

早稲田大学 × CTC

早稲田大学とCTCの協働で構築された「データ分析共有プラットフォーム」が

データサイエンスの新しい未来を拓く

「高度なデータ分析スキル」を持った研究者と、様々な領域における「深い専門知識」を持った研究者との融合による、データ駆動型最先端研究の推進と、高度IT人材の育成を目指し、早稲田大学によって2017年12月1日に設立されたのが、「データ科学総合研究教育センター」だ。同センターは伊藤忠テクノソリューションズ(CTC)との学術連携協定を礎とするパートナーシップにより、データ分析共有プラットフォームの試験的運用を推進している。

データサイエンス研究・教育における世界最先端の拠点を目指す

ビッグデータやIoTの普及、AI（人工知能）といった先進技術の進歩に伴い、データ解析技術も大きく発展を遂げる中で、実社会だけでなく、理工系・人文社会系を問わず、あらゆる学問・研究領域においてデータ科学の重要性が高まっている。これまでそれぞれの専門領域で積み重ねられてきた「理論」に「データによる実証」を融合させることで、新しい学問や研究の展開が期待されているからだ。

そうした時代の要請に応えるべく、2017年12月1日に早稲田大学が設立したのが、「データ科学総合研究教育センター（CDS：Center

for Data Science、以下、DSセンター）」である。DSセンターは、「高度なデータ分析スキル」を持った研究者と様々な領域における「深い専門知識」を持った研究者との融合によりデータ駆動型最先端研究を推進するとともに、専門知識とデータ分析スキルを兼ね備えた実践的な人材の育成を目指すもの。具体的には情報検索やデータベース、統計解析、機械学習やAI、画像・音声・自然言語処理といった、広義のデータ分析に関連する研究者と、政治、経済をはじめ、経営、マーケティング、金融、スポーツ、文学、芸術など、あらゆる専門分野の研究者が協力し合える、いわば“ハブ”となるプラットフォームを形成することで、新たなデータ分析と学際総合研究の創出を目指している。

DSセンターの所長を務める松嶋 敏泰氏は、「早稲田大学の創設者である大隈重信は、大学創立の1年前となる1881年に日本初の統計機関である統計院を設置するなど、早くからデータ科学の重要性を訴え、その発展に大きく貢献してきました。統計院の設置から136年後に創設されたDSセンターは、“国家の発展にはデータ科学に基づく施策が不可欠”という大隈の意志を



早稲田大学 データ科学総合研究教育センター 教授 小林 学 氏



早稲田大学 データ科学総合研究教育センター 所長 理工学術院 教授 松嶋 敏泰 氏

受け継いだ施設でもあります」と、設立に際して込められた想いを語る。

また、DSセンター 教授 小林 学氏も「DSセンターの最大の特徴は、データサイエンスを専門とする学生や研究者だけを対象とするのではなく、すべての学生や研究者に研究や教育の場を提供している点にあります。今後は早稲田大学内だけでなく、国内外の大学や研究機関、企業、公的機関ともネットワークを形成し、実践的教育の普及と最先端研究を推進する拠点に発展させていきたいと考えています」と語る。

「WIRP」を具現化するデータ分析共有プラットフォーム

DSセンターの担う役割の1つに、「WIRP (Waseda Integrated Research Platform)」の推進がある。WIRPは、国や地方自治体、公共団体や企業などが持つ様々なデータを、大学の研究・教育に有効活用するためのプラットフォームである。公共機関や各種企業は種々のビッグデータを保有しているが、昨今、研究や教育を目的にしたデータの有効な利活用が叫ばれている。また、大学においてはこれらの実データを扱うことによる研究力および教育力の向上が非常に重要となっている。

そして早稲田大学は、このWIRPを具現化するための基盤となる「データ分析共有プラットフォーム」の試験的構築運用をCTCとの協働により進めている。早稲田大学とCTCは2018年7月、データサイエンス力の高い人材の育成と、大学の教育研究活動および産学連携による共同研究の促進を目的に学術交流協定を締結。データ分析やAI開発の促進によりデータサイエンスの裾野を広げるべく、相互支援を行っている。データ分析共有プラットフォームの取り組みも、その一環となるものだ。



早稲田大学 データ科学総合研究教育センター 講師 安田 豪毅 氏

セキュアなデータアクセスを支える、NetAppのストレージソリューション

データ分析共有プラットフォームの構築にあたり掲げられた要件が、「国内外の大学や企業が共用可能なプラットフォーム」かつ「セキュアなデータアクセス基盤」の実現だ。

「学内だけでなく他大学や企業、公共機関などからデータを提供してもらったり、各利用者がデータ分析を行ったりするのに際して、公開データと機密データ、および各プロジェクトのデータを分離可能な、セキュアなデータアクセス環境の実現が求められていました」(松嶋氏)

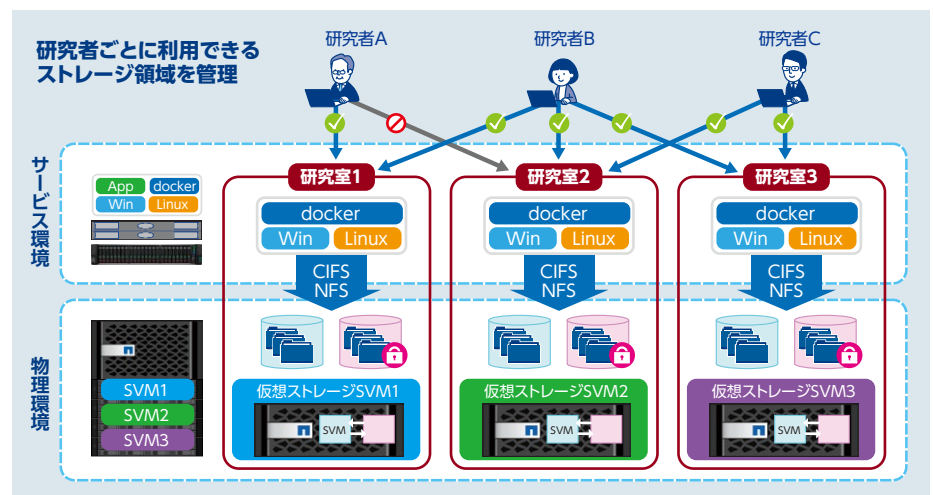
これらの要件に基づき、データ分析共有プラットフォームは、Dockerを用いてJupyter NotebookでPythonやRを実行する環境をコンテナ化し、kubernetesによるデプロイやスケールアップ、管理などの機能を実装することで、GPUを含めた分析・開発環境を利用者自身がカタログから容易に展開可能としている。また、セキュリティについてもコンテナを含むシステムのユーザー認証やアクセス管理はActive Directoryと連携し、データをDSセンター側のサーバで一元管理するとともに、専用のリモートデスクトップサーバ経由でアクセスを行わせることで、データがローカル側に残らないような仕組みを実現している。

こうしたシステムを具現化するにあたり、データ格納基盤として選択されたのが、NetAppのストレージソリューションだ。NetAppの「Storage Virtual Machine (SVM)」機能により、セキュリティ要件に応じて利用者ごとにス

トレージ領域を仮想的に分離しマルチテナント化を行うことで、セキュアなデータ活用も可能とする。さらに「NetApp Trident」機能により、コンテナ環境下における永続的なデータ保持やバージョン管理を可能とした。

現在、データ分析共有プラットフォームは学内展開に向けて、性能評価や課題の洗い出しを行っている段階にあるという。DSセンターで講師を務める安田 豪毅氏は、「本格的な稼働に向けて、CTCの協力を仰ぎながら物理的なシステムも合わせた改善に取り組んでいます。目標はいち早くこのプラットフォームの稼働を本格化し、WIRPを日本や世界に広げていくことですね」と語る。

今後の展望について小林氏は、「CTCとの協働により、私たちはこれまで経験したことのないようなチャレンジに取り組んでいます。先に述べたように、このプラットフォームを学内だけでなく国内外の大学、企業を結び付ける基盤へと発展させていきたいと考えています」と話す。そして松嶋氏も、「データ分析共有プラットフォームで実現される仕組みを、世界のデータサイエンス教育や研究におけるデファクトスタンダードへと成長させていきたいですね。そのためにもCTCにはこれまでと変わらない支援と協力をお願いしたいと考えています」と強調した。



早稲田大学データ科学総合研究教育センター(CDS:Center for Data Science)

2017年12月に設立。理工系・人文社会科学系の専門領域とデータ科学との知見の融合を図るプラットフォームの提供により、総合知の創造とグローバル社会の問題解決を担う人材の育成および大学の研究力向上を目指す。

<https://www.waseda.jp/inst/cds/>

伊藤忠テクノソリューションズ株式会社

〒105-6950 東京都港区虎ノ門4-1-1 神谷町トラストタワー
mrc-mvr@ctc-g.co.jp <http://www.ctc-g.co.jp>

記載されている会社名、製品名、サービス名は伊藤忠テクノソリューションズ(株)または各社の商標もしくは登録商標です。
©TOCHU Techno-Solutions Corporation

