

2025 年 5 月 20 日

報道機関 各位

国立大学法人東北大学

AI 時代の研究 DX を加速する研究データ基盤を構築 中核となる東北大学研究データレイク「IZUMI」の運用開始

【発表のポイント】

- 東北大学の研究データ基盤の中核となる研究データレイクを整備
- 仙台市民に親しみ易さがあり、知識の創出を想起する「IZUMI」と命名
- 研究者自身による研究データ管理やオープンアクセスの推進を加速

【概要】

東北大学は、本学での研究データ基盤の中核となる東北大学研究データレイク「IZUMI」の運用を 6 月より開始します。「IZUMI」という名称は、このシステムが知識の泉となるようにという思いと、仙台市民に親しみのある泉ヶ岳を由来として、命名されました。「IZUMI」は約 5PB の実効容量を持ち、各研究者や研究室が保有する研究データを保持します。また、研究データのグループ共有や一般公開する機能を提供します。「IZUMI」の整備にあたっては、アジリティ(直ぐに利用できること)、ユーザビリティ(簡単に利用できること)、インターオペラビリティ(他のシステムと繋がること)の 3 点を重視してシステム設計を行っています。「IZUMI」を中核とする研究データ基盤の整備により、研究者が生成する価値ある研究データの学内外での利活用が加速し、東北大学の AI 技術を活用した研究 DX がさらに発展することが期待されます。

【詳細な説明】

背景

G7 仙台科学技術大臣会合及び G7 広島サミットのコミュニケによるオープンサイエンス拡大に関する共同声明などを発端として、学術研究機関ではより高いレベルでの研究データ管理やその公開の推進の動きが高まっています。具体的には、科学研究費補助金課題についてのデータマネジメントプラン(DMP)の推奨、論文・研究データの即時オープンアクセスの推進などが挙げられます。

東北大学(総長：富永悌二、以下、本学)では、2020 年 7 月に策定した「東北大学コネクテッドユニバシティ戦略」^(注 1)に基づき、最高デジタル責任者(CDO)(青木孝文 理事・副学長)の主導により全方位の DX^(注 2)を展開しています。この全方位の DX には、データ駆動型研究とオープンサイエンスの推進である研究 DX が含まれています。

この研究 DX を実現するため、本学では学内外の様々なシステムとデータ連携する研究データ基盤の開発を進めています。これまで「研究データ管理アプリ(仮称)」、「統合データカタログシステム(仮称)」といったシステムを開発してまいりました。「研究データ管理アプリ(仮称)」では、研究者データベースなどとデータ連携し、DMP の作成や、外部資金・成果論文・DMP の関連付けが行えます。また、「統合データカタログシステム(仮称)」では、論文の根拠データに加えて学務・業務データを含む学内のさまざまなデータをデータ資産ととらえ、それらのデータのカatalog化が行えます。これらのシステムにより、研究者である教員や職員が生産性の高い研究・業務に集中できます。しかし、これらのシステムを実用的に連携させるためには、中核となる研究データの置き場(研究データレイク^(注 3))の整備が課題となっていました。

東北大学サイバーサイエンスセンター(センター長：菅沼拓夫、以下、本センター)では、共同利用・共同研究拠点として長年にわたり最新鋭のスーパーコンピュータの計算機資源サービスを学内外の研究者向けに提供しており、現在は「スーパーコンピュータ AOBA」を運用しています。さらに本センターでは、学内向けネットワークシステムである「TAINS」を本学構成員向けにサービス提供しています。本センターは本学の AI 技術を活用した研究 DX をさらに加速するため、研究 DX のための研究データ基盤の整備に向けた検討を推進してまいりました。

今回の取り組み

このたび、本学のデータ駆動型研究とオープンサイエンスをさらに加速させるため、本学研究者向けの研究データレイク「IZUMI」(以降、本システム)を整備しました。本システムは東北大学研究データ基盤の中核として、他システムと連携して動作します(図 1)。5 月からの一部利用者による仮運用を経て、6 月より正式運用を開始します(図 2)。本システムは約 5PB の実効容量を持ち、各研究者や研究室が保有する研究データを保持します。さらに、研究データのグループ共有や一般公開する機能を提供します。これまで特に大規模な研究データについては、各研究者や各学内組織が自ら保管先を用意する必要がありましたが、本システムの導入によりこれらの負担が大幅に軽減されます。

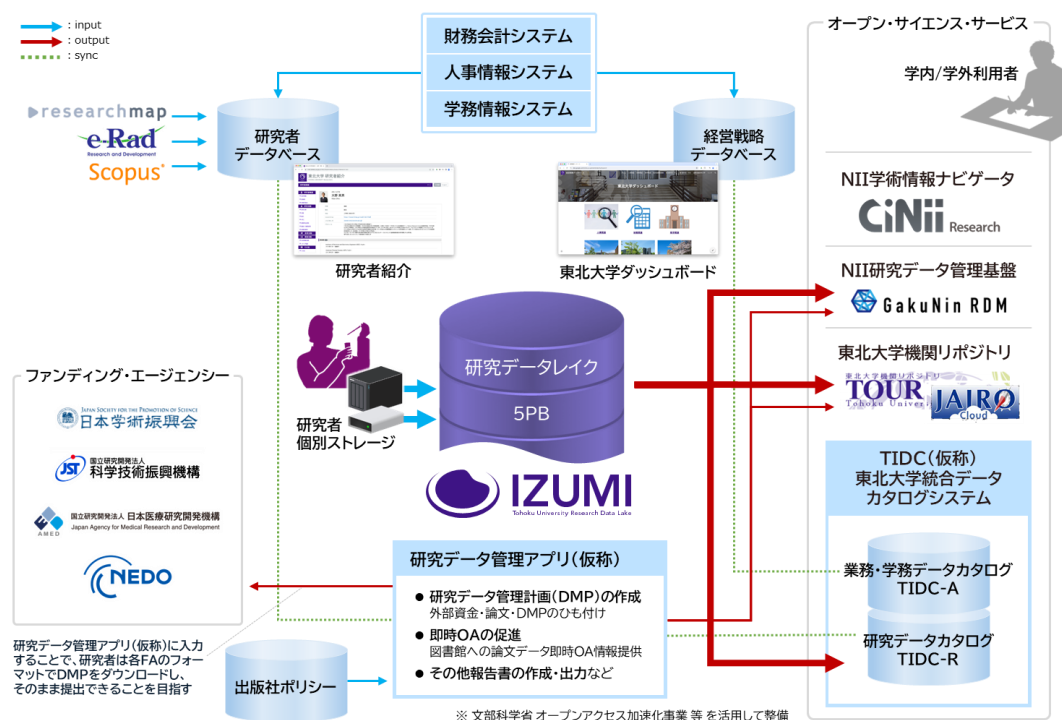


図 1. 東北大学における研究データ基盤のシステム構成図



図 2. 運用を開始する東北大学研究データレイク「IZUMI」の写真

本システムの整備にあたっては、アジリティ(直ぐに利用できること)、ユーザビリティ(簡単に利用できること)、インターオペラビリティ(他のシステムと繋がること)の3点を重視してシステム設計を行いました。アジリティの観点では、本学の約3500人の研究者らがサービスを使いたいと思った際に直ぐに利用できるように、利用申請の不要化や簡易化を重視した設計を行いました。ユーザビリティの観点では、本学の幅広い分野の研究者に簡単に利用頂くために、データレイクで一般的に用いられるオブジェクトストレージサービスに加えて、Webブラウザベースのデータサービスを採用しました。インターオペラビリティの観点では、国立情報学研究所が提供する研究データ管理サービスであるGakuNin RDMの外部ストレージとして連携可能なデータプロトコルを採用しました。

研究データレイクの名称として決定した「IZUMI」(図3:ロゴ)には大きく二つの思いが込められています。一つ目は、本学の豊富に蓄積された研究データから新しい知見や潮流を生み出すことができる知識の泉となるようにという思いです。二つ目は、仙台平野を広く俯瞰する泉ヶ岳にちなみ、仙台市民や本学構成員に広く親しまれるようにという思いです。



図3. 東北大学研究データレイク「IZUMI」のロゴ

本システムは、学内外の研究データ管理サービスとも連携することで、単なる研究データの置き場や共有としての利用だけでなく、研究者の主導による研究データ管理やオープンアクセスを簡単に進められる研究データ基盤の中核として、本学のAI技術を活用した研究DXの発展に貢献します。

東北大学のスーパーコンピュータはAOBAという名前なんだ。
AOBAもIZUMIも仙台市の区や山の名前が由来になっている
から、覚えやすいね！



今後の展開

東北大学研究データレイク「IZUMI」の運用開始により、本学の研究者が生み出す価値の高い研究データが集約されることによって、学内外での研究データの利活用が加速し、本学の研究 DX がさらに発展することが期待されます。今後は利用者からのフィードバックを通じて、さらに使いやすいシステムとなるように、課題整理と検討を進めてまいります。

また、本学の青葉山新キャンパスに隣接する 3GeV 高輝度放射光施設 NanoTerasu など本学の連携組織でも、価値ある大量のデータの創出が進んでいます。今後、学内研究者に留まらず、本学と連携する組織や研究者の研究データの利活用を支援できるように、本システムの拡張整備の検討を進めてまいります。

【東北大学研究データレイク IZUMI の説明サイト】

<https://sites.google.com/tohoku.ac.jp/oids-out/services/izumi>

(本学の研究者が利用できます)

【用語説明】

- 注1. 東北大学コネクテッドユニバシティ戦略: 教育、研究、社会との共創など、本学の諸活動のオンライン化を強力に進めるとともに、サイバー空間とリアル空間の融合的活用を通して、ボーダレスで多様性に富み、真にインクルーシブな大学を創っていく戦略。
- 注2. DX: Digital Transformation の略。デジタル情報やデジタル技術を活用して業務プロセスを変革すること。
- 注3. データレイク: 非構造化データを含む様々な形態のデータを一元管理できるデータの置き場。これらのデータを用いて機械学習などの手法でデータ分析を行うことで新しい価値を創造することができる。

【問い合わせ先】

(システムに関すること)

東北大学 総長特別補佐(情報基盤担当)・教授 中村隆喜
(サイバーサイエンスセンター データプラットフォーム研究部)

TEL: 022-752-2192

Email: takaki@tohoku.ac.jp

(報道に関すること)

東北大学情報部デジタルサービス支援課総務係

TEL: 022-795-3407

Email: cc-som@grp.tohoku.ac.jp