

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)



〒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



〒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。

データサイエンスの推進拠点

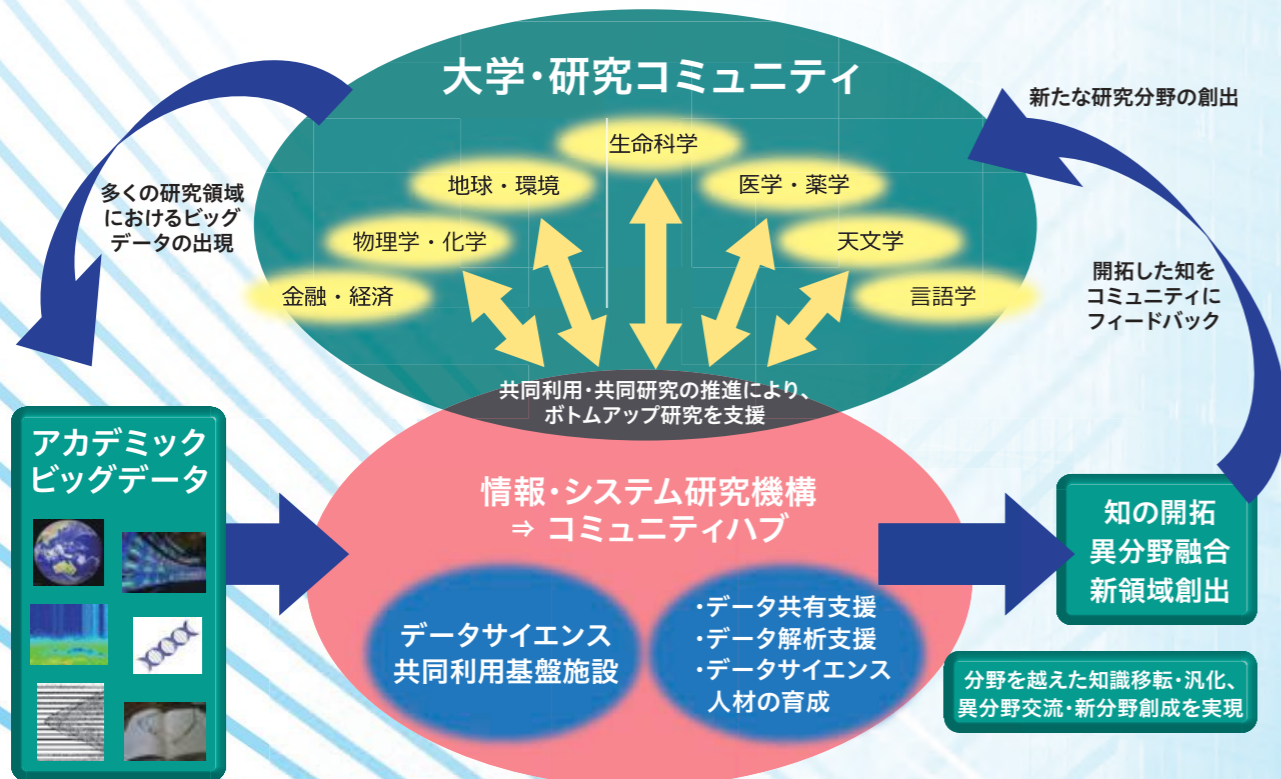
「データサイエンス共同利用基盤施設(DS施設)」は、大規模データの高度な解析により科学や社会の課題を解決する「データサイエンス」を全国規模で融合的に推進するための共同利用・共同研究拠点です。情報・システム研究機構は、データサイエンス(データ駆動型研究)を合い言葉に大学等との連携・協働の強化をする目的で、2016年度にDS施設を設置しました。データサイエンスといっても機構の各研究所が扱うデータは極めて多種多様で、DS施設各センターの活動進捗状況も対応する分野コミュニティの成熟度に依存してさまざまです。DS施設で扱うデータも、ゲノムや遺伝子に関する大量の生命情報データや大気レーダデータ等による観測データから、古典籍や社会調査、公的マイクロデータまで広い範囲に及ぶため、DS施設各センターは、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
共同利用基盤施設
施設長 藤山秋佐夫

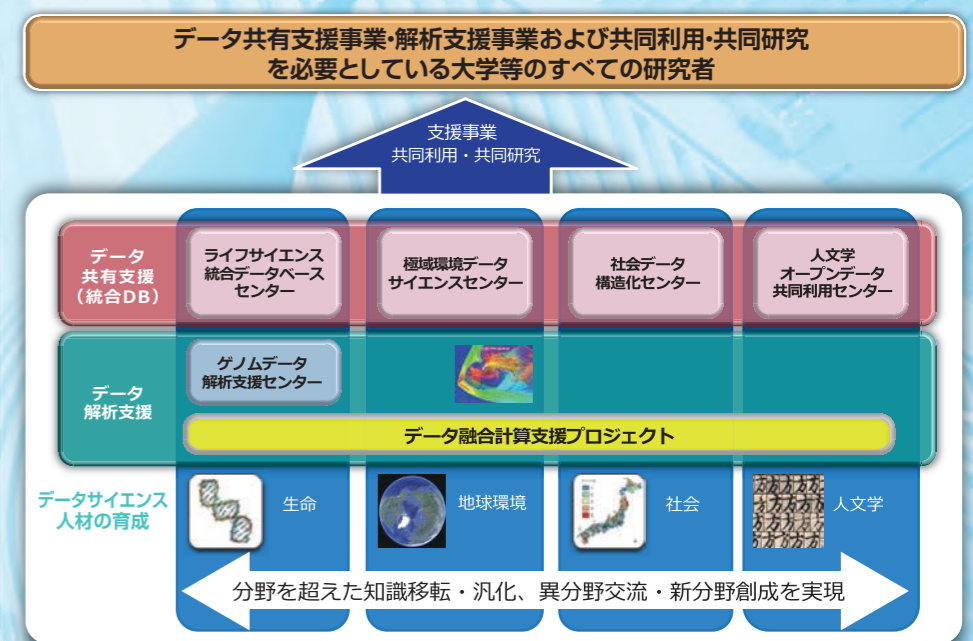


データサイエンス棟(立川キャンパス)

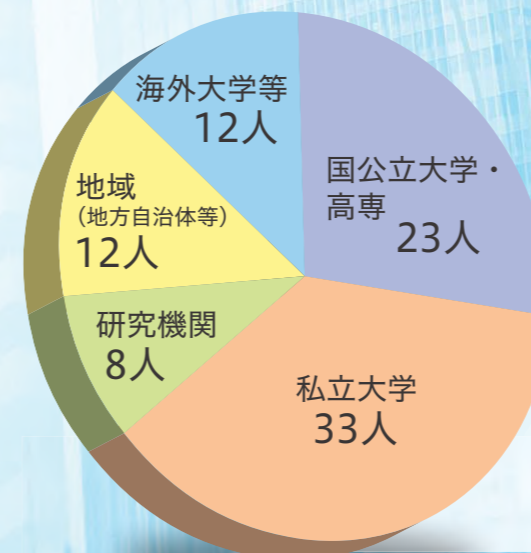


データ共有・データ解析支援・人材育成の3つの活動で研究を支援

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中心となるデータ共有支援、データ解析支援、人材育成の3本柱を軸に研究・支援活動を行っています。当施設は、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、2017年度にはそれぞれ20件、8件の計28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付しており、全国の研究者の皆様へに支援や共同利用・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。

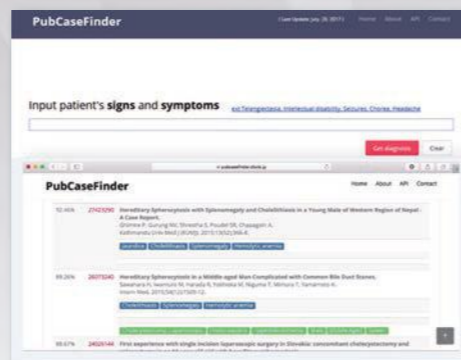
公募型共同研究 参加者所属機関内訳(2017年度)
 公募型共同研究には、国公立大学以外に、研究機関、高等専門学校、地方自治体、海外の大学など計46の機関から参加を得ることができ、参加者は全部で88名(当機構所属研究者を除く)となりました。

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。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)

2017年度の主な成果:

- ▶ 症例報告から希少疾患関連症状を収集したり、患者の症状から類似希少疾患・症例報告が検索できる希少疾患診断支援システムを公開しました。
- ▶ データベースから効率よくデータを取得し、利用者との接点となるアプリケーションへ渡す各種ミドルウェアを開発しました。
- ▶ 「BioHackathon 2017」を東京と盛岡で開催し、12か国・約140名が参加しました。その成果を国内に展開する国内版バイオハッカソンの他、SPARQLthon (毎月1回)、BLAH (Biomedical Linked Annotation Hackathon) 他多数のワークショップ開催により、データの標準化を促進しました。



希少疾患診断支援システム「PubCaseFinder」の画面例



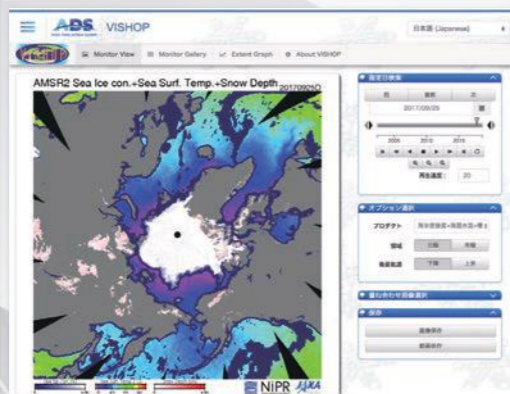
国際開発者会議「BioHackathon 2017」集合写真
*JST/NBDC(科学技術振興機構バイオサイエンスデータベースセンター)と共催

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)を统一的に整備しつつ、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)

2017年度の主な成果:

- ▶ 北極域データのアーカイブ・公開用に広く利用されてきているADSシステムにデータ処理を追加するとともに、両極域のデータを扱えるように拡張するためのシステム開発を行いました。
- ▶ 国内各機関が所有する超高層大気分野の多種多様なデータを、容易に横断的に検索・解析できる、大学間連携プロジェクト「IUGONET」メタデータベースシステムの機能強化を進めました。



北極域データアーカイブ(ADS)



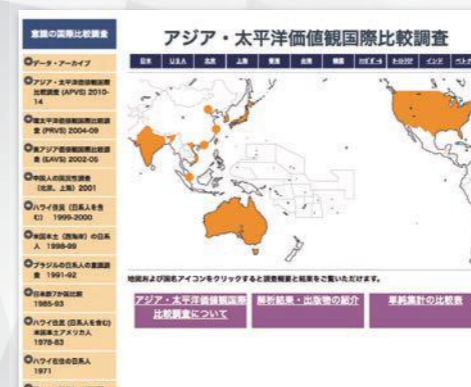
「IUGONET」

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、地域環境、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、人間・社会データ・コンプライアンス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」、オンラインで利用できる「公的調査のマイクロデータ」、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:吉野 諒三)

2017年度の主な成果:

- ▶ 日本人の意識に関する収集済みの社会調査データの属性別詳細集計表を作成・公開するとともに、アジア・太平洋価値観国際比較調査等の公開を開始しました。
- ▶ 公的マイクロデータの公開・利用について、国際会議の開催や研究コンソーシアムの事務局を担っています。昨年度は、政府・地方自治体の個票レベルのデータが利用できる「オンサイト室」をDS施設内に移転・開設しました。



アジア・太平洋価値観国際比較調査



第2回人間・社会データ構造化シンポジウム

- ▶ 話題の欧州のGDPR(一般データ保護規制制度)などを含め、ソーシャル・ビッグデータ時代の個人情報保護関連法律・倫理について広く啓蒙を図る「第2回人間・社会データ構造化シンポジウム」を開催しました。

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)

2017年度の主な成果:

- ▶ 「日本古典籍データセット」に、『伊勢物語図録』や『武鑑』など約1,000点を追加・公開しました。
- ▶ 世界のミュージアムやライブラリで使われる国際的画像配信方式「IIIF (International Image Interoperability Framework)」を活用し、任意の画像を切り出して収集可能な「IIIF Curation Viewer」を開発・公開しました。
- ▶ 「江戸料理レシピデータセット」で、江戸時代の卵料理を現代で活用するためのレシピと写真を追加し、ウェブサイト『クックパッド』にも掲載しました。



「IIIF Curation Viewer」を用いて画像を収集した「顔貌コレクション」を公開



くずし字チャレンジ!!:
くずし字解読のためのAIコンペを開催

ゲノムデータ解析支援センター

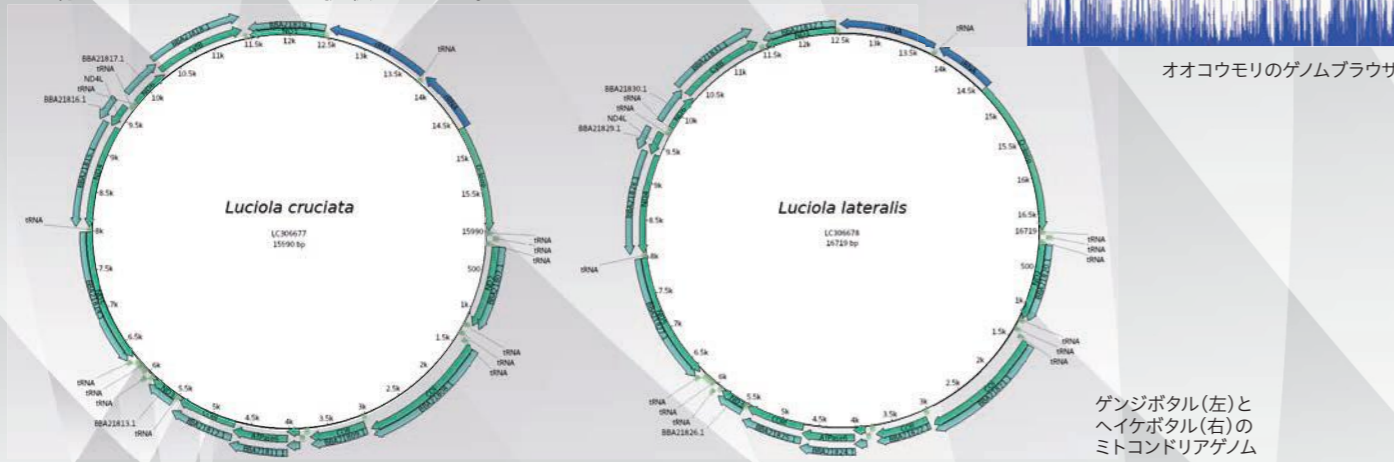
生命はゲノムDNAに書き込まれた情報を基に形作られています。そのため、ゲノムDNAに記録された情報の解読が生物学研究の出発点になっています。しかし、最新の技術から得られるゲノム情報データは、例えばヒトゲノム全体が30億文字あるのに対して、ひとつが300文字程度の短いものが数億本から成るデータセットとなっています。本センターは、ここから生物学、医学、環境学などさまざまな研究の発展につながる新たな知見を取り出すために、最先端のバイオインフォマティクス技術を開発・提供し、解析相談や共同研究を通じて研究者を支援します。(センター長:野口 英樹)

2017年度の主な成果:

- ▶ 昨年度は「ROIS-DS-JOINT」3件を含む、計22件の解析支援・相談を実施しました(継続課題含む)。
- ▶ 日本固有のホタル2種の全ミトコンドリアゲノムを解読し、25種類の昆虫を用いてその系統関係を明らかにしました(鹿児島大学との共同研究)。
- ▶ 支援の一環として、ゲノム情報を視覚化し、遺伝子名をはじめと各種検索も行えるゲノムブラウザを提供しました。



オオコウモリのゲノムブラウザ



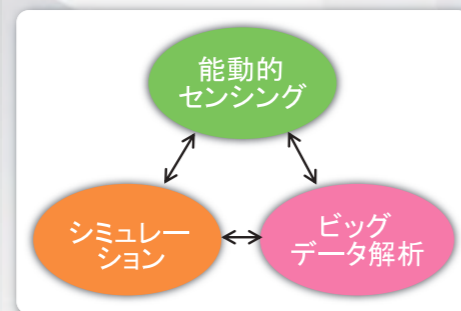
ゲンジボタル(左)とヘイケボタル(右)のミトコンドリアゲノム

データ融合計算支援プロジェクト

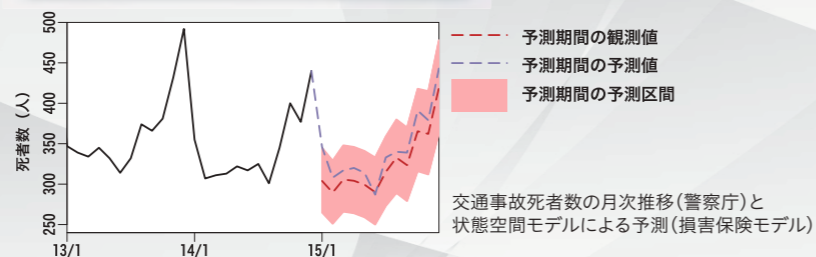
本プロジェクトでは、自然科学だけでなく、工学などの応用科学や社会科学等の広範囲にわたって、データとシミュレーションを統合する「データ同化」や、統計手法でシミュレーションを模倣する「エミュレータ」など、シミュレーションの可能性を広げる統計数理手法の研究開発を行っています。また、学術・産業界でシミュレーションの活用に関する課題をお持ちの方々に対し、シミュレーションと統計数理を融合させる諸技術のノウハウ等を提供し、課題解決を支援します。(代表:上野 玄太)

2017年度の主な成果:

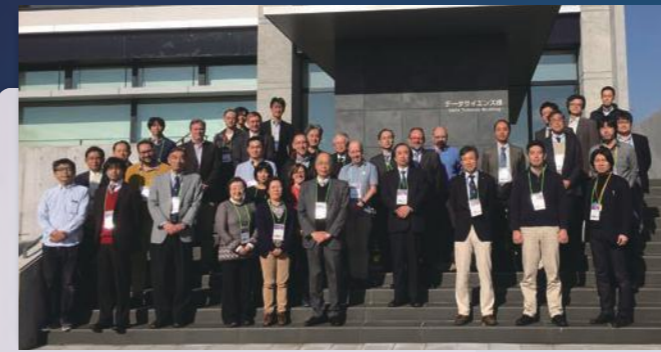
- ▶ 損害保険料算定のための、状態空間モデルを用いた時系列モデリング手法を開発・ソースコード公開し、公益社団法人日本アクチュアリー会の2017年度優秀論文として表彰されました。
- ▶ 海洋生態系モデルを元に、アンサンブルカルマンフィルタを用いたデータ同化システムを構築するための助言と技術指導を行いました。
- ▶ 太陽光発電で得られた電力をローカルに共有するためのマイクロ・グリッドの最適化手法に関する助言と技術指導を行いました。



プロジェクト概念図



交通事故死者数の月次推移(警察庁)と状態空間モデルによる予測(損害保険モデル)



国際ワークショップ

データサイエンスに関わる国際ワークショップでは、極域科学、ライフサイエンス、人文・社会科学を横断する学際的かつ国際的な視野に立ち、オープンサイエンス、オープンデータの将来ビジョンに関する議論を行っています。2017年には「分野を超えた科学データの共有・引用・出版」をテーマに開催し、約70名が参加しました。また社会データ構造化センターでは、アジア各国から統計部局関係者を招いた国際比較ワークショップ「国際マイクロデータワークショップ」を開催しています(年1回)。



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って、学会等での広報活動、問合せ対応、共同研究の仲介等を行っています。2017年度は、基礎生物学、農学、基礎医学の18学会でのブース出展等を通じて、特に要望の多いゲノムデータ解析支援を中心に幅広い研究分野での研究支援のコーディネーションを実現しました。



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(データ融合計算支援プロジェクト主催、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。

公開・関連ウェブサイトURL一覧

DBCLSウェブサイト	https://dbcls.rois.ac.jp/
サービス一覧ページ	https://dbcls.rois.ac.jp/services.html
希少疾患診断支援システムPubCaseFinder	https://pubcasefinder.dbcls.jp/
PEDSCウェブサイト	http://pedsc.rois.ac.jp/
北極域データアーカイブ(ADS)	https://ads.nipr.ac.jp/
IUGONET	http://www.iugonet.org/index.jsp?lang=ja
Polar Data Journal	https://pdr.repo.nii.ac.jp/
意識の国際比較調査	http://www.ism.ac.jp/~yoshino/
アジア・太平洋価値観国際比較調査	http://www.ism.ac.jp/~yoshino/ap2/index.html
オンサイト解析室	http://ds.rois.ac.jp/center3_micro/
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/
人文学オープンデータ共同利用センター(CODH)ウェブサイト	http://codh.rois.ac.jp/
ゲノムデータ解析支援センターウェブサイト	https://genome-info.nig.ac.jp/
シーラカンスゲノムブラウザ	http://coelacanth.nig.ac.jp/
MetaGeneAnnotator(メタゲノムアノテーター)	http://metagene.nig.ac.jp/
Platanus(ゲノムアセンブラ)	http://platanus.bio.titech.ac.jp/
データ融合計算支援プロジェクトウェブサイト	http://daweb.ism.ac.jp/support/
P ³ (Parallelized particle filter with Python 並列粒子フィルタ)	http://daweb.ism.ac.jp/support/software/P-cubed/P-cubed.html