



Title	JAPAN OPEN SCIENCE SUMMIT 2025開催報告
Author(s)	情報知識学会 オープンサイエンス・オープンデータ 研究部会; 芦野, 俊宏; Robertson, Mark 他
Citation	情報知識学会誌. 2025, 35(3), p. 368-388
Version Type	VoR
URL	https://hdl.handle.net/11094/105213
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

オープンサイエンス・オープンデータ研究部会報告

JAPAN OPEN SCIENCE SUMMIT 2025開催報告

Report on JAPAN OPEN SCIENCE SUMMIT 2025

情報知識学会オープンサイエンス・オープンデータ研究部会編

JSIK, Special Interest Group on Open Science and Open Data, ed.

E-mail: sig_os-od@ml.kanazawa-u.ac.jp

1 はじめに（芦野俊宏、東洋大学）

今年度から公的資金を得た研究の成果発表論文については即時のオープンアクセスが義務化された。また本年度は多くの学術機関にとってデータポリシー策定の期限とされている年度でもあり、多くの研究分野においてオープンサイエンスへの取り組みを具体化することが求められている。

このような中で、今回のJOSS2025では協賛セミナーを含めて30のセッションが開催され、平日5日間毎日6つのセッションが行われるというスケジュールとなった。今回の特徴はやはり、AIに関わる開発の開発が劇的に進む中で、AIの活用、AIとデータに関わるセッションが複数あったことであろう。FAIR原則（Findable, Accessible, Interoperable Reuseable）の提唱者の1人であるBarend Monsが数年前にFAIRを”Fully AI-Ready”とも述べていたが、これが重要性を増しつつある。

AIは学術研究にも大きな進歩をもたらすものと期待されるが、一方でデータポリシー、

データ倫理の議論にも多大な影響を及ぼしている。国際的には、700以上の機関が加盟するCoARA(Coalition for Advancing Research Assessment)にもワーキンググループERIP, Research Integrity Policy in Responsible Research Assessment for Data and Artificial Intelligenceが設けられるなど、今後ともデータ活用において議論が必要となるものと考えられる。

AIに関する議論は今回のJOSSにおいて新たに見られた話題であるが、やはり主要な話題は研究データマネジメントに関わる経験の共有やインフラに関わるものであった。

ほとんどのセッションがオンラインにて開催されたため、参加者数の算定は難しいが、zoomによる参加者は延べ2000程度、youtubeでの視聴(リアルタイム、動画視聴含む)もほぼ同数となった。30のセッションのうち9つについてセッション報告原稿が寄せられ、個人発表2件についてはいずれも予稿が投稿された。これらを部会報告として掲載する。なお、本報告の作成にあたって部会のメンバーである林正治氏、高田良宏氏、堀井洋氏のご協力を頂いたことを付記しておく

2 B4 JST-CHORUS Forum: Strategies to Ensure Transparency and Digital Tracking of Research Outputs (Mark Robertson, CHOR Inc.)

2.1 Overview

This JST/CHORUS Forum explored the evolving open science requirements for researchers. Under Japan's new open science policy, researchers seeking certain national awards must make their publications and research data openly accessible. Globally, funders are implementing similar mandates. To uphold the integrity of the scholarly record, researchers must confirm adherence to research protocols and disclose funding sources and grant details. During submission and publication, they must also include links and supporting data to ensure accessibility and transparency. This forum discussed strategies during the publication process to encourage compliance to ensure accuracy and to enable digital tracking of outputs.

Panelists

(Moderator) Kazuhiro Hayashi, Principal Senior Fellow and Director, Research Unit for Data Application, NISTEP

Haruki Nakamura, Professor Emeritus, Institute for Protein Research, Osaka University and EiC *Biophysics and Physicobiology*

Julie Glover, Executive Director, National Health and Medical Research Council, Australia (NHMRC)

Amanda French, ROR Technical Community Manager, Crossref

Tetsushi Ueta, Vice-Director and Professor,

Center for Administration of Information Technology, Tokushima University and EiC *Nonlinear Theory and Its Applications (NOLTA)*, IEICE

Jeff Lewandowski, Director of Publishing, American Physical Society

2.2 Presentations

Haruki Nakamura began by considering open data and science from the viewpoint of all stakeholders. He outlined the introduction of research in computational science in the 2000s to data-driven science in the 2010s, to AI-driven science now. For researchers he commented on some reluctance to pay APCs for open access, the current evaluation by publication metrics, and emphasized that research data must be open. For funders, universities, research institutes, and publishers he highlighted how research is evaluated, using as an example his own journal *Biophysics and Physicobiology*. Nakamura suggested that the public could become the largest future stakeholder in the evaluation of research output.

Next Julie Glover explained that NHMRC introduced a policy in 2014 requiring open access after a 12-month embargo and encouraging sharing research data was added in 2018. In 2022 immediate open access with CC BY licenses was made a requirement provided Aboriginal and Torres Strait Islander cultural rights are respected. She showed how open access publication by NHMRC-funded Australian-authors was above 70% in 2023. Currently NHMRC is considering how to incorporate more aspects of UNESCO's recommendations on open science policy, and

how to reward and recognize open science practices as a member of the Coalition for Advancing Research Assessment. Glover showed the Australian National Persistent Identifier (PID) Strategy to accelerate research impact through universally connected PIDs.

Next Amanda French explained why the Research Organization Registry (ROR) is a key part of open science. Open metadata enables tracking and analysis of open science and ROR as a global, community-led registry of open PIDs for research organizations has been established as a widely adopted standard. ROR makes information about research organizations clean, consistent, and easy to exchange among software systems. Not only can research outputs be reliably associated with those organizations, but also ROR is designed for research systems to be able to track research organization activity and for the analysis of open science. French highlighted how many scholarly systems are using ROR including CHORUS and NISTEP.

The improvement of open access and open data in journals published by academic societies in Japan was illustrated by Tetsushi Ueta with his *NOLTA*, *IEICE*. It was first published in 2010, is on J-STAGE, mandates ORCID IDs, publishes under CC BY-NC-ND licenses, and received its first Impact Factor in 2023. In its development the Information for Authors was updated. Open access requirements were implemented resulting in listing in the Directory of Open Access Journals (DOAJ) in 2024 and indexing in Scopus in 2025. Next will be an open data policy. Ueta summarized that the role of academic societies and publications is to increase the visibility of the research outputs

and to do so open access journals need to be properly structured with clear policies.

The final presentation addressed how the American Physical Society (APS) is using open science to maximize the return on investments in scientific research. Jeff Lewandowski introduced APS which has more than 50,000 members. Of its 17 journals 7 are fully open access and nine are hybrid. The Physical Review journals have been supporting open science for over 30 years since the launch of arXiv in 1991, the first fully open access journal in 1998, and a revised open access policy in 2009 allowing posting of accepted manuscripts after a 12-month embargo. Lewandowski said that further investment in open science will be with PIDs, enhanced metadata, open data policies, platform upgrades, and open peer review, with sustainable open access whilst maintaining quality and trust in peer review and workflows.

2.3 Discussion

Kazuhiro Hayashi showed a PID map as a potential strategy for Japan and moderated a discussion about the integration of PIDs globally and locally with collaboration involving all stakeholders across the scholarly research community. AI was raised as a risk to integrity and its value in tracking metadata.

Presentations and video recordings are available on the CHORUS event page here:

<https://www.chorusaccess.org/events/jst-chorus-forum-strategies-to-ensure-transparency-and-digital-tracking-of-research-outputs/>

3 B4 JST-CHORUS Forum: 研究成果の透明性確保とトラッキングのための戦略(浅野光基, 科学技術振興機構)

3.1 セッション概要

日本の新しいオープンサイエンス政策では、2025年度から新たに公募を行う競争的研究費によって創出された学術論文や根拠データの即時オープン化が義務づけられる。世界的にも、研究成果のオープンアクセス義務化の動きは加速している。このようなオープンサイエンスの流れにおいて、研究公正の維持が重視されており、研究者には、研究プロトコルの遵守、独創性の表明、資金の詳細の開示や研究に使用された設備などについて明示することが求められている。また、研究成果論文の投稿や出版の際には、研究成果の透明性やアクセス性を確保するために、リンク、永続的識別子 (PID)、根拠データの記載が求められている。

本フォーラムでは、国内外から講演者を招き、出版プロセスにおける研究者の責任について議論し、研究成果の透明性の確保と研究成果情報のトラッキング促進に向けた戦略を探った。

登壇者

林 和弘 (モデレーター)、文部科学省科学技術・学術政策研究所 (NISTEP) データ解析政策研究室長

中村 春木、大阪大学蛋白質研究所 名誉教授、*Biophysics and Physicobiology* 編集委員長

Julie Glover, Executive Director, National Health and Medical Research Council,

Australia (NHMRC)

Amanda French, ROR Technical Community Manager, Crossref

上田 哲史、徳島大学情報センター 副センター長・教授、*Nonlinear Theory and Its Applications (NOLTA)*, *IEICE* 編集委員長

Jeff Lewandowski, Director of Publishing, American Physical Society (APS)

3.2 講演

中村春木氏は、多様なステークホルダーの視点からオープンデータとオープンサイエンスについて考察した。冒頭では、2000年代の計算科学、2010年代のデータ駆動型科学、現在のAI駆動型科学という、科学の変遷について紹介した。続いて、オープンアクセス化のために論文掲載料 (APC) を支払うことに対する研究者の抵抗感や、論文出版に基づいた研究業績評価の現状を述べ、ステークホルダーが公正に評価できるよう、研究成果はオープンにすべきであると強調した。将来的には、市民が研究成果の評価における最大のステークホルダーになる可能性を示唆した。

Julie Glover氏は、NHMRCのオープンアクセスポリシーの歩みを紹介した。2014年に12カ月のエンバーゴ後に論文をオープンアクセスにすることを義務化し、2018年には研究データの共有を推奨する方針を追加。さらに2022年には、先住民族の文化的権利を尊重することを条件に、CC BYライセンスによる即時オープンアクセスを義務付けたと説明した。また、NHMRCから助成を受けたオーストラリア人著者による論文のオープンアクセス率が2023年には70%を超えたことを報告した。現在NHMRCは、UNESCOの「オープンサイエン

ス勧告」を政策に反映させる取り組みや、CoARAの一員として、オープンサイエンスの評価と促進のあり方を検討している。最後に、研究活動の加速を目指す国家PID戦略を紹介した。

Amanda French氏は、Research Organization Registry (ROR)がいかに関わりオープンサイエンスにとって重要な役割を果たしているかについて説明した。French氏は、RORが、研究機関を識別するためのオープンなPIDレジストリとして、幅広い領域で採用されていること、ソフトウェアシステム間で相互運用できるように提供されていること、研究成果と研究機関の紐付けだけでなく、オープンサイエンスに関する追跡・分析が可能となるように設計されていることを述べた。また、CHORUSやNISTEPを含む多くの学術システムがRORを使用していることを強調した。

上田哲史氏は、日本の学協会が発行するジャーナルにおけるオープンアクセスやオープンデータの改善の取り組みについて、自身が編集委員長を務めるNOLTA、IEICEを例に紹介した。2010年に創刊し、J-STAGEに掲載されている同誌が、論文投稿時にORCID IDの登録を必須としていること、CC BY-NC-NDライセンスで論文を公開していること、2023年に初めてJournal Impact Factorを取得したことを紹介した。その後、投稿規定の改定により、2024年にはDOAJ、2025年にはScopusに登載されたことを紹介した。次のステップとしてオープンデータポリシーの策定を挙げ、学協会と出版物の役割は研究成果の可視性を高めることであり、そのためにはオープンアクセスジャーナルが明確な方針で適切に設計・運営される必要があるとまとめた。

Jeff Lewandowski氏は、科学研究への投資効果を最大化するために、APSはオープンサイエンスを活用していると説明した。APSが、1991年のarXivの立ち上げ時からarXivでのプレプリント投稿を推奨していること、1998年に初の完全オープンアクセスジャーナルを出版したこと、2009年に12カ月のエンバーゴ後に原稿の公開を許可するオープンアクセス方針を導入したことなど、30年以上にわたりオープンサイエンスを支援してきたことを紹介した。最後に、今後のオープンサイエンスへの投資は、PIDの導入、メタデータの強化、オープンデータポリシーの策定、プラットフォームの改善、公開査読モデルに向けられると述べ、あわせて、査読やワークフローの品質と信頼を維持しつつ、持続可能なオープンアクセスを実現していくことの重要性を述べた。

3.3 ディスカッション

5名の講演後に、林和弘氏の司会によるパネルディスカッションが行われた。林氏は冒頭で、日本において将来的に戦略となり得るものとしてPIDマップを示した。続いて、学術コミュニティが協力してPIDをグローバルおよびローカルに統合していくことについて議論した。AIについては、研究公正に対するリスクと、メタデータの追跡における有用性の両側面が指摘された。

講演資料および録画は、以下のCHORUSイベントページから閲覧できる。

<https://www.chorusaccess.org/events/jst-chorus-forum-strategies-to-ensure-transparency-and-digital-tracking-of-research-outputs/>

4 B6 AI×文学研究の可能性を探る (村尾優子, 国立国会図書館)

4.1 セッション概要

近年、デジタル人文学の分野で、大規模言語モデル (LLM) をはじめとするAI技術の活用が進められている。国立国会図書館 (NDL) においても、AI技術を用いたOCR処理プログラムの研究開発を行うとともに、OCRによるデジタル化資料のテキスト化、本文検索サービスの提供等を進めている。これらの技術やその成果物が文学研究でも活用されつつあり、研究の効率化や深化の観点で、今後さらに大きな役割を果たしていく可能性もある。本セッションでは、デジタル技術を活用して文学を研究されている方々から、AI技術やLLMの可能性、研究のアイデア等を紹介いただいた。本セッションの資料は、当館の「NDLラボ」のページ (<https://lab.ndl.go.jp/event/joss2025/>) で公開している。

4.2 講演

最初に青池亨 (NDL) から、NDLOCR, NDL古典籍OCR等の技術や、次世代デジタルライブラリー, NDL Ngram Viewer等のテキストデータ関連サービスについて説明した。続いて、4人の研究者が順に登壇し、AI技術を活用した文学研究について紹介した。

近藤泰弘氏 (青山学院大学) からは、NDLが公開しているOCR全文テキストデータを取得し、生成AIを用いて活用する事例が紹介された。著作権保護期間が満了した『大日本国語辞典』初版を例に、テキストデータと位置を示す座標からなるOCRデータから本文テキストを見出し語と語釈が対応

する形式に復元する過程や、復元したテキストをMCP (Model Context Protocol) サーバに配置して生成AIから利用することでハルシネーションを避け、出所を明確にしつつ生成AIの古典知識の正確性を高められることが語られた。『大日本国語辞典』をAIで利用できることは後継辞書の改訂作業等に役立つ。またこの手法は辞書以外の資料についても有用との見解が示された。

北本朝展氏 (国立情報学研究所) からは、現代日本語での質問に対して江戸時代の古文風テキストで回答するチャットボット「からまる」について紹介された。現代日本語テキストを学習したLLMに、江戸時代の書物から構築したデータセットを継続学習させることで、江戸時代には存在しなかった事物についても古文風の表現に言い換えて回答できる。回答には江戸時代の世界観が反映されており、ある種の創造性や、元のLLMにはない謙虚さも見られる。AI技術により生成されたテキストを従来の作品と対比できる状況において、LLMとは何か、文学とは何かという根源的な問いにつながっていくことが指摘された。

黄晨雯氏 (大阪大学大学院) からは、Transformerに基づく深層学習モデルを用いて、小説投稿サイト「小説家になろう」のオンライン小説のジャンルを分類・可視化して分析する試みについて紹介された。BERT (Bidirectional Encoder Representations from Transformers) ベースの日本語に特化した深層学習モデルに、「小説家になろう」から取得したジャンル・あらすじ情報を学習させ、主成分分析等の手法で可視化した結果、ジャンルの特徴を捉えやすくなった。学習量の増加や、ジャンル・あ

らすじ以外の情報の学習により、考察の幅を広げられる可能性も示唆された。

日比嘉高氏（名古屋大学大学院）からは、NDLが公開している戦前期の小説を対象に、小説のジャンルを判定する試みが紹介された。OCR全文テキストデータから小説本文を復元した上で、本文データから小さな場面（チャンク）を切り出し、場面ごとにLLMで主題を自動抽出・分析した。この結果、トピックモデルLDAを用いた従来の分析に比べ、主題の抽出精度と解釈のしやすさが向上した。この手法は、従来の作品単位での研究とは異なる、場面主題の単位、間テキストの文学研究と位置付けられる。一方で、AIによる多数の作品の分析結果をどう検証するか、既存の知とどう交差させていくかといった課題も指摘された。

4.3 ディスカッション

近藤氏をモデレータとして北本氏、黄氏、日比氏の4人でディスカッションを行った。まず近藤氏から、適用しているAI技術が4人とも異なっていた点が指摘され、4人の研究者からAI技術選択の背景や意図について説明された。研究に関するデータの公開について参加者から質問が寄せられたことに対して、Hugging FaceやGitHubでの公開状況、今後の公開予定が紹介された。また、NDLが公開しているOCR全文テキストデータ（著作権保護期間満了分）から本文を復元して再配布することの可否について質問があり、NDLから、特に問題ない旨を回答した。

日比氏からは文学研究における作品テキストデータの精度が重要であることや、データ整備・共有が研究者の責務になると考えられることが指摘された。また、NDLが

公開しているOCR全文テキストデータが、レイアウトの構造化や読み順序の整序等の面で改良され、使い勝手の向上したOCRによるものに更新されることへの期待が語られた。

文学に活用できるAI技術の可能性や文学へのアプローチの多様性が感じられるセッションとなった。

参考情報

- [1] 次世代デジタルライブラリーのAPIについて、<https://lab.ndl.go.jp/service/tsugidigi/apiinfo/>（2025年7月10日参照）。
- [2] AI中納言を作る—ClaudeによるSQLコーパス操作—、<https://note.com/yhkondo/n/nac7f19127162>（2025年7月10日参照）。
- [3] 江戸時代の古文風テキストで会話できるチャットボット「からまる」を公開、<https://sakana.ai/karamaru/>（2025年7月10日参照）。
- [4] なろう小説APIから全ての作品情報を一括取得するPythonスクリプト、<https://github.com/karupoimou/NarouDataAllDownload>（2025年7月10日参照）。
- [5] 日比嘉高：トピックモデルを利用した近代日本小説のサブジャンル判定の試み、*跨境*, 19(1), 59-83, 2024, <https://doi.org/10.22628/bcjjl.2024.19.1.59>（2025年7月10日参照）

5 C1. 今、「オープンさ」を改めて問う(井出和希, 榎本啄杜, 長門裕介, 神崎隼人; 大阪大学)

5.1 セッション概要

近年, 「オープンサイエンス」の学術研究政策上の重要性が増し, 基盤整備と共に推進の機運が高まっている. 公的資金による研究に基づく学術論文や付随するデータの「即時オープンアクセス」も実現されようとするものの, 「オープンさ」のそもそもを考え語らう機会は少ない.

オープンであることは常によきことであろうか. オープン概念そのものを整理しながら, 結果としての「オープンさ」だけでなく, その過程についても議論を深めるべく本セッションを企画した.

注目され易い知的財産権や安全保障上の問題をこえて, 「社会への還元」といった論点に閉じることなく, これからの学術研究のあり方を考える機会とし, 参加者(最大111名)からのコメントも踏まえた討論を行った.

5.2 講演

最初に井出から導入として学術政策の動向を紹介すると共に問題提起を行った後, 「「オープンさ」の所在と倫理的・社会的課題」(榎本, 長門), 「「オープンさ」と知識の経済: ラテンアメリカから考える」(神崎)と題した話題提供に進んだ.

学術政策の動向を紹介するにあたっては, 前提にあるデータの扱いに関するFAIR原則[1]に触れ, G7科学技術大臣会合コミュニケや即時オープンアクセス方針の根底にある「すべきこと」としてのオープンさ向上を論じた. 近い将来これにインセン

ティブ付与された場合, 「よきこと」としてのオープンさが強化されることに対する懸念(=オープンでないことが悪しきことになっていく)を示すと共に, フレキシブルさの重要性を指摘した[2]. この背景には, 検証に要する時間の長さ, センシティブなトピックを扱うことによるハレーションの懸念, フィールドに住まう人々の生活への影響といったことが「オープンにしにくい研究を避ける」という動機を生み出し, 延いては学問の豊かさを失わせかねないといったことを想像することの肝要さがある.

所在と倫理的・社会的課題では, そもそもオープン概念とは一体何なのかということに注目し, オープンソースソフトウェアやオープンデータ, オープンアクセス, オープンサイエンスといったことが一文脈依存的ではあるものの—「アンチ囲い込み」という理念的モチーフとしての共通性を有すること—を述べた. その上で, アリストテレス的な中核依存的同名異義性 (core-dependent homonymy, CDH) の考えを用いることで秩序ある多様性を捉えることを提案した[3]. これを踏まえ, オープンさの部分(要素)や段階(程度)を許容することの重要性を強調しながら概念を定義した.

倫理的・社会的課題としては, 悪意ある利用、情報過多と品質管理の困難さ, 公共的利益と個別的利益のバランス, 不平等・格差の拡大といったことを挙げた.

「オープンさ」と知識の経済では, 人類学を背景にもつ神崎がラテンアメリカの社会, 経済, 文化からそのあり方を再考した.

サイエンスが西欧近代由来の特異的・特権的な知識形態の経済を形成してきたなかで, 他なる知識形態と価値から—ともすれば

無批判になり得る—オープンさを考える意義がある。独自に発展した／せざるを得なかった背景には、商業的学術情報出版の弱さ、リソースの乏しさ、スペイン語／ポストガル語の基盤といったことが挙げられる[4]。

そこでは、アマゾニア先住民の人々が自分たちに関する非／先住民研究者による学術論文を閲覧し、自分たちの権利要求のための証拠として活用するといった実践も為されている。これは、草の根運動と国家との部分的な結合（＝中央集権化に向かうと分散を志向し国家に抗すること）を示唆するものであり、「私たち」に対しても「義務化」のような強い力に対して自らその営みを問い直すことの意味の大きさを想起させる。

5.3 参加者からのコメントと応答、総括

以下に参加者からのコメントの内容を一部抜粋して応答を付す。

1) オープンアクセス比率は様々な課題を解決したり落としどころを探ったりしながら「100%を目指していくべき」かと問われたが、これについては「100%」であることを目的化したり、オープンであることが望ましいと決めつけたりせずに研究者の自律性に依拠した柔軟さを残すことが重要であると考えている。

2) 「市民科学」における上から目線のにおいに対して「もうひとつのサイエンス」（註：神崎発表に依る）はそれを乗り越える視点を感じたとの意見があった。これは正にオープンさ（の促進）においても同様に生じている課題であり、私たちがどのように在りたいかを示していくことの不可欠さを示唆するものでもあろう。

3) もっと人文社会科学の分野の人間が参画して多様なデータと価値観をそこ（オ

ープンサイエンスやリポジトリの促進）に組みこむべきとの指摘については、促進を前提におくかは留保するが、重要である半面、実務上で専門性を発揮してもよいと認識できるような施策が重要である。

他にも、オープンさが固着化する傾向を実感すると共にさまざまな具体の場におけるオープン化について語る場を広げていく必要があるように感じるとの見解も共有された。本企画がそのような場づくりと実践を行き来する機会の端緒になれば幸甚である。

謝辞

本報告の一部は、日本財団・大阪大学 感染症対策プロジェクト、科研費 若手研究（23K12845）の支援を受けて推進した研究に基づくものである。

参考文献

- [1] GO FAIR : FAIR principles , <https://www.go-fair.org/fair-principles/>, (2025年7月19日参照) .
- [2] Ide, Kazuki: “Japan can embrace open science — but flexible approaches are key”, *Nature*, Vol.629, No.8011, p.286, 2024.
- [3] 榎本啄杜, 長門裕介: 「多義的なオープン概念を整合的に理解する」, *Jxiv*, 2025, preprint, <https://doi.org/10.51094/jxiv.1299> (2025年7月19日参照) .
- [4] 村井友子: 「ラテンアメリカの学術情報プラットフォームの活動」, *ラテンアメリカ・レポート*, Vol. 38, No. 2, pp. 86-92, 2022.

6 D1「地域と連携した人文系学術研究データ生成の可能性と課題」 (堀井 洋, 合同会社AMANE)

6.1 セッション概要

古文書や民具など地域資料を対象とした学術研究データ生成については、これまで様々な議論・取り組みが行われてきた。特に、昨今の機械可読な学術研究データに対する需要の高まりにより、より大規模かつ効率的なデータ生成手法の確立が急務となっている。その一方で、地域の過疎化・人口減少が急速に進行する現状においては、市民の地域資料に対する興味・関心を喚起させ、地域文化に対する理解を継続的に促すことも重要な課題である。

合同会社AMANE（以下、AMANE）・長野県伊那市・みんなで翻刻プロジェクト（以下、みんなで翻刻）は、2024年9月に「地域資料デジタルアーカイブの公開にむけた包括的連携協定」を締結し、伊那高遠地域の古文書を対象とした「第1回古文書解読コンテスト」を実施した[1][2]。本セッションでは、本事例を中心に市民参加型の人文系学術研究データ生成の可能性と課題について、話題提供と議論を行った。

6.2 話題提供

本セッションでは、セッション提案者であるAMANE堀井による趣旨説明の後に、登壇者3名からの話題提供があった。

第一回：古文書解読コンテストについて (伊那市地域おこし協力隊 前田和弘)

前田氏からは、2024年10月から2025年1月末まで開催された古文書解読コンテストについて、話題提供があった。本コンテ

ストは、高遠藩の藩主内藤家やその家臣に関する古文書・古記録等資料834点を対象に、みんなで翻刻およびAMANE Archivesを利用してオンライン上で実施された[3]。その特徴としては、産学官連携に基づく実施体制や、上位優秀者が賞金を獲得する点などが挙げられる。加えて本発表では、地域おこし協力隊に所属し、かつクリエイターである前田氏が自治体と学術界を繋ぐ役割を担っていたこと、根幹となるコンセプトが「参加するだけで日本の課題解決になる知的な遊びを作ろう。」であったことなどが紹介された。前田氏は、今後に向けてイベントの全国展開や事業化を構想している。

みんなで翻刻と古文書解読コンテスト@伊那市（東京大学地震研究所 加納靖之）

加納氏からは、みんなで翻刻の紹介と同コンテストでの活動について報告があった。みんなで翻刻の初期は、京都大学古地震研究会に所属するメンバーが中心となり、地震や火山噴火等に関する史料を対象として遠隔地の専門家による翻刻（くずし字で記された文字を活字化）の実現を目指していた。その後、学術クラウドソーシングの考え方にに基づき、市民参加型史料翻刻プロジェクトとして2017年1月からサービスを公開し、2025年6月現在28プロジェクト、通算参加人数10,844人、翻刻文字数4,848万文字の成果を挙げている。現在もAI技術を適用した新バージョンのリリースやNDL古典籍OCRへのデータ提供など、積極的なデータおよびサービスのオープン化に取り組んでいる。市民参加による大量な翻刻データの獲得と、その成果のオープン化によるAI等の技術発展への貢献・フィ

ードバックは、“史資料・翻刻データの好循環”をもたらすことが期待される。

「地域資料デジタルアーカイブの公開にむけた包括的連携・協力に関する協定」におけるAMANE Archivesの役割（合同会社AMANE 小川歩美）

小川氏からは、多様な組織が連携する地域資料デジタルアーカイブの視点から、古文書解読コンテストにおける3組織の役割と連携の意義について発表があった。AMANEが2023年5月から社会貢献事業の一環として構築・公開しているAMANE Archivesは、2025年6月現在、6アーカイブ8コレクション約12,000点の資料について、IIIF形式で画像およびメタデータを公開している。アーカイブ公開の主な目的としては、学術的な試行および新たな知見の獲得や調査研究事業成果の逐次公開、資料公開を通じた地域・社会との連携が挙げられる。古文書解読コンテストでは、伊那市から提供されたメタデータおよび資料画像データをもとにAMANE ArchivesにおいてIIIF形式画像データを生成・公開し、みんなで翻刻がそれらを利用することで、イベントを実現した。従来の多くの事例では、資料所蔵機関がデジタルアーカイブの公開から活用事業の実施までの全てを担うことが一般的であったことに対して、本事例では多様な組織が互いに連携し役割分担をすることで、新たな取り組みを実現した。本連携形態は今後の発展・普及が期待される一方で、コンテスト開始時からの急激に増加するサーバーアクセスへの対応など、解決すべき課題が複数存在することも本発表において示された。

6.3 ディスカッション

ディスカッションパートでは、学術データ作成に参加する市民のハードルや、市民を学術プロジェクトに巻き込むための手法やブランディングについて、議論が行われた。さらに、生成AI技術が市民参加型学術プロジェクトに与える効果・影響はどのようなものか？という問いに対しては、加納氏からは、参加者がそれぞれにレベルに応じた翻刻をAI技術の支援により実現すること、前田氏からは、コンテストを実施する際のルールをAI技術の利用を前提としたものを検討する必要があること、などの発言があった。さらに、小川氏からは、デジタルアーカイブについても、対人向けのサービスと併せて、AI経由での利用も想定した構成を目指すことなどが示された。

6.4 まとめ

市民参加型の人文系学術研究データ生成については、学術研究分野のみならず一般社会からも非常に大きな関心と期待が寄せられている。その一方で、AI処理との役割分担など、新たに検討を進めるべき課題が山積している。今後もこれらの話題・課題について、議論する機会を設ける所存である。

参考文献

- [1] みんなで翻刻, <https://honkoku.org>
- [2] 古文書解読コンテスト, <https://www.komonjo-contest.com>
- [3] AMANE Archives, <https://ourarchives.amane-project.jp> (2025年7月31日参照)

7 E2 若手セッション: オープンサイエンスについて話そう! (大月希望, 東京大学史料編纂所)

7.1 セッション概要

JOSSは今回の開催で7回目を迎えるが、参加者、発表者とも若手が少なく、そもそもオープンサイエンスに関心を持つ若手も少ない。そこで、オープンサイエンスについて若手や若手を育てる人(教員等)の関心対象、取り組みを共有し交流したいという考えから、本セッションを企画するに至った。成果報告だけでなく経験や悩みを共有できる場をつくることを目的とした。

はじめに企画者の大月より趣旨説明を行った。若手セッションは今回初めての試みであり、まずは企画者の研究分野である人文系分野を中心に、発表、紹介、議論及び情報交換を行うことを目指した。理工系や社会科学系に比べ人文系のオープンデータは少なく、研究資源の独占も散見される。論文のオープンアクセス化など研究成果のオープン化も進みつつあるが不十分である。研究過程や研究データは一部でオープン化の動きは見られるものの、形式が定まっておらず参照可能性が低い。本セッションでは、オープンサイエンスに関わる実践や研究の内容と課題について、3組の登壇者より話題提供いただいた。

7.2 発表、紹介

第一に、合同会社AMANE(以下、AMANE)の小川歩美氏と佐々木紫帆氏から、AMANEが学術研究・社会貢献活動として構築・維持管理している地域と連携するデジタルアーカイブ「AMANE Archives」[1]についてと、

具体的な取り組みとして「奥州市学校資料」の調査・公開の流れの紹介が行われた。

第二に、総合研究大学院大学、国立情報学研究所他に所属する前山和喜氏[2]から、日本のコンピューティング関連資料について発表が行われた。資料が未保存のまま失われ、何が重要か不明なために残らないというオープン化以前の課題が指摘され、具体例として「使われた状態のパンチカード」が挙げられた。また、企業とNDA(秘密保持契約)を締結して研究を行う際の機密資料や流通に乗らない「グレー資料」の扱いについて、情報の信頼性確保やセキュリティ面での困難、公開ルールが未整備であることが課題として示された。

第三に、東京大学の小川潤氏から、「歴史研究プロセスの構造化による暗黙知の可視化と共有に向けて」と題した発表が行われた。小川氏は、歴史研究において膨大な作業やデータが明示されず、研究プロセス自体を記録する仕組みがないことを指摘した。この課題に対し、情報資源の生成過程を記述するためのデータモデル「PROV Ontology」を歴史研究プロセスに特化する形で拡張したデータスキーマの事例研究が紹介された[3]。

7.3 ディスカッション

ディスカッションでは、資料のオープン化にかかわる課題について議論が行われた。企業所有データに関して完全なオープン化の前に研究目的に限定した公開を行うことは可能かという質問には、コンピュータ関連企業と歴史研究者がNDAを締結して研究する事例がほぼなく、実践例を作り議論の促進を試みていることが前山氏から

示された。関連して AMANE の小川歩美氏、佐々木氏から、「AMANE Archives」での具体的な工夫が紹介された。資料群全体を把握しつつも公開可能な資料からデジタル化して公開する方針をとり、個人情報を含む資料については表紙のみの撮影やマスク処理、管理用と公開用のメタデータの分離などの工夫を行っている。一方で、建築図面などのように企業が権利を有している場合やセキュリティの問題で公開が難しい資料もある点が指摘された。

データ流通に向けた取り組みに関する質問には、特定地域に限定されない連携型アーカイブやメタデータについて、既存の横断型アーカイブでの扱われ方を検討している旨がAMANEから返答があった。

また、研究プロセスの構造化に関して、特定の史資料を直接的に用いた研究の前段階の、史資料収集・整理といった行為についても詳細な記述が可能なのかという質問が出された。これに対し小川潤氏からは、問いが明確でない段階での史資料収集や収集・整理を主目的とするケースについても記述し、コンテキストも含めて歴史研究のプロセスの中に取り込んでいくことができると考えていることが示された。

7.4 まとめ

本セッションでは、人文系分野におけるオープンサイエンスの推進には、資料の特性に応じた保存・公開ルールが必要であること、研究プロセスの構造化により分析過程の再現性を高め、研究行為の分析を可能にする仕組みが示された。また若手ならではの視点で、人材不足の原因として計算機系分野で学芸員資格を取得できる大学が

少なく、理工系資料の保存公開にはまず専門的な教育の基盤が必要なこと、歴史研究よりも最新技術が重視され就職も有利な分野であることが指摘され、それを踏まえて分野間の相互理解を深めたいという意見が出された。理工系分野における歴史・文化的資料の保存と公開には、人文系分野とはまた異なる困難があることが示唆された。今回紹介された逐次公開や研究プロセス構造化が何らかの形で役立つかもしれない、今後も交流ができればと思う。

本セッションでは若手として高校生や高校教員の発表・参加も期待されたが、今回は定期テストや学校行事により叶わなかった。また大学学部生にも積極的に参加していただきたいと考えている。来年度はより広く発表を呼びかけ、参加者獲得に向けた広報の拡大を試みたい。本稿をお読みの方で関心をお持ちの方はご連絡をいただければ幸いである。

参考情報

[1] 合同会社AMANE : AMANE Archives, <https://ourarchives.amane-project.jp> (2025/07/29参照) .

[2] 前山和喜 : 【連載】計算の歴史学とジェンダー—誰が計算をしていたのか？, <https://bungaku-report.com/computing.html> (2025/07/29参照) .

[3] 小川潤 : 「『行為 (Activity) 』概念に基づく歴史研究プロセス構造化オントロジーの検討：歴史研究の再現性・再検証性の確保に向けて」, *Library and Information Science*, No. 93, pp. 1-22, 2025, DOI : 10.46895/lis.93.1 (2025/07/29参照) .

8 E4 シビックテックにおける参加のデザイン～オープンソース・オープンデータが拓く共創～（瀬戸寿一・白松俊・白澤美幸，東京大学CSIS）

8.1 セッション概要

近年，オープンソース技術やオープンデータの活用が進み，行政や市民活動の現場においても新たな可能性が広がっている．そこで本セッションは，大学・行政・企業・市民団体との協力を通じて，オープンデータやオープンソース技術などを活用した新しい参加と共創の場をどのように形成できるか，シビックテック分野での活動実績を多数有する話題提供者の実践事例を通じて，そのあり方や課題を議論した．

8.2 話題提供

話題提供に先立って，瀬戸寿一（駒澤大学/東京大学CSIS）より，2024年度に開設された東京大学「シビックテック・デザイン学創成寄附研究部門（CTDI）」の活動概要と本セッションの趣旨説明が行われた．シビックテックは「市民が主体的に情報技術やデータを通じて身近な地域課題や社会課題の解決を目指す活動」と位置づけられ，日本でも東日本大震災やオープンデータ政策の推進を一つの契機として，全国各地で様々な実践型活動が行われている．

陣内一樹（一般社団法人コード・フォー・ジャパン）は，「シビックテックにおける場づくりと共創」と題して，2013年に設立されたコード・フォー・ジャパンの活動経過について振り返ったうえで，シビックテ

ック活動の実践事例として，これまで延べ70回以上の実績がある「ソーシャルハックデー（SHD）」における運営の仕組みや工夫について紹介し，継続的な活動の場を提供する重要性を強調した．SHDが契機となった事例として「アフターピル検索サービス」や，流山地域の市民主体であるエコポイント「もりポ」，さらには偽情報対策ツールとして開発された「BirdXplorer」など，シビックテックの社会実装モデルの諸要素を説明した．

榎本真美（パシフィックコンサルタンツ株式会社/一般社団法人 社会基盤情報流通推進協議会）は，「土木・まちづくり分野におけるデータドリブンな共創の挑戦」と題して，おもに社会インフラに関わるデータ利活用や，デジタルを駆使したコミュニケーションについて，広島県インフラマネジメント基盤「DoboX」などの例を基に言及した．社会インフラに関わるデータ・知識がよりオープンになることで，多様なステイクホルダーの協働が生まれ，透明性・再現性のある意思決定やエンパワーメントにつながる可能性がある一方で，専門性が高く大容量データという特性上，データ操作に関わる障壁や，専門家や行政と市民との意識ギャップをシビックテック的な取り組みからどう埋めていけるかが課題であると説いた．

小泉和久（シビックテック・ウラシマ）は「地域と一緒にシビックテック活動」と題して，浦安市を対象に様々なオープンデータを駆使しながらデジタルマップ（視覚）化することで，地域活動へ活用し対話のツールになる可能性を紹介した．この活動では，地元でのまち歩きや各種ワークショップを経て立ち上がってくる新たな（地理空

間) データや事例も多く、行政と住民という関係性だけでなく、地域のいろいろな諸団体とのコラボレーションが、シビックテック活動から創発されている例として「景観を巡るまち歩き」・「親子de通学路探検隊」・「うらやすドリームクエスト」などが例示された。

白松俊 (名古屋工業大学/東京大学CSIS 特任教授) は「LLMが拓く市民共創の新しい可能性」と題して、主に対話による合意形成や市民同士のコラボレーションを支援するためのAI技術として、LLM (大規模言語モデル) を活用したシステムの研究および実践例を紹介した。LLMを対話の場に導入することで、多様な立場のペルソナを生成することができ、例えば議論をシミュレーションすることや、対話の場に足りない観点 (ステイクホルダー) からの意見を生成できるなどの具体的な効果を挙げた。これらは、人同士の合意形成を行う上で課題となる対立感情を極力抑えた形で、LLM等の技術が「社会の結び目」になりうる可能性を示唆するものである。

8.3 ディスカッション

4 件の話題提供を受けて、白澤美幸 (東京大学 CSIS/一般社団法人コード・フォー・ジャパン) の進行により、主に参加者から登壇者への質問に答える形で「オープンな参加型プロジェクトを通じた社会変革の可能性と課題」を議論した。

このパートでは、「Slido」(オンラインの Q&A 投稿ツール) を介して、参加者から様々なコメントが寄せられた。その一例は以下のとおりである：

・シビックテックに取り組むには何 (ハード・ソフト・環境) が必要か？

・(登壇者が) シビックテックに興味を持ったきっかけはなにか？

・シビックテック始めたいけど、ハードルがまだあるなと思います。どうしたらいいですか？

これらの意見については、シビック「テック」という用語が付くことで、参加の心理的障壁が発生しうる点をどのように軽減できるか、各登壇者が実践を行う上での創意工夫について紹介し合った。この点はシチズンサイエンス活動とも共通する部分が多く、活動の場やきっかけをどのように開いたものにできるかが鍵となる。また、シビックテックを通じた参加のあり方を問う意見は次の通りであった：

・シビックテックのような参加型の文化では、意見をうまく表現できる人や十分な時間に余裕のある人など、いわゆる恵まれてる層の意見が吸い上げられやすい構造にありませんか？

・まちの課題を持っていない市民も参加できるような環境があればいいですね。いまは、まちの課題を持っている人たちだけの集まりのような気がします。

これらの意見に対しては、シビックテック自体の認知度・普及度が日本で途上である一方で、オープンデータやオープンソース文化といった活動の「透明性」を重視して取り組まれた点を踏まえて、継続的な活動を行う重要性が強調された。シビックテックを担う人材の発掘・育成については、東京大学「シビックテック・デザイン学創成寄附研究部門

(CTDI)」の一環として講座の試行や教科書を制作予定であることも付言する。

9 F1 研究データをより見つけやすくするためのメタデータ変換と学術機関リポジトリへの登録(能勢正仁、名古屋市立大学)

9.1 セッション概要

研究データの検索は、一般的に専門性の高いホームページからのみ利用が可能な状況であり、専門外のより広い研究コミュニティや一般市民に対しては、「Findability」はそれほど高くはない。研究境界領域における融合研究や市民科学を促進するためには、メタデータデータベースの被検索性を向上させる必要がある。こうした問題に対し、(1)専門分野のメタデータスキーマで記述されたメタデータを、広く利用されている汎用メタデータスキーマに従うメタデータ形式に変換し、(2)変換したメタデータを、学術機関リポジトリに登録する、ことで解決を試みた。このセッションでは、上記取り組みの報告に加え、メタデータ互換性に関わる話題を提供し、研究データを見つけやすくし、その流通・利用を促進するための方策について議論を行った。

9.2 講演

当日は、セッションの趣旨説明 (doi:10.18999/2013002) の後、次のような5つの講演が行われた。

- 新堀 淳樹、大平 司、直江 千寿子 (名古屋大学) 「名古屋大学における研究データの可視化・検索向上を目指したメタデータ変換と機関リポジトリへの登録について」 (doi:10.18999/2013003)

大学間連携事業IUGONETでは、太陽地球系物理学分野でよく用いられるSPASEデー

タスキーマを採用し、現在では1200以上のメタデータを作成している。名古屋大学はIUGONETに参画し、284件のメタデータを登録してきた。これらのメタデータをIUGONETや宇宙地球科学分野内にとどまらず、広く検索されるようにし、研究データがより発見・利用され易くなるように、メタデータ形式をSPASEからJPCOARに変換した。そして、変換したメタデータを名古屋大学機関リポジトリ (NAGOYA Repository) に登録した。これらのメタデータは、他のメタデータデータベースからハーベストされ、NII学術機関リポジトリデータベースやGoogle Dataset Searchなどでも検索可能となっている。また、メタデータスキーマのバージョンアップに対応して、メタデータ対応表を改修し、継続登録ができるようにメンテナンスを行っている。今後は、開発したツールの共用や異分野データベースの調査など、学術機関リポジトリへの登録に関して学術分野コミュニティ内の情報共有や連携を強化したい。

- 阿部 修司、芦北 卓也 (九州大学) 「研究データの発見・検索性向上に向けたメタデータ変換と学術機関リポジトリへの登録：九州大学における適用事例」 (doi:10.18999/2013004)

九州大学国際宇宙惑星環境研究センターでは、世界各地で取得した地磁気観測データについてSPASEメタデータを作成し、IUGONETに登録を行ってきた。これらのメタデータを、名古屋大学と同様の手法でJPCOARメタデータ形式に変換し、九州大学の機関リポジトリ (QIR) に180件登録した。QIRでは、国内の多くの機関リポジトリで採用されているJAIR Cloudとは異なるシステムを用いているが、そうした機関でも

展開可能であることを実証した。また、変換にあたってはGakuNin RDMやChatGPT等のツールを独自に活用し、迅速でスムーズな実施を行った。

- 三上 絢子（JPCOARコンテンツ流通促進作業部会/室蘭工業大学）「JPCOARと名古屋大学の研究用メタデータマッピングプロジェクトでの協働について」（doi:10.18999/2013005）

名古屋大学と九州大学で行っているメタデータ形式の変換と機関リポジトリへの登録に関して、JPCOARコンテンツ流通促進作業部会では、メタデータスキーマバージョンアップの提供とプロジェクト事例紹介により、協働を行ってきた。

研究データの利活用の促進は、JPCOARの活動でも意識されてきており、研究データ公開の研究者側のメリットを伝えるのは重要である。今回の事例紹介 (<https://schema.irdb.nii.ac.jp/ja/case-spase-nu>) では、実際に分野メタデータを公開した事例について、具体的な公開方法や公開によるメリットについて紹介できる貴重な機会であった。今後も様々なニーズに対応するJPCOARスキーマ活用事例の共有を行う計画である。

- 藤原 一毅（国立情報学研究所）「未公開のメタデータを流用したい～NBDCヒトDBから包括的未病DBへ～」（doi:10.18999/2013006）

ヒトに関する様々なデータを共有するためのプラットフォームであるNBDCヒトデータベースでは、BioProject形式のメタデータが作成される。一方、ムーンショット目標2「超早期に疾患の予測・予防をすることができる社会の実現」では、包括的未病データベースのメタデータがGakuNin

RDM上で作成されている。研究者がNBDCにメタデータを登録する際には、基本的な情報をエクセルシートに記入し、それをNBDCの登録担当者がBioProject形式のメタデータに変換するが、同時に、そのエクセルシートを使ってGakuNin RDM上の未病データベースへ登録する手順を開発した。変換プログラムは、pythonにより実装し、Jupyterノートブック上で実行する。メタデータの変換および登録をad hocなハードコーディングで実施した例であり、こうしたアプローチもありうる。

- 叢 艶、青木 学聡（名古屋大学）「分野別・汎用メタデータ互換性のスクリーニング調査」（doi:10.18999/2013007）

「研究データエコシステム構築事業」においては、ルール・ガイドライン整備の一つとして、機械可読データの標準化が挙げられている。異なる分野間でメタデータの変換効率を高め、相互流通を拡大していくには、様々なメタデータスキーマの共通項目の抽出と比較が不可欠なためである。そこで、分野別メタデータの仕様書のあり方を調査し、各仕様書に記載されているタイトル、作成者、作成日時などの記述項目を抽出し、それらとDataCiteメタデータスキーマとのマッピング対応関係の構築を目標とした。

125件の分野別メタデータスキーマから、対応関係が導けそうな8件のメタデータスキーマを対象とし、DataCiteメタデータスキーマの必須項目へのマッピングの可否を判断した。今後は、これに基づき、メタデータの互換性に関する作成方針の策定、マッピングにおける不足項目の補完手法、および変換ツールの開発について、検討を進めていく予定である。

10 F2 異分野の研究データをつなぐ: 統合検索システムの構築とメタデータの課題 (朝岡誠, 国立情報学研究所)

10.1 セッション概要

近年、オープンサイエンスの推進に伴い、政策や研究資金提供機関の要請、データ駆動型研究の発展などを背景に、研究データの公開と共有の重要性が高まっている。しかし、研究データは機関や分野ごとに異なるメタデータ標準で管理されており、横断的な検索や利活用は容易ではない。

本セッションでは、東京大学社会科学研究所の南山氏による司会のもと、異なる機関・分野の研究データを統合的に検索可能なデータカタログの運用をテーマに、統合検索システムの構築意義とその効果について議論を行った。メタデータの標準化などの技術的課題にとどまらず、多言語化やデータカタログの利用実績評価、そして異分野連携における法的・倫理的課題についても検討した。

10.2 講演

最初に国立情報学研究所の朝岡からは、人文学社会科学分野における統合データカタログ「JDCat」の事例が紹介された。社会科学では、構造化された量的データ(例: 調査票に基づく数値データ)が中心であり、変数情報や回答選択肢を含むメタデータの整備によって、再利用や二次分析が比較的容易に行える。一方、人文学分野では、非構造的な一次資料、特に画像データが中心であり、OCRによる文字認識や人物・地名のタグ付けといった精緻なメタデータ

付与が必要とされる。JDCatでは、DDI (Data Documentation Initiative) に基づいた独自のスキーマを採用し、31の項目を設定し、多言語対応(日本語・英語)、統制語彙による専門用語検索、Jupyter Hubによる分析環境の連携などが実装されている。今後の課題として、知名度向上や分析ツールとのさらなる統合、関連文書やコードのリンク、データ提供機関の拡大が挙げられた。

次に、国立遺伝学研究所の有田氏から、生命科学・生物学分野におけるデータカタログ整備の課題が報告された。すでに整備が進んでいる同分野においても、著作権や分類体系の変化といった運用上の課題が存在する。特に、和名と学名の対応付けや、学名の動的な変更への対応が実務上の大きな困難となっている。INSDC(国際塩基配列データベース連携)では、世界各国の論文に掲載される全塩基配列データが日米欧の連携により公開されており、現在20PBを超える大規模データが共有されている。一方で、生物多様性条約(CBD)による「利益配分(ABS)」の義務が、従来の無償公開を原則とするデジタルデータにも適用されることになった点は、学術データベースの運用方針に大きな影響を与える。特に、生物資源が国家の資産と見なされることで、国際的な合意形成と法的調整が求められる状況にあることが共有された。

最後にROIS-DS極域環境データサイエンスセンターの田中氏から、超高層大気データを対象とする大学間連携プロジェクト「IUGONET」と、その知見を活用した異分野データ統合への展開として「AMIDER」プロジェクトについて紹介された。IUGONETでは、太陽地球系科学における多

様な観測データを統合し、SPASE (Space Physics Archive Search and Extract) メタデータスキーマを拡張することで、学際的な解析に対応するデータ基盤を整備した。AMIDERではさらに、極域環境に関する生物・地質・気象データなどを統合的に扱うため、ISO 19139形式のメタデータにも対応したデータベースを構築した。ユーザーフレンドリーなインターフェース、キーワードによる検索支援、関連データのスコア表示など、多機能的設計が特徴である。今後は、ExcelからXMLを生成するメタデータ登録ツールや、自然言語処理を用いた多言語化支援、テキストマイニングによる関連性抽出の仕組みなども開発予定である。

10.3 ディスカッション

ディスカッションでは、能勢氏のモデレーションのもと、有田氏、田中氏、朝岡の3名が参加し、データカタログ運営に関する3つの課題について討論が行われた。

1つ目は、多言語化および翻訳の課題である。人文社会系では、文書タイトルや歴史的概念の翻訳が困難であるとの意見があり、むしろローマ字表記など原文保持が望ましい場合もある。一方、生命科学では学名と和名の厳密な対応が求められるなど、翻訳の要件そのものが分野ごとに異なる。いずれにせよ、専門知識を要する用語の揺らぎ、文化的背景の違いをどう扱うかが共通の課題として浮かび上がった。

2つ目は、データカタログの利用実績評価と維持に関する課題である。JDCatではアクセスログの大半が機械的（スクレイピング等）であるため、実利用者数の把握が困難であることが報告された。IUGONETでは1万人規模のユニークユーザーや論文で

の謝辞記載という実績がありながらも、安定的な財政支援の確保は難しく、データ基盤整備の価値が定量的に評価されづらいという構造的問題が指摘された。

3つ目は異分野連携の可能性と技術的課題である。生物学分野のデータカタログでは、学名を軸にして考古学文書や環境データとリンクさせるといった応用が議論される一方で、メタデータスキーマの多重管理やバージョン追従の煩雑さが連携の障壁となっている。特に、生物多様性条約により導入された利益配分原則は、従来の「自由な学術情報流通」とのギャップを際立たせ、異分野連携において予期せぬ法的・倫理的課題があることが浮き彫りになった。

本シンポジウムを通じて明らかになったのは、研究データ統合が単なる技術的課題ではなく、文化・制度・法制度を横断する複雑な営みであるということである。多様な分野の実践者が互いの前提を理解し、共通の語彙や技術を開発していくためには、継続的な対話と基盤整備のための支援体制が不可欠であることが、改めて認識された機会となった。

11 G2 データ管理専門職のペルソナ設計（南山泰之，東京大学社会科学研究所）

11.1 セッション概要

近年，政策的な側面から大学・研究機関によるデータ管理の要請が強まりつつある．2024年2月に公表された「学術論文等の即時オープンアクセス実現に向けた基本方針」では，所定の競争的研究費によって生成された学術論文及び根拠データは，学術雑誌へ掲載後即時に機関リポジトリ等の情報基盤への掲載が義務づけられており，機関リポジトリ実務者にとっては根拠データの登録が現実問題として迫ってきている．一方で，日本ではデータライブラリアン，データキュレーターなどの専門職が根付いておらず，機関リポジトリの担当者はデータ管理や公開といったタスクに対して，どのような態度で臨めばよいのか具体的なイメージを持てていない．さらに，データ管理は分野によっても流儀が異なるため，個別対応がどの程度発生するのか手探りの状況が続いている．

このような現状把握をもとに，本セッションではワークショップ形式で各分野のデータ管理者のペルソナを作成し，データ管理に求められる専門家としての人物像を明らかにすることを試みる．本セッションは，小林裕太氏及び熊崎由衣氏（千葉大学附属図書館）との共同企画である．

11.2 話題提供

本セッションの前半では，データ管理に明るい各分野の研究者として，八塚茂氏（独立行政法人製品評価技術基盤機構），福田陽子氏（国立研究開発法人国立環境研究所），杉原孝充氏（国立研究開発法人海

洋研究開発機構）の3名から，ご自身の日常的なデータ管理の実態に関する話題提供があった．八塚氏からは，生命科学系データベースの構築・運用やメタデータの付与といった業務経験に基づく話題提供があった．業務に必要な知識・スキルとしては，データベースの基礎知識やプログラミングスキルが挙げられた．また，バイオ専門でなくとも，利用者目線でのデータベース構築や，不明点の調査・コミュニケーション能力があれば対応可能な業務も多いこと，専門用語については利用者に分かりやすく伝える能力が重要であること，の2点が強調された．

福田氏からは，環境科学分野におけるデータ公開支援や基盤整備，問い合わせ対応といった業務経験に基づく話題提供があった．業務に必要な知識・スキルとしては，情報工学やプログラミングの知識に加え，ワークショップ参加者の分析からはコミュニケーション能力，業務における協働や分担，時間管理，学習能力といった総合的な側面が挙げられた．加えて，マニュアルに縛られない柔軟な対応や，ニーズに応じた情報提供など，研究経験に基づいたデータ利用者の視点が強みとなる旨言及があった．

杉原氏からは，海洋科学分野におけるデータ管理部署のマネジメント業務経験に基づく話題提供があった．業務に必要な知識・スキルとしては，全体フローの理解や分析方法に関する深い知見のほか，チーム内のリソースを俯瞰して作業を割り当てるマネジメント能力が挙げられた．また，データ管理業務をより客観視する観点から，過去に経験した近隣分野（宇宙探査）の知見も役立っている旨言及があった．

11.3 ディスカッション

セッションの後半では、各グループが作成したペルソナを会場全体で共有し、参加者を交えてディスカッションが行われた。

ディスカッションでは、大学図書館が多様な分野の研究データを扱う際のドメイン知識の深さや、マネジメント層の戦略的なアプローチの重要性など、実務的な関心を中心にした議論が展開された。前者の話題となる専門用語の理解に関しては、組織やコミュニティの段階によって必要なレベルが異なるとの意見があった。扱う専門分野が限定され、データ管理がルーチンとして確立された組織では一定以上の専門用語に対する理解が求められる一方で、幅広い分野を扱う組織では緩やかな連携やハブ的な役割のほうが重要視されるのではないかと指摘があった。

関連して、「研究データ管理業界」で用いられる専門用語を研究者に説明するためには、一定程度の経験値が必要であるという課題意識も共有された。この点、大学図書館員が研究データ管理に今後関わっていく際には、情報処理技術への関心や、レファレンスやガイダンスで培ったコミュニケーションスキルが生かせるのではないかと可能性が示唆された。

11.4 まとめ

本セッションでは、複数分野におけるデータ管理者の実務経験を題材に、データ管理に求められる専門家を明らかにすることを試みた。「専門用語を解釈し、利用者へ分かりやすく伝える」業務はどの話題提供者からも言及があり、データ管理専門職の重要な職務と言えそうである。データ提供者と利用者の間に存在するギャップを理解し、個人としてのドメイン知識や研

究経験、及び組織やコミュニティが保有しているノウハウの両方を駆使して課題解決にあたる態度が求められよう。本セッションでの議論が、大学・研究機関におけるデータライブラリアン、データキュレーターの在り方を探る契機になることを期待したい。

11.5 謝辞

本セッションの企画に当たっては、研究データ利活用協議会（RDUF）ジャパンデータリポジトリネットワーク推進部会（JDARN）から多大な支援を受けた。特に、白井知子氏（国立研究開発法人国立環境研究所）、福田和代氏（国立研究開発法人海洋研究開発機構）からは企画趣旨に関する貴重な助言をいただいた。ここに記して謝意を示す。

11.6 参考情報

- [1] 八塚茂. 生命科学におけるオープンデータの理想と現実 .
<https://doi.org/10.6084/m9.figshare.5682847.v3>
- [2] 福田陽子. NIES における研究データ管理の支援業務 .
<https://doi.org/10.34462/0002000238>
- [3] 杉原孝充. 海洋研究開発機構（JAMSTEC）でのデータ管理と体制 -データ管理専門職のペルソナ設計：JAMSTEC の場合-.
<https://jir.repo.nii.ac.jp/records/2000344>
- [4] 小林裕太, 熊崎由衣. 研究データ管理専門職のペルソナ検討ワークシート.
<https://opac.ll.chiba-u.jp/da/curator/900123436/>