

# 東北大学未踏スケールデータ アナリティクスセンターの 概要とミッション



TOHOKU  
UNIVERSITY

令和4年（2022年）12月26日（月）

センター長 中尾光之

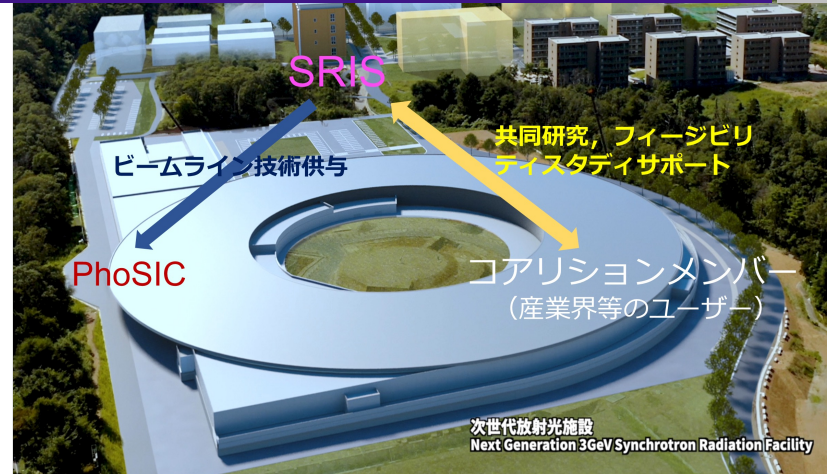
# データ駆動科学は未踏スケールの時代へ



東北メディカル・メガバンク：世界最大規模の15万人フルゲノム、7万人3世代コホート+情報科学と医学の分野融合研究者 => 数少ないビッグデータ研究の成功事例



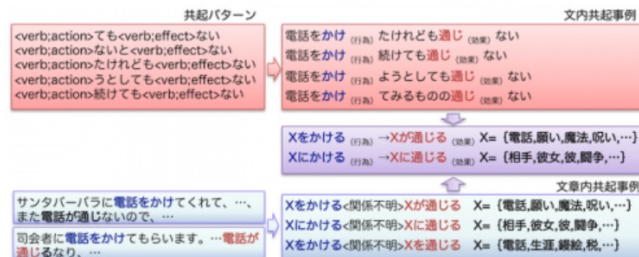
クライオ電子顕微鏡単粒子解析：生体機能分子立体構造の高解像度測定



次世代放射光施設（ナノテラス）：従来の軟X線領域の100倍の輝度・コヒーレンスで、ナノの機能を可視化することにより、農学、材料科学、理学、工学、環境科学、生命科学、創薬など様々な分野での画期的な研究成果が期待される。学術、産業界のデータ駆動型研究開発を支援すると共に新しい利用制度“コアリション・コンセプト”で、産学融合を加速する。2023年稼働開始。

Knowledge Acquisition from Large-scale Text Data / 大規模言語データからの知識獲得

数億文規模の大規模言語データから、例えばイベント間の因果関係や目的手段関係などの知識を自動獲得します。獲得した知識は意味・談話解析の高度化に利用します。



大規模言語データからの知識獲得



CDOを機構長としたデータシナジー創生機構を構築し、機構長の強力なガバナンスの下、3つの組織を一体的に連動させながら教育・研究活動を展開して、分野を越境したデータ・解析手法・人材・設備の一元化・共有化・強化を強力に推進

## サイバーサイエンスセンター(CSC)

- 民間企業との協働(共同研究部門)による次世代スーパーコンピュータの研究開発
- スーパーコンピュータを活用した実用システムの社会実装に関する研究開発
- 耐災害性、セキュリティ強度に優れた柔軟な高速ネットワークに関する研究開発
- 未踏スケールデータに対応した超大規模データレイクに関する研究開発

## データ駆動科学・AI教育研究センター(CDS)

- 学部新生全員に対するリテラシー教育 (AI & Data for All) 、学部専門教育でのスペシャリスト養成 (Specialists with AI & Data)、そして、大学院での高度AIプロフェッショナル人材育成 (AI & Data Top leader) による、全学的な教育体制
- 学修データ利活用による教育のデジタル変革への取り組み

## 未踏スケールデータアナリティクスセンター(UDAC)

未踏スケールデータ (サイズ、分解能、スピード、多様性において、従来のスケールを遥かに凌駕するデータ) のデータ科学を創生し、未踏領域の価値創造を実現する。その成果に基づいた産学連携、人材育成、起業を促進する。

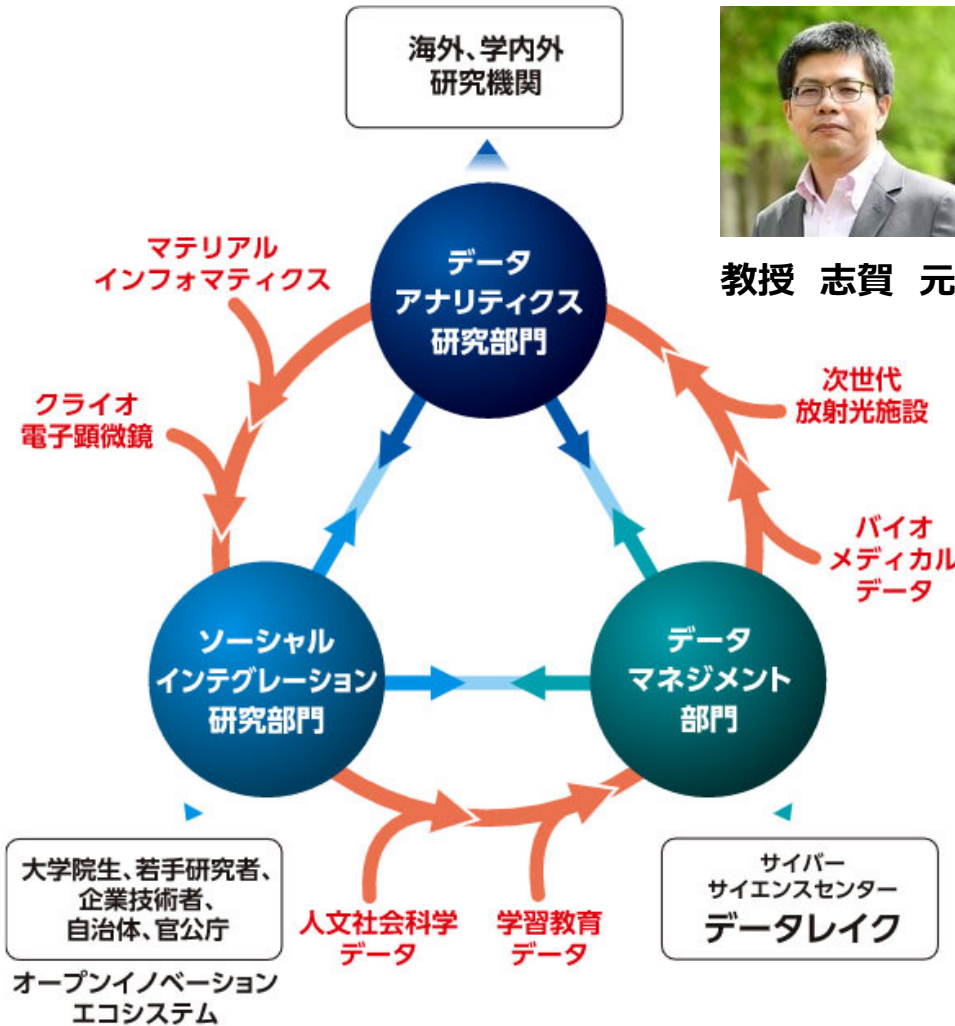


# 未踏スケールデータアナリティクスセンター

## Unprecedented-scale Data Analytics Center



教授 山田 和範

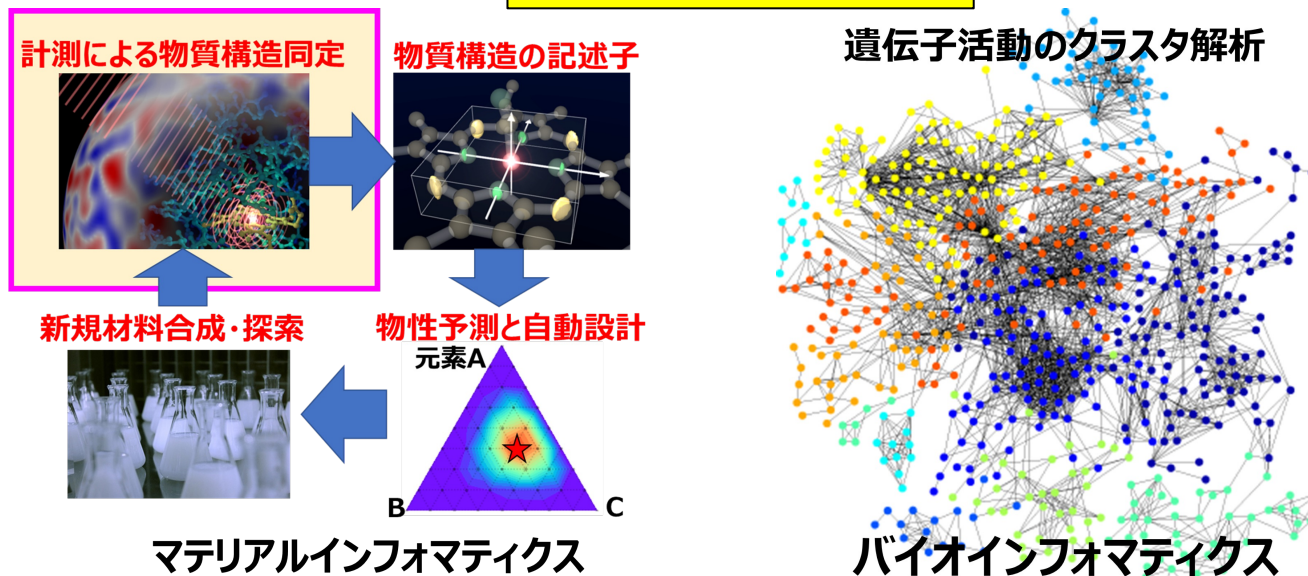


教授 志賀 元紀

**未踏スケールデータのデータアナリティクスを極め、学術・産業界と共有します。**

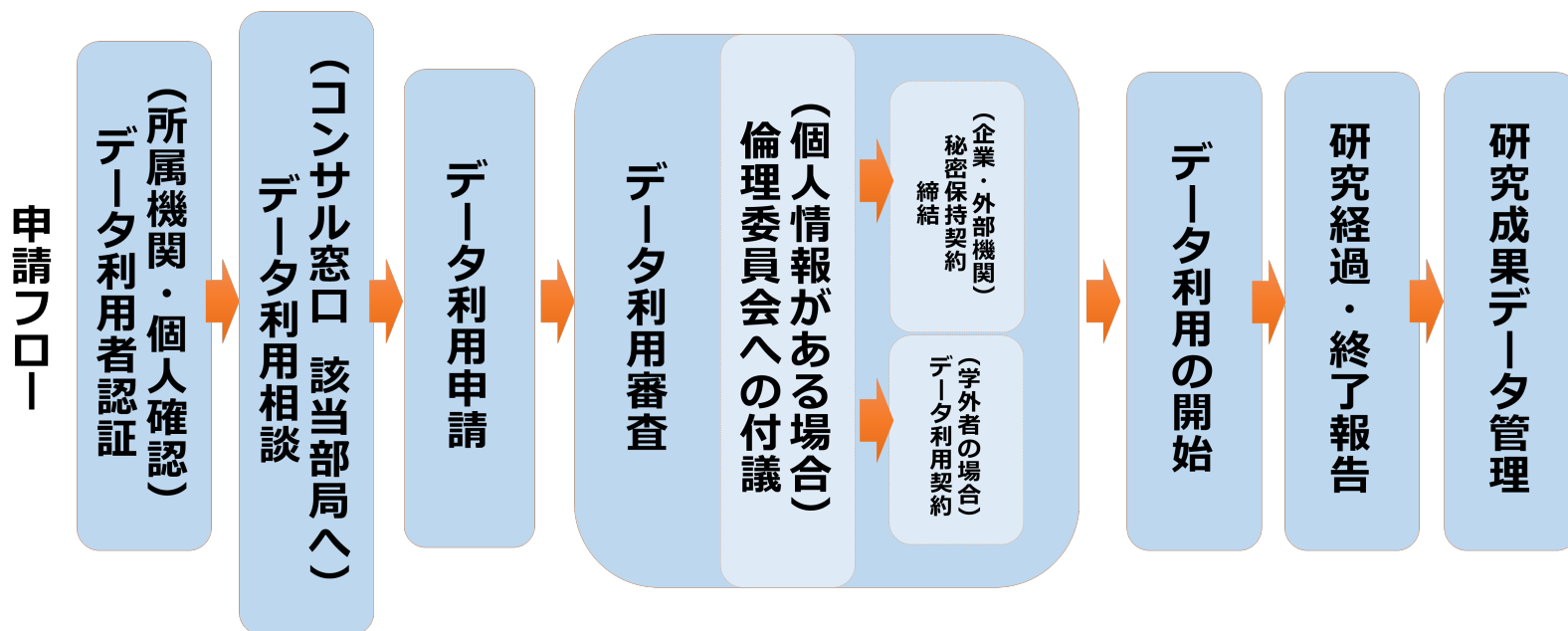
- 先端的なAI やデータ科学をさらに発展させることによって未踏スケールデータのハンドリング・分析手法の研究開発
- 分野を横断した研究領域融合による新たな分析技術の開発・共有、そして、新たな社会的価値の創造
- 研究・教育データの高度な加工・分析する技術に関する研究開発および研究教育拠点に対するコンサルティング
- データ駆動型の新しい教育改善手法の開発とその実践的な検証、およびそのサービス化に関する研究

## 志賀元紀教授の研究



**未踏スケール研究データの効率的でセキュアな利用を促進するマネジメントを行います。**

- 研究デジタルデータ保存サービスとアーカイブ
- データ関連の法令遵守、規程制定、契約業務
- 研究データ管理に係る機能・業務を集約化
- サイバーサイエンスセンターと連携し超大規模ストレージ(データレイク)、スパコン、高速ネットワーク等を基盤としたデータプラットフォーム構築支援、並びにそれに基づく未踏スケールデータの収集・格納・管理・共有利用



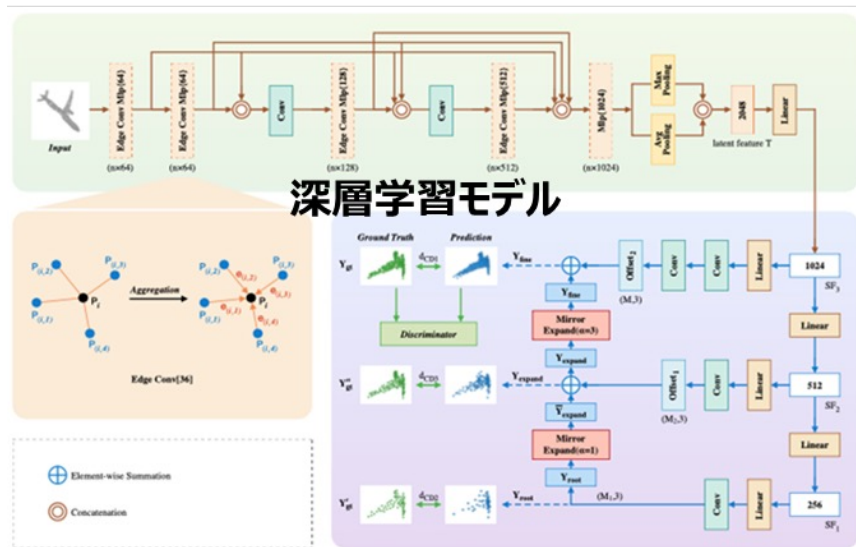


**未踏スケールデータに挑む学生・研究者・企業技術者を受入れ、オープンエコシステムによる価値創生や起業を支援します。**

- 学内の学部・研究科・研究所・センター等との連携、および海外研究教育機関、企業・政府・自治体との共同研究による社会実装の推進
- データ駆動科学・AI 教育研究センターとの協働による社会人特別プログラムの企画
- オープンエコシステムによる国内外からの大学院生、若手研究者、企業技術者の受入れ、価値創生への取組みや起業の支援

## 山田和範教授の研究

### AIアルゴリズム開発



### ドメインデータサイエンティスト養成

1. データ科学と機械学習
2. Pythonの基本的な使用方法
3. Pythonの発展的な使用方法
4. 教師あり学習法
5. 教師なし学習法
6. TensorFlowの基本的な利用方法
7. 多層パーセプトロン
8. 再帰型ニューラルネットワーク
9. アテンションネットワーク
10. Hugging Faceの利用方法
11. 敵対的生成ネットワーク
12. 強化学習法



## Senior Domain Specialist with AIMD

### データ科学国際共同大学院

- 情報科学研究科が6研究科と連携して提供する，留学生と日本人学生のための双方向・分野横断型学習環境
- 実世界のビッグデータ課題に取り組むプロジェクト型学習 (PBL) の実施
- 海外研究機関のメンターとの最低6ヶ月間の共同研究機会の提供

情報科学研究科および連携する大学院群

## Leading Researcher with Advanced AIMD

- 未踏スケールデータ科学を学び、研究を行う機会を提供
- アナリスト、ビジネスエキスパートとのユニットによる課題解決を行いデータ科学の利活用の仕方を理解する

未踏スケールデータアナリティクスセンター (UDAC)

## Domain Specialists with AIMD

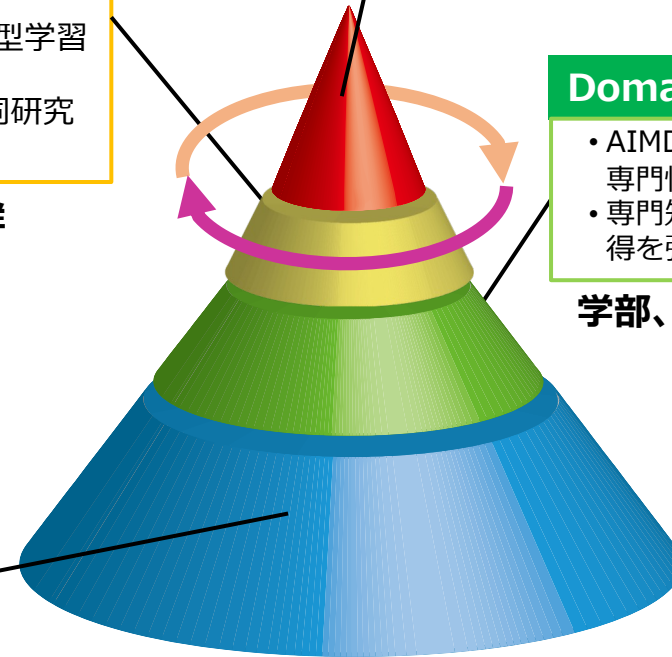
- AIMDについて学ぶ機会を文理問わず学生の専門性に依拠して提供する
- 専門知識とAIMDスキルのダブルメジャー修得を強く推奨する

学部、学科専門教育

## AIMD for All

- 新入生全員 (2,500名) にAIMDリテラシー教育を実施
- AIMDの先進的なトピックを学ぶと共に、現在進行中の研究とその最前線を知る機会を提供し、学生のモチベーションを高めるためのユニークなコース

全学教育







UDACには、それぞれがミッションを有する部門体制がありますが、まだ生まれたばかりの小さな組織である我々は、アナリティクスにも人材育成にも環境づくりにも全員総出で取組みます。

以下では、アナリティクスの研究内容を志賀教授が、人材育成と環境づくりを山田教授が説明します。