

国立国語研究所学術情報リポジトリ

バーチャル日本語情報資料館構築に向けて

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 熊谷, 康雄 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002933

バーチャル日本語情報資料館構築に向けて

熊谷康雄

(情報資料研究部第二研究室)

1. はじめに

様々な電子的なメディアやネットワークなど、急速に発展しつつある技術的、社会的な状況・基盤を背景として、「バーチャル日本語情報資料館」構築の計画（平成 13 年度より）を進めている。「バーチャル日本語情報資料館」は、電子化した情報・資料を元に、ネットワーク上に仮想的に実現する「日本語情報資料館」であり、インターネットを通して、国立国語研究所が蓄積する日本語に関する情報・資料を得ることのできるものである。

本発表では、「バーチャル日本語情報資料館」の構築に向けて、これまでの歩みを整理すると共に、来年度以降、本格的に展開する予定の「バーチャル日本語情報資料館」へ向けての構想と試みを紹介することとする。

国立国語研究所は 1948 年の設立以来、日本語および日本人の言語生活に関する科学的な調査研究を行って来た。これらは、漢字、語彙、文法、敬語、方言、音声、言語発達、言語と社会、地域の言語、言語の対照研究、日本語教育等の広い分野にわたる。これらの成果は研究報告、資料集、データ集等の刊行物の形で公刊されており、国語研究所には、ここで得られた様々な資料や研究報告等の文書などが蓄積されている。

これらの研究成果は、日本語に関する基礎的資料として、また、多くの先駆的な研究として、日本語研究の中で重要な位置を占めて来ている。また、長期的な歴史の流れの中に置いてみれば、多くの研究資料は貴重な日本語の歴史的な記録でもある。しかし、時代を超えて、研究の蓄積の中にある情報を埋もれさせることなく、新しい研究の中で有効に活用していくと共に、日本語の基礎的研究資料として後世に伝えていくには、組織的な取り組みが必要である。

そこで、これらの資料・研究成果や現在進行中の研究・事業による資料・研究成果の蓄積と利用を組織的に行うための仕組みを構築し、運用するための仕事をいっそう推進することが必要となっている。これらの情報や資料に対して、物理的にも情動的にも整理・保存・蓄積を組織的に行い、日本語に関する共通の知的な財産・資源として将来へ継承しつつ、活用していくための仕事を進めていきたいと考えている。そのためには、研究成果の組織的な蓄積・保存・管理の体制の整備と、その蓄積全体に関する情報検索・アクセス手段の整備が必要である。

国立国語研究所は平成 17 年度に東京都立川市への移転が予定されており、移転後の新庁舎では、大型の資料庫や一般市民へ向けての展示スペースも予定しており、資料の保管と利用に関しての基盤整備が進む予定でいるが、この移転に先立って、「バーチャル日本語情報資料館（仮称）」（デジタルアーカイブ）を立ち上げ、国立国語研究所の蓄積資料のみならず、研究所外の日本語に関する資料の収集も視野に収めなが

ら、研究成果・資料の保存、利用、普及を推進しようと計画している。日本語に関する研究資料を共通の財産として保存、利用、普及するための仕組み作りの必要性が高まっており、組織的にこれに対応していくことが求められている。

現在計画中の「バーチャル日本語情報資料館システム」は、電子資料館サブシステムおよび電子図書館サブシステムのふたつのサブシステムから構成される。電子資料館サブシステムは電子化資料をネットワークを通じて検索利用できるようにするためのシステムであり、電子図書館サブシステムは電子化した刊行物の本文をネットワークを通じて検索利用できるようにするためのシステムである。このふたつのサブシステムをネットワーク上で、統合的に運用し、日本語情報への総合的なアクセスを可能とするシステム構成とする。大容量記憶装置を中心とし、各種のサーバを配置し、テキストデータ、音声、画像、動画などの各種のデータの蓄積、管理、利用、配信が可能な性能を持ち、ネットワーク通じての情報提供が可能なものとして、システム構成を進めている。

2. 「バーチャル日本語情報資料館」へのあゆみ

2.1 ネットワーク

ネットワークによる情報の提供と言っても、ネットワークそのものが整備されなければ現実のものとはならない。

ネットワークによる情報の共有・提供の基盤としては、研究所では、1995年度には所内ネットワークおよびインターネット接続の利用をスタートしており、学術情報ネットワーク（SINET）経由によるインターネットへの接続を行っている。

1993年 パケット交換網により学術情報ネットワークに加入

文部省学術情報センターのNACSIS-CAT（学術情報センター目録所在情報サービス）に加入し、図書館の所蔵図書 of 遡及入力を開始

1994年 IPアドレス、ドメイン名（kokken.go.jp）を取得

マルチメディアによる研究用システム（音声映像解析システム）の導入

1995年 SINET（学術情報ネットワーク）経由でインターネット接続を開始

メインフレームを中心とした計算機システムの更新の機会に、メインフレームの更新だけでなく、ワークステーションと全所的なネットワークを導入

2000年 全面的にワークステーションのサーバを中心としたシステムに更新

研究所内では、ネットワークは日常的に不可欠の道具となっている。また、平行して、ほぼ同時期に、社会的な現象とも言えるほどのインターネットの急速な普及が進んでおり、ネットワークを基盤とする研究情報の利用ということが現実的なものとなってきた。また、1996年に試験的な運用を開始したホームページは、社会の情報化に添うように国内外からのアクセス件数の伸びを示している。国内外の研究教育機関関係はもちろん、一般市民、公共団体、地方自治体、企業等の各層からのアクセスがある。

2001年度（平成13年度）には「バーチャル日本語情報資料館」システムの導入

により、ハード、ソフト面の整備を進めるとともに、文化庁より「日本語教育支援総合ネットワーク」が国立国語研究所へ移管され、「日本語情報資料館」の一部となる予定である。(なお、今回の発表では「日本語教育支援総合ネットワーク」については、概略のみを示す。)

2.2 文献情報の電子化・文献情報システムの構築

文献情報は研究情報の中でも、基本的な位置を占めるものであることは多言を要しない。紙のカードを使った手作業による文献情報の収集によって『国語年鑑』という日本語研究文献情報に関する基礎的な文献が、1953年より国立国語研究所より継続的に編集・出版されていた。この蓄積があって、『フロッピー版日本語研究情報文献目録』の出版という形で国語学の全分野にわたる電子化された文献情報として1989年に最初の成果が世に出た。そして、そのような背景の中で、国語年鑑の編集システムの機械化という形で、データベースを中心として文献情報のシステムの構築を行ない、また、並行して、研究所の図書館システムの整備を進めてきた。紙のカードを使った文献収集に始まり、既存の情報の電子化、情報の蓄積過程と利用のシステム化、ネットワークの導入と進んできた。紙の目録カードを使った手仕事から、大型計算機、パソコン、インターネットと時代の流れと共に展開し、蓄積してきた情報が様々に利用の可能性の幅を広げてきていると共に、様々な電子的メディアやネットワークという新しい道具立てを手にすることができるようになり、これを土台とした新しい試みができるようになってきている。すでに、インターネットを通じた研究文献目録文の公開は多くの方に利用されている。

2.3 蓄積資料の情報のデータベース化

資料の保管・管理に対して、(1)実体としての資料と(2)その資料に関する情報のふたつの側面からのアプローチが必要である。観点としては、(1)既存の研究成果のストックの組織化、(2)知識の組織化、(3)成果の保存継承、(4)蓄積の利用や公開などである。研究資料・成果といっても、その内容は多岐に渡る。きちんとした形になった報告、論文というものから、中間的な資料やメモ的なもの、調査の原資料等さまざまな内容、形態のものがある。

(A)はっきりとした形になったもの

研究者(研究の実行主体)から離れて、研究が外在化されて独立して存在し、他者が利用できる知的産物(知識、技術、経験)

形態: 報告書, 論文, 資料, データ, プログラム : (外部公開, 内部公開)

媒体: 紙, 電子媒体(&装置, 機器)

(B)はっきりとした形にはなっていないもの

研究の実行主体自身と一体化されていて、外在化の程度が低い

中間的な各種の資料, 思い付き, メモ, 経験, ノウハウ, etc.

研究活動の節目は(B)を(A)に作り上げていくことにある。(A)は(B)から生まれて

くるものであるが、資料保存の立場から見ると、研究活動や事業によって生まれ、蓄積すべき研究資料・成果は(A)から(B)まで、いろいろな形で存在する。

また、(B)を積極的に外在化する仕事として、あるいは(A)が生まれた背景としての情報を資料化する仕事として、例えば、聞き取り調査(録音・録画)の試行を始めている。

アプローチの仕方として、いくつかのアプローチを組み合わせることを考えている。そのひとつとして、形になったものから入るというアプローチを考える。成果物として世に問うているものを入り口として整理する。

- (1) 資料としてはっきりしているものから整理する
- (2) 上のような枠とそれ以外の枠を組み合わせる

このような考えの上に立って、文献情報の位置付けをしている。文献情報は過去から現在に至る研究活動の蓄積へ接近するための手がかりとして捉えることができる。研究成果へのインデックスとしての機能や研究成果の蓄積に関する情報を組織化する機能を果たすものとして文献情報を捉え、研究情報システム中への組み込みの試みを行っている。そのために、書誌情報の他に、現在までに、目次のデータベース化、索引のデータベース化、図表目次のデータベース化、英文概要(英文による研究報告の要約を主要部分とする)のデータベース化を行なっている。また、これまでに刊行して来た国立国語研究所の報告書本文の電子化を計画している。

これらは、研究報告という書物としての研究成果へのインデックスであると同時に、研究資料のアーカイブにおける役割という視点も与えることができる。現在、国立国語研究所における調査資料やデータなど研究資料のデータベース化の仕事を進めているが、この中で、資料データベースとリンクしたものとして、文献情報を捉えており、その整備を進めている。

研究活動の過程において生み出される資料は、内容、形式、その意味など、多様なものである。この研究活動の成果を記録蓄積していくという活動、研究によって生産される資料を保存、蓄積すると同時に、新たな研究の中でも活用できるような仕組みを作っていくことが必要となっている。この資料の保存蓄積という観点から見たとき、研究報告などの文献は、記録として安定したものであり、研究者である著者によって組織化されたものであるという特徴があり、資料データベースとのリンクを図ることによって、研究活動の成果を記録蓄積していくという活動の中の重要な柱になるものと捉えている。

必要な研究情報へアクセスする仕組みをどのように作ればいいのか、埋もれている情報を表に出すにはどのような仕組みを作ればいいのか、ということに関して、様々な検討をしてきているが、そのような広い枠組みの中の重要な柱のひとつが、この文献情報の整備だと考えて、情報の整備を進めて来ている。

2.4 資料の電子化と電子媒体

資料の保存・利用・管理のシステムを実現するための基盤のひとつの側面として、急速な発達と普及が進行しているコンピュータ、電子媒体、ネットワーク、電子機器というものがある。これらに対する評価と利用を進めることが課題である。

研究資料の蓄積，保存と電子的なメディアやネットワークの利用について示す。データベースや電子化した資料，資料の保存管理のためのシステム作り等について行っていることを示す。

きちんとした物理的な物の保管システムが現状ではなかなか根本的な対応が難しい状況にあり，できる努力を続けているところである。原資料の蓄積・保管に関しても，資料の保存のための十分な手当が必要であるが，研究室に保管されているもの以外は，現状では資料保管スペースというものが十分整備されておらず，段ボールに詰めて「資料庫」や空きスペースにしまい込むというような最低限のことを実行しているにすぎないものが多い。また，カード等の保管に関しても，カードボックスにしまわれたまま，保存のための環境としては好ましくないところに置かれているものも多い。現段階で可能な範囲で，資料の散逸や環境の悪化に対しての注意は行っているが，物理的な資料のシステムティックな保管の体制が重要な課題である。近い将来に移転も控え，できる限り資料の保管体制を整えておくよう，物理的な保管方法についても，整理保管用具を整備し，情報のみならず，物理的な整理を行うことが，課題となっている。

電子化に伴う問題として，電子化資料の持つ脆弱性をシステムの中で考えておく必要がある。例えば，媒体の劣化，記録の劣化，物理的な変形・破損，再生装置の消失，記録フォーマットの時代遅れ（例えば，古いワープロの文書ファイルはそのソフトがないと読めない），プログラムの実行環境の消失（プログラムは実行できる計算機環境がなくなると使えない），ソフトウェア資産の継承の問題等である。電子化そのものが新しい問題を持ち込むことがあるし，すでに電子化されている資産については，このような負の側面があることを忘れてはならない。

電子化によって生じる新しいコストという観点でもを考える必要がある。

電子化した資料そのものが劣化に強い訳ではない。デジタル化した資料はコピーが可能である。デジタルからデジタルへコピーしている限りにおいては，複写による劣化はないが媒体は劣化する。ある媒体やフォーマットがいつまでも安泰である訳ではない。

電子化した資料の資料としての安定性という観点からは種々の問題を抱えている。電子化した資料を利用するためのプラットフォームという点からも気をつけるべきことは多い。資料保存という観点からは，電子化した資料と実物資料の双方を保存しておく必要がある。

さらに，電子化した資料の量が増加するにしたがって，その全体の蓄積，管理のシステムの重要性が増す。

2.5 権利の保護

著作権： 研究報告等のほとんどは国立国語研究所が著作権を持つ（執筆者が著作権を持つものが中にある）ものであり，その限りでは，公開に当たって著作権に関わる障害が少ないが，資料の中には調査目的に応じて，それぞれに著作権者との一定の契約のもとに作成した資料が少なからずある。これらについては，その著作権に関する情報を明示的に集中管理しておく必要がある。

被調査者の情報の保護： プライバシー・肖像権の保護： すでに公刊されている情

報に関しては問題はないが、もとの資料、データの扱いに関しては、被調査者の権利の保護に関しての注意が必要。調査票、録音資料、録画資料など、様々な資料があり、これらについても権利の保護に関する情報を明示的に集中管理しておく必要がある。

3. 研究成果の一次情報の提供

3.1 研究報告の電子化と研究資料の電子化

今日、世界的にコンピュータネットワークの急速な普及が進んでおり、インターネットによる世界的な規模での情報の共有と利用が可能になってきている。また、情報処理技術の発展によって新しいメディアによる大量の情報の蓄積利用の手段が一般に利用可能になってきている。一方、印刷物としての刊行物は、学術書の多くがそうであるように、品切れ等によって入手できなくなるものが多い。また、海外においては基本的な文献として国語研究所の刊行物を情報検索によって検索することができても、その本体に触れることは、多くの場合、非常に困難である。インターネット上の2次情報の整備により、研究や資料の存在を知ることが容易にできるようになってはいるが、あるいは、できるようになってきているからこそ、研究・資料の本体にも同様の容易さでアクセスできるようにすることの必要性は高い。

国立国語研究所が1948年の設立以来行って来た日本語研究の成果を、

- (1) その中心的な存在である研究報告を電子化し、これに対する情報検索手段を整備し、これを研究成果と資料へのアクセス手段として整備する。
- (2) インターネット等を利用して世界に向けて研究成果の公開を推進しする。
- (3) これらと対応する形で、その研究の元になった原資料やデータの組織的な管理・保管システムに関する研究と整備を行い、総合的かつ組織的な研究の蓄積・保存・管理・利用と公開を行うシステムに関する研究を行ないシステムの構築を行なう。

これにより、

蓄積した研究成果に新たな可能性を与える。

現在進行中の新たな研究成果の蓄積と公開のためのシステムの基盤ともなる。

現在進行中の新しい研究に関する研究情報の蓄積利用の高度化につながる。

研究報告の電子化：

国立国語研究所設立以来の研究成果の中心的なものは研究報告等の印刷物の形で世に出、利用されてきているが、日本国内においても、書店等では既に入手できなくなっているものも多い。また、インターネットを通じて、国内のみならず世界中で、研究成果に関する情報にアクセスできるよう、現在、2次的な情報（書誌、目次、索引等）については検索ができるようにデータ入力等の作業を進め、部分的に公開を開始している。しかし、研究報告自体を手にとって見ることは困難なところがほとんどであると思われ、研究成果物そのものへのアクセス手段が強く望まれている。

研究資料の電子化：

研究所設立以来、これまでに蓄積されている研究成果はネットワークによる研究成

果の共有というようなことが現実的なものとなる以前のものであり、このようなことは考えられていない。すでに機械可読となっているデータに関しては、ネットワーク上での公開が検討されているものがあるが、研究所の研究成果物の全体へのアクセスという観点からは部分的なものであり、新たな電子化が必要である。

3.2 資料の電子化の例

研究報告の電子化（資料，データ，報告）

国立国語研究所研究情報資料データベース

音声資料のデジタル化

映像資料のデジタル化（例．国際シンポジウムの映像記録）

情報検索：(a)書誌データ，(b)目次データ，(c)索引データ，(d)図表目次，(e)英文要旨，(f)，資料情報データ（所在，形態，内容，注記等），(g)関連情報，(h)文献情報〔日本語研究文献目録雑誌編（国語学会との共同事業），海外日本語研究文献目録（国語学会との共同事業），雑誌論文文献目録，刊行図書文献目録，データベース出版による国語年鑑，日本語教育年鑑〕

4. バーチャル日本語情報資料館システム

システムの概略の概念図を次ページに示す。

5. バーチャル日本語資料館：資料の電子化の例

5.1 文献情報の役割

文献情報は過去から現在に至る研究活動の蓄積へ接近するための手がかりとして捉えることができる。研究成果へのインデックスとしての機能や研究成果の蓄積に関する情報を組織化する機能を果たすものとして文献情報を捉え、研究情報システム中への組み込みを試みている。そのために、書誌情報の他に、目次のデータベース化、索引のデータベース化、図表目次のデータベース化を行なっている。また、これまでに刊行して来た国立国語研究所の報告書の電子化に取り掛かっている。

これらは、研究報告という書物としての研究成果へのインデックスであると同時に、研究資料のアーカイブにおける役割という視点も与えることができる。現在、国立国語研究所における研究資料のデータベース化の仕事を進めているが、この中で、資料データベースとリンクしたものとして、文献情報を捉えており、その整備を進めている。

研究活動の過程において生み出される資料は、内容、形式、その意味など、多様なものである。この研究活動の成果を記録蓄積していくという活動、研究によって生産される資料を保存、蓄積すると同時に、あらたな研究の中でも活用できるような仕組みを作っていくことが必要となっている。この資料の保存蓄積という観点から見たとき、研究報告などの文献は、記録として安定したものであり、研究者である著者によ

バーチャル日本語情報資料館

電子化した資料を用いてコンピュータ上に仮想的に実現した資料館。日本語研究資源情報基盤システム上に構築する。

- ・インターネットを通じて、国立国語研究所が蓄積した日本語に関する資料・情報を閲覧利用することができる。
- ・研究成果と研究者や市民を結ぶインターフェースを構築し、研究成果の普及を促進する。
- ・「日本語教育支援総合ネットワーク」の運用により、日本語教育への支援を促進する。

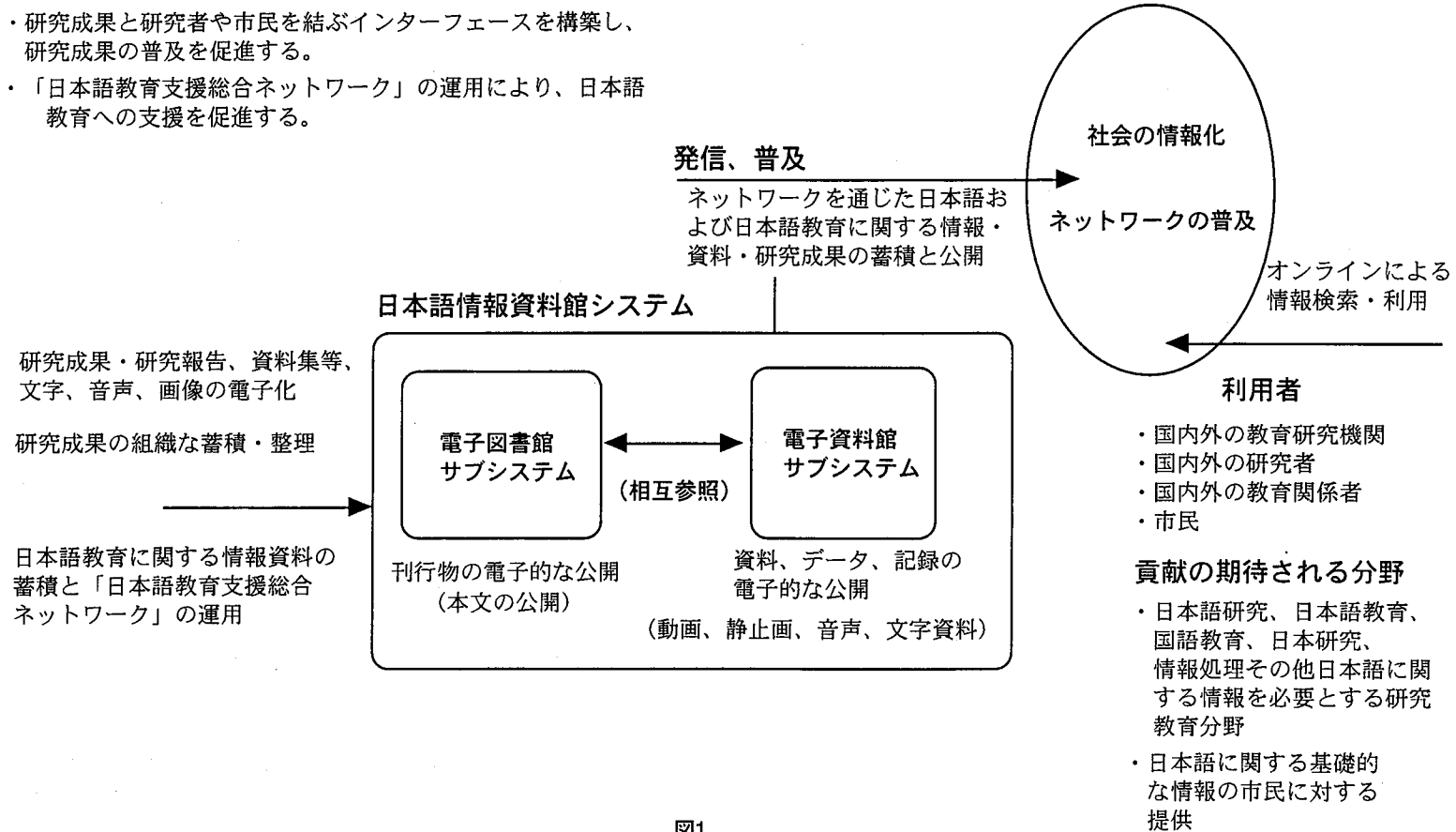


図1

って組織化されたものであるということがあり、資料データベースとのリンクを図ることによって、研究活動の成果を記録蓄積していくという活動の中の重要な柱になるものと捉えている。

5.2 資料の電子化の例1：日本言語地図データベース

電子化により、蓄積資料の利用の高度化と資料の保存の両方を狙うものとして、「日本言語地図データベース」を一例として取り上げることにする。

『日本言語地図』全6巻（国立国語研究所編，1966年～1974年）に地図化された方言分布に関するすべての情報、および、その原資料として国立国語研究所に唯一保管されている原資料のカード約54万枚の調査記録のすべてをデータベース化し、公開する計画であり、現在、本格的な作業に向けての準備的な入力作業を続けている。

『日本言語地図』は、全国各地の方言でどのような語形や発音がどこに現れるかを項目ごとに地図で表示したものであり、日本全国の方言の地理的分布を一望できる基礎資料である。調査は、1957年～1965年に行われ、調査地点数2400、調査項目数285からなり、すべて面接調査で実施した。

原カードには、調査者によって記入された各調査地点における個々の回答に関する原情報・原表記のみならず、『日本言語地図』を作成するために行った分類整理の情報が記録されている。『日本言語地図』データベースは、この記録を元に、分類整理され地図化された情報を文字データとし、それぞれの回答に対応する手書きの原表記や注記を含む原カードを画像データとすることにより、これまで原カードに遡らずには得られなかった情報や地図として公開されなかった項目も含め、『日本言語地図』のすべての情報をデータベース化し、日本の方言に関する基礎資料を広く提供しようとするものである。

このデータベースの作成により、(a)基礎資料である『日本言語地図』全6巻の全情報が広く共有されることになる、(b)地図化された情報が文字データとしてもデータベース化されることにより、地点や語形、回答パターンなどにより、『日本言語地図』の多面的な検索が可能となり、研究や教育上の利用の推進に寄与できる、(c)地図化の元になった原カードを地図化の結果と同時に参照できるようになることによって、『日本言語地図』の資料批判への道を開くことになり、研究の深化に寄与することができる、(d)語形一覧、注記一覧などの資料は、国立国語研究所の図書館に一揃いが製本され保管されているだけであるが、これらに対応する情報をデータベースの操作により、直接原資料より得ることができるようになり、公刊されている『日本言語地図』だけからは得ることができない情報が広く一般に参照可能なものになる、(e)これまで部分的に機械可読データ化されている『日本言語地図』の全体が機械可読データ化されることになり、『日本言語地図』に対する本格的な計量的方言研究が可能になる、(f)現在は、原資料は研究所に保管されている一揃いだけであるが、原カードを高解像度でスキャンすることにより、原カードの保存対策となる。さらに、『日本言語地図』の地図を画像ファイル化したデータとのリンクを図り、『日本言語地図』のデータベース化と公開の徹底を予定している。

5.3 資料の電子化の例2：録音資料のデジタル化の例

過去に録音された資料のデジタル化は、資料保存のための劣化対策と、新たなメディアによる利用への展開の両側面を持つ。方言談話資料、24時間調査の録音資料、言語生活の録音資料、その他各種の言語調査時の録音資料等、さまざまな過去の音声資料がある。これらのデジタル化は言語資料としての記録資料の保存であるとともに、過去の研究の再検討にも役立つものである。現在は、例えば、「話ことばの文型」等の研究のために集められ分析された過去のはなし言葉録音資料（オープンリールテープ）や、社会調査型の言語調査の録音テープのデジタル化を進めている。

6. おわりに

「バーチャル日本語情報資料館構築」の実現へ向けての試行や検討を進めている。冒頭にも述べたように、国立国語研究所全体における資料・研究成果や現在進行中の研究・事業による資料・研究成果の蓄積と利用を組織的に行うための仕組みを構築、運用するための仕事をいっそう推進することが必要となっている。これらの情報や資料に対して、物理的にも情報的にも整理・保存・蓄積を組織的に行い、日本語に関する共通の知的な財産・資源として将来へ継承しつつ、活用していくための仕事を進めていきたいと考えている。そのため、研究成果の組織的な蓄積・保存・管理体制の整備と、その蓄積全体に関する情報検索・アクセス手段の整備が必要である。「バーチャル日本語情報資料館構築」はそのためのひとつの柱となるものである。

情報・資料の蓄積保存と利用を両輪の輪とし、技術的な展開を反映させながら、発展させていきたい。

なお、本稿を補足するものとして、当日の配付資料および発表会の後の研究室公開のデモンストレーションを通じて、できるだけ具体的なイメージを提供したい。