

大阪大学における ONION を中心としたデータ管理基盤整備にむけて Towards Development of Data Management Infrastructure around ONION within Osaka University

田主 英之[†] 山下 晃弘[†] 細見 岳生^{†‡} 並木 悠太[†] 甲斐 尚人[†] 伊達 進[†]
Hideyuki Tanushi Akihiro Yamashita Takeo Hosomi Yuta Namiki Naoto Kai Susumu Date

1. はじめに

近年、オープンサイエンスへの関心が高まっている。オープンサイエンスとは、ICT によるデジタル化とネットワーク化された情報基盤、およびその基盤で公開される情報を活用して科学研究を変容させる活動であり、産業を含む社会を変え、科学と社会の関係も変える活動である[1]。また、オープンサイエンスには、あらゆるユーザーが研究成果である査読付き論文に自由にアクセスできるオープンアクセスや、研究成果に至った知見、および研究データにも自由にアクセスできるオープンデータという概念も含まれる。

大阪大学はオープンサイエンスを推進している。そして、大阪大学は情報基盤としてデータ集約基盤 ONION (Osaka University's Data Aggregation Infrastructure) とデータ公開基盤 OUKA (Osaka University Knowledge Archive) を整備している。ONION はデータ集約基盤であり、学内外で生み出されるデータを容易にかつ安全に集約する機能を持つ。そして、集約されたデータは高性能計算・データ分析基盤システム SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datascience) で活用されたり、国内外にまたがって行われる共同研究、産学共創活動における研究者間で共有されたりしている[2]。OUKA は機関リポジトリで、大阪大学の教育研究活動から生み出される学術成果を電子的に保管・公開するサービスであり、データ公開基盤としての役割も果たしている。取り扱う学術成果は論文だけでなく、画像、計測データなどの研究データも含まれる[3]。

今後 ONION 上にある研究データを OUKA で公開する必要性が高まることが予見される。現在、オープンアクセスにおいて自前の公的プラットフォームを持つこと、およびバックデータを含む学術論文の著者最終稿を公開することの重要性が活発に議論されている[4]。OUKA では著者最終稿のオープンアクセス化 (グリーンオープンアクセス) を進めており、ONION 上にある研究データが学術論文のバックデータとなるからである。

オープンサイエンスにおいても、学術論文や研究データの公開などの研究データ管理の責任を負うのは研究者自身である。そして、オープンアクセスやオープンデータを推進するための研究データ管理は研究者にとって多大な負担である。それゆえに、学術研究機関は研究者の研究データ管理の負担を減らすための環境整備を求められている。

大阪大学は ONION と OUKA という基盤は整備したが、それぞれは機能面で連携していない。つまり、データを所有する研究者とデータを公開する場所が接続できていない問題がある。また、研究活動における研究データ管理の「生成→加工→分析→保存→公開→再利用→生成→…」と

いう研究データライフサイクル[5]という観点で見ると、ONION から OUKA への「保存から公開をする」という研究データ管理のプロセスが研究者にとって負担になっている問題がある。

そこで本研究では、上記の問題に対して ONION と OUKA を連携し、新たな負担なく容易に、安全に研究者が研究データの公開申請をできる一つのプロトタイプを提案する。そして、その提案を実装し、研究データ公開プロセスをシームレスに繋げた仕組みについて報告する。

2. プロトタイプの提案と評価

2.1 課題

図 1 にデータ集約基盤 ONION 上にある研究データをデータ公開基盤 OUKA で公開する場合に想定されるワークフローを示す。研究者が公開用研究データを送付する場合、研究者 (A) が公開用研究データを ONION (C) から一旦自らのローカル環境にダウンロードし、OUKA 窓口 (B) 宛に公開申請とともに送付する「C→A→B」というフローをとる。つまり、研究者は「ONION から公開用研究データをダウンロードする (C→A)」「公開申請、および公開用研究データを OUKA 窓口へ送付する (A→B)」という 2 種のタスクを実行する必要がある。

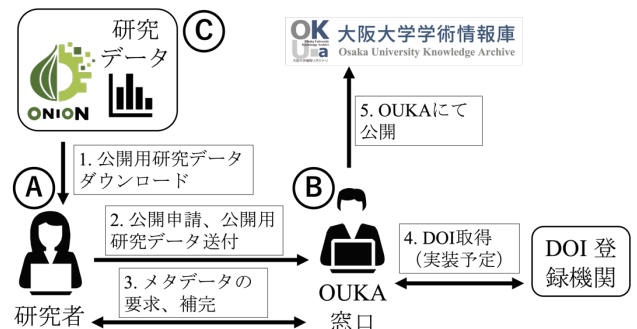


図 1 ONION 上のデータを公開する場合に想定されるワークフロー

図 1 の「C→A→B」というフローにおいて、公開用研究データが大量になれば、OUKA の申請方法を調べた上で自ら公開申請メールを作成し、OUKA 窓口宛に送付することは、研究者にとって煩わしい作業である。また、公開用研究データが大量、あるいは公開用研究データ自体の容量が非常に大きい場合、それらを研究者のローカル環境にダウンロードすることは研究者にとって負担が大きい。

本研究ではローカル環境を経由させることなく、ONION から直接 OUKA の窓口宛に研究者が公開申請、および公開用研究データを送付する方法を提案する。なぜなら、こ

[†] 大阪大学 Osaka University

[‡] 日本電気 NEC Corporation

の連携問題では、新たな負担なく容易に、安全に研究者が公開申請をできることが第一の目標だからである。

2.2 プロトタイプの提案

2.2.1 手段

本研究は既存の ONION の機能と OUKA での論文公開のワークフローを最大限に活用し組み合わせるアプローチを採用する。そして、図 1 で示した公開申請の「C→A→B」というフローを研究者 (A) を経由することなく「C→B」というフローに置き換える。

1. 研究者が ONION 上から直接公開申請をする。
2. OUKA 窓口 (B) が直接公開用研究データを ONION (C) からダウンロードする。

これらを実現することで、研究者は公開用研究データを自らの環境を経由させることなく ONION から直接 OUKA 窓口に向けて公開申請が可能になる。

2.2.2 実装

ONION はオンラインストレージ構築ソフトウェア Nextcloud[6]をポータルとして採用している。Nextcloud は PHP と Javascript をベースに開発されており、独自のプラグインを開発、実装することが可能である。本研究では手段 1 と 2 を実現させる連携プラグインを開発した。

開発した連携プラグインは Nextcloud の共有メニューに新たに「公開申請」メニューを追加するものである。図 2 は開発した公開申請メニューを示す。研究者は公開用研究データを選択し、共有メニューを選ぶことで図 2 に示す公開申請メニュータブの表示が可能になる。公開用研究データはファイル単位のみならず、複数ファイルをまとめたフォルダ単位も同様に公開申請をできるようにした。

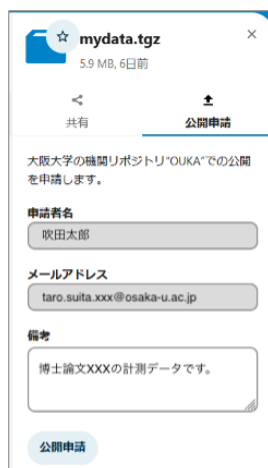


図 2 Nextcloud の公開申請メニュー例

Nextcloud にはパスワード付きの共有リンクを発行する機能がある。この機能を使うことで、ONION の非ユーザーも ONION にあるデータをダウンロードすることが可能になる。本研究ではこの機能を使い OUKA 窓口が公開用研究データを ONION からダウンロードできるようにした。

この連携プラグインでは公開申請ボタンを押すだけで、自動的にバックグラウンドで公開用研究データのパスワード付き共有リンクが生成される。そして、件名に送信日時が付き、本文に申請者名、メールアドレス、共有リンク、共有アイテム名、パスワードおよび備考が含まれる公開申

請メールが OUKA 窓口宛に送信される。この公開申請メールは申請者にもコピーが送信される。

連携プラグインから送信された公開申請メールを受け取った OUKA 窓口は、公開申請メールに記載されている共有リンクとパスワードを使用して公開用研究データをダウンロードする。そして、申請者のメールアドレスを使用し、以降の調整 (メタデータの要求、補完など) を行う。

2.3 評価

開発終了前の受入テストでは、連携プラグインを使ってファイルとフォルダの 2 種類のテストデータによる公開申請評価を行った。OUKA 窓口のメールアドレスに公開申請メールが届き、公開申請メールに記載されている共有リンクとパスワードを使用してテストデータをダウンロードし、OUKA リポジトリへ取り込めることが確認できた。

今回の連携プラグイン開発では手段 1 と 2 を実現することで、研究者が「データを選択して『公開申請』ボタンを押す」だけで公開用研究データを ONION から直接 OUKA 窓口に向けて公開申請することが可能になった。そして、このプロトタイプを提案、実装したことで ONION と OUKA を連携させることができた。

3. まとめと今後への課題

本研究ではデータ集約基盤 ONION とデータ公開基盤 OUKA を連携する一つのプロトタイプを提案、実装した。その提案した仕組みで、データ公開申請の際の研究者の負担を軽減し、容易かつ安全なデータ公開申請を可能にした。そして、「保存から公開をする」という研究データ管理における研究データライフサイクルの一部を効率化した。

本研究のプロトタイプは、図書館側に公開申請のメールを送るのみであり、今後重複申請を防ぐための公開申請情報を管理する機能や公開に必要なメタデータ送付の機能なども必要であると考えている。

本研究で開発した公開申請モジュールを使った実証試験はまだ実施できていない。今後実際に ONION 上からこの連携機能を利用して図書館宛に公開申請をして、OUKA 上でデータを公開する実証試験を実施する。

謝辞

本研究は文部科学省「AI 等の活用を推進する研究データエコシステム構築事業」の支援を受けたものです。

参考文献

- [1] 林 和弘, "オープンサイエンスの進展とシチズンサイエンスから共創型研究への発展", 学術の動向, 2018, 23 巻, 11 号, p. 11_12-11_29, (2018).
- [2] 伊達 進, 寺前 勇希, 勝浦 裕貴, 木越 信一郎, 木戸 善之, "ONION: 大阪大学のデータ集約基盤", 学術情報処理研究, 26 巻, 1 号, p. 87-96 (2022).
- [3] OUKA について, <https://ir.library.osaka-u.ac.jp/portal/about.html> (2023 年 5 月 26 日参照).
- [4] 内閣府 科学技術・イノベーション推進事務局, "論文等のオープンアクセスについて (論点とりまとめ)" (2023), <https://www8.cao.go.jp/cstp/gaiyo/yusikisha/20230525/siryoi1.pdf> (2023 年 6 月 2 日参照).
- [5] 吉田 幸苗, 天野 絵里子, 松本 侑子, 西園 由依, 山地 一禎, 南山 泰之, 尾城 孝一, 常川 真央, 大園 隼彦, "教材「研究データ管理サービスの設計と実践」第 2 版" (2022), <https://jpcoar.repo.nii.ac.jp/records/607> (2023 年 6 月 1 日参照).
- [6] Nextcloud: top page, <https://nextcloud.com> (2023 年 6 月 2 日参照).