

大阪大学のデータ集約基盤ONIONの誕生と成長経緯

大阪大学D3センター 先進高性能計算基盤システム研究部門

伊達 進



**Osaka university Next-generation Infrastructure
for Open research and open innovatioN**

<https://www.hpc.cmc.osaka-u.ac.jp/onion/>

Who am I?

伊達 進

- 博士(工学)
- 大阪大学 D3センター 先進高性能計算基盤システム研究部門 教授
- 大阪大学においてスーパーコンピュータ等の計算基盤・データ基盤のシステムプランニング、調達、運用、および、研究開発

E-mail: date@cmc.osaka-u.ac.jp

<http://ds-server.ais.cmc.osaka-u.ac.jp>

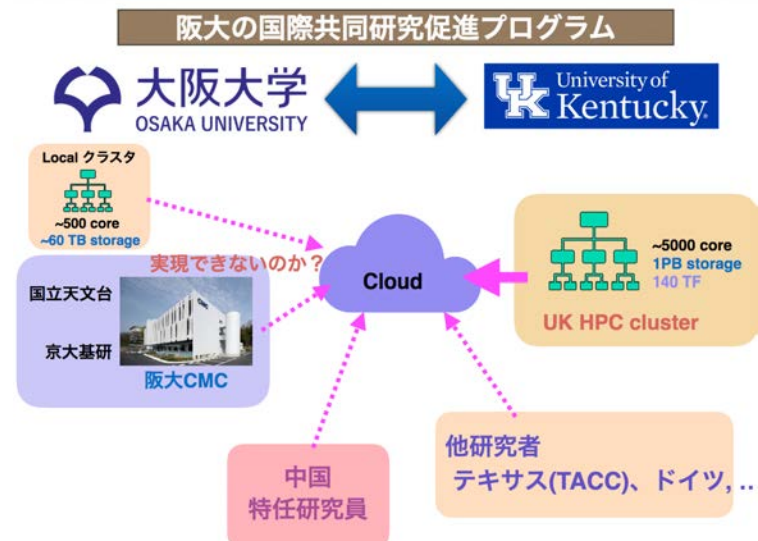


サイバーインフラ高度化への課題(1)

- ・ **学術研究の広域化・グローバル化、産学共創への期待**
 - ・ 共通の課題解決にむけて世界の研究者と協働する国際共同研究



地理的に分散する計算資源、データ資源を駆使して、
研究開発できる環境

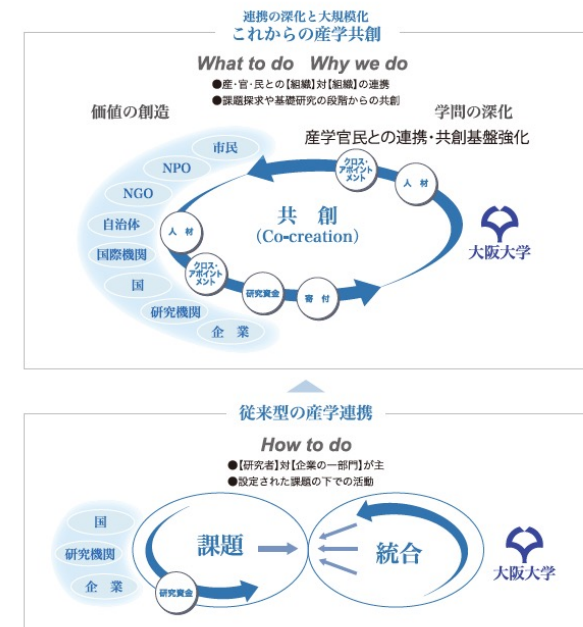


長峯健太郎, “理学研究とクラウド利用のニーズ:理論宇宙物理学の例”,
Cyber HPC Symposium 2019, Mar. 2019. 発表スライドより。

京都大学研究データ管理シンポジウム Feb. 27 2026



「産学連携から産学共創へ」をコンセプトに掲げ、
社会のニーズに基づく基礎研究の課題を発掘し、
新たな社会的価値の創出につながる
「Open Innovation」に挑戦します。

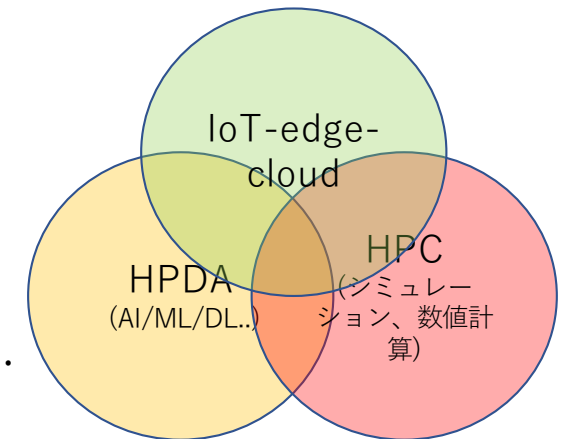


OUビジョン2021 - 社会変革に貢献する
世界屈指のイノベティブな大学へ -

サイバーインフラ高度化への課題(2)

• HPDA分野の計算ニーズの急速な高まり

- AI/MLを活用するデータ駆動型サイエンス
- 各種IoTセンサを利活用したビッグデータ解析
- HPC for AI, AI for HPC, AI for Science
 - データ同化シミュレーション, AIによる計算結果の判定, etc...



多様なデータをスーパーコンピューティング環境に“集約”できるデータ基盤
= 学術研究のデータフローを妨げないデータ基盤
= 利用者のデスクトップ環境、クラウド、スパコン環境を問わず
データ（計測データ、計算結果、等）を簡易に移動・管理できる環境

• 研究データ保存, Reproducibilityに向けた動きの活発化

- 公正な研究活動の推進にむけて、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策定
- 様々な科学データ計測機器がキャンパス内に存在しており、そのデータ管理・移動が課題
- 大容量・大規模データを取り扱うSCセンターとしての役割・責任



近い将来、スーパーコンピュータを提供する計算機センター（プラットフォーム）として、デジタル化された研究データを適正に管理するための計算・データ基盤を提供する必要がある

再現性の危機



スタンフォード 監獄実験は イカサマだった!

科学の信頼性を根底から揺るがす
「再現性の危機」に迫る真実の書!

- ・「ファスト&スロー」にもある“再現性の危機”
- ・論文不正件数の世界1位は、ある日本人の183件!
- ・権威ある心理学研究100件のうち、再現に成功したのはたった39%!?

ダイヤモンド社

・再現できない科学（学術研究）

- ・不正
- ・過失
- ・バイアス
- ・誇張



学術研究の主な出資者である国民からの不審・不満・疑問



大学設置の計算機センターとして、デジタル空間で行われる
解析・計算が正しく行われたことを提示・保証する方法はないか？

「Science Fictions あなたが知らない科学の真実」

Stuart Richie (訳: 矢羽野 薫), ダイヤモンド社

京都大学研究データ管理シンポジウム Feb. 27 2026

SQUID+ONION 5つのチャレンジ

データ集約基盤ONION

- Osaka university Next-generation Infrastructure for Open research and open Innovation

特徴5. テイラーメイド型計算

- 世界最高水準の基礎的、基盤的研究や学際融合研究が生み出す多様な知の創出と深化に寄与すべく、総合大学である大阪大学内で創出された「利用可能な超大量データを将来に渡る持続可能性を保持しつつ責任をもって活用」可能にするとともに、新たな社会的価値の創出を目指した「産学共創」「国際共同研究」のための学内外でのデータ利活用を支援するデータ集約基盤

- スーパーコンピュータSQUIDの調達に合わせてPoC (Proof-of Concept)実装として試験的に導入

- (1) 本格的に大規模なONIONを整備してしまうと、調達主目的のスーパーコンピュータの規模を大幅に縮減してしまう。
- (2) 構想したONIONが有用でなく、利用されない場合もありえる

特徴2. クラウド連動・連携機能

高負荷状態時にオフロード
新製品の早期実戦配備

民間クラウド(IaaS)事業者

ODINS網を活用した
学内部局からのデータ集約

クラウドストレージを活用した
海外研究拠点とのデータ共有

特徴4. データ集約環境

民間クラウド(ストレージ)事業者

データ集約基盤 ONION

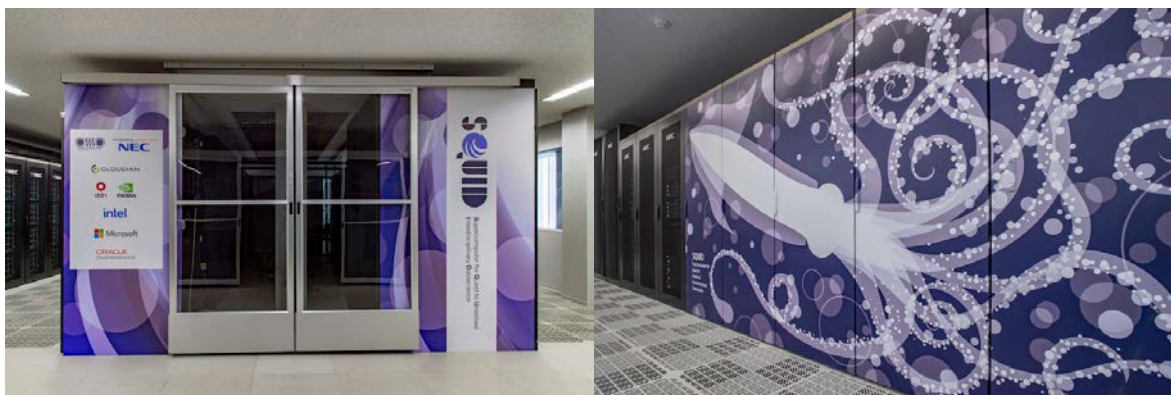
大容量高速ストレージ
20+1 PB (HDD+SSD)

SQUID since May 2021



Supercomputer for Quest to Unsolved
Interdisciplinary Datascience

- クラウド連動型高性能計算・高性能データ分析用スーパーコンピュータ
(Supercomputer for Quest to Unsolved Interdisciplinary Datascience)
 - 総理論演算性能 16.591 PFlops



最新 3 種混合プロセッサ・アクセラレータ搭載！

SQUIDは、2021 年 4 月に発表されたばかりの Ice Lake 世代のプロセッサをはじめ、最新 GPU アクセラレータ NVIDIA HGX A100 8 GPU ボード、最新ベクトルプロセッサ NEC SX-Aurora TSUBASA Type 20A を搭載しています。汎用 CPU ノード群は、Ice Lake 世代のプロセッサを搭載する国内最大級となります。

京都大学研究データ管理シンポジウム Feb. 27 2026

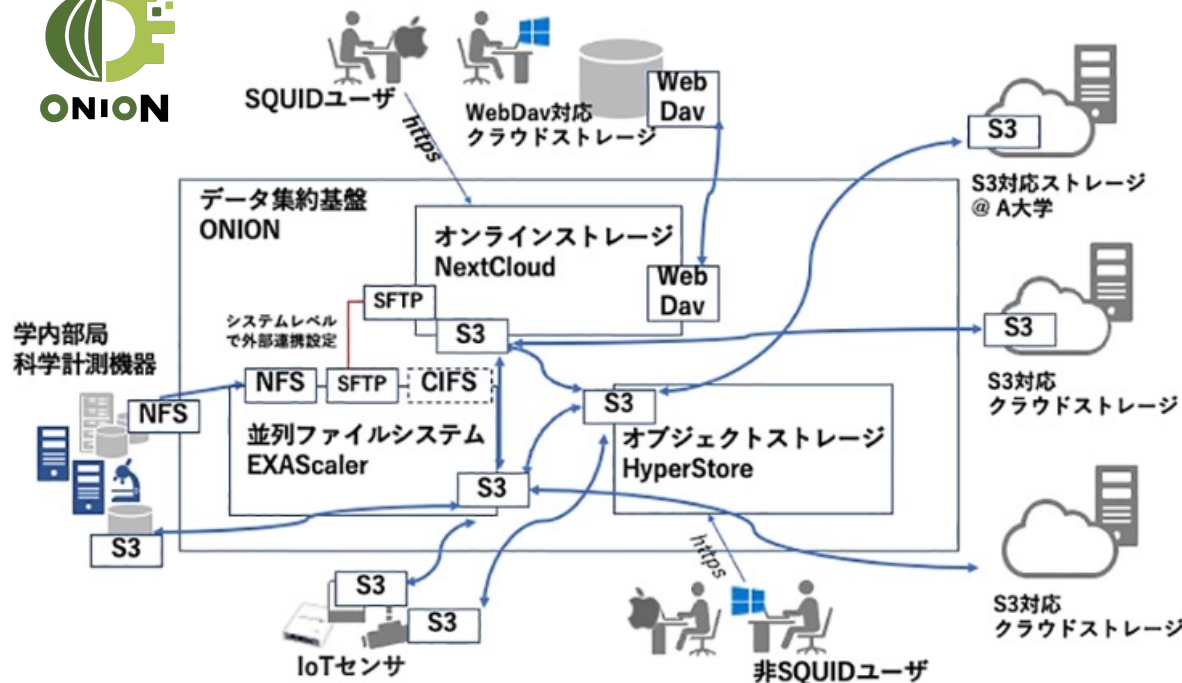


SQUID システム構成

汎用 CPU ノード群		
1,520 ノード x 理論演算性能 5,837 TFLOPS		8.871 PFLOPS
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38コア) 2 基	
主記憶容量	256 GB	
GPU ノード群		
42 ノード x 理論演算性能 161.836 TFLOPS		6.797 PFLOPS
プロセッサ	Intel Xeon Platinum 8368 (Ice Lake / 2.40 GHz 38 コア) 2 基	
主記憶容量	512 GB	
GPU	NVIDIA HGX A100 8 GPU ボード (Delta)	
ベクトルノード群		
36 ノード x 理論演算性能 25.611 TFLOPS		0.922 PFLOPS
プロセッサ	AMD EPYC 7402P (2.8 GHz 24コア) 1 基	
主記憶容量	128 GB	
Vector Engine	NEC SX-Aurora TSUBASA Type 20A 8 基	
ノード間接続		
ノード間接続	Mellanox InfiniBand HDR (200 Gbps)	
データ集約基盤 ONION		
S3 対応並列ファイルシステム		21.2 PB
ファイルシステム	DON EXAScaler (Lustre)	
HDD	20.0 PB	
SSD	1.2 PB	
S3 対応オブジェクトストレージ		500 TB
オブジェクトストレージ	CLOUDIAN HyperStore	
HDD	500 TB	

データ集約基盤ONION

(Osaka university Next-generation Infrastructure for Open research and open innovation)



- 総合大学である大阪大学内で創出された「利用可能な超大量データを将来に渡る持続可能性を保持しつつ責任をもって活用」可能にするとともに、新たな社会的価値の創出を目指した「産学共創」「国際共同研究」のための学内外でのデータ利活用を支援するデータ集約基盤
- 学内外で生成されるデータを計算基盤(スパコン)のストレージに“集約”する仕組み
- 異なる3種のストレージソリューションをS3プロトコルを中核として連携・連動

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(JACN),
2022. [[DOI: 10.24669/jacn.26.1_87](https://doi.org/10.24669/jacn.26.1_87)]

ONION内部構成

ONION内部構成

SQUIDユーザ

WebDav対応 クラウドストレージ

Web Dav

S3

要件(5) SDIの提供

要件(4) 外部ストレージ連携

データ集約基盤 ONION

オンラインストレージ NextCloud

Web Dav

要件(7) 将来計画 (GakuNinRDM)との接続性

要件(2) 非スパコン ユーザのデータアクセシビリティ

NFS

CIFS

並列ファイルシステム

要件(6) 高いI/O性能

S3

要件(1) 科学計測機器、IoTセンサからのデータ収容

要件(3) マルチテナント性

S3

要件(2) 非スパコン ユーザのデータアクセシビリティ

S3

IoTセンサ

学内部局 科学計測機器

京都大学研究データ管理シンポジウム Feb. 27

非SQUIDユーザ

S3対応クラウドストレージ @ A大学

S3対応クラウドストレージ

S3対応クラウドストレージ

11

ONIONの成長分岐点(1)



- 令和2年度先端研究設備整備費補助事業（第3次補正予算）「ONION駆動型科学計測データ自動収容・遠隔利用システム」
 - サイバーメディアセンターと科学機器リノベーションセンター・工作支援センターで提案・採択

文部科学省 MINISTRY OF EDUCATION, CULTURE, SPORTS, SCIENCE AND TECHNOLOGY JAPAN

> サイトマップ > English 文字サイズの変更 中 大 検索

会見・報道・お知らせ | 政策・審議会 | 白書・統計・出版物 | 申請・手続き | 文部科学省の紹介

トップ> 会見・報道・お知らせ> 公募情報> 「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）」の公募について

● 「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）」の公募について

令和2年12月17日

「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）」について、公募を開始いたします。

1. 公募期間

令和2年12月17日（木曜日）～令和3年1月18日（月曜日）12時00分

2. 事業の趣旨・目的

今般の新型コロナウイルス感染症の拡大の影響により、移動の制限や大学等への入構が制限され、研究活動に大きな影響を及ぼしました。研究活動の遅れを取り戻すとともに、ポストコロナを見据えた新たな研究環境を早期に構築するためには、研究の遠隔化や実験の自動化を推進する必要があります。

このため、研究機関等における基盤的及び先端的研究施設・設備・機器のリモート化・スマート化を図るために、幅広い研究者への共用体制を構築している研究機関において、研究者からのニーズの高い、遠隔地からの研究や研究の自動化が可能な共用研究設備・機器の新規導入や、各研究機関等が保有している共用研究施設・設備・機器に対して遠隔利用や実験の自動化を推進するための設備・機器の追加（既存の共用研究施設・設備・機器の高度化等）を支援します。

これらにより、時間や距離に縛られず研究を遂行できる新たな研究環境を各研究機関に整備し、魅力的な研究環境の実現や、研究現場の生産性向上、研究における飛躍的イノベーションの実現等を加速します。

(様式2別添1②)					
設備・機器の追加（詳細）					
				優先順位	1
部局名	サイバーメディアセンター				
遠隔利用や自動化を図る既存の共用研究施設・設備・機器	・工学研究科分析センター質量分析計 3台（分析室自動データ集約・配信システムを介して） ・共用電子顕微鏡 3台（大容量測定データの共有） ・学外利用対応全学共用機器 10台程度（大容量測定データの学外者へ提供）				
本事業で追加する設備・機器名 ^{※1}	ONION駆動型科学計測データ自動収容・遠隔利用システム	遠隔利用／自動化	遠隔利用&自動化	金額 ^{※2}	60,000（千円）

学内のデータ移動・集約需要を喚起

全学 機器共用

"部局"の研究リソースを "全学"でも共に活用する

大阪大の場合は、全部で300機器程度

約300機器・年間約8万件の利用

**これらの約300の共用機器が生み出すもの
→ 測定（研究）データ**



全体の取り纏め役（旗振り・調整役）がコアファシリティ

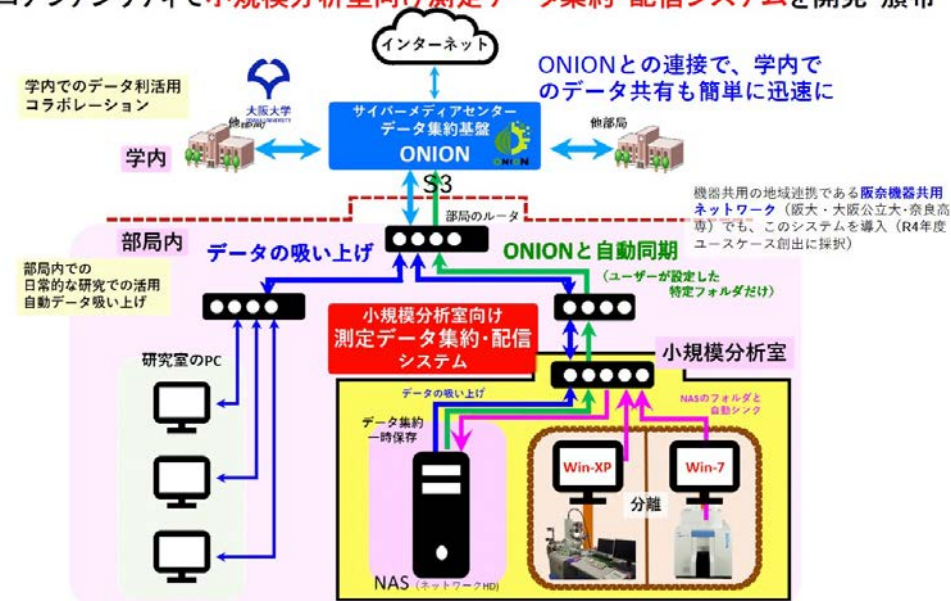


Courtesy of 古谷浩志 (大阪大学 コアファシリティ機構), "大阪大学コアファシリティ機構における測定データ流通・集約基盤の構築と阪奈機器共用ネットワークへの展開", 第1回北陸地区学術データ基盤セミナー ~コアファシリティ連携から研究データエコシステム構築を目指して~, 金沢, 2024年2月.

京都大学研究データ管理シンポジウム Feb. 27 2026

ネットワーク経由での(部局内での)測定データの流通を実現

コアファシリティで小規模分析室向け測定データ集約・配信システムを開発・頒布



ONIONの成長分岐点(2)

- 広域からのデータ収集機能と、データ集積・処理機能を企業や自治体との共同研究も含めた全国の大学・公的研究機関が関与する様々なデータ活用の取り組みに提供し、さらにはデータ活用のコミュニティを形成して分野・セクタを横断した連携を触媒するハブとなることを目指す。
- 9大学・2 研究機関で共同調達・運用. (学際大規模情報基盤共同利用・共同研究拠点 (JHPCN) 8大学で運用経費を共同負担)
- 全国の産学の研究者が利用するクラウドインフラ



CPU nodes
Installed in 2024.03



GPU nodes
Installed in 2025.03

京都大学研究データ管理シンポジウム Feb. 27 2026



2026年3月までにmdx II はさらに拡張される予定

mdx II (as of 2025.04)

General CPU computing nodes

60 nodes x Theoretical performance/node 7.168TFLOPS 430.08TFLOPS

processor	Intel Xeon Platinum 8480+ x 2 (Sapphire Rapids/ 56C, 2.0 GHz)
-----------	--

Main memory	512GB (32GB DDR5-4800 ECC RDIMM x 16)
-------------	---------------------------------------

GPU computing nodes

7 nodes x Theoretical performance/node 140.3 Tflops 982.1 TFLOPS

processor	Intel Xeon Gold 6530 (Emerald rapid / 38C, 2.1GHz,)
-----------	--

Main memory	1024GB
-------------	--------

GPU	NVIDIA H200 x 4
-----	-----------------

Interconnect

Inter-node connection	200GbE
-----------------------	--------

Storage

File system	DDN EXAScaler (Lustre)
-------------	------------------------

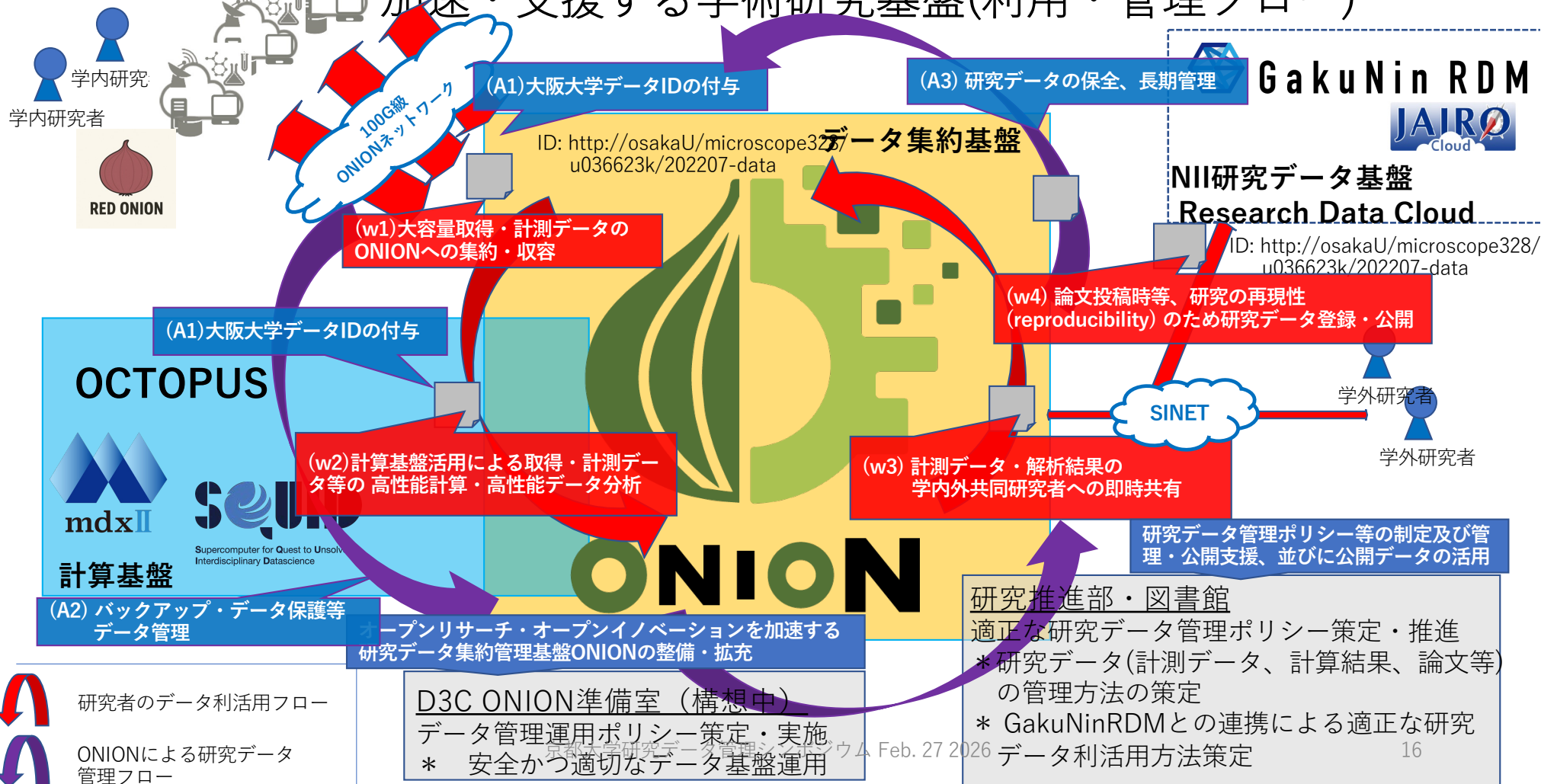
Actual size	1106.48 TB
-------------	------------

大阪大学が目指すOUエコサイクルとD3Cが 全国の大学・研究機関に提供する学術研究基盤の関係



データ源：キャンパス内
科学計測機器・センサー

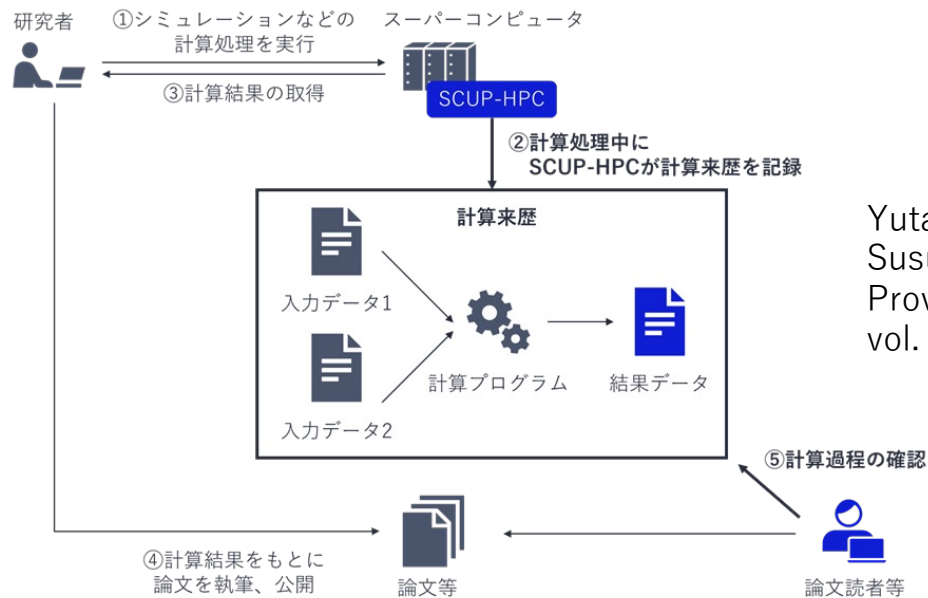
D3Cが目指す公正な研究活動を 加速・支援する学術研究基盤(利用・管理フロー)



さらなるオープンサイエンスの加速のために

• 計算来歴を統合した、オープンサイエンス時代の新しい科学計算の形 **“SCUP (Scientific Computing Unifying Provenance) -HPC”** の提案

- どのようなデータがどのようなプログラムにアクセスされ、どのようなデータを生成するかを追跡する計算来歴を、スーパーコンピュータの性能に与える影響を最小限に抑えながら、記録・管理・可視化する来歴管理システム (SCUP-HPC (System for Constructing and Utilizing Provenance on High-Performance Computing System)) によって実現

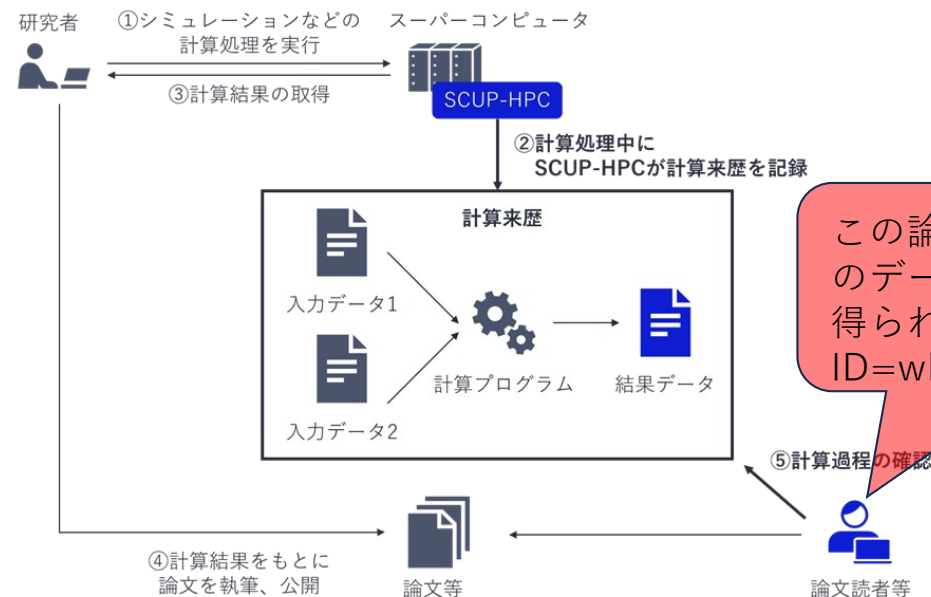


Yuta Namiki, Takeo Hosomi, Hideyuki Tanushi, Akihiro Yamashita, Susumu Date, “SCUP-HPC: System for Constructing and Utilizing Provenance on High-Performance Computing Systems”, IEEE Access, vol. 13, pp. 141090-141107, 2025. [DOI:10.1109/ACCESS.2025.3597361]

学術研究の適正性・公正性担保に向けた Open Scienceサービスの試験運用開始 2025年12月～



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 : Osaka university
Compute and sTorage Platform
Urging open Science)



この論文に記載のあるグラフ
のデータは本当に計算で
得られたのかな。
ID=whoami-123456789

本研究成果（の一部）は、大阪大学D3センターの
OCTOPUSを利用して得られたものです。（計算来歴
ID=whoami-123456789）

まとめ



- 本講演では、大阪大学のデータ”集約”基盤ONIONの誕生と成長経緯について報告した。
 - データ集約基盤ONIONは、計算基盤にデータを集約することを目的とし、スパコン調達での試験的な導入に端を発し、試験運用をしている。
(オープンサイエンスや研究データ管理を主目的として導入したわけではなかった)
- 令和2年度先端研究設備整備費補助事業（第3次補正予算）「ONION駆動型科学計測データ自動収容・遠隔利用システム」はONIONの全学利用への成長分岐点
- IaaS型計算基盤mdx II の大阪大学への設置は、ONIONのデータ利活用を促進する成長分岐点
- 附属図書館、オープンサイエンス推進室との連携は、ONIONのオープンサイエンス基盤にする成長分岐点
- ONION成長は大阪大学の研究支援の志ある現場研究者がボトムアップアプローチで積み上げてきたことにあり、それゆえに様々な利用事例が生まれつつある。