

## 中核機関群 活動・計画報告

# 研究データ基盤高度化チーム (NII)

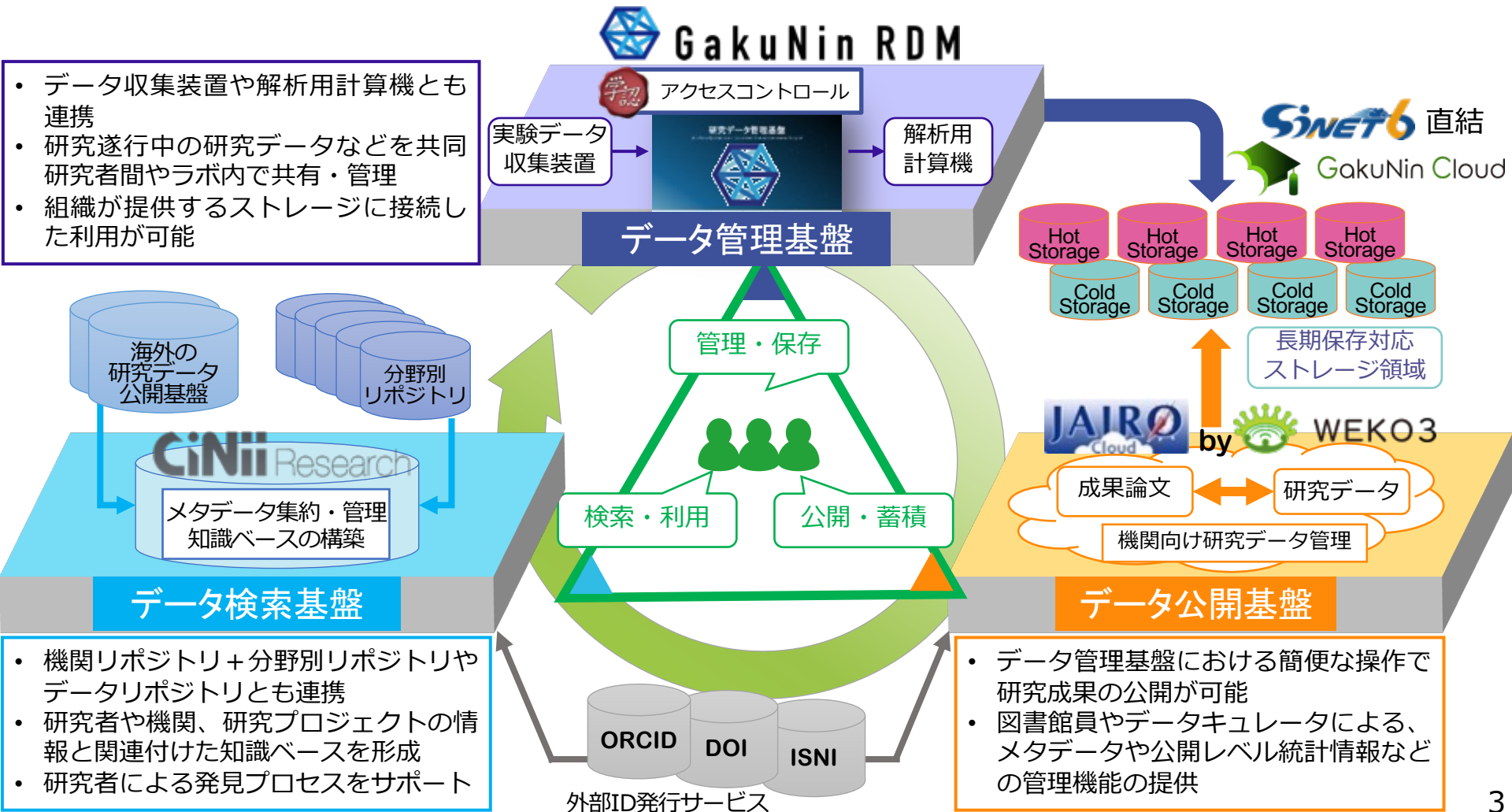
国立情報学研究所  
込山 悠介

AI等の活用を推進する研究データエコシステム構築事業

# 研究データ基盤高度化 (NII)

# 研究データ基盤 NII Research Data Cloud

2017年から開発開始 ⇒ 2021年から運用



- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能

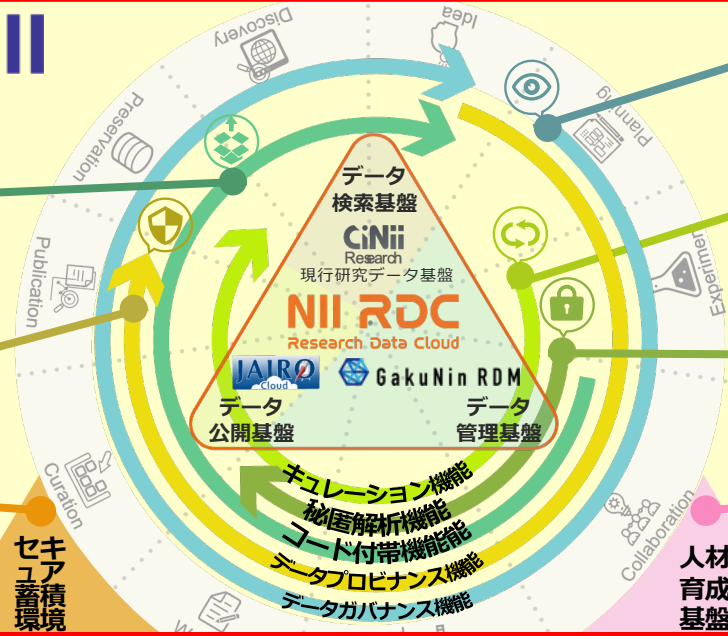
- 機関リポジトリ+分野別リポジトリやデータリポジトリとも連携
- 研究者や機関、研究プロジェクトの情報と関連付けた知識ベースを形成
- 研究者による発見プロセスをサポート

- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

# 研究データ基盤高度化チーム

研究データ基盤高度化チーム NII  
NII Research Data Cloudを  
7つの側面から機能拡張

研究データ基盤の機能実装



**データガバナンス機能 管理**  
計画に基づきデータ管理等を機械的に支援し、DMPをプロジェクト管理に不可欠な仕組みへと変革

**キュレーション機能 流通**  
専門的なキュレーションを実践できるエコシステムを構築し、データ再利用の促進に寄与

**秘匿解析機能 保護**  
秘密計算技術で機微な情報も安心して解析できる環境の提供で、新しいデータ駆動型研究の世界を開拓

**人材育成基盤 育成**  
RDMに必要なスキルを学ぶ環境を提供し、全ての研究者を新しい科学の実践者へと育成

**活用** **コード付帯機能**  
データ・プログラム・解析環境のパッケージ化と流通機能を提供し、研究成果の再現性を飛躍的に向上

**信頼** **データプロビانس機能**  
データの来歴情報の管理から利用状況を把握でき、データ公開へのインセンティブモデルを提供

**蓄積** **セキュア蓄積環境**  
安全で強固なデータの保存・保護機能を有する超鉄壁ストレージを提供し、機微な情報も安心して保全

## プラットフォーム連携チーム

**理化学研究所**

リーダ機関

- ・ 機関内サービス等とNII RDCの連携機能の整理と設計
- ・ 計測機器等からの大量データを効果的に管理するための要件整理と機能開発
- ・ 管理対象となるメタデータ的设计と実証
- ・ 関連する高度化機能との仕様調整と共同開発

## 融合・活用開拓チーム

**東京大学**  
THE UNIVERSITY OF TOKYO

リーダ機関

- ・ 異なる分野間でのデータ活用やデータ連携に発展する取り組みを精査
- ・ 異なる分野間でのデータ活用やデータ連携に関する具体的なユースケースを創出
- ・ ユースケースをまとめたツールキットの作成とそれを用いた広報活動

## ルール・ガイドライン整備チーム

**名古屋大学**  
NAGOYA UNIVERSITY

リーダ機関

- ・ 研究データの活用に適した機械可読データの統一的な記述ルール設計
- ・ 研究データの公開に必要な要項や作業フローの整備
- ・ 研究データを適切に取扱うための指針のまとめ
- ・ 学内整備のための事例形成

## 人材育成チーム

**大阪大学**  
OSAKA UNIVERSITY

リーダ機関

- ・ 人材育成を主とした研究データ管理体制の構築を推進する学内組織構築の事例形成
- ・ 研究データ管理人材に求められる標準スキルに関する検討
- ・ 研究データ管理人材育成のためのカリキュラムの作成、オンライン学習コースの整備

基盤の活用に係る環境整備

中核機関群の代表からなる運営委員会が全体を統括し研究データエコシステムの全国展開に向けて共同実施機関を随時拡大

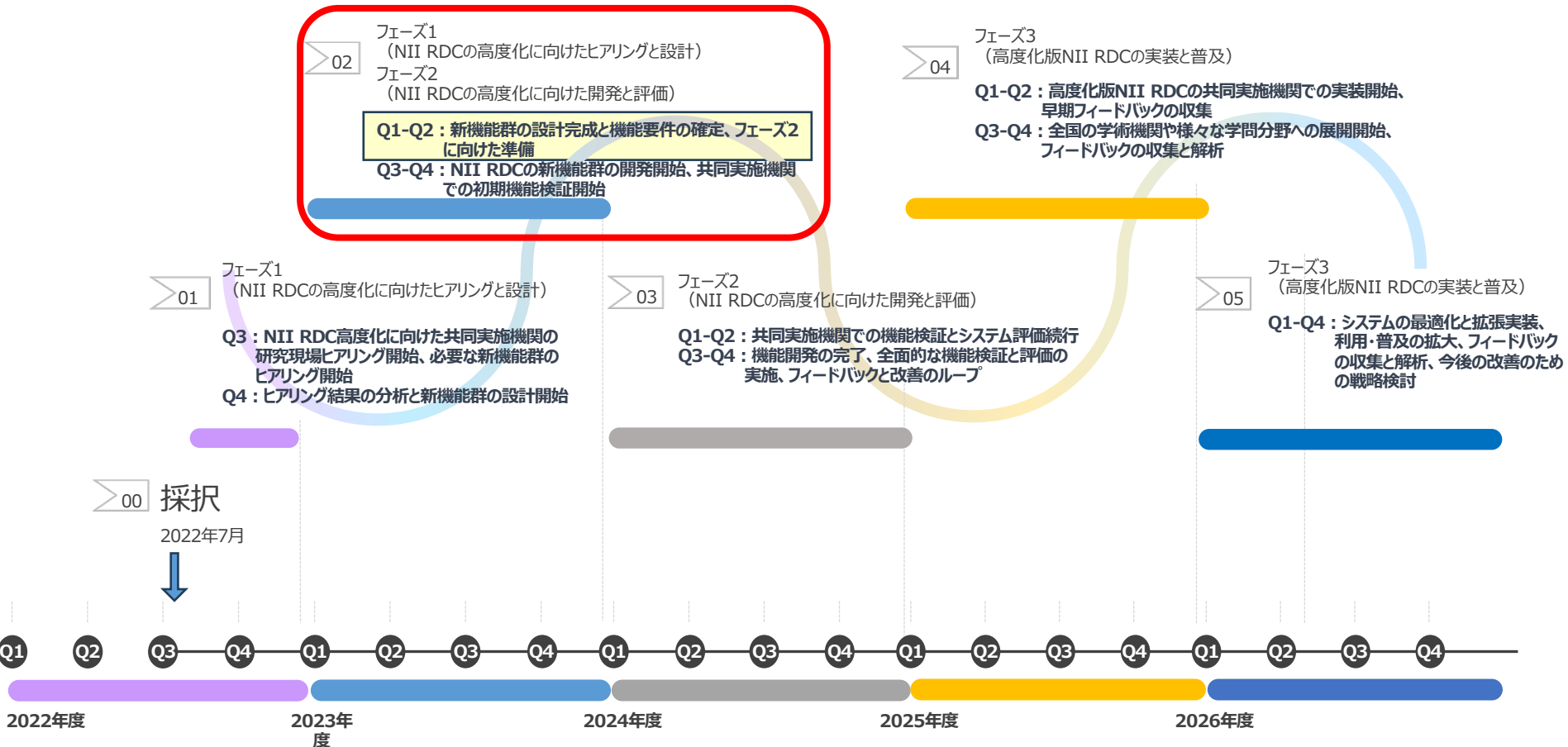


# 研究データ基盤高度化（NII）

## 2023年度

Q1-Q2：新機能群の設計完成と機能要件の確定、フェーズ2に向けた準備

Q3-Q4：NII RDCの新機能群の開発開始、共同実施機関での初期機能検証開始



# 研究データ基盤高度化（NII）

## — 先行開発機能群 —

---

1. データガバナンス機能（管理）
  - 実証実験版の開発
  - 機能評価版（2022年度開発）の運用・改善
2. コード付帯機能（活用）
  - GakuNin RDMプロジェクトのパッケージ化機能開発
  - 公開用RO-Crateメタデータスキーマの設計
3. 秘匿解析機能（保護）
  - お試し利用・実証実験（[トライアル実施中](#)）
  - 基盤サービス化に向けた要件検討
4. 人材育成基盤（育成）
  - 学認LMSに関する機能改修及び新機能開発
  - 教材の修正及び新規開発
  - 学認LMS利用機関の開拓

# データガバナンス機能の概要

## 課題

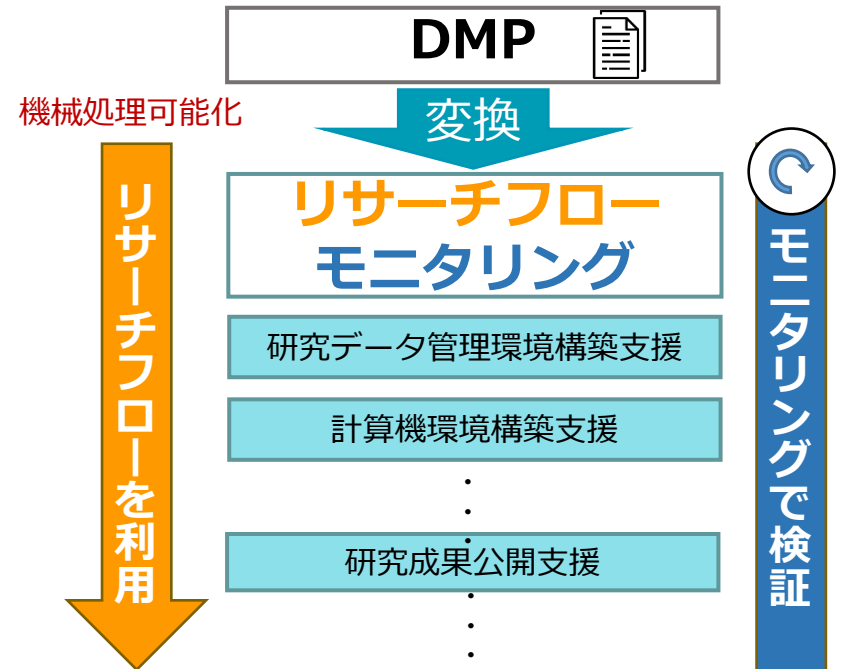
- 省庁や助成機関が要求する DMP を作成することが必要。しかしそれを研究サイクルに活かすことが難しい。
- ✓ 計画に沿った研究データ管理は研究者の裁量に...
- ✓ 研究管理部門も計画通りに研究データ管理がなされているか確認するのは手間...
- ✓ DMP が管理文書として蓄積される以外の使い道がない...

No.	データ名	データの説明	管理者	分類	公開レベル	秘匿理由
1	〇〇実証においてセンサより撮像したデータ及び関連データ	〇〇実証においてセンサより撮像したデータであり、道路の画像データ	〇〇研究所	委託者指定データ	レベル4 (広範な提供・利活用)	秘匿しない



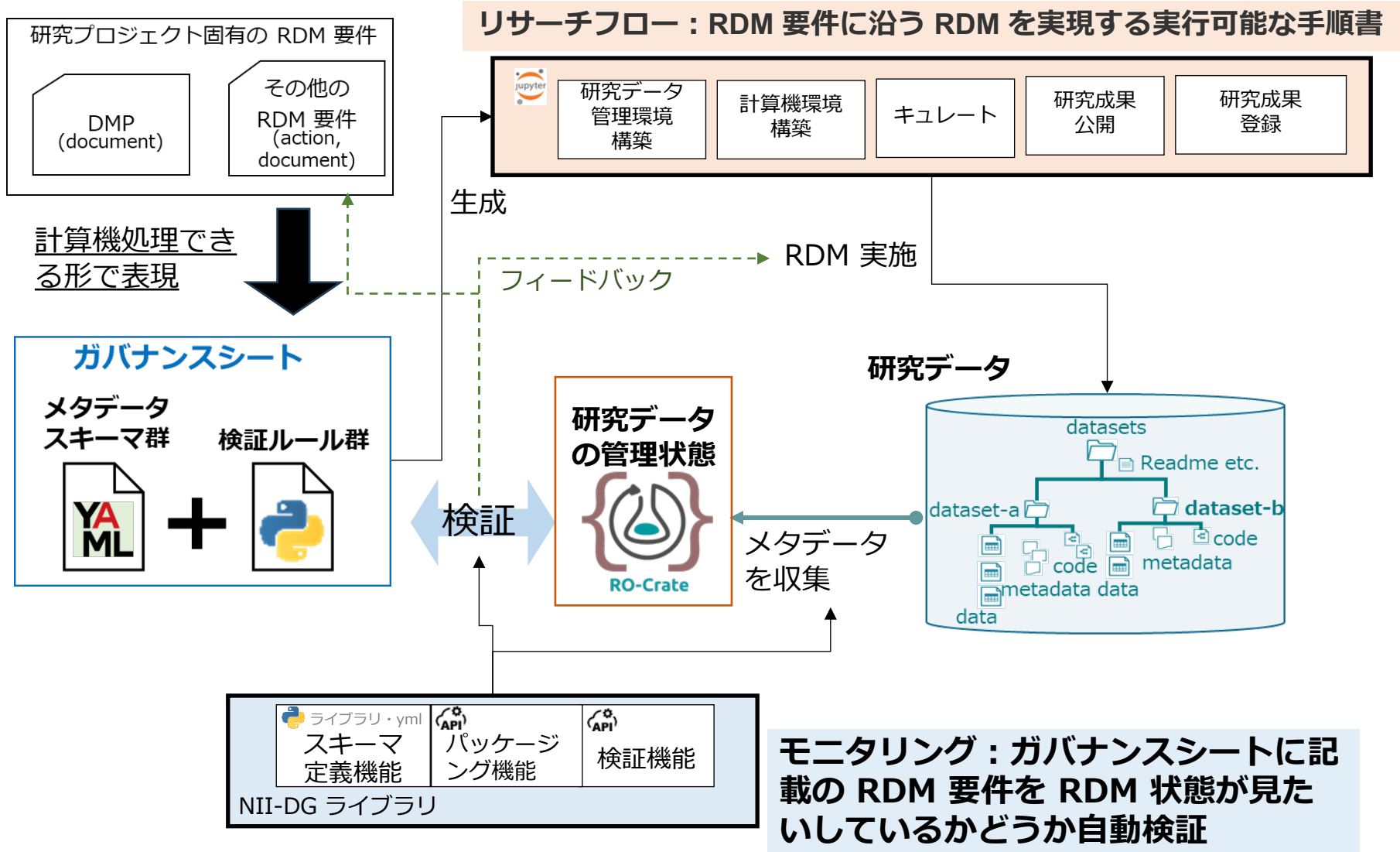
## 解決案

- DMP をもとに半自動で適切な研究データ管理環境を生成。
- ✓ DMP から生成されたリサーチフローで研究データ管理品質が担保される。
- ✓ DMP に沿った研究データ管理がなされていることをモニタリングにより継続的に検証し、研究データ管理品質が担保される。





# 現行のデータガバナンス機能



# 機能評価試験版サービス

ボタンクリック+簡便な入力により、リサーチフローを利用する研究および研究データ管理状態の検証が可能。

**研究実行環境用の手順書 (notebook)**

研究フロートップページ

研究プロセスで生じる研究者のタスクを半機械的に実行支援します。  
リサーチフロー  
しています。  
また、個々の  
Jupyter Note

▼ リサーチ  
「0.リサーチ  
援機能を実行  
初期セットア  
初期セットア  
を押下するこ  
「フェーズの

**実験実行環境用の手順書**

▼ 1. 実験実行環境にアクセスする

以下のセルを実行した後に出力されるボタンをクリックして実験実行環境を作成してください。  
作成に成功すると自動的に「実験フロートップページ」が表示されます。  
「実験フロートップページ」の説明を参考に、「フロー図を作成する」のセルを実行し、実験フロ  
実験フロー一回が表示されたら、実験フローの「0.実験フロー-実行準備」フェーズの「初期セット  
って上から順に必要なセルを実行し、実験実行環境のセットアップを行ってください。  
初期セットアップまで完了する途中で、実験実行環境のページを開いてしまった場合は、このタ  
実験実行環境について以下の点にご留意ください。

- 実験環境が維持されるのは最後に利用した日から30日間です。実行環境が消えてしまった場合  
タンを押下し、新しく作成し直してください。
- 作成できる実行環境は10個までです。

```

In [6]: In [0]: In [0]:
import os, urllib
from IPython.display import HTML
os.chdir(os.environ["HOME"] + "/WORKFLOWS")
from utils import display_util
os.chdir(os.environ["HOME"] + "/WORKFLOWS/FLOW")
from util_scripts import utils

# リモートリポジトリのURLを最新化する
is_new_private = utils.update_repo_url()

# リモートリポジトリのURLを取得する
remote_https_url = git.config.get(remote.origin.url)
for item in remote_https_url:
    remote_https_url = item
# 非公開リポジトリの場合ユーザー認証を行う
if is_new_private["is_private"]:
    utils.initial_sign_user_auth_without_email()
else:
    display_util.display_info("新規実験用の実行環境を作成します。<br>以下のボタンをク  
url = "https://binder.cs.rcos.nii.ac.jp/v2/git/" + urllib.parse.quote(remote_https_url)
    display(HTML(f"<button onclick='window.open(\"{url}\")>実行環境を作成する</butt

```

実験環境を作成する

**メタデータ検証用の手順書**

▼ 1. 研究リポジトリのメタデータを意用する

メタデータを検証するための準備として、研究リポジトリのメタデータの取得と整形を行います。

▼ 2. メタデータを検証する

メタデータの検証と検証結果の生成を行う検証サービスに、メタデータの検証を依頼します。

▼ 3. 検証結果を確認する

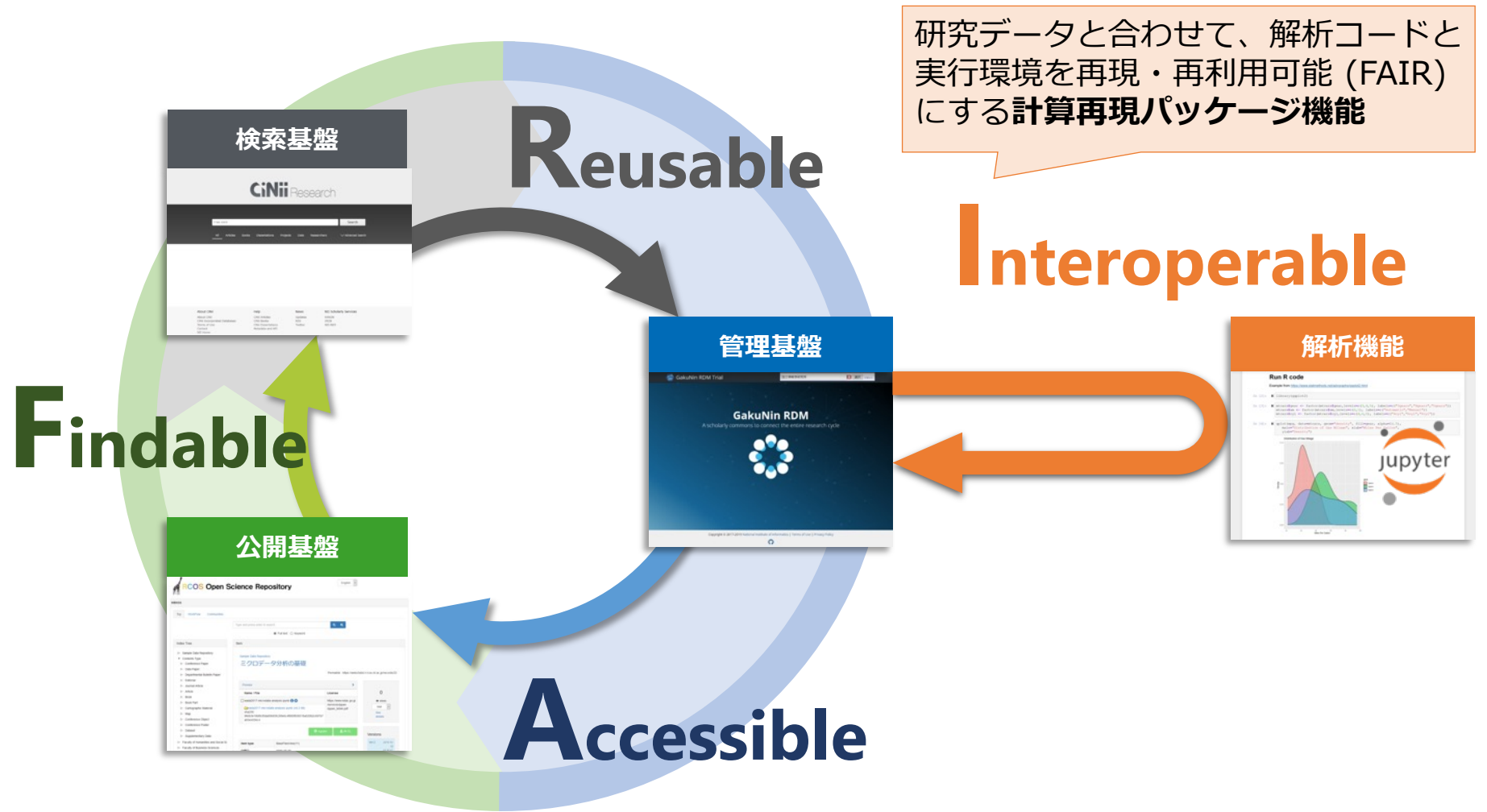
検証サービスから検証結果を取得し、結果を出力します。

▼ 4. 研究リポジトリに同期する

このタスクの実行結果を研究リポジトリに同期します。  
検証結果を同期するか破棄するかは、「4.1. 検証結果を研究リポジトリに同期するか破棄するかを選択する」で選択できます。

▼ 4.1. 検証結果を研究リポジトリに同期するか破棄するかを選択する

# コード付帯機能の概要



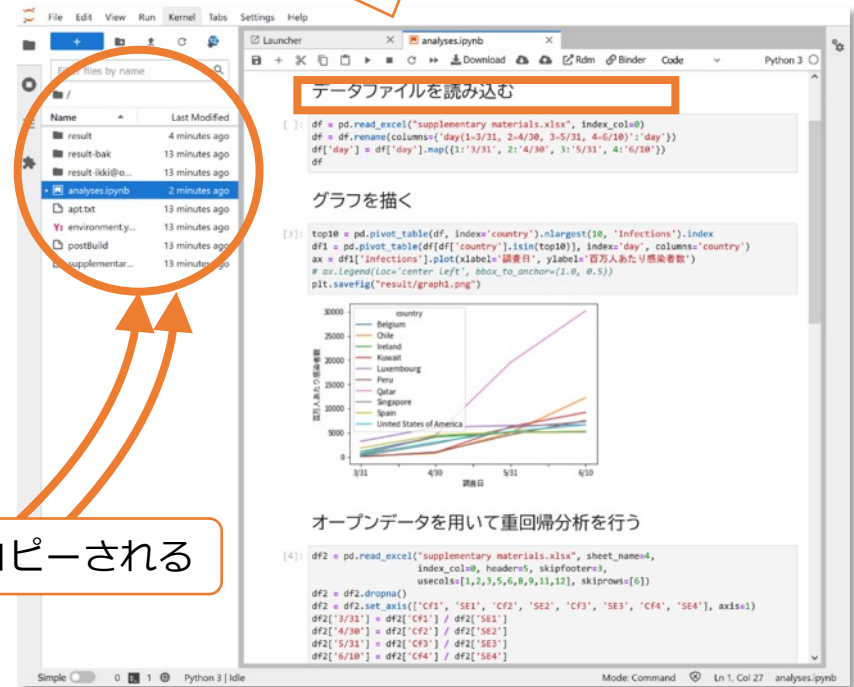
# 管理基盤のデータ解析機能として

コード付帯機能は管理基盤と連携し、データ解析環境を簡単に構築可能

1 分析環境を選んで作成ボタンを押す



3 分析サーバー上で、そのファイルを読み込むプログラムを書いて実行する



2 ファイルが分析サーバーにコピーされる

分析結果を管理基盤に書き戻す

4

# 計算再現パッケージ機能

公開基盤と連携し先行研究のデータ分析環境を再現、実行可能に

## 管理基盤

GakuNin RDM

(1) RDMパッケージエクスポートを選択

(2) エクスポート先を選択

(3) WEKOアドオンを経由してRO-Crateをアップロード

- 管理基盤で解析したデータセットと計算環境をパッケージとして公開基盤で公開。
- 先行研究のパッケージを管理基盤に取り込み再現実験する事も可能に。

## 公開基盤

WEKO3

RO-Crate ZIP

(4) プロジェクトがインポート可能に

# 秘匿解析機能の概要

## 秘匿解析機能とは

- データを秘匿（暗号化）したまま解析する機能

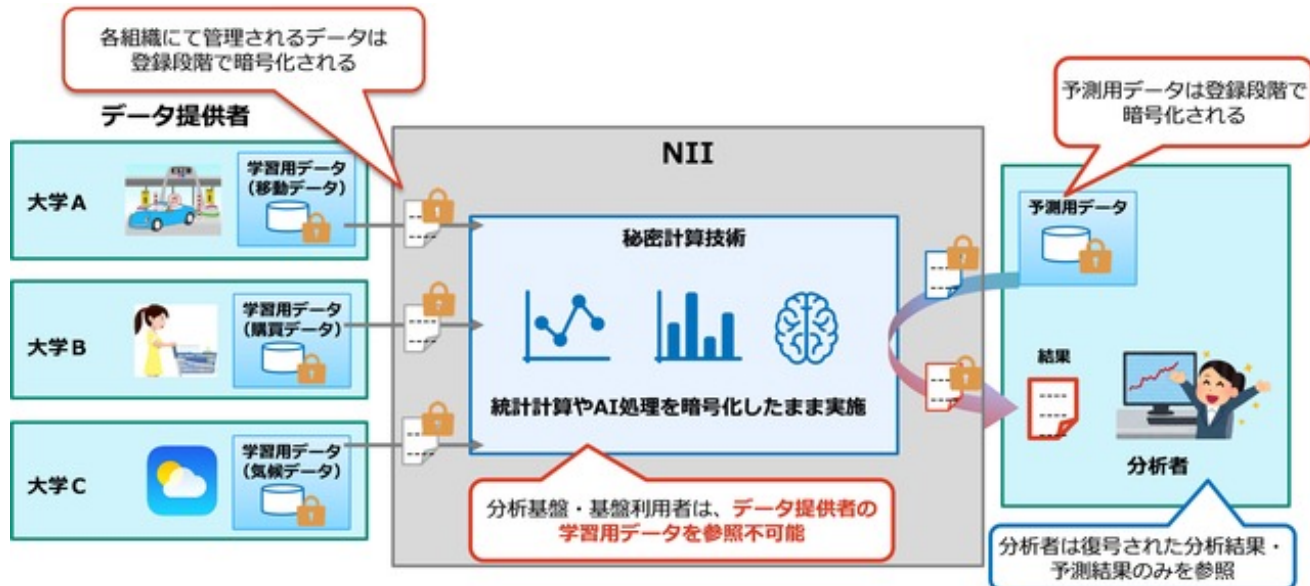
## 狙い

- データ駆動型研究では、幅広いデータが利用可能であるべき
- 個人のプライバシーや企業のビジネスに関わるデータは公開が難しい
- これらデータも安全・安心に活用できる環境を提供したい

**目標** 2027年までにRDCのサービスとして提供する

## 目標達成のためのアプローチ

- コア技術は民間の秘密計算技術等を評価・導入
- 実証を進めて価値を浸透
- 基盤サービス化の要件明確化



# 人材育成基盤の概要

## 学認LMS (https://lms.nii.ac.jp)

### ●開講コース

### 情報基盤スタッフ向けの研究データマネジメント

情報基盤スタッフにどのようなRDM支援活動が求められるか、またそのために必要な知識、技術について解説します。(学習時間の目安:約1.9時間)

### 研究者のための研究データマネジメント[2022年度版]

研究支援者としての目線から、大学や研究機関等に所属する研究者の方に向けて作成された教材です。研究データ管理の場面に応じた12のテーマ別に分かれており、研究者自身が本教材によって必要な知識を得ることを想定しています。(学習時間の目安:約4時間)

### 研究データ管理サービスの設計と実践[2022年度版]

研究データには、どのように生成され、保存され最終的に再利用されていくのかといったライフサイクルがあります。この講座では、研究データのライフサイクルに沿った形で、サービス設計や研究前の支援、研究中の支援、研究後の支援、そして日常的な支援について学びます。(学習時間の目安:約4.5時間)

### GakuNin RDM利用支援コース (公開準備中)

これからGakuNinRDMを使い始める方や、もっと活用したい方のためのコース。トピックを選んで学習します。

### オープンサイエンス時代の研究データ管理

入門編



- コース修了判定条件を満たすと、修了バッジ発行。
- 国立情報学研究所教育研修事業研究データ管理セルフラーニング教材の修了証書を取得可能。

### ●教材構成

マイクロコンテンツ教材 (合成音声)  
+ 確認テスト



### ●機関管理者機能

利用機関ごとに受講者の受講履歴を管理する機能を提供。

### ●その他のオプション機能 (テスト運用)

- 受講履歴取得API
- 自機関限定コース作成機能
- 機関限定コースの共有機能
- ラーニングアナリティクス機能
- マイクロコンテンツ教材作成機能
- 自機関LMSとのLTI連携

# 研究データ基盤高度化（NII）

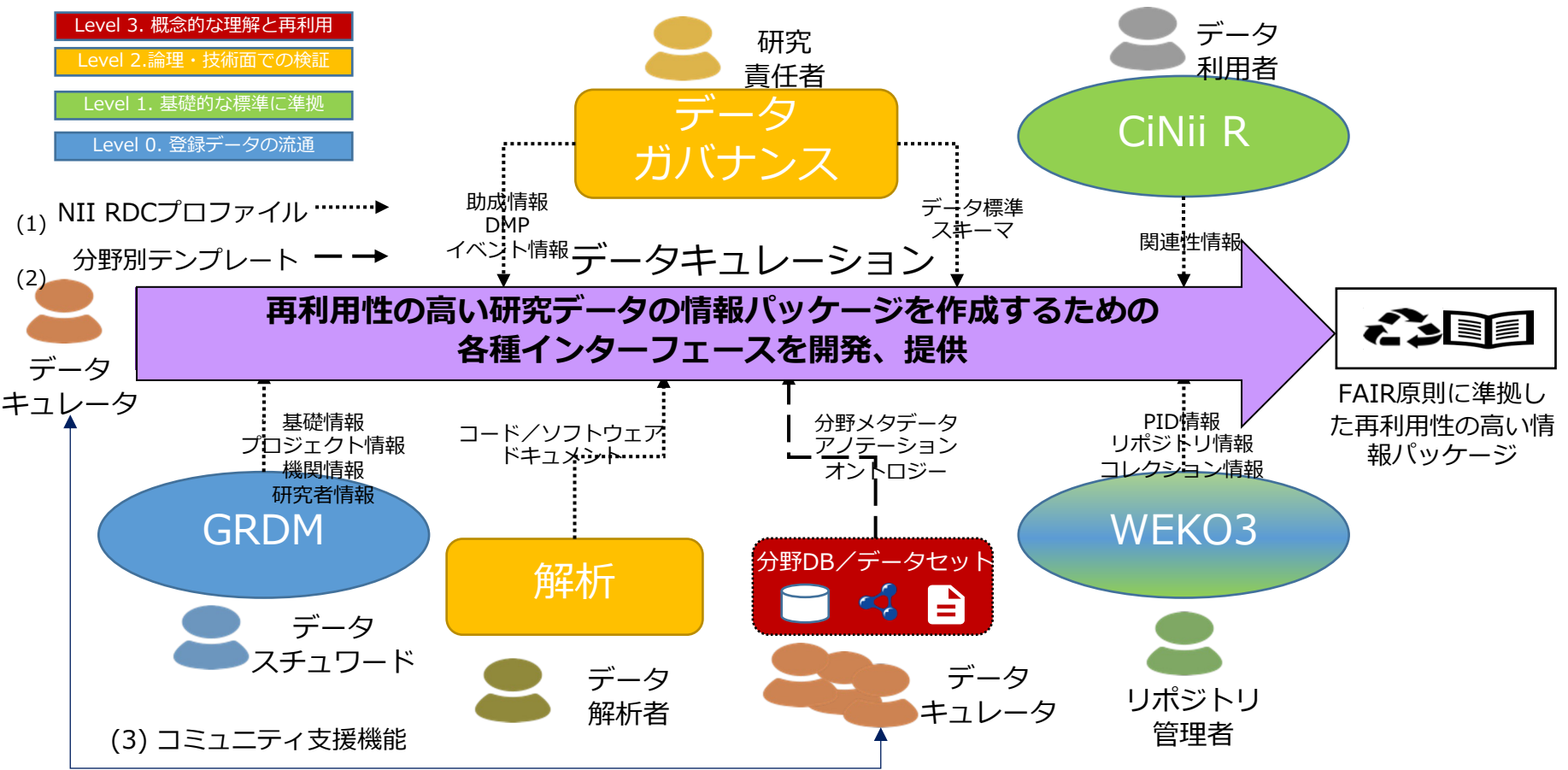
## —後発開発機能群—

---

5. キュレーション機能（流通）
  - プロトタイプシステムの開発
  - 10月からの機能評価実施に向け、共同実施機関を開拓
  
6. データプロビナンス機能（信頼）
  - 要件抽出のためのヒアリング
  
7. セキュア蓄積環境（蓄積）
  - セキュアストレージシステムの調達・構築
  - セキュリティ要件適合に向け、要件定義



# キュレーション機能の概要



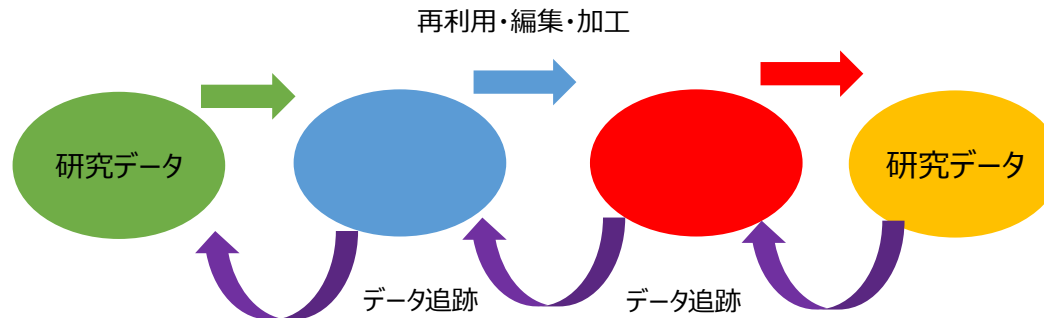
# データプロビナンス機能

## 主機能

研究データプロビナンス(来歴)情報の適切な管理

## 実施内容

来歴情報に関するメタデータの適切な収集・保存  
 データ原本、実験手順、分析・解析の記録  
 成果に対する正当性検証結果の保持



- 要件抽出のためのヒアリング 4機関 (JAXA, 極地研, 核融合研究所, NIMS)
- ヒアリングを受けて要件抽出中
- 既存のRDC基盤との連携のための調整中

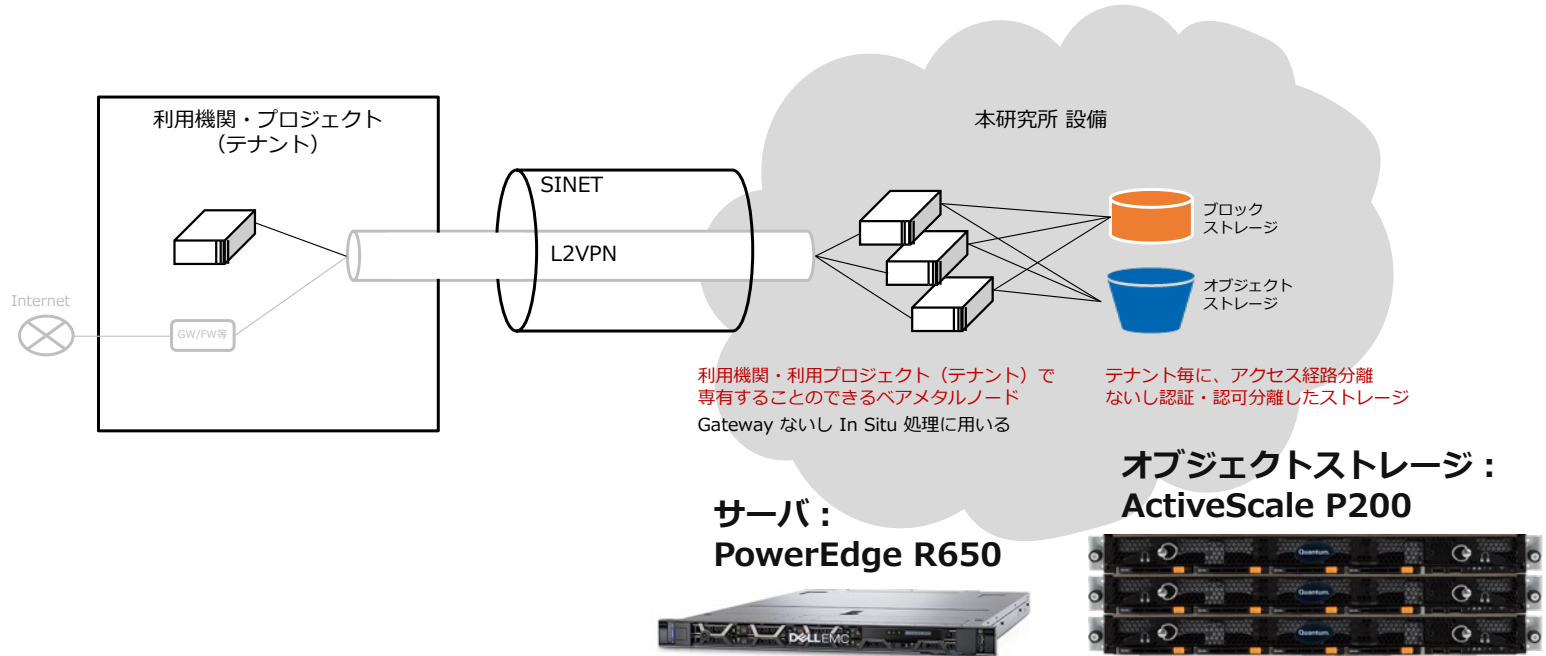
**研究データの真正性・来歴・出所を保証し利用をモニタリング**

# セキュア蓄積環境の概要

高セキュリティのオブジェクトストレージとサーバをセットで提供するクラウドサービスを実現

高セキュリティ要件のポイント

- 日本国内プロバイダ企業が運営する物理セキュリティの高い国内データセンタに設置
- サーバおよびストレージの物理資源に関してはNII専用(当面NII資産)
- SINET6 L2VPN環境に閉じたアクセスのみを提供



# 他機関への展開等の波及効果につながる取組み

- 機能改善要望への対応
  - NII RDC各基盤（機能）における、ヒアリング、機能評価試験、実証実験のフィードバックを受けた機能追加・改修
  - 融合・活用開拓チームと連携した、ユースケース課題公募事業の採択者からの、リクエストへの対応した分野対応
  - プラットフォーム連携チームと協働した、先端分野（生命科学、放射光科学等）向け機能の拡張
  - ルール・ガイドラインチームのポリシー整備と歩調を合わせたデータガバナンス機能の追加・拡張
  - 人材育成チームや地域コンソーシアムと連携した、分野データの集積方法の手法の検討

**RCOS**  
rcos-ext@nii.ac.jp