



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 (DS)



AMIDER

ROIS-DS共同研究

機構賛同型コンソーシアムで接続された研究機関・大学との連携によるデータ駆動型研究の促進

阿部修司(九州大学国際宇宙惑星環境研究センター)
共同研究メンバー

1. 共同研究について

AMIDERとは



AMIDERは、以下を提供・提案することを目的としたプロジェクト

1. 異種・異分野データを統合的に扱うデータベースの研究開発（技術観点）
2. 異種・異分野データを融合させて新しい知見を得る手法の研究開発（技術観点）
3. データサイエンスを全国で展開するための役割や分担の素地の策定（運営観点）

詳細、現状は
次の発表にて

これまでの成果

- 2016-2017年度 機構公募案件「機構未来投資プロジェクト」にて活動。多分野普遍型フレームワーク構築。
- 2018年度 AMIDER発足。極域データや古書籍データなどを1つのカタログDBへ纏める。
- 2019年度 共同研究設立。極域データ間(+α)の関係性の数値化による検索結果の満足度向上。
- 2020年度～ 展開に向けた準備。まずは研究促進を主軸に。

本共同研究について

これらの活動を進めている中で

「他の研究機関・大学の研究者が、こういった取組に興味を持っているようだ。」

「一体で推進していくために連絡会のようなコンソーシアムを作って欲しい。」

という声を聞き、今回のような共同研究を進めることとした。

- ・2019年度「データ駆動型研究を促進させる賛同型コンソーシアム形成に係る基盤研究」
- ・2020-年度「データ駆動型研究促進のための機構賛同型コンソーシアムによる研究機関・大学との接続」

2. 活動内容の紹介

概要

- 本共同研究で構築したコンソーシアムにより機構と各研究機関・大学を接続し、高度科学を一体体制で推進する試行(思考)をおこなう。
- データサイエンス基盤(ビジョン、データ変換ノウハウ、品質基準等)をコンソーシアムを通じて各研究機関と共有し、各機関におけるデータ業務そのものとの合流し、高度科学考察基盤へのフィードバックを通じて、データサイエンスをステージアップさせる。

背景と目的

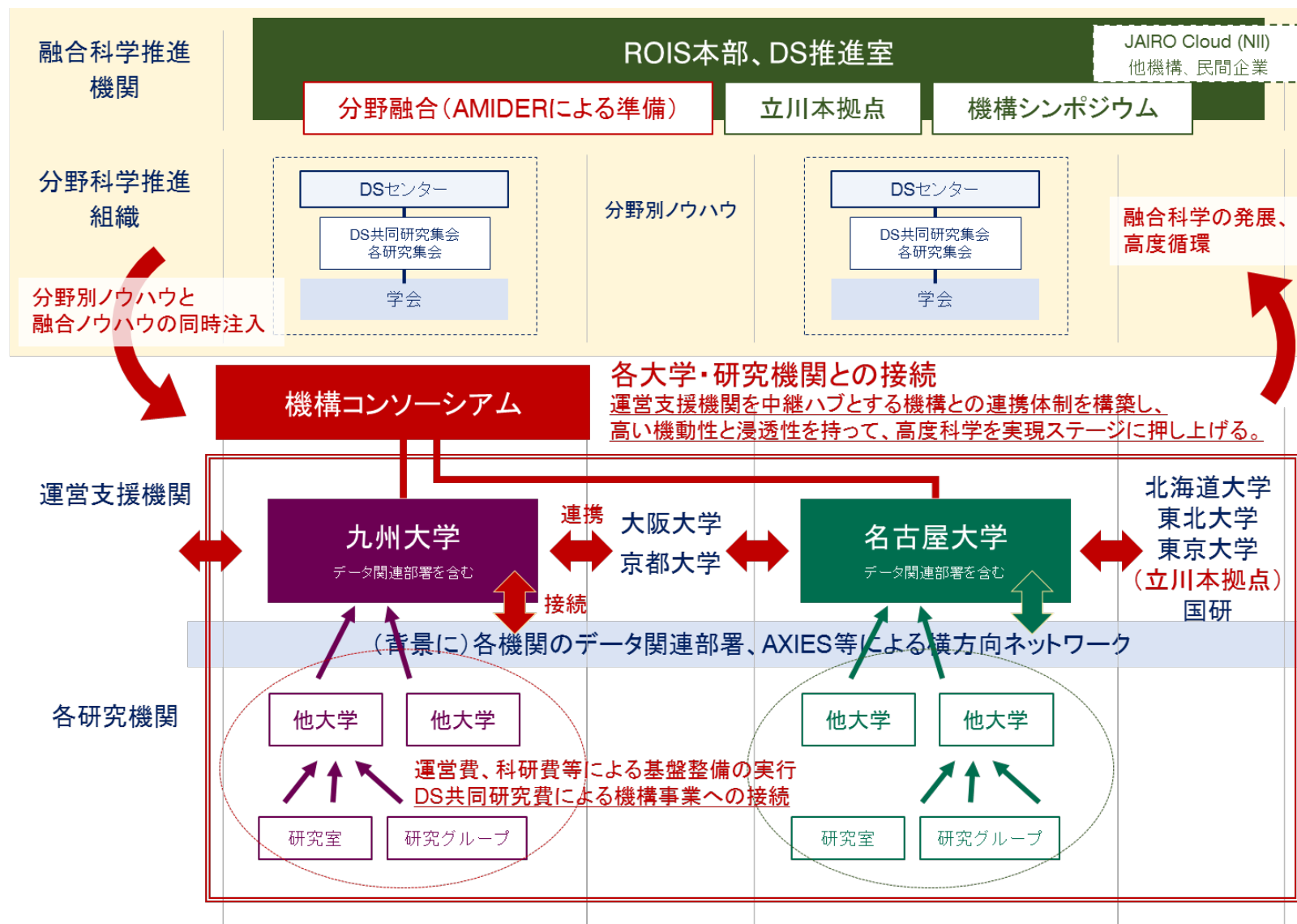
- 主要な研究機関では、データ関係部署の新設やデータポリシーの策定など、データサイエンスに関わる具体的な活動を急ピッチで進めている。しかしながら、これらの活動は、機関使命の観点から、研究データの保全と公開までを目的としているのが殆ど。
- 分野融合や新しい科学の創出といったビッグビジョンを共有し、新しい科学の到来を一体の体制で実行レベルに引き上げることが最終目的。また、機動性の観点からは、既存の横連携と同様の拠点体制を導入し、その相互連動によって、主要機関との接続のみならず、中小機関への波及も高める。

期待される効果

- 機構と研究機関の接続は、各所の中期計画や政府によるSociety5.0の実現に向けて、バックボーンとして機能することが期待される。
- 各機関においても、データ整備活動の本質は新しい科学や価値の創出という部分にあることを見出して底上げされることが期待できる。
- この一体の取組により、データサイエンスを、アカデミック全般でのビッグウェーブとして実行フェーズに移していくことが期待できる。

2. 活動内容の紹介

データサイエンスに関わる機関一体型の第一次運営モデルの構築



2. 活動内容の紹介

活動内容

- ◆ モデル検討と試行、整理
- ◆ 検討した推進モデルを対象の外部機関と共有し、試行して、以下の向上を図る。模擬実行とエラー出し。

1-a. ビジネスパートナー（データ事業者）

□ サービスのワンストップ化のための提携案の策定

- JAIRO Cloud、各大学および研究機関のデータ関連部署、データサイエンス教育強化拠点コンソーシアム等が該当

1-b. プロモーター（データ従事者）

□ 運営課題の選定とその相互解決と実現：

- 共同研究によるデータ融合活動を全体俯瞰して、運用を現実のものにするには5W1Hにどのような修正が必要かを提案してもらう。もらった改善サイクルを実施していく。

2. データプロバイダー

□ データの試験提供：

- 共有可能な研究データを用い、構築中のシステム上で処理して、他の研究データとの関連性などを算出。それがどのような発見に繋がるか、また、自身の今後の研究にどのような発展をもたらすかの提言。

3. ユーザー

□ データ駆動型システムの試験利用とその効果化：

- システムでの処理結果を実際の研究・教育活動に試験適用し、効果測定の試行（融合科学らしく異分野のデータと組み合わせがより良い）
- 具体的な利用場面の選定（＝ニーズ）、提案

2. 活動内容の紹介

活動内容

共同研究における前2ページ3種の割当

1. 運営者

- a. ビジネスパートナー: NII JAIRO-Cloud、数理・データサイエンス教育強化拠点コンソーシアム等の事業者
- b. プロモーター: 各機関・組織、図書館、研究グループ、希望従事者等

2. データプロバイダー: 極域データ(宙空圏、気水圏、地圏、生物圏)を取り扱う研究者

3. ユーザ: 項番2と同様

氏名	所属	1-a ビジネス パートナー	1-b プロモー ター	2 データ プロバイダー	3 ユーザー
阿部修司	九州大学国際宇宙惑星環境研究センター		○		○
小財正義	ROIS-DS 極域環境データサイエンスセンター		○		○
白井知子	国立環境研究所地球システム領域		○		
福田陽子	国立環境研究所地球システム領域		○		
北村健太郎	九州工業大学工学研究院宇宙システム工 学研究系			○	
古賀崇了	近畿大学産業理工学部情報学科		○		○
才田聡子	北九州工業高等専門学校情報システムコース				○
新堀淳樹	名古屋大学宇宙地球環境研究所				○
田中良昌	ROIS-DS 極域環境データサイエンスセンター		○	○	
...	その他、多数の協力研究者				

3. これまでの成果

1-a: 組織に対する活動の例

区分	相手先	テーマ
ビジネス	RCOS (NII)	事業提携の可能性の検討。データアーカイブとデータ融合サービス。
ビジネス	RCOS (NII)	提携を視野に入れてのデータフローの検討。
プロモーター	JpGU教育担当	教育活動への適用。
ビジネス	RCOS (NII)	キュレーションフローに関する基本役割の確認。
プロモーター	高専教員	既存体制＝高専ネットワークとのタイアップ検討。
ユーザ	大学・高専教員	研究データ整備に関する現場負担の実状の共有。
ビジネス	RCOS (NII)	メタデータフロー(サイエンス、アーカイブ)。研究実データの出し入れ。
プロモーター	AXIES	キュレーションの役割分担。
ユーザ	木質分野研究者	解析データの分類など、将来やりたい研究について。
ビジネス	RCOS (NII)	研究データの出し入れに関する進捗。
ユーザ	PI、大学・高専教員	棚卸し。実現していく事項の確認。
プロモーター	名大図書館	機関リポジトリとデータ活動の棲み分け。
プロモーター	機関データ担当者	外部機関のデータを使った融合計算の試行。
ビジネス	RCOS (NII)	キュレーションネットワーク構想。
ユーザ	PI、大学・高専教員	DX(デジタルトランスフォーメーション)再考。
プロモーター	名大図書館	データ活動とアーカイブのビジョン接続。
プロモーター	九大図書館	同上
プロモーター	機関データ担当者	テスト実行結果の戻し、DX精度向上。
ユーザ	PI、大学・高専教員	DX効果の確認
など		

3. これまでの成果

問題提起とそれに対するアプローチの例

提起	活動内容
地区ごとに拠点を設ける方がPIや研究者は動きやすい	オンライン(ハイブリッド)の敷居が下がったことを利用して、バーチャル拠点の開設を検討。共同研究打ち合わせもオンラインを主に進める。
コロナを機に、議論の場がますますオンラインになっていく。しかし分からないことをすぐに聞ける体制が好ましく、引き続き地域ごと(近いところ)に聞けると機動的	実態のある拠点体制は維持していく。
研究室単位では困ったことを解決するには小さい	他のネットワーク研究プラットフォームにて試験的なフォーラムを開設してみる(但し、このフォーラムは、まずは科学ディスカッションをする場として機能させる)。
データがラベル付けされていると、教育活動に導入しやすい	メタ情報からの単語抽出 and/or 実データ間の相関から、データを区分け＝クラスタリングしていくことを検討(この分布とラベルを対応付ける)。
DX(デジタルトランスフォーメーション)の深掘り検討。一例として、専門外分野のデータから事象を把握するにはどのような見え方が親和的か？	全体把握は遷移図のような俯瞰図、データごとにはExcelグラフのようなわかりやすいものが捉えやすい。→システムにおける可視化表現の多様化
データ整備、メタデータ作成のうち、サイエンスに関わる部分の体制をどうするか？	特定分野にフォーカスしたデータキュレーターを(学会等から)各機関に配置して、業務を集約的に遂行する「メタデータキュレーションネットワーク」の構想を検討。
リモート教育の普及により、研究室単位ではインフラの取り回しがますます困難となっている	観測・解析装置貸出に類似したような、計算機・ネットワーク貸出が考えられる。共同利用の役割に資するため積極的に考える。
機関としての業務は研究データの保全を主としており、新しい科学を切り開く手法開発は興味がある	構築中のシステム上に入れて処理。それを実際に利用してもらう。

3. これまでの成果

データ公開サービスフローに対する思考

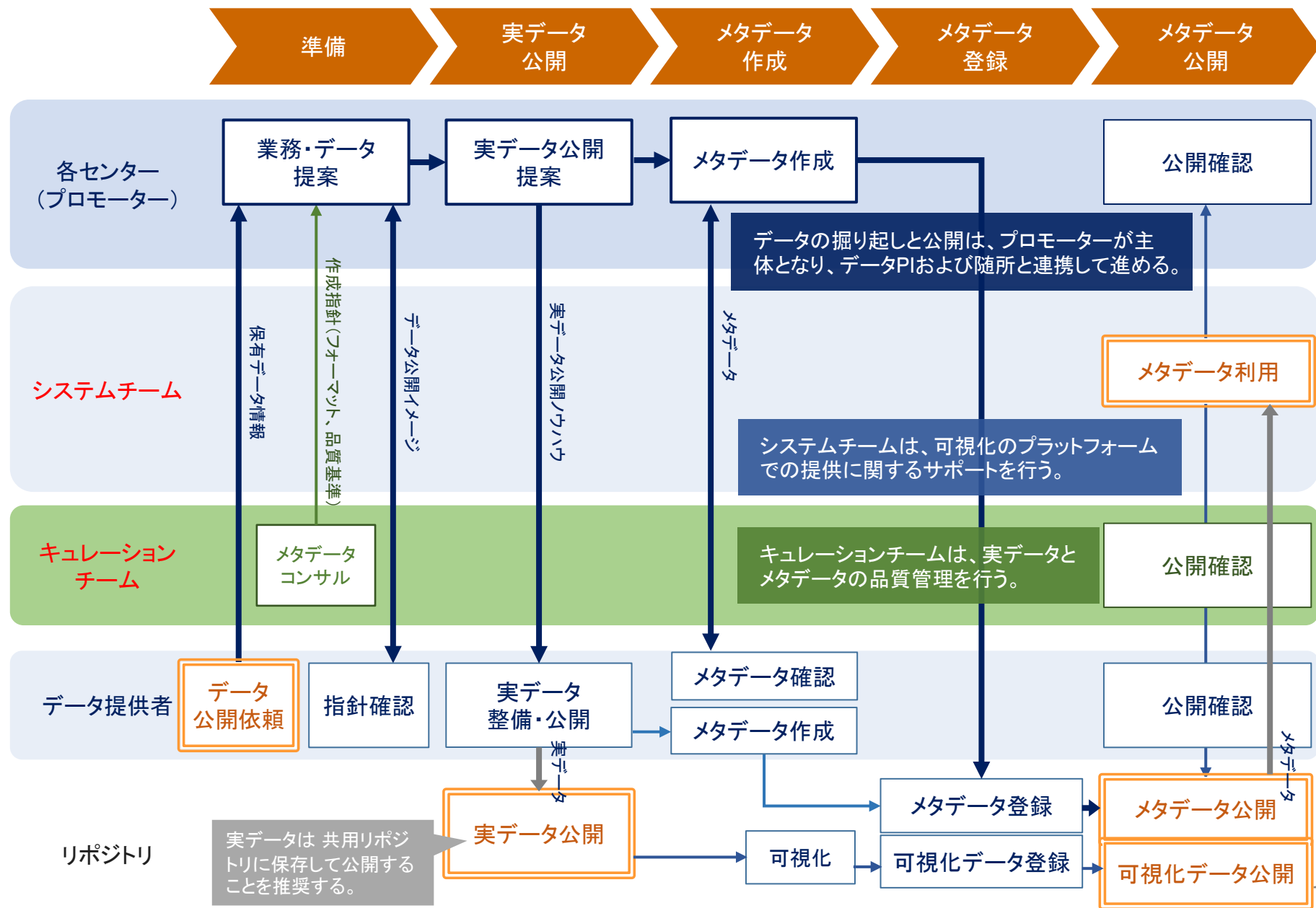
- 所有データ:
 - 関係者間公開中の観測データ(共有)。データポリシーの問題あり。
 - 未処理のデータ。公開範囲や時間分解能を考慮。
- 組織の連携:
 - 研究者の連携というよりは、機関間の連携、教育が重要。
 - ボトムアップは難しい。
- データの公開希望
 - 希望あり。ただし、公開のためのデータ処理等、マンパワー的に困難。
- 実データ公開について
 - インフラ(サーバ)を誰が用意するのか、という問題がある。
 - 基本的には、研究者が独力でやるしかない。
 - 個人のデータを公開する際のポリシーは明確でない。
 - 共用リポジトリでの公開はどうか？
 - 学校単位か、高専機構全体で契約するのか？(他の連携機関では既に契約しているとの話あり。)
 - 研究者間で共同研究を結んでデータ公開する方が簡単。組織全体で決めるとなると大変。
 - 公開リポジトリには、複数の選択肢が考えられる。

データ公開サービスフローに対する思考

- 実データ公開について(続き)
 - 公開リポジトリには、複数の選択肢が考えられる。
 - 実データのための公開用リポジトリを持っていない組織もある
 - メタデータの公開については、相談先がある
 - 所属組織から、あるサービスを使って公開することが推奨されているが、年によって契約が変わる。セキュリティの観点から、自前で勝手にサーバを立てての公開は不可。
 - 工学の実験データの公開の希望。情報系では、機械学習のためのラベル付きデータを公開したいという希望もある。
- 実データのフォーマット
 - 観測機器の吐き出すバイナリ、分野に特徴的なASCIIや自己記述型フォーマット、等
 - ある程度の変換はシステム側でも対応可能(後述)
- 誰がメタデータを作る?(自力?)
 - 図書館にてサイエンスメタデータを作ったり提案するのは難しい。
 - アーカイブ用のメタデータを作る必要がある、という認識は持っている。

3. これまでの成果

検討されるデータ公開サービスフロー(案)



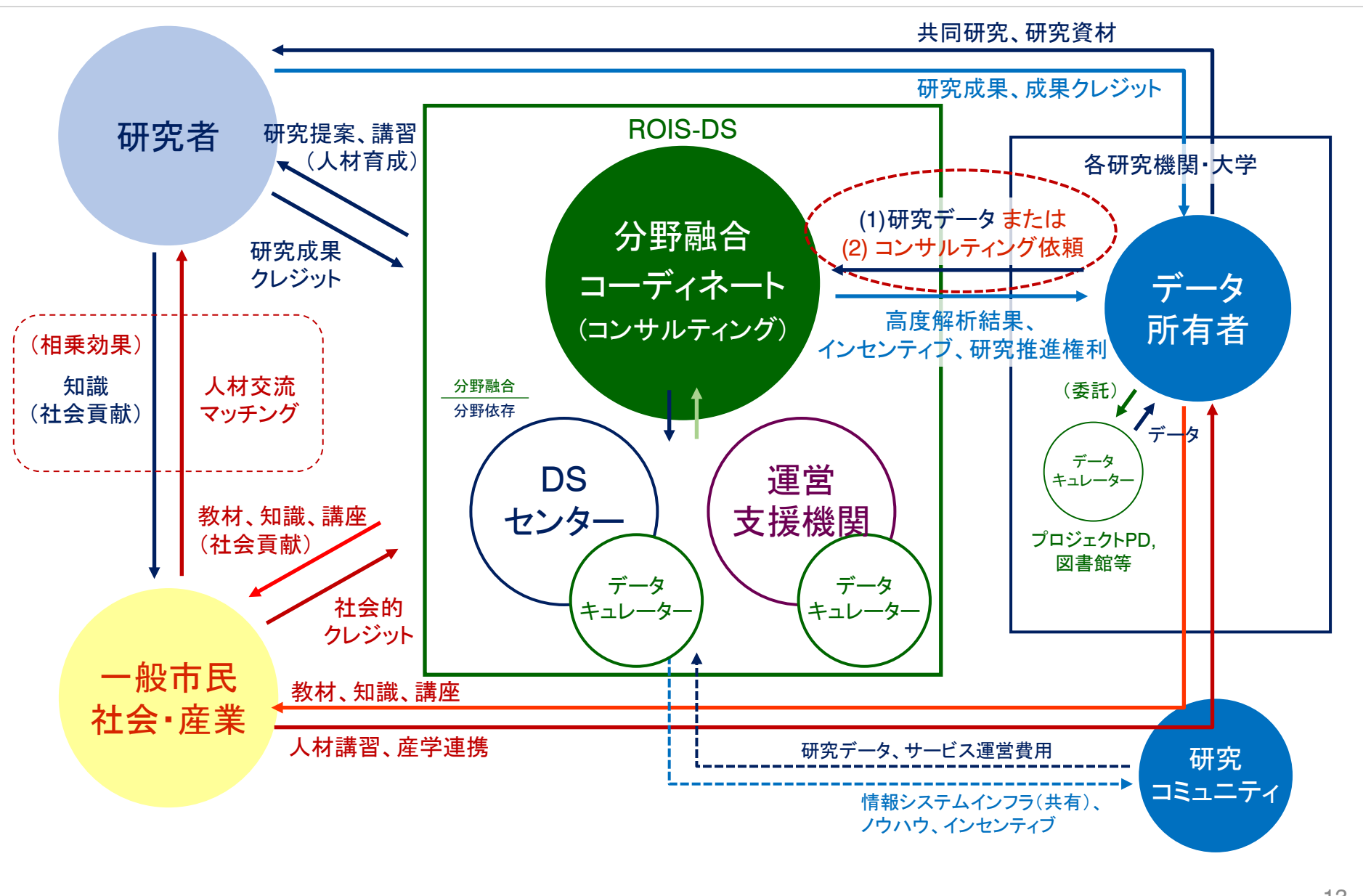
3. これまでの成果

プロモーション推進モデルに対する思考

- 現状、個々の組織でデータサイエンス推進活動は容易か？困難があるか？
 - データ公開について図書館に相談したが、やり方がわからなかった。
 - 研究者のニーズはあるが、所属機関にはその仕組みはない。
 - そのための予算も必要。
 - こちらから所属機関側のメリットを説明する必要がある。
- 解決策は？
 - データ公開について、相談できる部署があると良い。
 - データ公開の啓蒙活動が必要。
 - 機関内において、プロモータを育成。
 - FD(faculty development)活動等でデータ公開を紹介して、すそ野を広げる。

3. これまでの成果

検討されるプロモーション推進モデル(案)



3. これまでの成果

データ利用に対する思考

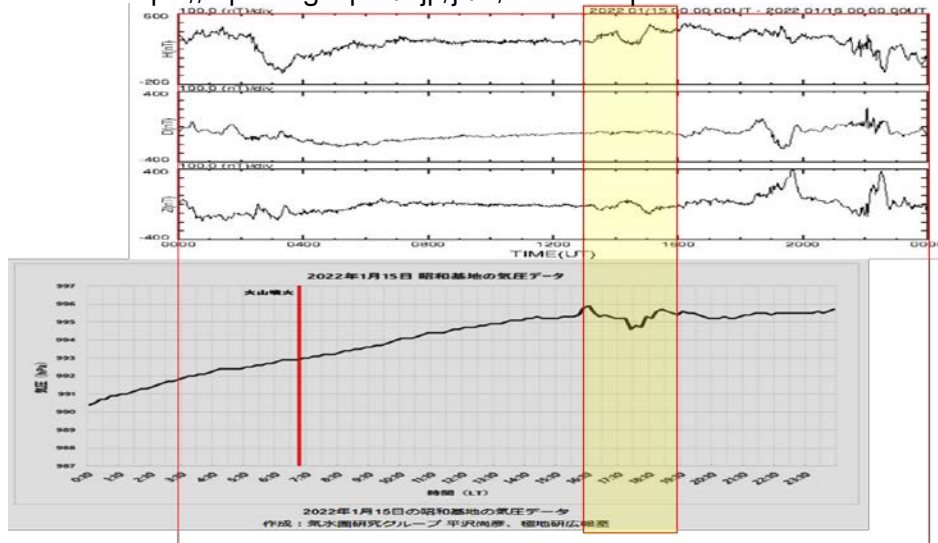
- 構築中システムのウェブUI操作レビュー、ユーザー視点での要望収集
 - リンクに対するポップアップ機能
 - APIの仕組み
 - フォーマット変換
 - ラベル付きデータ(機械学習への利用)、説明付きデータ(解釈への利用)の準備
- 2022年にフンガ・トンガ噴火
 - 日本上空でも2hPa程度の変動の波が伝搬するのが見られた
 - 気圧データの変動を九大(別の先生が置いていた試験機器にて)で捕捉。
 - 1チップの気圧計+RasPiで記録。
 - 宇宙線データの補正に気圧データを取得している
 - 昭和基地ではインフラサウンドのデータを取得している
- 複数のデータを使う際、サイエンス目的に対して、その専門外のデータの意味、重要性がわかりづらい
 - 各自が出せるデータを持ち寄って、ここで見比べてみる
 - 可視化の様式は各自に任せる

3. これまでの成果

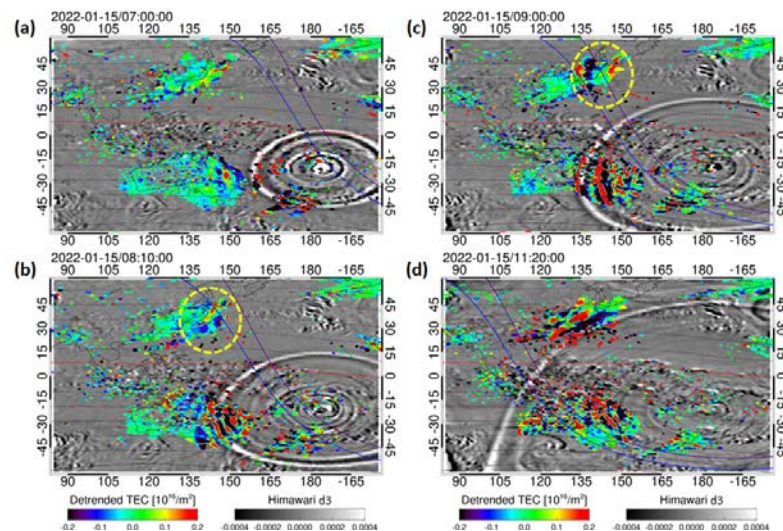
・昭和基地での気圧変化と地磁気変動

微気圧変動データ

<https://nipr-blog.nipr.ac.jp/jare/20220118post-203.html>

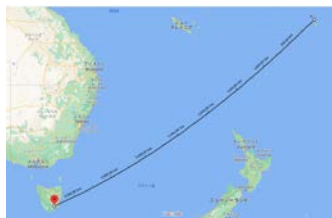
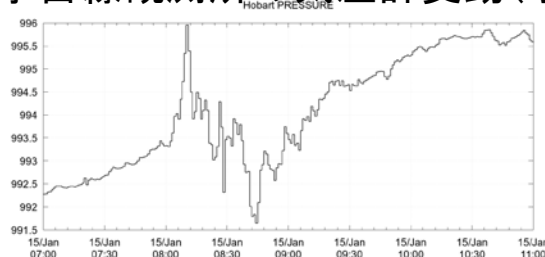


・人工衛星による下層大気擾乱とTEC変動



<https://earth-planet-space.springeropen.com/articles/10.1186/s40623-022-01665-8>

・宇宙線観測所の気圧計変動(オーストラリア)



トンガ噴火の影響が見えなかった例：
CO₂観測値

- ・長寿命でバックグラウンド濃度と突発的な濃度上昇の差が小さい
 - ・半球跨ぐ間にシグナルが薄まってしまう
- 等が原因と考えられる

ビジネス：データ活動の提携

- 多数の機関、大学組織等との意見交換。それぞれの教育活動とも連動していく必要性（機関連携と人材育成の連動）

プロモーション：データ事業の運営

- ビジネスモデルと業務フローの検討
- システム、機関リポジトリの稼働開始を受けて、一部機関とのジョイントの試行を始める
- データキュレーションを担うサイエンティストの確保。組織的な稼働に向けた検討

データプロバイダー：高度科学研究の芽生え、新しい知見の獲得

- フィードバックされる結果（見え方と使い方）を検討
- データ活動の体制は、既存のネットワーク体制と相乗して縦横方向に進める

ユーザ：教育、高度人材育成、研究

- インフラ（ハードサービス）は、共同利用機関がその提供を広げていくことも利点がある
- 教育推進（ソフトサービス）は、各拠点および中枢のデータサイエンティスト＝研究者を想定
- データ提供にあわせた特性や科学的背景の共有が、研究者ユーザーとしての統合解析には必要