

目 次

人間文化研究情報資源共有化研究会報告集10の刊行にあたって	永村 真 ……	3
-------------------------------	---------	---

第15回研究会報告集

第15回研究会プログラム ……		5
-----------------	--	---

趣旨説明

大内英範（人間文化研究機構本部） ……		7
---------------------	--	---

報告

「人文系研究データのライフサイクルを考える」

天野絵里子（京都大学学術研究支援室）

要旨 ……		9
-------	--	---

スライド ……		10
---------	--	----

「日本古典文学研究における「可逆性」とは」

佐々木孝浩（慶應義塾大学附属研究所斯道文庫）

要旨 ……		29
-------	--	----

スライド ……		30
---------	--	----

「話し言葉調査の「可逆性」」

高田智和（国立国語研究所）

要旨 ……		39
-------	--	----

スライド ……		40
---------	--	----

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」 ……		53
--------------------------------	--	----

はじめに

可逆性の検証実例について

デジタル化のメリット・デメリットについて

まとめ-人間文化研究機構／人間文化研究情報資源共有化システムに期待すること

人間文化研究情報共有化研究会報告書10の刊行にあたって

人文学関係の研究情報は多彩にわたり、その情報の質と量が常に問われることになりました。人間文化研究機構では、人文系の6研究機関（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）において構築された多様な形式・内容をもつ大規模なデータベース（以下DB）を共有化し、幅広い目的に対応できる情報公開事業を進めてきました。また本機構では、6研究機関のDBとは別に、機構独自に基盤DBの構築を進め、既存のDBとのデータ連結による、有効な情報提供の実現を目指しています。

ところで人文学系のDBですが、データ群をはさんでデータの生成・活用という両様の場面で、人文学分野の研究活動により支えられてきました。DBは単なる固定的なデータ群ではありません。不断の調査・研究活動のなかで見直され更新されるデータ群であり、日常的な活用のなかで修正を求められ新たな活用が提案されるデータ群であって、データとしての精度や活用効率率は、常に当該分野の研究に裏打ちされています。

またDBに格納されるデータ群は、それ自体が独自に権威をもち、その有効性が一人歩きするものではありません。データの生成段階、データの活用段階に目配りをして再検討がなされ、データにより導かれた結論が常に正しいか検証される必要があります。結論は疑い得るものとの姿勢を前提にした、「通説」をそのままに容認しない研究態度が不可欠なのです。

さて本研究会は、改めて「人文系研究データの生成と管理——『可逆性』の実現のために——」という課題のもとに開催いたしました。この「可逆性」という語句については、人文学系と理科系では厳密な定義は異なりますが、人文学系では提示された結論が、検討素材の活用により妥当に導かれるかを再度検証し、そのデータ・データ群が研究上で有効性をもつか再評価する必要があります。常に根本に立ち戻り、他の評価・解釈は可能ではないかを反芻する姿勢のもとで、素材と結論を結ぶ筋道を見直すことにより、硬直化した結論を排除することができます。本研究会では、このデータ群の性格を素材の次元から見極めるとともに、活用の有効性を導くデータ管理の意義についても貴重なご提言をいただきました。

2020年3月

大学共同利用機関法人 人間文化研究機構
総合情報発信センター情報部門長・客員教授 永 村 眞

人文系研究データの生成と管理

「可逆性」の実現のために

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、人文系の研究の中には、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を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。

- **日時** 2020年1月25日(土) 10:10~11:45
- **会場** 日比谷図書文化館 スタジオプラス (小ホール)
(〒100-0012 東京都千代田区日比谷公園 1-4)
- **プログラム**
 - 10:10 あいさつ・趣旨説明 大内英範 (人間文化研究機構)
 - 10:15 「人文系研究データのライフサイクルを考える」
天野絵里子 (京都大学学術研究支援室)
 - 10:35 「日本古典文学研究における「可逆性」とは」
佐々木孝浩 (慶應義塾大学附属研究所斯道文庫)
 - 10:55 「話し言葉調査の「可逆性」」 高田智和 (国立国語研究所)
 - 11:15 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・関野樹 (国際日本文化研究センター)
 - 11:45 閉会 永村眞 (人間文化研究機構総合情報発信センター 情報部門長)
- **主催** 大学共同利用機関法人人間文化研究機構 総合情報発信センター
- **連絡先** 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
〒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) 03-6402-9234 (FAX) 03-6402-9240 (Mail) cip-office@nihu.jp

趣旨説明

人間文化研究機構 大内 英 範

おはようございます。慌ただしくて恐縮ですが、時間になりましたので、始めさせていただきます。本日は第15回人間文化研究情報資源共有化研究会にお集まりいただきまして、ありがとうございます。人間文化研究機構の大内と申します。よろしく願いいたします。

本研究会は、人間文化研究に関わる研究情報資源共有化の推進について、皆さまとさまざまな意見交換をしたり、さらに連携を発展させる機会を得るという目的のために、ほぼ年に1回ほどのペースでこれまで開催させていただいております。今回のテーマは「人文系研究データの生成と管理——『可逆性』の実現のために——」です。研究データについては、3年ほど前にもこの研究会でオープンサイエンスの文脈からテーマとしたことがあります。今回はそのオープンサイエンスの一步手前の段階という形になるかと思いますが、人文系の各分野の研究データが、今回は特に国文学と国語学の分野においてどのようなものなのか、どのようにして作られていて、その中身はどのようなものか、あるいはその研究データからいわゆる一次資料などに本当に戻ることがどれくらい可能なのか、つまり研究自体の検証が本当に可能なのか、どうやって可能なのかということをもテーマとしました。テーマのサブタイトルに「可逆性」という言葉を使っております。辞書的なところでは、物質がその状態を変化させたときにまた元の状態に戻ることができる性質のことを可逆性と言うようですが、ここでは今申し上げたように研究データから一次資料に戻ることができるか、実際にその検証が可能なのかといった意味で使っているということをご了解いただければと存じます。

お手元に今日のプログラムをお持ちだと思いますが、まずは京都大学学術研究支援室の天野絵里子先生に、そもそも研究データとはどういうもので、どのように管理されるべきで、実際にどのように管理されているか、中でも特に人文系研究データについて触れていただく形でご講演をいただきます。次に慶應義塾大学斯道文庫の佐々木孝浩先生、そして国立国語研究所の高田智和先生から、それぞれ国文学分野・国語学分野において、特に国文学では写本などの一次資料、それから国語学では今回は「話し言葉」というモノですらないものを研究対象にした場合のお立場から、研究データの可逆性について、実際のところはどうか、問題点はどのようなところなのかをお話しいただきます。そして、最後に30分ほど議論の時間を取っておりますので、会場の皆さんからも積極的なご参加をいただければと存じます。短い時間ではありますが、どうぞよろしくお願い申し上げます。

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 リサーチ・アドミニストレーター 天野 絵里子

本発表では、研究データのライフサイクルと分野の多様性に着目しながら、人文系の研究においていかに研究データを管理していくかを研究者が自ら考えるための話題を提供する。

オープンサイエンス政策が推進される文脈では、自然科学系の研究を念頭に置いた研究データ管理の必要性が主張されてきた。研究データは実験や観測によって生成され、解析・編集され、それを元にした研究成果が論文などの形で発表され、第三者に再利用される。この研究データのライフサイクルの中で、利活用されたデータや研究成果の根拠となるデータをなるべく共有することによって新たな研究の創出を促進しようという取り組みが進められてきた。データ共有を進めるためには、研究のライフサイクルに沿った研究データ管理を効率よく、効果的に行う必要がある。そして、そのための基盤整備や人材育成、制度設計が、国や研究機関などさまざまなレベルで進められてきている。オープンサイエンスにおいてよりよいデータ共有とは何か、「FAIR原則」でわかりやすく示されてもいる。ただしこの文脈において、人文系の研究における研究データ管理について幅広く議論されてきたとは言えない。人文系の研究データは多様であり、研究データのライフサイクルについて個々の研究者の意識に上ることは少なかった。

本発表では他の分野における「研究の再現可能性」などをめぐる取り組みやツールを紹介し、人文系の多様な研究の実践における研究データ管理のあり方をめぐる議論のきっかけとしたい。

人間文化研究機構・研究情報資源共有化研究会
2020年1月25日（土）日比谷図書文化館



人文系研究データの ライフサイクルを考える



天野絵里子
博士（技術経営）

<https://researchmap.jp/amanoeriko/>
ORCID: 0000-0001-6234-152X

Jisc, 2020.
<https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle/>
CC BY-ND

1

自己紹介

<https://researchmap.jp/amanoeriko/>

リサーチ・アドミニストレーター（URA）

- 研究支援の専門職（新しい職種）
- 学内ファンド、URAの研修プログラム、人文・社会科学系研究の支援
- オープンサイエンスの推進

研究成果の
効果的な発信

×

元・ライブラリアン

- 京都大学、九州大学、国際日本文化研究センター
- 資料と人をつなぐ専門職
- オープンアクセスの推進

オープンアクセス
オープンサイエンス

2020/1/25

2

研究データ管理の教材作成・講師

オープンサイエンス時代の研究データ管理



NIIがJPCOARの協力を得て制作
2017年11月15日から2018年1月15日まで
gaccoにて提供

- 第1週：研究データ管理とは
- 第2週：保存・共有・文書化
- 第3週：メタデータ・法倫理的問題
- 第4週：運用に向けて

<https://www.nii.ac.jp/service/jmooc/rdm/>

研究データ管理サービスの設計と実践



大学図書館員、URA、技術スタッフ等の研究支援職員が、研究者の研究プロセスに沿って、研究データ管理サービスの設計と実践の方法を身につけることを目的とした講座。

※来年度より本格提供

- 第1章：序論
- 第2章：サービス設計
- 第3章：研究前の支援
- 第4章：研究中の支援
- 第5章：研究後の支援
- 第6章：日常的な支援

<http://id.nii.ac.jp/1458/00000107/>

2020/1/25

3

今日の話題



- 可逆性
- オープンサイエンス
- 研究公正、研究倫理



- 人文系研究データのライフサイクル



- RDMルーブリック
- 研究者コミュニティへの期待

2020/1/25

4

なぜ研究データ管理？

2020/1/25

5

再現可能性／可逆性

人文学における検証可能性

- 同じ現象・資料に依拠して研究を行っても人によって成果が異なる
- 記録・記述しただけで情報が劣化する
- NOT reproducibility（再現可能性）, replicability（複製可能性）
- 研究の発展の過程で、成果の根拠となる資料、文献（≡研究データ）に立ち戻ることができる道筋を示す必要がある

- 人文学における研究データ管理の必要性

Copyright © The Nobel Foundation 2006
<https://www.nobelprize.org/prizes/facts/the-nobel-medal-for-physics-and-chemistry-2>
<https://www.nobelprize.org/prizes/facts/the-nobel-medal-for-literature>

2020/1/25

6

オープンサイエンスとは (1)

政策の中での定義

オープンサイエンスとは、

公的研究資金を用いた研究成果（論文、生成された研究データ等）について、科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開くとともに、効果的に科学技術研究を推進することでイノベーションの創出につなげることを目指した新たなサイエンスの進め方

（国際的動向を踏まえたオープンサイエンスに関する検討会内閣府 2015）

オープンサイエンスとは、

オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である。

オープンイノベーションの重要な基盤としても注目されている。

- 研究者の所属機関、専門分野、国境を超えた新たな協働による知の創出を加速する
 - 社会に対する研究プロセスの透明化や研究成果の幅広い活用が図られる
 - 市民参画型のサイエンス（シチズンサイエンス）が拡大する
- （第5期科学技術基本計画. 2016-2020）

2020/1/25

7

オープンサイエンスとは (2)

より広い定義

研究者のような専門家だけでなく非専門家であっても、あらゆる人々が学術的研究や調査の成果やその他の発信される情報にアクセスしたり、研究活動に多様な方法で参加したりできるようにするさまざまな運動のことである。オープンサイエンスは、オープンアクセスの推進など科学的な知をもっとオープンにし、社会に伝えるというさまざまな活動を含む。

Wikipediaによる「オープンサイエンス」の定義

参考：KYOTOオープンサイエンス勉強会。

【レポート】Wikipedia日本語版の「オープンサイエンス」をみんなで編集してみよう！
<https://kyoto-open.science/?p=332>

2020/1/25

8

オープンサイエンスとは (3)

5つの school

学派	仮定	目的
民主 Democratic	知識へのアクセスが不平等な状態にある	誰でも自由に知識を入手可能にする
実用 Pragmatic	科学者間の協働が進めばより効率的に知識創造が進む	知識生産プロセスをオープンにする
設備 Infrastructure	効率のよい研究は、使えるツールや設備で決まる	科学者のためのオープンな基盤、ツール、サービスの提供
公共 Public	科学活動は公衆にアクセスされなければならない	科学活動に市民がアクセスできる
測定 Measurement	科学の現在の貢献を測るには別のインパクト指標が必要だ	科学のインパクトを測る新システム開発

山内保典, オープンな科学コミュニケーションが公正な研究に資する可能性と役割, 科学技術社会論研究, 14, 2017. による翻訳及び原文 Fecher B., Friesike S. Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) Opening Science. Springer, 2014. https://doi.org/10.1007/978-3-319-00026-8_2 より作成

2020/1/25

9

研究公正 (1)

研究公正に関するポリシーの指針



日本学術会議

『科学研究における健全性の向上について』 (2015年3月)

- 対象：基本的に論文の根拠資料（実験ノート、数値データ、画像）
- **原則、論文発表後10年間の保存期間**
- 必要に応じて開示／公開を推奨
- 教職員の異動・退職後も、機関として保存又はアクセスを確保

- 不正があった時に検証するために研究データを保存する

日本学術会議「科学研究における健全性の向上について」
(2015年3月)
<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第3章「研究前の支援」より

2020/1/25

10

研究公正 (2)

研究機関のポリシーの例



京都大学

『京都大学における公正な研究活動の推進等に関する規程第7条第2項の研究データの保存、開示等について定める件』（2015年7月）

- 論文等に疑義が呈された場合、その他必要に応じて保存する
研究データの開示

http://www.kyotou.ac.jp/ja/research/ethic/research_guide/documents/research_data150730.pdf



高エネルギー加速器研究機構

『公正な研究活動の推進に関する規程』（2015年3月）

- 研究活動によって得られた研究データ等の一定期間の保存、必要に応じた開示

https://www.kek.jp/ja/About/Compliance/Misconduct/1kitei_katsudou170213.pdf

2020/1/25

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第3章「研究前の支援」より

11

オープンサイエンス

アカデミア

社会

研究の発展

イノベーション
インパクト

研究データ管理



データ共有

研究公正
研究倫理

As open as possible
As closed as necessary

科学への信頼

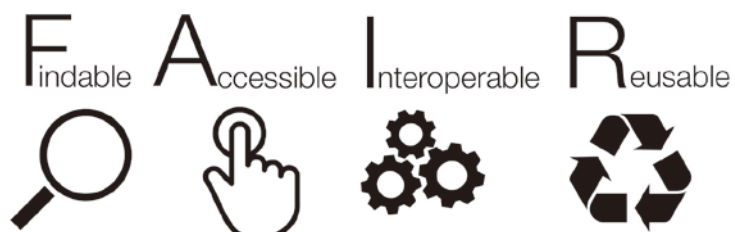
- オープンサイエンスを進める戦略の一つとしてデータ共有が推奨されている
- データ共有を含めた「研究データ管理」の基盤整備が進められている

2020/1/25

12

データ共有の原則

FAIR原則



To be Findable: (見つけられるために)

To be Accessible: (アクセスできるために)

To be Interoperable: (相互運用できるために)

To be Re-usable: (再利用できるために)



SangyaPundir, <https://commons.wikimedia.org/w/index.php?curid=53414062>
バイオサイエンスデータベースセンター. データ共有の基準としてのFAIR原則. 2018.
doi:10.18908/a.2018041901 <https://biosciencedbc.jp/about-us/report/fair-principle/>

2020/1/25

13

研究データ管理とは？

2020/1/25

14

研究データ管理とは

RDM: Research Data Management

研究データ管理 = ある研究プロジェクトにおいて使用された、あるいは生成された情報を、どのように組織化、構造化、保管、管理していくのか、を指す言葉



研究前に研究データの取り扱い計画を策定

研究中の日々の情報の取り扱い

研究後の長期的なデータの取り扱い

Research Data Oxford, About RDM
<http://researchdata.ox.ac.uk/home/introduction-to-rdm/>

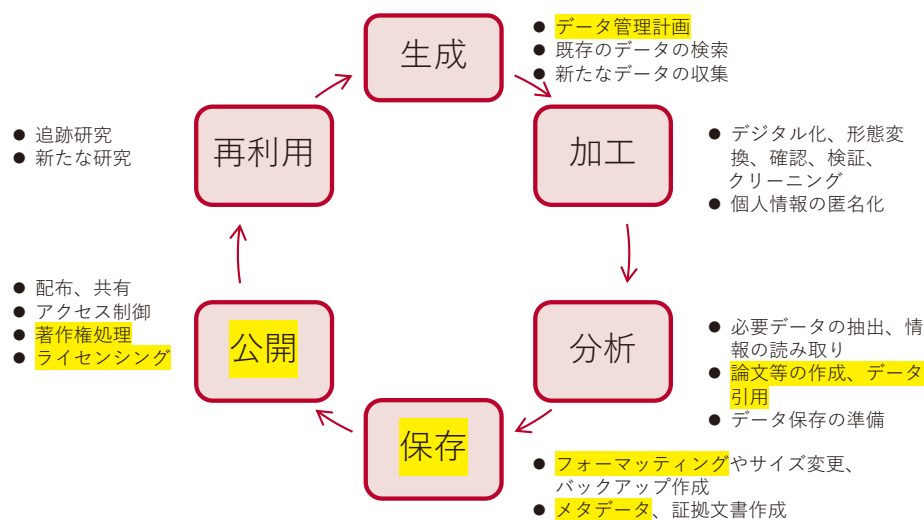
JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第1章「序論」より

2020/1/25

15

研究データ管理の要点 (1)

研究データのライフサイクル



The UK Data Archive, Research data lifecycle <http://www.data-archive.ac.uk/create-manage/life-cycle>
Mariette van Selm, RDM Support - basic training course for information specialists (2015)
<http://dx.doi.org/10.6084/m9.figshare.1285313> を元に加工

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第1章「序論」より

2020/1/25

16

研究データ管理の要点 (2)

データ管理計画 (DMP)

データ管理計画 (Data Management Plan, DMP)

研究プロジェクト等における研究データの取り扱いを定めるものであり、具体的にはデータの種類、フォーマット、アクセス及び共有のための方針、研究成果の保管に関する計画などについて記載されるもの



文部科学省「学術情報のオープン化の推進について（審議まとめ）」（2016年2月）
http://www.mext.go.jp/component/b_menu/shingi/toushin/_jcsFiles/afieldfile/2016/04/08/1368804_1_1_1.pdf

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第3章「研究前の支援」より

2020/1/25

17

研究データ管理の要点 (3)

データ管理計画 (DMP) の例

DMPに明記すべき項目：

管理対象となる研究データの保存・管理方針

研究データの公開・非公開に係る方針

公開可能な研究データの提供方法・体制

公開可能な研究データの想定利用用途

公開研究データの利活用促進に向けた取組み

JST「『オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針』運用ガイドライン」（2017年4月）
http://www.jst.go.jp/pr/intro/openscience/guideline_openscienc.pdf

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第3章「研究前の支援」より

2020/1/25

18

研究データ管理の要点 (4)

成果発表の際の研究データの引用

作成者 (出版年): タイトル. [バージョン.] 出版社. [データのタイプ.] 識別子

※[バージョン.]、[データのタイプ.]は追加的な要素

DataCite, Cite Your Data <https://www.datacite.org/cite-your-data.html>

分野やデータの種類に応じて以下の要素を追加する場合も：

- 経済・社会科学系：データがカバーする地理的情報、時間的範囲
- 生命科学分野の時系列データ：データベース名・アクセス番号 (例)
European Nucleotide Archive accession number CY115901
NCBI SNP database accession number ss472331023
- 継続的に更新されるデータベースの場合：ダウンロード日

研究で参照したデータセットは、他の利用者がアクセスできるように形で引用しなければならない。

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第4章「研究中の支援」より

2020/1/25

19

研究データ管理の要点 (5)

研究データの適切な保存場所

デスクトップ・モバイルPC



外部記憶デバイス



ネットワークドライブ



オンラインストレージ



JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第4章「研究中の支援」より

2020/1/25

20

研究データ管理の要点 (6)

ファイルフォーマット

推奨されるファイルフォーマット

独立した
フォーマット

国際標準化
されている
フォーマット

広く普及
している
フォーマット

フォーマット変換のリスクに注意



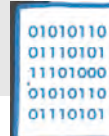
テキストファイル

- 環境に依存しない
- 可読性高

OR

バイナリファイル

- 豊富な情報を含む
- 人間可読性低



JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第4章「研究中の支援」より

2020/1/25

21

研究データ管理の要点 (7)

データリポジトリ

分野別
リポジトリ

データリポジトリの
検索サイトre3dataから
分野を絞り検索可能



<http://www.re3data.org/>

機関
リポジトリ



大阪大学学術情報庫



九州大学、九州大学学術情報
リポジトリ (QIR)



NII, IRDBコンテンツ分析
リポジトリ

汎用
リポジトリ

Zenodo

Figshare

Elsevier
Mendeley Data

データを取り扱う目的によって使い分けが必要

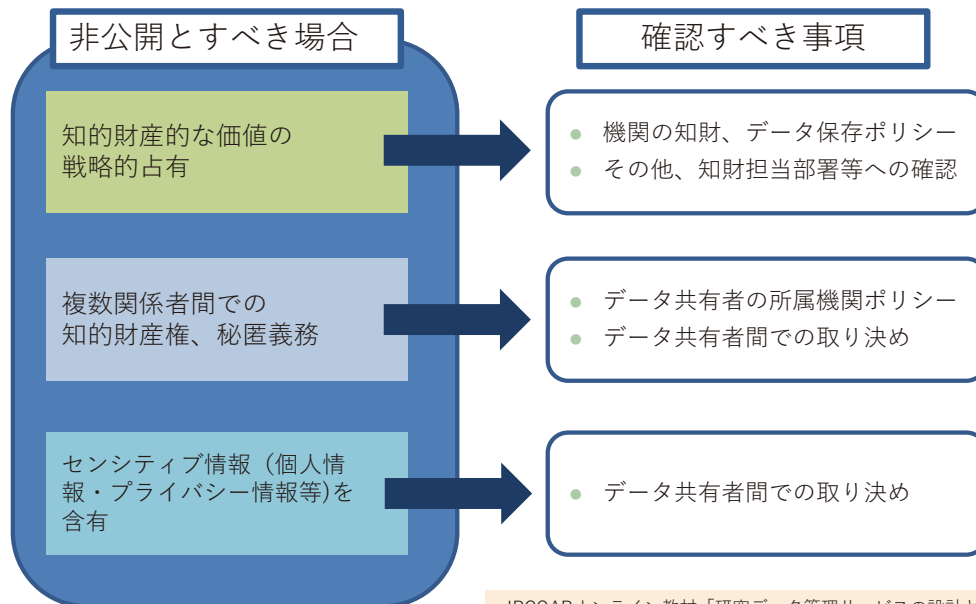
JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

22

研究データ管理の要点 (8)

研究データの公開の判断



JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

23

研究データ管理の要点 (9)

メタデータ

メタデータ = データに関するデータ

- データの詳細や文脈を記述
- 機械処理を想定（⇔文書：人の利用を想定）
- 一定のフォーマットに沿って記述することで、システム間の相互運用性を重視

データセットの検索・引用が可能に

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

24

研究データ管理の要点 (10)

著作権

著作物 = 「思想又は感情を創作的に表現したものであつて、文芸、学術、美術又は音楽の範囲に属するもの」をいう

(著作権法第2条第1項第1号)

判例：「データ自体は、仮にその集積行為に多額の費用、時間及び人員を費やしたものであったとしても著作権法の保護の対象となるわけではない」

名古屋地判平成12年10月18日、事件番号平成11(ワ)5181)

http://www.courts.go.jp/app/files/hanrei_ip/907/012907_hanrei.pdf

単なる客観的事実やデータは著作物とはみなされない
ただし、一定の考えのもとデータを整理・分析した場合は「創作性」が認められる場合も（地図、データベースの著作物etc.）。

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

25

研究データ管理の要点 (11)

ライセンス

ライセンス（利用許諾） = コンテンツの利用許諾の条件（ライセンス）を明示すること



研究データの公開時に、メタデータにライセンスを記載しておくことで、一律な利用が可能になる

※ライセンスが付与されていなくても、その他の知的財産権で保護されている可能性がある点に留意

JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

26

研究データ管理の要点 (12)

ライセンスの例

Creative Commons License
(クリエイティブコモンズ・ライセンス)

- 表示・非営利・改変禁止・継承の4条件を組み合わせて利用条件を表現するツール



Open Data Commons
(オープンデータコモンズ)

- 事実情報由来のデータ・データベースにも適用可能。



- ▶ 参考：ソフトウェアのライセンスツール
- GitHub, Choose an open source license



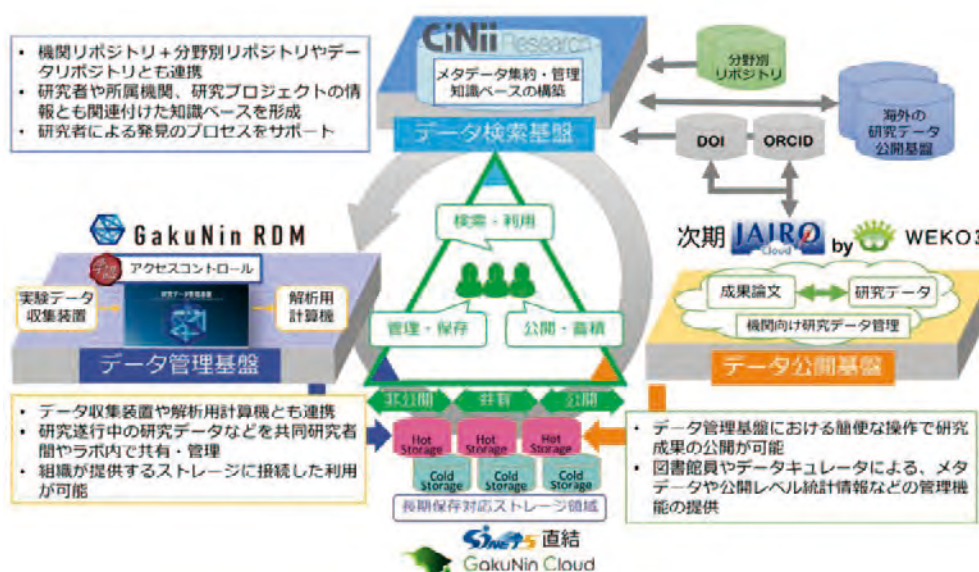
JPCOARオンライン教材「研究データ管理サービスの設計と実践」
第5章「研究後の支援」より

2020/1/25

27

参考：国立情報学研究所が提供する研究データ管理基盤

NII Research Data Cloud



込山悠介「研究データ管理サービス：GakuNin RDM」
第2回京都大学研究データマネジメントワークショップ発表資料より
<http://hdl.handle.net/2433/243961>

2020/1/25

28

研究データ管理、何から始める？

2020/1/25

29

自身で考えるためのツール

Support Your Data

<http://researchdata.org/>

スクリーンショット

2020/1/25

©Copyright 2018 The Regents of the University of California

30

RDMルーブリック(和訳)
https://github.com/kyoto-u/rdm_rubric

	その難点	僕のめ	常に通ず	評判のために最適化
データのオープン化	データの取り扱いが容易で、自分なりに考える。標準化や文脈はしていない。	最初に計画を練る。が、途中で見直すことは少ない。	詳細な計画を作成し、プロジェクトの進捗を定期的に確認する。	自分なりの見解をコメントやレビューを通じて共有できる。また、計画に即して修正できる。
データの保存とバックアップ	作業中にデータの保存やバックアップを忘れる。また、バックアップの場所や方法を覚えていない。	データの保存方法を整理している。損失のリスクを減らすために、作業前にバックアップをとる。	作業中でも定期的なデータの保存やバックアップを促している。複数のバックアップ方法を検討している。	再利用可能なフォーマットで保存し、アクセスしやすいようにしている。
データの整理と検索	整理が不十分で、必要なデータがすぐに見つからない。	事前に適切な形式と命名規則を設定し、整理している。定期的なメンテナンスを行っている。	データの整理方法を標準化し、一貫性を保っている。定期的なメンテナンスを行っている。	将来的に自分自身の研究が活用されることを想定し、適切な整理を行っている。
データの共有と公開	研究成果を共有しているが、その範囲や方法が不明確で、共有できない。	研究成果を共有する際の範囲や方法を明確にしている。必要に応じて、適切なアクセス権を設定している。	自身の研究成果やデータを利用できる形式で共有している。	自分の研究成果やデータを共有できる。また、第三者が利用できるようにしている。



2020/1/25

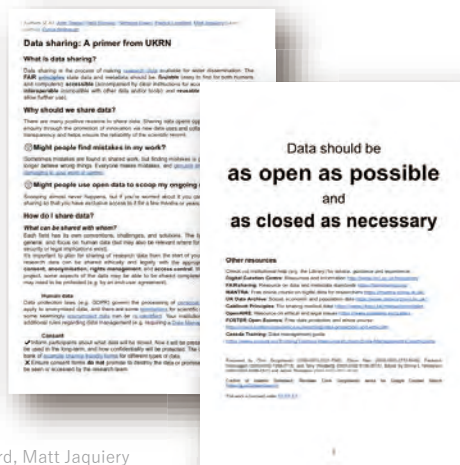
Borgh J, Abrams S, Lowenberg D, Simms S, Chodacki J (2018) Support Your Data: A Research Data Management Guide for Researchers. Research Ideas and Outcomes 4: e26439. <https://doi.org/10.3897/rio.4.e26439> より作成された日本語版

31

研究者コミュニティの活動

The UK Reproducibility Network (UKRN)

- 再現性の高い堅実な研究を行うための研修やベストプラクティスの共有を行うネットワーク
- Open Research Primers
初心者向けのガイド
- ReproducibiliTea
若手研究者による
ジャーナルクラブの
パッケージ



John Towse, Sally Rumsey, Nicholas Owen, Patrick Langford, Matt Jaquery [Lead author], Cylcia Bolibaugh. Data sharing: A primer from UKRN <https://www.bristol.ac.uk/psychology/research/ukrn/about/resources/>

2020/1/25

32

人文系研究者のデータ共有の実態

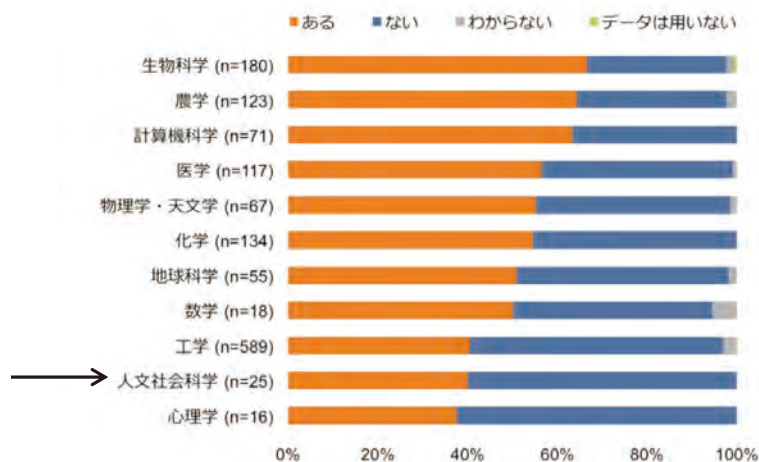


図 13 分野別データ公開経験 (n=1,395)



池内 有為, 林 和弘, 赤池 伸一. 研究データ公開と論文のオープンアクセスに関する実態調査. 調査資料, 263. 2017
<http://doi.org/10.15108/rm268>

2020/1/25

33

なぜデータ共有しないか

表 8 データ公開の障壁と新たな課題及び期待される効果

障壁	新たな課題	期待される効果
キャリアリスク 引用せずに利用される可能性 先に論文を出版される可能性 インセンティブの欠如	データ引用の標準化と徹底 データ公開の業績化 データ公開時期の(再)検討 先取権の保証	キャリアベネフィット 論文の引用増加 国際共同研究の契機 採用や昇進
データ公開のための資金 データ公開用のリポジトリ 公開データの信頼性	インフラへの投資と整備 発見可能性の向上 相互運用性の向上 データの評価・品質管理	研究成果の認知向上 公開データの利用促進 科学的利他性の醸成 低コストでの利活用
法的問題(知的財産権) 倫理的問題(機密等) 誤解・誤用・不正利用の懸念	リテラシー教育 ガイドラインの策定	再利用に関する懸念の低減 永続性の確保
データ公開のための人材 データ公開のための時間	データ管理支援体制の構築 人材育成とキャリアパス	研究コストの削減 時間や資金の節減



池内 有為, 林 和弘, 赤池 伸一. 研究データ公開と論文のオープンアクセスに関する実態調査. 調査資料, 263. 2017
<http://doi.org/10.15108/rm268>

2020/1/25

34

研究者コミュニティへの期待

考えるべきことの抽出

- 研究データとは何か？
- 再利用に必要な情報とは何か
 - データとともに残すべきメタ情報の特定

様々なレベルでの取り組み

- 研究機関
 - ポリシー
 - 情報基盤の整備
 - 関連部署との連携・協力
 - 大学院生への教育
- 学会、コミュニティ
 - 規範づくり
 - OSの浸透
 - 社会への発信

- 研究データ管理は喫緊の課題
- 「終活」では遅い。できることから始める
- 分野の専門家としての研究者の取り組みに期待

2020/1/25

35

日本古典文学研究における「可逆性」とは

慶應義塾大学教授・附属研究所斯道文庫長 佐々木 孝 浩

日本古典を研究する際に利用されるテキストは、大手出版社から刊行された、「大系・全集・集成」などと銘打たれた、その作品を専門とする研究者が校訂や注解を加えた本であることが多い。これらは、句読点や濁点、漢字への変換など様々に手が加えられたものであり、外国文学の翻訳に近いものであると言える。そこに不審がある場合には、使用された原典に当たる必要があるが、それは容易なことではなかった。近時のデジタル技術の発達は、高精細のカラー撮影を容易にし、出版やインターネットによる公開をも急速に促進した。この変化は、従来 of 古典文学研究を裨益しただけではなく、書誌学的研究の可能性を増大させもした。書物自体から得られる様々な情報を、作品解釈を始めとする諸分野に活かすことが積極的に行われるようになってきたのである。しかしそれにより情報量の格差という問題が生じている。原典の本文が読めれば十分であった時代から、表紙や紙質、細部の寸法などと言った詳細な情報が必要とされる時代となり、公開されている情報の斑が障害となっているのである。必要があれば直接原典の調査をすればよいというのは当然であるが、画像が公開されるとその本の閲覧が停止される例も少なくないのである。日本古典文学研究の「可逆性」とは、画像の影印やインターネットでの公開・現代の文字に改めた翻刻、そして校訂本文、その何れであっても、使用された原典に関する情報を直接・間接に確認できること、とまとめられるであろう。それを可能にするには今後どうすべきなのかを考えてみたい。

2020年1月25日

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斯道文庫

佐々木 孝浩

1

日本古典研究者の一般的なあり方

- 著名な古典作品を研究する場合には、大手出版社から刊行された、古典の全集や叢書などの、校注本を使用するのが一般的。

ex.: 源氏物語研究者は、小学館の『新編日本古典文学全集』の『源氏物語』(1994~8)を利用する。

理由: 時の学界を代表する専門の研究者が、(最)善本と考えられる伝本を利用して、従来の研究の蓄積を反映した、最新・最高の校訂本文と注解(・現代語訳)が提供されていると考えられているから。

・一歩引いた見方をすると、校訂本文というものは、外国文学の翻訳に近い存在であり、その本文は校訂者の学識(能力)や嗜好が反映されたものなのである。

* 問題のある箇所、疑わしいと思われる箇所などは、底本を確認するのが望ましく、他の伝本とも見比べるべきである。

2

『源氏物語』の校注本の多くは、 「大島本」と呼ばれる本を底本に選択している。 大島本を底本に採用した校注本

● 源氏物語大成	中央公論社	池田亀鑑編	1953～4	47帖
● 源氏物語評釈	角川書店	玉上琢弥編	1964～8	42帖
● 日本古典文学全集	小学館		1970～6	51帖
● 新潮日本古典集成	新潮社		1976～85	41帖
● 完訳日本の古典	小学館		1983	42帖
● 新日本古典文学大系	岩波書店		1993～7	53帖
● 新編日本古典文学全集	小学館		1994～8	41帖
● 岩波文庫	岩波書店		2017～	[53帖]

* 底本に使用した帖数が異なるのは、「定家手沢本」やその忠実な転写本とされる「明融本」等を使用するか否かに由来するものである。

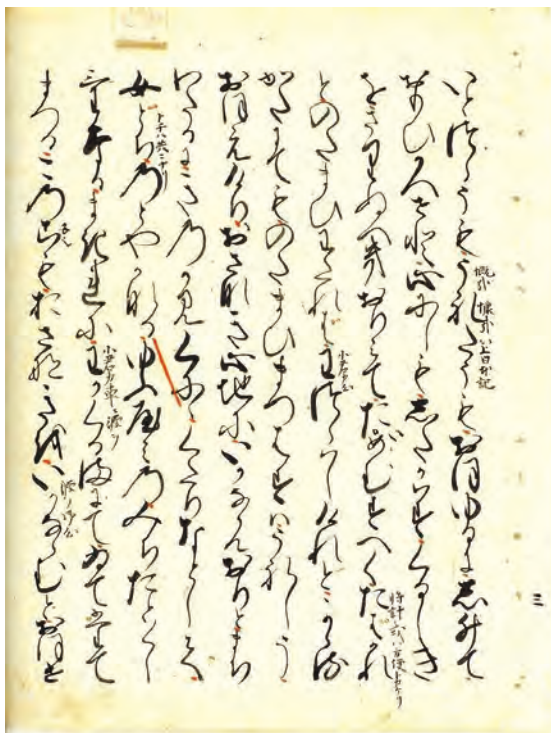
画像公開

- ・『大島本 源氏物語』(角川書店、1996～7)
- ・『CD-ROM版 角川古典大観「源氏物語」』(角川書店、1999)
- ・『大島本源氏物語DVD-ROM版』(角川学芸出版、2007)

3

底本と校訂本文

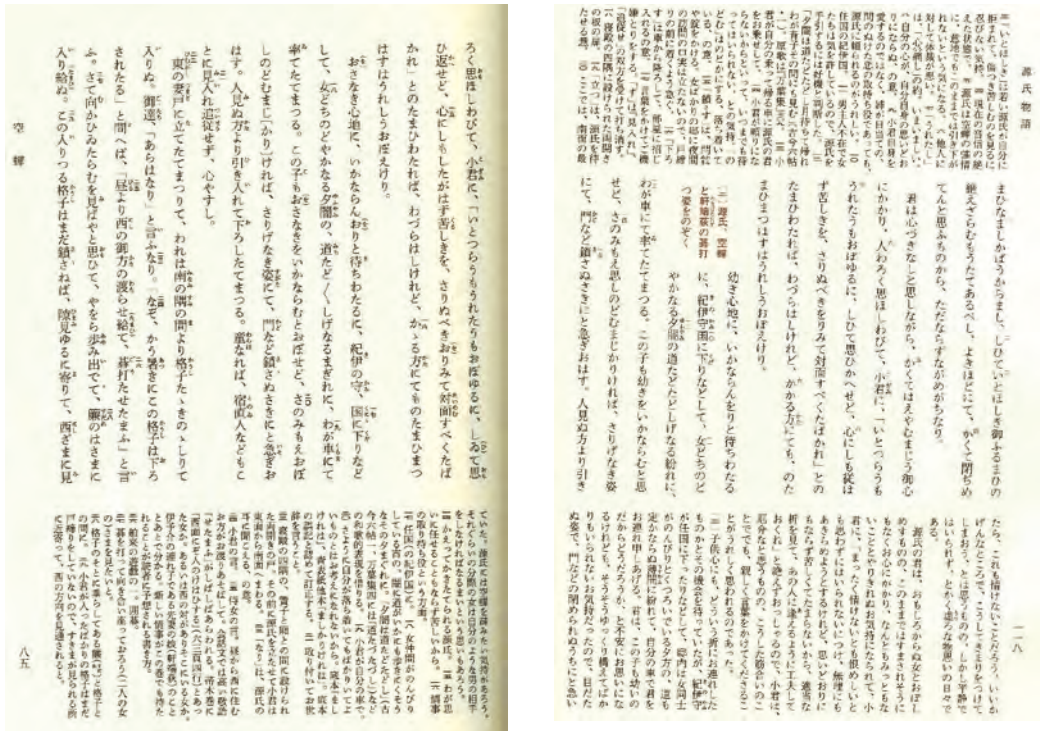
角川書店影印第1巻口絵と新古典全集当該部分



4

底本と校訂本文

岩波新大系と小学館新全集の当該部分の比較



5

底本(古典籍)確認方法の様々

- ◎: 所蔵者の元に出向き現物で調査する。
 - * 最も望ましいが、許可が下りない場合がある。お金がかかる。
 - : 複製で調査する。
 - * 高価であり、見られる場所が少ない。必ずしも忠実な複製ではない。
 - : 影印で調査する。
 - * 写真より鮮明ではない。白黒が多い。形態的に不明な点が多い。
 - △: マイクロフィルムで調査する。
 - * 見られる場所が少ない。白黒である。機械を使って見るのは大変。
 - △: 紙焼写真で調査する。
 - * 見られる場所が少ない。購入するとお金がかかる。白黒である。最近ではデジタル画像のカラー複写なども増えてきたのは○。
 - ? : インターネットで調査する。
 - * 簡便でお金もかからないが、画像の質にバラツキがある。
- @書誌情報については、原本調査以外は提供される情報が不足気味であるのが一般的な傾向である。

6

複製で調査する

一見本物ぽいので、本物もこういう物だと誤解する恐れ

『複製日本古典文学館 源氏物語
〈糸あはせ・みゆき・かしは木〉』
(ほるぷ出版、1972)

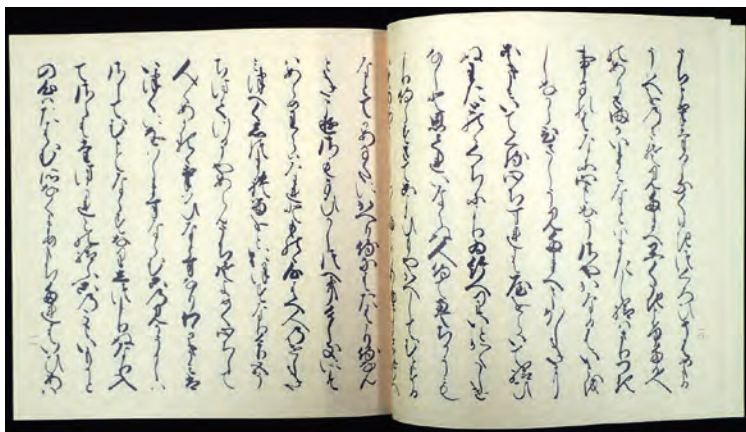


7

影印で調査する 1

読みやすいが、そのために消される情報もありうる

『宮内庁書陵部蔵青表紙本源氏物語』
(新典社、1968~70)



8

影印で調査する 2

写真的なものなので、やや読みにくいが実態を想像しやすい
『天理図書館善本叢書 源氏物語諸本集一』
(八木書店、1973)



9

マイクロフィルムで調査する

撮影の環境やフィルムの劣化などで、非常に見辛いものもある

例 『静嘉堂文庫所蔵 物語文学書集成』

(雄松堂書店、1980)

318リール

価格:5,088,000円(税別)



10

紙焼写真で調査する 紙焼き写真を製本したものの例

A5サイズのものが多く上に、ネガから製作した読みにくいものもある



11

インターネットで調査する時代の到来 早稲田大学図書館古典籍総合データベース

資料数も多く簡単に使用でき、ダウンロードも複数の形式で可能。理想的な
書物画像データベース。この登場が世界の日本研究の質を変えた。

[源氏物語]：若紫 / [紫式部] [著]



請求記号 Call No.	文庫30 A0010
タイトル Title	[源氏物語]：若紫 / [紫式部] [著] genji monogatari : wakamurasaki ●このタイトルで早稲田大学図書館検索(OPAC)試行検索●
著者/作者 Author	紫式部 murasaki shikibu ●この著者/作者で早稲田大学図書館検索(OPAC)作者検索●
出版事項 Imprint	写。[室町後期] sha
形態 Description	1冊；19.5×17.8cm
内容等 Notes	研形本 書名は内容による 朱点あり 欠丁・乱丁あり 箱入 列帖装 印記：九層文庫 中野益一旧蔵
キーワード Keywords	古典籍 / 九層文庫 / 源氏物語
公開権 Copyright	早稲田大学図書館 (Waseda University Library)



12

インターネットの画像公開

国立国会図書館デジタルコレクション

古典籍はもとより実見困難な複製のデジタル画像まで



13

インターネットの画像公開

国文学研究資料館新日本古典籍データベース

画像の質にばらつきがあるが、提供数はどんどん増えている



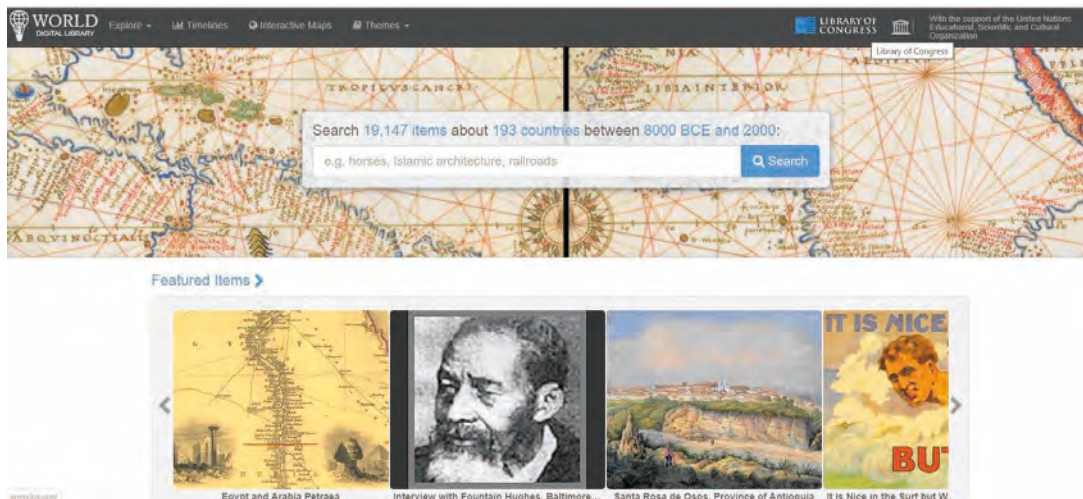
14

インターネットの画像公開

アメリカ議会図書館・ユネスコ

World Digital Library

The World Digital Library provides free access to manuscripts, rare books, maps, photographs, and other important cultural documents from all countries and cultures, in Arabic, Chinese, English, French, Portuguese, Russian and Spanish.



15

研究の質的变化に伴う必要とされる情報の変化

- 本文解釈研究が中心であった時代は、**とりあえず読めればよい**、という提供物が多かった。
- デジタル技術の発達や、インターネットの普及などにより、膨大な数の書物画像に接することができるようになって、日本国内のみならず、海外の日本研究の質が変化してきて、**画像の質や付加情報にも変化が求められるようになってきた**。
- **校訂本文や翻刻本文を底本によって確認する**といったタイプの「可逆性」からさらに進んで、**画像から「本物」の書物を理解する**タイプの「可逆性」が求められる時代になってきている。
- 従来の複製・影印などで提供されていた「書誌情報」は貧弱なものが多く、それでは新しい様々な研究のためには不十分であるという状況が生じている。
- 詳しく正確な書誌情報は、2次元画像を3次元の本物へと「可逆」する上で、不可欠な要素なのである。
- またインターネットの情報は、何時でも再確認できるという「可逆性」も必須である。リンク切れなどだけではなく、ブラウザが変わると、見ることができない状況が実際に起こっている。

16

これからの「可逆性」に必要な情報

- デジタルオブジェクト識別子(Digital Object Identifier、略称DOI)。ブラウザ等に入力するとコンテンツの所在情報(URI)に変換され、コンテンツにリンクを張ることで、リンク切れを防ぐ。
- 書物に保存されたテキストを理解するのに、その書物の物質的な特徴から得られる情報を活用して、テキストの性質をも判断する研究が進みつつある。それを行う為には、書誌情報は詳しくれば詳しいほどよい。デジタル画像に多くの情報を付与するのは、かなり困難なことであるのは確かである。

⇒ カラー画像を見て判断できることはあまり必要ない。

あらまほしい情報: 寸法・字面高さ・丁数(折数)・紙質・推定書写刊行年代

+α: 蔵書印の翻字(ネット検索のためにも)

後の手が加わっているかどうか(替表紙・改装) etc.

* 高望みであることは承知しているが、理想を提示することも大切であろう。

* またこのようなことを可能にする為に必要なのは、書物の本物を識っている書誌学を身につけた研究者である。今後バーチャルな世界が広がっていけばいくほど、それをフィジカル・リアルな世界に繋ぎ止める方法が必要になっていく。書物の世界では、その方法こそが「**書誌学**」なのである。

17

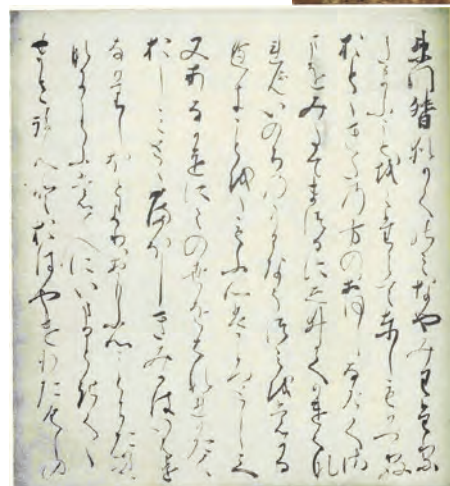
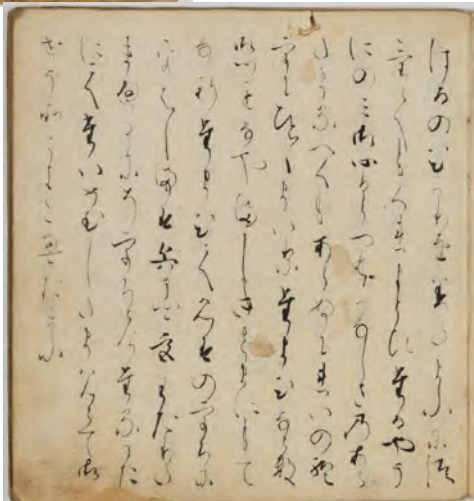
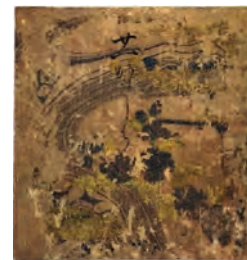
飯沼山圓福寺蔵「幻」と天理図書館蔵「柏木」



「地面高さ」の必要性

筆跡は異なるものの、表紙模様が共通

『天理図書館善本叢書和書之部30
源氏物語諸本集二』(八木書店、1987)



18

話し言葉調査の「可逆性」

人間文化研究機構国立国語研究所准教授 高 田 智 和

話し言葉（音声で表現される言）は、書き言葉（文字を使って書き留められた言語）に比べて不安定なものである。しかし、方言をはじめとする日常の言葉を調べようとするならば、話し言葉の不安定さを克服しなければならない。近年は、録音や録画をすることで、話し言葉を記録し、文字に書き起こしてインデックスとし、分析をすることができる。一回ごとに消えていく音声に「再現可能性」を与えることができる。また、記録された音声は、複数人で共有したり、時間を隔てて何度も聞くことができ、分析結果の検証や、新たな研究テーマのもとでの再利用の機会をもたらす。そのため、話し言葉調査では、音声を記録し、それを保存していくことが重要である。

過去の話し言葉調査においては、そもそも録音をしていなかったり、保存の過程で録音が失われたりすることがある。そのような場合であっても、調査票や書き起こしが残されていれば、「再現可能性」は後退するが、これらを利用して検証や再分析を行うことになる。調査票や書き起こしには、調査当時の調査者や分析者の書き込みが残されており、これらを適切に利用するためには、分析対象が話し言葉であるにもかかわらず、記録素材が紙であるがゆえに、文献研究的な手法も必要になる。

話し言葉調査の「可逆性」

高田 智和
国立国語研究所

1

発表の概要

- 話し言葉調査における言語の記録
 - 過去の話し言葉調査の利用
 - これらを通して「可逆性」について考える。
- ※ 発表者は「可逆性」を「再現可能性」あるいは「検証可能性」と解釈している。

2

話し言葉調査における言語の記録

3

話し言葉と書き言葉

- 話し言葉(口頭語)
 - 専ら音声言語による表現
 - 原則として、消える
 - 個人差、集団差が大きい
 - 主たる採取先は、人間
- 書き言葉(文章語)
 - 専ら文字言語による表現
 - 記録媒体の残存とともに、残る
 - 個人差、集団差が小さい
 - 主たる採取先は、書籍等の紙

4

話し言葉調査の対象と収集方法

- 対象
 - 地域方言
 - 社会方言 幼児の言葉、若者言葉
 - 待遇表現 敬語、配慮表現
 - 非母語話者(学習者)の言語
- 収集方法
 - 面接法
 - 談話収録
 - ロールプレイ

5

話し言葉の言語の記録

- 面接で得られた回答や談話を記録する。
- 近年は、録音や録画。
- 録音や録画を書き起こす(文字化)。
- 必要に応じて、タグをつけ、構造化。
- 調査対象者の属性(性別、年齢、出身地、居住地、母語など)、設定場面、収集日、調査者情報なども記録の対象。

6

書き起こしの例

- フヨナノ ナエデカンドー ホレハ ユギ フツド
アレダッス シャキショーシューダ ワラスコ°
ドツダナ オドゴデモ オナ〈ゴデモ〉。
- 冬など 何はともあれ ほれ 雪〔が〕 降ると
あれだし 百姓衆達〔は〕 藁仕事だよな 男
でも 女でも

『国立国語研究所資料集 10 方言談話資料 (1) ー山形・群馬・
長野ー』(1978年)より

7

録音・録画の保存の必要性

- 録音・録画は一次資料
- さまざまな記録媒体
 - 磁気テープ、カセットテープ、DAT、MDなど
 - 8mmフィルム、VHSビデオテープなど
- 問題点
 - 記録媒体は劣化する。
 - 再生機器は生産中止になる。
- 一次資料の保存・継承のための媒体変換

8

国立国語研究所研究資料室の場合

- 音声 WAV形式 MP3形式
- 映像 AVI形式 MP4形式
- 所蔵音声・映像データベース
 - 所内ネットワーク限定の試視聴システム
 - 申請により来館利用可能
- 国立国語研究所 研究資料室収蔵資料
 - <https://rnr.ninjal.ac.jp/>
 - 音声・映像ファイル目録を公開
- 今後の課題は、デジタル化音声・映像の保存・継承

9

音声・映像ファイル目録
Catalogue of Audio and Video Files

音声・映像ファイルID	種別	タイトル	備考	資料群ID	変換媒体	原媒体
hdl10004-91_001K	音声	鶴岡91-K: インフォーマント No.1		fo0244	dt04000	ct00029
hdl10004-91_001P	音声	鶴岡91-P: インフォーマント No.001		fo0244	dt05288	ct00728
hdl10004-91_002K	音声	鶴岡91-K: インフォーマント No.2		fo0244	dt04880	ct00830
hdl10004-91_002P	音声	鶴岡91-P: インフォーマント No.002		fo0244	dt05288	ct00729
hdl10004-91_003K	音声	鶴岡91-K: インフォーマント No.3		fo0244	dt04890	ct00831
hdl10004-91_003P	音声	鶴岡91-P: インフォーマント No.003		fo0244	dt05290	ct00730
hdl10004-91_004P	音声	鶴岡91-P: インフォーマント No.004		fo0244	dt05281	ct00731
hdl10004-91_005K	音声	鶴岡91-K: インフォーマント No.5		fo0244	dt04891	ct00832
hdl10004-91_005P	音声	鶴岡91-P: インフォーマント No.005		fo0244	dt05292	ct00732
hdl10004-91_006K	音声	鶴岡91-K: インフォーマント No.6		fo0244	dt04892	ct00833
hdl10004-91_006P	音声	鶴岡91-P: インフォーマント No.006		fo0244	dt05293	ct00733
hdl10004-91_007K	音声	鶴岡91-K: インフォーマント No.7		fo0244	dt04893	ct00834
hdl10004-91_007P	音声	鶴岡91-P: インフォーマント No.007		fo0244	dt05294	ct00734

<https://rnr.ninjal.ac.jp/audiovideofiles/>

10

過去の話し言葉調査の利用

11

鶴岡調査

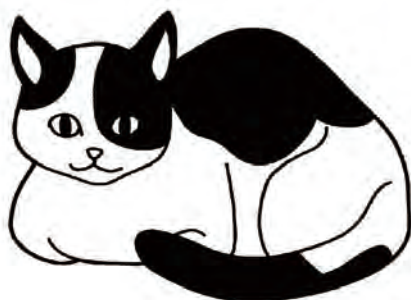
- 山形県鶴岡市
- 共通語化調査
 - 音声・アクセント、文法、語彙
 - 言語意識、言語生活
- 1950年、1971年、1991年、2011年の20年おき4回にわたる経年調査(同一の調査項目)
- ランダムサンプリング調査とパネル調査(同一人物の追跡調査)

12

言語形式を尋ねる面接調査

- 調査者：(刺激図を見せて)カードに何が書かれているかお答えください。 [調査録音例](#)

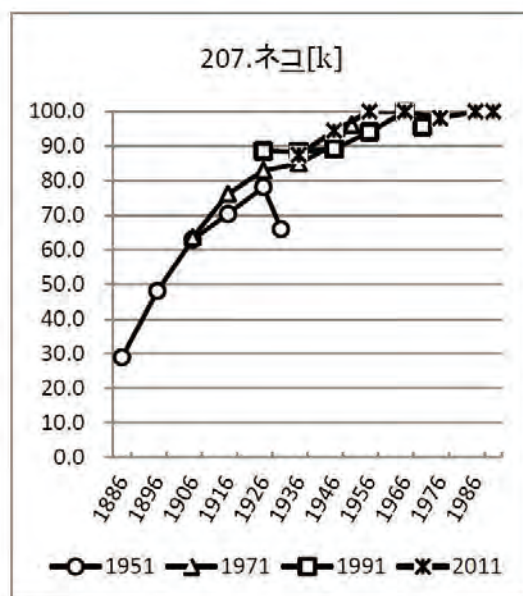
207



13

調査年別・生年コウホート別共通語化率

生年コウホート	1951	1971	1991	2011
1882-1891	28.9			
1892-1901	48.1			
1902-1911	63.0	63.6		
1912-1921	70.4	76.2		
1922-1931	78.2	83.0	88.6	
1932-1941	66.0	85.0	88.5	87.4
1942-1951		90.5	89.2	94.5
1952-1961		96.2	94.1	100.0
1962-1971			100.0	100.0
1972-1981			95.7	98.1
1982-1991				100.0
1992-1996				100.0

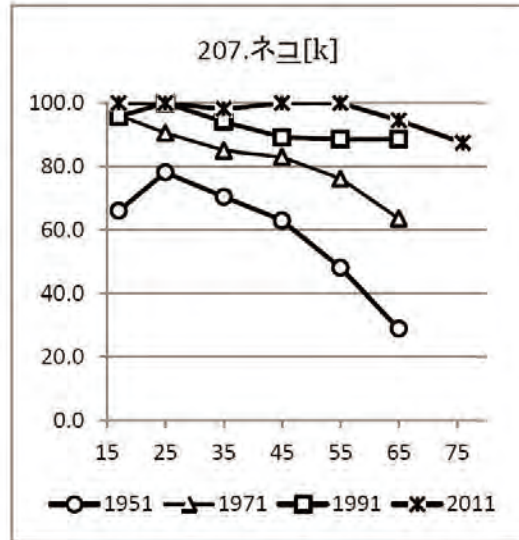


国立国語研究所編『地域社会の言語生活—鶴岡における20年間隔3回の継続調査—』(2007年)より

14

調査年別・年代別共通語化率

年代	1951	1971	1991	2011
15-19歳	66.0	96.2	95.7	100.0
20-29歳	78.2	90.5	100.0	100.0
30-39歳	70.4	85.0	94.1	98.1
40-49歳	63.0	83.0	89.2	100.0
50-59歳	48.1	76.2	88.5	100.0
60-69歳	28.9	63.6	88.6	94.5
70-79歳	-	-	-	87.4
15-69歳	63	82	92	98
15-79歳	-	-	-	96



国立国語研究所編『地域社会の言語生活—鶴岡における20年間隔3回の継続調査—』(2007年)より

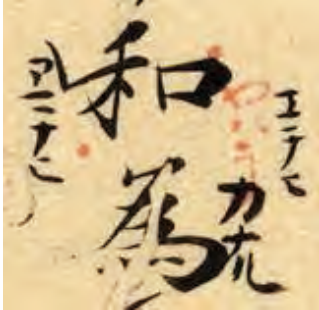
15

鶴岡調査の記録

	調査対象者数	調査票	調査票複製	録音	録音複製
1950年調査	496	現存	TIFF、PDF	録音していない	/
1971年調査	508	現存	TIFF、PDF	カセットテープ(数件)	DAT、WAV、MP3
1991年調査	713	現存	TIFF、PDF	カセットテープ	DAT、WAV、MP3
2011年調査	792	現存	PDF	音声ファイル(WAV、MP3)	/

16

訓点資料



書入れの順番推定

1. 朱点 ヲコト点、「ヤハラク」
2. 墨点① 「カナル」
3. 墨点② 「(左)アマナヒ」?
4. 墨点③ 「(右)アマナヒ」?
5. 墨点④ 合点(「(右)アマナヒ」)

17

1950年調査の調査票

「あの人はいつも遅れてくる」というとき、「いつも」ということをふつうなんとおっしゃいますか。	¹ itsumo. ² totsumo. ³ [ibudem]	1	66	1
「わたくしが留守番をしています」というとき、「留守番」ということをふつう何とおっしゃいますか。	¹ rusuban ² josiri ³ [rusu]	3	67	9
「どうぞこちらへいらっしゃい」といねいに言うとき、「いらっしゃい」ということをふつう何とおっしゃいますか。	¹ irajai ² gozake ³ []	1	68	1
「おながが一はいいになった。もうたくさんです」というとき、「もう」ということをふつう何とおっしゃいますか。	¹ mo: ² ado ³ []	2	69	3
「あんまり大きいので驚いた」というとき、「驚いた」ということをふつう何とおっしゃいますか。	¹ odoroita. ² obogeda. ³ [tamogeda] ¹ bi'kuuri sita ² kimoçeda	4	70	*2

18

1950年調査の調査票の書入れ

- 黒鉛筆 実査時の調査者
- 赤鉛筆(2種) 実査後の集計者・分析者
- 青ペン 誰？

19

1971年調査の調査票

24.1. 「あの人はいつも遅れてくる」という時「いつも」ということをふつう何とおっしゃいますか。

1	イツモ	2	シヨツチムー	3	トース、トーション	9[]	/
---	-----	---	--------	---	-----------	------	---

24.2. 「わたくしが留守番をしています」という時、「留守番」ということをふつう何とおっしゃいますか。

1	ルスパン	2	ルスイ	3	ヨスリ	9[]	/
---	------	---	-----	---	-----	------	---

24.3. 「どうぞとちらへいらっしゃい」といねいという時、「いらっしゃい」ということをふつう何とおっしゃいますか。

1	イラッシャイ	3	コザヘ	4	ハイッテクネヘン	9[]	4
---	--------	---	-----	---	----------	------	---

24.4. 「くたびれたもう歩けない」という時の「もう」ということをふつう何とおっしゃいますか。

1	モー	3	アド	9[]	3
---	----	---	----	------	---

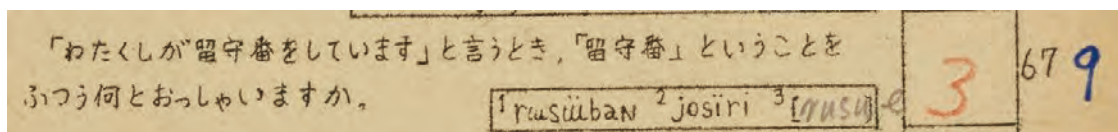
24.5. うしろから急にワッと大きい声をかけられた時の感じですが、「アッ、どりした」とおっしゃいますか。

1	オドロイタ	2	ビツクリシタ	3	オボゲタ	4	ハトモウ	9[]	2
---	-------	---	--------	---	------	---	------	------	---

↑老-2

20

青ペンの正体



242. 「わたくしが留守番をしています」という時、「留守番」ということをふ
つう何とおっしゃいますか。

1 ルスパン	2 ルスイ	3 ヨスリ	4 [rusu]	/
--------	-------	-------	----------	---

1971年調査時の回答分類コードを記入

⇒ 記入者は1971年調査時の集計者・分析者

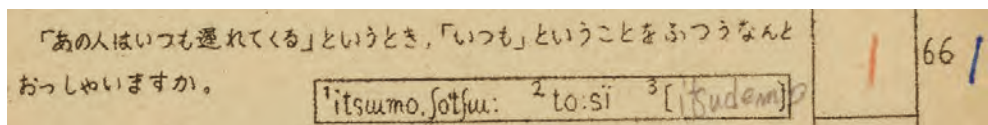
21

1950年調査の調査票の書入れ

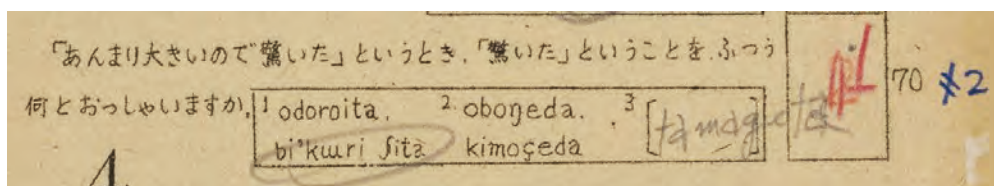
- 黒鉛筆 1950年実査時の調査者
- 赤鉛筆(2種)
1950年実査後の集計者・分析者
(濃い赤鉛筆で校正)
- 青ペン 1971年調査時の集計者・分析者

22

1950年調査時の判定基準(集計基準)



- 分類コード「1」は共通語的語形
(分類コード「2」は鶴岡方言的語形)



- 複数回答の場合、共通語的語形が回答されていれば、分類コード「1」を適用

23

鶴岡調査データベース ver 3.0

- 従来は報告書で集計データを提供
- 統計分析には回答データが有用
- <https://www2.ninjal.ac.jp/longitudinal/tsuruoka.html>
- 1950年調査(第1回)、1971年調査(第2回)、1991年調査(第3回)、2011年調査(第4回)の属性項目と音声・音韻項目を収録
- 調査票と録音は、申請により来館利用可能

24

まとめ

- 調査によって得られた話し言葉データの「可逆性」を担保するものは、一次資料である録音・録画と、その解釈データ(書き起こしや回答分類データ)。
- 過去の話し言葉調査で録音・録画がない場合、再現性は後退するが、解釈データが「可逆性」を担保するものとなる。
- 話し言葉調査データの保存・継承・公開・公開維持、すなわち「可逆性」の確保には、資本力、技術力、個人情報保護の遵守が必要。

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：関 野 樹（国際日本文化研究センター教授）
パネラー：天 野 絵里子（京都大学学術研究支援室 リサーチ・アドミニストレーター）
佐々木 孝 浩（慶應義塾大学附属研究所斯道文庫長・教授）
高 田 智 和（国立国語研究所准教授）
大 内 英 範（人間文化研究機構総合情報発信センター准教授）

〈はじめに〉

（関野） それでは、パネルディスカッションに入りたいと思います。司会を務めますのは、国際日本文化研究センターの関野です。よろしくお願いいたします。

まずは、ご登壇いただいた各先生方に対してフロアからご質問などがあれば、受け付けたいと思います。どなた宛てでも結構ですので、質問やコメントなどはありますでしょうか。

（質問者 1） どの話も面白くて、ありがとうございました。悩むところは一緒だなと思いました。僕は自然科学の方なのですが、記録としては紙のものなども結構あったりしますし、ペンの書き入れなども結構あって、これはどうしたらいいのだろうと、やはりそうやって「推理をしながら、当時の人の気持ちに返りながら」ということがあるなと思いました。

個々の研究はすごく面白くて、また勉強したいのですが、天野さんのご講演の最後に「終活では遅い」というお話がありました。僕はまだ終活にはだいぶあるつもりなのですが、前に僕がお世話になった先生などがやはり辞められるときにどっさりデータを残されるのです。ですから、僕がこれから考えるにしても、既に直面しているときにどうしたらいいかということで、「そうなる前に備えろ」ではなくて、今あるときにどういふことができそうかというのは、研究データマネジメントのお話の中で何かありますか。

（天野） 多分、研究成果をたくさん出された先生が退官するときになって一番まずい状態としては、その成果が自分の研究室にしかないもので出された場合だと思います。思い当たる節がある方もここにはいらっしゃるのではないかと思います。例えば自分が古本屋で買った古典籍など、自分が所有しているもので出した成果、それから自分が何年も通ってアクセスできた個人所有のものに依拠して出した成果などはあると思います。そういったものの中でも、日本にここしかないだろうというもの、海外に行って複製してきたような資料などはかなり貴重なものだと思いますから、まずそういう一番エクストリームなものから「終活」の対象に考えていただくことから始まるのではないかと思います。

（関野） 他にご質問はいかがでしょうか。フロアからないようでしたら、登壇者の間で質

問などがありますでしょうか。この後も、質問は適宜受け付けたいと思いますので、また何かありましたら、よろしくお願ひいたします。

〈可逆性の検証実例について〉

(関野) では、ディスカッションに入っていこうと思います。登壇された皆さまからお話を伺ってきたところ、高田先生の定義によれば、「可逆性」とは、検証可能性であるかどうかといった意味で捉えられるとの考えが示されました。そこで、「検証可能」という点について、皆さまからご意見をお伺ひしたいと思います。

たとえば、検証が実際に必要になった事例というのは本当にあるのかどうかをお伺ひします。そういった事例が今後、どのような形でデータを検証可能な状態に持っていけばいいのかという糸口になっていくのだろうと思います。そこで、実際にそういう検証が必要になるような事例が実際の現場で今までにあったのかをまずお伺ひしてみたいと思います。

(高田) 実は今日の最後にお話しした鶴岡調査は検証したのです。ですから、あのようなことをやっているわけです。つまり、かつての報告と数値がずれるのです。しかし、大勢では変わりません。あれは何もないと、報告書に書いた集計データを信じるしかないのですが、調査票が残っていたので、今回、全て見て検証しました。自分の研究で使ったということで、ある意味では再利用例と言ってもいいかもしれません。また、外部の研究者の方が時々来られて、「このように言っているけれども、音は本当にこうなのだろうか」「本当にこの音があるのか」と実際に聞いてみることはあります。

(佐々木) ご質問とは少しずれてしまうのかもしれませんが、高田先生のお話を伺っていて、われわれも書物を調査したときに調査カードというものを取って、要するに3次元のものを文字化して整理するのです。それで、その調査をどこまで取るかというデータの問題も画像とは別にあるわけですが、可逆性と言えるかどうかは分からないのですけれども、やはり優れた調査者が取ったカードは、その文字を読むだけで、頭の中にその本がどういう本かというイメージがふわっと浮かび上がってくるのです。それで、実際にその本に出会ったときに「ああ、これだ」と分かるようなことがあります。ですから、文字に落とす能力というのもすごいものだなと思ったりもするのです。

それにつけて必要というか、言い残したようなことなのですが、デジタル化がどんどん進んでいくと、デジタルしか見ない人がたくさん増えてしまうというか、それで済ませてしまうと思います。しかし、現物を見たことがあって、そこから画像を見て、自分の頭の中で復元するというか、元に戻す能力を持っていないと、あくまでも2次元の画像だけでやってしまったら、いろいろと問題が起こってくるのではないかと危惧しています。それで補足としてお話しさせていただきました。

(天野) 直接のお答えにはならないのですが、図書館員としてレファレンスを担当していた時に、検証の行為には日々接していました。例えば、研究者が引用文献を見てその成果の検証をしたい場合、その文献資料がどこにあるか探すのを手助けするのも図書館員

の仕事だったわけですが、引用文献の書誌事項の書き方が不十分な場合や間違っている場合が多いのです。ですから、書いてあるとおりに探しても見つからないことも多々あります。5が6になっているなど、数字が間違っていたり、巻（号）が間違っていたり、年が間違っていることがすごく多いです。今までそういう状態でしたから、もっときちんと他人が同じものにアクセスできる、検証できるということを意識して、今後、引用文献の書き方一つ取っても、研究者が身につけるべきメソッドとして良い方向になっていったらと思います。

〈デジタル化のメリット・デメリットについて〉

(関野) ありがとうございます。今、佐々木先生がおっしゃっていましたが、デジタル化でデジタルしか見ない人が増えてきたというのは、今の流れの一つであるかと思います。検証や可逆性という点では、デジタルになることによって落ちてしまうもの、失われてしまうものも恐らくいろいろあるのだろうと想像するのです。一方で高田先生のお話にもあったような、テープの再生環境がなくなったとかの問題がデジタル化によって解決するというメリットもあるのだと思います。

そういったメリットとデメリットについて、皆さまの分野、もしくは皆さまが扱っている資料については、どういったことが考えられるでしょうか。機構の中でも研究資源共有化事業の研究会がありますが、その辺が今後、資源を共有していく、もしくは生かしていくという点でも重要になっていくと思います。デジタル化によって失われていくものや、デジタル化のメリットをもう少し整理してお伺いできたらと思うのですが、いかがでしょうか。大内さんからお願いします。

(大内) 登壇者でもないのに端っこに座ってしまして、すみません。私自身はデジタル化されるのはすごくありがたいと思っています。私も佐々木先生と割と分野が近くて、写本を見ないといけない立場なのですが、やはりなかなか見に行くことができないものがデジタル化されるというのは、すごくありがたい話ではあるのです。ところが、先生もおっしゃっていたのですが、「デジタル化されたので、そちらを見てください」となって現物が見られなくなることが割と多いのです。最近写本をたくさん持っている図書館でも「この本はデジタルがあるから、こちらで画像を見てください」という、図書館の中でデジタル画像を見るコーナーがあったりします。

なぜそれが困るかという、やはり現物を見ないと分からない情報がたくさんあるのです。例えば袋とじになっている写本だと、その袋の中まで見たいわけですが、それがデジタルになると、できなくなってしまいます。それはわれわれからすると非常にデメリットです。一般的にはすごくメリットが多いと思うのですが、中にはそういうデメリットもあると私は思っています。

(高田) 音の場合ですが、やはりメリットとしては、今はデジタル化したので、所内にいれば、取りあえずさっと聞くことができるようになったというのが一番のメリットです。実は、以前は職員でも聞けなかったのです。目録にはあるのだけれども、これは一体どういう録音なのかというものがすぐ分かるようになったのがメリットです。

もう一つのメリットは、実は調査の最初から最後まで録音しているので、調査票をたどっていくと、その人がどのように聞いたかということも分かるのです。つまり、標準語で問い掛けたのか、方言ちっくに問い掛けたのかによって引き出し方が違うので、そのようなことが分かるようになったのがメリットです。

デメリットは、自分の調査が聞かれてしまって恥ずかしいということです。あまり学術的ではなくて、すみません。

(佐々木) 今、大内先生が「袋の中まで見たい」とおっしゃったことに近いのですが、現物を見ないと分からないことにどういうことがあるのかということ、例えば表紙が付け替わっているのかどうかは、現物を見ると、いろいろとチェックポイントがあつてすぐに分かるのです。しかし、画像だけだと判断に悩んでしまったりすることがあります。表紙が元かどうかは、その本の制作年代、制作された場所やグループなどを判断するのに結構重要な情報なのですが、結局、テキストが読めさえすればいいという発想ではそういうところまで分からないことがあつたりします。

あとは簡単に、やはり3次元というか、立体のものなので、和本などは下小口という冊子の下の部分に書名が書いてあります。昔は棚に保存していたので、そこを見て取り出してきたりしたのですが、下小口を撮ったデジタル画像はないと思うのです。しかし、本によっては下小口にしか書名がないことがあつて、ここに意外に重要な情報があつたりするのです。

そういうものはやはり調べた段階で、この本は単に上から水平に写すだけではなくて下小口なども必要だということをつかっていた上で、そのような画像を併せて撮影する必要があります。そのようなきめ細かい対応はなかなか大変だと思うのですが、特に良い資料や重要な資料は、最近では美術品などは立体画像で撮影されていたりするので、書物もそのようなことが必要なのかもしれません。

(天野) 図書館という研究基盤を整備する立場から言うと、デジタル化を進めてきたおかげで、研究機関が所蔵している資料・資源を研究者や一般市民の方がいろいろな形で使ってくれるようになった、利活用が進んだということがメリットだと思っています。ニーズも広まってきて、いろいろな使われ方ができるようになってきたことは私たちの喜びです。

デメリットとしては、出せば出すほど「もっと、もっと」と言われるというところです。先ほどの佐々木先生のお話されていた「字面高さ」が欲しい情報だと言われると、それをメタデータに付けるというニーズが発生するわけですが、それに応えるのは結構大変です。「まだまだ」「もっと、もっと」と言われるのがデメリットかと思います。

(関野) ありがとうございます。フロアからもご意見を伺ってみたいと思います。それでは永村先生、お願いいたします。

(永村) 人間文化研究機構の永村です。実は今回の「『可逆性』の実現のために」というテーマを決めた理由としては、人文系の世界では判断の拠り所が個性的なのです。私は歴

史が専門ですが、これは佐々木先生と同じように、調査をして、史料の時代判定をするわけですね。時代判定では紙の質を見て、かざしてみても、字体や墨の色で判断していくわけですね。今までは時代判定で「これは奈良時代です」と言われたときに、実は誰がそれをやったかでほとんど通用してきたのです。「〇〇先生が奈良時代と言ったから、奈良時代なのです」ということです。これをきちんと検証するためにはどうすればよいのかが重要な課題になっています。そのときに、人文系の分野でおのおのどういうやり方で、その聖域を壊していくのかが重要な問題になってきます。

国文学、国語学、さまざまな分野でその聖域の破壊の重要な道具としてのデジタル化があります。デジタル化というのはその成果を共有できるわけで、これほど重要なことはありません。今まで見るができなかった史料を簡単に見ることができるのです。ただし、大きな問題があります。どこのデータベースも詳細な画像を提供していないのです。1 MB 以下というのが大体の基準で、それは要するにウェブで流すということが一つあります。それと同時に、やはりその二次的利用・三次的利用を嫌がる体質が文科系の分野にはあります。このような文科系の分野にある聖域を破ると同時に、その一種の抱え込みの体質をどういう形で破壊していくのか、そういうことでこのテーマを選ばせていただきました。

(関野) ありがとうございます。他にフロアからございますか。

(質問者 2) どうもありがとうございました。お伺いしたいところとして、デジタル化の話が可逆性の中で入ってきましたが、デジタル化を前提にして研究データの話をする、今、原本の維持が難しいと思うのですけれども、そうすると、そもそもどこまで原本の維持にエフォートをかけていくべきかということがすごく気になっています。要はデジタル化でできる限界を基準にすべきなのか、それとも原本がないと可逆性は担保できないから、原本はどうしても必要だという話になるのかということです。特に先ほど高田先生のお話にあったような、過去のものを見て、あれを全部たどって、字面高さなどを見ていくというのは、ものすごく研究者の能力をもってして可逆性がようやく担保されているという状況に見えるのですが、果たしてそこまでやらなければいけないのかという基準のところについて、もしご意見があれば、教えていただきたいと思います。

(関野) 高田先生、佐々木先生、いかがでしょうか。

(高田) 焼けるなどといったことでなければ、原本（オリジナル）がある以上は維持したいのです。私も本来は佐々木先生とかなり近い分野なので、原本が大好きなのです。そういうこともあり、話し言葉のデータであっても、ある以上は維持していきたいと考えています。それで、可能ならばアクセスして、データの修正・訂正をしていって、それが恐らく研究の発展なのだろうと思います。場合によっては過去の調査報告書を修正しなければいけないことが出てくるのですが、それは研究の発展であり、むしろリスクであると考えて、今はやっております。

(佐々木) なかなか難しいのですが、もちろん本物というか、原本を大事にして、できるだけそれを保存するというのは大前提であって、そのためにデジタル公開しているということがあるのですけれども、先ほど大内先生もおっしゃったように、公開したらもう現物を見せなくなるというのが本当に一番問題です。それは確かに守ることになるのかもしれませんが、やはり理解を妨げることにもなっていたりします。

「どこまで」とおっしゃいましたが、本当に研究はどんどんマニアックになるというか、進んでいけばいくほど要求度が高くなるという先ほどの天野先生のお話もそうですけれども、そういう要求は原本を見せてくれる状況があれば、別にデジタルには求めないわけですから。しかし、現物を見せない以上はデジタルで判断せざるを得なくて、そうするとそれを要求してしまうということで、どちらが先かという問題になったりするところもあるかと思います。

(関野) この点に関して、他に何かご意見などはございますか。

(質問者3) 今のデジタル化うんぬんと原本の関係について、デジタル化というのは文字が読み取れるように画像化するのが基本だと思うのですが、それだけではないというところが一つ大きいと思っています。例えば料紙研究、紙の研究などでは、斜光照明で見て、この紙はいつ作られたかを判断するための方法などもありました。重さや法量もそうなのですが、今は顕微鏡で撮影していたりして、そういったことをやるためには、今、公開している画像だけでは当然できないわけですから、原本をもう一回見せてほしいということで、所蔵元に史料を撮影させてほしいと言ったら、費用がかかるということでした。もう原本を出したくないというわけですから。例えばそのようにデジタル化することによって原本に触れなくなるという事例が実際にあるのですが、実はまだ原本でやることのあるのだということです。

もう一つ、今回は考古の観点の方がいらっしやいませんでしたが、例えば平城京の発掘現場へ行くと、掘っているところをずっとムービーで撮影しています。そして、3DCGで現場を再現するのです。発掘した木簡やごみなど、発掘現場について、掘った後の様子だけを図面に起こすのではなくて、実は掘っている段階から、これは何だったのかを記録していたりします。そういうことを考えると、もしかしたらデジタル化は単に今やったものだけでは終わらない世界が実は待っているかもしれないということを言いたいと思います。

(関野) ありがとうございます。

(質問者2) 先ほどの話で文脈の共有が不足していたところをお伝えするのですが、可逆性について、私は初めの天野さんのお話にあったように研究不正の文脈で可逆性の話をしているものだと思って質問させていただいたのですが、もちろん研究倫理というか、研究者として研究に努めていくというところで、紙の必要性を否定するものでは全くありません。不正を検証する上で、それをたどっていかねばいけません。ただ、集計データを信頼できないから全部たどっていくというと、際限なくその作業が発生して

いくわけですね。それはどこまでやるのかというのが元々の趣旨でした。ありがとうございます。

〈まとめ——人間文化研究機構／人間文化研究情報資源共有化システムに期待すること〉

(関野) ありがとうございます。そろそろパネルディスカッションを締めなければいけない時間なのですが、最後にご登壇いただいた皆さまにお伺いしたいと思います。人間文化研究機構としてはこうした研究会をこれまでも開催してきましたし、また、資源共有化システムという形でさまざまなデータを共有するような仕組みもつくってきました。それが何らかの形で、今回のテーマである「可逆性」に貢献してくれるのかもしれませんが。そういった今回のテーマも踏まえながら、人間文化研究機構や資源共有化システムに求めていくものが何かありましたら、一言ずつで構いませんので、ご意見を頂ければと思います。

(天野) 私の発表の最後の部分で申し上げたことでもあるのですが、やはり研究者それぞれが考えて行動しないといけない部分がすごく大きいです。その中で人間文化研究機構は博物館など、社会につながっている組織も含まれますし、あとは地球研のような学際的なことをやっている研究機関もあるということで、波及効果を考えるとその役割は大きいと思います。

聖域の打破というお話もありましたが、聖域という部分があるがために捏造が起こったり、不正が起こったりすることもあります。それを打破するためのオープンサイエンスや研究データ管理が必要ですし、それを裏返せば可逆性の担保ということになります。機構やここにいらっしゃる研究者の皆さんには、オープンサイエンスや研究データ管理の推進という面で期待するところはとてもたくさんあります。

(佐々木) 人間文化研究機構にはお世話になるばかりで、つつい甘えてしまうのですが、やはり近い研究をしているみんなが協力する、手助けするというか、先ほどのデータを取ることや、自分のところの範囲でできるものは情報を提供するなど、そういうことをもっと呼び掛けて、一体化してやっていけばいいのではないかと思います。今は「みんなで翻刻」など、いろいろありますが、そういうこともネットで簡単にできるような状況になってきたので、全体で発展していけばいいのではないかと考えています。あとは国文学研究資料館のビューアの使い勝手の改善は、最大の課題です。

(高田) 私は身内なので、人間文化研究機構に要望はできないのですが、先ほどの電子化のメリット・デメリットの話で、私はデメリットの方をはぐらかしたのですけれども、実はオリジナルの保存・維持と同時にデジタル化したものもまた維持していくので、コストが増えたということが、便利になった代償なのかもしれないのですが、実はそれがデメリットかと考えています。

(関野) ご登壇いただいた皆さま、どうもありがとうございました。それではディスカッ

ションを縮めたいと思います。

(大内) どうもありがとうございました。それでは最後に、人間文化研究機構総合情報発信センター情報部門長永村からご挨拶します。

(永村) 大変短い時間の研究会でしたが、先生方、どうもありがとうございました。大変ご無理を頂いたのですが、このように短い時であってもきちんとやっておかなければいけないテーマであると思っていました。先ほどなぜこのような企画をしたのかということでお話ししたように、例えばデジタルデータにしても、活字のデータでも、いったん公になりますと、それが独り歩きしてしまいます。しかし、今日の天野先生のご発言の中で情報の劣化という言葉があったのですが、これはどきっとする言葉で、やはり私たちは情報をいつもリフレッシュしながら活用していかなければいけません。

今日も国文系の話で、特定の出版社が出した本が研究の底本になっていくわけですが、それは情報の劣化ということの一つだと私は思っています。そのような形で情報を固定化していくことを避けていきたい、元に戻れということのをこれからも考えていきたいと思っています。

先生方には短時間でご講演とディスカッションをしていただき、関野先生にはまとめをしていただいたわけですが、これは年度内には報告書を出しますので、そちらをぜひともご覧いただければと思います。

今日は短時間ですが、本当に内容の濃い研究会となりました。皆さん、どうもありがとうございました。

人間文化研究情報資源共有化研究会報告集 10
**Proceedings of the Study on Information
Resources of the Human Science Vol.10**

令和2年3月発行

編集・発行 大学共同利用機関法人 人間文化研究機構
総合情報発信センター
〒105-0001 東京都港区虎ノ門4-3-13
ヒューリック神谷町ビル2F
Tel 03-6402-9200
URL : <https://www.nihu.jp/>

