

2023年3月22日

報道機関 各位

国立大学法人東北大学

クラウドを活用した次世代型研究者データベースを構築 — AWS上のデータレイクでエビデンスベースの業績評価を効率化 —

【発表のポイント】

- ・ 学内外の研究者情報を AWS 上のデータレイク^(注1)で効率的に集積します。
- ・ 多面的でエビデンスベースな研究・教育業績評価を自動化・効率化します。
- ・ 2023年4月よりデータベースの本格稼働を開始します。
- ・ 世界と伍する研究大学のためのデータベースとして運用します。

【概要】

現在、大学には国際的プレゼンス向上、若手研究者の登用、多様性と包摂性、産学連携推進など、様々な課題があります。これら課題解決の一助として、研究者の多様な業績を効率的に集積し、エビデンスベースで評価・分析するシステムの構築が求められています。東北大学（総長：大野 英男）は、学外の論文情報データベースと、学内に蓄積された人事・研究費・教育実績等の多種多様なフォーマットのデータをアマゾン ウェブ サービス（以下、AWS）^(注2)のクラウドサービスで構築したデータレイクに集約し、瞬時に評価・分析・可視化が可能なシステムを開発しました。これにより、論文被引用数等の画一的指標に限らず、共同研究、学生指導、管理運営の実績が自動的に集積され、より多面的でエビデンスベースな研究者の評価・分析が可能となりました。同時に、大学経営戦略のDX、研究者の研究時間確保への貢献も期待されます。

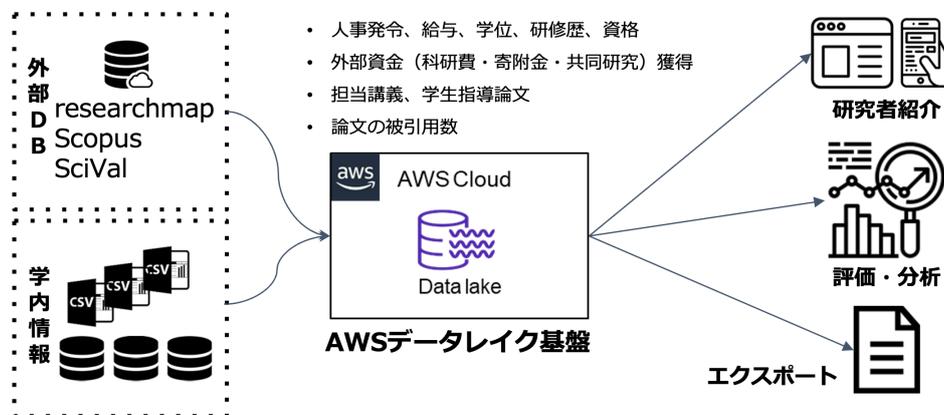


図1 AWS上のデータレイクで構築された研究者データベースの模式図

【AWS コメント】

アマゾン ウェブ サービス ジャパン合同会社

執行役員 パブリックセクター 統括本部長 宇佐見 潮 氏

「AWS 上のデータレイクを基盤としたサービスは世界中の大学・研究機関で活用されています。この度東北大学様におかれましては、AWS のサービスの活用によりエビデンスベースの研究者評価・分析、可視化が進み、より世界への発信力が向上し、世界に伍する研究大学へさらに邁進されることとなるでしょう。そのような東北大学様のご発展に寄り添いつつご支援できることを大変光栄に思います。」

【詳細な説明】

国際卓越研究大学の公募が開始されるなど、我が国の大学を取り巻く環境は大きく変わろうとしています。現在、大学では国際的プレゼンスの向上、若手研究者の登用、多様性と包摂性の促進、産学連携推進など様々な課題に取り組む必要がありますが、そのためには、研究者や組織の教育・研究アクティビティを多面的かつエビデンスベースで評価・分析し、大学経営戦略にフィードバックする必要があります。近年、国内の研究者データベース (researchmap^(注3) や KAKEN^(注4))、海外の論文情報データベース (Scopus^(注5) や Web of Science^(注6)) により、論文数や被引用数、科研費獲得額等は比較的容易に集積・分析が可能であり、内閣府エビデンスシステム(e-CSTI^(注7))では、研究者数当たりの論文被引用数や寄付金額等が大学・研究機関別に可視化されています。

一方、大学では、秘匿性がある企業との共同研究や、論文指導、管理運營業務など外部データベースに集積しにくい教育・研究活動があり、従来、それら実績は研究者が自身の研究時間を割いて学内のデータベースに入力しています。またデータの inputs は研究者に依存するため、量や質に大きな差があり、エビデンスベースの評価・分析が難しい要因になっていました。さらに大学や研究者に求められる評価基準は固定されたものでなく、学問分野により異なり、時代でも変化します。そのため入出力データの形式を事前に決めて設計する既存のデータベースでは、これら多面的な評価に迅速に対応できない課題がありました。

そこで学内外に散在する研究者の情報を自動的・効率的に集積し、必要なときに必要な情報を可視化できる次世代型研究者データベースの開発に取り組みました (図 1)。このシステム開発では、基盤としてセキュリティ・堅牢性・可用性の観点から AWS のクラウドサービスを用い、データレイクを構築しました。そこに学外の論文情報データベース、学内の人事給与統合システム・財務会計システム・学務情報システム等の様々なフォーマットのファイルを投入すると、研究者や評価担当者が必要なときに必要なフォーマットで情報を加工・取り出すことができ、ビジネスインテリジェンス(BI)ツールで可視化できます (図 2)。

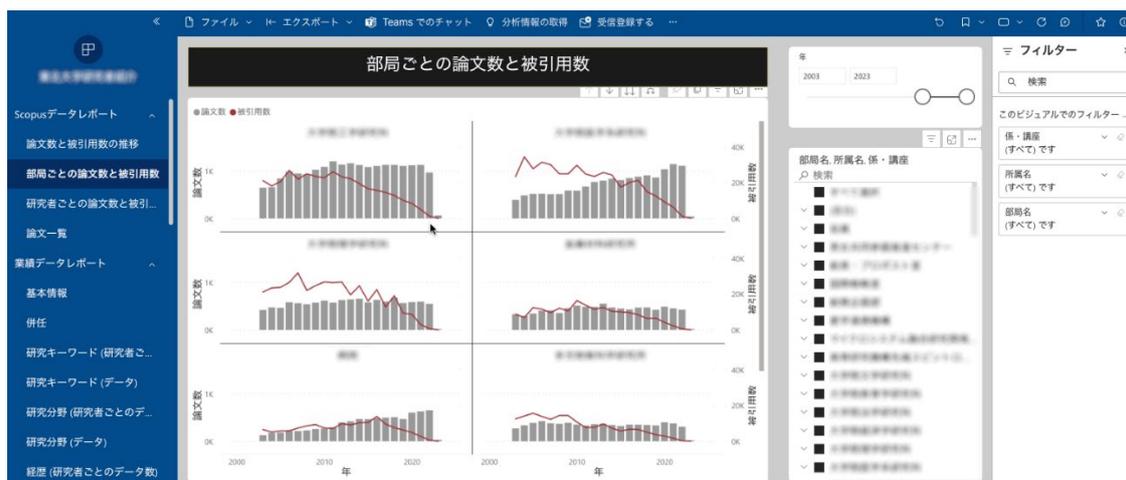


図 2 BI ツール（Microsoft PowerBI）で分析結果を可視化した例

東北大学では、この AWS 上のデータレイクを用いた研究者データベースの試験運用を開始し、2023 年 4 月より本格稼働させます。これにより、外部データベースでは集積ができない企業との共同研究費、論文指導実績、講義担当数、留学生受け入れ実績、管理運營業務などが自動的に集計され、研究者・研究グループの教育・研究活動の評価がより多面的かつエビデンスベースになると期待できます。同時に、研究者自身の入力を全く必要としないため、研究時間の確保にも貢献します。また、リアルタイムでこれらアクティビティが可視化されるため、世界と伍する研究大学を目指すための大学経営戦略にも迅速なフィードバックが可能と期待されます。

- 注1) AWS 上のデータレイク: AWS が提供するクラウドサービスを活用した、すべての構造化データと非構造化データを保存することができる一元化されたレポジトリ。様々なデータをそのままの形式で保存し、データ処理、可視化、リアルタイム分析、機械学習などを実行することでの確な意思決定に役立てることができる環境。
- 注2) アマゾン ウェブ サービス (AWS) : 2006 年より同名の企業がグローバルに提供するクラウドサービス。
- 注3) researchmap: 科学技術振興機構(JST)が運用する研究者データベース・ポータルサイト (<https://researchmap.jp>)。
- 注4) KAKEN: 国立情報学研究所(NII)が運用する科学研究費助成データベース (<https://kaken.nii.ac.jp>)。
- 注5) Scopus: Elsevier 社が提供する論文データベース。書誌情報に加えて、被引用数などが参照可能(<https://www.scopus.com>)。

- 注6) Web of Science: Clarivate 社が提供する論文データベース。論文情報に加えて、論文査読数なども参照可能(<https://www.webofscience.com>)。
- 注7) e-CSTI: 内閣府が運用し、国立大学法人や国立研究開発法人の研究力等を分析・可視化するシステム (<https://e-csti.go.jp>)。

【問い合わせ先】

(取材に関すること)

東北大学 総長・プロボスト室

担当 安孫子 寛樹

電話 022-217-4805

E-mail hyoka@grp.tohoku.ac.jp

(技術に関すること)

東北大学大学院工学研究科

総長特別補佐 (大学改革担当)

教授 高村 仁

電話 022-795-3938

E-mail takamura@material.tohoku.ac.jp