



# 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

## - 事例紹介 -

### ROISデータサイエンス事業の取り組みに おける図書館員の役割

梅村宜生[1]、田中良昌[2][3]、中野慎也[4]、南山泰之[3]、阿部修司[5]

[1] 名古屋大学宇宙地球環境研究所

[2] 情報・システム研究機構

[3] 国立極地研究所

[4] 統計数理研究所

[5] 九州大学 国際宇宙天気科学・教育センター

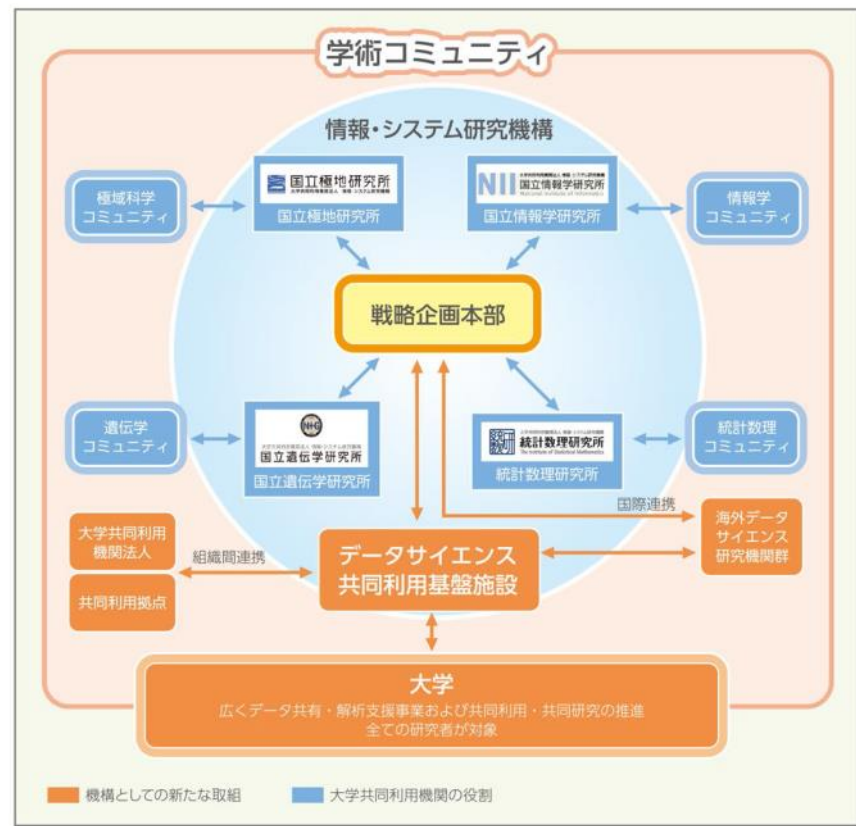
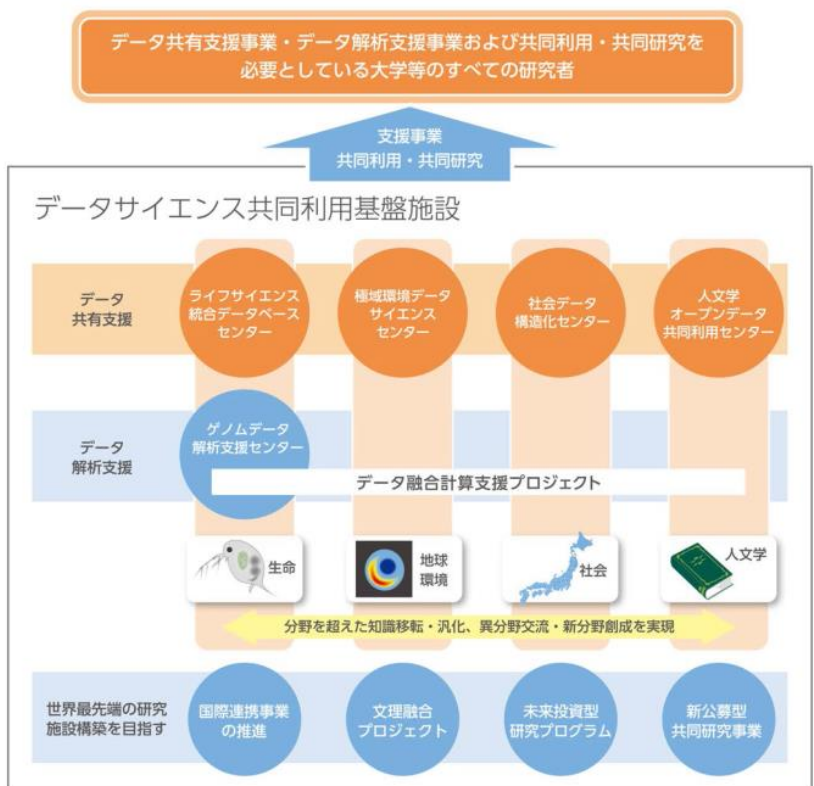
Joint Support-Center for Data Science Research

## データサイエンス共同利用基盤施設

大規模データから高度な解析により科学や社会の課題解決をめざす 融合的な共同利用・共同研究拠点

### 三位一体の活動により、大学のデータ駆動型研究を積極支援

- 1 データ共有支援 ..... 生命、地球環境、社会、人文学における大規模データの共有支援
- 2 データ解析支援 ..... 最先端の解析支援、分野を横断する高度な数理的手法の開発と支援
- 3 データサイエンスのための人材育成 ..... 大学等におけるデータ駆動型の研究と異分野交流・新分野創成の促進



- ライフサイエンス統合データベースセンター ..... 生命科学分野のデータベース統合化によるオープンサイエンスの推進
- 極域環境データサイエンスセンター(平成29年度 設置予定) ..... 極域の環境データに関するデータベースの整備及び共同利用の推進
- 社会データ構造化センター ..... 社会調査データ、公的統計マイクロデータ、ソーシャルビッグデータのデータ共有の推進
- 人文学オープンデータ共同利用センター(平成29年度 設置予定) ..... データサイエンスに基づく人文学(人文情報学)の創生及びデータのオープン化の推進
- ゲノムデータ解析支援センター ..... バイオインフォマティクス技術を駆使したゲノムデータの解析支援
- データ融合計算支援プロジェクト ..... データ融合計算技術を駆使した諸科学・産業界における課題の解決支援

## これまでの歩み

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)

戦略部門を強化し、当機構ならではの取組  
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積  
極的に推進

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

データサイエンスの観点から  
大学等の研究活動を支援

「大学におけるデータ駆動型学術研究力強化のための共同利用推進事業」年次計画

年度 (西暦)		平成28 2016	平成29 2017	平成30 2018	中間 評価	平成31 2019	平成32 2020	平成33 2021	
データサイエンス 共同利用基盤施設		各分野別センター設置により、DB基盤を形成					第4期に向け、統合化を開始		
データサイエンス支援	データ共有支援	生命科学分野	DB統合化への新規データの 産出機関の参加を増強				中間評価を経て、DS施設内 機能ユニット（センター） の事業内容・実施体制を見 直し、必要に応じて再編  分野横断型統合DB 複数分野統合検索機能  分野を超えた、 知識移転・汎化、異分野 交流・新分野創成の実現  手法の統合化、高度化		
		地球環境分野	支援体制整備 (IUGONET,PANSY)	各種データアーカイブ、 統合データベース化					
		人間社会分野	構造化データの 設計と構築	構造化データ の運営と改良					
		人文学 オープンデータ	支援体制構築	データ利用基盤構築、 内容分析、オープン化					
	データ解析支援	ゲノムデータ 解析支援	先進ゲノム支援課題 を対象に解析支援	大学等への公募、選定、実施 OJT訓練生受入れ等					
		データ融合計算 支援	支援体制 構築	諸科学・産業界でのデータ融合計算支援に 関する相談受付、共同研究の実施					
戦略プログラム	分野横断型研究 ネットワーク形成	データサイエンスに関する国際ワークショップ開催、MoUに基づくインターンシップ生の受け入れ等							
		制度設計、 公募準備	大学等への共同利用公募			制度の再検討、大学等のニーズに 合った共同利用			

学問変化に柔軟かつ戦略的に対応できる大学共同利用機関法人へと進化  
↓  
大学共同利用の新モデル

データ統合化

ROIS機構長、DS施設長



プロジェクトリーダー  
田中良昌(極地研究所)



データ融合スペシャリスト  
中野慎也、上野玄太(統計数理研究所)



データ融合アルゴリズム  
開発チームリーダー  
阿部修司(九州大学)



システム開発  
チームリーダー  
梅村宜生(名古屋大学)



メタデータ作成・キュレー  
ションチームリーダー  
南山泰之(極地研究所)

外部  
アドバイザー

データ融合  
アルゴリズム開発

システム開発

メタデータ作成・  
キュレーション

検証・フィードバック

## 📄 プロジェクトタスク(～FY2019テスト活動)

---

極地研が所有する多種多様な極域データ(宙空圏、気水圏、地圏、生物圏)を対象に、

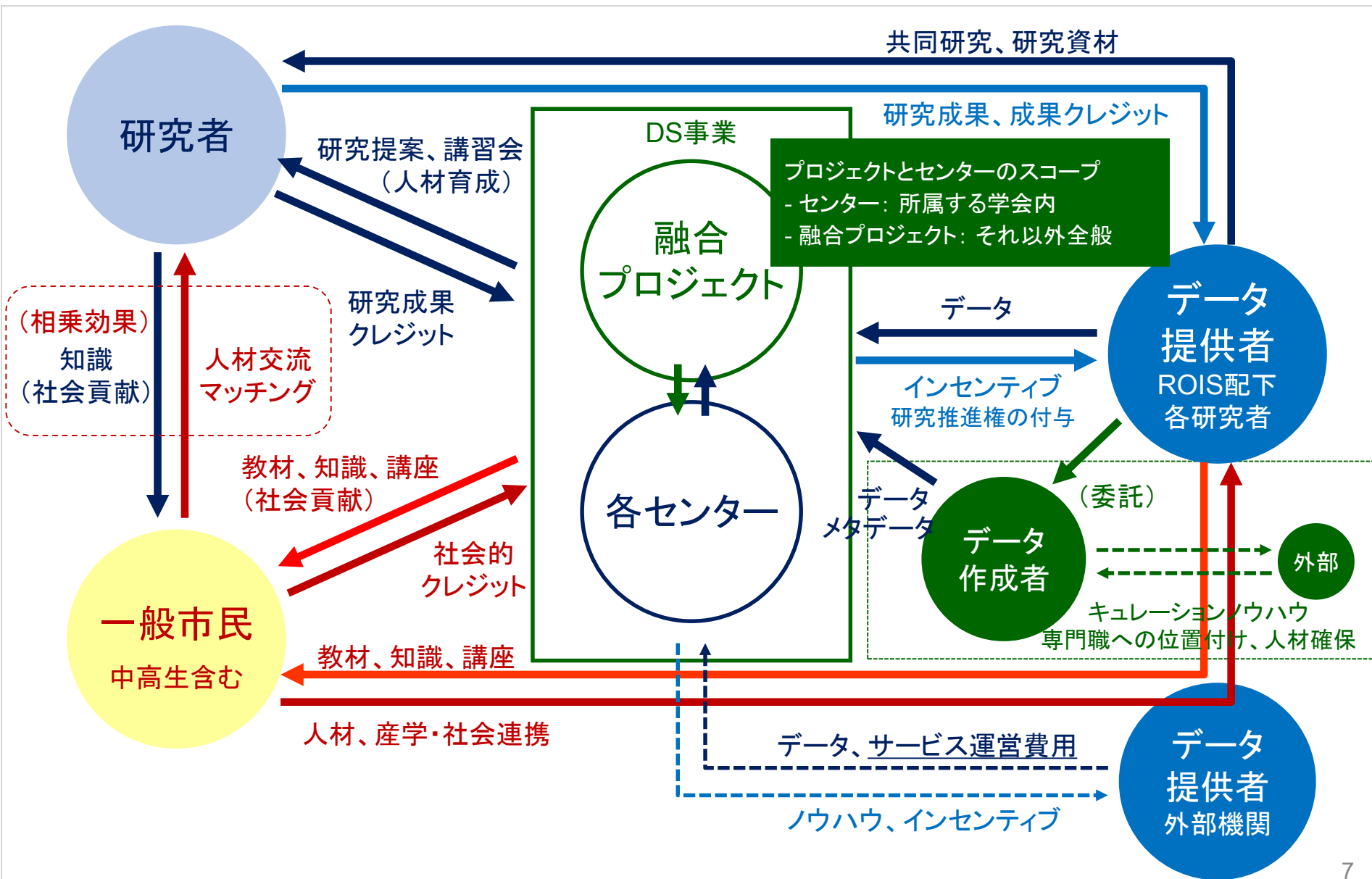
- 公開済データ(のメタデータ)の質の均一化と底上げを狙う。未公開データについては予め品質ポリシーを導入して作成する
- メタデータ、及び、実データを統合的に扱うことができるデータベースを開発する
- 上記の異種のデータ間の関連性を計算機により導出し、「データ駆動型科学(Data-Driven-Science)」に繋がるデータ融合の手法を確立する
- 2020年度以降に実施するROIS全体および全国大学展開に向けた総合システム(体制・サービス・活動)のスマールモデルを作り検証する

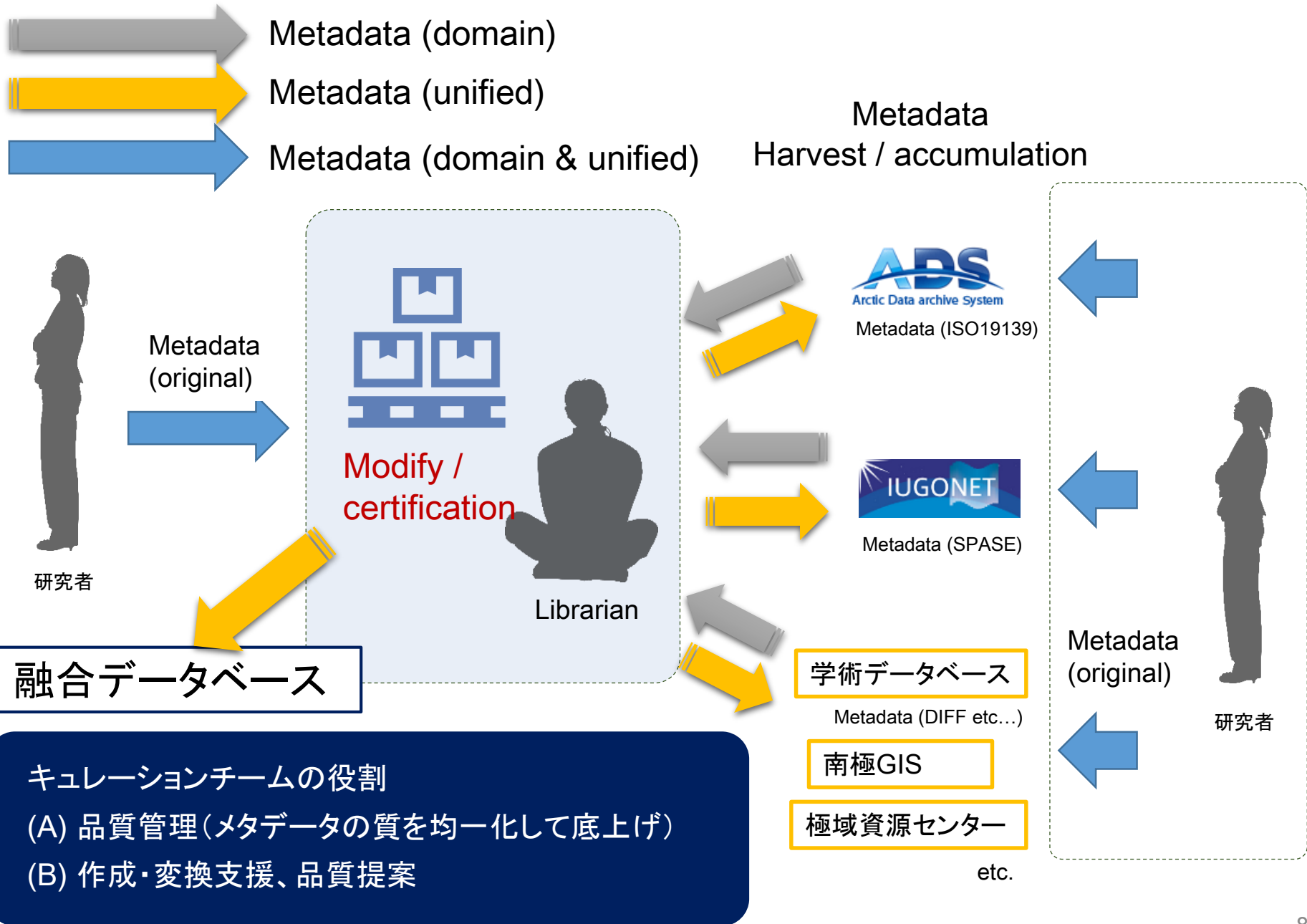
## 📄 波及効果

---

- 本プロジェクトで得られる多様なデータに対するデータ公開、データ融合の手法やノウハウ、ツール等を、大学共同利用機関として、将来ROIS以外の全国の研究機関や大学に展開することで、オープンデータの促進、データ駆動型科学の発展に貢献できる。

📁 メタデータ作成・キュレーションチームの位置づけ





**キュレーションチームの役割**

(A) 品質管理 (メタデータの質を均一化して底上げ)

(B) 作成・変換支援、品質提案



## 図書館員による活動の一例

### サイエンスメタデータ(DIFフォーマット)

項目名	注意事項
概要についての項目	
研究分野のカテゴリ (複数選択可)	必須分野と任意選択分野があります。必須分野は一覧から1つ以上必ず選択してください。必須分野に加えて、任意選択分野も複数選択できます。任意選択分野は登録するメタデータが任意選択分野には含まれないため、選択の必要はありません。(選択可能な分野については表4: p30 参照)
サイエンス キーワード (Science Keywords) [GCMD]	一覧から選択してください。(複数選択可)
ISO トピック カテゴリ (ISO Topic Category) [GCMD]	一覧から選択してください。(複数選択可)
Entry ID(GCMD用タイトル) [GCMD]	(英語で記入、特殊記号不可、80文字まで) (詳細はこちら <a href="http://gcdm.nasa.gov/add/difguide/entry_id.html">http://gcdm.nasa.gov/add/difguide/entry_id.html</a> )
プロジェクト(Project) [GCMD]	(英語で記入、特殊記号不可、255文字まで)
タイトル	(255文字まで)
Title [GCMD] [JaLC/DataCite *]	(英語で記入、特殊記号不可、255文字まで) JaLC/DataCiteへのDOI登録申請時必須
(サブタイトル)	(255文字まで)
(Sub-Title) [JaLC/DataCite]	(英語で記入、特殊記号不可、255文字まで)
データ概要	
Data Summary [GCMD] [JaLC/DataCite]	(英語で記入、特殊記号不可、1000文字まで)

### DOI登録用メタデータフォーマット(JaLC)

項目	属性	必須	検索	文字種	長さ	コード	内容	備考/注意事項
url		O	O	-	-	-	ルート配下URL(ウェブ資源)を示す要素を保持する	
error_process		O	O	1	-	-	エラー処理方法	コンテンツ登録時にエラーが発生したとき、登録の進捗状況を監視するためのフィールド。登録完了後は「中心研究」タブでエラー処理方法を登録する必要がある。
result_method		O	O	1	-	-	登録結果の方法	コンテンツ登録時にエラーが発生したとき、登録の進捗状況を監視するためのフィールド。登録完了後は「中心研究」タブでエラー処理方法を登録する必要がある。
content_classification		O	O	1	-	-	コンテンツ種別	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
request_kind		O	O	1	-	-	アクセス種別	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id		O	O	1	-	-	ファイルID	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
content		O	O	1-N	-	-	コンテンツ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
sequence		O	O	1	-	-	シーケンス	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file		O	O	1	-	-	ファイル	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size		O	O	1	-	-	ファイルサイズ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_mime		O	O	1	-	-	ファイルMIMEタイプ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_md5		O	O	1	-	-	ファイルMD5ハッシュ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_checksum		O	O	1	-	-	ファイルチェックサム	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size_text		O	O	1	-	-	ファイルサイズテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_mime_text		O	O	1	-	-	ファイルMIMEタイプテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_checksum_text		O	O	1	-	-	ファイルチェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size_text_mime_text_checksum_text		O	O	1	-	-	ファイルサイズ、MIMEタイプ、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
creator		O	O	1	-	-	作成者	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
sequence		O	O	1	-	-	シーケンス	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file		O	O	1	-	-	ファイル	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size		O	O	1	-	-	ファイルサイズ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_mime		O	O	1	-	-	ファイルMIMEタイプ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_checksum		O	O	1	-	-	ファイルチェックサム	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size_text_mime_text_checksum_text		O	O	1	-	-	ファイルサイズ、MIMEタイプ、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
description		O	O	1	-	-	説明	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
keywords		O	O	1	-	-	キーワード	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
subject		O	O	1	-	-	主題	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id		O	O	1	-	-	ファイルID	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size		O	O	1	-	-	ファイルサイズ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_mime		O	O	1	-	-	ファイルMIMEタイプ	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_checksum		O	O	1	-	-	ファイルチェックサム	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_size_text_mime_text_checksum_text		O	O	1	-	-	ファイルサイズ、MIMEタイプ、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサム	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text_checksum_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text_checksum_text_mime_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text_checksum_text_mime_text_checksum_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text_checksum_text_mime_text_checksum_text_mime_text_checksum_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。
file_id_mime_size_checksum_text_mime_text_checksum_text_mime_text_checksum_text_mime_text_checksum_text_mime_text_checksum_text		O	O	1	-	-	ファイルID、MIMEタイプ、サイズ、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト、MIMEタイプテキスト、チェックサムテキスト	コンテンツ種別を登録する必要がある。システム上の登録バッチ処理時に、登録結果が「登録失敗」の場合、エラーメッセージが表示される。

項目	必須	検索	文字種	長さ	コード	内容	備考/注意事項
項目名							
研究分野のカテゴリ(複数選択可)							
サイエンスキーワード(Science Keywords)	O						datacite:subject
ISOトピックカテゴリ(ISO Topic Category)	O						
Entry ID(GCMD用タイトル)	O						
プロジェクト(Project)	O						
タイトル	O						
Title (サブタイトル)	O						
(Sub-Title)	O						
データ概要							
Data Summary	O						
観測地域(Observation area)	O						
観測領域(Spatial Coverage)	O						
観測点名							
Data Location	O						
観測期間/運用期間(Observation Period / Temporal Coverage)	O						datacite:date type:collected
データ取得期間(Data Period)	O						
観測風景(Observation Photo)	O						
データ取得者情報(Investigator)	O						
データ名(詳細)							
Data Name (Detail)	O						datacite:description
データ形式							
Data Format	O						datacite:format
データ取得サンプリング							
Data Recording / Acquisition Sampling	O						datacite:description
観測機器							
Observation Instruments	O						datacite:description
サンプルデータ(図、グラフ)(Sample data)							
データ公開方法(URL等)							
Publication method(URL, etc.)	O						
データポリシー							

現在、設計工程が進行中

DataCite4.0 スキーマへのマッピングにあたり、記述が足りない箇所をピックアップ

↓

PLsへフィードバック

## 業務量の実際

メタデータ総数：6,000～8,000

データ種（大気データ、気水データ、地質データ、隕石・岩石データ…）

## チェック業務


かかる時間：1個あたり3～10分（類似データごとにまとめて作業）

チェック項目数：30～50（スキーマによって異なる）

チェック内容：テキスト記述箇所を中心に

（誤字・脱字、年月日の入力間違い、存在しない観測点名、リンク先の内容等）

機械的なvalidationは済んでいるため、確認すべきポイントは絞られる。

 技術者からの(一般的な)メッセージ

- こういうツールが欲しい！業務を楽にしたい！と言ってみてください。作れます。  
業務を画一化し、品質を管理して、その次のステージへ(PDCA)

 プロジェクト全体からのメッセージ

- メタデータ作成・キュレーションチームの実務担当を募集 → プロジェクトメンバーへ
- 検証・フィードバックチームへの参加は自由 → プロジェクトメンバーへ

 **まずはその取り掛かりが大切**

やってみて、初めて分かることがある。

- 図書館が本当に得意とするところは何か？
- グループを作り(共同研究、プロジェクト)、専任業務として認知され、処遇される

(参考資料)

## パイロット (6月～)

## 図書職員とのコラボレーション

No.	難易度	内容	所感 (図書)	所感 (IUG)
1	低	既存のメタデータ(XML)を機械的に修正する	◎ 容易	◎ OK
2	中-	既存のメタデータ(XML)を新しいスキーマバージョンに則したものに变更する	○ 出来た	○ OK
3	中	PIにインタビューしながら新しくサイエンスメタデータを作成する	進行中	○ OK
4	中	外部サーバを使って単純なデータ公開・更新を行う	(予定)	
5	中+	自前サーバを使って単純なデータ公開・更新を行う	(予定)	
6	高	IUGONETのようなサービスのデータ公開部分を運用する	<del>(予定)</del> (予定)	
7	高+	IUGONETのようなサービスのデータ登録部分を運用する	(予定)	

サイエンス成果の創出に向けたデータアーカイブと解析環境の融合. 梅村宜生ほか  
第3回オープンサイエンスデータ推進ワークショップ@京大より抜粋