

OU 研究データマネジメントエコシステム構築を目指した実験系研究データメタデータ管理の検証モデルシステム構築

田主 英之 (D3 センター)

伊達 進 (D3 センター)

古谷 浩志 (コアファシリティ機構)

春本 要 (D3 センター)

甲斐 尚人 (D3 センター)

細見 岳生 (D3 センター)

1 研究の背景

オープンサイエンスへの流れが加速する中で、研究データを流通させ、共有、公開、そして利活用することが、日常的に行われる時代へと向かっている。しかし、現実には日常的に行うための手法や基盤整備が未だに不十分であり、その確立が喫緊の課題となっている。

D3 センターは学内外で生み出される研究データを容易に集約し、本センターが運用する計算・分析基盤での解析に利活用、および学内の研究者間でのデータ共有を可能にする機能を持つデータ集約基盤 ONION (Osaka university Next-generation Infrastructure for Open research and open innovatiON) を試験的に運用している [a]。D3 センターの共同研究部門の一つである高性能計算・データ分析融合基盤協働研究所は、学術研究者が産学・国際共同研究等を通じて新たな社会的価値を創生するための計算基盤・データ基盤構築について研究開発を行っており、本研究所の研究データ管理チームは、ONION を活用しオープンサイエンスを推進するための部局間連携を行っている。その一つが学内の分析機器の共用を推し進めているコアファシリティ機構との共同研究である。

コアファシリティ機構は、ネットワークから隔離されている共用分析機器からの計測データを分析室内の NAS (ネットワーク HDD) に集約し、その NAS を経由させることで、部局内でのネットワークを介したデータ共有を可能にするシステムを開発し、全学に対して頒布している (図 1)。そして、その測定データ集約・配信システムの NAS と ONION を S3 (Amazon Simple Storage Service) プロトコルで繋げることで、学内の研究者間での計測データの共有を可能にし、学内における研究データの流通の効率化に寄与してきている。

しかし、メタデータが付与されていない研究データは、真正性の保証もなく、その来歴も不明である。研究データを集約しても、メタデータが不十分であれば、研究データの利活用を進めることは不可能である。そこで、令和 5 年度 IDS 学際共創プロジェクトにおいては、学内データ基盤に集約される研究データの利活用に向けて、共用分析装置からの計測データにメタデータおよび固有識別子を付与し、ONION まで流通させるコンセプトを構築した (図 2)。また、メタデータの付与方法や ONION までのデータ流通の方法に関して技術的検証も行った。

令和 6 年度の本プロジェクトにおいては、この令和 5 年度にお

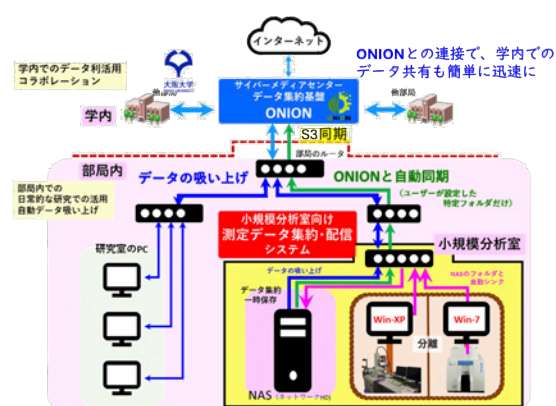


図 1: 小規模分析室向け測定データ集約・配信システム、およびデータ集約基盤 ONION との連携概要図

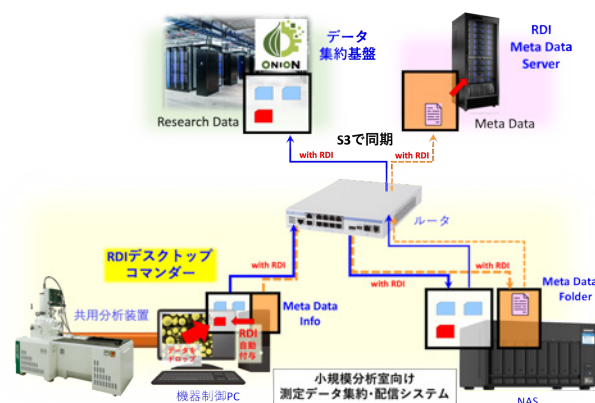


図 2: 小規模分析室向け測定データ集約・配信システムにおけるメタデータ、固有識別子自動付与・流通システムコンセプト図

いて設計したコンセプトを実証するための、モデルシステムの構築を行い、その試行テストを行った。

2 研究の目的

本プロジェクトでは、R5 年度 IDS 学際共創プロジェクトにおいて作成したコンセプトを「メタデータ管理システム」として具現化すべく、以下の研究課題に取り組んだ。

1. 計測データへのメタデータ・固有識別子付与手法の確立。

2. メタデータの登録・検索・閲覧を扱うカタログサーバ構築.
3. メタデータをカタログサーバに送信登録を行うモデルシステムの確立.

このメタデータ管理システムは、共用機器で生成される計測データ全てに研究者の負担を軽減した形で、何らかのメタデータ、固有識別子を付与し、収集、管理することを可能にするものである。また、このモデルシステム構築により、研究データの公開と共有のための原則である FAIR 原則の Findability (検索可能性) と Accessibility (アクセス可能性) を高め、計測データの利活用促進を目指した。

3 研究の方法

本プロジェクトでは、上記の三つの研究課題において以下の手法を用いて進めた。

1. 共用分析装置の制御 PC から NAS 間における計測データの真正性担保に必要なメタデータ・固有識別子付与手法の確立。
 - 計測データの識別に際し最低限必要であるメタデータの選定
 - 計測機器制御 PC でのメタデータ・固有識別子付与におけるフローの確立
 - 付与されたメタデータ・固有識別子の NAS への転送、保存方法の確立
2. メタデータの登録、検索、閲覧を可能にするメタデータサーバの構築。
 - データカタログツール、およびサーバの選定
 - アドホックで作成したメタデータの登録、検索、閲覧を可能にするカタログサービス構築
 - カatalogサービスとデータ集約基盤 ONION へのポータル機能を提供する Nextcloud との連携手法の検討
3. 計測データのメタデータをメタデータサーバに送信、登録する流通方法の確立。
 - 1. で作成されたメタデータ情報と 2. で構築したカタログサービスとの連携
 - メタデータ情報のカタログサービスへの自動登録
 - 登録されたメタデータと ONION に保存された計測データの関連付け

4 研究成果

特許申請予定のため、非公開。

引用文献

- [a] 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, 2022.

発表論文等

(学会発表)

- [1] H. Tanushi, H. Furutani, T. Hosomi, N. Kai, K. Harumoto and S. Date, "Towards Development of University-wide Data Aggregation and Management

Infrastructure for Research Data Utilization," 2024 IEEE 20th International Conference on e-Science (e-Science), Osaka, Japan, 2024, pp. 1-4, doi: 10.1109/e-Science62913.2024.10678692.

(その他)

- [1] 田主 英之, 古谷 浩志, 細見 岳生, 甲斐 尚人, 春本 要, 伊達 進, “大阪大学におけるオープンサイエンス情報基盤の実現に向けて: 研究データ利活用を促す全学データ・メタデータ集約管理基盤構想”, xSIG 2024 (ポスター発表), Aug. 2024.”
- [2] Hideyuki Tanushi, “The Development of Data Infrastructure for the Advancement of Open Science”, PRAGMA 40 (講演), Oct. 2024.
- [3] Hideyuki Tanushi, “Metadata Registration and Management System for Experimental Scientists”, SC2024 (ポスター展示), Nov. 2024.
- [4] 古谷 浩志, “実験系研究者が使う ONION ベースの研究データ基盤を作る - 全学機器共用の立場から”, Cyber HPC Symposium 2025 (講演), Mar. 2025.

(外部資金)

- [1] 2025-2028 年度, 科学研究費助成金 基盤研究 C, “次世代研究データ基盤: メタデータ連携型データ管理システムの構築とその応用”, (代表) 田主英之