



中核機関群 活動・計画報告

プラットフォーム連携

工藤知宏、小林紀郎、實本英之

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推進
- 研究データエコ事業は情報統合本部基盤研究開発部門を中心に実施：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

TRIP 事業

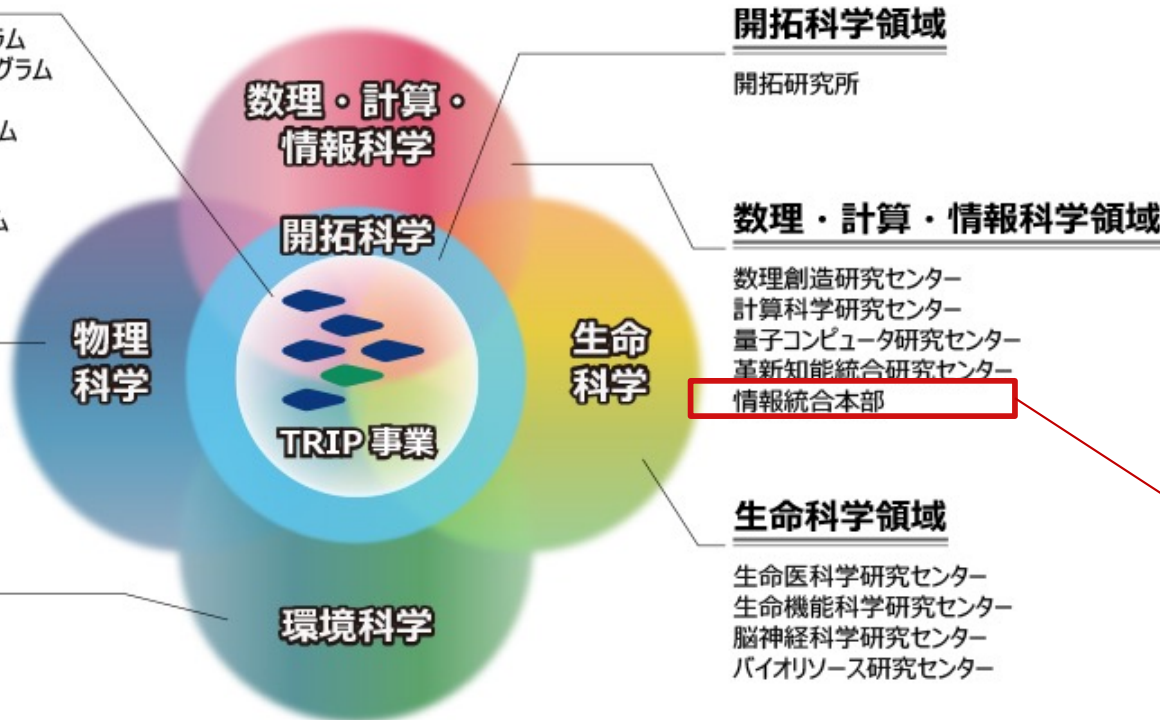
統合データ・計算科学プログラム
科学研究基盤モデル開発プログラム
基礎量子科学研究プログラム
創薬・医療技術基盤プログラム
先端半導体科学プログラム
理研産業協創プログラム
バトンゾーン研究推進プログラム

物理科学領域

創発物性科学研究センター
光量子工学研究センター
仁科加速器科学研究センター
放射光科学研究センター

環境科学領域

環境資源科学研究センター
(バイオリソース研究センター)



開拓科学領域

開拓研究所

数理・計算・情報科学領域

数理創造研究センター
計算科学研究センター
量子コンピュータ研究センター
革新知能統合研究センター
情報統合本部

生命科学領域

生命医科学研究センター
生命機能科学研究センター
脳神経科学研究センター
バイオリソース研究センター

基盤研究開発部門

- 情報環境の整備・運用・維持管理の支援、研究開発
- オープンサイエンス環境の整備・運用・研究開発

※TRIP (Transformative Research Innovation Platform of RIKEN platforms)

最先端研究プラットフォーム連携

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

研究データエコシステム・プラットフォーム連携チーム

- NII RDCをコアとする全国規模の研究データ連携プラットフォームの構築技術確立
 - 大学・研究所等が持つデータ管理プラットフォームと、NII RDCを連携させる
- 全国で統一された研究データ管理の方法やルール・ガイドラインにより、研究者のデータ管理コストを低減しつつ相互データ利活用を実現
 - 理研は産学連携や共同研究を強力に推進している
→ガイドラインやルールに応じたプラットフォーム連携のベストプラクティス構築につなげる

→ **研究データの共有・利活用を推進**



良質な研究データの生成

RARS

RIKEN MetaDatabase



GakuNin RDM



理研内のルールやオープン・
アンド・クローズド戦略でデータを
管理・利活用

- データセットカタログ
(メタデータ) 提供
- 研究データの相互参照
- 公開データの提供

NII

CiNii Research

NII RDC
Research Data Cloud



GakuNin RDM



国内外の研究所・
大学・企業等

- NII RDCと、大学・研究所・企業等のデータ管理基盤が
連携し、全国的なデータの集約・管理・公開を実現
- メタデータを活用した的確なデータ検索

■ ① システム(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
- 利用者環境とNII RDCとの連携

■ ② 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

研究目標・計画

- ① 機関で所有するRDMや情報基盤／実験機器と NII Research Data Cloud (NII RDC) とのシステム連携の方法論を確立する
- ② データ検索を含めたデータ流通とデータ管理(データ連携)に必要なメタデータの定義およびその管理の方法を確立する



システム連携の目標

機関で所有するRDMや情報基盤／実験機器と NII Research Data Cloud (NII RDC) との連携のベストプラクティスを確立する

課題 1：NII RDCの利用／拠点RDMのNII RDCへの移行

従来：NII RDCによるデータ管理は可能

→NII RDCを拠点情報基盤と、

「どのように導入・連携させ研究フローを支援するか」

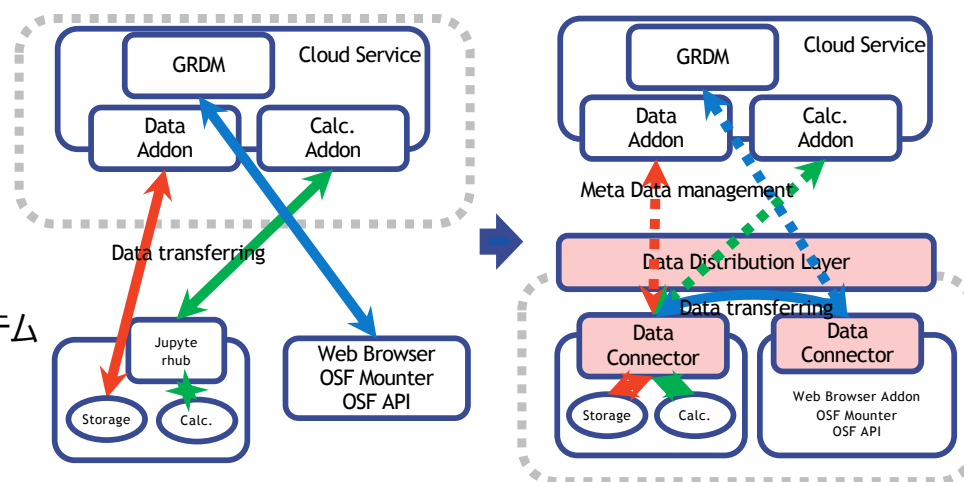
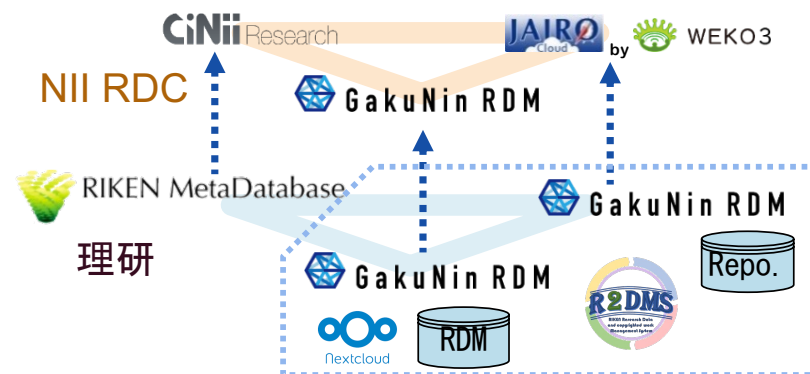
- 拠点RDMを機能分解しNII RDCにマップ
 - ・ 既存事例から標準的に必要な機能を抽出、
現行のNII RDC運用で問題発生していない部分
は今後のNII RDCの検討案件とする
- ・ 機関の情報基盤との連携
 - ・ スパコン・実験機器からの利用
- ・ ユーザ管理
- ・ データの保管場所とNII RDCの接続

課題 2：複数RDM連携と大規模データ転送

従来：小規模データに関するRDM／拠点IoT連携は可能

→大規模データ転送を考慮したRDM間のより深い機能連携や拠点情報基盤連携の実現

- ・ RDM間の機能連携に関しては、RDM間に限らず、複数システムでデータ・メタデータ共有を助けるという目標に昇華統合
 - ・ ドメインRDMの利点を生かす
- ・ ユーザの手元や計算基盤への大規模データ転送の効率化
- ・ メタデータ共有手法の提案



システム連携 : GakuNinRDM 活用ツール

10月9日 オープニングデモセッションでご紹介

- GakuNinRDM と各ユーザのデータ用法やアプリケーションを簡易に連携させることを目的とするツール群
 - GakuNinRDM ファイル連携ツール
 - GakuNinRDM上のデータをユーザPCのファイルマネージャ(相当)から利用できるようにする
 - ローカルファイルシステムにファイルがあるように見えるため、リモート機能のないアプリケーションでもデータを捕捉・利用できる
 - Google DriveやOne Drive などと同じ感覚で使える
 - GakuNinRDMプロジェクト大規模作成ツール
 - GakuNinRDM上に一括して複数のプロジェクトを作成する
 - 大量のデータセットについてデータセット毎にプロジェクトページを作
って管理できる
 - 史料がたくさんあるときなど、その史料毎に解析や説明補足、共有を行える

システム連携 : GakuNinRDM ファイル連携ツール

◎ OSに合わせて2通りのツールがある

- GakuNinRDM マウンタ
ファイルシステムから直接GakuNinRDM のファイル进行操作する

- Linux/Mac のみ

- マウントとは：別の場所にあるデータを、あたかも自分のパソコンの一部として見せる仕組み
- 起動時マウント処理が可能

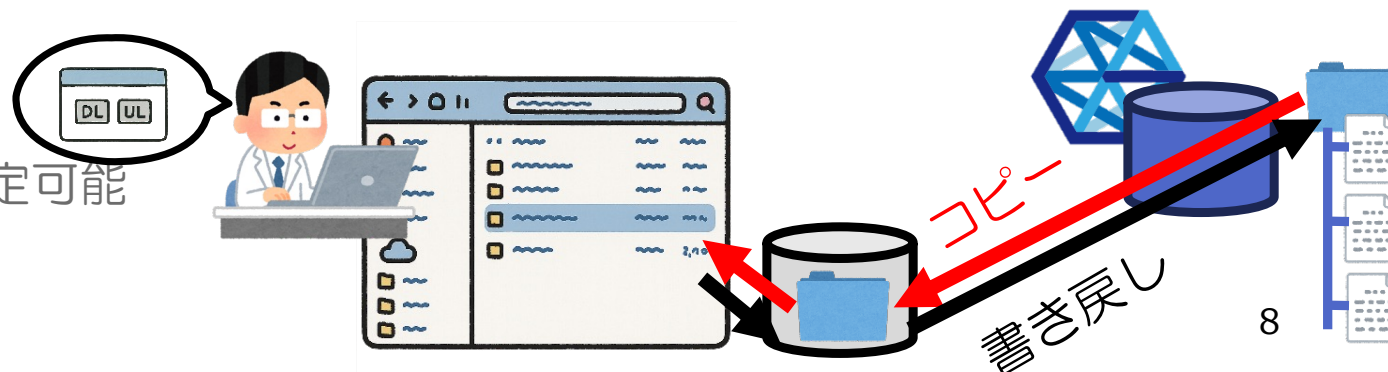
(こっそり)Pyfuse3->winfsp 移植が出来れば..... Windowsでも



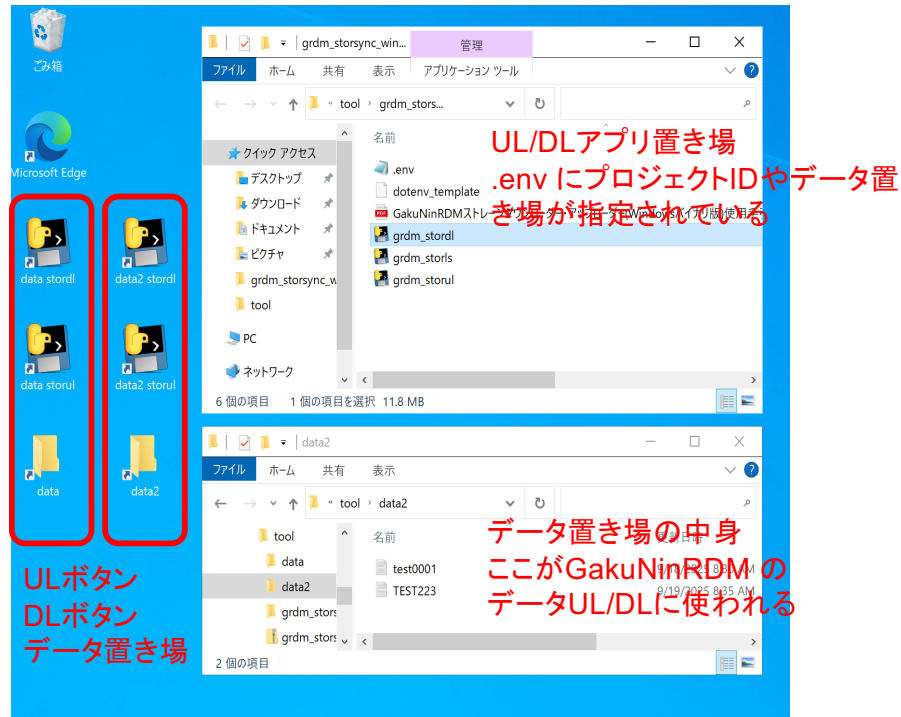
- GakuNinRDM アップローダ／ダウンローダ
UL/DLアプリでファイルをローカルに自動取得・書き戻し

- Windows/Linux/Mac

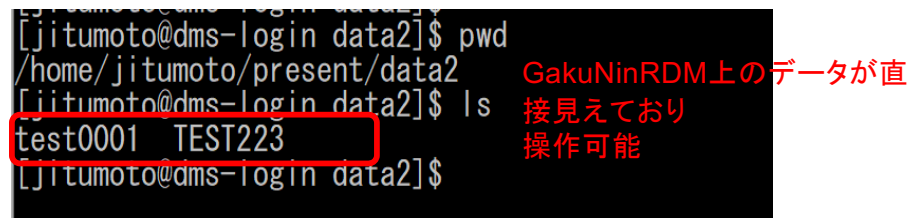
- UL/DLボタンでファイル操作
- 起動・終了タスクとして自動処理設定可能



実際の画面例



UL/DLツール on Windows



マウンタ on Linux

GakuNinRDM by WebUI (通常利用)



GakuNinRDM上のデータ
ここからデータのダウンロードやアップロードも可能
Webベースのログイン手続きや、プロジェクト選択が必要

システム連携： GakuNinRDMプロジェクト大規模作成ツール

- Windows/Linux/Mac で利用可能
- CSVファイルに行ごとにプロジェクトページの設定を書く
 - 既存親プロジェクトのID (省略可能), タイトル, データのローカルでの置き場所
- コマンドラインツールを使って
CSVファイルを1行ずつ処理しながら
GakuNinRDM にページを作り、
データがアップロードされる

```

vrbfz,土偶1,C:¥¥Data¥¥Yayoi¥¥dogu1
vrbfz,土偶2,C:¥¥Data¥¥Yayoi¥¥dogu2
vrbfz,壺1,C:¥¥Data¥¥Yayoi¥¥tsubo1
vrbfz,壺2,C:¥¥Data¥¥Yayoi¥¥tsubo2

```

設定ファイル(CSV)



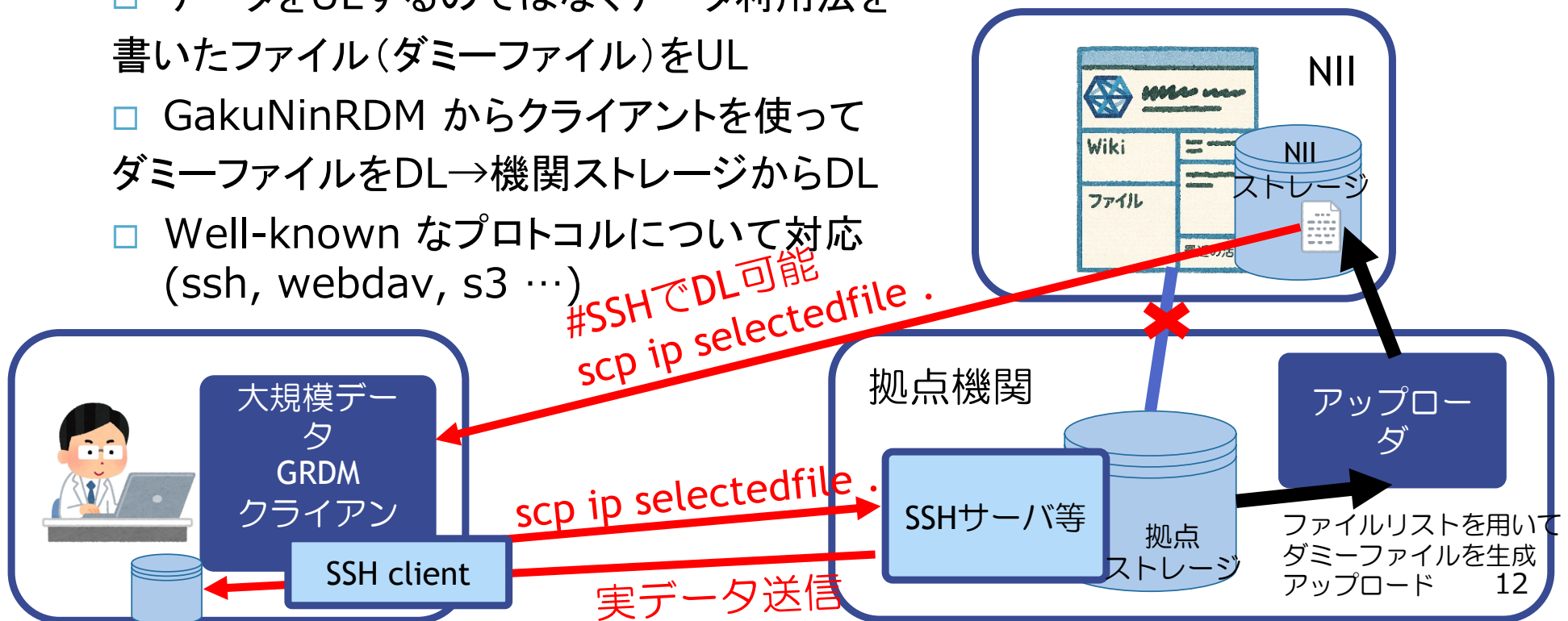


システム連携：今後に向けて

- GakuNinRDM 活用ツールの利用支援と改良
 - ユースケース構築と利用促進
(東大ユースケース構築事業との連携予定)
 - 機能改善
 - 是非利用していただけると助かります
 - 実験機器や他のデータシステムとの連携などでも利用検討したい
- 大規模データを含むシステム間データ参照・流通連携
 - 大規模データを対象に含む、データ流通を効率化するサービスと NII RDC の連携を行うための設計及び実装
 - 次ページ

システム連携：大規模データの転送効率化ツール

- GakuNinRDM に拠点ストレージを追加した場合、データが常にGakuNinRDM (NII環境) を経由して転送される
 - 拠点ストレージ→GakuNinRDM→利用ユーザ
- クライアントツールを構築し、GakuNinRDM からの情報を元にクライアントと拠点ストレージ間で直接データをやり取りするツールの開発中
 - データをULするのではなくデータ利用法を書いたファイル(ダミーファイル)をUL
 - GakuNinRDM からクライアントを使ってダミーファイルをDL→機関ストレージからDL
 - Well-known なプロトコルについて対応 (ssh, webdav, s3 ...)



データ連携：メタデータと研究データ管理

- メタデータ仕様策定指針確立（NIIのデータ検索基盤CiNii Researchを想定）
 - データを管理・検索・利活用する際には、
 - 何についてのデータで
 - どのような方法・機器により生成されたデータで
 - 誰が、どの実験・プロジェクトで生成されたものか など、
研究の背景を必要十分に記載しておくことが必要となる。
- 分野に跨ったデータ連携をサポートするメタデータの要件検討
- メタデータに関する課題
 - 記述すべき項目の決定・メタデータ設計の難しさ
 - 研究コミュニティ・分野等での共通化・標準化の多くの努力との整合
 - 研究者がメタデータを作成する手間を考慮することが重要
 - メタデータの検索・推論・表示機能開発の難しさ
- データのライフサイクル(生成から利用まで)をカバーするメタデータ
 - 従来のメタデータ研究開発は、データ公開・検索にフォーカスしたものがほとんどだった

データ連携：メタデータ設計ツールの実装と公開

現状: メタデータ定義には、統制語彙の利用においてメタデータの専門知識が必要
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して現状課題の解決を試る。

システム上でプロジェクトを作成し、プロジェクト内で共同作業

研究者



データの深い理解
スキーマ基本設計

図書館員、RA



データ流通・検索の知見
標準メタデータとの関連づけ

メタデータ専門家
(情報)



メタデータ技術
オントロジー・既存スキーマを
利用した意味情報の付与

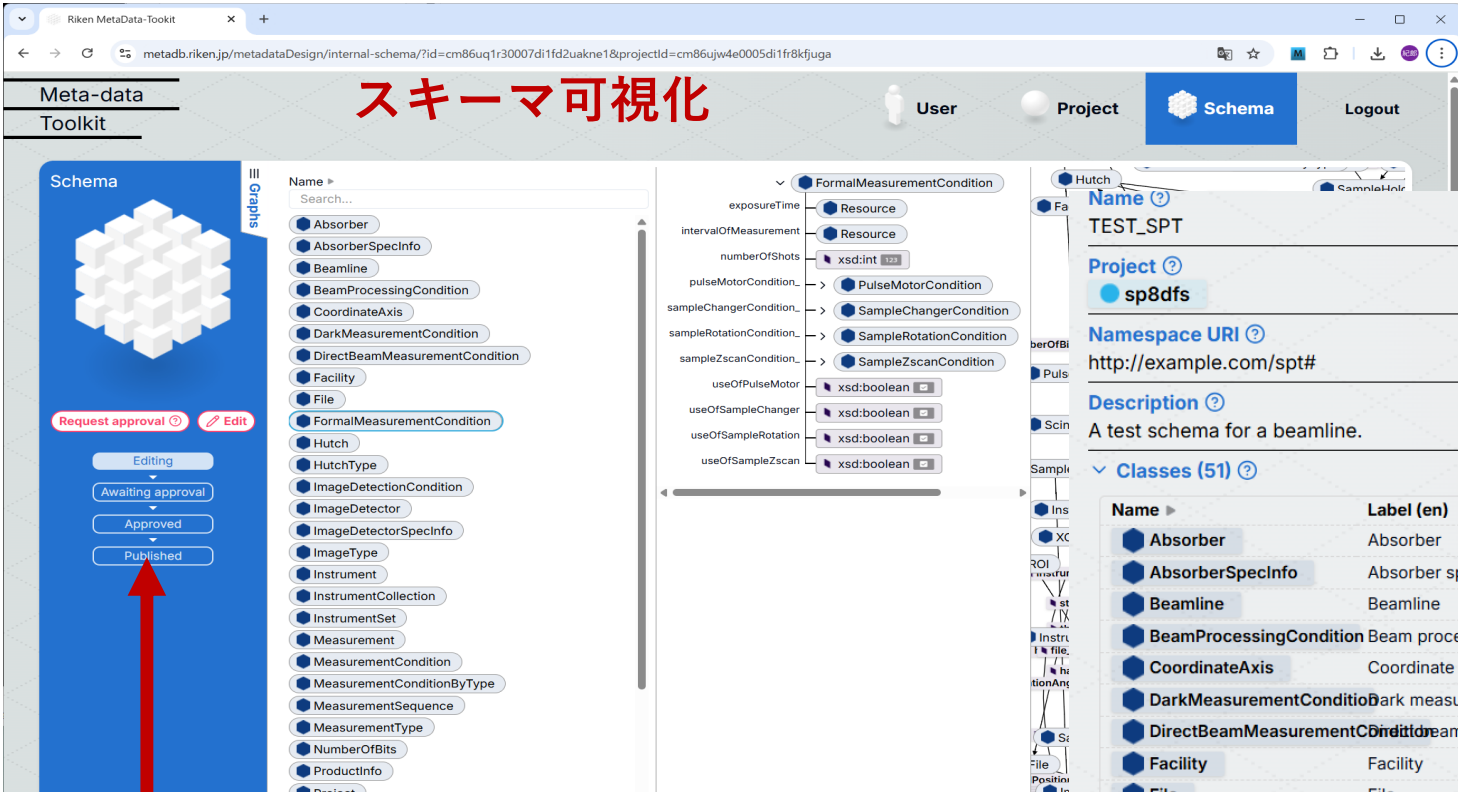
設計結果はJSON Schema, RDF形式で出力

学認による認証機能を導入し学認SPとして公開中

ユーザフィードバックを得て、特にGUIの改良を進めている

データ連携：メタデータ設計ツールの実装と公開

<https://metadb.riken.jp/metadataDesign>



The screenshot displays the Riken Metadata-Toolkit interface. On the left, a sidebar shows a list of schema elements like Absorber, Beamline, and Facility. The main area is titled 'スキーマ可視化' (Schema Visualization) and shows a detailed view of a schema element, 'FormalMeasurementCondition'. On the right, a panel titled '概念・語彙定義' (Concept/Vocabulary Definition) shows the details for a specific schema element, 'TEST_SPT', including its project, namespace URI, and a list of classes with their labels in English and Japanese.

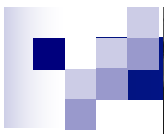
スキーマ可視化

概念・語彙定義

スキーマ設計・評価・公開の制御が可能

- 理研では、分野を問わず開発済メタデータスキーマ・オントロジの登録を進めている
- 大阪大学では、コアファシリティ機構に試験導入し、共用機器から生まれる測定データの全学での共有・管理を可能とするメタデータスキーマの設計を進めている

10月10日 13:45～ 研究データエコシステム事例紹介のセッションでもご紹介



Meta-data Toolkit



Manage Users



Manage Projects



Manage Schemata

To Login page

大規模言語モデルを活用したメタデータ検索 理研メタデータベース・チャットの試作

現状: メタデータを詳細に検索するにはクエリ言語の記述が求められ、経験やノウハウが必要

目的:

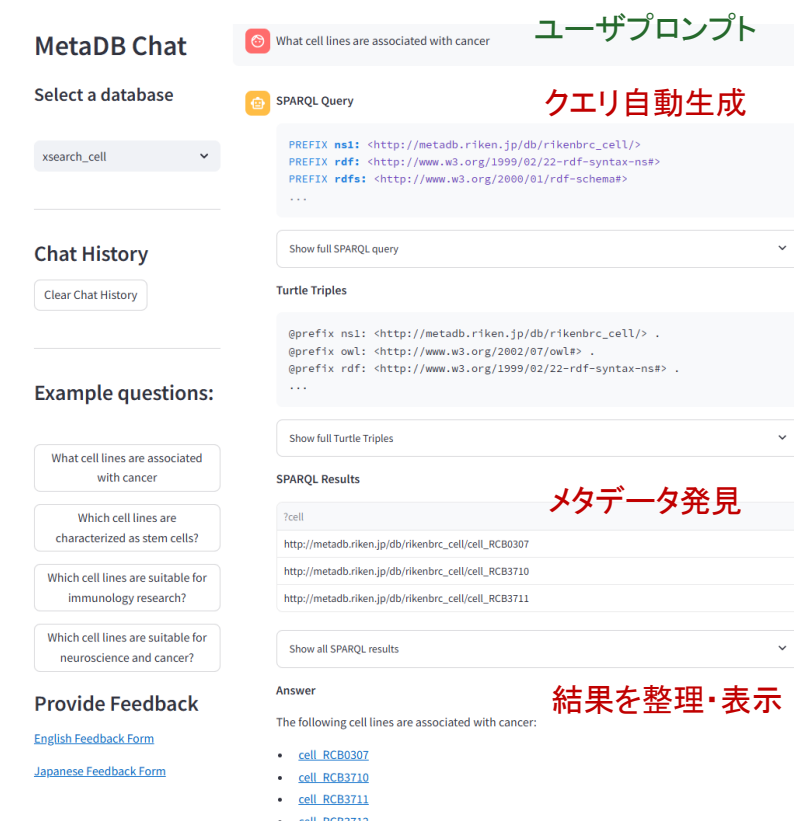
生命科学公開メタデータ(RIKEN MetaDatabase) を用いて LLM+RAG による検索枠組みを試作・検証し、その成果を実際のデータ管理基盤における研究データ検索へと展開できる要素技術を確立する。

進捗:

- 初期版を公開: <https://metadb.riken.jp/chat/>
- 大規模言語モデルopenLLaMA を活用し実装中
- スループット向上を目指した最適化を試験中
- 順次生命科学メタデータの取り込みを実施中
- AIとデータをつなぐ標準プロトコルとしてMCP (Model Context Protocol)を採用し、MCPサーバとして実装中

今後の計画:

十分なメタデータが付与されていないデータセットについても検索できるよう、論文データ等も取り込んで拡張版を試作し、その検索能力を評価する。



MetaDB Chat

Select a database: xsearch_cell

Chat History: Clear Chat History

Example questions:

- What cell lines are associated with cancer
- Which cell lines are characterized as stem cells?
- Which cell lines are suitable for immunology research?
- Which cell lines are suitable for neuroscience and cancer?

Provide Feedback: [English Feedback Form](#), [Japanese Feedback Form](#)

ユーザプロンプト

What cell lines are associated with cancer

クエリ自動生成

SPARQL Query

```
PREFIX ns1: <http://metadb.riken.jp/db/rikenbrc_cell/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
...
```

Show full SPARQL query

メタデータ発見

Turtle Triples

```
@prefix ns1: <http://metadb.riken.jp/db/rikenbrc_cell/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
...
```

Show full Turtle Triples

SPARQL Results

| ?cell |
|---|
| http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB0307 |
| http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB3710 |
| http://metadb.riken.jp/db/rikenbrc_cell/cell_RCB3711 |

Show all SPARQL results

結果を整理・表示

Answer

The following cell lines are associated with cancer:

- [cell_RCB0307](#)
- [cell_RCB3710](#)
- [cell_RCB3711](#)
- [cell_RCB3712](#)

MetaDB Chat

Ask a question about your data...



Select a database

xsearch_animal



Chat History

Clear Chat History

Example questions:

What cell lines are associated with cancer

Which cell lines are characterized as stem cells?

Which cell lines are suitable for immunology research?

Which cell lines are suitable for neuroscience and cancer?

Provide Feedback

[English Feedback Form](#)

[Japanese Feedback Form](#)

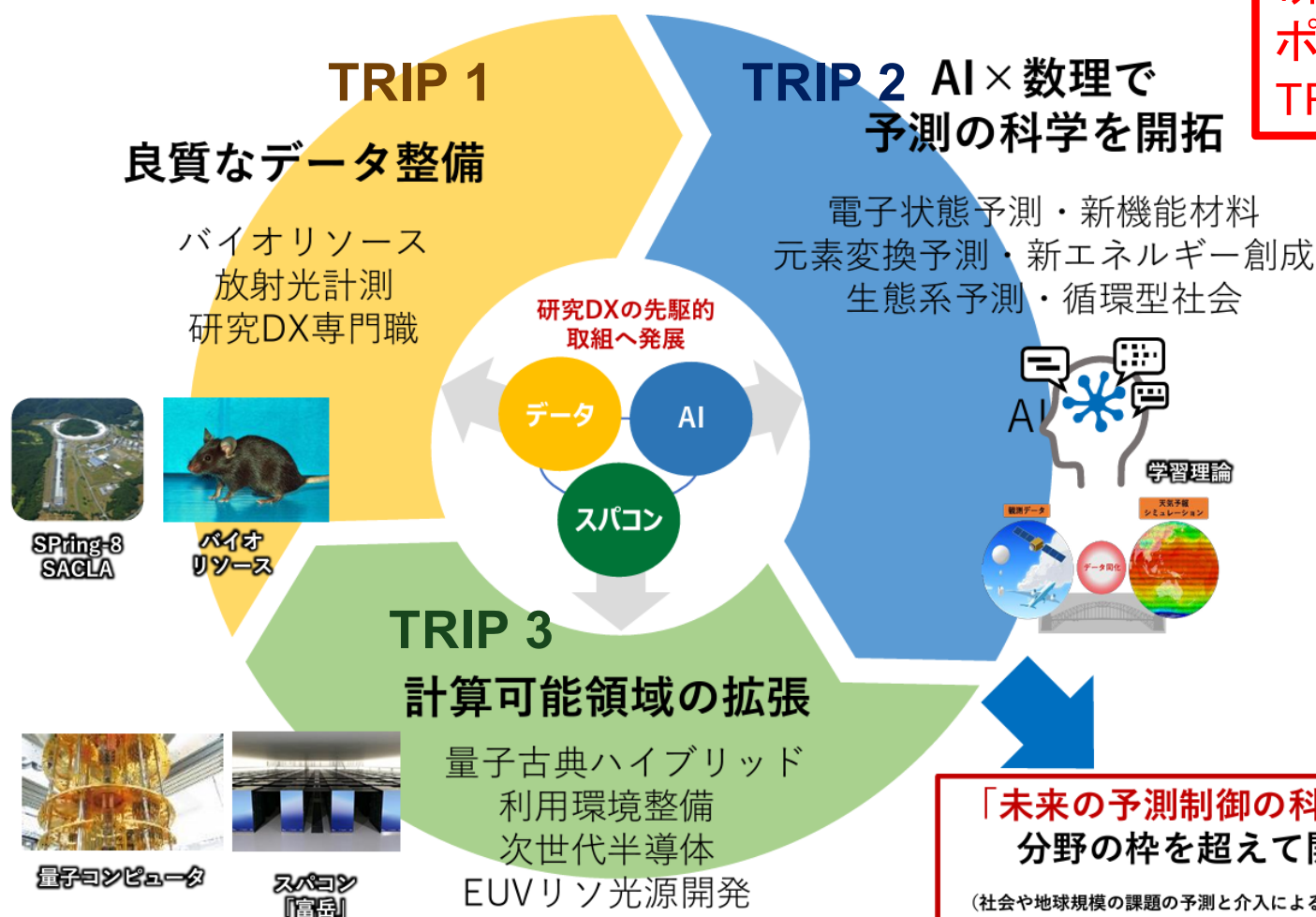
clide

理研TRIP:

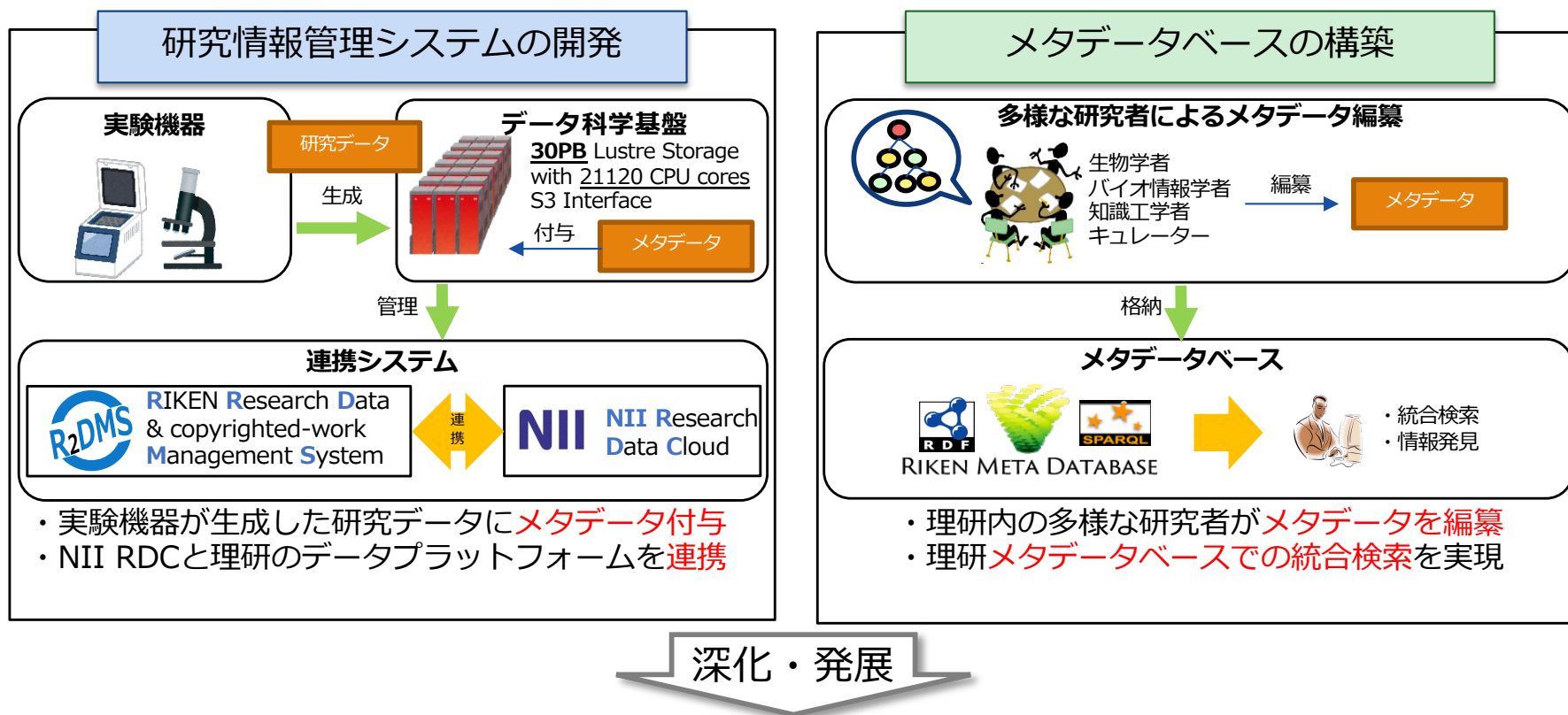
Transformative Research Innovation Platform of RIKEN platforms

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

研究データエコシステム、リ
ポジトリ・データ基盤連携は
TRIPを支える重要な基盤



成果を我が国の基盤構築構築に繋げる



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