

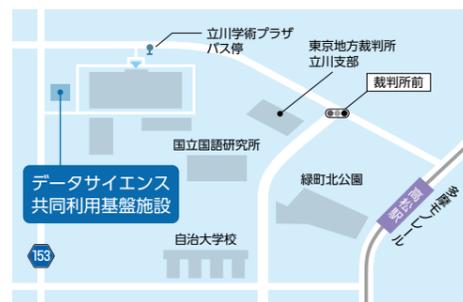
【ロゴマーク紹介】



データサイエンスの略であるDSの文字で地球(地軸)の傾きを、そして周囲は4つの研究所を表現しています。
全体では、枠に収まらずに形を変えながら発展するさまを表すとともに、親しみのあるシルエットにより社会貢献を意味する組織をイメージしています。

データサイエンス共同利用基盤施設

〒190-0014 東京都立川市緑町10-3 データサイエンス棟
URL: <https://ds.rois.ac.jp/>



詳しくはWEBへ

大学共同利用機関法人 情報・システム研究機構

データサイエンス 共同利用基盤施設

Joint Support-Center for Data Science Research

ROIS-DS

Data for All

パンフレット 202410

データサイエンス
共同利用基盤施設

施設長

喜連川 優



Data for All ～データで未来を切り拓く新たな挑戦～

私たちの周囲には、無数のデータが存在しています。現代社会が直面する複雑な課題を解決し、より効果的で持続可能な意思決定を可能とするには、データを読み解く技術であるデータサイエンスの重要性が一層高まっています。2022年にChatGPTが誕生して以来、生成AIの技術の進展は加速度的に進んでおり、これまで以上に多様かつ膨大なデータが日々生成されています。それに伴い、ハルシネーション、倫理的な課題、不適切な表現や権利侵害など、新たな問題も露呈しています。当施設は「Data for All」という理念のもと、社会課題の解決に寄与するべく、皆様とともに未来を切り拓く挑戦を続けていきたいと考えております。

まず、時代の要請に先駆けた新しいサービスを創り出す組織として、柔軟な運営により革新的かつ挑戦的な取組を推進します。従来より広く学術界に提供してきたデータ共有・データ解析支援に加え、今後はデータサイエンスを軸とした、さまざまな学問領域をカバーするインキュベーション組織を目指します。

2023年11月には「人工知能法学研究支援センター（人工知能法学研究センターに改名）」を設置しました。当センターは「人工知能による法学研究支援」と「法による人工知能制御」を2つの柱とする「人工知能法学」という新たな学問分野を提唱し、関連データベースの作成とその活用方法に係る研究による人工知能法学の世界的拠点となることを目標としています。

また2024年4月には、質の高い多種多様な大規模データを系統的に収集し、安心・安全かつ持続的にデータ提供できる枠組みの構築を目標とする「データレイク研究開発センター」を設置しました。2024年6月には内閣府戦略的イノベーション創造プログラムに採択され、これを受け、医療データならびに医療LLM/LMMの利活用を促進する医療データ基盤を構築していきます。

このようなセンターの活動により、研究現場からのフィードバックを直接活用し、社会ニーズに即応した新しい科学分野の探求を進めることが可能となり、迅速かつ効率的に資源を活用し、革新的な研究を支援するための重要な基盤となることを目指します。

これまでにない「新しいサービス」の創出、それを実現できる「人材」の育成、データサイエンスを軸とした「共同研究の場」の構築を柱とし、当施設が持つ潜在力を国内外に向けて最大限に引き出していく所存です。

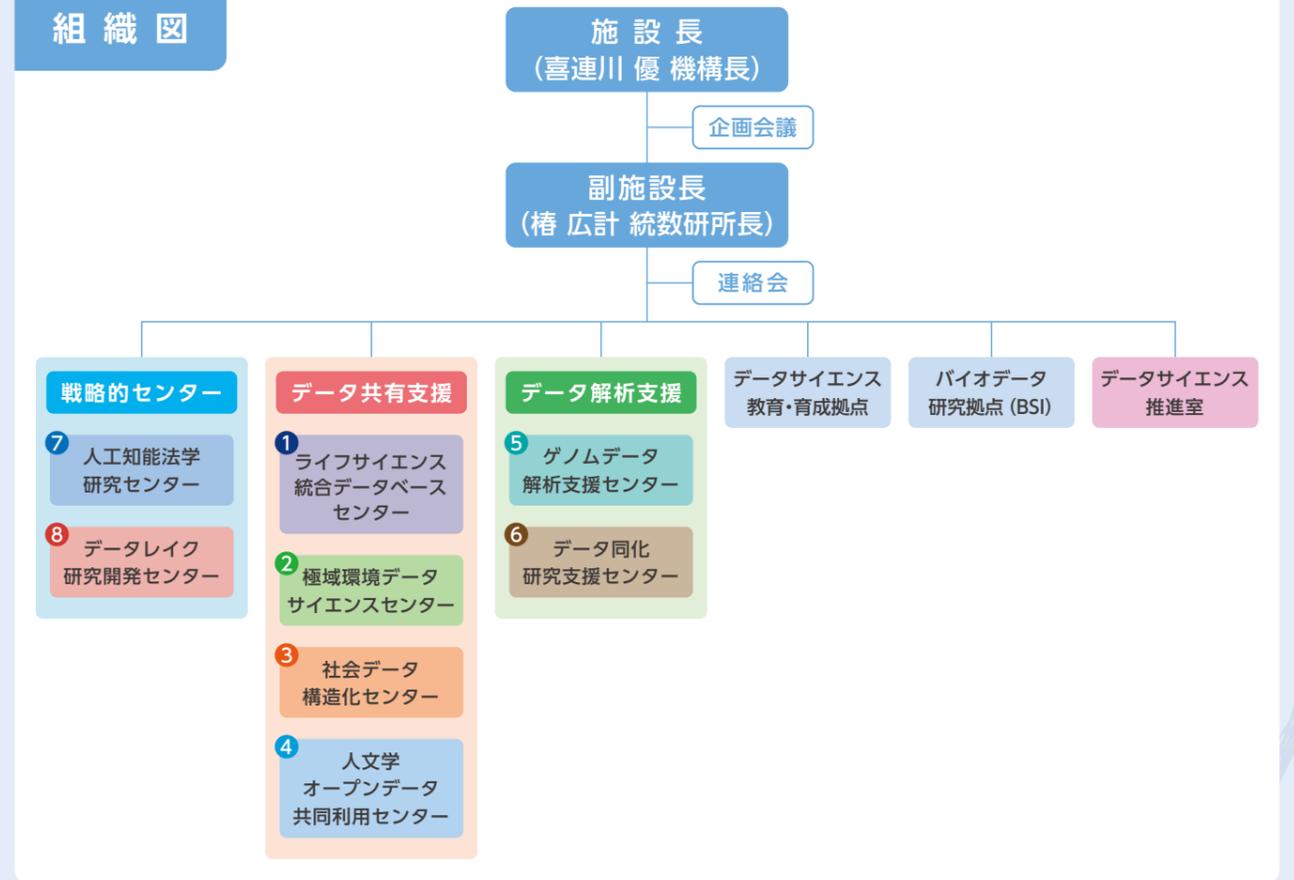
皆さまとともに、共に学び、共に創り出すことができる環境を整え、多くの革新的なプロジェクトを成功に導くべく、一層努力してまいります。今後とも、皆さまのご指導とご協力を賜りますよう、お願い申し上げます。

目次

Contents

▶ 施設長着任のご挨拶	01	▶ データ解析支援	
▶ 組織の構成	02	ゲノムデータ解析支援センター	09
▶ 事業内容紹介/戦略的センターについて	03	データ同化研究支援センター	10
▶ 拠点活動	04	▶ 戦略的センター	
▶ データ共有支援		人工知能法学研究センター	11
ライフサイエンス統合データベースセンター	05	データレイク研究開発センター	12
極域環境データサイエンスセンター	06	▶ 公募型共同研究	13
社会データ構造化センター	07	▶ DS推進室の活動・他	14
人文学オープンデータ共同利用センター	08		

組織図



1 ライフサイエンス統合データベースセンター Database Center for Life Science (DBCLS) P05へ

生命科学分野のオープンサイエンス推進、ライフサイエンス・データベース統合化のための基盤技術開発を推進

2 極域環境データサイエンスセンター Polar Environment Data Science Center (PEDSC) P06へ

過去から現在に至る長大な時間軸を持った極域環境変動・地球システム変動に関する貴重なデータと、その分析・解析支援を提供する共同利用を推進

3 社会データ構造化センター Center for Social Data Structuring (CSDS) P07へ

大学研究者のための社会調査データ、公的調査マイクロデータ、ソーシャルビッグデータに関するデータベースを整備。また、データ利用コミュニティを形成し、環境、安全、経済を含む各種の社会的課題の解決のための実証的研究を促進

4 人文学オープンデータ共同利用センター Center for Open Data in the Humanities (CODH) P08へ

データサイエンスに基づく人文学（人情情報学）という新たな学問分野を創生するとともに、データを中心としたオープン化を推進することで、組織の枠を超えた研究拠点を形成・強化

5 ゲノムデータ解析支援センター Center for Genome Informatics (CGI) P09へ

最先端のバイオインフォマティクス技術を駆使して大量のゲノム・トランスクリプトームデータから生物学的に重要な情報を得るためのデータ解析支援

6 データ同化研究支援センター Center for Data Assimilation Research and Applications (CARA) P10へ

データ同化計算技術による諸科学・産業界での課題解決

7 人工知能法学研究センター Center for Juris-Informatics (CJI) P11へ

「人工知能による法学研究支援」と「法による人工知能制御」を2つの柱とする「人工知能法学」という新たな学問分野の創生、および、関連データベースの作成とその活用方法の研究による人工知能法学の世界的拠点の構築

8 データレイク研究開発センター Center for Research and Development on Data Lake (DLRD) P12へ

質の高い多種多様な大規模データが求められる今日、体系的なデータ収集と安心・安全かつ持続的なデータ提供を目指し、生成系AI、医療など各分野の研究を加速する枠組みを構築

支援事業(データ共有支援)

生命科学分野におけるデータ共有支援事業
 極域環境科学分野におけるデータ共有支援事業
 人間・社会分野におけるデータ共有支援事業
 人文学オープンデータ共有支援事業

支援事業(データ解析支援)

ゲノムデータ解析支援事業
 データ融合計算支援事業

人材育成事業(データサイエンティスト育成)

データサイエンス高度研究教育者 育成事業

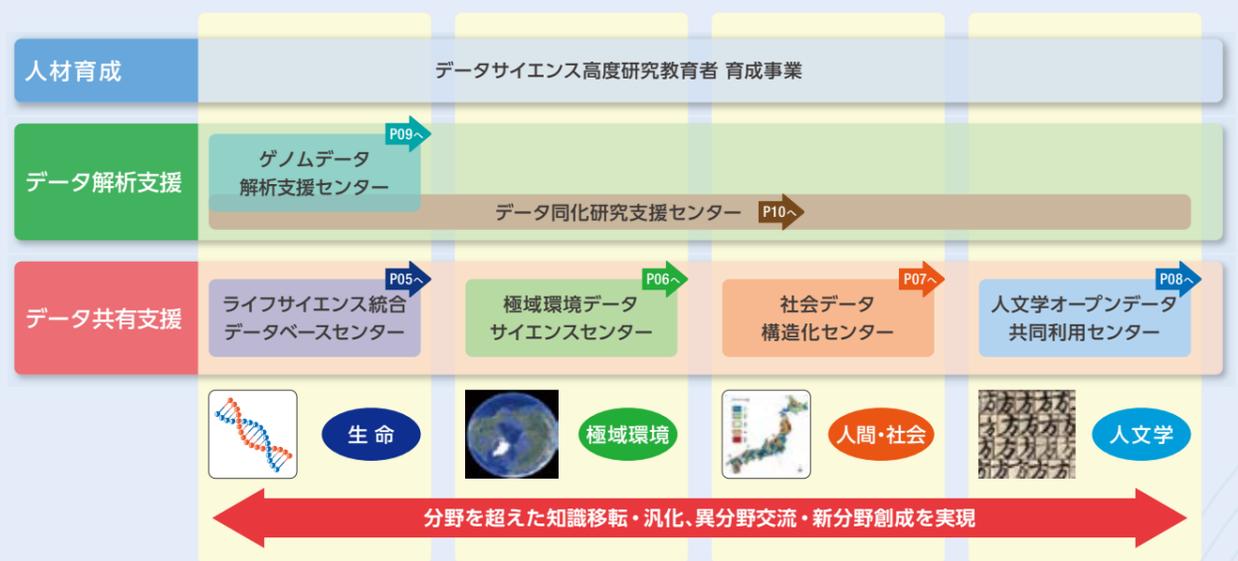
公募型共同研究(ROIS-DS-JOINT)

一般共同研究
 共同研究集会

P13を
 ご覧ください

データ共有支援事業・解析支援事業および共同利用・共同研究を必要としている
 大学等のすべての研究者

支援事業
 共同利用・共同研究



戦略的センターについて

戦略的センターは、国家戦略や社会の多岐にわたる課題解決に向けて、先進的かつ多面的な研究活動を展開し、広く社会に貢献する新しいサービスや技術の創出を目指して、2024年に設置された新たなDS施設内の革新的な組織です。このセンターは、外部資金の活用を通じて、機動的かつ柔軟に質の高い研究環境を整備していくことを目指しています。DS施設における革新的なサービスの創出と、社会からの多様な要請に迅速かつ的確に応える研究を力強く推進していきます。

データサイエンス教育・育成拠点

全国の大学等の若手研究者をデータサイエンスの高度研究教育者に育成

最近、データサイエンス系の学部・学科を新設する大学が急増しています。このような中で、深刻な問題となっているのが、データサイエンスの基盤となる統計科学の専門教員の不足です。この問題に対応するため、2021年に、全国の大学とともにデータサイエンス共同利用基盤施設が参画する「統計エキスパート人材育成コンソーシアム」(中核機関:統計数理研究所)が発足しました。コンソーシアムでは、様々な学術分野の若手研究者に対して研修を行うことにより、大学院修士学生に統計学を講義し、統計を活用した研究を指導することのできる大学教員を育成しています。

データサイエンス教育・育成拠点では、ここで育成された若手大学教員に対して、統計科学の教育・研究指導と必要な支援を行う特別研修を実施することにより、更に高度な統計教育・研究指導と統計活用研究を行うことができる高度研究教育者を育成しています。このほか、総合研究大学院大学の「データサイエンス型研究者人材養成事業」に当拠点所属の教員がメンターとして参画し、総研大基盤機関の若手研究者をデータサイエンス型研究者に育成しています。



大学統計教員育成研修 中間報告会 (2024年6月14日)

バイオデータ研究拠点

生命科学における「知識の体系化」の推進とデータサイエンスの基盤形成

バイオデータ研究拠点(BioData Science Initiative: BSI)は、分野横断型研究ハブ組織としてDS施設と国立遺伝学研究所を中心とした機構内連携のもと2022年4月に設置されました。バイオデータとは、生物の特性、機能、状態、さらには生物集団の構造とダイナミクス、生物間の相互作用および生物と環境との相互作用に関する定量的かつ定性的な情報です。BSIは、多様な生物種・実験条件で得られた大量のゲノム情報をはじめとする広範なバイオデータと文献などに含まれる知識を標準化しバイオ知識の体系化を推進します。バイオデータを一体的に登録・集積・整理・標準化し提供する日本およびアジアの中核拠点として、バイオデータサイエンスの発展に貢献する研究インフラを構築し国内外の連携ネットワークを通して広く提供します。また、分野別リポジトリとしてデータの登録拠点の機能も持ち、研究者への一体的窓口として、オープンサイエンスを促進します。BSIは国内唯一のバイオデータサイエンティストの集積地であり、この人材によりAIと生命科学の融合を進めるとともにバイオデータサイエンティストの育成も行います。



図: BSIと機関外ネットワーク

ライフサイエンス統合データベースセンター

データベース統合化を通じた生命科学分野のオープンサイエンスの推進

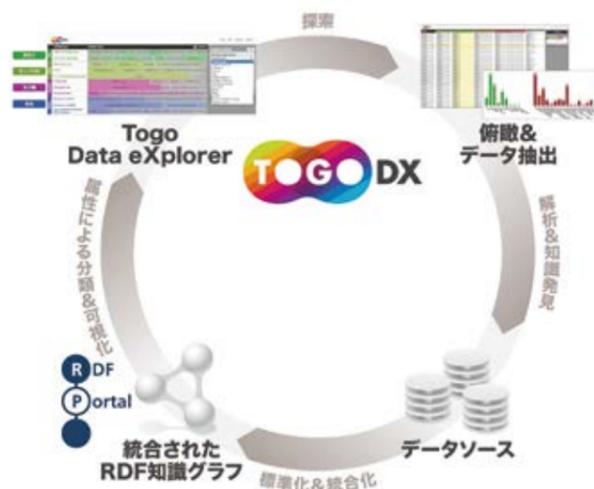
当センターは生命科学分野のオープンサイエンスを目指し、全国の大学、研究機関などが所有・生産する多様かつ急速に増加するデータベース(DB)を一元的に利活用するための「データベース統合化」に関する研究開発を行っています。データの記述に用いる用語とその分類体系(オントロジー)を標準化した知識グラフ(Knowledge graph)の構築、構築された知識グラフから必要な情報を効果的に抽出するための技術開発、文献情報の有効利用などに力を入れています。



小原 雄治 センター長

データベースの統合的利用環境の構築

▶生命科学DBを統合的に利用できる知識グラフ基盤を構築するために、各種DBの Resource Description Framework (RDF) 化支援とその集積を行い、RDF Portalから提供しています。また、知識グラフ基盤を利活用するための様々なアプリケーションを開発しています。例えば、ヒトに関する情報をワンストップで探索できるアプリケーションTogoDX/Humanは、研究対象を絞り込んで得られる結果を統合解析に応用可能です。また、実験結果の解釈や考察の一助として利用できます。今後はヒトに関するデータの充実とともに、TogoDXを他の生物種へ展開していきます。



▶様々なデータベースを統合した知識グラフとデータを統合的に探索するためのフレームワークTogoDXによるデータ解析プラットフォーム(ヒト関連データへの応用) TogoDX/human [https://togodx.dbcls.jp/human/]

統合的利用の実現のための国内外の研究機関との連携

▶知識グラフ基盤の構築と利活用に必要な対外連携を深めるために、約1週間の合宿形式で共同開発作業を集中的に行う国際版BioHackathonを始め、各種のハッカソンを10年以上開催しています。これらにより、データ相互利用のためのルール、仕組みやツールができ、国内外の機関で採用され、国際的な標準化を進めています。2022~2024年度には科学技術振興機構NBDC事業推進室が運用していたRDF Portal、日本人ゲノム多様性データベースTogoVar、NBDCヒトデータベースについて運用主体を引き継ぎました。日本人のゲノムや疾患を解析する研究機関や環境微生物を扱う研究機関などとも連携しています。



▶2024年8月開催の国際版バイオハッカソン集合写真@福島・土湯温泉



Address : 〒277-0871 千葉県柏市若葉178-4-4
東京大学柏の葉キャンパス駅前サテライト 6階
URL: https://dbcls.rois.ac.jp/



極域環境データサイエンスセンター

南北両極域から得られた様々な科学データの公開と共同利用、データサイエンスを推進し地球環境研究に貢献することを目指しています。

当センターは、南北両極域での観測・研究活動により取得された、多種多様な科学データの公開と共同利用を促進し、より多くの研究成果を創出するために、大学等外部コミュニティに対してデータサイエンス面からの支援を行うことを目的としています。また、国際的には極域観測・研究のナショナルデータセンターとしての役割も果たし、極域科学におけるデータ活動の中核を担うとともに、データに基づく新しい極域科学を創出し、地球環境研究に貢献することを目指しています。



門倉 昭 センター長

当センターが扱うデータ

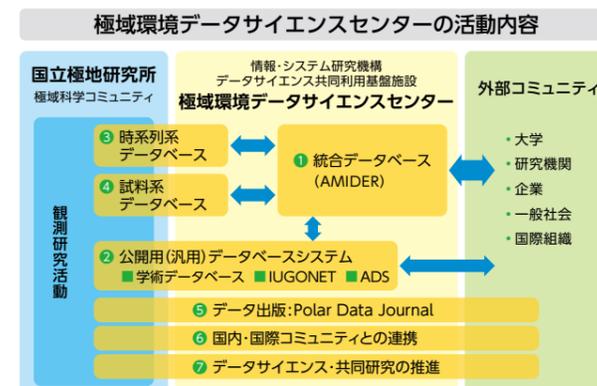
▶当センターが取り扱うデータは、極域科学の全ての分野のデータになります。両極域では、超高層、大気、海洋、雪氷、地学、生物など様々な分野の観測・研究活動が行われていて、様々な媒体に記録されたデジタルデータや、採取されて保管された試料系データなど、多種多様なデータが取得されています。それらのデータは、取得後に、様々な処理、解析、分析をされ、物理的に意味のあるデータとなった後に、科学的な成果が生み出されます。信頼される科学的成果を生み出すためには、データが確実に保管されていて、失われたり、劣化したり、改ざんされたりしないこと、そのデータが誰にでも利用可能で、同じ科学的成果の再現性が保たれること、など、データの信頼性の確保が求められます。



▶極域環境データサイエンスセンターでは、南北両極域での科学観測・研究活動によって取得された全ての分野のデータを扱います。

当センターの活動内容

▶また一方で、地球環境変動のような研究では、多分野の多種多様なデータを同時に用いることによって全く新しい成果が生み出される、ということもあります。その場合は、様々なデータの所在情報、属性情報などのメタ情報(メタデータ)を統一的に扱う必要があります。また、ある分野のデータが、予想も出来ない分野に応用され、予想もされない新しい成果や価値が生み出される、ということもあります。そのためには、そのデータの公開性や所在の分かり易さが重要になります。当センターでは、こうした、極域科学データの、処理、解析、保管、共有、公開、共同利用、についての活動支援を行っています。



▶極域環境データサイエンスセンターは、極域科学コミュニティと外部コミュニティとの間の様々なデータ活動の橋渡しをします。
<http://pedsc.rois.ac.jp/ja/activity>



Address : 〒190-0014 東京都立川市緑町10-3 データサイエンス棟 1階
URL: http://pedsc.rois.ac.jp/ja/



社会データ構造化センター

社会を対象として得られる様々なデータの整備と 共同利用を通じて社会的課題の解決に貢献

当センターは、社会で生じる様々な現象を測定・計測することで得られる「社会データ」—例えば個人や組織を対象とする社会調査のデータ、官庁等の公的セクションが実施する統計調査に基づく公的マイクロデータ、様々な機器を通じて社会活動をリアルタイムで計測するソーシャルビッグデータ、等—を整備し、様々なステークホルダーに提供することを通じて、社会的課題の解決につながる実証的学問を促進すること、社会データの整備・活用に資する基盤技術を開発することを目標に活動しています。



前田 忠彦 センター長

当センターは、データが得られる分野に対応して、主に3つのグループに分かれて、次のプロジェクト・事業を展開しています。

▶ 社会調査関連事業

全国共同調査ネットワーク形成によるデータ収集、及び社会調査データの整備と公開を進めます。統計数理研究所から継承した大規模学術調査データの整備と公開、他機関の研究者と共同で実施する調査の企画やデータ共有、社会調査の実施に伴うコンプライアンスに関わる課題の研究と普及などのプロジェクトを推進しています。(図1参照)

▶ 公的マイクロデータ事業

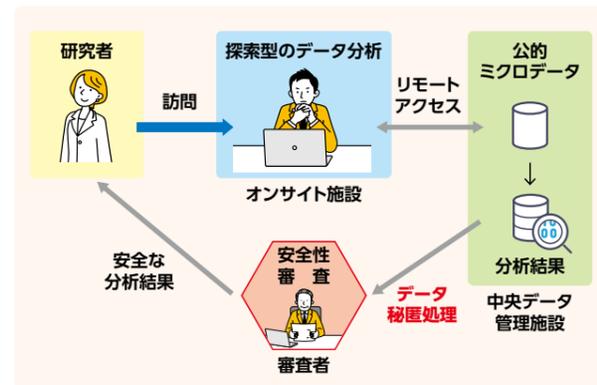
公的統計データの整備と共同利用システムの発展、及びオンラインのデータ解析システムの研究開発、オンサイト施設の運営などを担当します(図2参照)。公的マイクロデータの安全な公開に関わるマイクロデータ・セキュリティ、公的統計二次的利用の推進、経済・金融分野でのリスク管理に関わる高度信用リスク、政府統計を用いたEBPM(証拠に基づく政策決定)、等の研究プロジェクトを推進しています。公的統計マイクロデータ研究コンソーシアムの事務局機能も担っています。

▶ ソーシャルビッグデータ事業

異なる組織の研究者間によるソーシャルビッグデータを用いる研究活動で必要となる共同利用データの管理方法ならびにプラットフォームの整備を進めています。また、ソーシャルビッグデータを用いた共同研究も実施します。実社会データ共有基盤の開発を目指すプロジェクトでは、道路や交通などの社会インフラで、常に変動する実社会の状況の効率的な収集、状況把握、分析を可能にするデータ共有基盤システムを、自治体等と連携して実証的に開発します。



▶ 図1: データライフサイクル



▶ 図2: オンサイト解析のプロセス (施設利用についてはP.14を参照ください)



人文学オープンデータ共同利用センター

人文学におけるオープンサイエンスとデジタル変革の推進： データ駆動型人文学と人文学ビッグデータの展開

人文学オープンデータ共同利用センター(CODH)は、人文学分野におけるオープンサイエンスとデジタル変革の推進を目指しています。情報学・統計学における最新のデータ駆動型技術の導入により人文学の研究方法を革新する「データ駆動型人文学」や、人文学分野で生み出されたビッグデータを他分野の研究に活用する「人文学ビッグデータ」の研究を通して、人文学におけるAI活用や歴史ビッグデータに基づく分野横断型歴史研究などを推進し、その成果として様々な研究資源を公開しています。



北本 朝展 センター長

AIくずし字認識アプリ「みを」

スマホでくずし字画像を撮影すると、AIが数秒で現代日本語文字に変換して表示するアプリを、iOSおよびAndroidで無料公開しています。



IIIF Curation Platform

IIIFの世界に、キュレーションという新しいコンセプトを導入し、利用者主導型のIIIFプラットフォームを実現します



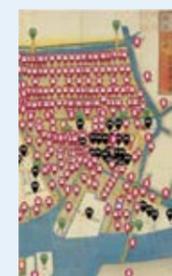
顔貌コレクション(顔コレ)

IIIF Curation Viewerを用いて、美術作品に出現する顔の部分を取り取って収集し、美術史研究に活用します。



江戸ビッグデータ

都市「江戸」に関する地理情報と紐づく、商業や観光・人物などの歴史ビッグデータを統合し、データを基に過去の世界を分析します。



武鑑全集

江戸時代の200年続いたベストセラー「武鑑」を網羅的に解析し、大名や幕府役人に関する中核の情報プラットフォームを構築します。



れきすけ

歴史資料を利用した研究のために、歴史資料に関する知識や経験を、様々な分野の研究者で共有します。



ゲノムデータ解析支援センター

大量のゲノム・トランスクリプトームデータから生物学的に重要な情報を抽出するための情報科学的解析を支援します。

次世代シーケンシング (NGS) 技術の発展に伴い、さまざまな生命科学の分野でNGSを用いた全ゲノムレベルの解析が広く行われるようになってきました。しかし、NGSデータは膨大な量の断片配列データであり、解析には生物学の知識に加えてバイオインフォマティクスの知識と技術が不可欠です。ゲノムデータ解析支援センターでは、大量のゲノムデータを高精度に解析するための情報科学技術の研究開発や、実データの解析支援、またそのための人材の育成を行なっています。



野口 秀樹 センター長

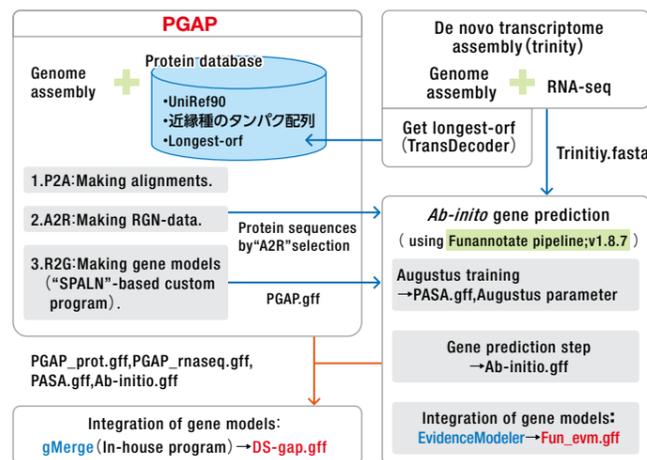
▶ゲノムデータ解析支援センターでは、大学・研究機関等の研究者を対象にさまざまな種類のゲノムデータ解析支援を行なっています(図1)。

<p>• De novo ゲノムシーケンス</p> <p>De novo ゲノムアセンブリ 参照配列のない新規生物種のゲノムを、NGSデータを用いて構築します。</p> <p>ゲノムアノテーション ゲノム配列上の遺伝子の位置やエクソン-イントロン構造を同定し、注釈付けします。</p>	<p>• トランスクリプトーム解析</p> <p>遺伝子構造・発現解析 RNA-seqデータの <i>de novo</i> アセンブル・マッピングを通して、遺伝子構造同定や発現量解析を行います。</p> <p>non-coding RNA 解析 RNA2次構造予測、miRNAのターゲット検索など。</p>
<p>• ゲノムリシーケンス</p> <p>全ゲノムリシーケンス 全ゲノムの配列リードを参照ゲノム配列と比較し、SNVや構造多型を検出します。</p> <p>ターゲットゲノムリシーケンス・エピジェネティクス解析 エキソーム、RAD-seq、ChIP-seq、HiC-seq 等々。</p>	<p>• メタゲノム解析</p> <p>メタゲノムアセンブリ メタゲノム配列リードを <i>de novo</i> でアセンブルします。</p> <p>種分類、遺伝子予測 メタゲノム配列のクラスタリング、遺伝子予測、パスウェイ解析など。</p>

▶図1: 当センターで行なっているゲノムデータ解析

▶取り扱うゲノムデータは主にNGSの配列データですが、研究の目的や実験条件などは研究ごとに大きく異なります。また、対象の生物種も哺乳類やその他脊椎動物を中心に、昆虫、植物、真菌、原核生物などさまざままで、ゲノムサイズや構造、進化的背景などに応じて適切な解析手法を選択する必要があります。当センターでは豊富な解析経験を活かして研究目的に応じた柔軟で高精度な解析支援を行なっています。

▶また、解析支援を円滑に実施するためにゲノムアノテーションパイプライン(図2)やゲノム再シーケンスパイプラインなどの各種解析パイプラインの開発を行なっているほか、遺伝子予測手法やRNA-seqアセンブラ、メタゲノムの種分類手法などの新規解析手法の開発も行ない、最先端の解析手法を提供できる環境を整えています。



▶図2: ゲノムアノテーションパイプライン

データ同化研究支援センター

シミュレーションと観測データの統合による問題解決: 合わないシミュレーションをまだ続けますか?

データ同化とは、観測データと数値シミュレーションを統合する方法です。データ同化により、高精度の予測が可能なシミュレーションである「データ同化システム」や、計算時間を大幅に短縮できるシミュレーションである「エミュレータ」を開発できます。データ同化研究支援センターは、諸科学・産業分野でシミュレーションの活用に関する課題をお持ちの方々に対し、経験豊富な研究者がデータとシミュレーションを融合させる諸技術のノウハウ等を提供し、課題解決を支援します。



上野 玄太 センター長

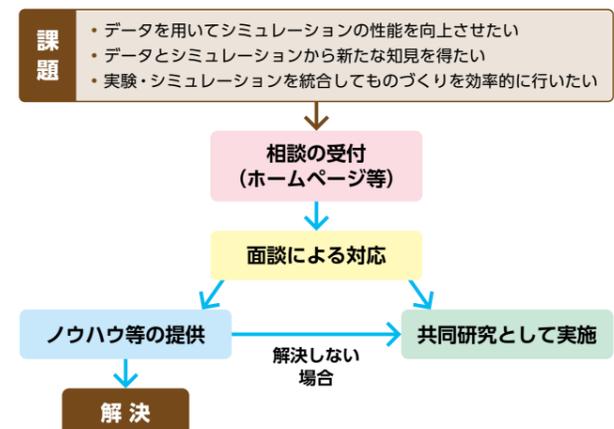
当センターの詳細

▶当センターは、データ同化研究の相談窓口を開いており、面談で助言や技術指導を提供し、問題解決の支援を行っています。統計科学を基盤とするデータ同化の方法から、データ同化を応用する現場の観点からの相談対応、ならびに共同研究が可能です。データ同化を新しく導入したいがどうすればよいのか、データ同化の計算を完了したがこの先どうすればよいのか等、ご相談をお待ちしております。

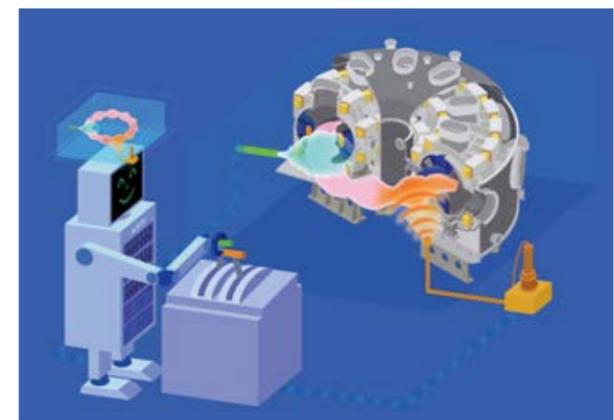
研究内容の紹介

- ▶データ同化システムの開発と方法
- ▶データ同化を活用した細胞・発生生物学におけるモデル評価とパラメータ推定
- ▶超高層物理学におけるデータ同化や関連統計手法の応用研究
- ▶磁気圏-電離圏系のデータ同化
- ▶津波予報の向上を目指した津波データ同化の研究
- ▶時空間回帰と都市・地域問題への応用

当センターのメンバーは、情報・システム研究機構統計数理研究所、国立遺伝学研究所、国立極地研究所の教員であり、統計科学を基盤とするデータ同化の方法から、データ同化を応用する現場の観点からの相談対応、ならびに共同研究が可能です。



▶研究相談は、ホームページをご覧のうえメールにてお申し込みください。



▶データ同化により計算機上に再現した仮想プラズマを通して、現実のプラズマを制御するデジタルツイン制御のイメージ図。



人工知能法学研究センター

司法のIT化とAI時代の到来：
技術と法律を融合させた新たな学問分野
－人工知能法学－の確立が急務

研究の背景

▶ Chat-GPTをはじめとする大規模データを使う人工知能技術が知的作業に導入され始めています。海外では大手弁護士事務所がGPTを活用が進みつつありますが、日本では2022年に民事訴訟法が改正され、IT化が推進され始めたばかりです。世界の技術レベルに追いつくため、早急な対応が必要です。一方、機械学習の普及に伴い、個人情報保護や著作権侵害などの法的問題が生じています。信頼できるAIを作るには、法的規範を理解する技術者やAI技術を理解する法曹、理想的には両方を理解する人材の育成が不可欠です。そのため、AIと法律を融合させた新しい学問分野の確立が必要です。



佐藤 健 センター長

人工知能法学の創成

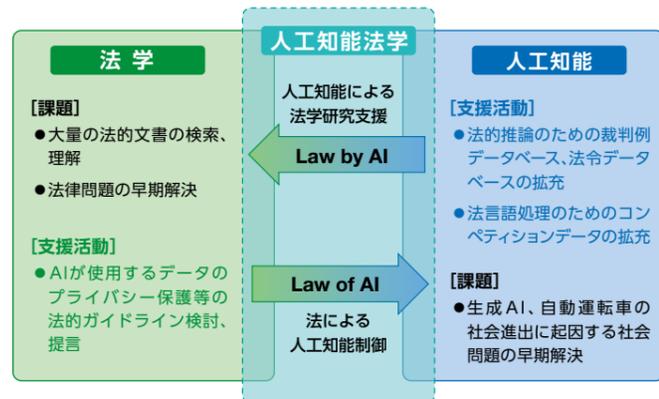
▶ このため「人工知能による法学研究支援」と「法による人工知能制御」の双方を実施し、大量な法的文書の処理を可能とするとともに、法学における法律問題および生成AIや自動運転車の社会進出等に起因して発生する社会問題の早期解決を図る「人工知能法学」の研究分野を創成します。

(1) AIによる法律の支援に係る研究(LAW by AI (LAW supported by AI))

法学研究データの収集・管理、当該データ活用方法の研究を実施するとともに、法的推論のための裁判例データベースや法令データベース等を拡充します。さらに、大量な法的文書の検索を可能にし、法学における法律問題の解決を支援します。各国の法令や条例(原文、論理式)、判例要約、判決理由生成等のタスクのコンペティションを支援するために、データの収集・運用や、データの応用研究により、各タスクに必要な技術開発を行なうとともに、広範囲な分野からの参画を促し、人工知能法学研究コミュニティを新たに構築します。

(2) 人工知能をどうやって
法により統御するかというテーマ
(LAW of AI (LAW control of AI))

法学者と共同で、データサイエンスおよび人工知能における法律問題の対処方法について研究し、AIが使用するデータのプライバシー保護等の法的ガイドラインを検討、提言する。さらに、生成AI、自動運転車の社会進出に起因する社会問題の早期解決を図ります。また、法規範や倫理規範を考慮した技術を開発し、データサイエンティスト教育を支援します。



データレイク研究開発センター

データ駆動社会における新たな情報基盤の創出：
進化する人工知能(AI)モデルと増大する多様なコンテンツを一つの
情報基盤として利用可能なフレームワークの確立と運用

センターの背景と目的

▶ 近年の人工知能(AI)技術は数年で大きく進展し、chatGPT等の生成系AIのサービスが提供され、現代社会の在り方に多大な影響を与えています。大量・多様なデータを学習するAIモデルでは、利用する立場からの分かりやすさ、透明性、安心・安全の担保が求められます。当センターでは、進化するAIモデルとデータを一つの情報基盤として管理し、モデルおよびデータの来歴を明示し、最新のコンテンツ参照が容易なフレームワークを構築します。



中野 美由紀 センター長

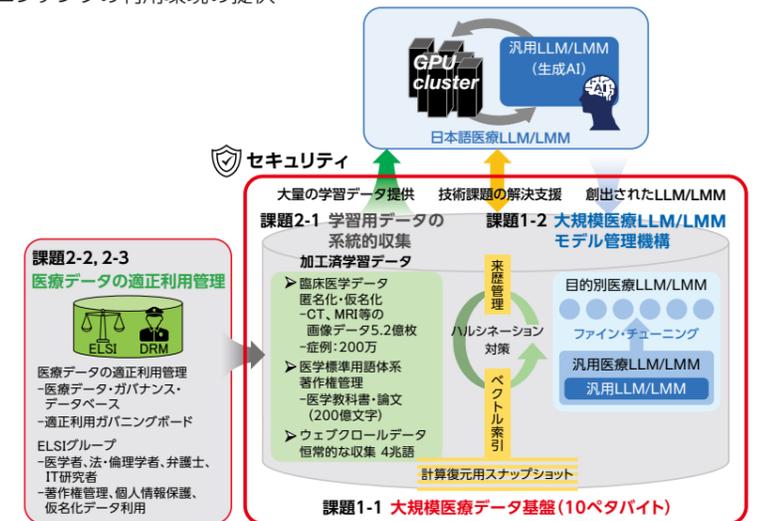
(1) データレイク情報基盤フレームワークとは

データレイクでは、構造化データ、非構造化データを問わず、多種多様なデータを一元的に参照、利用が可能とし、コンテンツの収集、管理を容易にします。当センターでは、さらに、データから学習されたAIモデルも、コンテンツと同様に管理し、参照可能とします。コンテンツとモデルの一元的な管理環境を提供することで、データの利活用とAI技術の研究、開発を共に促進するフレームワークを構築し、我が国に将来の情報基盤をとして、以下のような機能を実現します。

- －コンテンツの継続的蓄積とコンテンツの来歴管理、ベクトル索引技術の確立
- －モデルの持続的管理とモデルの学習コンテンツ情報、学習過程も含めた来歴情報の管理等モデル運用時に関する先端的技術の提供
- －大学共同利用機関として、セキュリティ技術に支えられた堅牢なシステムを構築し、産学官における様々な最先端研究分野へ安心・安全な利用環境の提供
- －データ利活用における社会的な課題への対応として、個人情報保護法、著作権法等の法律課題、生命に関連した社会倫理課題を洗い出し、安全、安心な人工知能技術およびコンテンツの利用環境の提供

(2) 医療用LLM/LMMの研究開発を促進する
医療データ基盤

我が国では医療分野の研究促進に加え、適正な医療技術の提供等、学術的・社会的に、医療データの共有、生成系AIの利活用が求められています。当センターでは、データレイク情報基盤を具体化し、SIP「統合型ヘルスケアの研究開発」2023年度補正予算により、医療データ基盤(図1)の研究開発に着手しています。「10ペタ規模の大規模利用データ基盤」を構築し、「大規模医療LLM/LMMモデル管理機構」および医療DBを構築する。また、医療データも含めた「学習用データの系統的収集を行う」と共に、医療データの適正利用管理を行います。



▶ 図1 医療用LLM/LMMの研究開発を促進する医療データ基盤



公募型共同研究「ROIS-DS-JOINT」

DS施設では、全国の研究者等に共同利用／共同研究の機会を提供するため、データサイエンスに関する公募型共同研究プログラム「ROIS-DS-JOINT」を毎年募集しています。このプログラムには、DS施設各センターの研究者および研究リソースを活用して共同研究を行う「一般共同研究」と、各センターを中心に研究交流、講習会等を行う「共同研究集会」の2種類があります。

一般共同研究

データサイエンスに関する特定の課題について、DS施設に所属する研究者及びDS施設のリソース等を活用し、少人数で行う研究を対象とします。

- 代表者は当機構以外の教育・研究機関の研究者（図書館司書や博物館等の学芸員、研究員を含む）
- 共同研究経費：最大100万円
- 研究の実施に必要な旅費、役務費、物品等を計上可能（機関が通常備えるべき物品等は対象外です）

共同研究集会

共同研究に向けた討論やデータサイエンスに関する研究交流、講習等を比較的少人数で行う集会を対象とします。

- 代表者は当機構以外の教育・研究機関の所属者（研究・開発、調査等を主たる業務としない方も申請が可能です）
- 共同研究経費：最大50万円
- 共同研究集会に参加するために必要な旅費および開催に要する経費を計上可能

応募に関する注意点

- 申請に先立ち、**DS施設の教員との十分な事前相談をお願いいたします**
- 共同研究経費はDS施設で執行します。**代表者に直接研究費を配分するものではありません
- DS施設はデータの共有・解析支援を目的としているため、データ生産（調査等を含む）のみを目的とする課題は本公募の対象外です
- その他、詳細な募集情報については、**ROIS-DSホームページ** (<https://ds.rois.ac.jp/>) をご覧ください

共同研究参加者の所属機関件数および共同研究参加者数の推移

2017年度の募集開始から現在に至るまで、共同研究参加者の人数・所属機関数ともに増加傾向であり、2023年度には共同研究参加機関数が100件を超えました。大学・高等専門学校や国立研究機関をはじめと様々な機関の研究者等から研究課題の応募をいただいております。アジア諸国を中心とした海外の大学・研究機関からの研究参加者も増加しています。

公募型共同研究「ROIS-DS-JOINT」参加機関数・共同研究参加者数の推移（一般共同研究）



近年の共同研究参加状況（一般共同研究）

研究代表者・共同研究者の所属機関件数	2022	2023	2024
大学・高等専門学校	53	58	61
国立研究開発法人・省庁系研究所等	19	21	22
民間企業・NPO・自治体等	4	12	10
海外大学・研究機関	8	15	16

DS推進室の活動について

■ 成果報告会・国際シンポジウム

各センターの活動報告・公募型共同研究の成果発信として開催しています。4回目となる2023年度についてもオンラインで開催しました（当日の講演動画はDS施設ウェブサイトで公開）。第1部の基調講演・センター活動報告、第2部ではビデオチャットツールを使用したポスター発表の場を設け、共同研究の成果について活発な議論が行われました。2023年12月には「データサイエンス国際シンポジウム 2023 - アジア・オセアニア地域におけるオープンデータ協力体制の構築」を開催しました。



■ 研究支援・コーディネート

当施設の研究コーディネーターが中心となり、学会等での広報活動、問合せ対応、共同研究のスタートアップ支援等を行っています。これまでに生物学、医学、薬学、工学、農学、環境学、地球惑星科学、統計学、金融工学など、30以上の学会でのブース出展を通じ、幅広い研究分野のデータサイエンス推進を支援しています。

- | | | |
|------------|----------------|------------|
| 日本分子生物学会 | 日本人類遺伝学会 | 日本農芸化学会 |
| 日本進化学会 | 日本薬学会 | 日本生態学会 |
| 日本癌学会 | 日本生物工学会 | 日本地球惑星科学連合 |
| 日本アクチュアリー会 | 日本保険・年金リスク学会、他 | |
- 過去の実績はこちら → <https://ds.rois.ac.jp/information/exhibitions/>

情報発信

■ ハンズオン(体験的)講習会

統合データベース講習会：AJACS (JST 主催DBCLS共催)、RDF講習会 (DBCLS主催)、CODHチュートリアル (CODH主催)、データ同化ハンズオン (CARA主催、写真参照) 等の各種ハンズオンを随時開催しています。またデータ比較のプログラムを含んだIUGONET講習会 (PEDSC主催) を、国内・海外で実施しています。



■ 出版物

DS施設のこれまでの歩みをまとめた記録冊子「ROIS-DS データサイエンスへの歩み」、DS施設講演集「データサイエンスという未来」を発行しています。

- ROIS-DS データサイエンスへの歩み
- DS施設パンフレット
- DS施設講演集「データサイエンスという未来」



施設利用

■ オンサイト施設

データサイエンス共同利用基盤施設 (ROIS-DS) では、社会データ構造化センター (CSDS) が中心となり、調査票情報や秘匿データなどを安全に利用するために入退管理システム・監視カメラ等の高度なセキュリティ環境を整えたオンサイト利用施設を整備し、研究利用に提供しています。



<開室日時> 平日 10時～17時

- ※ 土曜日、日曜日、国民の祝日、12月29日～1月3日、及び、予約状況カレンダーにて休業告知されている休業日を除く
- ※ 休業日の利用、時間外利用については応相談

〒190-8562 東京都立川市緑町10-3
 大学共同利用機関法人 情報・システム研究機構
 データサイエンス棟 2階
 メールアドレス: tokumeid(at)ism.ac.jp