

大阪大学のデータ集約基盤 ONION

伊達 進¹⁾, 寺前勇希²⁾, 勝浦裕貴²⁾, 木越信一郎²⁾, 木戸善之¹⁾

1) 大阪大学サイバーメディアセンター

2) 大阪大学情報推進部

date@cmc.osaka-u.ac.jp

ONION: Osaka University's Data Aggregation Infrastructure

Susumu Date¹⁾, Yuki Teramae²⁾, Yuki Katsuura²⁾, Shinichiro Kigoshi²⁾, Yoshiyuki Kido¹⁾

1) Cybermedia Center, Osaka University.

2) Department of Information and Communication Technology Service, Osaka University

概要

近年、学術研究がますます広域化・グローバル化する傾向にある。さらに、様々な科学分野で高性能データ分析の利活用に対する関心と期待が急速に高まりつつある。このような背景から、大学、学術研究機関、企業の研究者らがデータを移動・共有しつつ、共同研究、産学共創を高効率に推進することを可能にするデータ基盤の必要性・重要性が高まっている。大阪大学サイバーメディアセンターでは、このような視点から、2021年5月のスーパーコンピュータ SQUID 導入に合わせ、データ集約基盤 ONION の試験運用を開始した。本稿では、われわれがどのような考えに基づき ONION を着想・構想したのかとともに、ONION の概要を概説する。その後、ONION で想定するユースケースについて検証を行う。

1 はじめに

大阪大学サイバーメディアセンター (CMC) では、2014年11月に導入されたベクトル型スーパーコンピュータ SX-ACE の後継機として、2021年5月に高性能計算・データ分析基盤システム SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datasience) [1] を導入し、3ヶ月間の試験運転を経て、8月より正式な運用を開始した。われわれは、この SQUID の実現に向け、2017年度後半期より約4年間の歳月を費やし、近年の高性能計算環境を取り巻く環境、CMCの大規模計算機システム事業の現状、大阪大学の戦略目標等を鑑み、CMCが導入・運用すべきスーパーコンピュータシステムのあり方について検討を通じて、SQUID 調達に全力で取り組んできた。その結果、上記目的を実現するために設定した5項目のチャレンジ、すなわち、(1) 高性能計算・高性能データ分析 (HPC・HPDA) 融合、(2) クラウド連動・連携機能、(3) セキュアコンピューティング環境、(4) データ集約基盤、(5) テイラーメイド型計算環境を

SQUID で実現することができた。

本稿では、上記のうち、CMC が実現した(4)データ集約基盤 ONION についてまとめる。本稿の構成は以下の通りである。まず2節では、ONION 導入に至る背景と要求要件についてまとめる。その後、3節において、データ集約基盤 ONION について概説する。4節では、ONION 上で想定されるユースケースを検証する。5節で本稿をまとめる。

2 ONION 導入に至る背景とアプローチ

2.1 ONION 導入の背景

近年では、様々な科学分野における学術研究が広域化・グローバル化傾向にある。今日の学術研究では、異なる研究機関や大学の研究機関、および、企業の研究者らが、特定の科学問題の解決のためチームとして研究活動を行うことが日常的なものとなりつつある。2019年3月に CMC が主催した Cyber HPC Symposium 2019 における本学所属研究者の講演内でも、University of Kentucky、CMC、理学研究科、国立天文台等に所属する研究者らがクラウドストレージ等を活用しながら相互に研究

データを共有・交換している例が示されている [2]。また、今日では様々な学術分野の知識、専門性、技術の集約が不可欠となる学際融合領域での研究も盛んであり、今後の学術研究の発展には地理的に分散した研究者がチームとなり研究活動を推進していくことが不可避となることが予見される。

また、大阪大学では、OU ビジョン 2021 [3] のもと、専門分野を超え、広く世界と協働する新たな知の創出を目指す「Open Research」、社会のニーズに基づく基盤研究の課題を発掘し、新たな社会的価値の創出につなげる「Open Innovation」が示され、より一層国内、国外、産学の枠を超えて連携・協働することが全教職員に求められている。こうした状況から、大阪大学の教育・研究を支える情報基盤の整備・安定運用を担う CMC には、学内の研究者らが産学共創、国際共同研究を高い生産性をもって推進することを支援できる計算基盤およびデータ基盤の実現への高い期待と関心が寄せられている。

さらに、今日では、学術研究の再現性 (Reproducibility) の確保にむけた動きが活発化している。事実、国内外の大学や研究機関において、公正な研究活動の推進に向けて、研究活動に伴い作成・取得した研究データの保存期間および管理方法等についての基準を定めたガイドラインの策定が進められている状況にある。その一方、学内には様々な科学データ計測機器が存在しており、そのような科学計測機器から収集される計測データを適正に大学として保存・管理する仕組みは存在していない現状がある。さらに、そのような科学計測機器から収集される計測データをはじめとした大容量・大規模データの解析・計算結果もまた適切に保存・管理する仕組みは未整備である。そうした現状と背景から、今後ますます大容量・大規模化していくことが予想される研究データを適正に管理するための全学規模でのデータ基盤の整備・運用は、全学の情報基盤を整備・安定運用する責務を担う CMC にとって喫緊の課題である。

さらに、近年の AI、ML、DL といったキーワー

ドに代表される高性能データ分析 (HPDA: High Performance Data Analysis) の利活用に対する期待と関心の高まりは、これまでスーパーコンピュータを利用していなかった研究者のスーパーコンピュータ利用を促進しつつある。その結果、従来のスーパーコンピュータ調達で整備していた、SCP や FTP といった CLI (Command Line Interface) ベースのプロトコルを用いて利用者環境とスーパーコンピュータ間のデータ移動のみを考慮するのではなく、スーパーコンピュータに関する詳細な知識や技術を有していない利用者がより直感的かつ容易にデータを移動・操作できる環境の必要性・重要性が急速に高まりつつある。

2.2 ONION 実現へのアプローチ

このような背景と現状を鑑み、われわれは、広域・グローバル化する学術研究において、チームとなって協働する研究者が自在にデータを共有・交換・移動できるデータ基盤の必要性・重要性に着目し、2017 年後半期より、本学の教育・研究の情報基盤を支える責務を担う CMC としてどのようなデータ基盤が必要であるか？について検討を行ってきた。その結果、本学のスーパーコンピュータが具備すべきデータ基盤のあり方を検討し、本学のビジョンに示される Open Research および Open Innovation を加速・進展させることのできるデータ基盤 ONION (Osaka university Next-generation Infrastructure for Open research and Open innovatioN) の構想に至っている。

しかし、この ONION の構想を実現するためには、重要な 2 点の解決しなければならない問題がある。1 点目は ONION 実現のための予算、2 点目は人的資源である。

スーパーコンピュータを利用する研究者のストレージニーズ・要求に加え、全学規模の研究データを収容可能なストレージニーズ・要求を充足可能なデータ基盤を整備するための予算は、CMC には手当てされていない。また、そのような全学規模のデータを安全かつ安定的に管理できるデータ基盤の運用は、現状の CMC の大規模計算機シス

テム事業の人員が当該事業に加えて片手間にできる業務量ではない。

しかし、2.1 節に記載した通り、広域化・グローバル化する国際共同研究、産学共創を加速させることのできるデータ基盤の整備は、本学だけでなくわが国の学術研究の発展のための急務である。そうした背景から、われわれは、2021 年 5 月に高性能計算・データ分析基盤システム SQUID の一部として ONION を PoC (Proof of Concept) 実装として試験的に導入するアプローチを選択した。SQUID 調達で ONION を試験的な導入に留める主な理由としては、(1) 本格的に大規模に ONION を整備することは、主目的である高性能計算・データ分析基盤システムで必要となる計算資源のサイズを大幅に縮減してしまい、ますます拡大傾向にある計算ニーズ・要求を満たすことができなくなる恐れがあること、(2) 本センターで構想した ONION が有用ではなく、利用されない場合もありえること、の 2 点があげられる。このような理由から、われわれは、ONION を試験導入し、利用者実際に利用していただきながら、改善、機能拡張・増強を経て成長させていく方針である。

2.3 ONION に向けた要求要件

2.1 節で記した通り、ONION では、スーパーコンピュータの利用者によるストレージ容量や I/O 性能の充足だけではなく、広域・グローバル環境で行われる国際共同研究、産学共創で発生する研究者間のデータ移動・交換・共有ニーズを充足する必要がある。それゆえ、ONION では、SQUID で導入するストレージに学内外のデータを集約するとともに、また必要に応じて学外の研究機関・大学および民間クラウドの提供するクラウドストレージ間で容易かつ直感的にデータ移動が可能であることが求められる。

ONION の実現にむけ、ONION が満たすべき要求要件として、われわれは下記 7 項目を設定した。

- (1) キャンパス内に存在する科学データ計測機器、IoT センサ機器等の各種データ生成源からのデータを収容可能であること。

本学は総合大学であり、クライオ電子顕微鏡等の様々な科学データ計測機器が存在する。また、近年の IoT 技術の発展により、インターネットに接続可能な様々な IoT センサが存在する。今日では、データ同化シミュレーション、ビッグデータ解析など、こうしたデータ源から生成されるデータを即座に利活用する大規模計算・データ分析に対する需要・ニーズがある。

- (2) スーパーコンピュータ利用者以外の研究者が ONION 上の研究データにアクセスできること。

国際共同研究、産学共創活動では、スーパーコンピュータを利活用する研究においても、必ずしも研究グループの全ての研究者がスーパーコンピュータのアカウントを有していない。この場合においても、計算で利用する入力データ、あるいは計算結果を、アカウントを有していない研究者に対してもアクセスを提供したい場合がある。

- (3) 研究者グループ代表者にストレージ管理権限を委譲でき、スーパーコンピュータを利用しない研究者にもアカウントを発行できること。

計算基盤であるスーパーコンピュータを利用しない研究者が ONION を利用するためだけにアカウントを発行することは非現実的である。そのため、CMC はストレージ容量を切り分け、そのストレージの管理権限を各研究者グループに付与し、各管理者にアカウントの発行を委譲したい。

- (4) 利用者個別に利用者のストレージを ONION と連携可能であること。

研究者は、その研究者の所属する研究環境でストレージ環境を保有している場合がある。この際、当該ストレージ環境と ONION が連携できることが望ましい。

- (5) ONION、および、ONION と連携する各種ストレージ群のシングルディスクイメージ (SDI: Single Disk Image) ビューを利用者に直感的に提供できること。

ストレージ構成に詳しくない利用者に対して、それらストレージへのアクセス情報等を意識させることなく、利用者が直感的なデータ操作を行え

ントを作成できる。ONION では、この機能を利用することにより、CMC の管理者によるアカウント作成・管理業務の低減をねらう。

3.2 ONION の構成

以下では、ONION を構成する並列ファイルシステム EXAScaler、オブジェクトストレージ HyperStore、および、オンラインストレージソフトウェア NextCloud について概説する。

・ 並列ファイルシステム EXAScaler

2021 年 5 月に導入された SQUID は、汎用 CPU ノード群、GPU ノード群、ベクトルノード群の合計 1598 ノードが 200Gbps の相互結合網 InfiniBand HDR で接続されたスーパーコンピュータである。それゆえ、SQUID のストレージシステムには、これまでのスーパーコンピュータ以上に、SQUID 上で同時実行される多数のジョブによるファイルアクセスの高速性が要求される。上述した通り、SQUID では、このようなファイル I/O の高速性を十分に満たすことのできる並列ファイルシステムとして、Lustre ファイルシステムをベースとする DDN 製 EXAScaler を採用した。

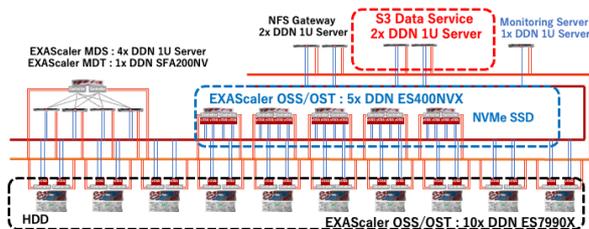


図 3: ONION 構成並列ファイルシステム EXAScaler の構成図.

ONION 構成ファイルシステム EXAScaler では、限られた予算制約のもと 2.3 節で記した要件(6)を実現するために、われわれは HDD と SSD を併用する構成を採用した(図 3, 表 1)。具体的には、ONION-EXAScaler では、HDD 用に DDN ES7990X を 10 基、SSD 用に DDN ES400NVX を 5 基導入し、HDD および SSD 領域それぞれのファイルシステムを構築している。利用者が直感的に使いやすい

ように、これらのファイルシステムはシングルディスクイメージのサブツリー構造として提供される。すなわち、利用者は SSD 領域のディレクトリにデータを配置し、ジョブ実行を行うだけで、SSD の提供する I/O 性能を享受できるようにしている。

さらに、ONION-EXAScaler では、2 節で示したわれわれの ONION 構想の実現のために、DDN S3 データサービス(S3DS)を導入した。この S3DS サーバを通じて、利用者が S3 に対応したクライアントから並列ファイルシステム EXAScaler 上にデータをファイルとして格納できるよう設計している。さらにいえば、この仕組みを利用することで、学内に設置された科学計測機器あるいはセンサ等からデータを自動的に集約する仕組みもまた構築可能である。

表 1: ONION 構成並列ファイルシステムの性能.

実効容量(HDD)	20.00 PB
実効容量(NVMe SSD)	1.20 PB
最大合計 inode 数	約 88 億個
最大想定実効スループット (HDD)	160 GB/s 以上
最大想定実効スループット (NVMe SSD)	write : 160 GB/s 以上 read : 180 GB/s 以上

・ オブジェクトストレージ HyperStore

Cloudian 製 HyperStore は、前述した S3DS 同様に、S3 API をサポートする。そのため、外部のクラウドや、S3 対応ストレージとの相互のデータ交換を容易に行うことができる。さらに、この HyperStore は、高信頼性、高拡張性、高可用性を特徴としている。この HyperStore では、最小 3 ノード以上からストレージシステムを構成でき、必要に応じてサーバ 1 ノード単位で拡張を行うことができる柔軟な拡張性を備えている。それゆえ、2.2 節で記したように、ONION を PoC (Proof of Concept) 実装として試験的に導入するわれわれのアプローチとの親和性が高く、今後、試験導入した ONION の利用実績の向上に合わせ拡張を行っ

ていくことを考えている。

また、この HyperStore は、データが自動的に保護される機能を有しており、高信頼性を提供している。その方式として、レプリカ(複製)方式と、イレージャー・コーディング(EC)方式が利用可能である。前者は HyperStore にデータが格納される際に、自動的に書き込まれるデータのコピー(複製)をノード間に分散する方法であり、後者は格納されるデータに加えて、データを復元するための符号を付加して、それらを複数のサーバ間に分散する方法である。このようなデータ保護機能をもたせることで、オブジェクトストレージを構成する一部のサーバの障害時においてもデータを失うことのない仕組みになっている。ONION では、データ冗長性より実効データサイズ容量を優先して、後者の EC 方式を採用した。

本学は、主要キャンパスとして豊中、吹田、箕面キャンパスを有している。HyperStore は、一部のノードが障害により利用できない場合においても、システムは停止することがなく、可用性が高い。それゆえ、ONION での試験導入後、本格運用に移行することになれば、大学キャンパスの地理的な分散を利用し、キャンパスの停電時にもデータが利用できる環境を構築することも可能となる。

・ オンラインストレージ NextCloud

ONION では、オンラインストレージソフトウェア NextCloud を、データ集約基盤 ONION へのポータルとして採用している。ONION では、利用者はスーパーコンピュータシステムのアカウトとパスワードを用いて、NextCloud ポータルよりログインする。ログイン後、ONION-EXAScaler で提供されるホームディレクトリの閲覧が可能である。

この ONION へのポータルとなる NextCloud は、3.1 節に記したが、外部ストレージ機能を提供している。ONION-NextCloud では、利用者は WebDAV と S3 プロトコルに対応したストレージを連携させることを可能としている。例えば、研究室で運用している WebDAV 対応ストレージサーバと連携させ、ONION-NextCloud 上から当該ストレージデ

ータにアクセスすることができる。また、上述した ONION-HyperStore、AWS S3 などの S3 対応ストレージとも容易に連携させることができる。さらに、NextCloud 上では、任意のファイルやディレクトリを選択し、それらに対して URL とパスワードを発行し、スーパーコンピュータを利用していない利用者に対してダウンロード、アップロードさせることもまた可能である。

4 想定ユースケースの検証・評価

本節では、3 節で概説した ONION について、想定されるユースケースの検証・評価を行う。

4.1 SQUID を利用してない研究者とのデータ共有

CMC の従来環境では、スーパーコンピュータ利用者がプログラムや計算結果をスーパーコンピュータを利用してない研究者と共有するためには、スーパーコンピュータのフロントエンドサーバから SCP や FTP を用いて自身のローカル環境にダウンロードした後、メールで送付する、自身の設置したウェブサーバにアップロードする、あるいは、Google Drive などのクラウドストレージにアップロードするなどの方法に限定されていた。一方、ONION では、利用者は NextCloud にログインし、対象となるデータをブラウザ上で選択し、共有設定を行うことで可能である。



図 5: スマートフォンからの NextCloud 利用例。

図 5 は ONION を用いて実際に検証利用した例である。SQUID ではスケジューラシステム NQSV を導入しており、ジョブ投入時にジョブ完了後にメールが送信されるように設定しておくことができる。この機能と ONION の機能を併用すること

で、スーパーコンピュータのフロントエンドサーバにアクセスできる端末が利用できない外出中においても、メール受信後(①)即座にスマートフォンのブラウザ経由で ONION-NextCloud にログイン(②)し、SQUID 上に生成された計算結果ファイルを確認(③)する。その後、海外の共同研究者に計算結果ファイルの共有設定(④)を行い、当該ファイルにアクセスするための URL とパスワードをメールにて送付可能である。また、紙面の関係で詳細は省略するが、計算結果やプログラムを ONION-HyperStore に移動した後、上述の方法と同様に ONION-HyperStore 上で共有設定を行うことで SQUID を利用していない利用者とデータ共有することも検証済みである。

4.2 S3 クライアントからのデータ集約

次に、科学計測機器から生成されたデータを收容した S3 対応ストレージあるいは S3 対応 IoT データを想定して、S3 に対応したクライアントから ONION へのデータ移動が可能であるかどうかを検証する。S3 クライアントには、S3 プロトコルでのデータ転送を可能にするツール *s3curl* をインストールした Localpc(MacBookPro, Intel Corei7 3.5GHz, 16GB; Mac OS 10.13.6)を用いる。検証実験では、ONION-EXAScaler および ONION-HyperStore それぞれの S3 エンドポイントに *s3curl* でアクセスし、Bucket を作成、当該 Bucket に put 処理が正常に行われることを検証する。

図 6 に S3 クライアントを想定した Localpc から *s3curl* ツールを用いて ONION-HyperStore にデータ転送した際の手順を示す。クライアント上で、ONION-HyperStore の S3 エンドポイント、アクセスキー、シークレットキーを含む設定ファイル *.s3curl* を作成する(0)。データ転送用 Bucket *transfertest* を作成(1)し、検証用アップロードファイル *testtransfer.txt* ファイルを put する(2)。本検証試験では、この手順により正しく Localpc から ONION-HyperStore へのデータ移動が正しく行われることを確認した。その後、ONION-EXAScaler 側の S3 エンドポイント(S3DS)に対しても同様の

手順により動作確認した。この仕組みを利用したスクリプトなどの作成により、図 7 に示すような S3 に対応したデータ源から ONION にデータを自動的に集約する仕組みが構築可能である。

(0) 設定ファイル.s3curlの作成

```
Localpc:~ date$ vi .s3curl
-----
%awsSecretAccessKeys = (
dateoncloudian => {
  id => '7fcbf6xxxxxx5adbb4',
  key => 'xxxxxxxxxxxxxxxxxxxxxxxx',
},
);
push(@endpoints, 's3-osakau.oniongw.hpc.cmc.osaka-u.ac.jp');
```

(1) Onion-HyperStore上にtransferTestバケットを作成

```
Localpc:~ date$ s3curl.pl --id dateoncloudian --createBucket -- -s https://s3-osakau.oniongw.hpc.cmc.osaka-u.ac.jp/transfertest
```

(2) transfertestバケットにtesttransfer.txtファイルをアップロード

```
Localpc:~ date$ s3curl.pl --id dateoncloudian --put testtransfer.txt -- -s https://s3-osakau.oniongw.hpc.cmc.osaka-u.ac.jp/transfertest/testtransfer.txt
```

図 6: S3 クライアントから HyperStore へのデータ移動手順。

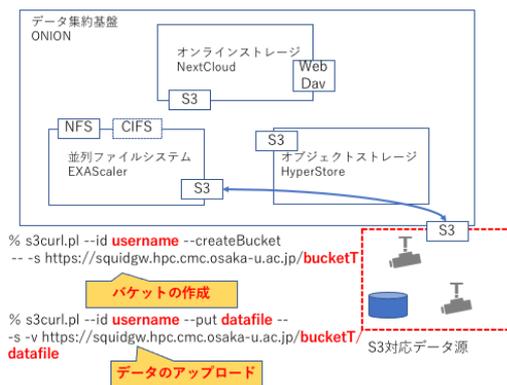


図 7: S3 クライアントから S3DS へのデータ移動。

4.3 EXAScaler – HyperStore 間のデータ移動

次に、ONION 内に配備されている、S3DS および HyperStore の S3 エンドポイント間でのデータ移動の動作確認を行う。SQUID のフロントエンドサーバに *s3curl* ツールをインストールした後、HyperStore 上にデータを put、HyperStore 上のデータを get した。その結果、S3DS および HyperStore 間の双方向データ移動を確認できた。

4.4 ONION-ExaScaler の性能評価

最後に、ONION-EXAScaler の I/O 性能を計測する。3 節で記した通り、ONION を構成する並列ファイルシステムは、HDD および SSD を併用した

構成である。本検証実験では、HDD および SSD 領域上で IOR による I/O 性能評価を行う。

計測には、SQUID の CPU ノード(Intel Xeon Platinum 8368(2.40GHz/38C), 256GB)群を用いて、SSD 領域のファイルシステムを用いた場合および HDD 領域を用いた場合の read 性能、write 性能を計測する。

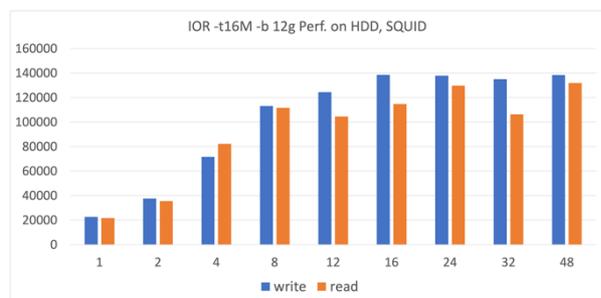


図 8: EXAScaler の HDD 領域での IOR 性能。

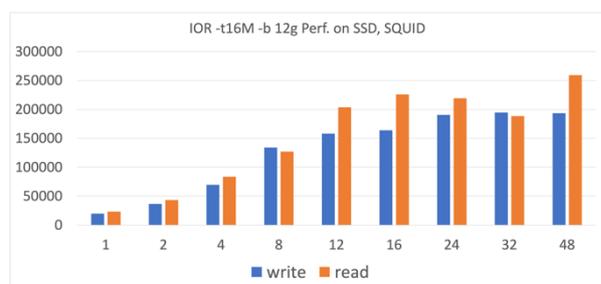


図 9: EXAScaler の SSD 領域での IOR 性能。

図 8、図 9 にそれぞれ HDD 領域および SSD 領域での IOR(ver.2.10.3) [5] の性能計測結果を示す。計測は SQUID 試験運転中の 6 月中に実施した。いずれの場合も、利用するノード数を 1 ノード(76 コア)から 48 ノード(3648 コア)まで並列度数を変化させて計測した。この結果、HDD 領域を用いた場合は、最大性能としてそれぞれ 138.3 GB/s(write)、131.8 GB/s(read)、SSD 領域を用いた場合は 193.3 GB/s(write)、259.3 GB/s(read)が得られ、HPC および HPDA に要求される実用上十分な I/O 性能を確認することができた。

5 まとめと今後の課題

本稿では、2021 年 5 月に導入した SQUID にあわせて試験導入した、大阪大学のデータ集約基盤 ONION についてまとめた。並列ファイルシステム

EXAScaler、オブジェクトストレージ HyperStora およびオンラインストレージソフトウェア NextCloud を相補的に相互作用させることにより、ONION は HPC・HPDA に要求される I/O 性能と広域下で推進される共同研究や産学共創でのデータ移動・共有を支援する機能性を実現した。

この ONION に対しては、本稿執筆時点においても、すでに利用者からの問い合わせが数多く寄せられている。令和 2 年度第三次補正予算「先端研究設備整備補助事業（研究施設・設備・機器のリモート化・スマート化）」において採択された本学のプロジェクトにおいては、学内の科学計測機器から集約された計測データを ONION へ自動収容するシステムの構築が現在進行しており、また「先端研究基盤共用促進事業(先端研究設備プラットフォーム)」に採択されたパワーレーザー DX プラットフォーム事業においても ONION の試用が開始されている。また、従来のスーパーコンピュータ利用者からも、I/O 性能だけでなく、4 節で示したような利用方法について評価いただく声をいただいている。

今後、ONION が真の意味で本学だけでなく我が国の学術研究を支えられるよう、ONION をお使いいただく利用者の声に耳を傾けながら、われわれは、改善、機能拡張・増強を経て ONION の発展に全力で取り組んでいく。mdx [6]、GakuNin RDM との連携は今後の重要課題として位置付けている。

参考文献

- [1] 大阪大学サイバーメディアセンター, SQUID, <http://www.hpc.cmc.osaka-u.ac.jp/squid/>, 2021.
- [2] 長峯健太郎, “理学研究とクラウド利用のニーズ: 理論宇宙物理学の例”, Cyber HPC Symposium 2019, 大阪, Mar. 2019.
- [3] 大阪大学, OU ビジョン 2021 – 社会変革に貢献する世界屈指のイノベーティブな大学へ -, <https://www.osaka-u.ac.jp/ja/oumode/OUvision2021/>, 2021.
- [4] 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, GakuNin RDM, <https://rcos.nii.ac.jp/service/rdm/>, 2021.
- [5] hpc/ior, <http://github.com/hpc/ior/>, 2021.
- [6] データ活用社会創成プラットフォーム共同研究基盤 mdx, <https://mdx.jp>, 2021.