)外	1点		
1 一つの箱の中に複数のデータが収録している場合、メディア種別ごとに記入を要するが、箱は一つなので、メディア番号は同じ					
3. カード・その他の場合					
カード番号	形態	内容	数量・単位	備考	所在
1	カード	[A] 品カード	(カードケース数3×1×5)×100枚		
2	地図	[B] 調査地沿線用地図	N枚	マップケースでの保管 が必要	

図 10 移管票（旧）：資料移管リスト記入例

- ⑦ 研究・調査の概要（成果刊行物・関係調査等があれば記載）
- ⑧ 保存・利用にあたっての留意事項（資料の媒体・個人情報の扱い等）
- ⑨ 調査 ID (=fond ID。受取者の側で記載）

B. 資料移管リスト（カード資料を除く）

- ① 保存箱（メディア箱）番号
- ② 内容（メディアの場合、その種別）
- ③ 所在（資料庫内の配架位置）
- ④ 備考（メディアの場合のみ）

上記のように移管票のフォーマットは、移管資料の概要とともに、文書箱 1 ケースずつの情報が記載可能なものとなっている。資料受取者（資料管理者）は、上記 B を配架リストに反映する。

ただし、現在の配架リストの項目は移管票に対応しておらず、下記のようになっている。

(2) 配架リストの項目

- ① 棚位置（配架位置）
- ② 作成部署
- ③ 調査名
- ④ 資料内容（文書箱 1 ケースの内容）
- ⑤ 箱 ID (fond ID + box ID)
- ⑥ 移管者
- ⑦ 備考

4.4 新データベースの考え方

前項で記したように、EAD と紙・メディア目録・移管票の三つは別々に整備されており、ともに関連性の薄い状態で並存してきた。ただ、配架リストの記述レベルに関してはフォンド（レコードグループ）—シリーズ—アイテム（箱単位）と、すでに階層的な記述レベルに分類されている。この記述レベルに対応した記述要素のうち、上記 4.3 の（2）で挙げた配架リストの記述項目のうち①～③がフォンドに相当する。新データベースでのフォンド階層は、検索時にキーワードの組み合わせや絞込ができない場合、目次的な「資料ガイド」の機能を担う。

同じく配架リストの項目④は、シリーズおよびアイテムに当たる。資料ガイドの下位レベルであり、研究組織に基づいた個別の分類を表す。箱単位のレコードのため、④の現状は「○○（シリーズに相当する）関係資料」のような記載があったり、かなり詳細に記載したりするなど精粗がみられる。これは資料現物の状態によるところが大きい。したがって、アイテムは文書 1 件ごとではなく、箱単位で上位レベルと結びつける必要があると思われる（次頁図 11）。

これら各階層のリレーションシップは、片方のテーブルの 1 情報（行）に対して、もう片方では複数の情報が関連付けられた「1 対多」である。検索のためには、1 対多の階層構造で結びついた資料目録を、階層的な画面体系で表現する必要がある（寺島・星野 2015）。

記述 レベル	フォンド(レコードグループ)						
	シリーズ			アイテム			
	資料群		シリーズ・アイテム				
記述 要素	棚位置	作成部署	調査名	内容	箱ID	移管者	備考
	002-C-1R	コーパス開発センター	現代日本語書き言葉均衡コーパスの構築(BCCWJ)	台帳(書籍) BCCWJのサンプリング用台帳。	fo0208-02	○○	
	002-C-2L	〃	〃	〃	fo0208-03	○○	
	002-C-2R	〃	〃	〃	fo0208-04	○○	
	002-C-3L	〃	〃	〃	fo0208-05	○○	
	002-C-3R	〃	〃	〃	fo0208-06	○○	

リレーションシップ

資料ガイド

移管資料(シリーズ・アイテム)

図 11 「配架リスト」に関する情報の構成（大城（2000）所載表を改変）

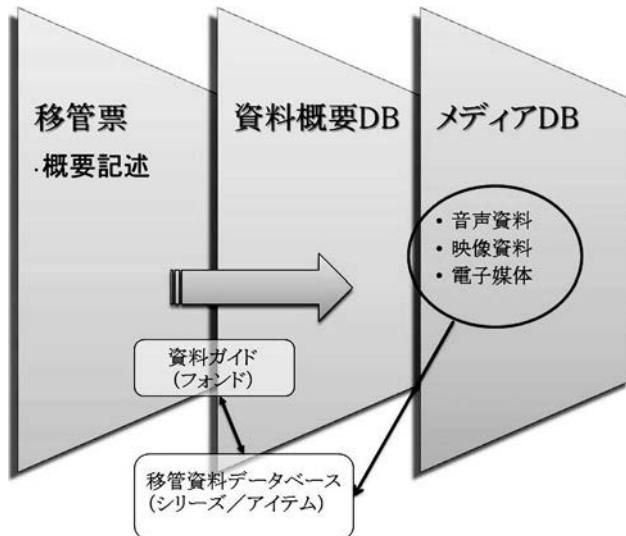


図 12 研究資料室収蔵データベース（概念図）

このように、既存の階層関係を生かしながら、手持ちの移管票・目録を活用したリレーションナルデータベースが研究資料の管理としては適切である（図 12）。

4.5 「研究資料室収蔵データベース」の所内公開

4.4 に述べた点を踏まえ、研究情報資料センターでは、フォンド ID をキーにして概要情報と紙・メディア資料の目録とを関連付けたリレーションナルデータベースを構築した。この「研究資料室収蔵データベース」は 2015 年 7 月に所内公開された（寺島・星野 2015）。

本データベースは、国語研 EAD における大きな問題であった資料群全体のみえにくさの解決を目的とする。これまで別々に存在していた概要情報と資料情報とを関連づけ、全体像をより簡単に、さらに文書箱 1 点レベルと結びつけて把握できるようにしたものである。以下で簡単に本データベースの構築手順を説明し、データベースがどのように構成されたかを示しておく（図 13）。なお検索手順は図 14（次頁～）を参照されたい。

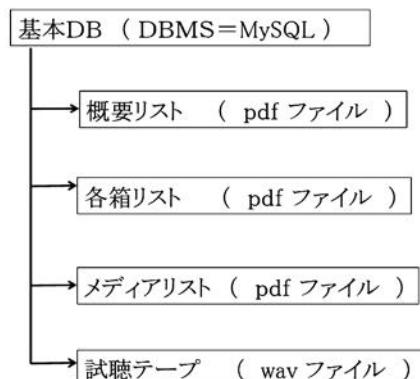
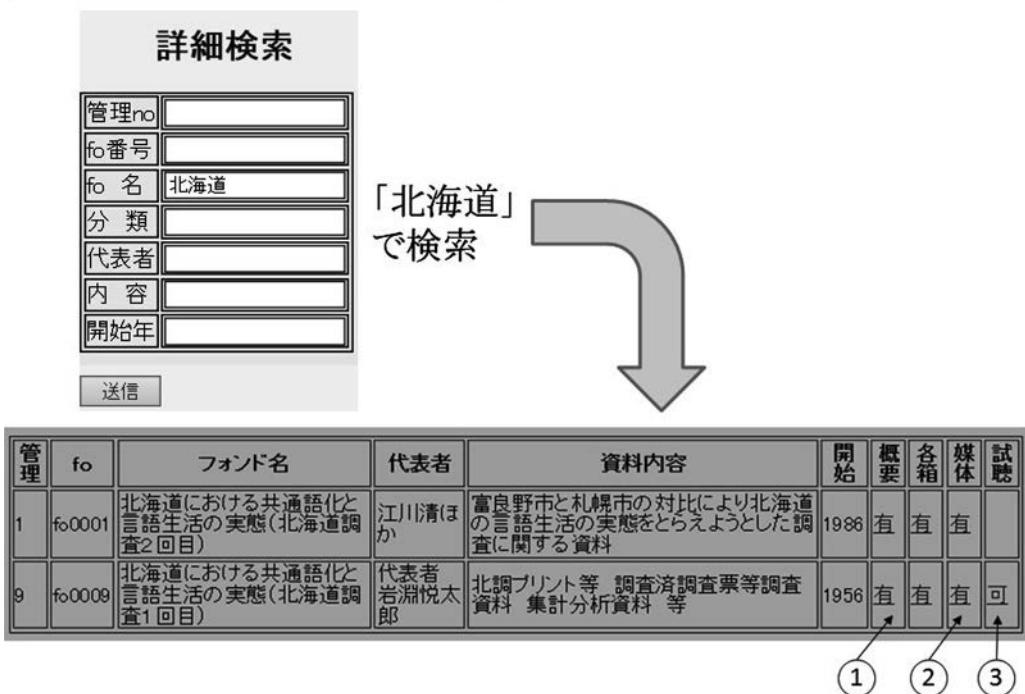


図 13 データベースの構成（寺島・星野 2015）

- ① フォンド ID で管理している資料群については、EAD の第 1 層の記述を使用して、基本的な内容に絞って「概要リスト」を Excel で作成する。これは EAD 記述によって作成した資料概要の記述項目を再選定し、移管票に準じた簡易記述に改めたものである。このリストはフォンド ID ごとに pdf 化して、検索時に表示される。
- ② 次に「概要リスト」から、さらに基本的情報に絞って「基本 DB」を MySQL で構築する。
- ③ Excel で別途作成されていた配架リストをもとに簡略化して、Excel で「各箱リスト」を作成する。さらにフォンド ID ごとに pdf 化し、検索時に表示する。
- ④ Excel で別途作成されていた各種メディア ID 一覧をもとに、dat テープに着目して簡略化し、Excel で「メディアリスト」を作成する。これをフォンド ID ごとに pdf 化し、検索時に表示する。なお、各フォンドに属する音声メディアの一部については今後、検索結果画面において試聴可能としていく（図 14）。

研究資料室 データベース



①「概要」の「有」をクリックすると、概要を表示

管理番号	fo0009
名称	北海道における共通語化と言語生活の実態(北海道調査1回目)
調査年	1956/1986
資料内容	北調プリント等 調査済調査票等調査資料 集計分析資料等
数量	紙資料 22箱, オープンリールテープ 9本(os1436～os1439, os00787～os00791) DATテープ, 一部音声ファイル
配架位置	中央資料庫 045-C-2L～045-D-6R 中央メディア保管庫 10-D-4L, 10-G-5L
管理歴	西が丘庁舎旧図書館、利用後の資料は、第1資料庫および第4資料庫から。移転により合流して、保管庫へ。
録音テープ	4本は野元氏より移管。2006年12月、中央資料庫へ移管。
調査組織	科研費「北海道の言語の実態と共通語化の過程」研究代表者:岩淵悦太郎
調査内容	東京語に近いといわれる「北海道共通語」が、どのようにして成立しつつあるかを明らかにして、日本全国共通語化の方策と、共通語教育の方法をたてるのに有効な知識を得ることにある。
公開条件	被調査者の個人情報は原則として非公開
利用条件	被調査者の個人情報は複写不可
関連資料	fo0001 北海道における共通語化と言語生活の実態 2回目 『昭和33年度国立国語研究所年報10』1959 p41-70 『言語生活』第90号1969-03(「北海道に生まれた共通語」柴田武) 『昭和34年度国立国語研究所年報11』1960 p57-72 『昭和35年度国立国語研究所年報12』1961 p36-45
出版物	国立国語研究所報告27『共通語化の過程－北海道における親子三代のことば－』 『北海道における共通語化と言語生活の実態』(中間報告)1997
記録者	磯部よし子
記録年月	Apr-09

②「媒体」の「有」をクリックすると、メディアリストを表示

一連番号	フォンド	棚位置	Reel ID	タイトル	複製ID
1	fo0009	10-G-5L	os01436	北A:宮城県亘理郡、俱知安町八幡 調査風景	dt06427～dt06430
2	fo0009	10-G-5L	os01437	北B:北海道俱知安町八幡 上野家 のアクセント	dt06430～dt06434
3	fo0009	10-G-5L	os01438	北C:北海道俱知安 上野本家2世 アクセント	dt06435～dt06438
4	fo0009	10-G-5L	os01439	北D:北海道俱知安 上野本家2 世・3世アクセント	dt06439 dt06440
5	fo0009	10-D-4L	os00787	北海道アクセント(1)	dt06115 dt06116
6	fo0009	10-D-4L	os00788	北海道アクセント(2)	dt06117 dt06118
7	fo0009	10-D-4L	os00789	北海道アクセント(3)	dt06119 dt06120
8	fo0009	10-D-4L	os00790	北海道アクセント(4)	dt06121 dt06122
9	fo0009	10-D-4L	os00791	北海道アクセント(全体)	dt06123 dt06124

③「試聴」の「可」をクリックすると、試聴テープを表示



図 14 「研究資料室収蔵データベース」検索画面（寺島・星野 2015 より改変）

4.6 「研究資料受入れ申込書（仮）」の作成

最後に、新たな資料受入れ体制のための準備、特に移管票の見直しについて述べておきたい。4.3 でみたように移管票の項目は、配架リストのそれとの間に齟齬を来している。この状態を解消すべく現在、移管票の改訂作業中である。新たに「研究資料受入れ申込書（仮）」として検討されている主な改訂内容を、下記に示す。

- ① 新しい移管票は上記のように「研究資料受入れ申込書（仮）」とし、作成者・提出先を明確にした、申込書の形式を採用する。上図に掲げた旧移管票の記入項目では、各部署から研究情報資料センターへの移管であることを示す情報が反映できない。旧移管票は資料受取者の側での管理を重視して作成されており、配架リスト・各種メディア ID 一覧とともに、いわば作業資料としての性格が強かった⁵。受入れ申込書では受渡者をプロジェクト

⁵ 今回、第4回鶴岡市における言語調査に関する資料（fo0214）の移管に際し、旧移管票のフォームで受取者名のみ改めた。これまで資料の移管体制が、作成部署と研究情報資料センターとの間の授受として規定されてこなかった点は、同センターないしは国語研全体において、研究資料室の組織的な位置がまったく不明確であったことを意味する。移管体制を再整備した上で、さらに所内講習会を通じて移管体制を周知していく必要もあり、ことば資料のアーカイブズはオープン化とともに内部管理上の課題も抱えると言わねばならない。

リーダー名、受取者を研究情報資料センター長名として、組織間での資料授受である点の明示に重きを置く。

- ② 個人情報の有無について明確化する。受渡者の側で、個人情報を含む資料の「あり・なし」にマルつけできるようなフォーマットとする。
- ③ デジタル情報の受け入れ方法を確立する。新しい受け入れ申込書では、これまでのようにデータの数量ではなく、データ形式を最重視したフォームを作成する。

上記②について、国語研の各資料群は、各種の社会調査において作成されたインフォーマントに関する情報を多く含んでいる（記入済み調査票・フェイスシートといった資料）⁶。個人情報については、各資料群に含まれる公文書にも当てはまる⁷。資料閲覧の際は、所内および所外共同研究員に限定し、部外者の閲覧は職員立会で行なうことが望ましい。

5. おわりに

最後に、以上に述べてきた諸点を踏まえ、これから検討課題について述べる。ことば資料のアーカイブズの構築に際しては、貴重資料は除くとしても、歴史的文書の扱いとは異なった処置を探るべきであろう。さらに、今後は電子データの移管の増大が予想され、その管理は極めて重要な課題である。既に保管してある電子媒体についても、経年劣化によるデータ消失は避けられない。今後の再調査・研究の素材の一つである、デジタル資料の媒体変換（dt からハードディスク等へのコピー）とその定期的な更新を図らなければならない。

次に、今後の「研究資料室収蔵データベース」のあり方についてである。同データベースに追加される概要情報では、新たな「研究資料受け入れ申込書（仮）」に移管者の側で情報をできる限り多く盛り込んでもらい、これを概要記述に代えることが望まれる。ことばの研究資料にとって、概要情報と資料目録とで2層化したデータベース（資料群によっては各種メディアの3層目を立てる）運用が有用であろう。さらに国語研のウェブサイトでは様々なデータベースが公開されており、これらとの連携も模索しなければならない。

国語研所蔵資料を、学術の諸領域において公開利用を図っていくための環境は未だ整備の途上にある。本稿に記したように、国語研の財産というべき様々な日本語研究資料やデータに関しては、その記述形式をにらみつつアーカイブズをいかに構築するかという課題を残したままであった。資料情報と紙・メディアとを相互にリンクさせ、言語学さらには他領域に開かれた形で国語研の知を共有し、遺していくためのアーカイブズが必要となっているのである。

⁶ 個人情報の取扱いについては2015年4月に施行された「国立国語研究所における人を対象とした研究倫理規程」を参照のこと。

⁷ 公文書管理のあり方は現在のような資料庫内の文書保存箱ではなく、別の保存方法を考慮する必要がある。原議書、庶務・会計・契約などの事務書類関係など、あるいは単に保管されているにすぎない資料などは別途管理すること、あるいは評価選別のうえ破棄することが必要であろう。

参照文献

- 青山英幸 (2002) 『記録から記録史料へ—アーカイバル・コントロール論序説』 東京：岩田書院。
- 五島敏芳 (2003) 「アーカイブズ情報の電子化とネットワーク—電子的検索手段の国際規格」 国文学研究資料館 (編) 下, 261–277。
- 五島敏芳 (2007) 「アーカイブズ情報の電子化・保存と共有化の動向」 『情報知識学会誌』 17(4): 217–224。
- 国文学研究資料館 (編) (2003) 『アーカイブズの科学』 (上・下) 東京：柏書房。
- 国立国語研究所 (1949–1994) 『国立国語研究所年報』 1–45。
- 国立国語研究所 (2000–2006) 『国立国語研究所研究活動一覧』。
- 国立国語研究所 (2014) 『国立国語研究所要覧 2014–2015』。
- 森本祥子 (2003) 「アーカイブズの編成と記述—国際動向を中心に」 国文学研究資料館 (編) 下, 236–260。
- 森本祥子 (2006) 「EAD を用いた資料記述システムの開発について—国立国語研究所の事例」 『アーカイブズ学研究』 4: 92–102。
- 永田治樹 (2003) 「アーカイブズ学と図書館情報学」 国文学研究資料館 (編) 上, 219–244。
- 大城博光 (2000) 「公文書目録情報のデータベースモデル—階層構造を持つ目録情報のリレーションナルデータベースでの実装」 『沖縄県公文書館研究紀要』 2: 117–123。
- 坂口貴弘 (2010) 「アーカイブズの編成・記述とメタデータ」 『情報の科学と技術』 60(9): 384–389。
- 寺島宏貴・星野雅英 (2015) 「日本語研究データ利活用のための資料整備」 第 121 回 NINJAL サロン発表, 2015 年 1 月 27 日。
- 山崎圭 (2003) 「アーカイブズの編成と記述—近世史料を中心に」 国文学研究資料館 (編) 上, 199–214。
- 全国歴史資料保存利用機関連絡協議会 (1997) 『文書館用語集』 大阪：大阪大学出版会。

Practical Uses for Research Materials Owned by NINJAL

TERASHIMA Hirotaka

Adjunct Researcher, Center for Research Resources, NINJAL [-2015.08]

Abstract

Since its inception, The National Institute for Japanese Language and Linguistics (NINJAL) has been accumulating research documents and media for linguistics (e.g., phonetic data recordings on cassette and digital audio tapes). This paper aims to explain the contents of these materials owned by NINJAL. A further important point of consideration is creating a database for these abundant materials. I note the advantages of using a Relational Database Management System (RDBMS) for the description of research materials and the contents of previous research projects, which is an improvement on XML as formerly used in our laboratory.

Key words: archives, EAD, RDBMS, audio-visual data, digital data