

利用系グループ活動報告

JASRI 研究DX推進室/XFEL利用研究推進室 先端計測・解析技術グループ
理研・放射光科学研究センター・制御情報・データ創出基盤グループ

公益財団法人高輝度光科学研究センター
研究DX推進室 / XFEL利用研究推進室 先端計測・解析技術グループ

国立研究開発法人理化学研究所
放射光科学研究センター 制御情報・データ創出基盤グループ 城地保昌

国立研究開発法人理化学研究所
放射光科学研究センター 制御情報・データ創出基盤グループ

公益財団法人高輝度光科学研究センター
XFEL利用研究推進室 先端計測・解析技術グループ 初井宇記

1. はじめに

高輝度光科学研究センター研究DX推進室/XFEL利用研究推進室 先端計測・解析技術グループと理化学研究所放射光科学研究センター（理研RSC）制御情報・データ創出基盤グループは、共同で先端計測・解析技術に関する開発、高度化および運用をおこなっている。これらは、X線検出器、ビームライン制御システム、データ解析基盤に大別できる。本稿では、これらの活動のうちSPring-8のデータ基盤の現状について紹介する。

2. SPring-8データセンター

SPring-8データセンター構想

SPring-8では、日々国内外の研究者がさまざまな実験を行っており、貴重な実験データが大量に生み出されている。しかしながら、特に1試料あたりのデータが1TBを超える大容量データについては、データの転送や解析に時間がかかり、成果創出までの時間がかかっていた（解析律速）。加えてSPring-8では、理研RSCが開発した高感度で高速撮像が可能なX線画像検出器CITIUSの導入が進められている。それにより、今後生み出される実験データの量は以前の1,000倍以上になると想定される。この「解析律速」を解決することを主眼として2021年に「SPring-8データセンター構想」を提唱した。欧州・米国の大型放射光施設ではデータ基盤に大規

模な投資を行っている。SPring-8はCITIUS検出器の開発に成功したこともあり、欧米施設よりも10倍から100倍の大量のデータが得られることが確実である。したがって、単純な大規模化では対応できず、新しいアプローチが必須である。この理解に基づき「SPring-8データセンター構想」では、ビームライン近傍（Edge）でデータを圧縮すること、およびSPring-8外の「富岳」や国立情報学研究所などのデータ関連基盤とネットワークを介して接続した仮想的なデータ基盤を開発・整備することを特徴としている。ハードウェアの概略を図1に示す。

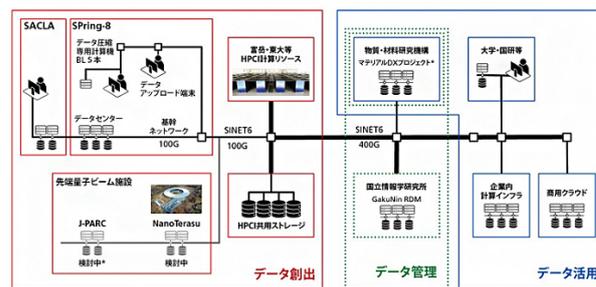


図1：SPring-8データセンター構想の概略図

これまで構内基幹ネットワークの通信速度を100Gbps（ギガbps）へアップグレードした。データは確実に保存されるだけでは不十分で、その品質を実験中に確認する必要がある（データ品質評価, Data Quality Assessment）。データに問題があれば、直ちに実験条件を変更しなければならないか

らである。多くの大容量データ実験では並列計算を多用した解析がデータ品質評価のために必要になってきている。そこでクラスタ計算機システムをSPring-8内のデータセンターに整備した。さらに大規模な計算が必要な実験のために、「富岳」などのSPring-8外のスーパーコンピュータも利用できるようHPC（High Performance Computing）向けデータ管理・転送・解析環境Globusの導入検討を米国アルゴンヌ研究所および理研・計算科学研究センター（R-CCS）とともに実施している。

日本では、文部科学省がデータ関連の研究開発の役割を整理しており、その中で国立情報学研究所はデータ管理サービスGakuNin RDM^[1]を提供している。SPring-8データセンターはこのサービスと連携して利用できるように設計されており、2025年4月から試験運用を行っている^[2]。このサービスが本格運用すれば将来的に、研究室のデータ、J-PARCおよびNanoTerasuのデータなどが一体的に管理できるほか、データが改ざんされていないことを証明することが可能となる。

CITIUS検出器のうち最大のシステムは、2,020万画素、データ帯域10Tbps（テラbps = 1,000Gbps）、データ量が1年で6エクサバイト（= 6,000,000テラバイト）と見込まれる。ネットワーク転送も保存も極めて困難な量である。10Tbpsのデータ処理を実現するために、理研計算科学研究センター（R-CCS）との共同研究により、データ圧縮技術の開発に取り組んでいる。

クラスタ計算機システム

理研RSCは、「SPring-8データセンター構想」に基づき、電算機室（図2）を新たに整備し、クラスタ計算機システムを2023年10月から共用している。クラスタ計算機システムは、CPUノード64台、GPUノード16台、データ転送ノード6台、S3ゲートウェイ（後述）4台、共有ストレージシステムなどで構成される^[2]。クラスタ計算機の各ノードは高速ストレージシステムに広帯域で接続しており、高いI/O性能が要求されるSPring-8の多くの解析に最適化された構成となっている。



図2：SPring-8データセンターの電算機室

SPring-8データセンター上のサービス

2025年4月現在、下記に紹介する6つのサービスがユーザーに提供されている。

(1) SPring-8 Data Flow Service (DFS)

利用実験で生成される少量データの流通を目的として、SPring-8 Data Flow Service (DFS)^[3]を開発し、2023年10月にサービスを開始した。1課題あたり1TB以下、1ファイルあたり2GBのデータが対象である。ビームライン担当者が申請することにより利用が可能となる。DFSが利用可能なビームラインの場合、ユーザーは、ビームライン担当者に申請するとDFS利用の権限が付与される。ユーザーは、実験代表者もしくは共同研究者である課題の実験データを、Webブラウザを介してアップロード/ダウンロードすることができるようになる。

(2) クラスタ計算機サービス

CITIUS検出器などにより取得される大容量実験データを迅速に解析するためのクラスタ計算機サービスを2023年11月に開始した。データセンターへの高速転送が可能なビームラインネットワーク^[4]が導入されたビームラインのユーザーが利用可能である。

(3) Open OnDemandサービス

クラスタ計算機システムで計算を実行するにはジョブ管理システムなどの知見が必要である。これらに不慣れなユーザーが簡便に利用できるようにSPring-8データセンターではOpen OnDemand^[5]を導入している。ユーザーはWebブラウザから

ファイルアクセス、Jupyter^[6]、ImageJ^[7]など様々なツールを、施設側で最適化した環境で利用できる。Open OnDemandでは特定解析プログラムを実行するために最適化された環境を開発することも可能で、開発環境を提供している。

(4) ビームラインからデータセンターへのデータ転送サービス

CITIUS 検出器などにより取得される大容量実験データをビームラインからデータセンターへ転送するツールを開発し提供している。ユーザーは、WebGUIから転送元ディレクトリと実験課題番号を設定することで取得データを、自動でデータセンターに転送することができる。このサービスでは、データが確実に転送されたことをチェックサムにより確認している。

(5) 自動解析サービス

データ転送サービスによりデータセンターに転送されたデータを自動で解析する基盤技術を整備した。このサービスでは、ユーザーアカウントではなく専用アカウントにより解析が実行され、ユーザーは、実験代表者もしくは共同研究者である課題の解析結果を、Webブラウザを介してダウンロードすることができる。

(6) 所外への広帯域データ転送サービス

データセンターの計算機システムは、Amazon S3 API^[8]と互換性のあるオブジェクトストレージ機能を有している。クラスタ計算機サービスの利用ユーザーは、s3cmd^[9]などを利用して、データセンター上の大容量データをSPring-8所外から高速（2025年4月現在100Gbpsが上限）にダウンロードすることが可能である。

3. SPring-8データセンターの活用事例

高分解能粉末回折装置（BL13XU第3ハッチ）

BL13XUの第3ハッチにおける高分解能粉末回折実験では、装置の自動化が進み多数の少容量データが測定できることから、自動でDFSにアップロードできる仕組みを導入した。この自動アップロード

サービスを他のビームラインにも展開していく予定である。

高エネルギーX線CT装置（BL28B2）

BL28B2における自動CT計測装置^[10]における測定代行では、測定データをデータ転送サービスによりデータセンターに転送し、自動解析サービスを利用して3次元再構成してユーザーに提供している。ユーザーは、解析された結果を、Webブラウザを介してダウンロードすることができる。

準弾性散乱実験（BL35XU）

BL35XUにおいてCITIUS検出器を用いた準弾性散乱実験が行われている^[11]。この実験方法は東北大学齋藤真器名准教授が提案した新しい手法で、試料のナノ秒スケール揺らぎを解析できるという特徴を持つ。この実験のための検出器、データ基盤は、SACLA/SPring-8基盤開発プログラム^[12]において開発・整備された。CITIUS検出器は84万画素をもち、17.4kframes/sでデータを出力している。データ帯域は5.1PB/dayに達し、1ビームタイムあたり35PBといったデータ量になる。このデータはビームライン脇に設置されたデータ圧縮用のサーバー群によって直ちに1000分の1以下にデータが圧縮されたのち、データ転送サービスを利用してデータセンターに転送されている。さらにOpen OnDemand上に準弾性散乱実験データ解析アプリが開発されており、2025年度に共用される予定となっている。

4. おわりに

本稿では、「SPring-8データセンター構想」の概要とその進捗、提供中のデータ関連サービスについて紹介した。本稿がユーザーの皆様のご参考になれば幸いである。

5. 謝辞

本稿で紹介した活動は多くの皆様に支えられており、深く感謝申し上げます。特に理研RSCの本村幸治氏、中町将貴氏、平木俊幸氏（現在NII）、矢橋牧名氏、JASRIの河口彰吾氏、小林慎太郎氏、渡邊佳氏、上杉健太郎氏、星野真人氏、竹田裕介氏、西野

玄記氏、杉本崇氏、中嶋享氏、国立情報学研究所の山地一禎氏、込山悠介氏、下山武司氏、平原孝明氏、亀田武氏、理研情報統合本部の實本英之氏、J-PARCの大友季哉氏、稲村泰弘氏、岡崎伸生氏、理研R-CCSの松岡聡氏、庄司文由氏、佐藤賢斗氏、佐野健太郎氏、東北大学の齋藤真器名氏らには主要な貢献、重要な助言を頂きました。また実装・運用においてグローリーテクニカルソリューションズ株式会社、両備システムズ、日本技術センターの支援を頂いています。この場をお借りして感謝申し上げます。

初井 宇記 HATSUI Takaki

(国研) 理化学研究所
放射光科学研究センター
制御情報・データ創出基盤グループ
(公財) 高輝度光科学研究センター
XFEL 利用研究推進室
先端計測・解析技術グループ
〒679-5198 兵庫県佐用郡佐用町光都 1-1-1
TEL : 050-3502-3235
e-mail : hatsui@spring8.or.jp

参考文献

- [1] <https://rdm.nii.ac.jp/>
- [2] <https://dc-portal.spring8.or.jp/>
- [3] <https://dc-dfs.spring8.or.jp/>
- [4] <https://dncom.spring8.or.jp/networkcnfg/beamlinenetwork/>
- [5] <https://openondemand.org/>
- [6] <https://jupyter.org/>
- [7] <https://imagej.net/ij/>
- [8] <https://docs.aws.amazon.com/s3/>
- [9] <https://s3tools.org/s3cmd>
- [10] 上杉健太郎、星野真人: SPring-8/SACLA 利用者情報 **28** (2023) 331.
- [11] M. Saito *et al.*: *Phys. Rev. Lett.* **132** (2024) 256901.
- [12] http://xfel.riken.jp/topics/sacla_basic_development_2025.html

城地 保昌 JOTI Yasumasa

(公財) 高輝度光科学研究センター
研究DX推進室 / XFEL 利用研究推進室
先端計測・解析技術グループ
(国研) 理化学研究所
放射光科学研究センター
制御情報・データ創出基盤グループ
〒679-5198 兵庫県佐用郡佐用町光都 1-1-1
TEL : 050-3502-7383
e-mail : joti@spring8.or.jp