

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)

ROIS-DS

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 (ROIS-DS)

〒190-0014 東京都立川市緑町10-3 データサイエンス棟
<https://ds.rois.ac.jp/>
本パンフレットに関するお問合せ先：042-512-9254

大学共同利用機関法人 情報・システム研究機構 (ROIS)

〒105-0001 東京都港区虎ノ門4-3-13 ヒューリック神谷町ビル2階
TEL 03-6402-6200 (代表) <http://www.rois.ac.jp/>

— 連携、協働、そして発展へ —

データ駆動型研究で研究者を支援する 分野融合的な共同利用・共同研究拠点です。



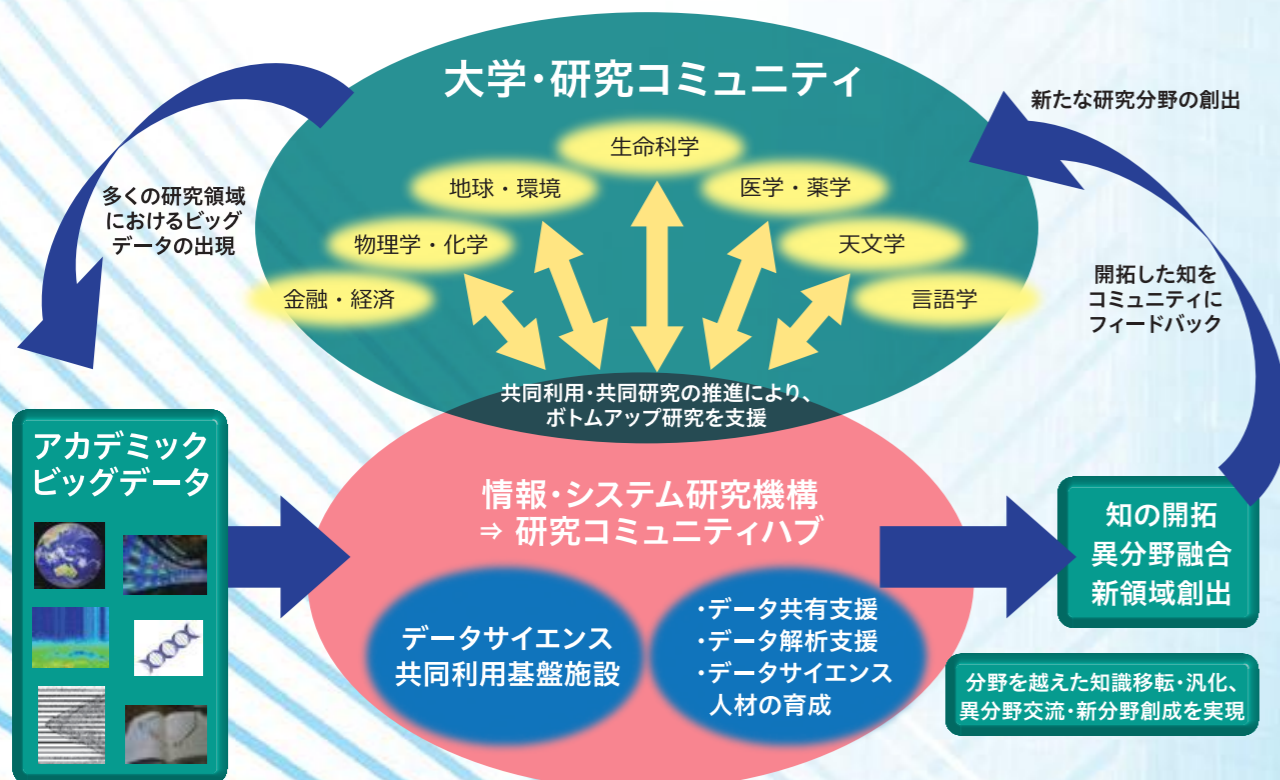
データサイエンス
共同利用基盤施設
施設長 藤山秋佐夫

データサイエンスの推進拠点

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(DS施設)」は、大規模データの高度な解析により科学や社会の課題を解決する「データサイエンス」を全国規模で融合的に推進するための共同利用・共同研究拠点です。データサイエンス(データ駆動型研究)を合い言葉に大学等との連携・協働を強化する目的で、2016年度に情報・システム研究機構により設置されました。2019年現在、ライフサイエンス統合データベースセンター(DBCLS)、極域環境データサイエンスセンター(PEDSC)、社会データ構造化センター(CSDS)、人文学オープンデータ共同利用センター(CODH)、ゲノムデータ解析支援センター(CGI)、データ同化研究支援センター(CARA)の計6センターで構成され、大学等の研究力強化に貢献しています。当施設が扱うデータは、ゲノムや遺伝子に関する大量の生命情報データや大気レーダデータ等による観測データから、古典籍や社会調査、公的ミクロデータまで極めて広範囲に及ぶため、情報・システム研究機構を構成する4つの研究所——国立極地研究所、国立情報学研究所、統計数理研究所、国立遺伝学研究所——や、他の大学共同利用機関法人の研究所等と協力して活動しています。

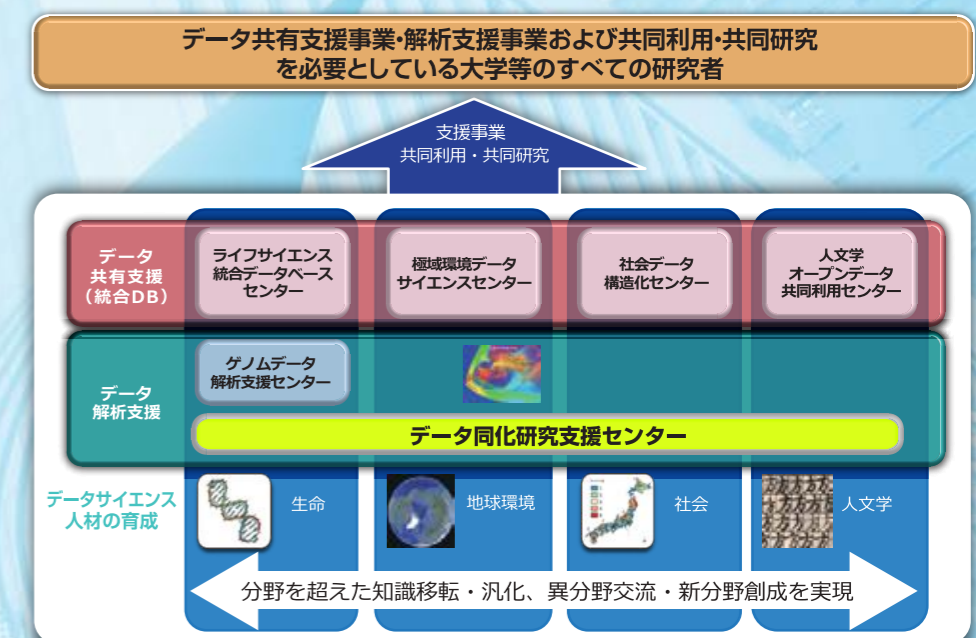


データサイエンス棟(立川キャンパス)

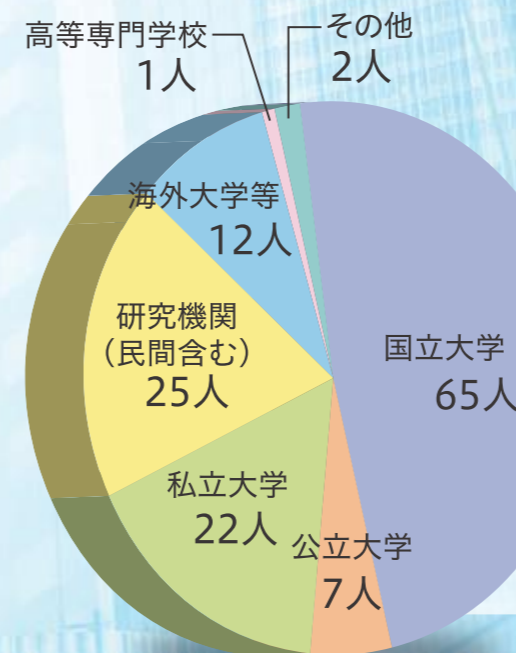


データ共有・データ解析支援・人材育成の3つの活動で研究を支援

DS施設では、生命科学、環境、統計数理から人文・社会科学に至る広範な分野をカバーする6つの研究センターが、データサイエンスの基盤となるデータ共有支援、データ解析支援、人材育成を3本柱に研究・支援活動を行っています。学術や社会のさまざまな分野でコミュニティと連携しながら大学等の機能強化に施設全体として取り組み、学術の発展と社会のイノベーションに幅広く貢献してまいります。



公募型共同研究



DS施設では、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」を実施しています。このプログラムには、DS施設各センターの研究者と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、2019年度にはそれぞれ37件、6件の計43件の課題を採択・実施しました。この公募型共同研究プログラムに限らずDS施設各センターでの相談は随時受け付けており、全国の研究者の皆様様に支援や共同利用・共同研究の機会を広く提供できるよう努めています。DS施設の活動内容については、ホームページや研究コーディネーターの活動、シンポジウム等を通じてお知らせしています。

公募型一般共同研究 参加者所属機関内訳(2019年度)
公募型一般共同研究には、国公立大学、研究機関(民間含む)、高等専門学校、海外の大学など計75機関、134名(当機構内を除く)が参加し、研究を推進しました。

ライフサイエンス統合データベースセンター (DBCLS)

本センターは、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、研究機関などが保有・生産する多様かつ急速に増加するデータベースを一元的に活用するための「データベース統合化」に関する研究開発を行っています。用語とその分類体系(オントロジー)や、データ記述方式の標準化に基づいた連結化データである「Linked Data」の整備など、統合化に必要な技術開発に力を入れています。また、世界各国のデータベース開発機関からエキスパートを招き、バイオハッカソン*などの開発者会議を毎年開催し、統合化のための技術開発と標準化を国際的に主導しています。(センター長:小原 雄治)

センターの主な活動:

- ▶ ウェブ上に分散しているデータベース (DB) を統合的に利用できる環境を実現するために、生命科学DBのResource Description Framework (RDF) 化を推進しています。RDF化支援と各機関からのRDFデータの受け入れを継続的に実施しており、2019年度には製品評価技術基盤機構 (NITE) と理研からそれぞれ登録された微生物データを統合検索できるサイトがNITEから公開されました。前年度に公開した日本人ゲノム配列のバリエーションと関連疾患情報などのDB「TogoVar」ではより多くの情報が見やすくなるよう、インタフェースを改良しました。
- ▶ 統合的利用の実現には、国内外のDB関係者との連携が必要です。そのために共同開発作業を集中的に行う国際開発者会議「BioHackathon」を開催しています。12回目となる「BioHackathon 2019」には13か国・約120名が福岡に参集しました。そのほか、国内版バイオハッカソン、SPARQLthon、BLAH (Biomedical Linked Annotation Hackathon) 他を通じて、DBの標準化を促進しています。



TogoVarトップページ



国際開発者会議「BioHackathon2019」集合写真 *科学技術振興機構バイオサイエンスデータベースセンター(NBDC)と共催

極域環境データサイエンスセンター (PEDSC)

本センターは、データに基づく新しい極域科学の創生と地球環境研究への貢献を目指しています。そのため南極域、北極域を対象とする調査・観測活動で取得された貴重なデータの公開と共同利用を進め、地球環境分野でのデータサイエンスの推進に貢献します。従来の調査・観測活動により多分野から多種多様な研究・観測データが得られていますが、データセットごとにデータベース化の程度と公開状況に差があるため、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)を統一的に整備しつつ、アーカイブ化とデータベース化を推進しています。(センター長:門倉 昭)

2019年度の主な成果:

- ▶ 北極・南極データアーカイブシステム (AADS) や学術データベース、IUGONETシステムによるメタデータ登録やデータ公開、データへのDOI付与などを継続して進めました。
- ▶ DS施設の公募型共同研究に対応し、データ公開支援を行いました。

共同研究成果:昭和基地宇宙線観測データ公開用WEBシステム作成

▶ 各分野のデータ・資料のアーカイブ・公開を進めました。



北極・南極データアーカイブシステム (AADS)



学術データベース



オーロラ地上多点観測データ表示用WEBサイト作成



極域歴史的資料デジタルアーカイブシステム構築



一般向け極域データ検索サイト公開

社会データ構造化センター (CSDS)

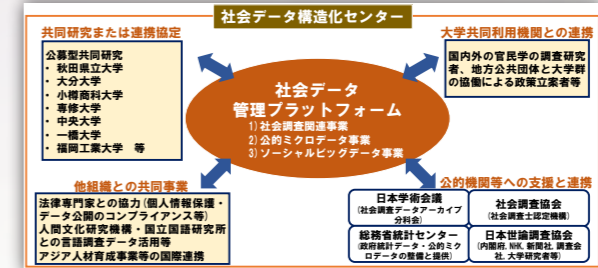
構造化され、整備された社会データは、複雑な現代社会を解明し、地球環境、治安、経済を含むさまざまな課題解決に役立ちます。本センターは、日本学術会議、総務省統計センターをはじめ、国内外の調査機関、研究機関等と連携し、社会データ管理プラットフォームの構築を目指しています。具体的に日本人の国民性調査などの「社会調査データ」、オンサイトで利用できる「公的統計のマイクロデータ」、人の社会行動をリアルタイムで表す「ソーシャルビッグデータ」の整備・向上に取り組み、データに基づく人文・社会科学と政策立案の進捗を支援しています。(センター長:前田 忠彦)

2019年度の主な成果:

- ▶ アジア・太平洋価値観国際比較調査等の詳細な集計結果を公開しています。「日本人の国民性調査」等の国内調査については、共同研究の形で個票データの共同利用を行い、論文等の成果を発表しています。
- ▶ 政府・地方自治体の個票レベルの公的マイクロデータが利用できる「オンサイト室」をDS施設内で運営しています。
- ▶ センターが目指す「社会データ管理プラットフォーム」について試験運用を行いながら機能の整理を行い、プラットフォームを通じて発信するコンテンツの蓄積に努めました。
- ▶ 調査データの取得から公開に至る様々な段階におけるコンプライアンスについて、弁護士に法的立場から論じていただくチュートリアル・セミナーを開催しました。



アジア・太平洋価値観国際比較調査



社会データ管理プラットフォーム

人文学オープンデータ共同利用センター (CODH)

本センターは、人文学分野におけるデータのオープン化と共同利用に基づく研究・支援活動を推進するため、大規模に電子化されつつある日本の文化、例えばくずし字で書かれた江戸時代の古典籍や活字印刷された明治以降の書籍などを対象に、情報学・統計学の最新技術を活用した自動テキスト化や内容分析などを行っています。しかし人文学研究分野におけるデータサイエンスの発展に不可欠な、データのオープン化はまだ十分に進んでいません。そこで、世界の人文知を共有する情報基盤を構築するとともに、異分野の研究者、市民、企業等とコラボレーションしながら、オープン化の推進を図っています。(センター長:北本 朝展)

2019年度の主な成果:

- ▶ 深層学習(ディープラーニング)による物体検出アルゴリズムを改良し、一ページの画像に含まれるすべてのくずし字を1秒程度で認識し現代の文字に変換するシステム「KuroNetくずし字認識」を開発しました。
- ▶ 日本美術の絵巻や絵本から顔貌を収集した「顔貌コレクション」を拡充し、美術史学と情報学の連携研究だけでなく、機械学習の訓練データやエンターテインメントへの応用など新しい活用法を開拓しました。
- ▶ 国際的な機械学習コンペ「Kaggleくずし字認識」を開催し、300人以上の参加者がくずし字認識アルゴリズムの精度を競いました。またそれに合わせて「くずし字データセット」を100万文字以上に増やすなど、各種データセットの拡充や新規公開などを進めました。



KuroNetくずし字認識によるくずし字認識の例。『百人一首万葉鑑』(国文学研究資料館蔵)の元画像と認識画像。



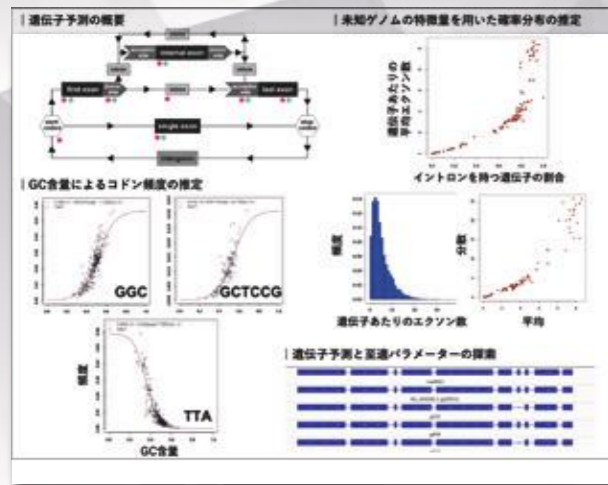
顔貌コレクションにおける『大黒舞』(国文学研究資料館蔵)の顔貌表現。収集した顔貌を検索し一覧表示可能。

ゲノムデータ解析支援センター (CGI)

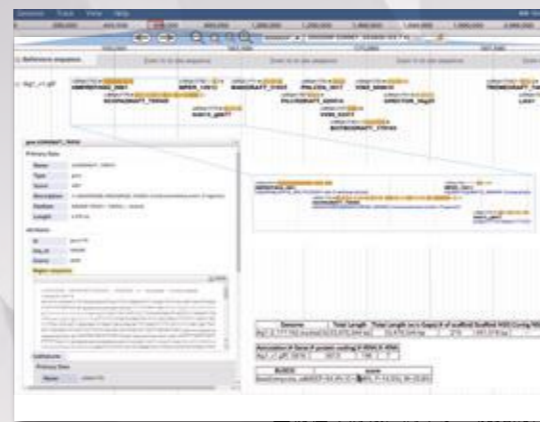
生命はゲノムDNAに書き込まれた情報を基に形作られています。そのため、ゲノムDNAに記録された情報の解読が生物学研究の出発点になっています。しかし、最新の技術から得られるゲノム情報データは、例えばヒトゲノム全体が30億文字あるのに対して、ひとつが300文字程度の短いものが数億本から成るデータセットとなっています。本センターは、ここから生物学、医学、環境学などさまざまな研究の発展につながる新たな知見を取り出すために、最先端のバイオインフォマティクス技術を開発・提供し、解析相談や共同研究を通じて研究者を支援します。(センター長:野口 英樹)

センターの主な活動:

- ▶さまざまな生物種(動物、植物、菌類、原核生物)を対象に、新規ゲノム決定や変異解析などのゲノム解析支援を実施しています。
- ▶ゲノム解析パイプラインやゲノムブラウザの整備、新規解析手法(遺伝子予測、RNA-seqアセンブラなど)の開発を行っています。



新規解析手法による真菌の遺伝子予測。未知ゲノムの特徴をあげ出す



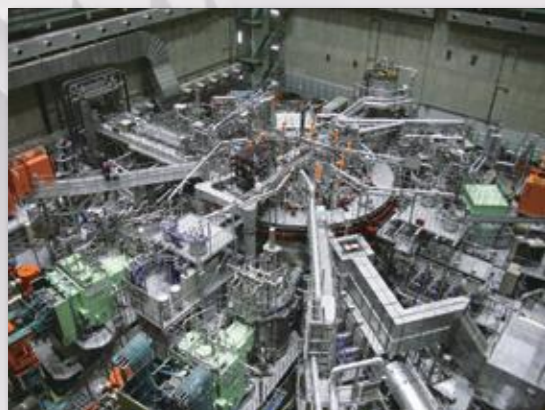
アクセスできるゲノムブラウザ

データ同化研究支援センター (CARA)

本センターでは、自然科学だけでなく、工学などの応用科学や社会科学等の広範囲にわたって、データとシミュレーションを統合する「データ同化」や、統計手法でシミュレーションを模倣する「エミュレータ」など、シミュレーションの可能性を広げる統計数理手法の研究開発を行っています。また、学術・産業界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、シミュレーションと統計数理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。(センター長:上野 玄太)

2019年度の主な成果:

- ▶地震の予測における時系列モデル(非定常更新過程モデル)の導入に関する助言と技術指導を行いました。
- ▶液膜流れのシミュレーションへのデータ同化手法の導入についての助言と技術指導を行いました。
- ▶大型ヘリカル装置(LHD)のデータ同化システムの開発に成功し、プラズマ・核融合学会において招待講演を行いました。
- ▶地理空間データの解析手法の高速化に成功し、統計ソフトウェアRのパッケージとしての整備・公開を行いました。



大型ヘリカル装置(LHD) 写真:核融合科学研究所提供



学会にブース出展し、データ解析手法相談を実施

研究支援コーディネーション

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



ハンズオン(体験的講習会)

統合データベース講習会(JST NBDC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、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。



センターウェブサイトURL一覧



DBCLSウェブサイト
<https://dbcls.rois.ac.jp/>



PEDSCウェブサイト
<http://pedsc.rois.ac.jp/>



CSDSウェブサイト
<http://csds.rois.ac.jp/>



CODHウェブサイト
<http://codh.rois.ac.jp/>



CGIウェブサイト
<https://genome-info.nig.ac.jp/>



CARAウェブサイト
<http://daweb.ism.ac.jp/cara/>

DBCLSサービス一覧ページ	https://dbcls.rois.ac.jp/services.html	オンサイト解析室	http://ds.rois.ac.jp/center3_micro/
TogoVar	https://togovar.biosciencedbc.jp/	公的統計マイクロデータ研究コンソーシアム	http://ds.rois.ac.jp/center3_micro/moc/
BioHackathonウェブサイト	http://www.biohackathon.org/	シーラカンスゲノムブラウザ	http://coelacanth.nig.ac.jp/
北極域データアーカイブ(ADS)	https://ads.nipr.ac.jp/	MetaGeneAnnotator(メタゲノムアノテーター)	http://metagene.nig.ac.jp/
IUGONET	http://www.iugonet.org/index.jsp?lang=ja	Platanus(ゲノムアセンブラ)	http://platanus.bio.titech.ac.jp/
極域データ一覧「学術データベース」	https://scidbase.nipr.ac.jp/	P3(Parallelized particle filter with Python)	http://daweb.ism.ac.jp/support/software/P-cubed/P-cubed.html
意識の国際比較調査	http://www.ism.ac.jp/~yoshino/	並列粒子フィルタ	
アジア・太平洋価値観国際比較調査	http://www.ism.ac.jp/~yoshino/ap2/index.html		