

ROIS-DS共同研究#018RP2020



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

データ駆動型研究促進のための機関・大学間型
コンソーシアムによる研究機関・大学間の接続

代表

梅村宜生(名大ISEE)

副代表

阿部修司(九大ICSWSE)

機構担当者

田中良昌(ROIS-DS)

2021年1月14日「極域データの保全・公開と利活用に関する研究集会」

1. 活動内容の紹介

共同研究申請書から抜粋

概要

本共同研究では、**機構事業に賛同する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こと**を目的とする。

機構が進めてきた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、機構による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一気に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である。

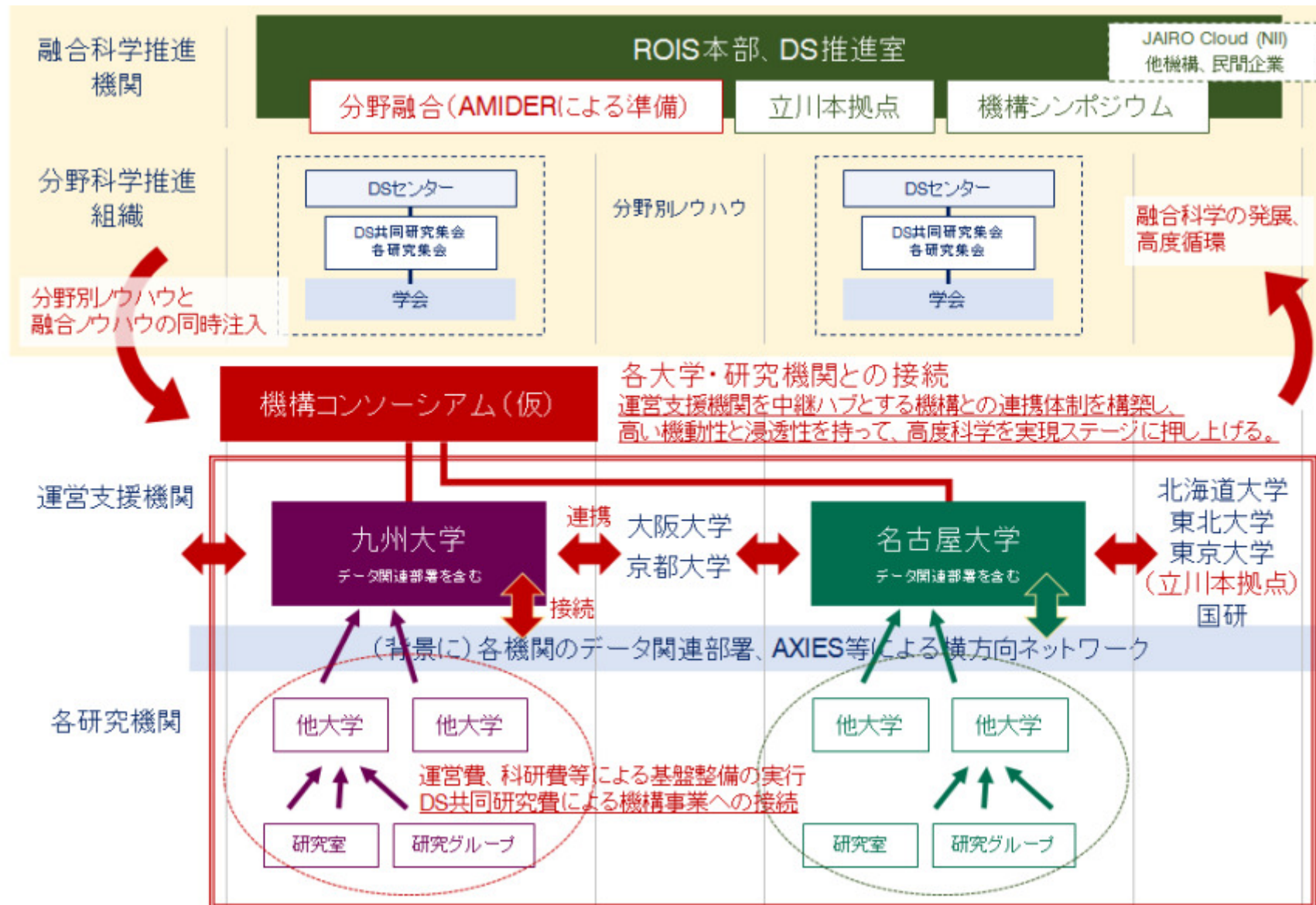
機構との接続により、その先にある分野融合や新しい科学の創出といったビッグビジョンを共有し、新しい科学の到来を一体の体制で実行レベルに引き上げることを目的とする。また、機動性の観点から、AXIESや図書館等の既存の横連携と同様の拠点体制を導入し、その相互連動により、主要機関との強固な接続のみならず中小機関への波及も最大に高めていく。

期待効果

機構と研究機関の接続は、機構によるジョイント構想、次期中期計画、および、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。**

共同研究のゴール

機構と調整しながら、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を築く。



1. 活動内容の紹介

活動内容

What: モデル検討と試行、整理

推進モデル2018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): AMIDERが随所と折衝。

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: AMIDER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ご提案ください。AMIDERはその改善サイクルを実施していきます。また、各種研究集会やカンファレンスに参加して、データ業界の動向を共有させてください(一緒に進めていくことに意義。特に前頁にて*1を付した方。)

2. データプロバイダー

データの試験提供: お持ちの研究データを貸してください。AMIDERで融合計算してみて、他の研究データとの関連性をお見せします。それがどのような発見に繋がったか、また、ご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をお願いします。

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: AMIDERによる融合計算を実際の研究・教育活動に試験適用し、効果測定を行ってください。融合科学らしく異分野のデータと組み合わせをお勧めします。具体的な場面の選定(=ニーズ)もご提案ください。

2. これまでの成果(一部抜粋)

(2019年度)

No	共同研究での提起内容	決定内容
1	地区ごとに拠点を設ける方がPIや研究者は動きやすい	拠点体制を導入し、2020年度以降も継続する。
2	研究者自らがデータのフォーマット化やメタデータ作成をするのは手間が掛かる。手伝える体制はないか？	データ整備等に関わるキュレーション体制をどう構築するか？ a) 図書館 実データやメタデータを定常運用(保管運用)するには適任。 b) サイエンスに関わる部分の体制は、学会など分野に精通した組織と総合的に作っていくのはどうか？
3	研究室単位では困ったことを解決するには小さく、機構やIUGONETのようなところに頼りたい。	IUGONETでフォーラムを試験的に開設する。
4	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	カテゴライズに加え、データを分類＝クラスタリングしていくことを検討。この分布とラベルの対応付けもする。
5	異なる機構法人配下の機関との連携の試行。	環境研メンバーを追加した。
	など	

(2020年度)

No	共同研究での提起内容	現時点での検討
1	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？DXの深掘り。	全体把握は遷移図やネットワーク図のような俯瞰方式、データごとはExcelグラフのようなシンプルなものが見えやすい。可視化ソフトの選定へ
2	(前年度No.2の続き)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？	他のデータ従事グループも同じ課題を認識している。学会等を主とする「メタデータキュレーションネットワーク」の構想も新たに検討に加えている。
3	試みに計算してもらいたい。	やってみましょう。2月に結果を渡す予定。
4	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。分からないことをすぐに聞ける体制が好ましい。	拠点体制を維持していく。
5	リモート教育の普及により、研究室では限りあるインフラの取り回しが困難となっている。リモート環境があると助かる。	極地研の解析装置の貸出しをヒントに、共同利用機関からのクラウド型計算機の貸出しもひとつに考える。
6	修士・博士課程の活動で使った研究データをきちんとアーカイブするために大学の機関リポジトリに預けたい。	図書館と一緒に進めていきましょう。
	など	

3. 今後の展望 (and discussion)

全体感

データサイエンス活動への関心は日に日に広がっている実感がある。機構と各研究機関・大学等との連携の素地もできつつある。

ビジネス データ活動の提携に関わる部分

1. データアーカイバーとの業務提携に向かう可能性もある。中長期的には。
2. 大学等の教育活動とも連動していく可能性がある。機関連携と人材育成の連動。

プロモーション データ事業の運営に関わる部分

1. ビジネスモデルと業務フローは大まかには成立しつつある。
2. 機関リポジトリの稼働開始を受けて、一部の機関とはデータフローおよび活動フローの接続の試行を始めている。未接続の機関もビジョン共有によりステージングされる日も近い。
3. データキュレーションを担うサイエンティストの確保は引き続きの課題。各機関にとり機構はone of them であるが、1つ実績が作られることで、その組織的な稼働が開始されることを期待する。(例えばAXIES)。

データプロバイダー 高度科学研究の芽生え、新しい知見の獲得に向けて

1. フィードバックされる融合計算結果(見え方と使い方)は精査されつつある。研究に有益であるとの意見が増えている。
2. データ活動の体制も、既存の図書館ネットワーク(AXIES)や高専ネットワークのような既存体制と相乗して縦横方向に進めることが効果的であり、それを推奨する方向。

ユーザ 教育、高度人材育成

1. インフラ(ハードサービス)は、機構のような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。
2. 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。

データサイエンスは科学の変革とも言われている。機構や我々の取組みが全てでないし、将来はどんどん変わっていく。今できる "100分の5" をきちんと実行に移していけばよい。