

# 情報・システム研究機構のこれまでの振り返って

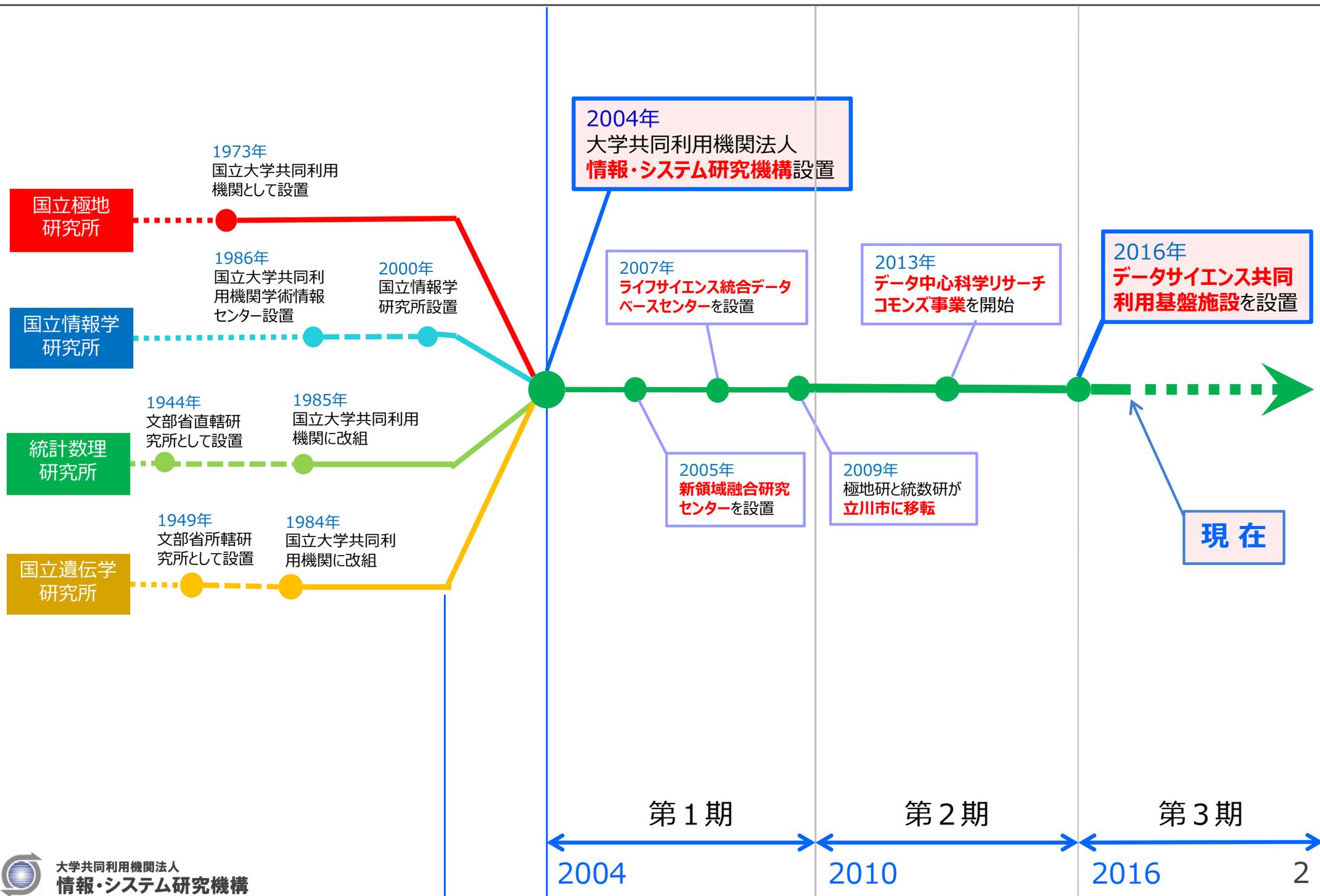
---

情報・システム研究機構

北川 源四郎

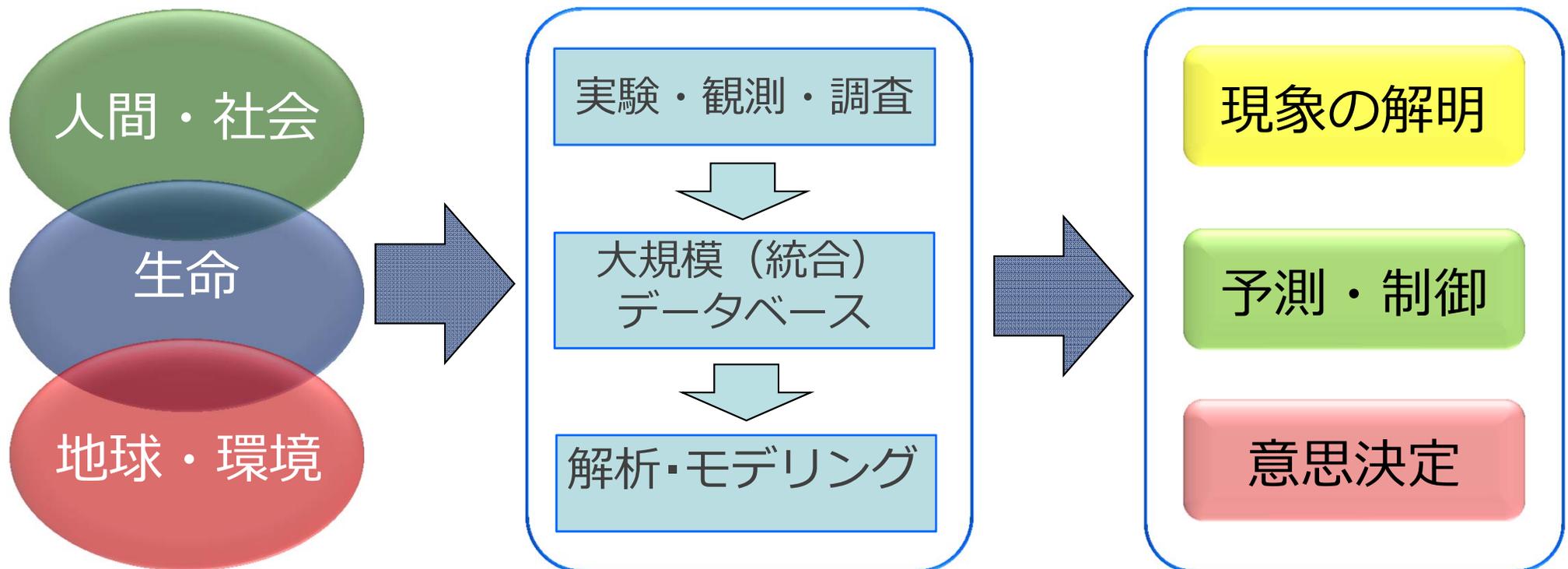
一橋講堂 2/20/2017

# 機構の設立からこれまで



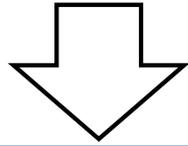
# 情報・システム研究機構の理念

生命，地球，環境，社会などの複雑な問題を，物質とエネルギーの観点に替って**情報とシステム**という立場から捉えるための，方法の研究，研究基盤の整備および融合研究による新分野の開拓を行なう。



# アカデミック・ビッグデータ

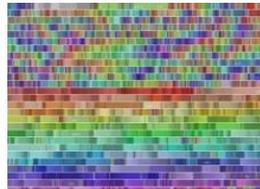
## 情報通信技術の飛躍的發展



## 大量・大規模データの集積

- 生命科学: DNA, Micro-array data
- マーケティング: POS data
- ファイナンス: High frequency data
- 環境科学, 防災 (河川)
- 地球科学 (地震学・気象学)
- 農学: Drone
- 天文学 (Whole-sky CCD camera)
- 高エネルギー物理学 (LHC)
- 物質・材料科学
- 言語学: コーパス

国立天文台 ALMA

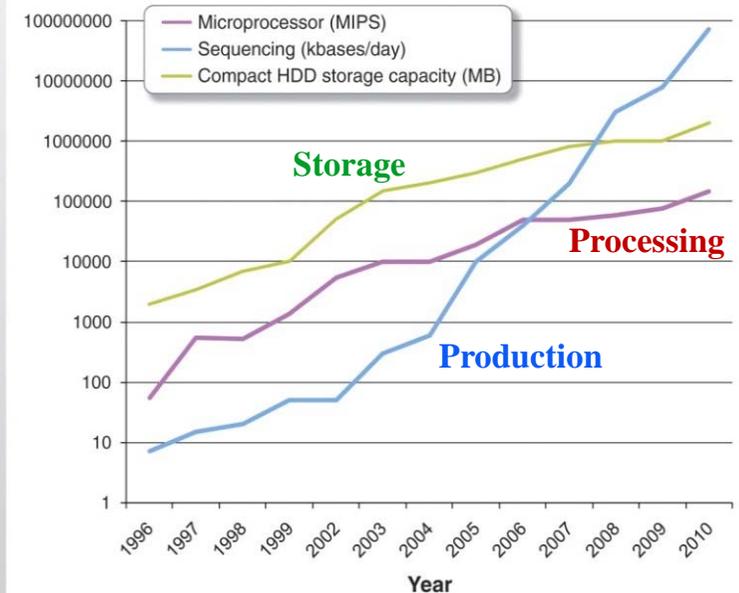


## 情報通信技術

- 観測機器、センサー
- インターネット、通信
- データベース
- 並列計算機

### Sequencing Progress vs Compute and Storage

Moore's and Kryder's Laws fall far behind



(TECHTILIS Feb. 24, 2011)

# ソーシャル・ビッグデータ

## 人間社会の活動を精細・網羅的に記録し、デジタル化できるようになった結果がビッグデータ

- インターネット： Web, SNS, メール, モバイルフォン
- センサーデータ： 家電, 自動車, GPS, RFID
- ドローンデータ： 画像, スペクトル, 温度
- 取引データ： POS, 株取引, 不動産情報
- マルチメディア： 画像, 音声,
- ログデータ： ソフトウェア・ログ, ライフログ

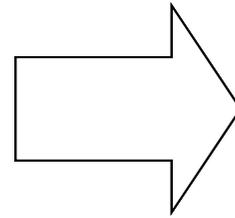
## 活用例

- マーケティング（市場予測、顧客行動予測モデル）
- オンラインショッピングにおける推奨機能
- データ駆動型産業
- 医療・創薬・保健における個人化対応
- 社会インフラのスマート化
- センサーデータ活用（防犯, 防災, 故障検出）
- Evidence Based Policy Making, Data Journalism



# 社会における科学の役割

専門家の  
経験と勘



科学的方法  
根拠に基づく



- 占星術, 航海術, 錬金術
- 工業生産過程 (ものづくり)
- 天気予報, 経済予測
- マネジメント, マーケティング
- リスク管理, ファイナンス
- 科学的発見 (発見科学)
- サービス
- 政策決定

# ビッグデータのインパクト

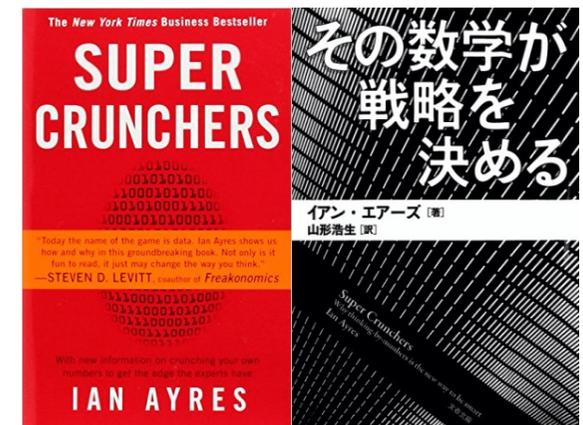
- **超スマート社会**が実現する。
- すべての研究は**データサイエンス化**する。

## 「大量データ分析」が「専門家の経験と勘」を凌駕する

Ian Ayres : **Super Crunchers**,

Why thinking-by-numbers is the new way to be smart

- EBM 医療診断支援
- 保険料の設定
- ネット販売の個別価格設定
- ワインのヴィンテージ評価
- スポーツ（野球等）のリクルーティング
- 人事採用
- 航空会社顧客サービス
- 判決予測、取引業者評価



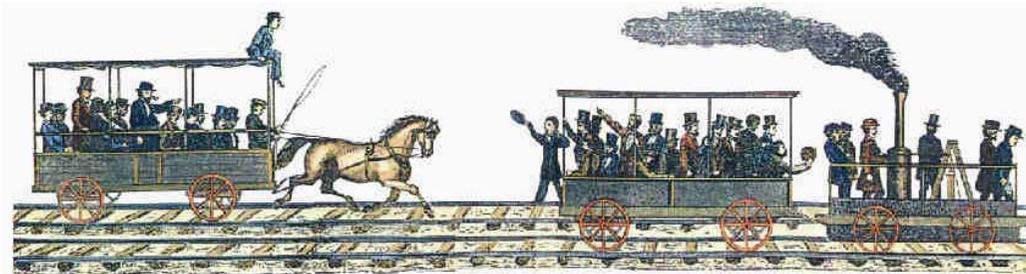
(翻訳本)

# 歴史的瞬間の再現

- 鉄道馬車 vs. 蒸気機関車

蒸気機関 ⇨ 工業化社会

1776年 ワット蒸気機関



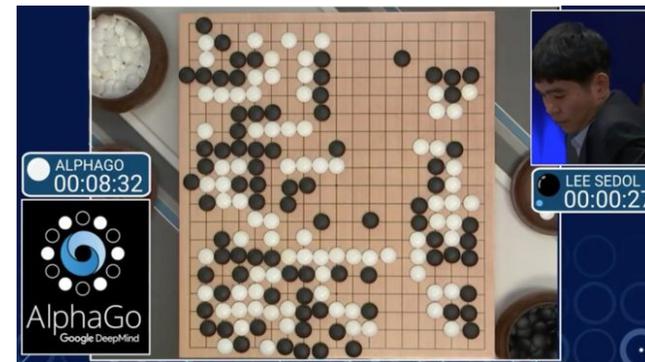
ボルチモア・オハイオ鉄道  
1830年12月25日

佐野正博「技術の生存競争 ---「動力」に見る進化論」『週刊朝日百科 世界の歴史』第110号, 1991,p.695

- 経験と勘にもとづく専門技能 vs. 大量データ分析  
ビッグデータ解析・人工知能 ⇨ 超スマート社会



(日本将棋連盟・ドワンゴ)



(アルファ碁 << Master・Google DeepMind)

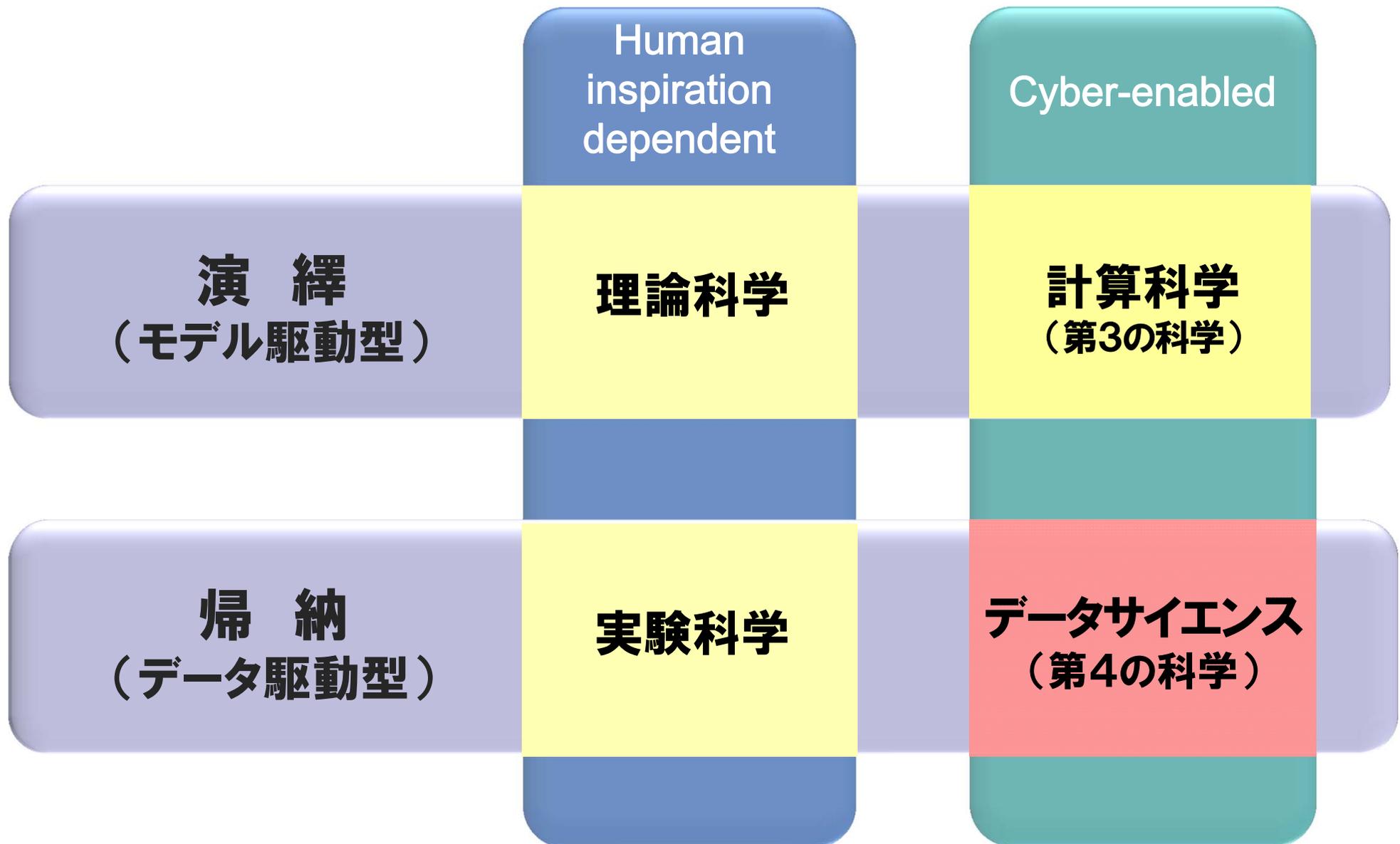
# ビッグデータ活用における課題

- ビッグデータには膨大な知識や価値が埋もれている。
- しかし、現在の方法・技術では有効活用は不十分。
  - ・多くは構造化されていない
  - ・価値密度が低い
  - ・不均一（形式、精度、観測頻度、非定常性）、スパース
- ビッグデータを効果的・効率的に集約し、知識発見や価値創造を行うための革新的な方法が不可欠。

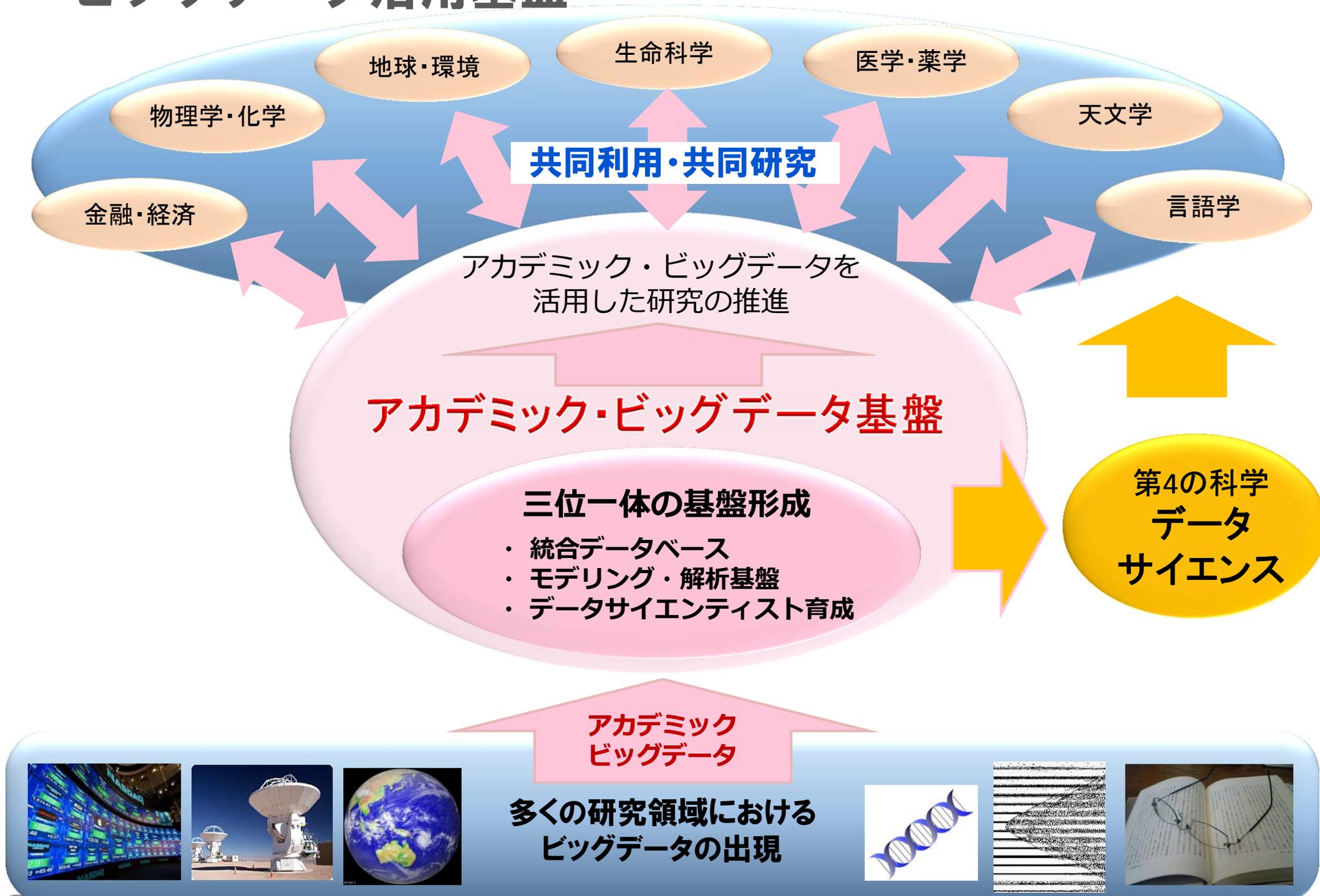
⇒ 大規模データ活用の方法論：第4の科学的方法論が必要

- 大規模データ処理 大量の散在するデータをリアルタイム処理するための技術
- データ可視化 膨大な高次元データや計算結果を人間が把握できるようにするための技術
- データ解析 ビッグデータからの深い知識獲得のための方法

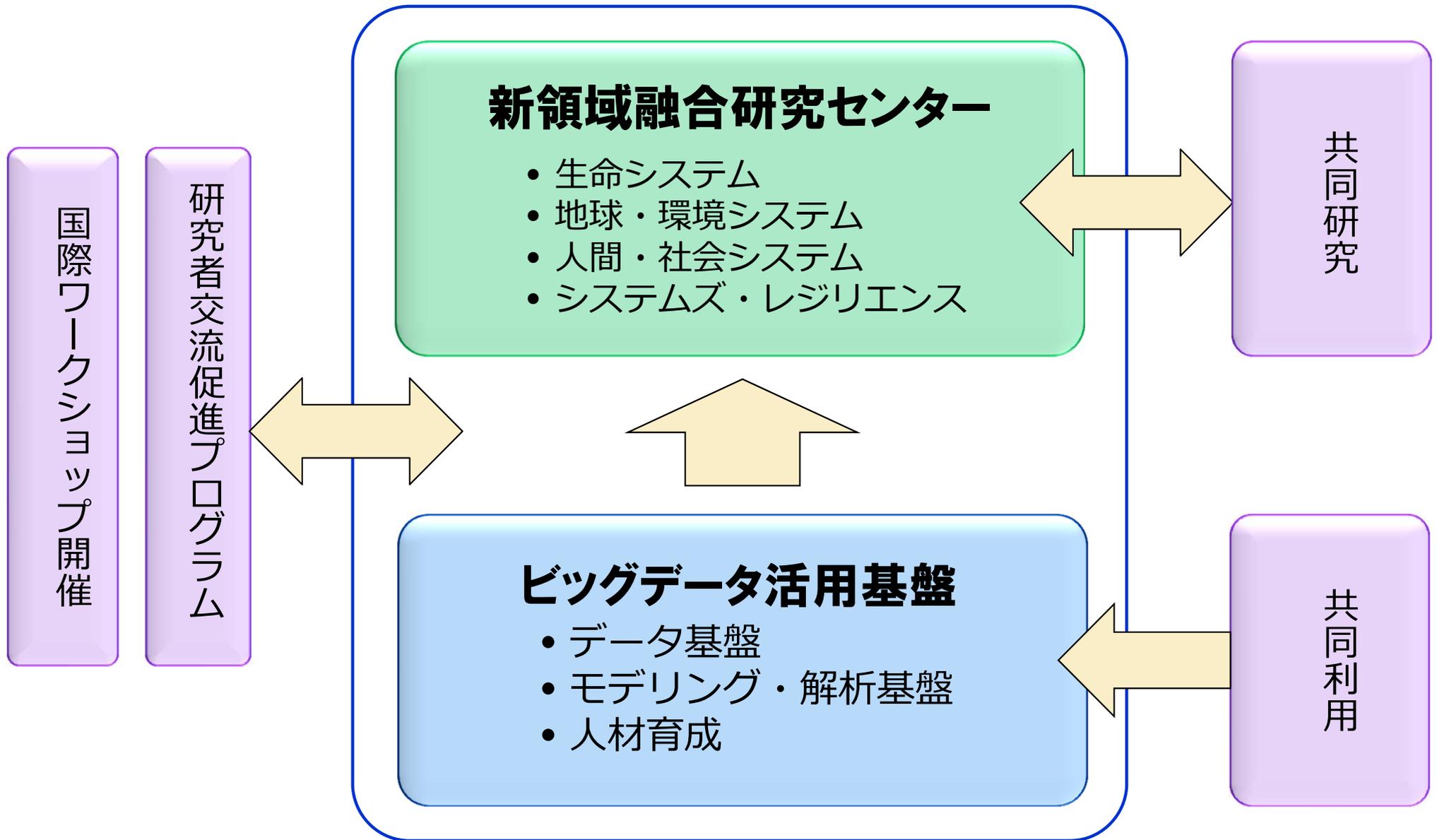
# 第4の科学：データサイエンス



# ビッグデータ活用基盤



# データ中心科学リサーチコモンズ（H25～H27）



# ビッグデータ活用基盤の形成

## 人材育成

- データサイエンティスト（T型， $\Pi$ 型人材）の育成
- 統計思考院，新領域融合研究センターにおけるOJT

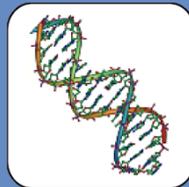
## モデリング ・解析基盤

- 可視化，構造探索，機械学習
- モデリング，知識発展，データ同化
- シミュレーション，予測，意思決定

## データ基盤

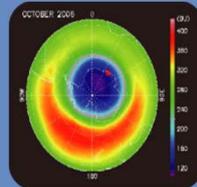
ライフサイエンス  
統合データベース  
センター

生命科学



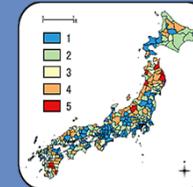
PANSY データ  
解析センター

地球環境科学



人間・社会科学  
データコモンズ

人間・社会科学



埋蔵化学  
物質 DB

その他

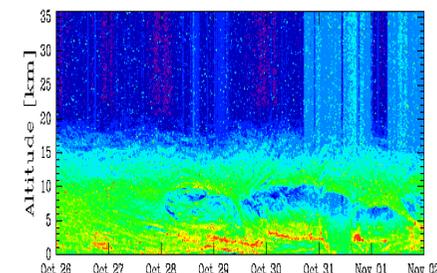


# データ基盤整備

Date: October 28, 2014 - November 01, 2014  
Beam:1, Channel:1

## 地球環境データ

- PANSY観測データの国内へのリアルタイム転送とアーカイブのシステムを開発
- PANSY観測の多量のデータを高次解析処理して物理量データとして国内外に配信
- レーダーのデータ処理技術を駆使して、観測の品質を向上



## ライフサイエンスデータ (DBCLS)

- DB間のデータ関連付けのための用語の整理と共通化 (オントロジー整備)
- 多様なデータの表現方法の国際標準化を進め、格納する仕組みを構築
- データ利用を効率的に検索・表示するためのツールの開発と公開



## 人間・社会科学

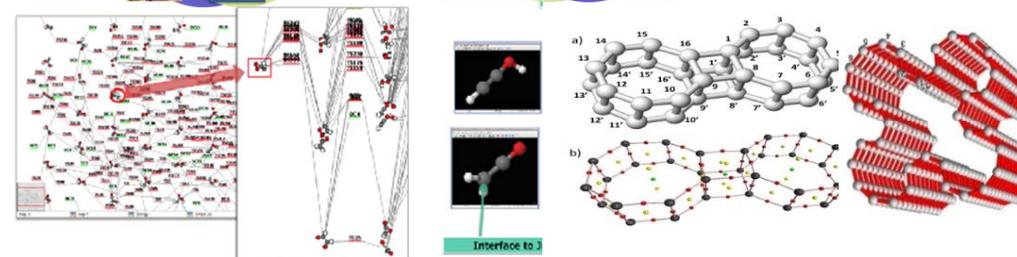
人間・社会データの収集・分析・活用基盤

- 公的統計データの二次利用オンサイト拠点の拡充
- 公的統計とWeb情報の統合
- ライフログデータ、モバイル統計データ



## データ中心ケミストリ

- 理論的に存在しうる化学物質と化学反応経路を、量子化学の理論に基づき、探索結果を解析するためのソフトウェア RMapViewを開発、公開
- 理論的に存在しうる新規炭素物質 (埋蔵分子) を発見



# モデリング・解析基盤整備

## データ同化・シミュレーション

- シミュレーションと観測データを統合するデータ同化手法の高度化
- 未適用分野の発掘と問題解決（地震による微気圧変動分析、細胞質流動を引き起こすせん断力分布の推定、感染症流行予測）

## e-サイエンスデータ基盤技術

- サイエンス3.0基盤としてresearchmap(ver 1)を構築
- Linked Academic Open Dataに関する日本唯一の研究プロジェクト
- 実文書の解析手法の開発とツール公開、言語処理学会等に全国大会論文閲覧システムを提供

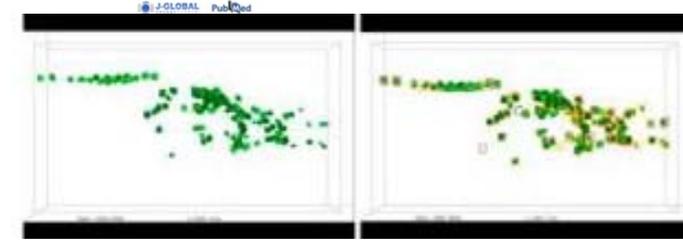
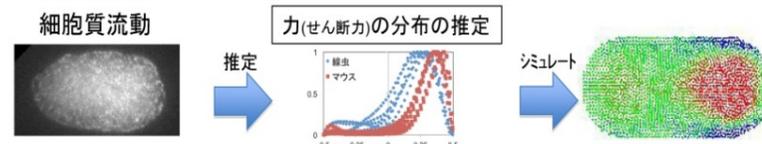
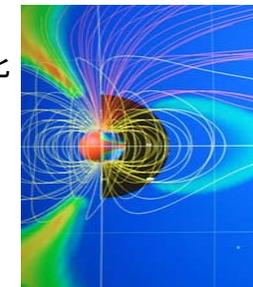
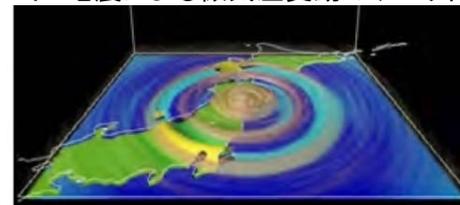
## イメージデータ解析

- バイオイメージング・インフォーマティクス
- オプティカルイメージングデータによる脳の動作原理の解明
- 人間・社会の視覚情報データ解析とMCDCツールの開発

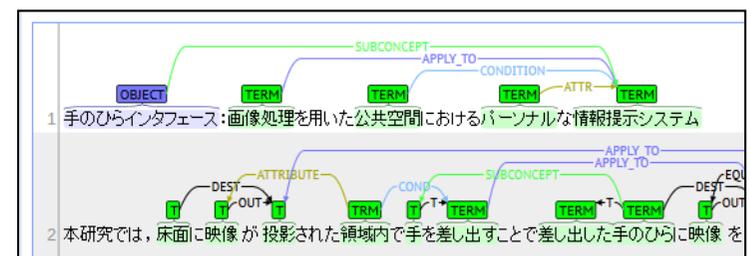
## メタ知識構造解析（データマイニング）

- メタ知識構造コーパスの開発
- メタ知識構造認識技術の開発、かしこい検索システムへの応用

右: オーロラデータ同化  
下: 地震による微気圧変動のデータ同化



SPF-CellTracker



メタ知識構造

# データサイエンティストの育成

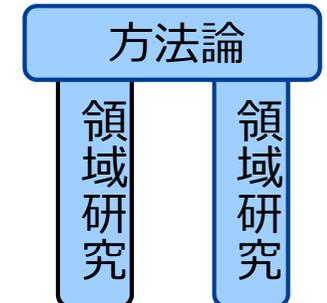
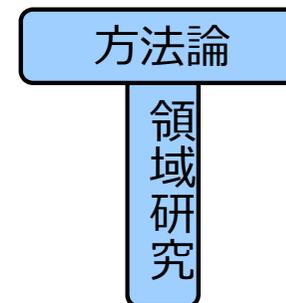
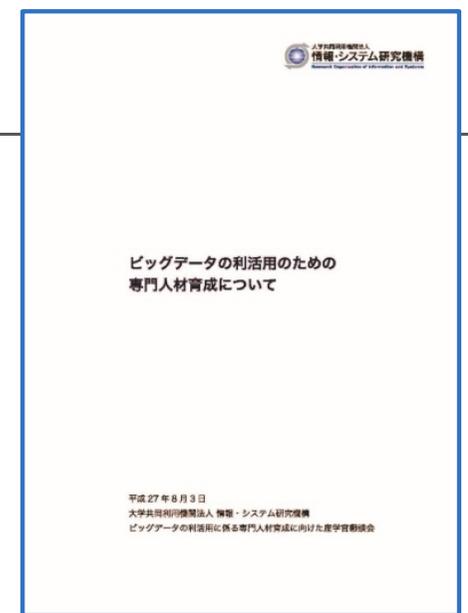
## ● データサイエンスと融合研究の推進に必要な人材

- データ解析、可視化，大規模データ処理
- 領域科学の知識と理解
- 課題設定，企画立案能力
- コミュニケーション能力（異分野交流）
- 研究倫理，個人情報保護

⇒ **T-型，Π-型研究者**

## ● データサイエンティストの育成方法

- **主専攻**：統計・数理・情報， **副専攻**：領域科学
- 領域科学の博士取得者の統計・数理・情報**再教育**

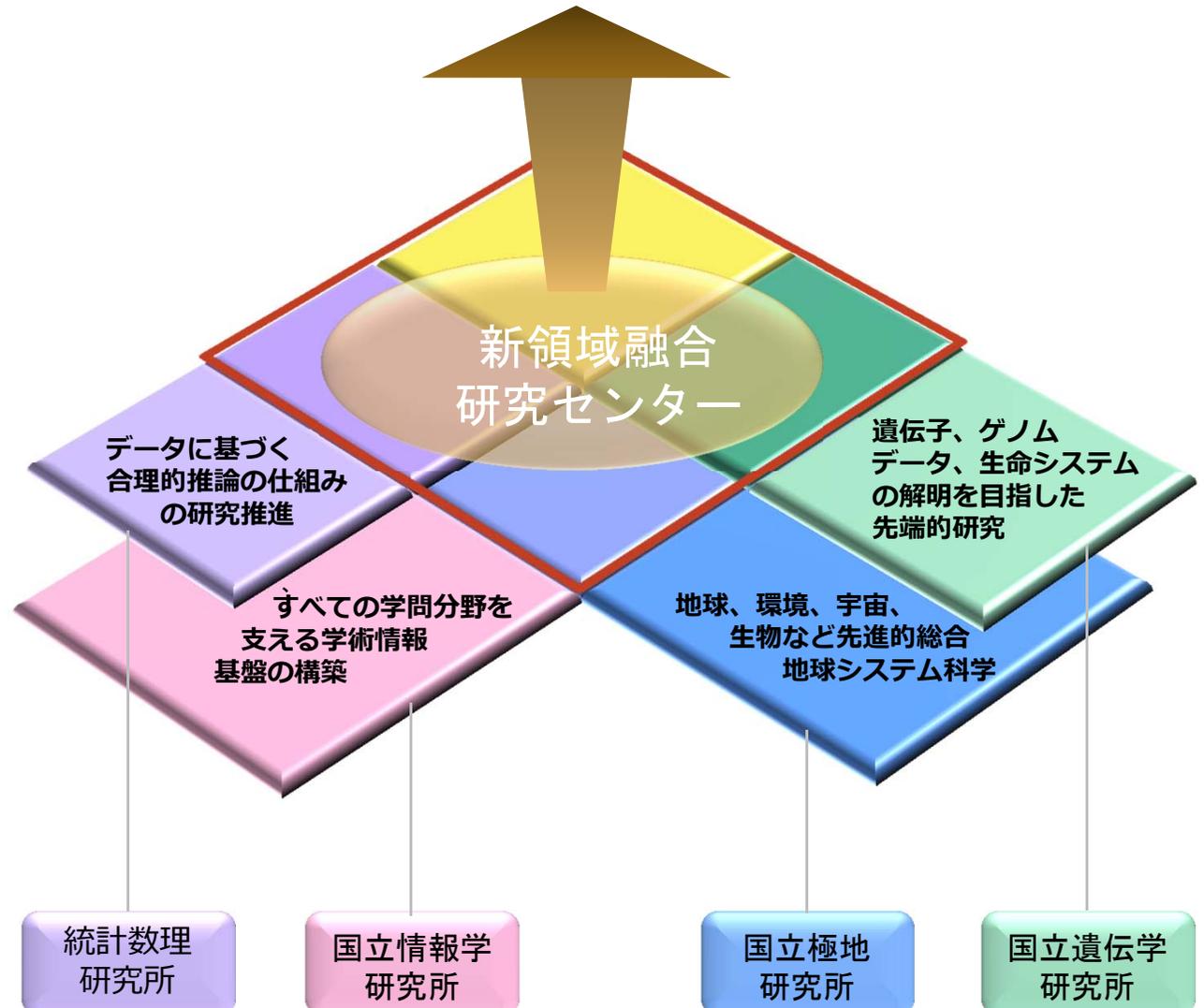


# 新領域融合研究センター

## 設置目的

極地研、遺伝研において得られる多種大量の地球科学・生命科学のデータ及び知見を**統数研**で開発されるモデリング・計算技術および**情報研**において開発される情報技術・情報基盤と結合させ、データとモデルに基づく真理の発見と予測の独創的な手法を生み出し、地球、生命、人間・社会システムの各研究分野において、**新しいパラダイムを創造**することを目指す。

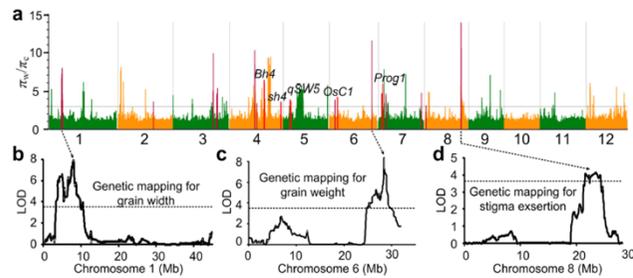
## 新パラダイム創成



# 新領域融合研究センターのプロジェクト

## 遺伝機能(生命)システム融合研究

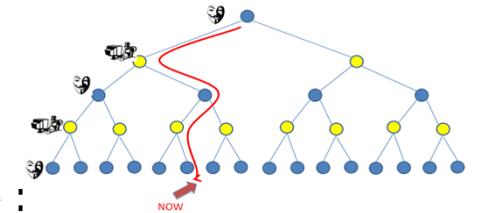
- ゲノム関連情報の大規模生産とその情報解析手法の開発
- 遺伝的相関構造描出のための統計手法の開発と最適化
- 大量で多元的なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出



## システムズ・レジリエンス融合研究

「想定外」の障害から柔軟に回復するレジリエントなシステムを設計・運用するための知識体系 (Body of Knowledge, BOK)の構築

堅牢なシステムからレジリエントなシステムへ

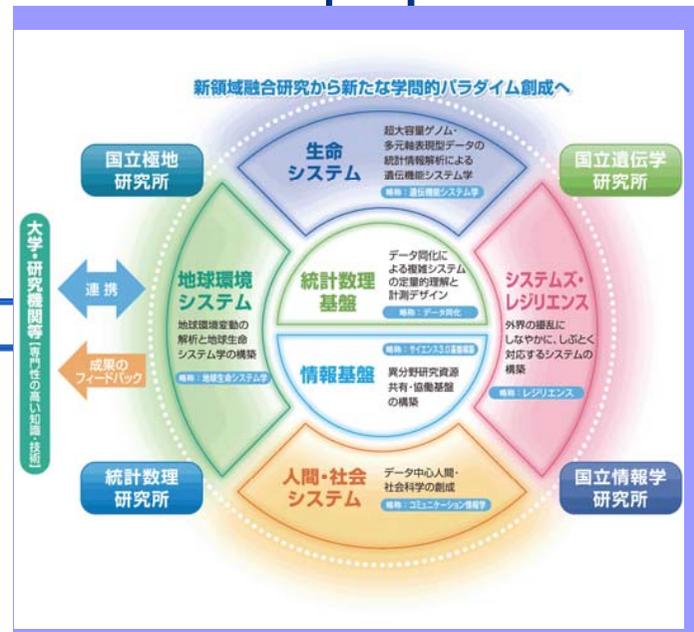
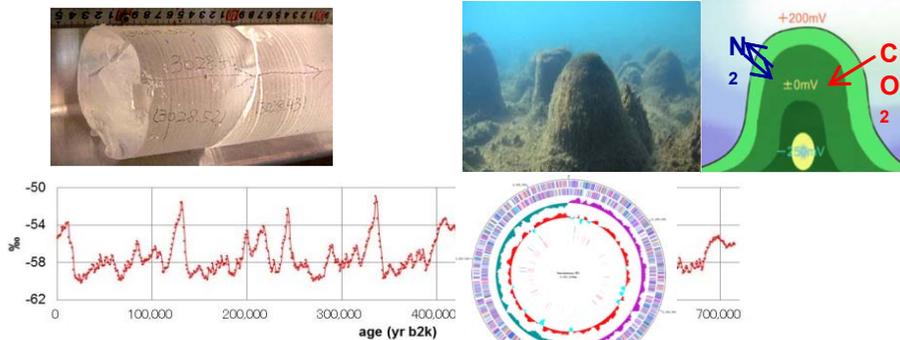


研究内容：

- 想定外の事象のリスクマネジメント
- 生態系におけるレジリエンス機構
- 社会システムにおける秩序形成
- システムのレジリエンス性評価の計算モデル

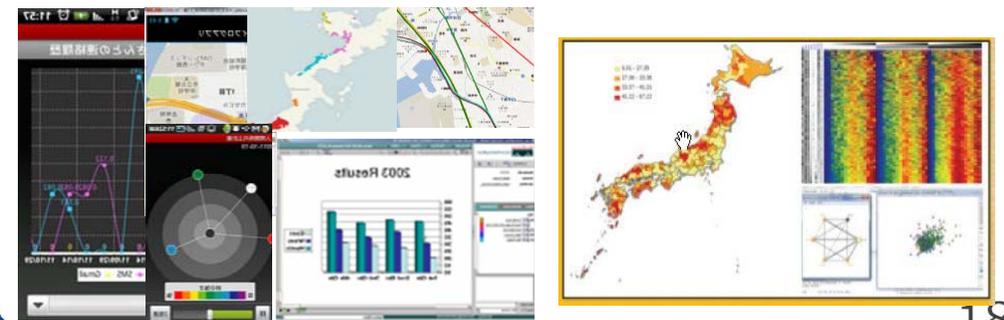
## 地球・環境システム融合研究

- 地球環境変動と微生物の進化・多様性の相互作用を解明
- 環境変動下での生命の適応戦略のメカニズムを解明



## 社会コミュニケーション融合研究

ネットワーク型の人間・社会データ共同利用・共同研究基盤の構築を通じて、人間・社会を適切なデータに基づいて、人やその集合体としての社会の相互情報コミュニケーションを理解・設計する新たな情報科学を創成。



# 新領域融合研究(1)

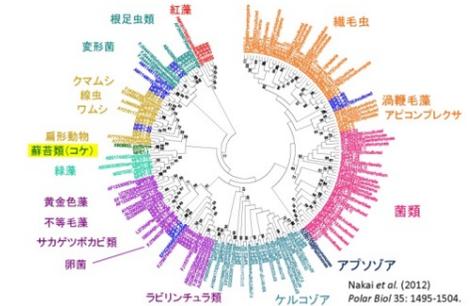
## 地球・環境システム融合研究

- 氷床コア中の微少なゲノム解析を行うための技術開発
- 氷床コア解析により氷期-間氷期の気温変動に硫酸塩エアロゾルが寄与していたことを解明
- 南極湖沼底の生物共同体「コケ坊主」の真核微生物コミュニティ解析

### 【H26共同研究先】

北海道大学、京都府立大学、玉川大学、東京工業大学、京都大学、広島大学、札幌医科大学、慶応大学、長浜バイオ大学、新潟大学、筑波大学、千葉大学、琉球大学、海洋研究開発機構、ノースウェスタン大学、復旦大学

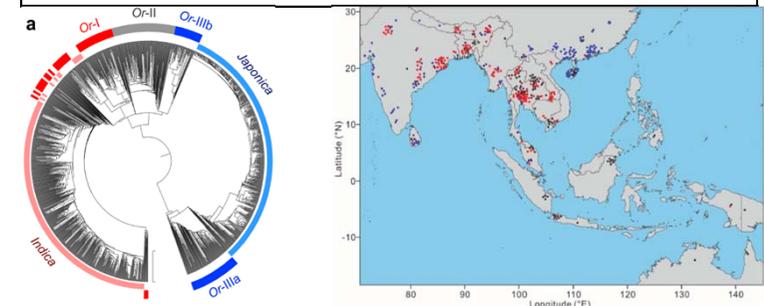
### 2. コケボウズ内の生物多様性の全容解明へ



## 遺伝機能 (生命) システム融合研究

- 次世代シーケンサによるゲノム関連情報の大規模生産とその情報解析手法の開発
- 大量ゲノム関連データと多様な生物表現型多様性データの統合による遺伝的相関構造抽出のための統計手法の開発と最適化
- 大量で多様なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出

野生イネ446系統と栽培イネ1083系統のゲノム情報を用いて、栽培化の道筋と栽培化遺伝子の選抜がもたらしたゲノム構造変化を明らかにした。

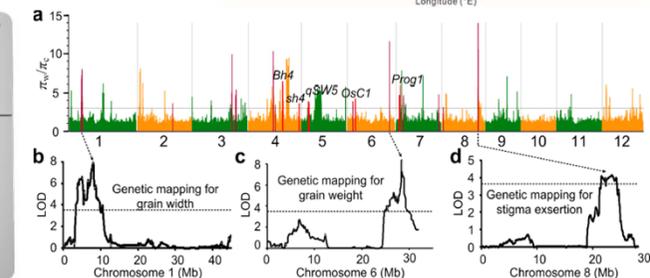


### 【H26共同研究先】

東京工業大学、京都大学、慶応義塾大学、基礎生物学研究所、理化学研究所、新潟大学、山形大学、大阪府立大学、九州大学、京都工芸繊維大学、首都大学、愛知工科大学、大阪府立大学



次世代シーケンサ



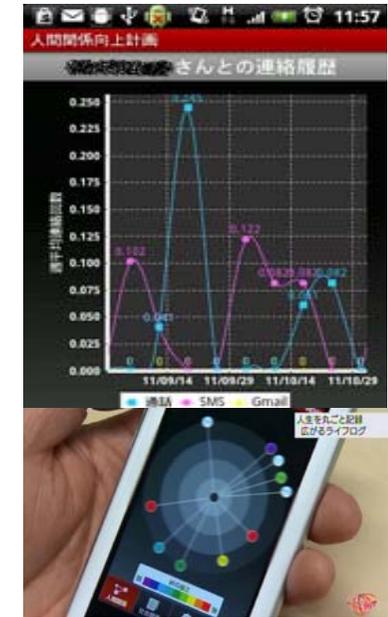
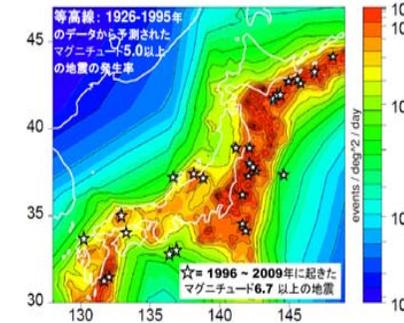
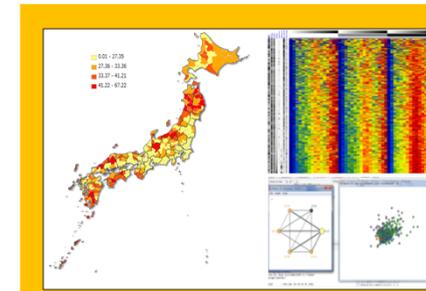
# 新領域融合研究(2)

## 社会コミュニケーション融合研究

- 公的統計データの二次利用オンサイト拠点の拡充
- 公的統計データの国際拠点構築、オープンデータ化
- 自殺予防のための時空間統計データ分析（公的統計）
- 観光政策支援システム（観光統計・Webデータ）

【H26共同研究先】

広島県、広島大学、山梨県、山梨大学、高知県、高知大学、高知工業大学、観光協会、UQコミュニケーション、NTT 等

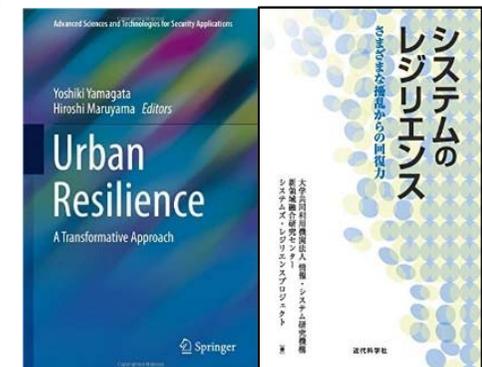


## システムズ・レジリエンス融合研究 —東日本大震災を契機とした取組

- 平成23年度に機構長裁量経費によりFSを実施、平成24年度から本格的な研究プロジェクトとして推進
- レジリエンス方略の計算モデルSR-Modelを提案
- 国際的な研究者ネットワークの形成
  - Int. Workshop on Systems Resilience (2013)、X-Center Workshop (2014), 「NII湘南会議」(2015)
- プロジェクトで得られた知見を体系化し、ハンドブック形式にまとめ、広く一般向けに発刊、国際シンポジウム成果を海外出版
- 今後の分野横断型研究の1つのモデルケース
  - 異分野の研究者が集まって全く新しい分野を切り開いた
  - 今後のレジリエンスに関する科学的解明や、減災や持続可能性のために応用が期待される



さまざまな分野の研究者が集まった湘南会議



# データサイエンスに関する国際的研究拠点の形成

- 平成25年度から「データサイエンス」を進めている主要海外研究機関を訪問し、**国際ネットワーク形成**に向けた事業連携等のための情報収集を行い、平成27年2月に欧州機関（フィンランドCSC，英国インペリアル・カレッジ・ロンドン（ICL））の研究所長及び研究者を招聘して「データ中心科学に関する国際ワークショップ」を開催し、CSC，ICLデータ科学研究所（DSI），EUDATと計3件の**MoUを締結**した。
- 平成27年度には、CSCとのMoUに基づくフォローアップミーティング実施のため、関連する研究者を海外に派遣し、同年度3月に、EUDATとのMoUに基づく**合同ワークショップ**「EUDAT/ROIS Collaborative Data Infrastructure Workshop」を東京で開催、共同データインフラのための国際協力について議論を形成するなど国際ネットワーク形成を推進した。



「データサイエンスに関する国際ワークショップ」



Kimmo Koski氏（CSC Managing Director, EUDAT Coordinator）とのMoU調印

# 法人中期目標期間と機構の取組み

## 法人第一期 (2004-2009)

新領域融合研究センター  
およびDBCLSを設置

機構内中心の融合研究と  
ライフサイエンス分野の  
データ基盤構築を推進

ライフサイエンス  
統合DBセンター

新領域融合  
研究センター

## 法人第二期 (2010-2015)

データ中心科学リサーチコモンズ事業に  
よりデータサイエンスの三位一体の基盤  
整備事業を推進

データ中心科学  
リサーチコモンズ事業

ライフサイエンス  
統合DBセンター

新領域融合  
研究センター



## 法人第三期 (2016-2021)

戦略部門を機能強化し、機構の強み・特長を活か  
した取組みを立案・推進できる仕組みを確立

大学等への支援事業（データ共有、データ解析、  
人材育成等）を中心とした活動を推進

### データサイエンス共同利用基盤施設

#### データサイエンス支援

- データ共有支援
- データ解析支援
- T型・II型人材育成

#### 戦略プログラム

- 戦略的研究ネットワーク形成
- 未来投資型研究プログラム
- 文理融合プログラム

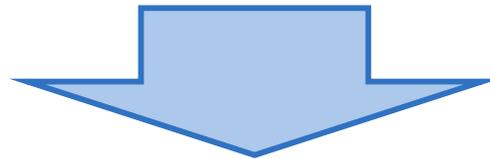
データサイエンスの観点か  
ら大学等の研究活動を支援

# 今年度からの取組み

---

## データ中心科学リサーチコモンズ（H25～H27）

- データ中心科学研究基盤構築
- 新領域融合研究



## データサイエンス共同利用基盤施設

- 5つのセンターを設置（準備室を含む）
- データ中心科学リサーチコモンズ事業の成果と経験を活かして大学等の研究者の**データサイエンスを支援**



# データサイエンス共同利用基盤施設（H28に設置）

**大学等へのデータサイエンスに関する支援事業**  
（データ共有、データ解析、人材育成）を推進

## データ共有支援

ライフサイエンス統合データベースセンター

地球環境データサイエンスセンター（準備室）

社会データ構造化センター

人文学オープンデータ共同利用センター（準備室）

## 戦略プログラム

### ● 戦略的研究ネットワーク形成

- 国際会議、MOU推進等
- データサイエンスの公募型共同利用
- 研究者交流事業
- 文理融合研究プロジェクト

### ● 未来投資型研究プログラム

### ● 文理融合プロジェクト

## データ解析支援系

ゲノムデータ解析支援センター

データ融合計算支援プロジェクト

データサイエンティスト人材育成

