実験系データの集約と連動してメタデータ生成・集約・管理を行う概念実証システムの構築

田主 英之 [1)], 古谷 浩志 [1),2)], 細見 岳生 [1),3)], 甲斐 尚人 [4)], 春本 要 [4)], 伊達 進 [1)]

1) 大阪大学 D3 センター 高性能計算・データ分析融合基盤協働研究所
2) 大阪大学 コアファシリティ機構　3) 日本電気株式会社
4) 大阪大学 D3 センター データビリティプラットフォーム研究部門

h-tanushi.cmc@osaka-u.ac.jp

Development of a Proof-of-Concept System for Metadata Generation, Aggregation, and Management in Conjunction with Experimental Data Integration

Hideyuki Tanushi[1)], Hiroshi Furutani[2),1)], Takeo Hosomi[1),3)], Naoto Kai[4)], Kaname Harumoto[4)], Susumu Date[1)]

1) The Joint Research Laboratory for Integrated Infrastructure of High Performance Computing and Data Analysis, D3 Center, The University of Osaka
2) Core Facility Center, The University of Osaka　3) NEC Corporation
4) Datability Platform Research Division, D3 Center, The University of Osaka

概要

大阪大学 D3 センターは，データ駆動型研究の促進と並行し，オープンサイエンスの推進に取り組んでおり，全学データ集約基盤 ONION を試験運用している．この全学データ集約基盤 ONION は，学内外の研究データ集約・共有などのストレージ・利活用の面では大いに利用されているものの，オープンサイエンス視点で研究データを利活用するために必要なメタデータ生成・集約・管理に関しては十分な機能を備えてはいない．本稿では，研究データの ONION 等への集約と連動して，メタデータ生成・集約・管理を行うコンセプトに示し，実験系研究データを対象として，研究者のメタデータ付与の負担軽減も含めて実験系研究データの集約・管理・利活用の基盤となる「実験系研究データマネージメントエコシステム」の構想を紹介する．次いで，その実現に向けた第一歩として，研究データを ONION へ格納すると同時にメタデータも自動登録しカタログ化を行う「概念実証メタデータ管理システム」の構築について報告する．最後に，システム開発を通じて明らかになった将来課題を整理し，オープンサイエンス推進に資するデータ基盤 ONION の将来的な役割を考察する．

1　はじめに

大阪大学 D3 センターは，高性能計算，および，高性能分析を可能にするスーパーコンピュータ SQUID（Supercomputer for Quest to Unsolved Interdisciplinary Datascience）[1]，OCTOPUS（Osaka university Compute & sTOrage Platform Urging open Science）[2] をはじめとした大規模計算機システムを，学内外を問わず，高等教育，学術研究，産業利用に向けて提供している．本センターでは，データ駆動型研究を促進するため，学内外で生み出されるデータを集約し，計算機システムから AI 分析に代表される高性能分析で不可欠となる大容量データへ容易にアクセス可能にするためのデータ集約基盤 ONION（Osaka university Next-generation Infrastructure for Open research and open innovatioN）[3] を SQUID の一部として試験的に運用している．

研究データを公開・利活用することで，研究の加速化や新たなイノベーションを促進するオープンサイエンスの動きが学術界だけでなく，産業界においても活発になっている．大阪大学 D3 センターにおいても，そのようなオープンサイエンスに向けた潮流にいち早く呼応し，「オープンサイエンスを促進する計算・ストレージプラットフォーム」OCTOPUS には，本センターの高性能計算・データ分析融合基盤協働研究所において研究開発した新技術 SCUP-HPC（System

for Constructing and Utilizing Provenance on High-Performance Computing system）[4] が搭載されており，高性能計算システム上で計算来歴を記録・管理する新たなサービスの開始が予定されている [5]．これにより計算結果の真正性と再現性が担保され，研究成果の利活用が促進されることで，新たなイノベーションの創出が期待される．

データ集約基盤 ONION は，学内外の研究データを集約し，研究データの保存，分析，管理，共有の機能を提供し，本学におけるストレージ基盤としての重要な役割を担っている．公開データとなり得る研究成果の根拠データだけでなく，多くの非公開データの保管，管理も ONION の重要な役割である．しかし，現状の ONION は，学内外の研究データ集約・共有などのストレージ利活用機能を提供する一方，研究データをオープンサイエンス視点で利活用する上で十分な機能を提供できていない．オープンサイエンス推進に向けた研究データの公開，利活用において，国際的に求められているデータ共有の原則に FAIR（Findability, Accessibility, Interoperability, Reusability）原則がある．ONION に集約された研究データの利活用を促進し，適切な管理を行うためには，データに関する情報を記述するメタデータを研究データに付与することで，見つけやすく，アクセスしやすく，相互運用性を持たせ，再利用しやすくすることが必要不可欠となる．本稿では，その実現に向けた第一歩として実験系研究データを対象として行った，研究データの集約と連動してメタデータ収集・管理を行う概念実証システムの試みについて報告する．

2 実験系研究データマネージメントエコシステム構想

2.1 データ集約基盤 ONION

データ集約基盤 ONION は異なる 3 種のストレージソリューションで構成される [3]．SQUID と高速なデータ通信を可能にする並列ファイルシステム EXAScaler, 研究データのアーカイブとしての役割も担っているオブジェクトストレージ HyperStore（ONION-object），および Web ブラウザ経由で異なるストレージをシングルビューで表示し，ユーザと ONION 間のデータのやり取りを可能にするオンラインストレージ Nextcloud（ONION-file）である．各ストレージは Amazon Simple Storage Service（S3）プロトコルをサポートしており，ONION のストレージ間のみならず，外部の S3 をサポートする機器，ストレージとの連携が可能である [6].

ONION-file 上では，Nextcloud のパスワード付き URL 発行機能を使用し，学内外の研究者とのデータ受け渡しができる．また，ONION-file から直接機関リポジトリ OUKA（the university of Osaka institUtional Knowledge Archive）[7] へ公開申請を可能にする，Nextcloud の独自プラグイン開発機能を使った ONION-file 向けの研究データ公開申請モジュールを開発し，試験運用を行っている [8]．ONION-object は拡張ストレージとして GakuNin RDM にマウントでき，GakuNin RDM のプロジェクトに ONION-object のバケットをマウントさせることで学術認証フェデレーションに参加している学外の共同研究者とのデータ共有を容易にすることができる．ONION-object では，ユーザ，およびユーザグループが S3 のアクセス制御機能を利用することで，オブジェクト単位からバケット単位まで柔軟なアクセス制御が可能である．ONION にデータを集約させることで，安全に，かつ容易で円滑にデータが流通し，シームレスにデータ公開，および高性能計算システムでの計算・分析を行うことが可能である．

2.1.1 オープンサイエンスデータ基盤としての ONION

ONION は，研究成果を得る過程の研究活動の中で生み出されるさまざまな研究データを集約，分析，管理，共有をすることが主な位置付けである．FAIR 原則に沿ったデータ公開における重要な要素に，データに関する情報を記述するメタデータがある．現状，ONION には研究データと共にメタデータを保存･管理する機能は実装されておらず，FAIR 原則に沿って研究データを利活用する上で十分な機能を果たしているとはいえない．ONION を研究データの公開・利活用を推進するオープンサイエンスデータ基盤として発展させるためには，研究データの集約と連動してメタデータを収集し，管理するシステムと共にデータ管理をする機能を備えることが必要である．

ONION は研究成果の根拠データだけでなく，多くの非公開データの保管，管理も重要な役割である．したがって，ONION に必要なメタデータ管理は，公開データに必須であるメタデータ（資金配分機関など）とは別の，研究者個人，あるいは共同研究者内での研究データそのものの識別，管理，利活用において必要なメタデータ管理機能である．そのようなメタデータ管理システムを備えることで，非公開，公開データの識別も容易になり，共同研究相手との柔軟なデータ共

有，そして研究成果の根拠データの公開が可能になると考える．

2.2　小規模測定室向け測定データ集約配信システム

大阪大学コアファシリティ機構は，学内の部局で小規模単位に運用されている科学計測機器の共用を推進し，計測機器の共同利用のための支援を行う役割を担う．各分析室に設置されている計測機器は，Windows 7 のようなレガシー OS で運用されている場合が多々ある．さらに，セキュリティ面を考慮し，ネットワークから隔離したスタンドアロンで運用されている．それゆえ，利用者は計測データを取得する際には計測機器が設置されている分析室へ直接出向き，データを取得し，その後，USB メモリ等の外部メディアへデータを移す作業が必要であった．この作業は，安全かつ円滑なデータ取得とは程遠く，研究者にとって大きな負担となっている．

本学のコアファシリティ機構は，計測データ取得の際の研究者の負担を軽減するために，小規模分析室向け測定データ集約･配信システムを開発し，全学に展開している．測定データ集約・配信システムでは，分析室内のネットワークルータを介して，共用計測機器から集約される計測データの流れを共用機器から NAS（ネットワーク HDD），NAS から分析室の外へデータを集約・配信する．このデータの流れの方向性は一方向であり，かつセキュリティ上の安全を確保しながら，分析室に出向くことなくネットワーク経由で計測データの取得を可能にしている．

測定データ集約・配信システムを構成する NAS は S3 プロトコルをサポートし，S3 プロトコルを通して NAS の特定フォルダ内の計測データがデータ集約基盤 ONION に同期されるよう設定されている．この D3 センター（旧サイバーメディアセンター）との部局間連携により，部局内で閉じられていたデータ流通が，部局間で行うことが可能になった．また，ONION は GakuNin RDM と連携しており，学術認証フェデレーションに参加している学外の共同研究者ともスタンドアロンの共用機器からの計測データを共有することが可能となっている．コアファシリティ機構は，共用機器の利用促進，および円滑なデータ流通を推進する目的で，開発した測定データ集約・配信システムを全学に向けて頒布している．

2.3　課題

これまで研究データに対するメタデータ付与が十分に発展してこなかったのは，メタデータ付与が研究者にとって大きな負担となる点である．メタデータ付与が研究者にとって大きな負担である理由には，往々にしてメタデータ付与が研究者の手作業に頼らざるを得ない点があげられる．また，研究データのメタデータは書誌データと異なり，分野によって千差万別であり，分野ごとに異なるメタデータスキーマを構築する必要がある点である．その他の理由には，研究者間に研究成果を公開し，利活用してもらい，次のイノベーションに繋げるオープンサイエンスの重要性が十分に浸透していないこともあげられる．特に，実験系の研究においては，大量の計測データの中から研究成果に繋がる計測データが得られる割合が少ない上に，「研究活動において生み出されたすべての研究データを 10 年間保存」などの研究データ管理ルールにおいて，研究成果に直接結びつかない膨大なデータにもメタデータを付与する必要が生じる．この点が，メタデータ付与の普及を妨げる一因になっていると考えられる．

2.4　実験系研究データ向けメタデータ付与・収集コンセプト

メタデータ付与における研究者の負担軽減に向け，実験系研究データ向けメタデータ付与・収集コンセプトを作り上げた [6]．当該コンセプトでは，共用計測機器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 とメタデータ管理システムまで流通させる実験系研究データ向けエコシステム構築を目的とする．

図 1 に構築した実験系研究データ向けメタデータ付与・収集コンセプトの概要図を示す．コンセプトでは，RDI（Research Data Identifier）デスクトップコマンダを使い，固有識別子，計測者などの計測データ識別のために必要最小限なメタデータを付与する（①）．続いて，研究データとメタデータを連動させて転送する（②）．その後，研究データは ONION に送られ，一方でメタデータはメタデータ管理システムに集約され管理される（③）．研究データのオーナーは，メタデー

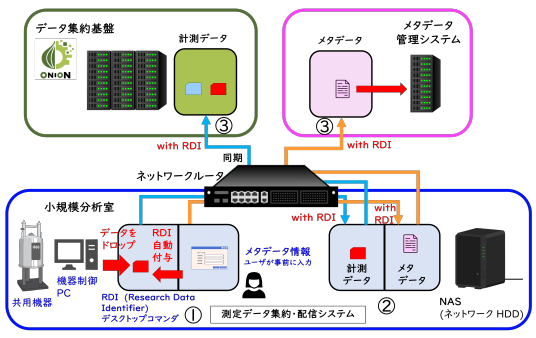

図 1　実験系研究データ向けメタデータ付与・収集コンセプト図

タ管理システム上で必要に応じてメタデータの追記・修正を行う．

3　メタデータ生成・集約・管理を行う概念実証システムの構築

3.1　目的

データ集約基盤 ONION は，研究成果を得る過程の研究活動の中で生み出されるさまざまな研究データを集約，分析，管理をすることが主な位置付けであり，公開データとなり得る研究成果の根拠データだけでなく，多くの非公開データの保管，管理も ONION の重要な役割である．したがって ONION に必要なメタデータ管理は，公開データに必須であるメタデータ（資金配分機関など）とは別の，研究者個人，あるいは共同研究者内での研究データそのものの識別，管理，利活用において必要なメタデータ管理機能である．そのようなメタデータ管理システムを備えることで，公開，非公開データの識別も容易になり，共同研究相手との柔軟なデータ共有が可能になると考える．

メタデータ生成・集約・管理を行う概念実証システムの構築においては，2.4 節の実験系研究データ向けのメタデータ付与・収集コンセプト（図 1）にあるように，研究データを ONION に同期するのと連動して，メタデータ管理システムにメタデータの登録をし，ONION と連携したメタデータ管理システムの構築が必要であると考え，まずは，研究データを ONION へ同期するのと並行してメタデータを登録し，その情報を検索できることを確認する，という技術的な実現可能性の検証を優先し，PoC (Proof of Concept) プロジェクトとして開発を行った．

3.2　システム概要

本研究プロジェクトでは，メタデータ管理プラットフォーム Apache Atlas [9] を使用し，データ活用社会創成プラットフォーム基盤高度化システム mdxII [10] 上に概念実証システムを構築した．実験系データの集約と連動してメタデータ生成・集約・管理を行う概念実証システムを以下のように設計した（図 2）：

- 研究データのメタデータ作成，および研究データとメタデータのセットを一時保管領域に転送.
  - ① 研究者がメタデータ出力プログラムを使い，json 形式のメタデータファイルを作成.
  - ② メタデータファイルを OpenSSL を用いて暗号化し，研究データとメタデータファイルをセットで，SCP プロトコルを用いて NAS 上のデータ一時保管領域に格納.
- 研究データとメタデータを ONION とメタデータ収集・管理システムへ転送.
  - ③ メタデータ収集・管理システムが，定期的にデータ一時保管領域の新規データセットの有無をチェック.
  - ④ 新しいデータセットが配置されたことがトリガーとなり，研究データを ONION に，メタデータをメタデータ収集・管理システムのストレージに転送.
- 研究データの格納先アドレス取得，メタデータに追記.
  - ⑤ 研究データを ONION に転送後，格納先アドレスを取得し，格納先アドレスをメタデータファイルに追記.
- Apache Atlas にメタデータを登録.
  - ⑥ メタデータファイルを複号化し，メタデータ管理プラットフォーム Apache Atlas にメタデータを登録.

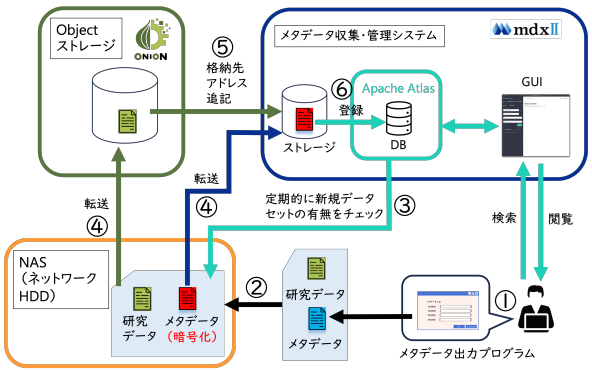

図 2　実験系データの集約と連動してメタデータ生成・集約・管理を行う概念実証システム図

研究データはローカル PC で作成したテキストファイルで代用した．メタデータの項目は，計測機器から得られる計測結果を参考に設定し，その項目に従ってテスト用のメタデータファイルを作成することで，システムの開発および検証を行った．通常，ONION やメタデータ管理システムへのファイル転送に用いるデータ一時保管領域（NAS 等のネットワークドライブ）は，各研究室の管理下に置かれることが想定される．しかし本プロジェクトでは，mdxII 上のメタデータ管理システムとは別の領域に NAS サーバを構築し，これを仮想的なデータ一時保管領域として扱った．また，メタデータ管理システムへのメタデータの登録，メタデータの検索，閲覧機能の開発を優先させるため，本プロジェクトではユーザ認証機能は組み込んではい

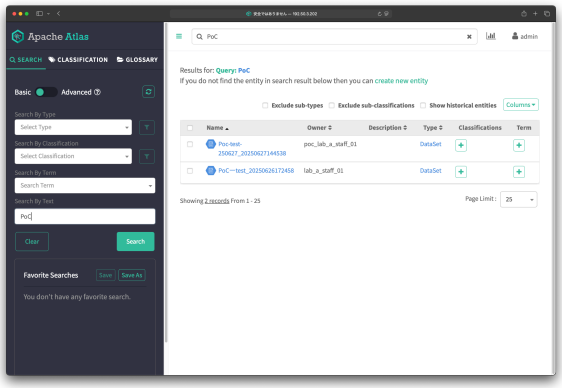

図 3　ApacheAtlas によるメタデータ検索結果の例

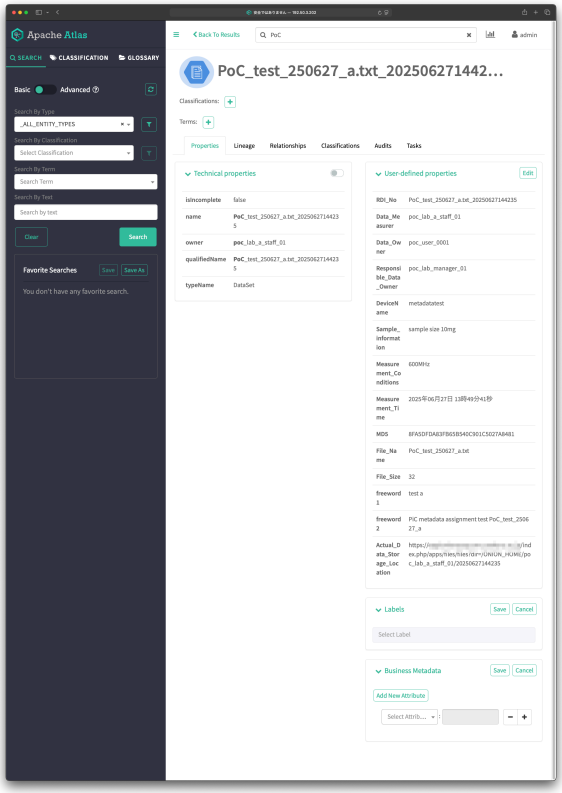

図 4　ApacheAtlas によるメタデータ一覧の例

ない．

メタデータ管理システムに登録されたメタデータは，当該研究データのオーナーが Apatche Atlas にログインすることで閲覧できる（図 3）．また，一覧の中からメタデータの項目名等や内容等を参照して該当する研究データを検索可能である（図 4）．そして，検索結果の中の研究データ名をクリックすることで，その研究データのメタデータ一覧が表示される．研究データの ONION 上の格納場所は自動的に取得され，URL の形式でメタデータの一覧に表示される．

4　考察

データ集約基盤 ONION は，研究活動の中で生み出されるさまざまな研究データを集約し，研究データの保存，分析，管理，共有の機能を提供し，本学におけるストレージ基盤としての重要な役割を担っている．研究データがオープンサイエンス視点で利活用されるためには，研究データが見つけやすく，アクセスしやすく，相互運用性があり，再利用しやすいことが重要である．単にデータを保存するだけでは，所有者以外がデータの所在や内容を把握することは困難であり，データに関する情報を記述するメタデータの付与がオープンサイエンス視点での研究データ利活用実現のためには必須である．しかし，現在の ONION には，学内外の研究データ集約・共有などのストレージ利活用機能を提供する一方，メタデータと連動して研究データを管理する仕組みは導入されていない．

本研究プロジェクトでは，データ集約基盤 ONION と連携したメタデータ生成・集約・管理システムを PoC プロジェクトとして開発した．開発したメタデータ生成・集約・管理システムでは，研究活動により生み出された研究データ，およびそのメタデータが書き込まれたメタデータファイルを，メタデータ管理システムと連携したネットワークドライブ内の特定の格納場所にアップロードすることで，自動的に研究データが ONION へ転送され，メタデータの内容がメタデータ管理プラットフォーム Apache Atlas に登録されることを実現した．転送時に研究データの ONION における格納先アドレスを自動的に取得してメタデータへ追記することで，Apache Atlas 上でメタデータの値から目的の研究データを探索可能にした．

しかし，本プロジェクトではメタデータ管理システムへのメタデータの登録，メタデータの検索，閲覧機能の開発を優先させたため，実用化に向けていくつかの解決すべき課題が明らかになった．まず，ONION に保存される研究データの多くは研究途中の非公開データであり，メタデータも同様に研究データの所有者以外には非公開である必要がある．その実現のためには，メタデータ管理システムにもユーザ管理機能が必須である．また，そのユーザ管理も ONION のアクセス制御と連携させることで，Apache Atlas 上からシームレスに ONION 上の研究データにアクセスすることが可能になると考える．

次に，学内外のさまざまなデータソースからのデータを集約する ONION において，メタデータ管理シス

テムは研究分野ごとに異なるメタデータを柔軟に受け入れる必要がある．研究分野ごとに明確なメタデータスキーマ定義を作成することは，メタデータの管理負担を軽減することに繋がるが，各研究分野にとって最適なメタデータスキーマはそれぞれが定義することである．理想としては，システム側で厳格なスキーマを規定するのではなく，利用者がそれぞれの目的に沿ったメタデータスキーマを柔軟に作成・適用できる仕組みを提供することが最善であると考える．メタデータスキーマに関しては，今後 ONION へのデータ保存，およびメタデータ管理システムに関心を寄せていただける共同研究者とともに，構築に向けた研究を進めていきたい．

5　まとめと将来課題

本稿で提案，試験構築した ONION と連携した生成・集約・管理システムは，研究活動の中で生み出され，ONION に集約される研究データを，研究者に負担を掛けずに研究データと対応するメタデータを連動させて一括して適切に管理する上で，非常に重要なステップである．研究成果としてデータ公開へ繋げていく上で，明確なスキーマに従ったメタデータが付与されていることは，FAIR 原則に沿った研究データ公開，利活用の推進に必須である．しかし，本稿でも示したように，利活用に必要なメタデータ作成は，研究者にとって非常に大きな負担である．本稿で試験構築した概念システムは，これらの課題を解決する 1 つのアプローチである．

多くの学術機関が同時に高等教育機関でもあり，研究室は，数年単位の短い周期で入れ替わる学生が学位取得のために行う研究で生み出すデータを管理する責任を負っている．学生が生成する研究データが公開に至ることは稀であるものの，研究室がこれらのデータを適切に管理する必要がある．ONION と本メタデータ管理システムはこれらの用途にも有用なツールとなり得る．今後は，現概念実証システムに必要な機能等を追加していき，実際の実験系研究データと対応するメタデータの一括･連動管理に試みたいと考えている．

次世代計算・ストレージ基盤 OCTOPUS に搭載された計算来歴記録・管理システム SCUP-HPC の来歴情報とメタデータ管理システムの連携，そして学内外の研究室・研究機関の ONION とのデータ連携も今後推し進めていく課題である．D3 センターはオープンサイエンス推進を目的に，積極的に他部局，あるいは他研究機関との連携を推し進めており，今後，メタデータ管理システムを充実させ，オープンサイエンスデータ基盤としての ONION の発展に取り組んでいく所存である．

謝辞

本研究の成果は，データ集約基盤 ONION，および，データ活用社会創成プラットフォーム基盤高度化システム mdxII を利用して得られました．また，本研究の一部は，以下の助成をうけ行われました：

- R6 年度大阪大学データビリティフロンティア機構（IDS）学際共創プロジェクト
- JSPS 科研費 25K15811

参考文献

[1] Susumu Date, Yoshiyuki Kido, Yuki Katsuura, Yuki Teramae, Shinichiro Kigoshi. Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) and its Five Challenges, Sustained Simulation Performance 2021, 2022. [DOI: 10.1007/978-3-031-18046-0_1]

[2] 大阪大学 D3 センター大規模計算機システム：OCTOPUS. https://www.hpc.cmc.osaka-u.ac.jp/octopus2（2025 年 9 月 21 日参照）

[3] 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，Vol. 26, No. 1, pp. 87 – 96, 2022. [DOI: 10.24669/jacn.26.1_87]

[4] Yuta Namiki, Takeo Hosomi, Hideyuki Tanushi, Akihiro Yamashita, Susumu Date. SCUP-HPC: System for Constructing and Utilizing Provenance on High-Performance Computing Systems, IEEE Access, vol. 13, pp. 141090-141107, 2025. [DOI:10.1109/ACCESS.2025.3597361]

[5] プレスリリース: 大阪大学 D3 センター，NEC が構築した新たな計算・データ基盤の運用を開始～計算来歴を記録・管理する技術の開発・導入によりオープンサイエンスを促進～, https://www.hpc.cmc.osaka-u.ac.jp/wp-content/uploads/2025/09/20250912-OCTOPUS2PR.pdf （2025 年 9 月 21 日参照）

[6] Hideyuki Tanushi, Hiroshi Furutani, Takeo Hosomi, Naoto Kai, Kaname Harumoto, Susumu

Date. Towards Development of University-wide Data Aggregation and Management Infrastructure for Research Data Utilization, NRDPISI-1, eScience2024, Osaka Japan, Sep. 2024. [DOI: 10.1109/e-Science62913.2024.10678692]

[7] 大阪大学学術情報庫：OUKA について．https://ir.library.osaka-u.ac.jp/portal/about.html（2025 年 9 月 22 日参照）

[8] 田主 英之，山下 晃弘，細見 岳生，並木 悠太，甲斐 尚人，松浦 かんな，伊達 進． 研究データ管理を支える学内情報基盤連携の実現に向けて，学術情報処理研究，Vol. 27, No. 1, pp. 98-105，2023. [DOI: 10.24669/jacn.27.1_98]

[9] Apache Atlas. https://atlas.apache.org/（2025 年 9 月 21 日参照）

[10] mdxII. https://mdx.jp/mdx2（2025 年 9 月 21 日参照）