Slide 1｜Trends in Open Science in the Humanities

【00:00–00:30】

皆様こんにちは。国立情報学研究所の朝岡誠です。私は、オープンサイエンス基盤研究センターに所属し、研究データ公開のためのインフラの設計、運営を行っています。本日は、昨今の人文学における研究データ公開の動向について、国内外の取り組みをもとにお話したいと思います。なお、本日お話する内容は朝岡の個人的見解であり、国立情報学研究所の見解ではありません。

Slide 2｜報告の流れ

【00:37–02:00】

本日の流れは以下のとおりです。

まず背景として、学術界全体のオープンサイエンスの動向について押さえ、人文学において「どのような場面でデータ公開が求められているか」を押さえます。そこで、ジャーナルの対応、政策としてのデータ公開について確認します。

次に、人文学分野データ公開のための研究基盤を俯瞰します。欧米のインフラと、日本における研究インフラについて見ていきます。

そのうえで本題の研究データの公開について整理していきます。今日の研究データの公開にはDMP（データマネジメントプラン）の作成が大きく関わっており、計画段階から研究データ公開について検討する必要があること、研究データの公開には研究データの文脈などを記載したメタデータを作成する必要があること、当該研究データを公開できるかどうかは、研究者が自分で判断する必要があることについて説明します。最後に、研究データのインフラを設計する立場から、これからの研究データ公開の展望についてお話できますと幸いです。

Slide 3｜オープンサイエンス（OS）とは

【02:00–03:50】

ここでは、オープンサイエンス（OS）とは何かを、UNESCOが挙げる要素に沿って確認します。

まず、科学知識へのオープンアクセスです。オープンサイエンスといえば、論文、研究データなどをオープンに公開することを指しますが、これらの取り組みはオープンサイエンスの一部であり、オープンサイエンスには、他にも3つの要素があります。

2つ目の要素は研究インフラの整備となります。リポジトリ（アーカイブ）、研究データ管理システム、ディスカバリー用のデータベース、そして解析システムといった器が揃って初めて、公開された知識が実際に使えるようになります。本日の発表は、この研究成果物の公開と、それを支えるインフラに焦点を当てたものになります。

3つ目は、社会的アクターの関与。市民の参画や、政策立案者・企業との連携がOSの射程に含まれます。最後に、他の知識システムとの対話――先住民・地域の伝統的知識をはじめ、学術以外の知の取り込みも重要であり、人文学はこの要素と大きく関わる学術分野だと考えられます。オープンサイエンスとは、科学者と社会全体の利益のために、あらゆる分野の研究を誰もが利用できるようにするための実践です。

この定義を前提に、次のスライドからデータ公開とインフラの変化について説明します。

Slide 4｜研究データ公開環境の変化

【03:50–05:30】

ここでは、研究データ公開をめぐる環境の変化を確認します。ポイントは「IT以前」と「IT以後」です。

まず IT発達以前。分野の代表的データは研究界の公共財として保護されつつ、国際比較に資する一部データが共有される、という枠組みでした。そのため、共有される研究データは分野を代表するデータの共有が中心で、研究データは研究者個人、または研究グループの資産という意識が強かったと考えられます。

次に IT発達以後。地域や分野を越えて共有する前提が整い、電子化した文化資源の長期保存、そして研究データのダウンロード提供ができるようになりました。

研究者個人が他の分野、地域の研究データを利用できるようになった一方で、学術出版社や助成機関主導によるデータ公開が浸透してきております。

学術出版社は、自分たちが出版する研究論文の査読、公開要件の明確化のため、国は公的助成の正当性を示すために、論文とその論文の根拠となるデータを公開する動きが強まり、結果として、研究データ公開は“科学の常識”になりつつあります。

ここからは、出版社による研究データ公開の推奨、公的助成機関による研究データ公開義務の広がりについてみていきます。

Slide 5｜データ共有に言及する雑誌の推移（フランス）

【05:30–06:30】

このグラフは、「データ公開を推奨する学術雑誌が増えている」ことを示すデータです。

一部の学術雑誌では、その論文で引用した研究データにアクセスする方法を投稿論文に記載することを推奨、もしくは義務としていますが、この論文で引用されたデータアクセスについての記述をDAS（Data Availability Statement）とよびます。

フランスのオープンサイエンス指標を公開している French Open Science Monitorでは、このDAS を含む雑誌の割合の推移を公開していますが、このグラフを見てもわかるように、DASを含む雑誌が年々増加しており、2023年では、37％の雑誌がDASを採用しています。

Slide 6｜人文学分野でもDAS言及誌が増加

【06:30–07:00】

このグラフは、2023年にDASを含む雑誌の割合を分野別に示したものです。人文学の雑誌のうち22%がDASを採用しており、学術分野全体として、論文の根拠となる研究データの所在を明示する動きは広まっていることがわかります。

Slide 7｜Data Availability Statement（DAS）

【07:00–07:50】

このスライドは、Nature社のHumanities and Social Sciences CommunicationsにおけるDASの公開例です。DASは引用文献の前に書かれており、この論文の根拠となるデータがどこにいけばアクセスできるかを示しています。この論文の作者は自分のGitHub上に上げておりますが、機関リポジトリや分野リポジトリに上げるケースもあります。なお、この例では、研究データを公開していますが、問い合わせて相談次第で共有、権利関係上、提供できないなどと公開を制限する記述をする著者もおります。

Slide 8｜公的助成を受けた研究成果の即時公開

【07:50–09:50】

次に、公的助成をうけた研究成果物についてみていきます。結論から言うと、欧米だけでなく日本でも研究成果物の公開義務化が進展しております。

研究成果の公開義務化は EUから始まりました。EUは2018年9月にPlan Sの原則を発表します。Plan Sとは公的助成を受けた研究論文は出版後すぐに公開するという原則であり、欧州11カ国の助成機関の公的助成を受けた研究が対象となっており、2021年から実施されています。

つぎに 米国 OSTPが2022年8月に即時オープンアクセスの義務化を打ち出します。即時オープンアクセスとは、論文出版後、エンバーゴ期間なしに公開することを指し、論文だけでなく、その根拠データが含まれており、2025年度からの実施となります。

そして 日本も内閣府が 2024年2月にアメリカと同様に即時オープンアクセスの基本方針を発表し、アメリカと同じく2025年度から新規公募の競争的研究費について、査読付き学術論文および根拠データを、掲載後ただちに機関リポジトリ等へ掲載することを義務づけています。

なお、ここで言う根拠データは、投稿先の執筆要領・出版規程で公表が求められる研究データを指し、執筆要領で根拠データと明言されていない研究データについては、公開義務はありません。

　ここまで、研究データ公開が進んでおり、人文学分野の研究者も無視できない状況になっていることを説明しました。次からは、これら研究データ公開をはじめ、研究活動を支援する研究インフラについて紹介します。

Slide 9｜欧米の人文学分野の研究インフラ

【09:50–11:00】

まずは欧米の人文学インフラについて説明します。研究データ公開のためのインフラの要は、研究データの「文脈情報のフォーマット」と研究データと文脈情報を公開するための「受け皿」の２つです。情報学では、研究データの文脈情報のことをメタデータ、データの受け皿のことをリポジトリと呼びます。まずはメタデータについて説明します。人文学分野のデータはデータの生成方法、解釈の方法、他のデータとの関連など複雑な文脈を持っていますが、他のデータと比較し、検索しやすくするために、これらの文脈情報は標準化して記述する必要があります。そのためにテキストデータの公開には TEIという標準規格が使われ、画像データの公開のためにIIIF、文化遺産の出来事・関係記述に CIDOC-CRM、そしてウェブ発見性の向上のためにschema.orgという標準規格が使われています。

次に、公開・長期保存のための基盤について説明します。研究データは、先に説明したメタデータともに、リポジトリで公開されています。欧米の人文学分野の代表的なリポジトリとして、欧州の文化資源を検索するためのシステム Europeana、考古・文化遺産データの出版型プラットフォーム Open Context、米国の博物館などの横断ポータル DPLAがありますが、これらのプラットフォームはAPI や他のプラットフォームとの連携によって、機関間をまたぐ横断利用ができるように設計されています。

以上、人文学分野のインフラですが、以下の三つの課題があります。①著作権等の権利処理、②多言語対応、③持続可能な運営。どれも基盤の規模が大きいほど重くなる論点です。これらのことを踏まえ、次は日本の研究インフラについて説明します。

Slide 10｜日本の社会科学研究データインフラ

【10:10–11:20】

日本では、各研究機関が独自に運用しているインフラを中心にデータ共有は進んでいますが、メタデータの標準化や他機関の基盤との連携が遅れているため、公開データの検索がされにくく、再利用環境も十分ではありません。結果として、外部研究者との共同研究の機会を取りこぼす場面が生じがちで、人文・社会系の基盤整備の遅れに対する危機感がありました。そこで日本学術振興会は、2018年度から2023年度の5年間で「人文学・社会科学データインフラストラクチャー構築推進事業」を実施し、各研究機関の研究データを共通の標準メタデータ規格を用いて整備しました。現在はこの後続事業として、2023年10月から「強化事業」がスタートし、前事業で整備したインフラの強化を行っています。次に、この事業で構築したインフラを説明します。

Slide 11｜人文社会科学研究者に必要なインフラ

【11:20–12:30】

本事業のインフラ構築は国立情報学研究所が担当しましたが、その際、以下の3つをコンセプトに開発に着手しました。

まず1つめは横断検索プラットフォームです。横断プラットフォームを構築するにあたり、各分野のリポジトリと、人文学社会科学の分野標準メタデータを用いて連携できるプラットフォームの構築を行いました。

次に、ブラウザで使える研究データの実行環境（Jupyter／RStudio 等）を整備しました。授業や共同研究では、同じノートブック・同じ手順を共有できることが重要です。そこでサンプルデータとノートブックを一緒に用意する環境の整備を行いました。

最後に人文学・社会科学のためのリポジトリの開発です。国立情報学研究所は、WEKO3という汎用データリポジトリシステムを開発していましたが、主に社会科学分野との連携のためにData Documentation Initiativeという社会科学の標準メタデータ規格への対応と研究データのアクセス制限に対応しました。

Slide 12｜データインフラ事業の成果

【12:30–13:40】

先の3つのコンセプトをもとに構築した研究データのエコシステムがこの図となります。研究データを利用する研究者は、 ①検索 → ②データ入手 → ③分析 の流れで研究データをシームレスに再利用できるように設計されています。

人文学の研究者は、国立情報学研究所が運用するCiNiiなどの検索プラットフォームを使って研究論文を検索していましたが、この事業で構築したJDCatという統合検索システムを用いて横断的検索を行うことができます。ここに各機関が運用しているデータリポジトリのメタデータ情報が集約され、タイトル・作成者・時期・対象地域などの基本メタデータで研究データを検索することができます。

この検索システムで探索した研究データは、各機関が運用しているデータリポジトリから入手することができますが、この事業で構築したオンライン分析システムと連携しており、研究データはプログラムコードとともにJupyter Hubというオンライン環境を使ってそのまま再利用できます。

本事業では、分野データリポジトリに研究データを提供するためのガイドラインを用意しており、人文学社会科学分野の研究データの公開を希望する研究者が研究データを提供するための手順が整理されています。

Slide 13｜どうやって研究データを公開するの？

【13:40–16:10】

ここまで、研究データ公開のためのインフラについて説明しました。次からは、データインフラ事業で作成したガイドラインをもとに人文学分野の研究データの公開方法について紹介します。

Slide 14｜どうやって研究データを公開するの？

（約1分）  
それでは、人文学・社会科学におけるデータ共有の手引をもとに人文学データのデータ公開について紹介します。このような人文学向けのデータ公開ガイドラインは、欧州アカデミー連合ALLEAでも報告書の形で公開しています。これらのドキュメントは、このQRコードから確認することができます。

この2つのドキュメントでは、データマネジメントプラン、DMPに基づいた研究データ公開の重要性が強調されています。そこで、データマネジメントプランについて簡単に紹介したいと思います。

Slide 15｜データマネジメントプラン（DMP）

【21:50–23:10】

DMPとは研究で生じるデータの「取得・管理・保存・公開」方針を記述した計画書です。。DMPはプロジェクト終了後の報告書ではなく、プロジェクト進行に合わせて更新する前提の文書です。DMPは研究プロジェクトの管理、プロジェクトで生成された研究データの保存のために必要な資源を明確にするために必要なため、プロジェクト開始時にDMPの提出を義務としている助成機関もあります。日本では2024年度より日本学術振興会が採択された研究課題の研究代表者に対してDMPの作成を必須としています。現在のところ、提出義務はありません。

DMPに記載する内容は、研究中に生成した研究データについて、その形式、生成者、管理者、機微情報の有無、公開方針であり、研究プロジェクトが進行する中でDMPが更新され、DMPに沿ってデータ公開が行われます。

Slide 16｜研究データの公開の手順（計画〜収集・生成）

【23:10–24:50】

それでは、研究プロジェクトの進行に合わせて、具体的な研究データ公開の手順について紹介します。まず計画段階です。

まずは、研究前にこれからどのようなデータを生成する予定なのかを計画します。そのデータは一次資料なのか、一次資料を加工した二次加工データなのか、そのデータはテキストなのか、画像なのか、どのフォーマットで保存するのかについて計画します。次にそれらデータの法的・倫理的制約の確認をします。これから生成するデータの著作権、肖像権、文化財保護法などを事前に点検します。人文学では文化遺産データや個人データが絡むため、ここがボトルネックになりがちです。

以上をDMP（データマネジメントプラン）として文書化します。研究プロジェクトの進行によって、生成するデータの変更がある場合、その都度DMPを修正します。

つぎに収集・生成段階の手順について説明します。研究プロジェクトの進行にともない、研究データが収集・生成されます。その際、研究データの保存フォーマットを決めておくのが望ましいです。加工、解析中はともかく、保存するデータのフォーマットは、CSV、TIFF、WAVなど持続可能なオープンフォーマットを優先します。

また、研究データの文脈情報を記述するメタデータ規格を決め、収集・生成に合わせてメタデータを記述することが望ましいです。メタデータ規格には、大学図書館で利用するDublin CoreやJPCOARといった書誌ベースの規格がありますが、人文学データをより詳細に記述するために様々な規格があります。後で詳細に説明しますが、テキストデータではTEI、高精細画像データは IIIFなどの規格があります。

以上が研究中のデータ公開準備の要点です。続くスライドでは、メタデータの作成について補足します。

Slide 17｜公的助成機関が求めるメタデータ

【24:50–26:30】

まずは、公的助成の要請に応えるために最低限そろえるメタデータを確認します。2025年度より即時オープンアクセスがスタートし、今後公的助成を受けた研究成果のうち、根拠データは原則公開となっております。公開が難しい場合でもメタデータは公開が基本です。

この表は、実際に公開する項目を示しています。助成を受けたプロジェクト情報、データの名称、データの説明、分野、データの大きさ、公開先、データの作成者、管理者などの情報であり、DMPに沿って入力ができる項目となっております。ただこれらのメタデータは研究プロジェクトで生成したデータを説明する情報でしかなく、人文学データとしての詳細な文脈情報をもっていません。次に、人文学分野で用いられるメタデータ規格について紹介します。

Slide 18｜人文学のメタデータ

【26:30–27:50】

この表は、人文学分野で利用されている主要なメタデータを示しています。

大きく2つに分けると、Dublin Core（DCMI）、schema.org、JPCOARの３つは汎用的なメタデータ規格であり、残りの5つは人文学分野に特化したメタデータ規格です。

まず Dublin Core（DC）について説明します。最小限の記述要素（題名・作成者・日付・主題・権利など）で、機関リポジトリやOAI-PMHの“共通語彙”として最も広く使われており、古くからの書誌情報はこの規格で記述されています。

schema.org はウェブ検索向けの語彙で、JSON-LDという機械可読のためのフォーマットで記述することでGoogle等の検索エンジンで発見されやすいため、近年利用されております。Dublin Coreより細かに記述することができます。

JPCOARは日本の大学図書館などが運用する機関リポジトリで利用されていうメタデータ規格です。Dublin CoreやDataCiteという汎用研究データ用メタデータ規格との互換性を持ち、この規格で記述されたメタデータは、国立情報学研究所が運用するCiNii Researchに流通されます。京都大学学術情報リポジトリKURENAIもこのメタデータ規格を利用しています。次に人文学分野に特化したメタデータ規格について紹介します。

TEI（Text Encoding Initiative） はテキスト資料に特化した規格、CIDOC-CRM は博物館・文化遺産向けのモデル、IIIF はメタデータというより画像配信APIです。これら3つのメタデータはこの後くわしく解説します。

EAD（Encoded Archival Description） はアーカイブズの目録・ファインディングエイド用規格です。階層構造（ファンド→シリーズ→ファイル→アイテム）で集合体としての文脈を残します。欧米の公文書館で広く利用されています。

MODS/METS は図書館においてセットで使うことが多い組み合わせ規格です。MODSがリッチな書誌記述、METSがパッケージと構造マップ（どのファイルがどのページか）を担い、複合的なデジタル資源の整理、保存に向いた規格です。

次にTEI（Text Encoding Initiative）について詳細に紹介します。

Slide 19｜TEI: テキストデータのための標準規格

【27:50–29:10】

TEI（Text Encoding Initiative） はテキストの構造・注釈・版差を機械可読にする標準規格であり、そのテキストデータの書誌情報というよりも、テキストデータの背景情報を付与し、機械可読するための規格です。テキストデータに対して、章、段落、ページなどの論理構造を付与したり、注釈、校訂情報、人名や地名などの固有名詞にIDを付与、翻訳情報を追加することができます。XMLというタグ言語で記述し、国際的な人文学コミュニティで採用されている規格です。

次にCIDOC-CRMについて詳細に紹介します。

Slide 20｜CIDOC-CRM（文化遺産）

【29:10–30:50】

CIDOC-CRMは、文化遺産を記述するためのモデルであり、正確にはメタデータではありません。この規格は、それぞれの文化遺産に対して博物館や分野をまたいだ意味的なつながりを明確にすることを目的としています。

この規格では、作品・資料そのものの属性ではなく、制作（Production）、取得（Acquisition）、展示（Exhibition）、学術記述（Documentation）など、対象をめぐる出来事を記述します。

これらの出来事には常に誰（Actor）が関与し、いつ（Time-Span）、どこで（Place）起き、何を生んだ（Object/Information）が紐づきます。人物、地名、資料などをIDと紐づけ、各博物館がこの規格を用いて文化遺産をWEB上で公開することで、博物館をこえて文化遺産をつなげることができます。

最後にIIIFについて詳細に紹介します。

Slide 21｜IIIF（画像）

【30:50–32:20】

IIIF は高精細画像を「どこでも同じ方式で配信・閲覧・共有できるようにするAPI群」であり、これもまた正確にはメタデータではありません。画像ファイルを研究利用する際、この画像の右下に写っている人物のように、画像全体ではなく一部だけを拡大し、指したいことがありますが、この規格を用いることで、データの書誌情報を付与するだけでなく、この規格に対応した画像ビューアーを使って画像を拡大縮小部分表示、並べて比較などを行うことができます。この規格の記述にはJSON-LDという機械可読のためのフォーマットを利用しており、日本でも国立国会図書館や歴史民俗博物館や京都大学をはじめとする大学図書館などで活用されています。

メタデータについての説明は以上です。次は、研究が一段落し、研究データを実際に公開するかどうかをどのように判断するのかについて説明します

Slide 22｜研究後：公開判断

【32:20–33:40】

この図は　研究データ利活用協議会が作成した「研究データの公開・利用条件表示ガイドライン」の中で示されている研究データ公開フローチャートです。

研究データ公開を検討する際、このフローチャートに沿って検討していただけると幸いです。

データ公開を行う際、まずそのデータはリポジトリ等で公開の対象なのかどうかを検討し、その後、データ取得の経緯から、公開についての制約を確認します。

次のスライドでは、Q1の公開対象の特定について詳細にみていきます。

Slide23

ここでは **Q1：公開対象の特定について説明します**。リポジトリ等で公開の対象になるのは、基本的に **デジタルデータ** です。数値、テキスト、画像、音声、動画が該当します。

このデジタルデータは**三つに区別することができます**。  
**1つ目は根拠データです。これは**論文の図表や再現に必要なデータ、あるいは成果そのものとして公開するデータです。検証可能性の観点から、公開の優先度が高い領域です。  
**2つ目は一次データです。これは**観測・収集によるオリジナルのデータです。自分の保有物か、第三者が保有するかを明確にし、第三者保有の場合は所在と許諾の確認が必要です。  
**3つ目は派生データです。**クリーニング、翻刻、注釈、集計など **ソースを加工して生成** したデータです。**引用方法・権利の帰属・バージョン情報** を必ず記録します。

一方で、**公開対象外** もあります。  
メタデータそのもの、研究ノートや日誌、標本・試料などの **実体物**、既に刊行された **著作物**（論文・書籍・二次著作物等）、そして **使用環境**（データベースや分析ソフト等）は、この枠組みでは公開対象に含めません。

なお、**公開が義務** になる場合があります。助成機関、出版社、所属機関の方針で、根拠データ等の公開が求められるケースです。  
**迷ったら**、法務・知財・倫理の担当や外部の専門家に相談し、必要に応じて **組織のデータポリシー整備** も検討してください。

以上がQ1の要点です。次のステップでは、こうして特定したデータについて **公開を妨げる制約条件** を確認します。

Slide 23｜研究後：研究データの公開判断1

【33:40–34:50】

次のQ2のステップでは、そのデータの特性や取得経緯から公開可能かどうかを判断します。

人文学分野の研究データの場合、文化財保護や著作権保護の観点から公開が難しいことがあります。そのようなデータは、助成機関や出版社が公開を求めていても公開できません。

また、個人情報保護の観点から公開が難しいデータも存在します。

次のQ3のステップでは、制約がある場合の対処方法を検討します。

制約がある場合でも、公開までの猶予期間を設ける、匿名加工を行う、制限付きアクセスを行うことで共有を行うことができる可能性もあります。どうしても公開ができない場合は、判断プロセスを文書化し、メタデータのみ公開し、原データは適切に保管するという方法があります。

制約に対する対応策はが分からないときは、機関の倫理審査委員会等に確認する必要があるでしょう。

Slide 24｜公開先の選択

【34:50–36:10】

データ公開の制約について確認した後、そのデータをどのリポジトリで公開するのかを判断します。

研究データを公開するリポジトリには分野リポジトリ、機関リポジトリ、汎用リポジトリの3種類があります。

まずは分野リポジトリについて紹介します。分野リポジトリは、公開したデータが利活用できるように様々な機能を有しており、先に紹介したTEIやIIIF規格を使った整備を行うことができ、引用や再利用しやすいという長所があります。ただし、特定のフォーマットの研究データのみ受け入れるといった問題や、公開サービスを試験的に行っている機関もあるので、その永続性が不透明な点もあります

次に機関リポジトリについて紹介します。機関リポジトリは大学図書館などが運用していて、その機関に所属している研究者であれば、どのような種類のデータであっても原則公開できます。機関リポジトリで簡単なメタデータ整備が行われ、リポジトリの永続性が期待できます。ただし、付与されるメタデータはJPCOARなどの汎用的なメタデータに限定され、条件を満たした利用希望者にのみ公開に対応している機関リポジトリは少数です。

最後に汎用リポジトリです。汎用リポジトリは研究機関や学術出版社が運用しており、Figshare、Zenodo、OSFなどがあります、GitHubもこの汎用リポジトリに上げることができると考えられます。汎用リポジトリの長所は、研究者自身で即座に研究データを公開できるところです。また、研究者自身で公開しているデータに対してアクセス制限をかけることができます。汎用リポジトリは研究者にとって便利ですが、付与できるメタデータは汎用的なメタデータに限定される点、個人のドキュメント置き場として見られ、研究データリポジトリとしてのブランドが弱い点が問題に挙げられます。また現在これらの汎用リポジトリは原則無償で利用できますが、有償化への転換、またはサービス停止の可能性もあります。

Slide 25｜ライセンスの指定

【36:10–37:20】

最後に研究データへのライセンスの付与について説明します。ライセンス付与のポイントは以下の3つです。

まず、対象のデータに対してライセンスを付与できる権利があるかを最初に確認します。

写本や絵画を撮影した画像ファイルなどは、その撮影対象の所蔵者への確認が必要になると考えられます。、またPublic Domain相当の素材には新たなライセンスを付けることができません。第三者コンテンツが混在する場合は、その部分を別条件にするか、除外します。

特に画像とテキストが混在するデータの場合、画像は権利が重い場合が多いため、テキストと異なる条件を付与することも考慮すべきです。

次にライセンスの選択ですが、できるだけ自由度の高いライセンスを選択するべきです。

研究データの場合、クリエイティブ・コモンズ・ライセンスの一つ CC BY 4.0が望ましいと考えられます。CC BY 4.0は原作者の表示を条件に、資料の複製、配布、改変、商用利用などが許可されるライセンスのため、研究データのライセンスとしては自由度が高すぎるようにも思われますが、商用利用禁止の CC BY NCにすると、商用利用の定義が曖昧なため、教育目的や研究目的の利用が阻害される恐れがあり、改変禁止のCC BY NDにすると、その研究データの翻訳や修正、一部抜粋などに利用されにくくなり、TEI規格の適用ができなくなります。

また、ライセンスを記述する際はURIをメタデータに明記しておくことで、機械的アクセスに対応しやすくなります。一般のリポジトリシステムはURLも明記されるように設計されていますが、自分のサイトで公開する際は注意が必要です。

Slide 26｜データ公開FAQ

【37:20–38:50】

ここまで一通り研究データ公開について解説しました。今までの説明を3点に整理したいと思います。

まず「どこに」公開するのかですが、候補は 分野リポジトリ／機関リポジトリ／汎用リポジトリの3つがあります。利活用が見込まれる分野リポジトリが理想的ですが、データの種類によっては受け入れられないことがあります。永続性を考えると、大学機関が運用する機関リポジトリが望ましいと思いますが、即時性やアクセス制限を行う場合は汎用リポジトリの利用も検討できると思います。

次に公開のためには「何を」するのかですが、研究実施前からDMPを想定した研究を行うことが重要だと考えられます。研究プロジェクトの進展によって計画の変更はありますが、どのようなデータを収集・生成して、そのデータにはどのような権利関係が発生するのかをDMPとして文書化しておけば、データ公開に対応できると考えられます。

最後に「どこまで」研究データを公開するのかですが、現在公開が義務化されている研究データは査読論文のエビデンスデータのみです。公開するべきではない、または公開してはならないデータは公開してはいけません。オープン・アンド・クローズド戦略を踏まえて公開が必要です。特に人文学分野では、先住民資料や宗教的資料など、個人情報ではないが文化的感受性の高いデータもあるので、そのようなデータを公開する際は関係者との合意形成が必要となります。

Slide 27｜研究データ公開の展望

【38:50–40:40】

最後に、これからの研究データ公開の展望について述べたいと思います。

現在、研究データ公開がトップダウンで決定し、データ公開環境が整備されています。公的助成機関を中心に研究論文と根拠データの公開が義務化され、学術出版社もDASの記載をはじめ根拠データの公開が当たり前になりつつあります。その中で研究データ公開のためのインフラが急ピッチで整備されている状況です。

今後、機関リポジトリが最低限の公開体制――エビデンスデータ（論文図表）と研究概要メタデータの公開――を担い、分野リポジトリが再利用に強い公開を牽引すると考えられます。IIIF や TEIといった標準で記述・配信することで、利活用指向の公開が広がることが期待されますが、研究者がおいていかれている状況になっているのも事実です。

そこでボトムアップからのデータ公開を促進していく必要があると考えれます。

各分野でデータ公開ガイドラインが整備され、何を最小に、どこまで公開するかの判断がそろい、研究データを公開する研究者に対するインセンティブ、自分が公開した研究データに関係する研究を把握し、新しい共同研究の機会が広がるような状況が設計されると、研究者からの自発的な公開が広がるのではないかと思います。

Slide 28｜まとめ

【44:05–44:45】

最後に本日のセミナーの内容を振り返ります。

オープンサイエンスの浸透により、研究成果の公開は全分野で前提化し、国際・異分野連携の機会が広がっています。その一方でジャーナル側の査読・公開要件の明確化し公的助成に伴う研究データ公開義務が研究者に課せられるようになってきております。

研究インフラの整備で研究環境が変化しつつあり、DMP による研究管理が今後当たり前になりつつありことについて説明しました。

データ公開を進める上で重要んことはこのDMPを最低限のメタデータとして、公開時の判断資料として使うことです。また、最後に今後のデータ公開環境について話させていただきました。より良い公開環境には、公開インセンティブとルールの見直しが欠かせないと思います。

Slide 29｜連絡先・クロージング

【44:45–45:00】

それでは発表を終わらせていただきます。ご清聴ありがとうございました。研究データ公開についてご相談があればNII RCOSまで気軽にお問い合わせください。