

政府関係機関事業紹介

研究データ公開基盤WEKO3

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（NII）が構築を進めている研究データ基盤NII Research Data Cloud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公開基盤「WEKO3」について紹介します。

2. これまでのJAIRO Cloud

これまでNIIでは、大学等研究機関における学術成果の公開と流通を目的とした機関リポジトリの構築支援を実施してきました。現在は、機関リポジトリの構築・運用が可能なクラウド型の機関リポジトリ構築サービス「JAIRO Cloud」を提供しています。

機関リポジトリとは、学術機関の構成員による研究成果を保存し、流通させることを目的としたサービスの総称です。主に、学術論文、博士論文、紀要論文などの公開の場として利用されています。学術機関リポジトリデータベース「IRDB」^[1]の調査によると、2020年1月末時点で809の学術機関で機関リポジトリが導入されています（図1）。日本の大学は、2019年度時点で786機関あるため^[2]、数字だけみると、殆どすべての大学で機関リポジトリが導入されていることとなります。この普及度合いは、世界的にみても類をみないものであり、日本は機関リポジトリの先進国と言えます。

日本における機関リポジトリの始まりは「オープンアクセス運動」をきっかけとしています。オープンアクセス運動とは「ブダペスト・オープンアクセス・イニシアチブ（BOAI）」^[3]により提唱された、研究者のだれもが障壁なく学術論文へのアクセスを可能とする「オープンアクセス」の実現を目指した取り組みです。この取り組みの背景には、増え続ける雑誌購読料による学術情報流通の危機「シリアルズ・クライシス」への危機感があります。オープンアクセス運動では、学術論文へのアクセスをオープンにすることで研究者による研究成果へのアクセスを確実にすることを目指しています。BOAIでは、そのための手段として、研究者自身が学術論文をアーカイブする「セルフアーカ

イブ」、出版時に学術論文へのアクセスをオープンにする「オープンアクセス出版」を提唱しています。機関リポジトリでは前者によるオープンアクセスを実現します。

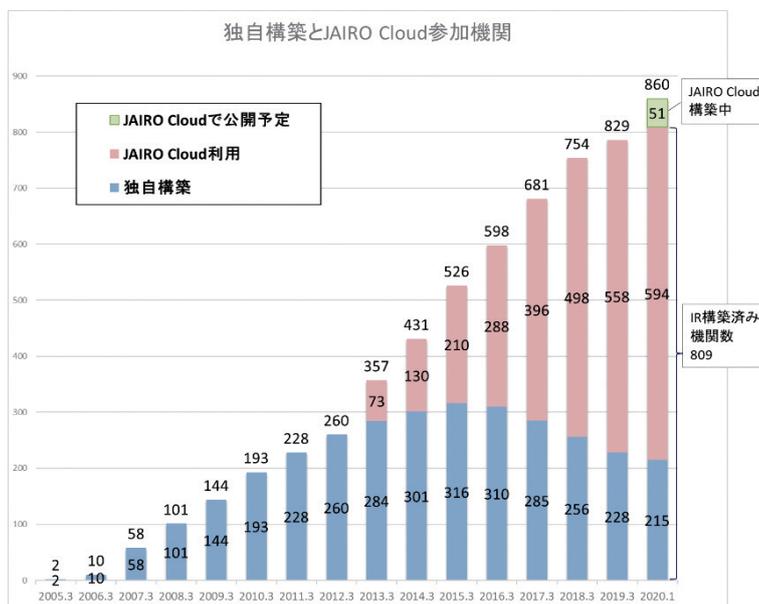


図1 機関リポジトリ導入機関数の推移

機関リポジトリの普及に影響を与えたものとして、2013年4月から施行された「博士論文のオープンアクセス義務化」があります。学位規則の改正により、論文の要旨および審査結果を学位授与後3ヶ月以内にインターネットで公開し、本文も1年以内に公開することが義務付けられました。この規則改正に伴い、国会図書館による機関リポジトリ経由での博士論文収集事業が開始されています。2020年1月末時点で、国立国会図書館デジタルコレクション^[4]に収録されている約6万5千件の電子化された博士論文のうち約7割が機関リポジトリ経由で収集されています^[5]。今や博士論文は機関リポジトリの主要なコンテンツの一つとなっています。

そして、機関リポジトリの普及に強い影響を与えたと考えられるのが「紀要論文」の電子化です。紀要論文とは、大学等の研究機関、機関に属する研究所、博物館などが定期的に発行する学術雑誌

をいいます。従来は紙媒体での出版が殆どだったのですが、近年は経費削減のために電子出版に移行するケースが増えています。IRDBによる統計値によると、日本の機関リポジトリが保持するコンテンツのうち最も多いのが紀要論文であり、コンテンツ全体の約5割を占めています。

JAIRO Cloudはこのような状況の中、約600の機関に対して、学术论文、博士論文、紀要論文を公開、流通させるための機関リポジトリ機能を提供しています。このJAIRO Cloudのリポジトリ機能は、同じくNIIが開発するリポジトリ基盤ソフトウェア「WEKO」^[6]で実現されています。WEKOの特徴は、メタデータや画面レイアウトの設計など、機関リポジトリの構築に必要なすべての機能が、ウェブブラウザ経由で手軽に利用できる点にあります。WEKOに実装されている機能については、JAIRO Cloudユーザの意見・要望を積極的に取り入れ、JaLC DOIやERDB-JP^[7]との連携機能など、機関リポジトリとしての機能の拡張を続けてきました。現在では、WEKOは、リポジトリソフトウェアとして著名なEPrints^[8]やDSpace^[9]と比較しても遜色のない、場合によってはそれらを上回る機能を有するに至っています。

3. 公開基盤のこれから

近年、機関リポジトリに求められる役割に大きな変化が起ころうとしています。従来の学术论文に対するオープンアクセスの実現だけでなく、研究データに対するオープンアクセスの実現も求められるようになってきています。NIIが開発するWEKOやJAIRO Cloudについても、これからの環境変化に柔軟に対応していく必要があります。

研究データ公開基盤「WEKO3」は、現在のJAIRO CloudのベースソフトウェアWEKOの後継ソフトウェアです。①JAIRO Cloudの運用で洗練されてきた機能を踏襲・強化すること、②JAIRO Cloudサービスとしての可用性・信頼性・保守性を確保すること、③文献リポジトリだけでなく研究データリポジトリとしての機能要求に対応すること、を基本方針として開発を進めています。特に③については、研究データの多様性を鑑み、システムの拡張性とスケーラビリティを重要視したシステム設計を目指しています。

また、WEKO3は研究データの多様性に対応するため、①ファイルプレビュー機能、②メタデータ管理機能、③ワークフロー機能の強化を実施しています。①では、ファイルプレビュー可能なファイル形式をプラグイン機構で拡張できるように

なっており、メタデータだけでは内容の把握が難しい研究データへのアクセスを支援します。②では、階層構造を持つメタデータスキーマをサポートし、研究データの多様性に対応するメタデータ表現を実現するだけでなく、メタデータおよびファイルのバージョン管理機能も提供します(図2)。WEKO3は研究データに対応するJPCOARスキーマ^[10]にも対応しています。③では、機関における研究データの受入れフローの多様化を想定し、フロー内容や実行権限の変更がカスタマイズ可能なワークフロー機能を実装しています。WEKO3は、これらの機能を核として、機関における研究データの公開を支援していきます。

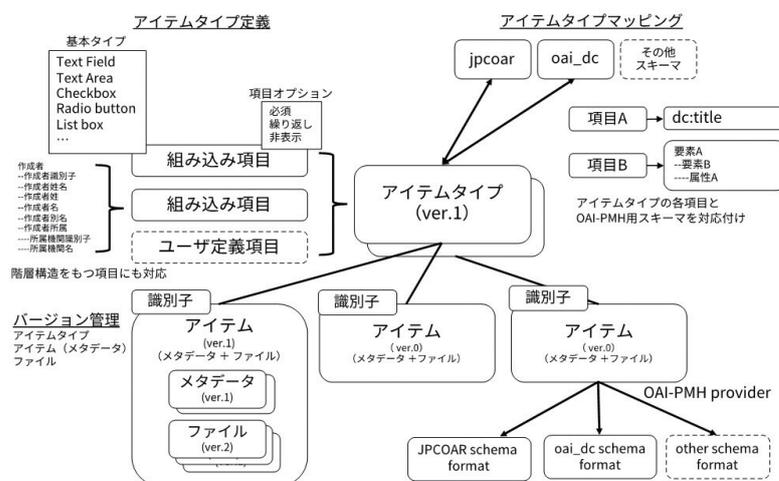


図2 WEKO3のデータモデル

4. おわりに

WEKO3をベースにした次期JAIRO Cloudは2020年度のサービス提供を計画しています。2019年9月からはJPCOARの協力を得て、WEKO3を用いた移行実験を開始し、具体的な移行作業に向けて動き出しています。JAIRO Cloudの利用機関は約600あり、大規模な移行作業となります。移行作業は当然のこと、WEKO3/次期JAIRO Cloudにより、機関による研究データの公開が促進され、オープンサイエンスの実践に繋がるよう、鋭意開発を進めていきます。

参考文献及び関連URL

- [1] <https://irdb.nii.ac.jp/>
- [2] 文部科学省「学校基本調査」(2019年)
- [3] <https://www.budapestopenaccessinitiative.org/>
- [4] <http://dl.ndl.go.jp/>
- [5] 2020年1月末時点で収集された電子化博士論文は65,781件 (IRDB連携: 47,646件, 送信: 18,135件)
- [6] <http://weko.at.nii.ac.jp/>
- [7] <https://erdb-jp.nii.ac.jp/>
- [8] <https://www.eprints.org/>
- [9] <https://duraspace.org/dspace/>
- [10] <https://schema.irdb.nii.ac.jp/ja>