

## 中核機関群 活動・計画報告

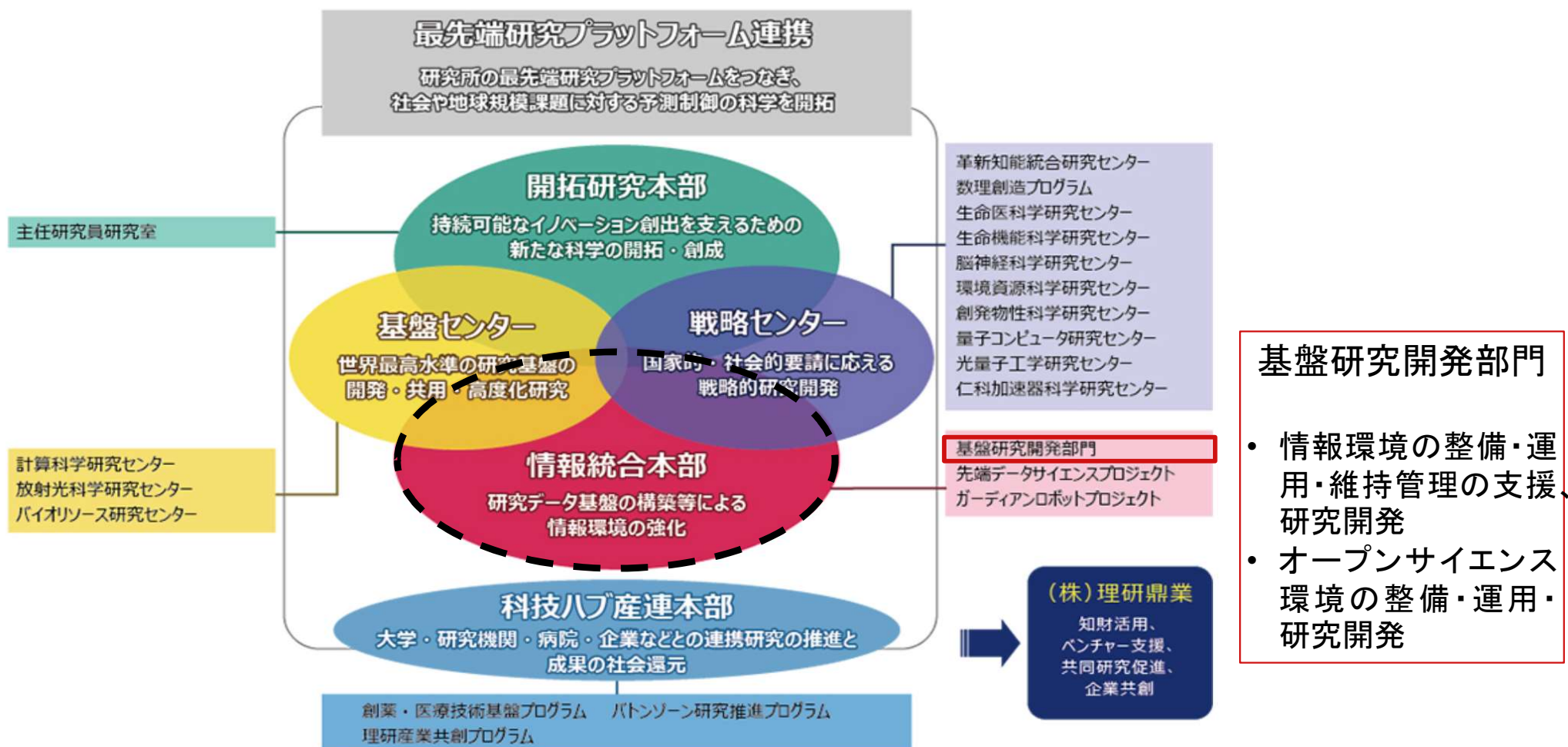
# プラットフォーム連携

美濃 導彦

理化学研究所 情報統合本部

# 理化学研究所の研究体制

- 日本で唯一の自然科学の総合研究所  
物理学、工学、化学、数理・情報科学、計算科学、生物学、医科学などに及ぶ広い分野で研究を進めています。

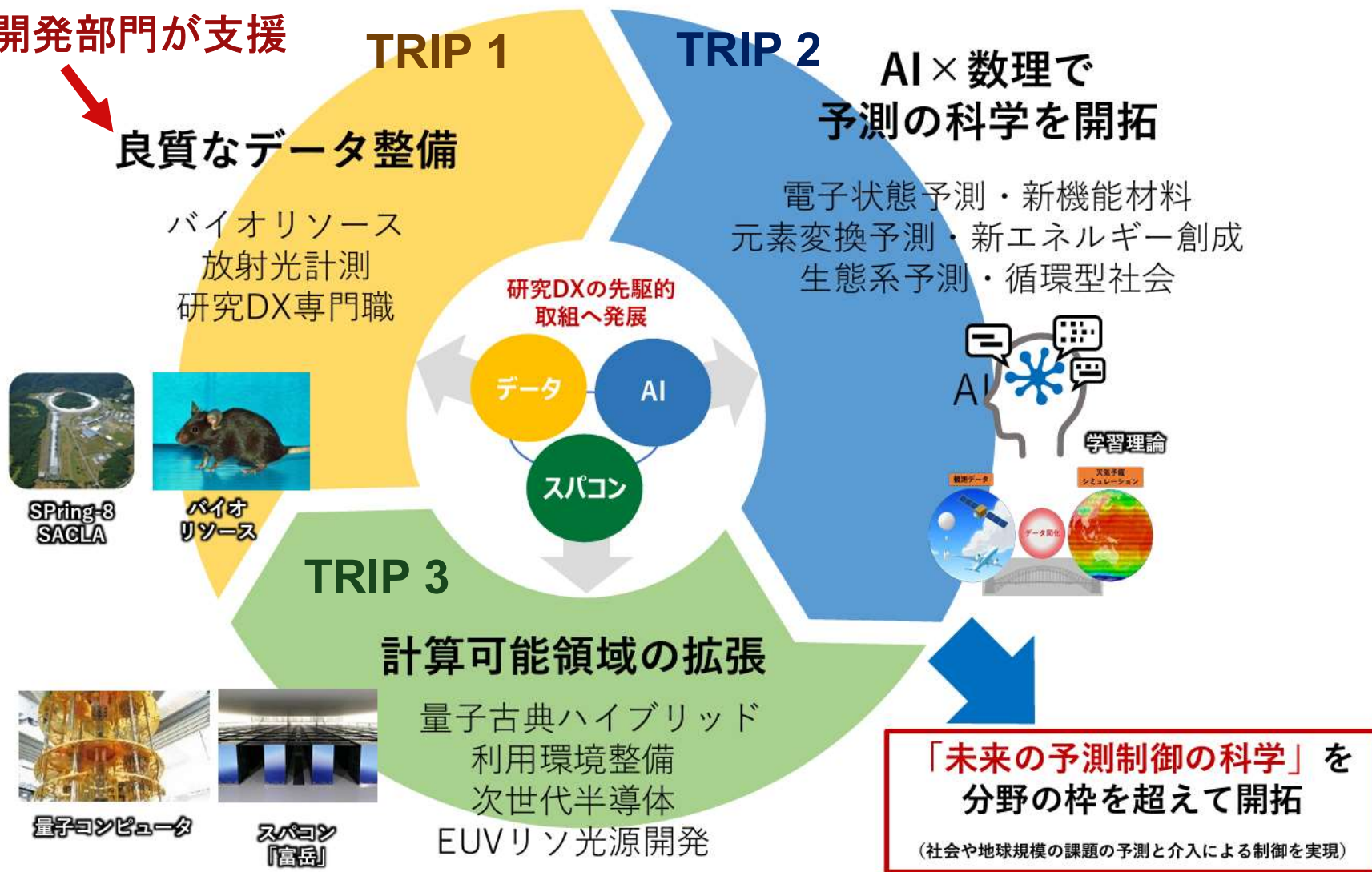


情報統合本部は、理研全体の情報基盤構築・運営を統合的かつ戦略的に推進  
データ駆動型研究、オープンサイエンス、研究デジタル変革(DX)の加速を支援

# TRIP: Transformative Research Innovation Platform of RIKEN platforms

新型計算機と予測アルゴリズム、データ整備を連携させ、  
未来の予測制御の科学を開拓

基盤研究開発部門が支援



# TRIP1における基盤研究開発部門の役割

オープンサイエンスの推進によるTRIPへの貢献

国内外の機関や企業と  
データ共有・公開

## ➤ 【利活用推進】 データ利活用・研究推進の支援

【データ知識化開発ユニット】

- 利用しやすいデータへと統合・再編（研究ニーズに沿ったデータ・メタデータ再編）
- データの検索・推論、可視化技術の開発・提供
- 実験機器からの自動的なメタデータ収集と広範な利用が可能となる良質なメタデータの付与支援

## ➤ 【蓄積・管理】 データ共有・社会還元のための研究情報基盤の構築

【データ管理システム開発ユニット】

- 研究者が研究データを扱いやすいRDCをNIIと共同で開発
- 研究データ管理の高度化に関するサポート（DMP、データ持続性、最適配置）
- 個人情報等を安全に管理するセキュアデータレジストリの構築
- 大規模なデータを出力する観測装置のエッジプロセッシング環境とIoT化

## ➤ 【収集】 良質な研究データの収集支援

- 理研が産出する多種多様な研究分野のデータの集積に関する研究開発
- 各研究分野におけるデータ標準化、メタデータ標準化支援
- 集積したデータの2次利用の推進のための研究開発

生命科学、医科学、物理学、工学、化学・・・【データ共有開発ユニット群】

良質なデータを生み出し  
幅広い科学分野で利活用を促進

広範なデータの  
使い方についての  
新しい学理

科学の計算領域を  
拡張する新たな技術



# 理研研究情報管理基盤

(理研オープンサイエンス基盤)

理研研究情報管理基盤

理研研究者



論文・著作  
証拠データ

RIKEN RARS  
Bib

成果・発表管理

理研 情報統合本部が整備する範囲

研究者により作成されたデータ・メタデータ

共用計測機器等から生成されたデータ・メタデータ

Data storage

理研拠点エッジシステムでのデータ集約・加工

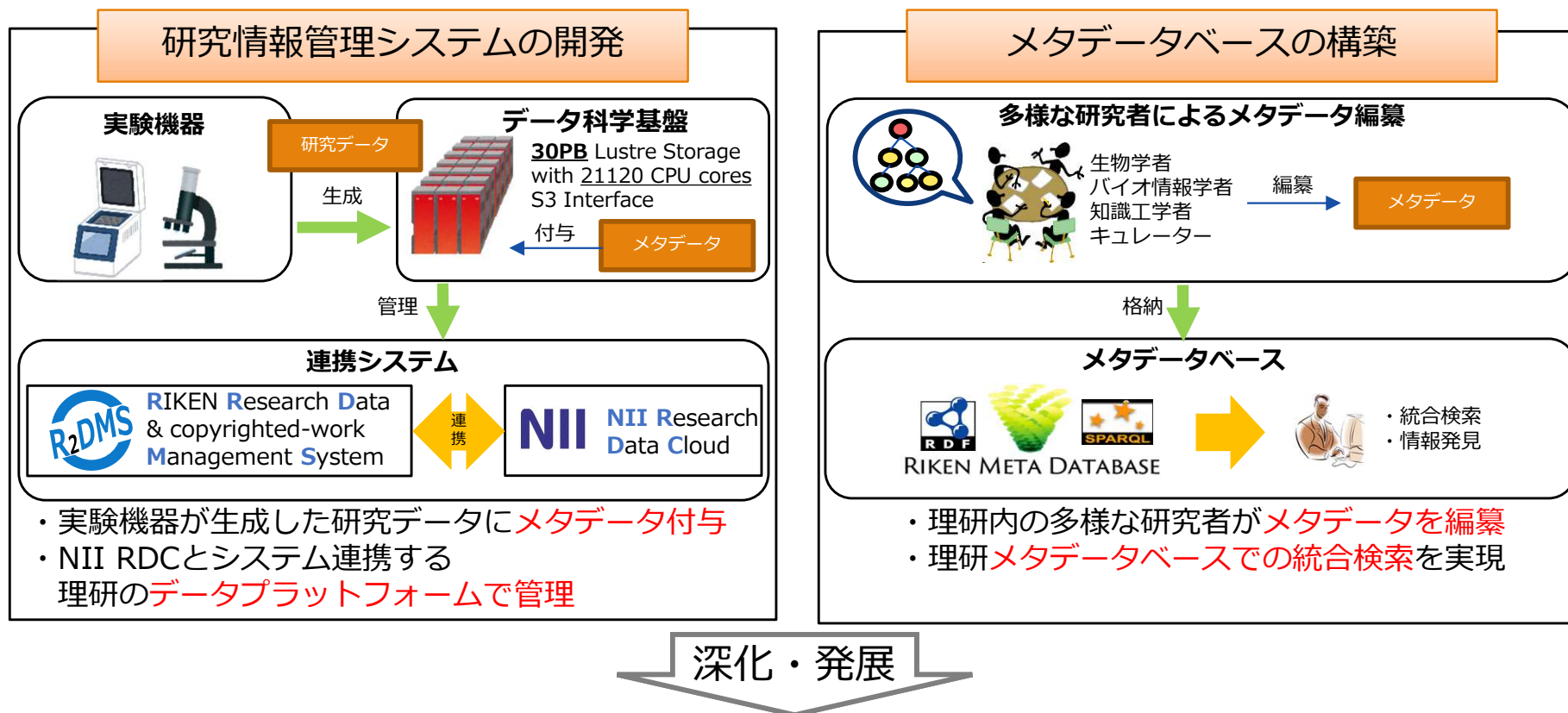
RIKEN Research Data & copyrighted-work Management System  
box nextcloud  
GakuNin RDM  
RIKEN MetaDatabase  
Metadata  
SEARCH

データ共有・解析・検索

NII Research Data Cloud  
CiNii Research  
JAIRO Cloud  
GakuNin RDM

研究機関・大学等

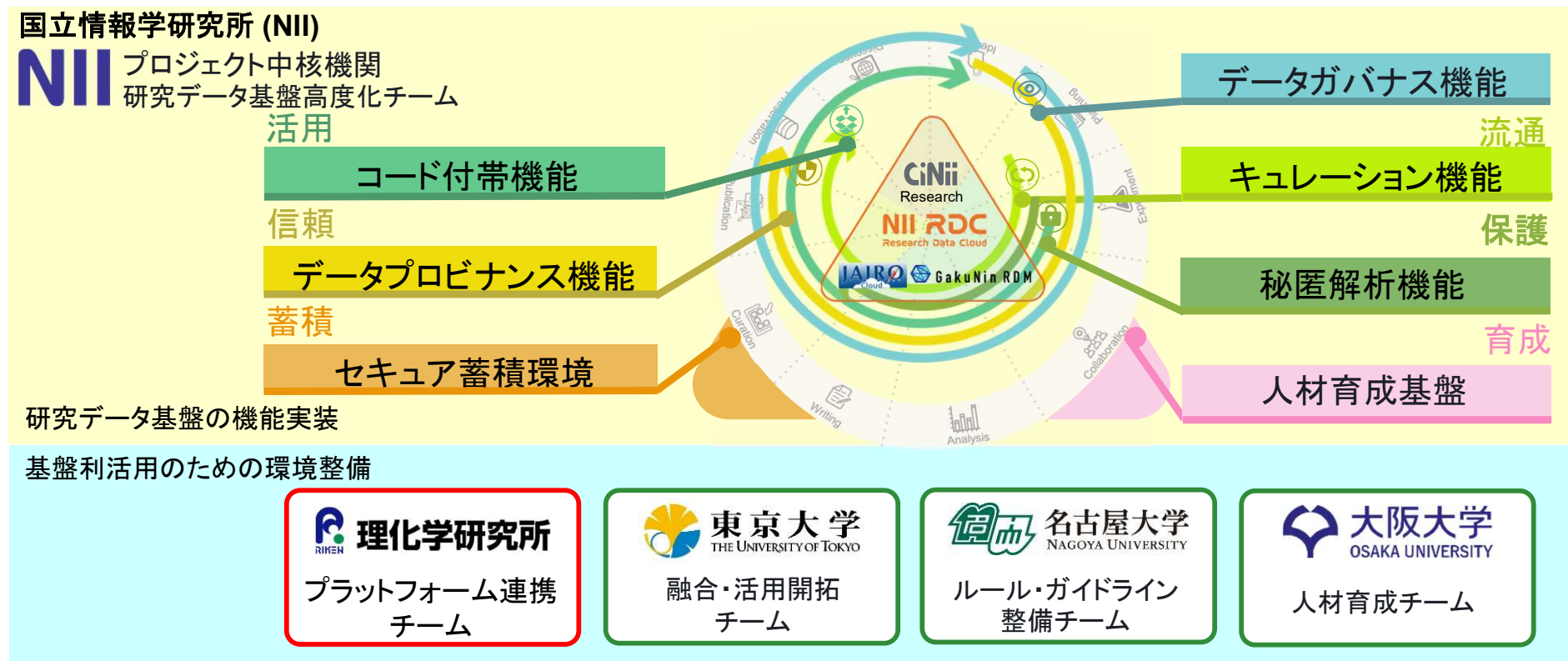
# 基盤研究開発部門の成果を我が国の基盤構築に



研究データエコプロジェクトにおいて、これまでの知見を深化・発展させ、NIIが進める我が国の研究データ管理基盤構築を加速させる

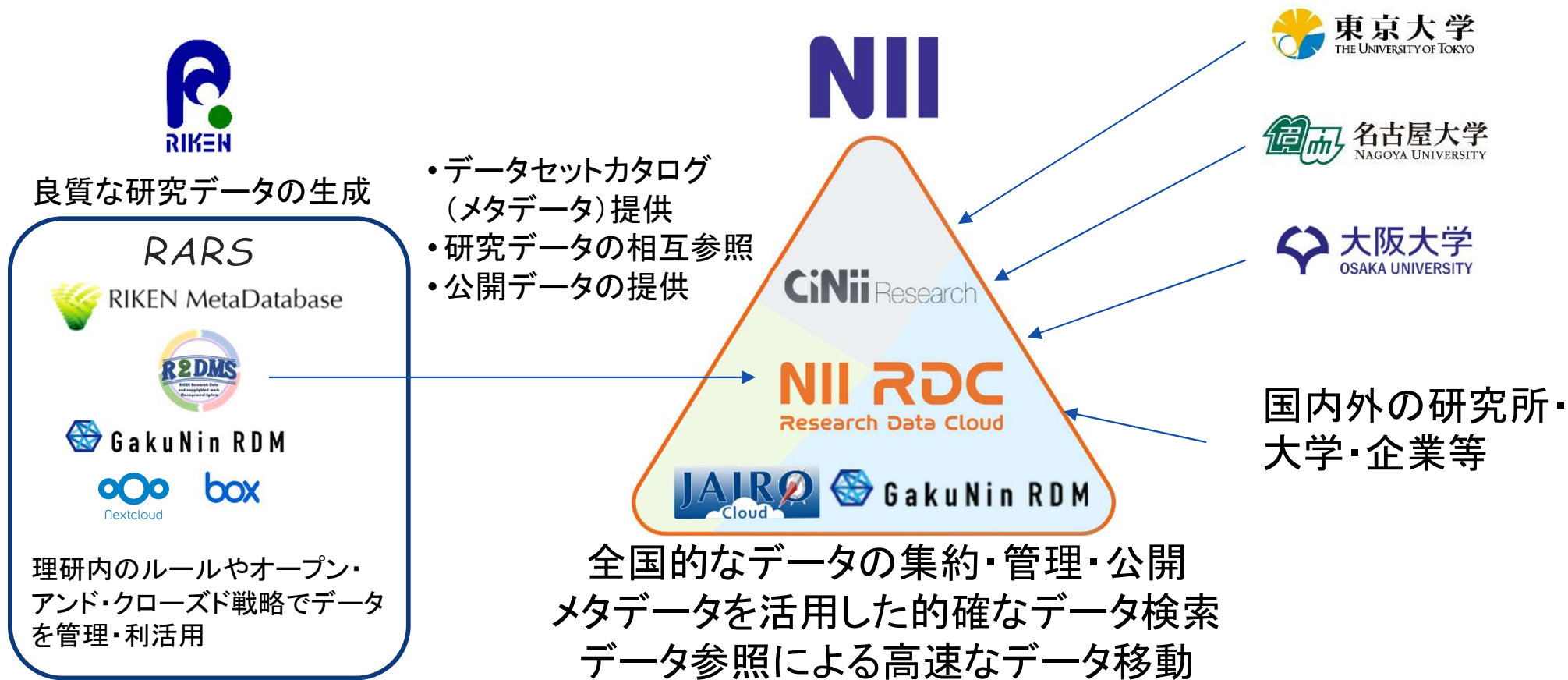
# 我が国・世界で研究データの共有・利活用を推進するために

## 研究データエコプロジェクト プラットフォーム連携チームの役割



- **プラットフォーム（研究基盤）連携の実現**
  - 理研・他機関のプラットフォームとNII Research Data Cloud (NII RDC) との連携
- **データ連携のためのメタデータ設計**
  - 検索やデータを説明するためのメタデータ設計、標準化推進

# プラットフォーム連携の理想形



- 全国で統一された研究データ管理の方法やルール・ガイドラインが整備され、研究者のデータ管理コストの低減を図りつつ、データ利活用を加速
- 研究者は所属組織の基盤や設備を活用し良質なデータを生成しつつ、データが所在する場所(組織)を意識せずともデータの管理・利活用が可能



# プラットフォーム連携チーム研究計画

プラットフォーム連携のための  
機能・メタデータに関する研究開発を推進

## I.設計フェーズ (2022-2023)

プラットフォーム  
連携のための  
機能・メタデータ設計

## II.開発・評価フェーズ (2024-2025)

プラットフォーム連携を  
実現する機能開発

## III.実装・普及フェーズ (2025-2026)

プラットフォーム連携を  
実現する機能の  
実装と普及

プラットフォームの  
リポジトリ機能開発

2 理研内の  
機関リポジトリと  
NII RDCとの連携機能  
要件を検討

1 ガイドラインや運用  
ルールを考慮した  
プラットフォーム  
連携機能開発

5 連携プラットフォームの  
試行運用、ルール  
や規定整備による  
普及推進

学術機関リポジトリの高度化・利用拡大



プラットフォームの  
メタデータ開発

3 データ利活用推進や  
統合検索に資する  
メタデータの要件定義

4 付与コスト低減や  
品質向上を考慮した  
メタデータ付与  
システムの開発

6 データ利活用を推進  
する新たなメタデータ  
仕様を利用し、NII  
データ検索基盤と連携

検索基盤における統合的な発見



持続的・安定的に運用可能な  
プラットフォーム

# これまでの成果

## 1. プラットフォーム連携に関する設計

- データ参照・移行を行うための検討
  - RIKEN R2DMS(理研研究情報管理システム)や、理研CBS(脳神経科学研究センター)のドメインリポジトリと、GakuNin RDMとの連携手法等を検討
  - メタデータおよび大規模データをハンドリングする機能要件検討

## 2. データ連携のためのメタデータ設計

- 組織、立場を超えてメタデータを設計できるツールの整備
- ユーザの専門性に従ったわかりやすいGUIを設計

## 3. 運営体制整備

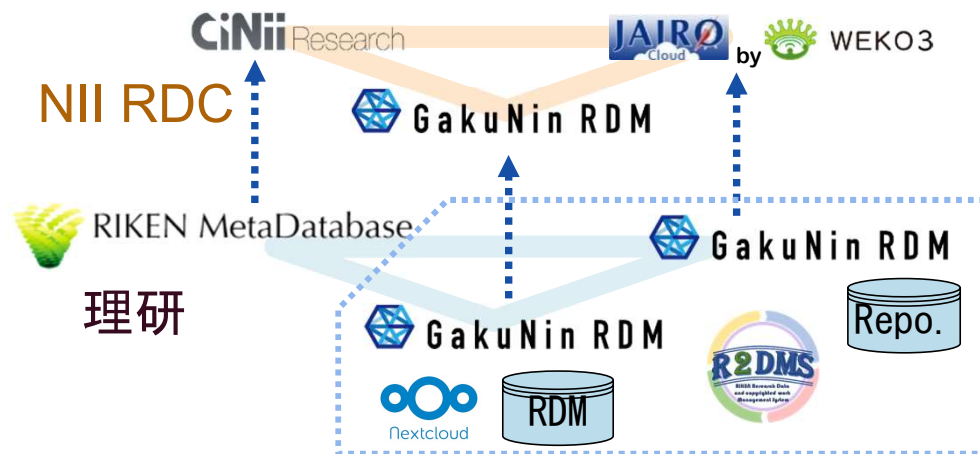
- 理研の生命科学を中心とする各センターから選ばれた委員が参加する事業推進委員会を設置
- プロジェクト参加機関の委員が参加する、プラットフォーム連携チーム連携会議を設置

# プラットフォーム連携に関する設計

機関で所有する研究データ管理基盤(RDM)、情報基盤、実験機器と NII Research Data Cloud (NII RDC) との連携の方法論を確立する

## 方法1：機関のRDMを終了し、データをNII RDCに移行させる方法

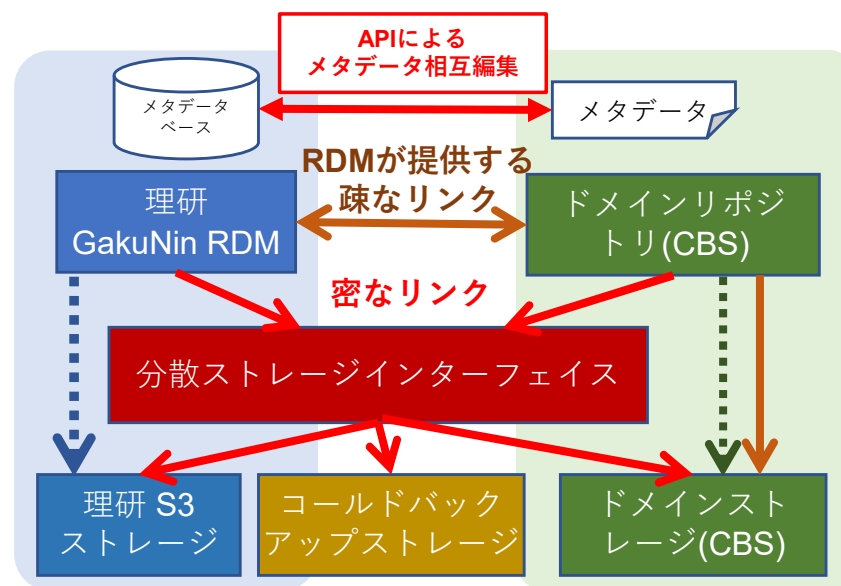
- データの移行方法
- データがNII RDCにある時、機関の情報基盤をどう活用するか、どういう機能を残すか？
  - スパコンの利用
  - データを生成する実験機器の取り扱い
- ユーザ管理方法



## 方法2：複数のRDM間や、RDM-リポジトリ間で互いにリンクを張って分散管理する方法

(事例：理研の情報統合本部のRDMと脳神経科学研究センター(CBS)のドメインリポジトリの連携)

- RDM間リンクの方法は基本的に確立している
  - HTMLリダイレクションなど
- 小規模レポジトリが方法1に向かうための過渡的運用
  - コンテンツの相互編集機能の検討
  - 機関の情報基盤との連携方法



## NII RDCへのデータ移行を行うための検討

理研内の2つの情報基盤 SPring-8基盤 および R2DMS の運用体制から、NII RDCとの連携に必要な機能の一般化と抽出

- 共用施設におけるロール(役割:管理者・利用者等)の抽出と適切な権限の設定
  - 施設(機関)内のすべての研究プロジェクトを管理できるロールが必要
- 大量プロジェクト・メンバーの一斉管理機能の検討
  - 各部署でのまとまったプロジェクト発足のタイミング(部署で公募したプロジェクトの開始など)に必要

→検討結果に基づき、研究データ基盤高度化チーム(NII)が実装

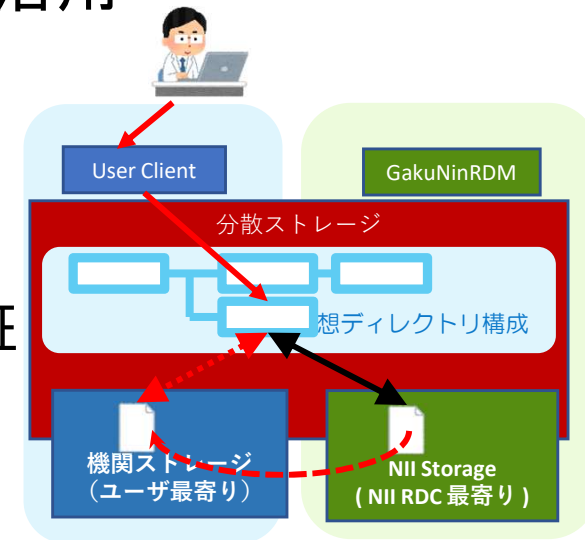
- 以下の項目をさらに検討中
  - 大規模データの GakuNin RDMからの見せ方・使い方についての検討
  - 理研の各拠点(神戸・横浜等)の情報基盤との連携手法



# 大規模データやメタデータを扱うための機能要件検討

## 大規模データ管理に適したシステムでのデータ蓄積

- 分散型ストレージによるデータ共有・送受信機能を活用
  - データ配置効率化による転送速度改善
  - データ転送・共有に特化したプロトコルの利用
- GakuNinRDMによる分散型ストレージ利用法の検討
  - 組織横断的サービス作成の難しさ
  - ストレージでのデータ編集におけるデータの正当性保証



## 柔軟なメタデータスキーマをサポートする メタデータ入力システム

- GakuNinRDM アドオン形式
- メタデータスキーマファイルのアップロードと共有が可能
- スキーマファイルに基づいた編集インターフェースの提供



## 脳科学ドメインレポジトリとの密連携による相互編集型レポジトリの検討

- 上記2つの検証環境およびデータの密参照の検証
- テスト環境の設計→現在構築中

# メタデータの定義:

## 1. カタログデータ

- データセットを検索するために活用されるメタデータ
- データセットの位置を含む説明が記載される (書誌的情報)

## 2. 研究知識データ

- 研究知識の共有・より詳細な検索・知識推論に活用されるメタデータ
- データが生成されたときの、実験環境、機器及びその設定、バイオリソース等の研究の詳細が記載される

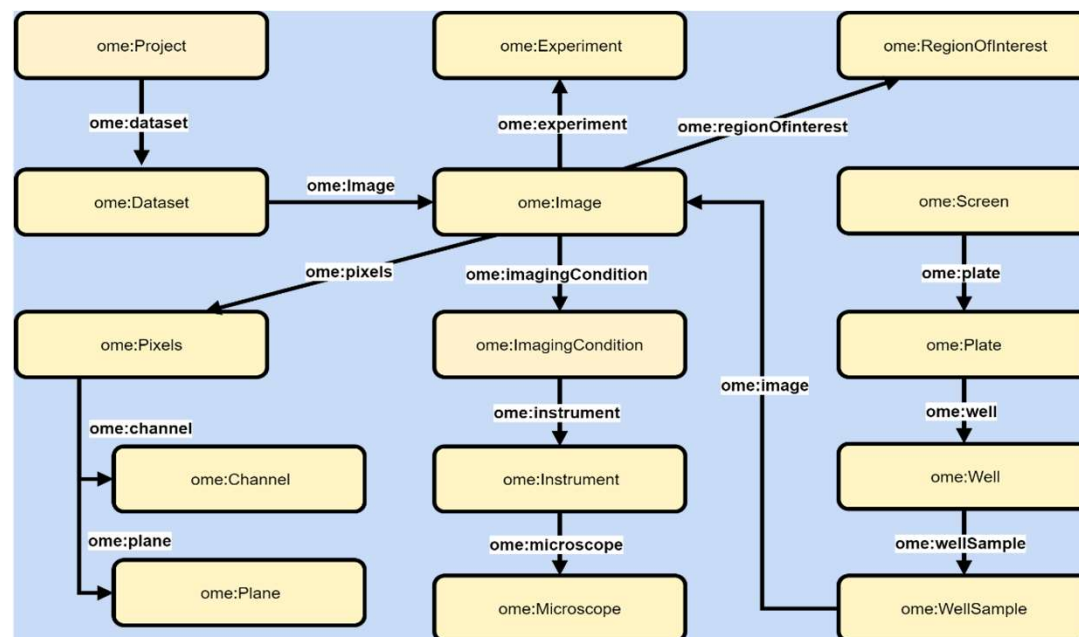
1.2.をどう組み合わせ、研究データセットの検索に最適なメタデータを定義するかが課題

### 1. カタログデータの例

#### plantMetabolomics

title	RIKEN Plant Metabolome MetaDatabase	タイトル
alternative	RIKEN PMM	
homepage	<a href="http://metabobank.riken.jp/">http://metabobank.riken.jp/</a>	URL
Metadata link	<a href="http://metabobank.riken.jp/ommm/db/plantMetabolomics">http://metabobank.riken.jp/ommm/db/plantMetabolomics</a>	
publisher	RIKEN Center for Sustainable Resource Science Head Office for Information Systems and Cybersecurity RIKEN	発表者
description	RIKEN Plant Metabolome MetaDatabase (PMM) is used to provide integrated plant metabolome data generated by RIKEN Center for Sustainable Resource Science (CSRS). This database provides data in the bases of standardized metadata techniques including semantic web and Resource Description Framework (RDF).	
language	English language Japanese language	
species	<i>Ophiorniza pumila</i> <i>Jatropha curcas</i> <i>Dicksonia antarctica</i> <i>Lotus japonicus</i> <i>Spinacia oleracea</i> <i>Arabidopsis thaliana</i> Show more (14 items)	生物種
theme	drug or chemical metabolite, metabolism or metabolome	対象
dataType	bioresource interaction / pathway others	データの種類

### 2. 研究知識データの例 (顕微鏡イメージング)



130の概念(クラス)と210の関係(プロパティ)が定義されている

## メタデータについての課題

- 研究データ(セット)を見付つために必要不可欠なデータ  
研究データの検索・再利用に必要な事項を記載したもの
  - 研究データが生成されたときの、実験条件、機器、機器設定、実験者、実験日などの事項
  - 研究データを説明する論文への参照
  - 研究データ取得の目的や科学的仮説などの実験背景

しかし、

- メタデータ作成には、記述するための語彙の選定やスキーマの設計において、情報学的知識を要する
- メタデータ作成は、研究者には負担が大きな作業
- メタデータを検索するにも、クエリ作成技術が必要
- メタデータ付与は分野に特化するが、他分野のメタデータとの統合に戦略が必要

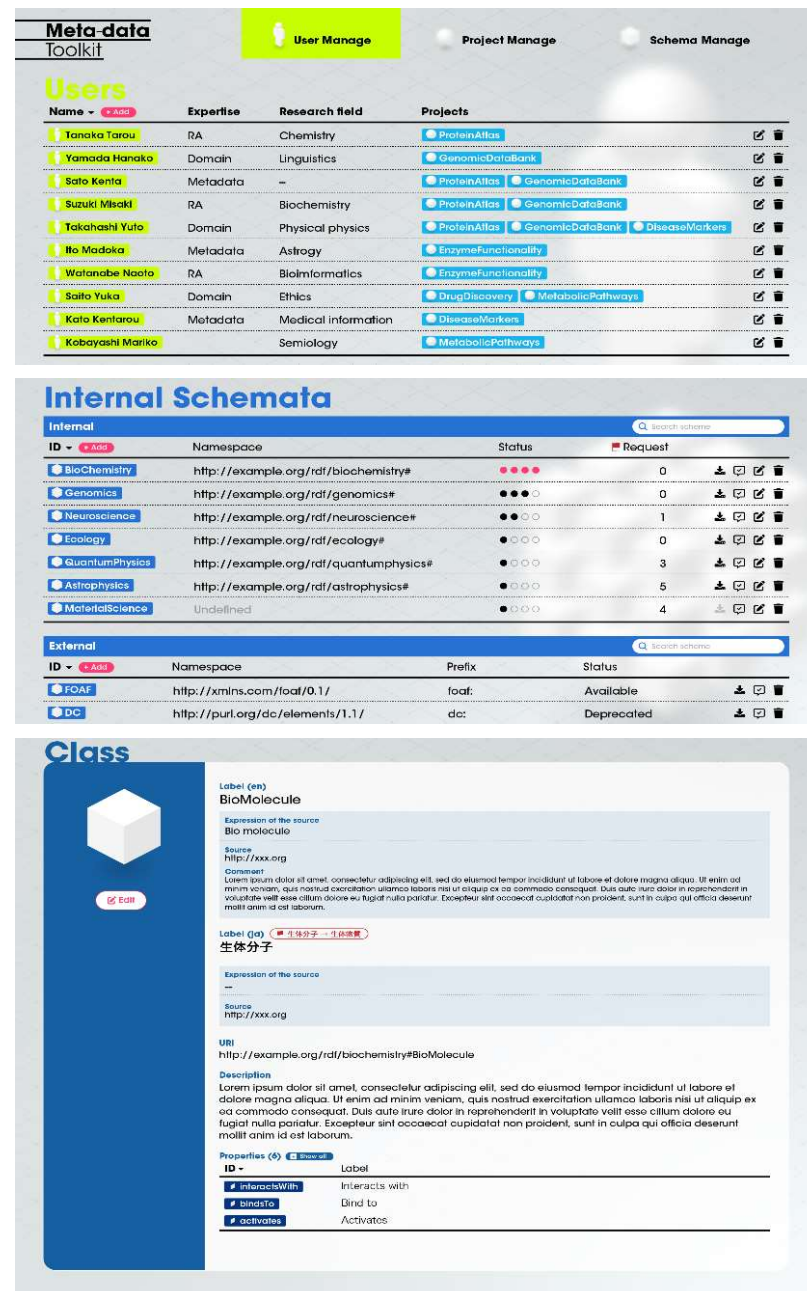
# データ連携のためのメタデータ設計

組織や職制(立場)を超えて、複数人で協力してメタデータを設計できるウェブツールを整備

下記ユーザの専門性に従ったわかりやすいGUIを設計した:

1. セマンティックウェブ技術をよく知る、メタデータ専門家
2. データを産出した研究についてよく知るドメイン専門家
3. メタデータの入力や関係者との連携を行う、メタデータ作成支援員

設計結果はJSON Schema形式で出力(プラットフォーム連携のメタデータ管理機能から取り込み可能)



The screenshot displays the 'Meta-data Toolkit' interface with three main sections:

- Users:** A table listing users with columns for Name, Expertise, Research field, and Projects. Users include Tanaka Tarou (RA, Chemistry), Yamada Hanako (Domain, Linguistics), Sato Kenta (Metaddata, -), Suzuki Mitsuki (RA, Biochemistry), Takahashi Yuto (Domain, Physical physics), Ito Madoka (Metaddata, Astrogy), Watanabe Naoto (RA, Bioinformatics), Saito Yuka (Domain, Ethics), Kato Kentarou (Metaddata, Medical information), and Kobayashi Mariko (Semiology).
- Internal Schemata:** A table listing internal namespaces with columns for ID, Namespace, Status, and Request. Examples include BioChemistry, Genomics, Neuroscience, Ecology, QuantumPhysics, Astrophysics, and MaterialScience.
- External Schemata:** A table listing external namespaces with columns for ID, Namespace, Prefix, and Status. Examples include FOAF and DC.
- Class:** A detailed view of a class named 'BioMolecule' with fields for Label (en), Expression of the source, Source, URI, and Description. It also shows properties like 'interactsWith', 'bindsTo', and 'activates'.



## 他機関への展開等の波及効果につながる取組み

- データ連携のためのメタデータ設計
  - 研究所、大学等の研究データ管理支援体制の多様性を考慮したメタデータ設計ツールを開発
  - 組織を超えて、複数人の協調によりメタデータ設計を行っていく
  - 実装完了次第、下記のメタデータについて理研案を入力し、組織を超えた共同作業を実施予定
    - データ管理計画を含む、研究プロセスメタデータ
    - 生命科学系データカタログ
- メタデータ標準化プロセスの加速に貢献