

# ビッグデータ時代に対応する人材の育成

## 日本学術会議 情報学委員会 E-サイエンス・データ中心科学分科会(2014年9月)

1. はじめに
2. 海外の動向
3. 我が国の現状と人材育成に関する課題
4. ビッグデータ活用に必要な要素技術と人材育成
5. ● 提言1 データ中心科学を専門とする教育組織の設置
  - 提言2 基幹的研究組織内における恒久的なデータ解析部門の設置
  - 提言3 日本版インサイト・プログラムの早急な設置
  - 提言4 データサイエンティストの資格の制定



## ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会 (2015年7月)

1. データサイエンティスト育成の必要性と我が国の課題
2. 我が国におけるデータサイエンティストへの要請
3. データサイエンティスト人事育成のあるべき姿と実現に向けた仮説
4. 具体的施策
5. ● 提言1 500名規模の「棟梁レベル」人材育成とトリクルダウン
  - 提言2 主要10大学程度での人材育成による大学教育加速
  - 提言3 全学的教養教育の実施, 国家レベルフラグシッププロジェクトの推進



31

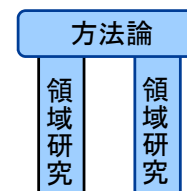
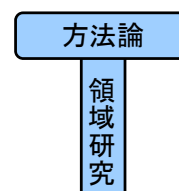
# データサイエンティストの育成方法

## ● データ中心科学と融合研究の推進に必要な人材

- データ解析、可視化、データ処理、モデリング、知識発展の方法
- 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: データサイエンス(統計数理), 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

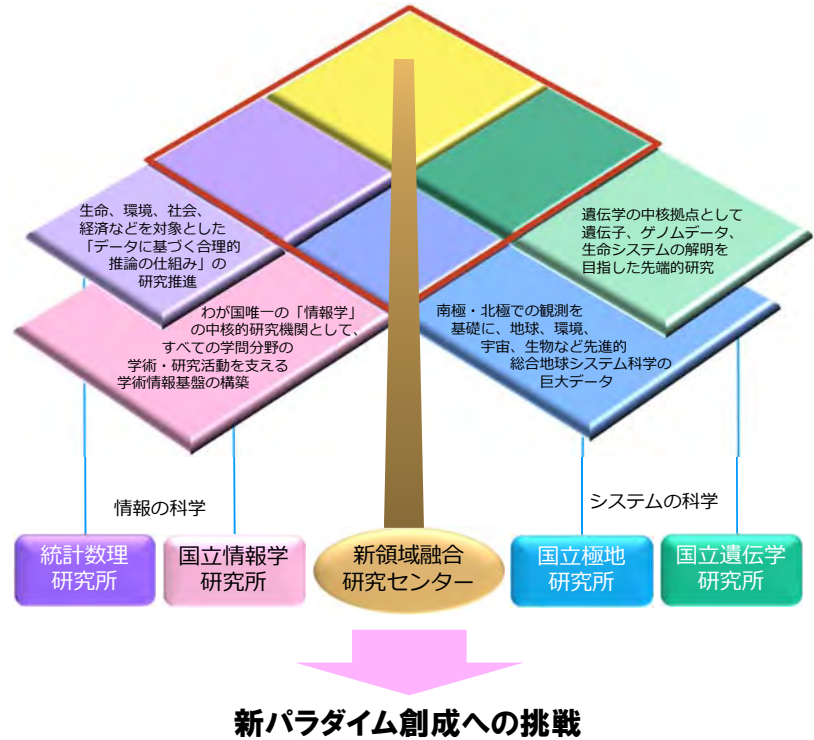
32

# 新領域融合研究センター

## 運営概念図

### 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**することを旨とする。

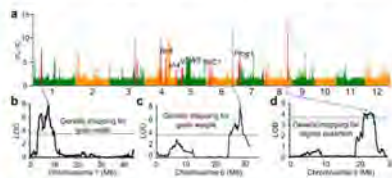


33

# 新領域融合研究センターのプロジェクト

## 生命システム学

- 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

「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

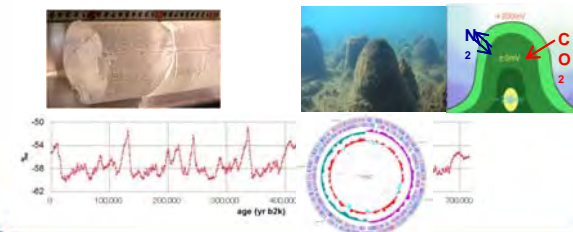


研究内容：

- 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

- 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



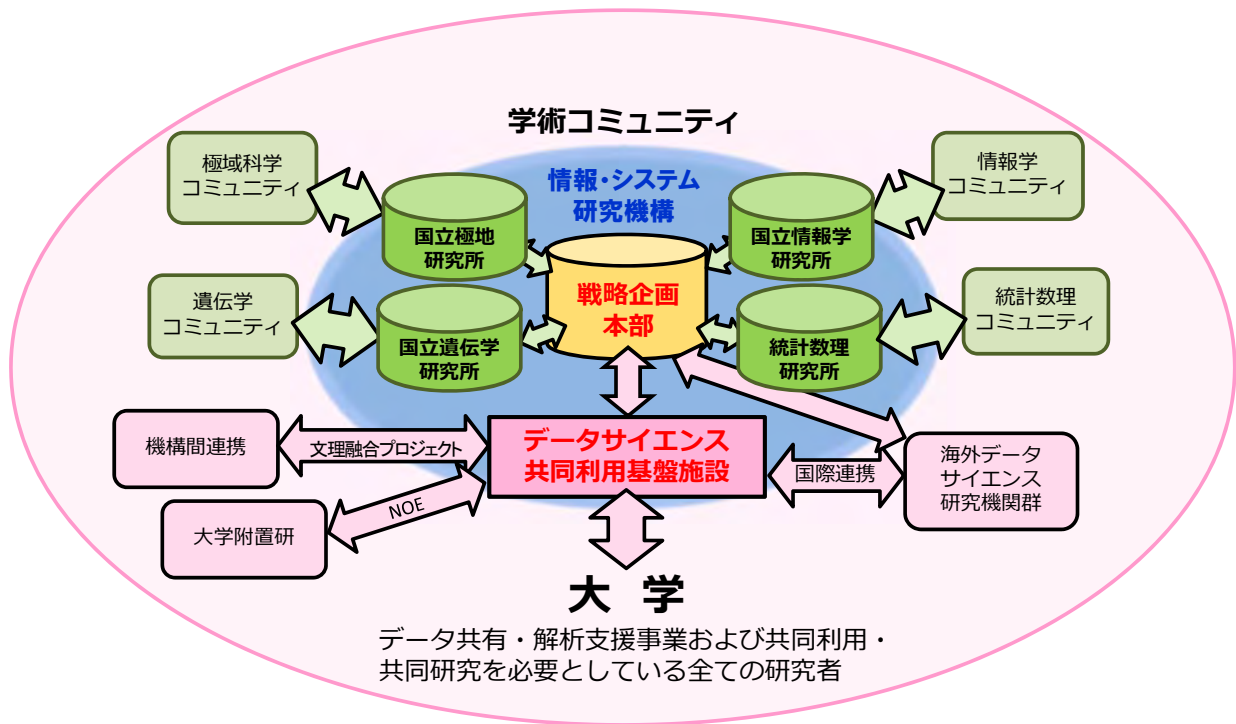
## データ中心人間・社会科学の創成

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。

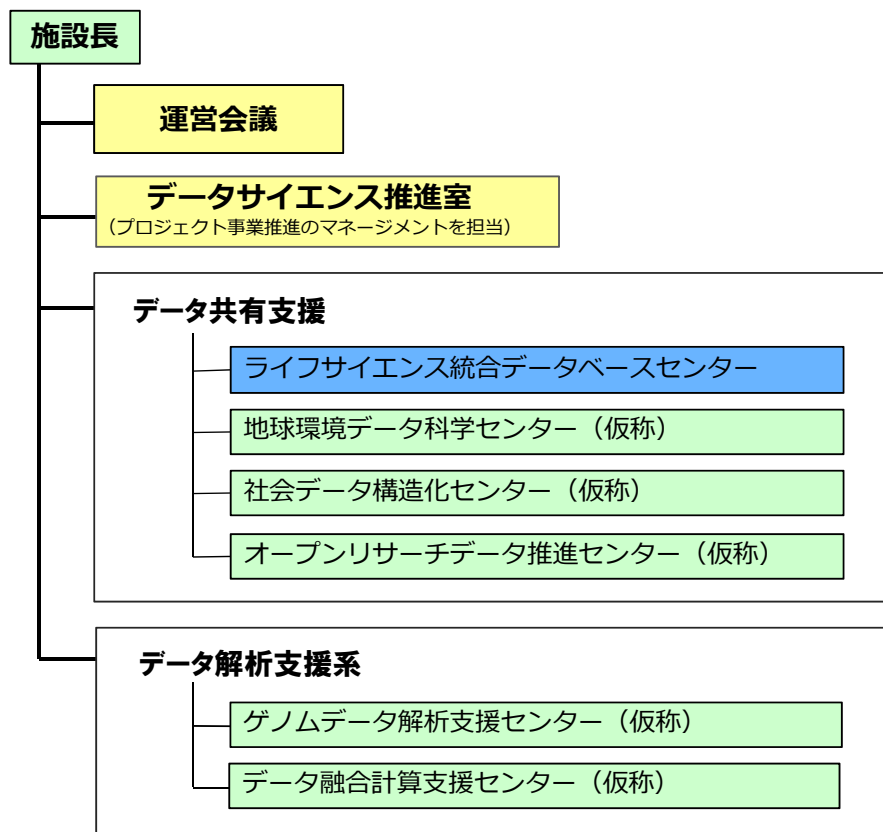


34





# データサイエンス共同利用基盤施設



# 横串を通すことの重要性

## データ基盤において

1. 汎化支援技術の活用する
2. ハブ機能を利用した知識移転が行える



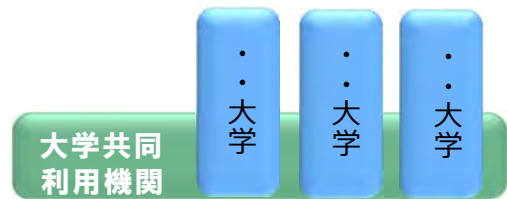
## モデリング・解析基盤において

1. 解析手法の汎化により本質を捉える
2. ハブ機能を利用した知識移転が行える
3. 新分野の創成（方法と領域のクロスで自然に新領域が成立する）

情報〇〇、数理〇〇、計量〇〇

## 大学間において

- 大学共同利用機関法人：  
コミュニティハブ（ボトムアップ研究で分野別の学術研究を支える）  
DS共同利用基盤施設： 各大学に**分野を越えて**横串を通した支援
- 研究開発法人：イノベーションハブ（トップダウン研究を推進）



39

## 現場の研究者の意見 ～情報・システム研究機構シンポジウム パネル討論から～

### 1. オープンデータに関する研究者の意識

- 研究者からすると「**下手なデータ公開はできない**」という気持ちがあり、こだわりがOpen Dataの邪魔をする
- データを公開する時期が重要。データが旬のときがベストといっても、そのときは研究者としては公開したくないとの気持ちが入り、obsoleteになってから公開ということになりかねない。自分で**公開時期を正しく判断することは難しい**。
- 研究者にとってデータの管理は大変。DBに入れて安全に管理されるのであればインセンティブになる
- データジャーナルにデータを出して、先のキャリアにどう活きるかのパスが見えない

### 2. データの加工者の貢献を正當に評価する仕組み

- データの**中間加工をしている人への正當な評価**が重要。前処理コストという言葉があるが、Big Dataの90%は前処理

### 3. 公開されるデータの特性について

- 論文の**仮説が通るようなデータだけが公開されて、バイアスがかかる**恐れ
- **失敗データも公表される**ことが重要

40

### 1. データサイエンス・オープンサイエンスにむけて

- 三位一体の基盤構築
- 横串を通すことの重要性

### 2. データサイエンティスト

- 育成体制（棟梁レベル）
- キャリアパス構築（評価・人事）
- 研究機関、研究プロジェクトへの配置

### 3. ジャーナル

- データジャーナルよりむしろジャーナル自体の変革

### 4. 研究コーディネーションのマシン化