

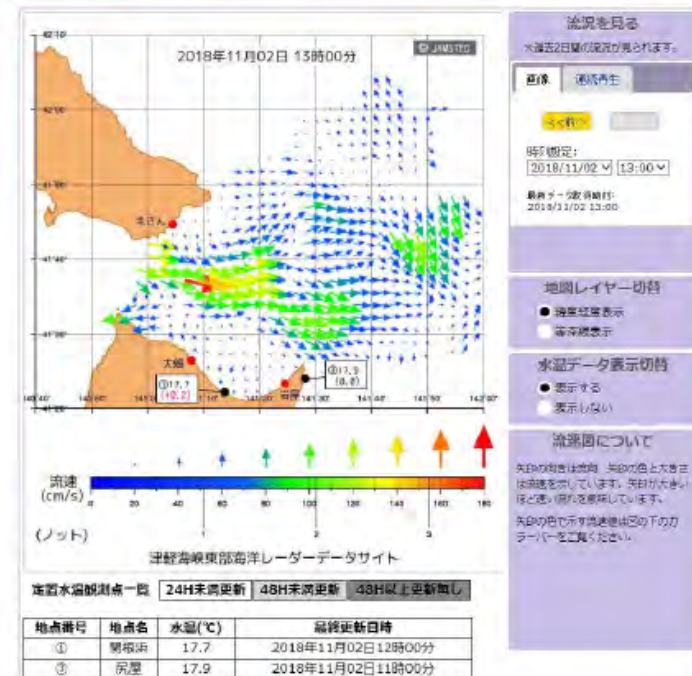
表面流マップの準リアルタイム公開・・・漁業者にも有用な情報提供



津軽海峡東部海洋レーダーデータサイト

MIO Ocean Radar data Site for the Eastern Tsugaru Strait (MORSETS)

<http://www.godac.jamstec.go.jp/morsets/j/top/>



表面流マップは、研究データという位置づけのみならず、おそらく漁業者や現場を航行する船舶などにとっても有用な情報。

Web公開に対する反応



NIMSの研究データマネジメント事例

アカデミア

- ・ きらりと光る方法論
- ・ 世界がハッとする成果

やり方・課題 解決シナリオ

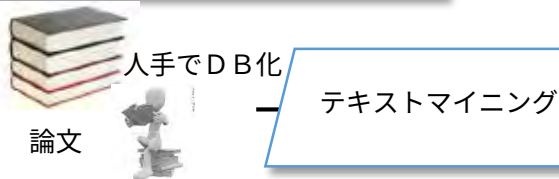
産業界

- ・ ビジネス展開の主役
- ・ 国際競争力強化

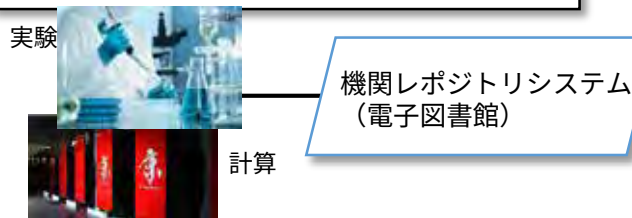
実験装置・計測機器からのデータ収集



文献からのデータ収集



大学等の成果からのデータ収集



アカデミックデータ収集
(信頼できる標準データ)

材料情報統合データ プラットフォーム

高機能物質・材料データベース



高速サーバー
(解析ツール)

MI²データプラットフォームを核に実用版を構築



研究者・技術者コミュニ
ティーによる研究推進

化学企業A
(カーボン素材)

自動車企業B
(車体素材)

鉄鋼企業X
(高性能鋼)

電機企業Y
(燃料電池)



企業内データの
徹底活用

人間系データバンク

• 人体寸法・体形データベース

- 収集データを国際標準化ISO TR 7250-2 (世界の人体寸法DB)制定のインシヤティブ
- 国際協力により計測技術の供与とデータの交換で蓄積
- 3次元形状DB (収集) をもとに少ない測定点から人体形状のモデル化 (相同モデリング mHBM、体形統計処理 HBS)
- データを活用するシミュレータの構築(Dhaiba)



• 歩行・歩容データベース

- 歩行研究で標準データ (裸足歩行) を計測 ($n > 500$)
- 年間200万件のアクセス
- コンサルティングによる支援
- 転びやすい人の歩行特徴の解明糸口
- 日本転倒予防学会推奨靴などの開発



• データポリシー

- 人間計測データ(オープン), 解析・二次データ(クローズ)
- 歩行データ (オープン), 転倒歴, 転倒スコア (クローズ)





ヒトセルアトラス (Human Cell Atlas) プロジェクト



背景・目的・期待される成果

- ヒトゲノム情報の多くは細胞の集合（バルク）に由来する平均値的な情報であり、生命現象を本質的に理解し将来の医療をさらに発展させるためには、生体の究極の基本単位での数百種類にのぼる細胞種を「1細胞レベル」で解析することが必要とされる。
- こうした背景をふまえ、ヒトの全細胞（約37兆個）について、種類・状態・系統などを最新の1細胞解析手法によって分類し、カタログ化することを目的とした国際プログラム「セルアトラス（ヒト細胞百科事典、HCA）プロジェクト」が、主要な4研究機関（英国：サンガー研究所、米国：ブロード研究所、欧州：カロリンスカ研究所、アジア：理研）を中心に進められ、世界中の参画研究者が集まる全体会議で、統一的なデータ解析手法やデータ公開などの議論が行われている。
- 本プロジェクトの推進のためには、系統的かつ高精度でヒトを構成する全ての細胞について個々の細胞固有の遺伝子発現状態を網羅的に解析する技術が必要であり、理研をはじめとする日本の持つ1細胞ゲノミクス解析技術が重要な役割を果たし、アジア地域のハブとしての機能が期待されている。

各国で HCA 全体会議が開催され、ヒトセルアトラス作成に向けた議論が進んでいる



研究が進められている解析対象



- 免疫系
- 肝臓
- 腎臓
- 肺
- 皮膚
- 腸
- 心臓
- 神経系（等）

米国では、NIHがこの二年間でセルアトラス関連のプロジェクトに約200億円の予算を投じている。欧州でも、特定の臓器等を対象とした複数の国際研究プロジェクト（約35億円）が進められている。

FANTOMプロジェクト

FANTOM = Functional ANnoTation Of Mammalian genome

理化学研究所の主導による国際共同研究プロジェクト

理化学研究所で配列決定したマウス完全長cDNAや独自技術であるCAGE法を用いた哺乳類のゲノム機能研究

FANTOM1 - マウス完全長cDNA (約2万個) の機能アノテーション

FANTOM2 - 約6万個のマウス完全長cDNAの機能アノテーション

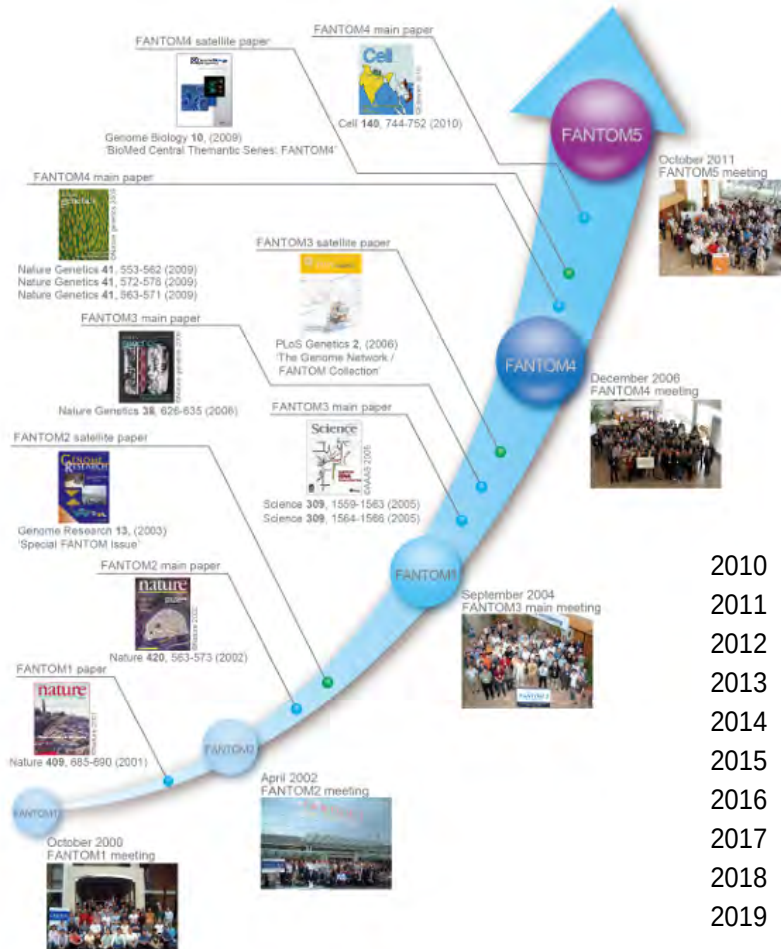
FANTOM3 - 哺乳類ゲノムの転写ランドスケープ

FANTOM4 - 発現制御ネットワークの理解

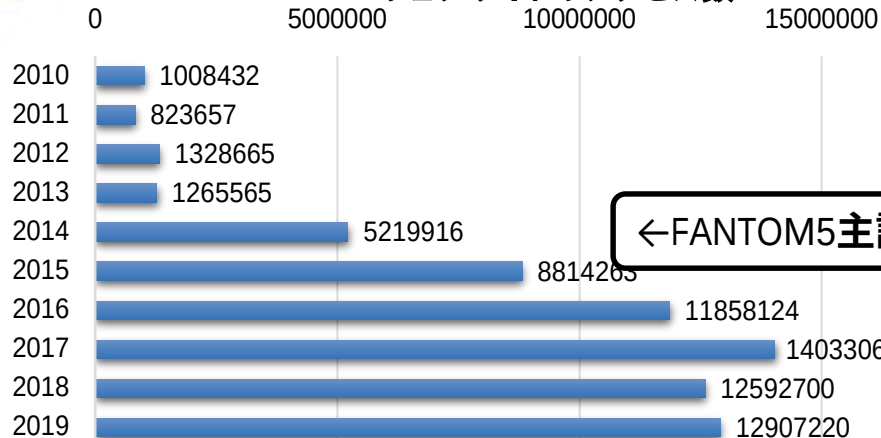
FANTOM5 - 哺乳類プロモーター・エンハンサー・長鎖ノンコーディングRNA(lncRNA)、マイクロRNA(miRNA)の発現地図

FANTOM6 - ノンコーディングRNAの機能解析

研究成果、解析に用いたデータやその解析結果はリソースとして全世界に無制限で公開(Creative Commons Attribution 4.0 International License)



FANTOMウェブサイトのアクセス数

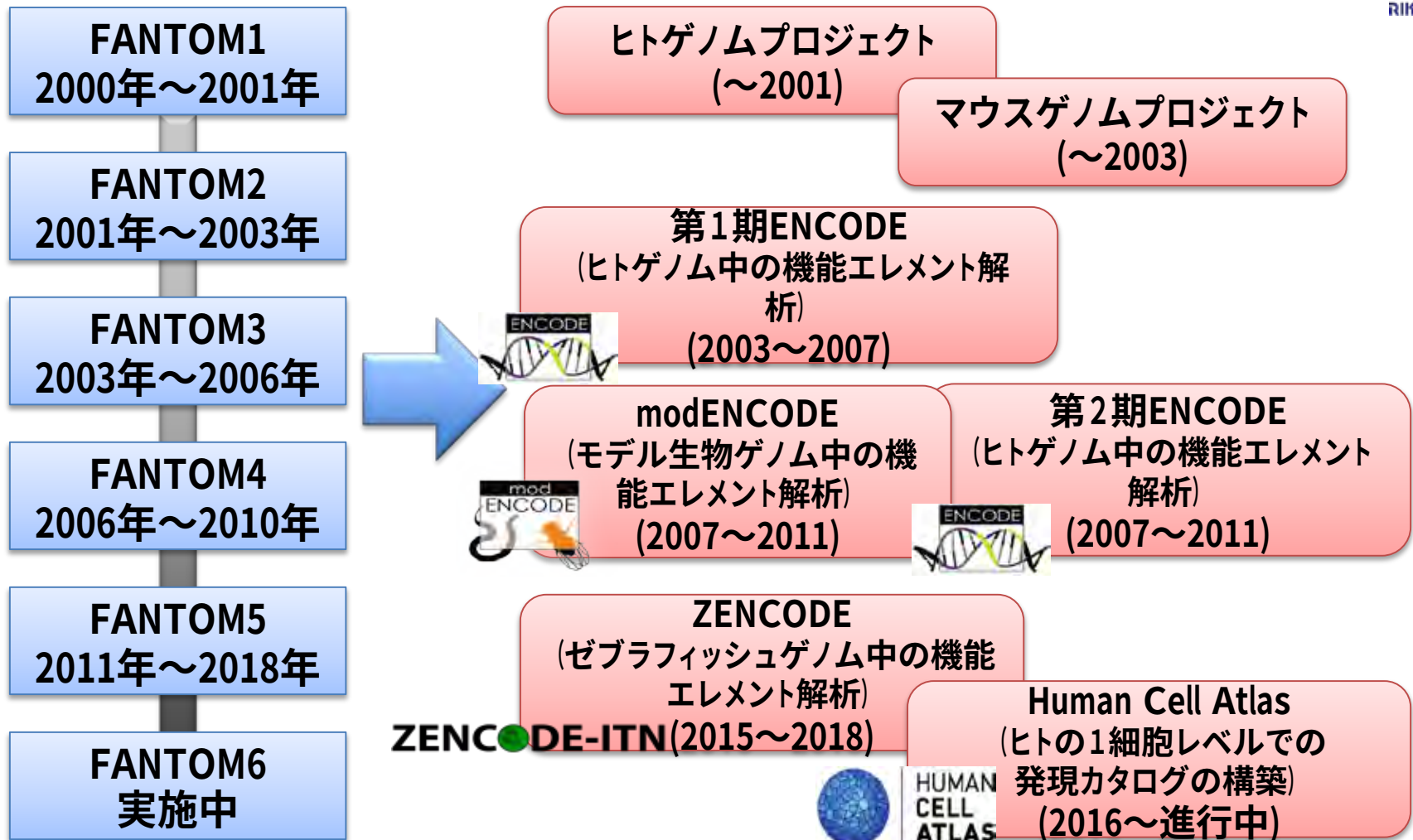


←FANTOM5主論文公開

※2019年は1~3月実績より推定

出典：理化学研究所

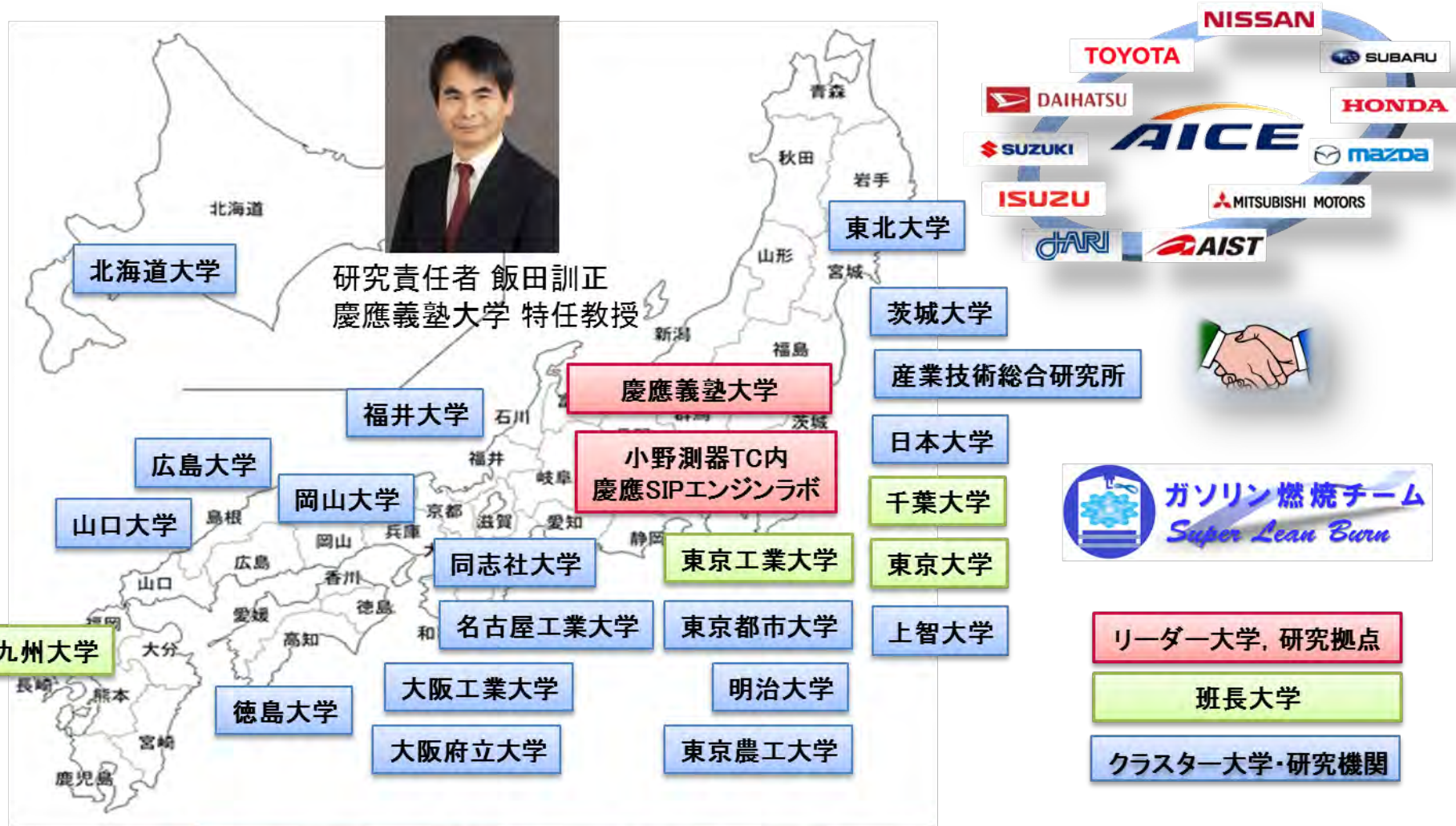
FANTOMプロジェクトの波及効果



様々な国際共同研究プロジェクトに招待、参画へ

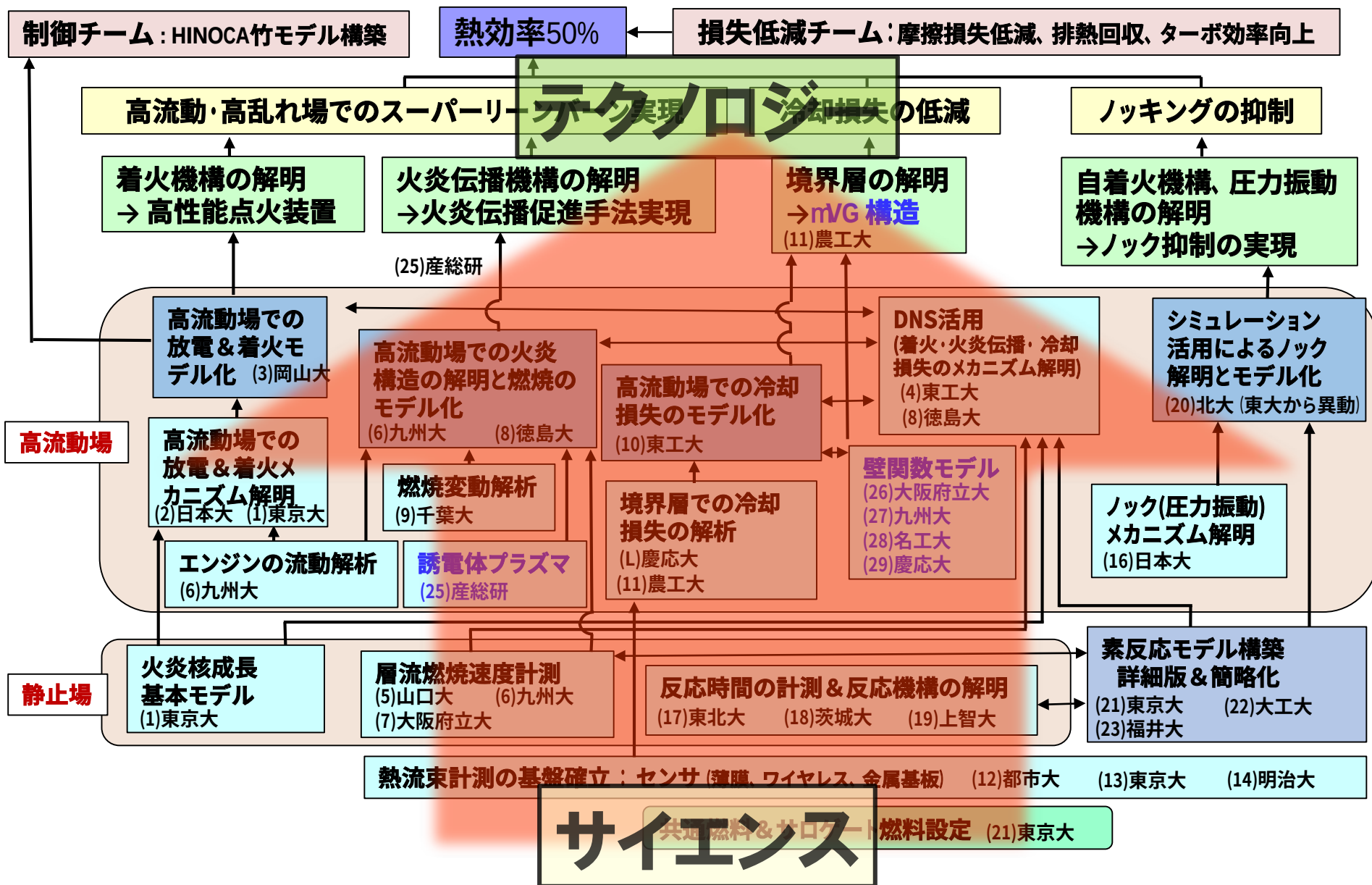
3. その他の研究データマネジメントの事例

SIP革新的燃焼技術（ガソリン燃焼チーム）の研究データマネジメント：全体像



- ・日本の知恵を結集し、世界を牽引する要素技術を創出
- ・産業界の発展にも貢献

SIP革新的燃焼技術（ガソリン燃焼チーム）の研究データマネジメント：チーム内編成



SIP革新的燃焼技術（ガソリン燃焼チーム）の研究データマネジメント：要点

◆目的：プロジェクト内における20を超える研究チームの異分野研究者間の創発的研究推進

◆手法：プロジェクト内の研究者間で、日々の研究データを適切に共有できるシステムを構築するとともに、対面のみならず、Web会議システムを介したディスカッションを促進

1. 柔軟なアクセスコントロール	PMがプロジェクトメンバーの研究データへのアクセスを柔軟にコントロールできること
2. 幅広いデータへの対応	研究データの種類は、生データ、加工データ、メタデータのいずれも対応できること ＊）プロジェクトメンバー間であれば、生データでもある程度理解できるため、生データであっても共有できること
3. コミュニケーションツールとの連動	遠隔にいる複数のプロジェクトメンバーどうしが、データを閲覧しながら円滑にコミュニケーションが取れること ＊）プロジェクトメンバー間であれば、データを見ればある程度理解できるとは言え、やはり直接話してデータについて議論する機会を柔軟に設定できることが重要
4. ストレスフリーなメタデータ付与とそれに基づく検索機能	各データにメタ情報をストレスなく付与することができる仕組みと、それを基に効率よく検索できる仕組みがあること ＊）ガソリン燃焼チームでは、研究者に対して、決められた画一的な項目のメタ情報を記入させるのではなく、1000文字のフリースペースを与えて、研究者が必要と思うキーワードを好きなように記入させる方法を採用
5. セキュリティ	十分なセキュリティが担保されていること
6. ユーザーフレンドリーな操作性	以上の仕組みがストレスなくユーザーフレンドリーであること ＊）システムに不得手な研究者にいかに関感的に使ってもらえるようにするかが重要
7. アジャイルな開発体制	以上のシステムを研究者に提供したときに、使っていく上で新たなニーズや不具合が出てくるので、これらにアジャイルに対応できる体制を整えること
8. 適切な目標	メンバー全員の知恵を集結させないと達成できないような野心的で実現可能な目標の設定 ＊）ガソリン燃焼効率50%の達成

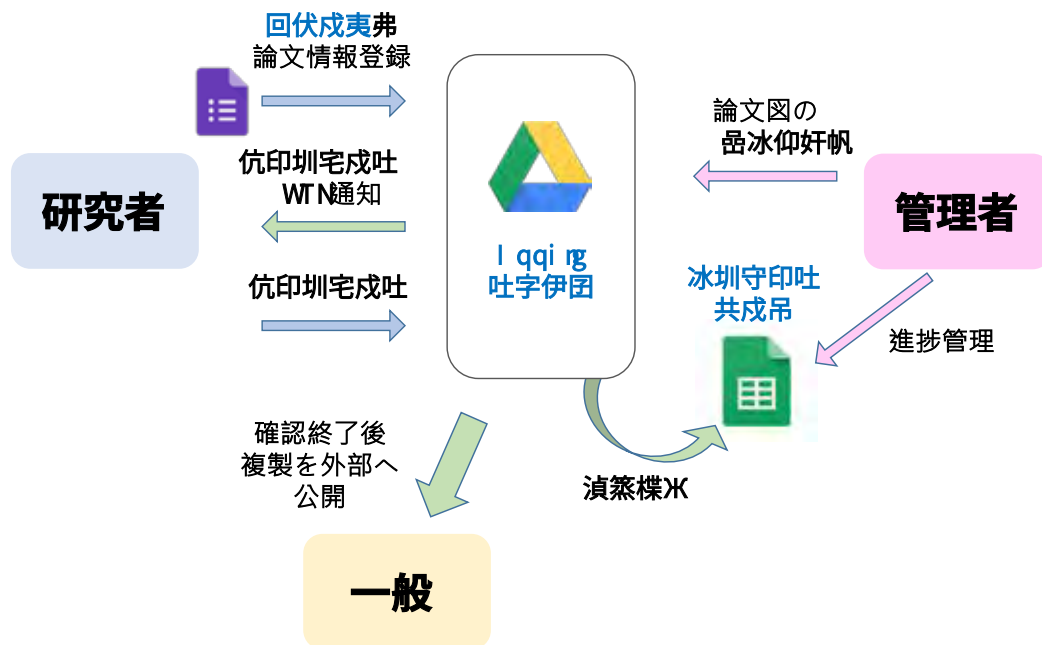
東京大学定量生命科学研究所のMODシステム

◆目的：研究論文における研究公正の確保

◆手法：受理された論文の「最終原稿と図」「論文に使用した全てのデータ（生データ・加工データ）」「論文作成に関するチェックリスト」を管理する独自の管理システムを構築

MODシステム（Manuscript scan & Original data Deposition）

全体概要図



データアップロード画面



データ公開画面



出典：東京大学 定量生命科学研究所

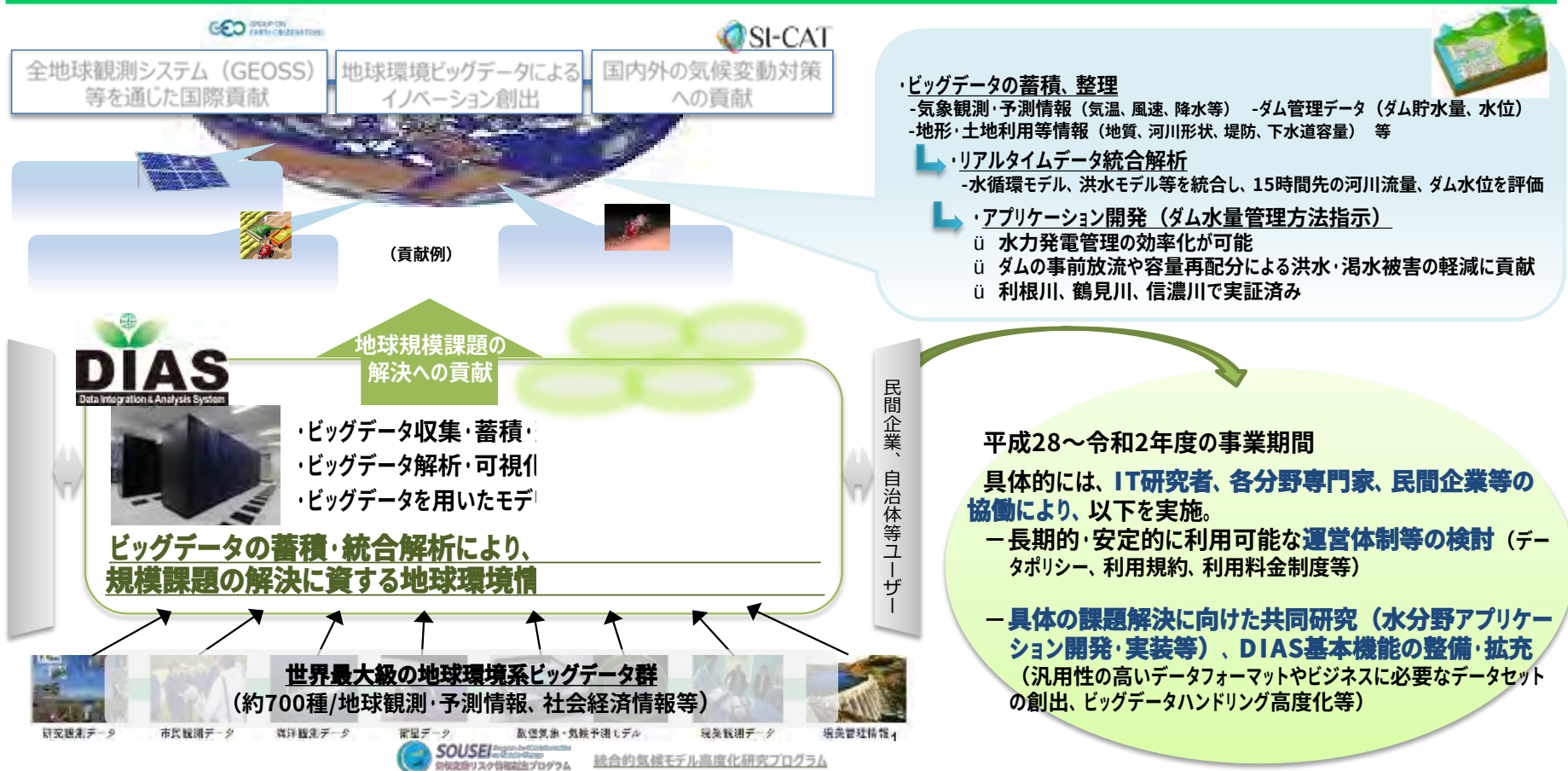
◆備考：1年3か月の運用期間で60報を超える論文情報を登録。データ容量は約200GB。管理のために、教員1名、事務員1名が本務の間に対応。画像スキャンは専任職員1名が対応。

データ統合・解析システム (DIAS)

気候変動適応戦略イニシアチブ

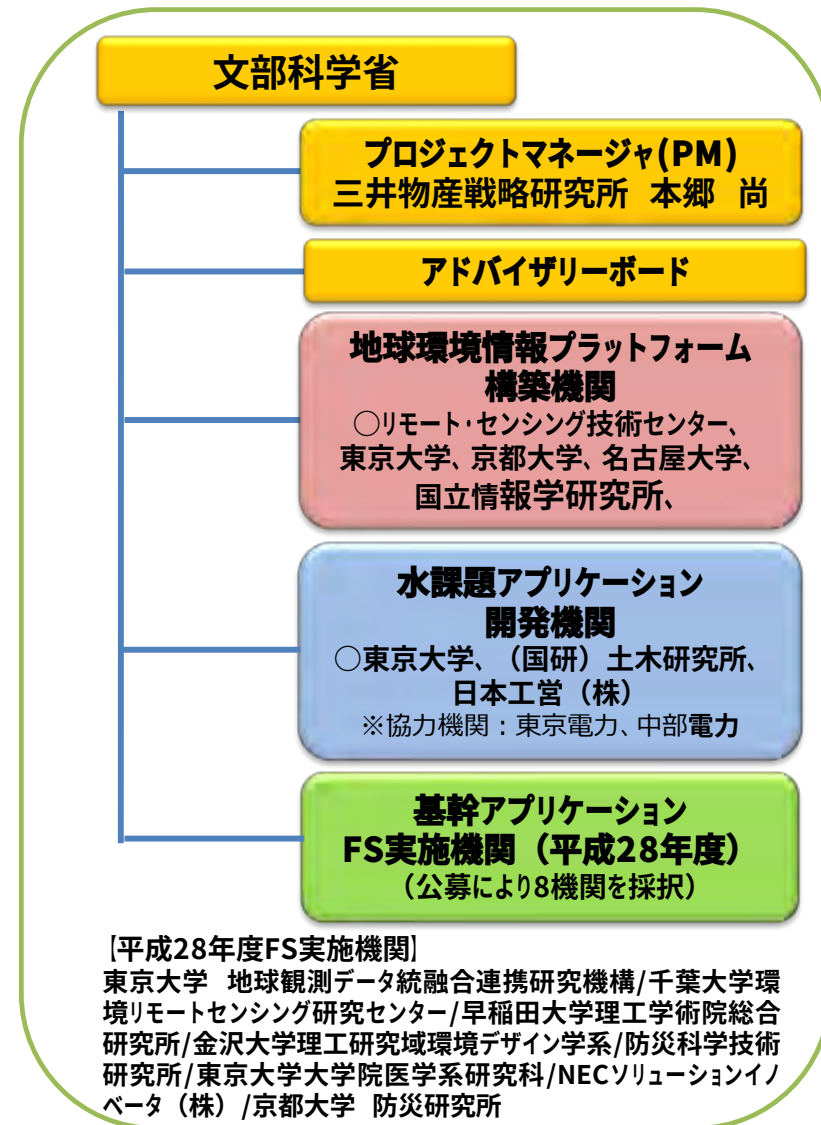
地球環境情報プラットフォーム構築推進プログラム

- 地球環境ビッグデータ（観測情報・予測情報等）を蓄積・統合解析し、気候変動等の地球規模課題の解決に資する情報システムとして、「データ統合・解析システム (DIAS)」を開発。
- GEOやIPCC等を通じた国際貢献や学術研究の場面でも活用されており、今後は、ニーズ等を踏まえた長期的・安定的な運用体制の構築や具体的な課題解決に向けた共同研究等の研究開発を推進することで、産学官で活用が可能な地球環境情報プラットフォームの構築を一層推進。



データ統合・解析システム (DIAS) の概要①

実施体制



DIASシステム構成



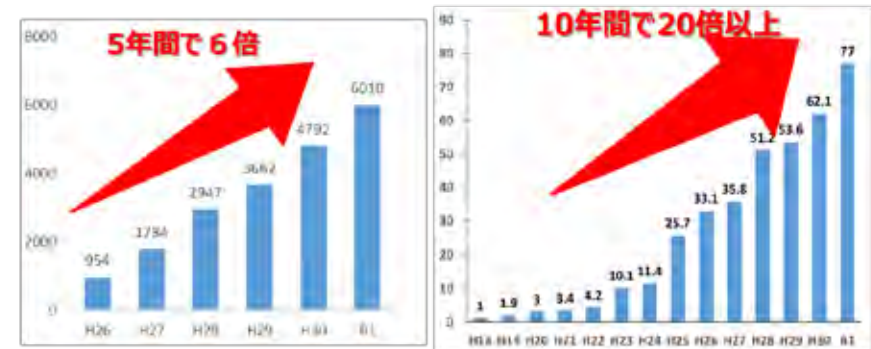
DIASデータポリシー

- DIASのデータ利用ポリシーは、第1にデータ提供機関のデータ利用規約を優先することとしている。データ提供機関の定めがない場合は、DIASのデータ利用規約を適用することとし、研究及び教育目的に利用することができる。
- なお、DIASのビジネス利用に向けて、営利目的のデータ利用が可能となるよう検討を行っている。

データ統合・解析システム (DIAS) の概要②

ü ユーザー数が直近5年で6倍になるなど、利用者が国内外で拡大

ü DIASの保管データ容量は、10年間で20倍以上に増加。



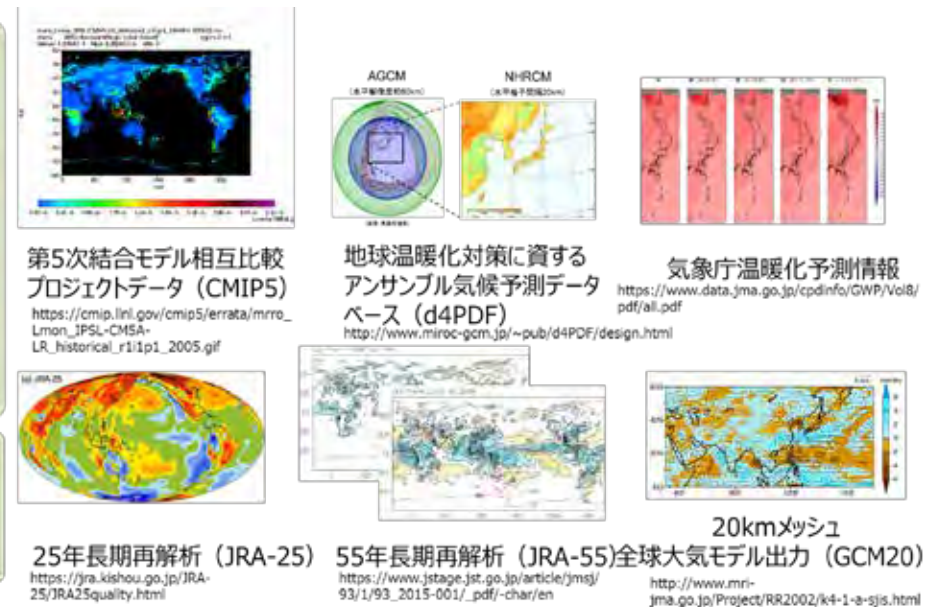
年度ごとのDIASのユーザ数の増加 (人)

DIASの保管データ容量の年度ごとの増加 (2006年のデータ容量を1とした比率)

リアルタイムデータの例



データセットの例



○ 人間文化研究機構 国文学研究資料館 新日本古典籍総合データベース

大規模学術フロンティア事業「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」によって作成されている、日本の過去の書籍（典籍）を集めたデータベース。日本国内30万点の典籍のデータ化を行い、国内外における文学・日本文化研究に資することに加え、データを活用した異分野連携への活用も目的としたデータである。

URL : <https://kotenseki.nijl.ac.jp/>

○ 人間文化研究機構 国立国語研究所 コーパス

言語を分析するための基礎資料として構築されたデータセット。書き言葉・話し言葉のデータ等を語彙ごとに分類しつつデータ化しており、自然言語処理等のもっとも基礎的なデータとしても用いる。奈良時代から明治、現代までのデータを蓄積しているとともに、アイヌ語のデータ、Web上のデータなど、対象も多岐にわたる。

URL : https://pj.ninjal.ac.jp/corpus_center/

○ 人間文化研究機構 国立歴史民俗博物館 khirin (Knowledgebase of Historical Resources in Institutes)

日本歴史資料の総合的なデータインフラ構築のために作成されている。東京大学史料編纂所が古代～近世の有力な寺院や家のデータを中心に集積しているのに対して、khirinは近世～近代の地域歴史資料を対象としている。

URL : <https://khirin-ld.rekihaku.ac.jp/>

○ 人間文化研究機構 国立歴史民俗博物館 みんなで翻刻

日本の、くずし字といわれる過去の字体で書かれた文字をクラウドソーシングで現代の文字に変換（翻刻）するプロジェクト。クラウドソーシングプロジェクトではあるが、令和3年2月20日時点で800万文字を超える文字数の古文書・古典籍が翻刻されており、テキストデータとしては一大データ基盤としても位置付けられつつある。

URL : <https://honkoku.org/>

○ 国立文化財機構 東京文化財研究所のデータセット

日本美術に関する研究を行うための基礎的なデータセットを提供している。展覧会に関するデータ、美術家・美術関係者に関する主に明治期以降の人名データなどが代表例となる。また、これらの人名を国内外で連携して発見できるように、Getty研究所とも協定を結び、美術関係のデータ基盤の構築を行っている。

URL : <https://www.tobunken.go.jp/japanese/database.html>

○ 国立文化財機構 奈良文化財研究所 全国遺跡報告総覧

日本全国の考古遺跡に関する発掘調査報告書のデジタルデータ。発掘調査の成果でもあり、研究のある段階におけるデータとしても位置付けられる発掘調査報告書を、自治体・研究所等の作成元を問わず、全国的に検索・発見・閲覧することができる。

URL : <https://sitereports.nabunken.go.jp/ja>

○ 立命館大学 アートリサーチセンター 浮世絵ポータルデータベース

Web上で公開されている浮世絵を統合的に研究レベルにおいて検索可能としたデータベース。国内外の複数機関の浮世絵を同時に検索し、分析可能。

URL : https://www.dh-jac.net/db/nishikie/search_portal.php

○ 漢字字体規範史データセット (HNG)

各地域・時代における漢字字体（文字の骨組み）の標準の変遷を追い、過去にアナログカード化されていた情報をデジタルデータ化したデータセット。漢字の様々な字体を検索可能なデータベースを提供するとともに、漢字の部品から、字体を検索するシステム、および字体を集積したカードをデータセットとして公開している。

URL : <http://www.hng-data.org/>