

公募シンポジウム

公募シンポジウム7

歯科医療情報共有化と異分野融合によるデータ駆動型時代の歯科医療

2023年11月24日(金) 13:30 ~ 16:00 C会場 (EX1-B)

[3-C-3-03] 歯学医療データ meets HPC &AI 基盤 ~大阪大学の事例~

*伊達 進¹ (1. 大阪大学サイバーメディアセンター)

*Susumu Date¹ (1. Cybermeida Center, Osaka University)

キーワード : Open Science, Reproducibility, S2DH, SQUID, ONION

近年、“オープンサイエンス”に向けた取り組みが活発化している。今日の学術研究では、デジタル化された研究データはストレージに蓄積され、計算機で解析・処理される。さらに、解析・処理されたデータは蓄積・公開され、他の研究者の研究データとして利活用される。このような研究データの利活用を促進させることで、研究者間の知識格差の解消、研究プロセスの再現性担保を通じて、学術研究のさらなる発展が期待される。大阪大学サイバーメディアセンターでは、スーパーコンピューティングシステム SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datascience)の導入に伴い、利用者である研究者が大容量な研究データに対して、大規模な計算資源により高性能計算(High Performance Computing)あるいは高性能データ分析処理(High Performance Data Analysis)を迅速に行えるよう、学内外の研究データを容易に収容可能とするデータ集約基盤 ONIONを試験運用している。本講演では、データ集約基盤 ONION及びスーパーコンピュータ SQUIDを中核とするサイバーメディアセンターのサイバーインフラストラクチャの現状を紹介するとともに、そのサイバーインフラストラクチャを利活用し推進中のソーシャル・スマートデンタルホスピタル構想についても言及する。その後、講演者らが構想するオープンサイエンス実現に向けた基盤整備計画についても紹介したい。

歯学医療データ meets HPC & AI 基盤

- 大阪大学の事例 -

伊達 進

大阪大学サイバーメディアセンター

Dental Clinical Data meets HPC & AI Infrastructure

- A Case in Osaka University -

Susumu Date

Cybermedia Center, Osaka University, Japan

Recently, Open Science has gathered much attention and concern from scientists and researchers and thus academic activities has been accelerating. In today academic research, a variety of digitalized research data are accumulated in storages, and analyzed or processed on computer systems. Also, such analyzed or processed data are utilized as research data by other researchers. Academic research expects more advancement through alleviation of knowledge among researchers and reproducibility assurance of research process brought about by the further utilization of research data. In Cybermedia Center, Osaka University, which is a supercomputing center and has a responsibility to deliver high performance computing services to researchers and scientists in Japan, we have started the operation of ONION (Osaka university Next-generation Infrastructure for Open data and open innovatioN) as a data aggregation infrastructure in conjunction with the introduction of our new supercomputing system named SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datascience). ONION facilitates researchers and scientists to aggregate data onto it and then to perform HPC (High Performance Computing) and HPDA (High Performance Data Analysis) on data. In this talk, the speaker briefly introduces our Cyberinfrastructure of which ONION and SQUID plays a central role. Also, Social Smart Dental Hospital (S2DH) which we envision through the collaboration of Osaka University Dental Hospital is explained as well as our infrastructure planning towards Open Science.

Keywords: Open Science, Reproducibility, S2DH, SQUID, ONION

1. はじめに

著者の所属する大阪大学サイバーメディアセンターは、大阪大学だけでなく、全国の研究者に高性能計算機環境を提供するミッションを有する全国共同利用施設である。本センターでは、そのようなミッションに基づき、2021年5月より16.591 PFLOPSの理論演算性能を有するスーパーコンピューティングシステム SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datascience) [1] を導入している。当該スーパーコンピューティングシステムは、汎用 CPU ノード群(Intel 製プロセッサ(Ice Lake 世代)), GPU ノード群(NVIDIA 製 GPU A100)、ベクトルノード群(NEC 製ベクトルプロセッサ NEC SX-Aurora TSUBASA)の異なる3種のアーキテクチャから構成され、利用者である研究者の多種多様な HPC(High Performance Computing)および HPDA(High Performance Data Analysis) 計算ニーズを収容可能としている。さらに、DL (Deep Learning)、

ML (Machine Learning)、AI (Artificial Intelligence)等のキーワードに代表される HPDA(High Performance Data Analysis)用途でのスーパーコンピューティングシステムの利用促進の視点から、本センターでは、SQUID 導入に合わせ、学内外のデータを容易に SQUID に収容可能とするデータ集約基盤 ONION (Osaka university Next-generation Infrastructure for Open research and Open innovatioN) [2] を試験運用・導入している。

本講演では、これら ONION および SQUID を中核とする、サイバーメディアセンターのサイバーインフラストラクチャについて、そのねらい、仕組みについて紹介する。その後、本センターが大阪大学歯学部附属病院および日本電気株式会社との共同研究を通じて目指す Social Smart Dental Hospital (S2DH) [3] について概説する。また、本センターが今後計画する OpenScience に向けたサイバーインフラストラクチャ整備計画についても概説する。

表 1: SQUID の概要.

ノード構成	汎用 CPU ノード 1520 ノード(8.871 PFlops)	プロセッサ: Intel Xeon Platinum 83686 (Ice Lake / 2.4 GHz 38 コア) 2 基 主記憶容量: 256 GB
	GPU ノード 42 ノード(6.797 PFlops)	プロセッサ: Intel Xeon Platinum 8368 (Ice Lake / 2.4 GHz 38 コア) 2 基 GPU: NVIDIA HGX A100 8GPU ボード 主記憶容量: 512 GB
	ベクトルノード 36 ノード(0.922 PFlops)	プロセッサ: AMD EPYC 7402P (ROME / 2.8 GHz 24 コア) 1 基 主記憶容量: 128 GB ベクトルプロセッサ: NEC SX-Aurora TSUBASA Type20A 8 基
相互結合網	Mellanox InfiniBand HDR (200Gbps)	
ストレージ	DDN EXAScaler (Lustre/ HDD 20.0 PB, NVMe SSD 1.2 PB)	

2. SQUID & ONION

表1に SQUID の概要を示す。SQUID は Intel 製 Xeon Platinum 8368 (Ice Lake 世代)のプロセッサを2基、256GBの主記憶を搭載した汎用 CPU ノードが1520ノード、汎用計算ノードと同様のプロセッサ2基と512GBの主記憶、NVIDIA製 HGX 8GPU ボードを搭載した GPU ノードが42ノード、AMD製 EPYC 7402P (ROME 世代)のプロセッサを1基、128GBの主記憶、NEC製 SX-Aurora TSUBASA Type20Aを8基搭載したベクトルノードを36ノードから構成されるハイブリッド型スーパーコンピューティングシステムである。これらのノードは Mellanox 製 InfiniBand HDR で200Gbpsで相互に接続されており、広帯域低遅延でのノード間通信をサポートしている。ストレージとしては、DDN製 EXA Scalerを採用し、HDD及びSSD領域をシングルディスクイメージとして、Lustreファイルシステムで提供し、利用者は高速にread/write操作が可能となっている。

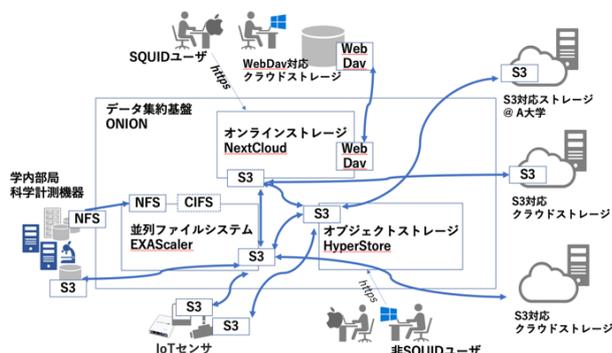


図1: ONIONの内部構造。

上述したように、本SQUIDの導入に伴い、SQUIDではデータ集約基盤ONIONを試験的に導入している。図1にONIONの概要を示す。基本的には、ONIONは、上述した並列ファイルシステムLustreの機能を提供するEXAScaler、オブジェクトストレージクラウド社製HyperStore、および、オンラインストレージNextCloudから構成される。これら3種の異なるストレージソリューションをクラウドストレージでのデファクトスタンダードプロトコルとなっているS3プロトコルを中核として統合する。これにより、学内外部局の所有する科学計測機器、学内に設置されたIoTセンサー、あるいは学外に設置されたS3対応ストレージからのデータ集約を可能としている。また、オンラインストレージNextCloudは、HyperStoreおよびEXAScalerの詳細を隠蔽し、シングルディスクイメージを直感的なGUIで提供する。

本稿執筆時点においては、令和2年度第三次補正予算「先端研究設備補助事業(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大阪大学のプロジェクト、後述する大阪大学歯学部附属病院の推進するSocial Smart Dental Hospital、「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プログラム)」に採択されたパワーレーザーDXプラットフォーム事業[4]等での試験的な利用が開始されている。また、国立情報学研究所がOpen Scienceの実現に向け運用する研究データ管理基盤GakuNinRDM[5]との試験的な接続もすでに開始されている。

3. Social Smart Dental Hospital (S2DH)

2018年より大阪大学歯学部附属病院、日本電気株式会社、および大阪大学サイバーメディアセンターは、歯学医療データ

の知識や情報をできる限りデータ化し、その膨大なデータを高性能なスーパーコンピュータを前提としたAIを駆使し、歯学研究及び臨床に役立てることを目的とした共同研究を推進してきた。矯正歯科[6]、歯周病診断[7]等への応用等々、さまざまな共同研究を通じたAI開発とともに、そうした医療データを安全にスーパーコンピューティングシステムで利用するための基盤研究開発[8, 9]を行ってきた。その一例であるセキュアステージングは、歯学部附属病院に設置されたデータに対して、スーパーコンピューティングでの計算処理が行われる際に安全にアクセスさせる仕組みを提供する。本講演ではこのようなS2DHの活動および成果について報告する。

4. おわりに

Open Scienceは、今日研究者だけでなく、一般市民にも浸透するキーワードとなっている。AI時代においては、より多くのデータを解析処理することが極めて重要であり、それゆえに、スーパーコンピューティングシステムの利活用はますます重要になっている。逆に言えば、大量のデータを効率的に処理できる計算基盤・データ基盤は我が国の学術研究に必要な不可欠となっている。また、デジタル化された研究データは、研究プロセスの再現性を担保する形で管理・公開される必要がある。本講演では、S2DHなどの大阪大学のケーススタディから、大阪大学のOpen Scienceを視野に入れた活動を紹介する。

参考文献

- [1] Susumu Date, Yoshiyuki Kido, Yuki Katsura, 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] 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]
- [3] ソーシャルスマートデンタルホスピタル (S2DH), <https://s2dh.org/> (cited 2023-Aug.20).
- [4] パワーレーザーDX プラットフォーム, <https://powerlaser.jp> (cited 2023-Aug.20).
- [5] GakuNin RDM, <https://rdm.nii.ac.jp/> (cited 2023-Aug.20).
- [6] Seiya Murata, Chonho Lee, Chihiro Tanikawa, Susumu Date, “Towards a Fully Automated Diagnostic System for Orthodontic Treatment in Dentistry”, The Thirteenth IEEE eScience Conference (e-science2017), pp.1-8, Auckland, New Zealand, Oct. 2017. [DOI 10.1109/eScience.2017.12]
- [7] Yusuke Moriyama, Chonho Lee, Susumu Date, Yoichiro Kashiwagi, Yuki Narukawa, Kazunori Nozaki, Shinya Murakami, “Evaluation of Dental Image Augmentation for the Severity Assessment of Periodontal Disease”, Proceedings of 6th International Conference on Computational Science and Computational Intelligence, Dec. 2019 [DOI: 10.1109/CSCI49370.2019.00176]
- [8] 伊達進, 吉川隆士, 野崎一徳, 渡場康弘, Lee Chonho, 木戸善之, 下條真司, “医療データを高性能計算機システムで利用するためのダイナミックセキュアなステージングシステム”, 第37回日本医療情報学連合大会, 大阪, Nov. 2017.
- [9] Takashi Yoshikawa, Susumu Date, Yasuhiro Watashiba, Yuki Matsui, Kazunori Nozaki, Shinya Murakami, Chonho Lee, Masami Hida, Shinji Shimojo, “Secure Staging System for Highly Confidential Data Built on Reconfigurable Computing Platform”, 2019 IEEE International Conference on Computational Science and Engineering (CSE), New York, USA, Aug. 2019. [DOI:10.1109/CSE/EUC.2019.00066]