

政府関係機関事業紹介

研究データ検索基盤CiNii Research

国立情報学研究所 オープンサイエンス基盤研究センター

1. はじめに

国立情報学研究所（以下、NIIとします）が構築を進めている研究データ基盤NII Research Data Cloud (NII RDC) は、研究データの管理基盤、公開基盤、検索基盤という3つの基盤から構成されています。今号では、この中から研究データ検索基盤について紹介します。

これまでNIIは、日本の学術情報流通に関わる様々な検索サービスを大学の研究者や学生、図書館員などに提供してきました。NII学術情報ナビゲータCiNii (<https://ci.nii.ac.jp>) は、学術論文や会議抄録、大学図書館の蔵書、博士論文といった学術成果情報についての検索サービスです。科学研究費助成事業データベースKAKEN (<https://kaken.nii.ac.jp>) では、文部科学省および日本学術振興会が実施する科学研究費助成事業により行われた研究の採択課題や研究成果報告書などをデータベースに収録して、検索サービスとして提供しています。学術機関リポジトリデータベースIRDB (<https://irdb.nii.ac.jp>) では、日本の機関リポジトリに登録されたコンテンツのメタデータを集約して、データベースとしています。

学術情報の検索サービスに新たに求められているのが、研究データについてのサポートです。オープンサイエンスでは、論文だけでなく研究データを含む研究成果を積極的に公開することで、論文だけではなし得なかった再利用や分野横断型研究への発展を促進しようとしています。別の側面としては、論文のもととなった研究データを論文の補足として公開することは、研究成果の再現性や透明性にとって重要です。以前より生物学や臨床医学といったいくつかの研究分野では、雑誌論文に投稿する際に研究データを公開することを求められることがありましたが、最近は多くの分野で雑誌論文の投稿規定として研究データの公開について言及されるようになってきています。

そのような状況変化の中で、学術論文の検索を提供しているCiNiiにおいても、論文の詳細情報として補足物が表示できたり、研究成果物の一種として研究データそのものが発見できるようになることが求められてきました。そこで現在新しく開発を進めているのが、CiNii Researchです。

2. 新しい学術検索サービス CiNii Research

従来のCiNiiは学術成果情報を主な検索対象としていましたが、CiNii Researchでは視点を変えて広く研究活動に関わるものを検索対象とします。論文や図書といった伝統的な研究成果物だけではなく、研究データやソフトウェアといった新しいタイプの研究成果物や、KAKENが扱っているような研究プロジェクトや研究者といった研究活動に関わる様々な事象をつなぐようにデータベースの整備と検索サービスの開発を進めています。これによって、今読んでいる論文に使われているデータを探したいといったように、様々な切り口から検索して、関連する情報を辿ることで目的のものを発見できるようにします。

CiNii Researchでは様々なデータソースからメタデータを収集して活用しています。CiNii の論文2,000万件や図書1,200万件、KAKENの研究者20万人・助成60万プロジェクトに加えて、IRDBを介した各大学機関リポジトリ、JaLCやDataCiteといった研究データのためのDOI登録機関、ならびに分野データリポジトリなどからメタデータを収集しています。これらのメタデータの形式はデータソース毎にすべて異なるため、CiNii Research用に統一した形式に合わせて変換することで利用しています。現時点で約6,760万件収録しています。

CiNii Researchでは、単に集めたメタデータをそのまま検索できるようにしているのではなく、論文、研究データ、研究者といった検索対象として想定している学術情報資源の間を名寄せして関係

リンクを構築することで、緩やかなデータ統合をしています。これをCiNiiナレッジグラフと呼んでいます。名寄せとは、例えば複数のデータソースに同じ論文や研究者がある場合に、それを統合して扱うようにすることです。関係リンクとは、例えばある論文を書いた研究者や、ある研究データを用いて書かれた論文、ある論文を引用している論文といったように、学術情報資源間にある何かしらの関係を明示したものです。一つ一つは単純な関係ですが、それを集めると巨大なグラフとなるのです。

図1にCiNiiナレッジグラフを可視化した例を示します。ネットワーク図の円形のノードは個別の論文や研究データ、研究者などを表しています。また、エッジはノード間の関係を表しています。

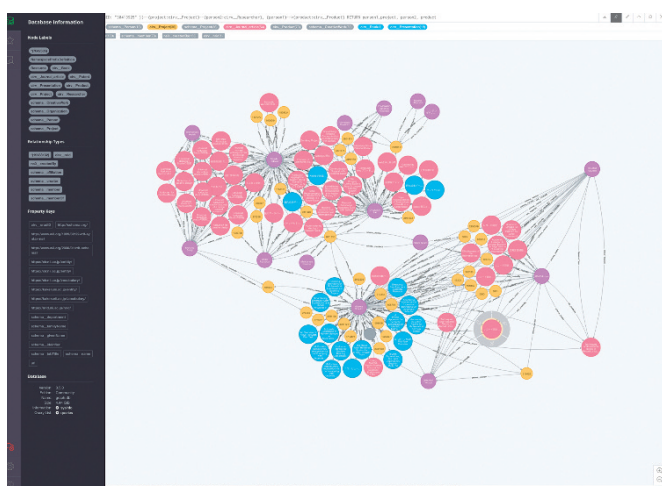


図1 CiNiiナレッジグラフ可視化例

3. 大学や研究機関との連携

CiNii Researchをより良いものにしていく上で、大学や研究機関と連携することは大切です。これまでのNIIの検索サービスにおいても、大学機関との連携でデータ作成が行われてきました。論文検索であるCiNii Articlesや博士論文検索であるCiNii Dissertationsでは、大学側で管理されている機関リポジトリのデータをIRDBに集約することで、利用しています。大学図書館の蔵書検索であるCiNii Booksは、大学図書館の方々に書誌や蔵書についてNIIのデータベースに入力して頂いたものを、ウェブから検索できるようにしたものです。

まず想定していることは、機関リポジトリとの連携を強化することです。機関リポジトリは大学

の研究者が研究データを公開する場所の一つとして考えられています。連携を強化するために、2019年4月にIRDBの完全リニューアルを行いました。新しいIRDBは、日本の機関リポジトリコミュニティであるオープンアクセスリポジトリ推進協会（JPCOAR）が策定した新しいメタデータスキーマ（JPCOARスキーマ）をベースに作られています。これにより研究データが扱いやすくなるほか、研究者や学術資料を識別するためのIDを流通しやすくなるといったメリットがあります。また、メタデータのエラーチェックや正規化も可能となっており、問題があった場合は機関側と連携して修正をするということも行っています。なお、機関リポジトリ自体の強化については、次号で紹介する予定です。

次に想定しているのが、各大学や研究機関が持っている研究データリポジトリやアーカイブとの連携です。研究データ共有は、すでに分野内の研究者間では行われていることが多く、分野で有名な研究機関が研究データリポジトリをホストしていることがあります。国立情報学研究所では、情報学分野の研究向けに企業や研究コミュニティが提供するデータセットを集約した情報学研究データリポジトリ (<https://www.nii.ac.jp/dsc/idr/>) を提供していますし、東京大学社会科学研究所附属社会調査・データアーカイブ研究センター (<https://csrda.iss.u-tokyo.ac.jp/>) では、統計調査・社会調査の個票データを収集・保管・提供しています。CiNii Researchでは、分野別データリポジトリのメタデータを収集してナレッジグラフに統合することで、分野横断的な検索を提供したいと考えています。これはオープンサイエンスの目標の一つである分野横断型研究への発展にとって重要な一歩です。

4. おわりに

CiNii Researchは2020年度内の公開に向けて、鋭意開発中です。CiNii Researchによって広く日本の研究成果をつながりから探せるようになることで、新しい研究のきっかけになることができれば幸いです。