

# オープンサイエンス と 研究データ

国立情報学研究所 山地 一禎

琉球大学 研究データ管理セミナー 2024年2月15日

(於) 琉球大学附属図書館2階ラーニング・コモンズ



# 学術論文における「研究データ」の取り扱い

Elsevier出版学術誌の データ共有ポリシー率

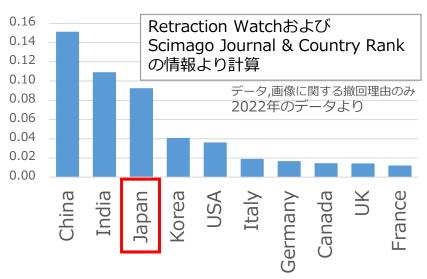


(n=2986件、2023年3月)

%

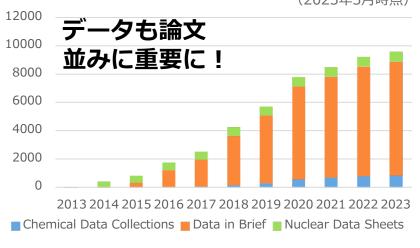
### 研究公正への対応 研究成果の信頼性確保

#### 論文出版数に対する同年の撤回通知数の割合



### 研究成果の再利用 研究の発展・加速化

Elsevierのデータジャーナル累積論文数



(データジャーナル:データとその解説を成果とする論文)

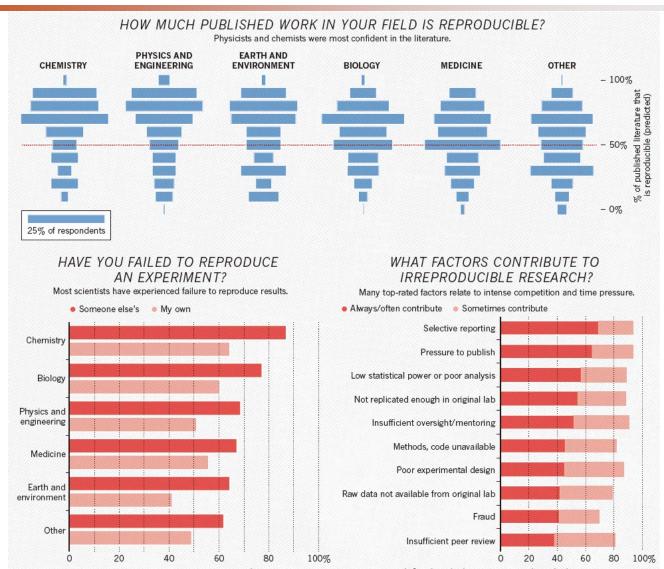
### 研究データの管理・公開が不可欠!



**Reproducibility Crisis** 7% Don't know 1,576 3% **Researchers Surveyed** No, there is no crisis IS THERE A REPRODUCIBILITY A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. BY MONYA BAKER 38% 52% **Yes**, a slight crisis **Yes**, a significant crisis



# **Reproducibility Crisis**



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016). https://doi.org/10.1038/533452a



# オープンサイエンスとは?

「オープンサイエンスとは、<u>公的研究資金を用いた研究成果(論文、生産された研究データ等)</u>について、科学界はもとより産業界及び社会一般から広く容易なアクセス・利用を可能にし、知の創出に新たな道を開くとともに、効果的に科学技術研究を推進することでイノベーションの創出につなげることを目指した新たなサイエンス」

内閣府『我が国におけるオープンサイエンス推進のあり方について〜サイエンスの新たな飛躍の時代の幕開け〜』(2015年3月30日)http://www8.cao.go.jp/cstp/sonota/openscience/



# オープンサイエンスの意義

### 1. 情報流通の活性化による学問の進展の飛躍的加速化

分野を越えた新たな知見の創出や効率的な研究の推進等に 資するとともに、研究成果への理解促進や研究成果の更な る普及につながる

### 2. 研究の重複回避による研究開発投資の効率化 作業の重複を避け、コストの削減・研究の効率化を図る

### 3. 研究成果の再現性、透明性の確保

社会からの信頼を確保、研究不正を回避することで科学技術の進展と研究活動の透明性、公正性を確保する

### 4. 科学技術外交

科学技術先進国として地球規模研究への対応、国際的共同研究・研究交流を図り、国際的競争力を強化する



### オープンサイエンス時代の 研究データ基盤構築に向けた国内の政策的経緯

2015年3月:内閣府「国際的動向を踏まえたオープンサイエンスに関する」

検討会|報告書

- 2016年1月:政府「第5期科学技術基本計画」
- 2016年2月: 文部科学省 科学技術・学術審議会 学術分科会 学術情報委員会 「学術情報のオープン化の推進について」
- 2016年5月: G7茨木・つくば科学技術大臣会合 つくばコミュニケ (共同声明)
- 2016年5月: **政府**「科学技術イノベーション総合戦略2016」
- 2016年7月:日本学術会議「オープンイノベーションに資するオープン

サイエンスのあり方に関する提言」(提言)

- 2017年6月:<mark>政府</mark>「科学技術イノベーション総合戦略2017」
- 2018年6月:<mark>政府</mark>「統合イノベーション戦略」
- 2019年6月:<mark>政府</mark>「統合イノベーション戦略2019」
- 2020年6月:日本学術会議「オープンサイエンスの深化と推進に向けて」 (提言)
- 2020年7月:政府「統合イノベーション戦略2020」
- 2021年3月: 政府「第6期科学技術・イノベーション基本計画」
- 2021年6月:<mark>政府</mark>「統合イノベーション戦略2021」
- 2021年7月: **政府**「公的資金による研究データの管理・利活用に関する

基本的な考え方」

- 2022年6月:政府「統合イノベーション戦略2022」
- 2022年12月:日本学術会議 「回答・研究DXの推進 特にオープンサイエンス、

データ利活用推進の 視点から - に関する審議について」

• 2023年5月: G7仙台科学技術大臣会合 G7科学大臣コミュニケ(共同声明)

2022年6日・東京「紘令ノノベーニーン戦略2022」



# 2021年3月26日 第6期科学技術・イノベーション基本計画

#### 「新たな研究システムの構築(オープンサイエンスとデータ駆動型研究等の推進)」 における目標

#### 【目標】

・ オープン・アンド・クローズ戦略に基づく研究データの管理・利活用、世界最高水準のネットワーク・計算資源の整備、設備・機器の共用・スマート化等により、研究者が必要な知識や研究資源に効果的にアクセスすることが可能となり、データ駆動型研究等の高付加価値な研究が加速されるとともに、市民等の多様な主体が参画した研究活動が行われる。

# 【**科学技術・イノベーション政策において目指す主要な数値目標**】 (主要指標)

• 機関リポジトリを有する全ての大学・大学共同利用機関法人・国立研究開発法人において、2025年までに、データポリシーの策定率が100%になる。公募型の研究資金の新規公募分において、2023年度までに、データマネジメントプラン(DMP)及びこれと連動したメタデータの付与を行う仕組みの導入率が100%になる。

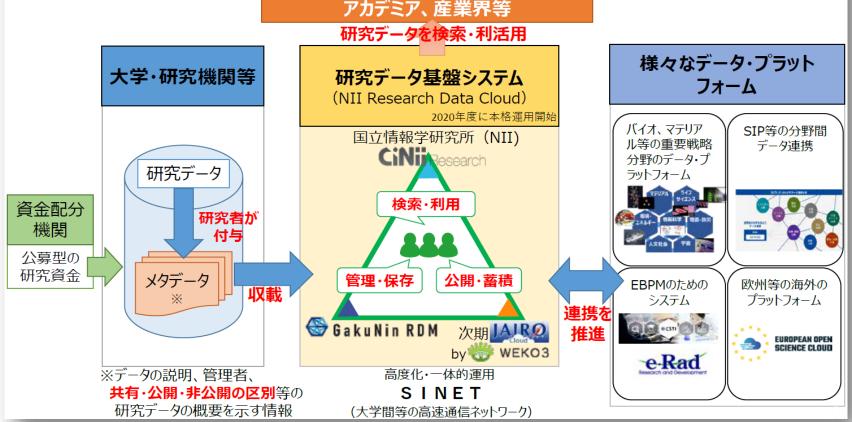


# 2021年4月27日 統合イノベーション戦略推進会議

#### 公的資金による研究データの管理・利活用に関する基本的な考え方について

研究データ基盤システムを中核としたデータ・プラットフォームの構築

- ○研究データの公開・共有を推進、産学官のユーザが<u>データを検索可能</u>
- ○<u>ムーンショット型研究開発制度</u>における試行(2020年度開始)、その後、次期<u>SIP</u>に導入
  - <u>全ての公募型の研究資金の新規公募分に導入(2023年度まで)</u>



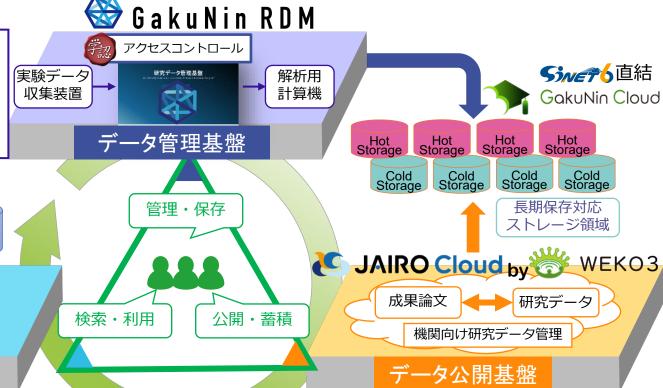


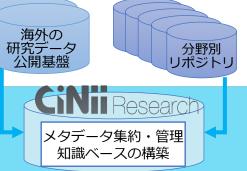
### 研究データ基盤: NII Research Data Cloud

### 2017年から開発開始 ⇒ 2021年から**運用開始**

(必要とされる機能を順次拡充)

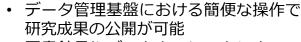
- データ収集装置や解析用計算機とも 連携
- 研究遂行中の研究データなどを共同 研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能





#### データ検索基盤

- 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



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



# 検索基盤

- 研究データなどの研究成果の総合的な検索サイト
- 関連情報を統合的に発見できる探索機能の提供





| データソース       | 成果物数       | 研究者数       | Project数  |
|--------------|------------|------------|-----------|
| CROSSREF     | 10,344,535 | 63,705,306 | 0         |
| JALC         | 11,137,631 | 18,783,938 | 0         |
| KAKEN        | 2,160,330  | 513,620    | 1,032,388 |
| IRDB         | 1,784,754  | 2,619,821  | 0         |
| CIB          | 13,011,327 | 8,299,962  | 0         |
| CID          | 714,435    | 701,885    | 0         |
| NDL          | 12,350,521 | 2,172,722  | 0         |
| NINJAL       | 391        | 948        | 0         |
| IDR          | 961        | 3,411      | 0         |
| RUDA         | 70         | 75         | 0         |
| ERAD         | 0          | 177,128    | 0         |
| INTEGBIO     | 1,917      | 817        | 0         |
| JDCAT        | 5,828      | 23,258     | 0         |
| PUBLIC_DATA  | 104,153    | 0          | 0         |
| LSDB Archive | 852        | 497        | 0         |
| MDR          | 14,295     | 3,751      | 0         |
| PubMed       | 35         | 654,599    | 0         |

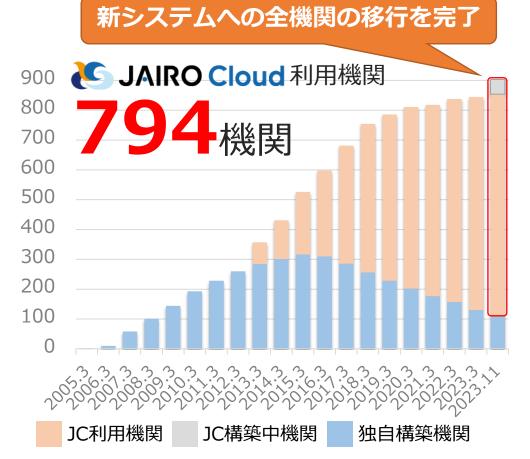


データソースを拡大中



# 公開基型

- オープンアクセスの推進を加速化させるための機能拡張
- 研究データの公開を支援するための機能提供



JSPSの人文学・社会科学データインフラ ストラクチャー構築推進事業でも活用

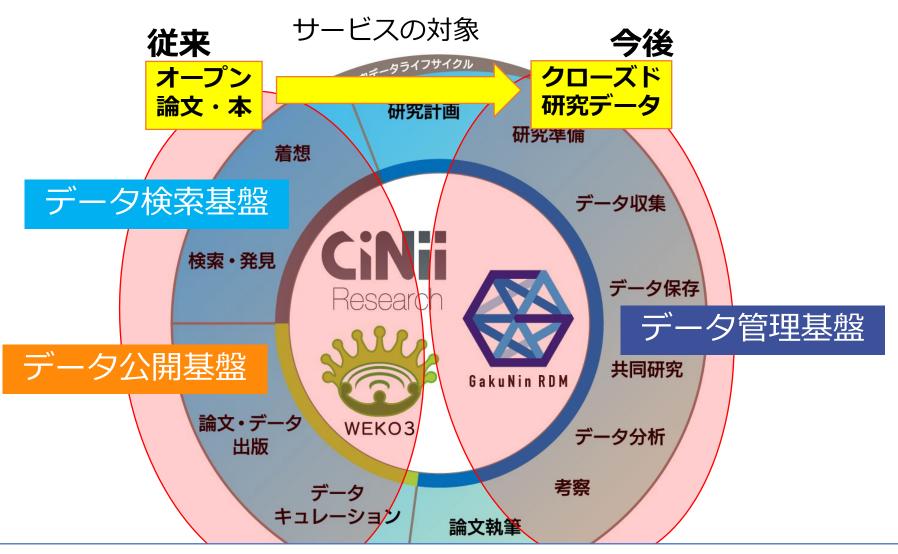


#### 他の研究分野での利用

- 生命科学(理研)
- 医学(AMED)
- 気象科学 (極地研)
- 複合科学(東大mdx) など



# Open/Closed戦略とNII RDC



オープン・クローズド戦略のもとでオープンサイエンスが実現できる基盤



# **NII Research Data Cloud**

データ管理基盤





# 研究データ管理とは?

### **RDM: Research Data Management**

ある研究プロジェクトにおいて使用された、 または生成された情報を、どのように組織化・ 構造化・保管・管理していくのかを指す言葉



研究データの取り扱い 計画の策定

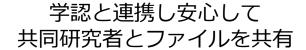
研究中の日々の情報の取り扱い

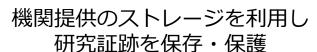
研究後のデータの公開 や長期的なの保管

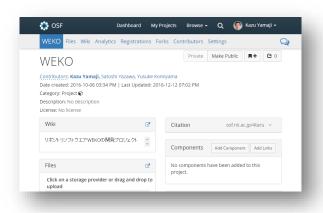


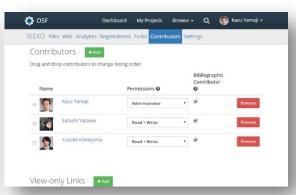
# 管理基盤 🍪 GakuNin RDMとは

#### 研究プロジェクト単位で ファイルなどを管理











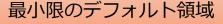
WebアプリケーションはNIIが提供

研究データ管理サービス

#### 持続的運用のための役割分担

エクストラストレージ

機関が管理するストレージに接続

















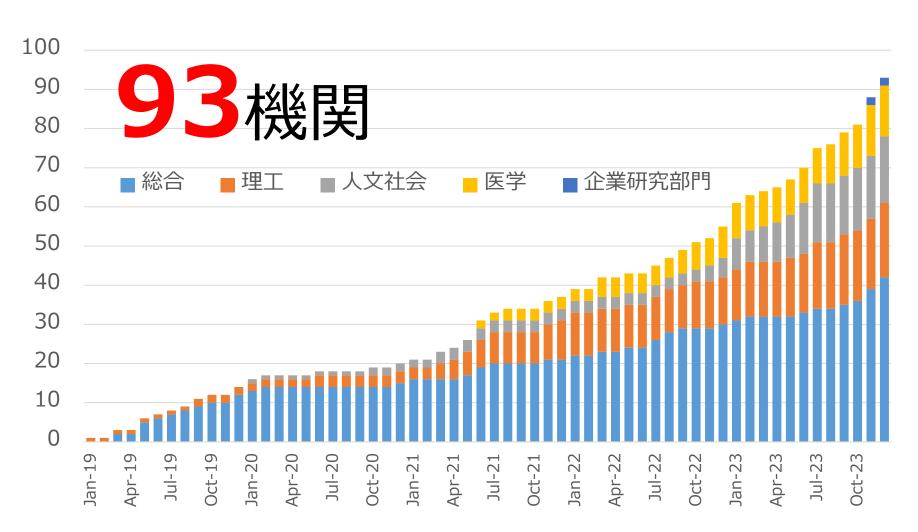


パブリッククラウド (プロバイダーDC)





### GakuNin RDMの利用機関数の増加



(2023年12月現在) 『GakuNin RDM 利用機関数』 https://support.rdm.nii.ac.jp/about/#a2



# RDM学内推進の取組み状況

### RDM、GakuNin RDM利用に関する全学研修等の開催

- 2022年···北海道国立大学機構(北見工業大学、小樽商科大学、 帯広畜産大学)、名古屋大学、広島大学、東邦大学、琉球大学
- 2023年…東北大学、京都大学、大阪大学、九州大学、鳥取大学、三重大学、電気通信大学、滋賀医科大学、同志社大(NIIで把握できたものに限る)

### • GakuNin RDM利用に関する学内サポートページの開設

- 名古屋大学、北海道大学、九州大学、信州大学、北陸先端科学技術大学院大学、三重大学、琉球大学、立命館大学、同志社大学
- GakuNin RDM:試行利用(特定部局)
  - →全学利用へ拡大申請
    - 2022年…慶應義塾大学、大阪電気通信大学、理化学研究所
    - 2023年…東京大学、京都大学、同志社大学、核融合研究所



オープンサイエンス & 研究データ管理 の

さらなる全国的な展開に向けて



# 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

我が国の研究力の飛躍的発展を図るため、NIIと大学・研究機関が強力に連携し、以下の事業を実践。

- 研究データ基盤の更なる高度化
- 多角的なアプローチにより研究データ基盤の全国展開

#### 中核機関

事務局担当 (専従PMの配置)



信頼性・再利用性・秘匿性の側面も考慮した上で、研究データの管理・蓄積・利活用・流通を実践するために必要となるNII RDCの高度化

#### 共同実施機関

プラットフォーム 連携チーム



#### 理化学研究所

共同実施機関が運用 するデータPFのデー タをNII RDCから検 索可能とする連携 融合・活用開拓 チーム



分野間でのデータ連携を前提とした複数のシーズ・ユースケースを創出

ルール・ガイドライン 整備チーム



名古屋大学

研究データの効果 的・効率的な活用の ためのルールやガイ ドラインの整備 人材育成 チーム

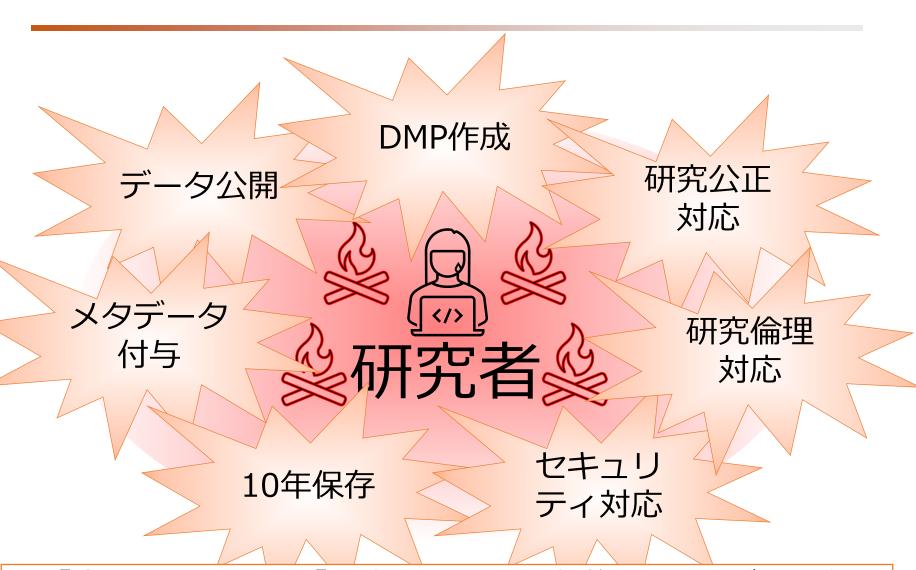


大阪大学

データマネジメント 人材のスキルセット の整理および必要な 教材等の整備



# 研究DX?に向けた研究者への過大な要請



「余計な作業」を「研究の一部」に転換するDXが不可欠



# 責務を意欲に変えるための新機能

#### 機関の責任

- 4-1 データポリシーの策定
- 4-2 メタデータの付与
- 4-3 人材・支援体制の整備
- 4-4 セキュリティの確保

#### 研究者の責任

- 6-1 管理対象データの決定
- 6-2 メタデータの付与
- 6-3 DMPの作成
- 6-4 研究データの保存
- 6-5 研究データの公開・共有
- 6-6 プロジェクトの終了後の取扱い

「公的資金による研究データの管理・利活用に 関する基本的な考え方について」より要請事項を抽出

DMP作成

データ公開データガバナンス

研究公正対応

コード付帯



キュレーション



研究者



秘匿解析



セキュア蓄積



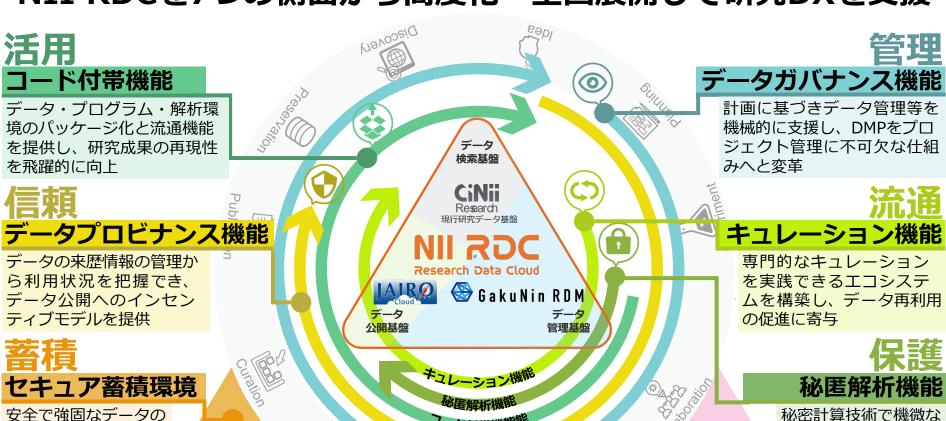
人材育成

「余計な作業」を「研究の一部」に転換するDXを実現



# 次世代研究データ基盤としてのNII RDC

### NII RDCを7つの側面から高度化・全国展開して研究DXを支援



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

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

セキュア

蓄積環境

#### 人材育成基盤

人材育成

基盤

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



# 次世代研究データ基盤としてのNII RDC

### NII RDCを7つの側面から高度化・全国展開して研究DXを支援

検索基盤

CiNii Research

Research Data Cloud

GakuNin RDM

# 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

#### データプロビナンス機能

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



#### ニュニ データガバナンス機能

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

#### には キュレーション機能

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

### 保護

#### 秘匿解析機能

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

**手**从

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

人材育成基盤

人材育成

基盤



# 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 を作成すことが必要。<u>しかしそれ</u> <u>を研究サイクルに活かすことが難</u> <u>しい</u>。
- ✓ 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<br>レベル                   | 秘匿<br>理由 |
|-----|---|--|-----------|-----|-----------------------------|----------|
| 1   | 〇〇実証におい<br>てセンサより撮<br>像したデータ及<br>び関連データ | 〇〇実証におい<br>てセンサより撮<br>像したデータで<br>あり、道路の画<br>像データ | 〇〇<br>研究所 | 上 _ | レベル4<br>(広範な提<br>供・利活<br>用) |          |



提出

それで終わり



#### 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。

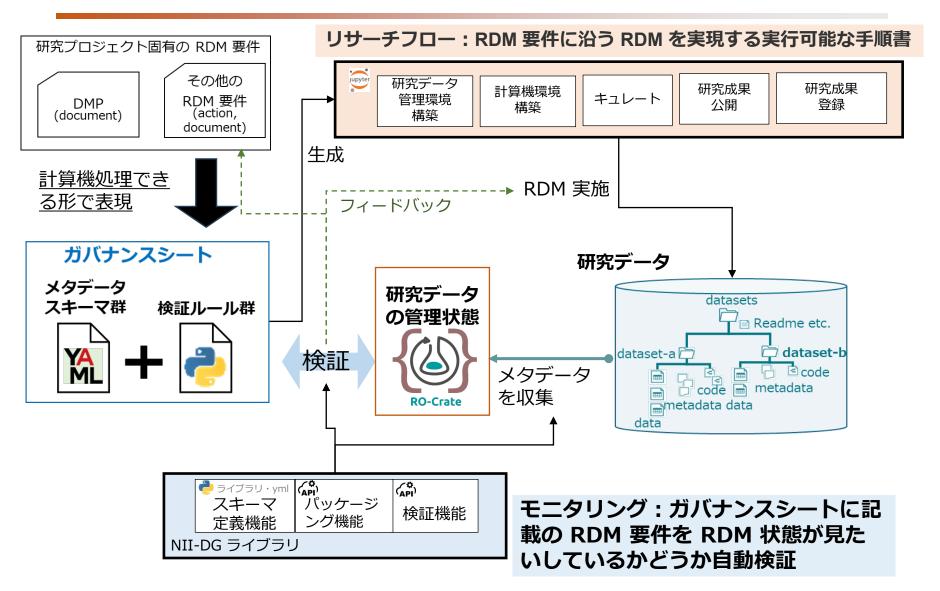


研究成果公開支援

モニタリングで検証



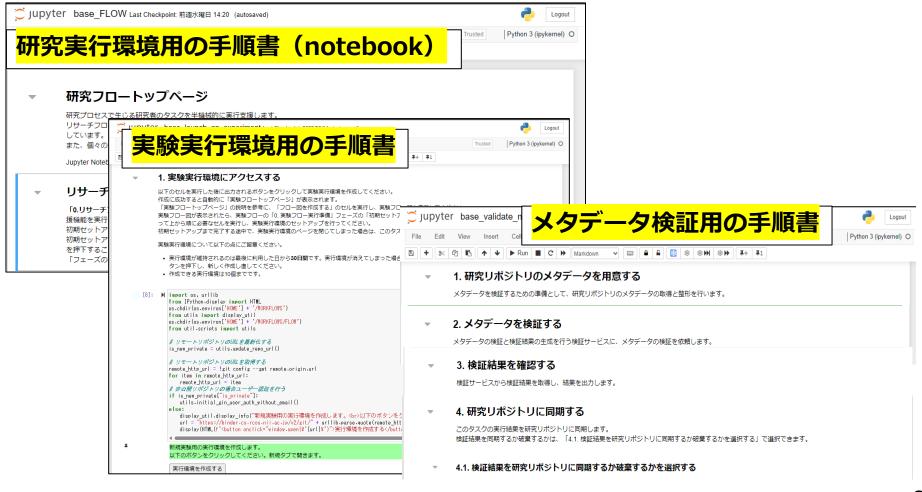
# 現行のデータガバナンス機能





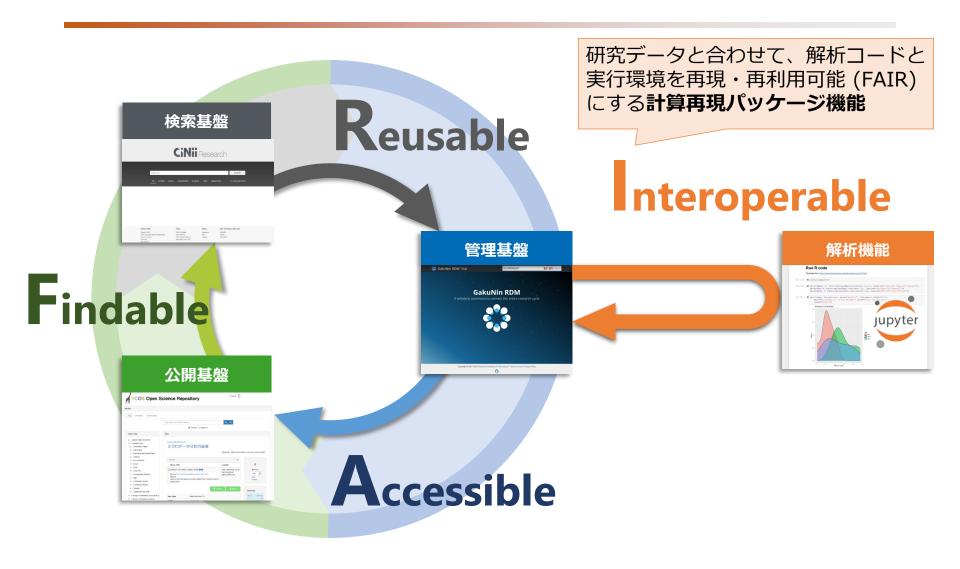
# 機能評価試験版サービス

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。





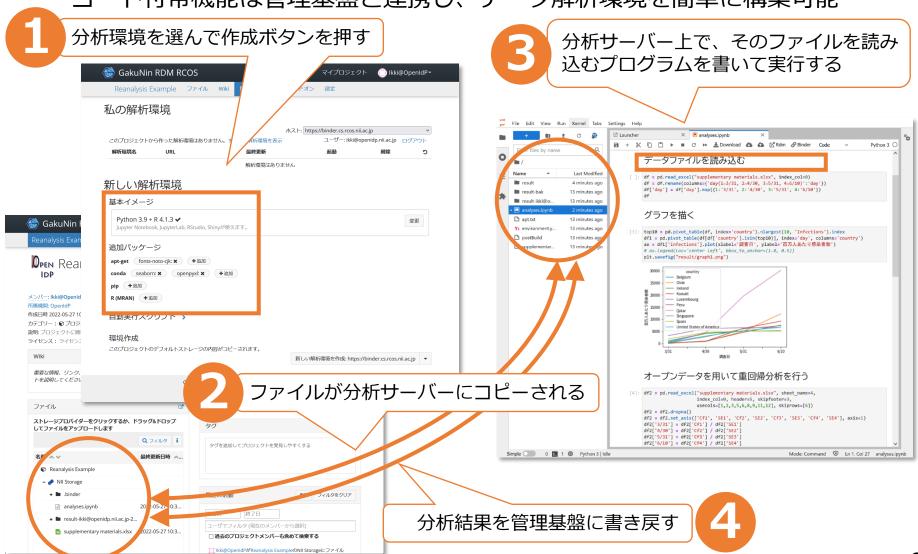
### コード付帯機能の概要





# 管理基盤のデータ解析機能として

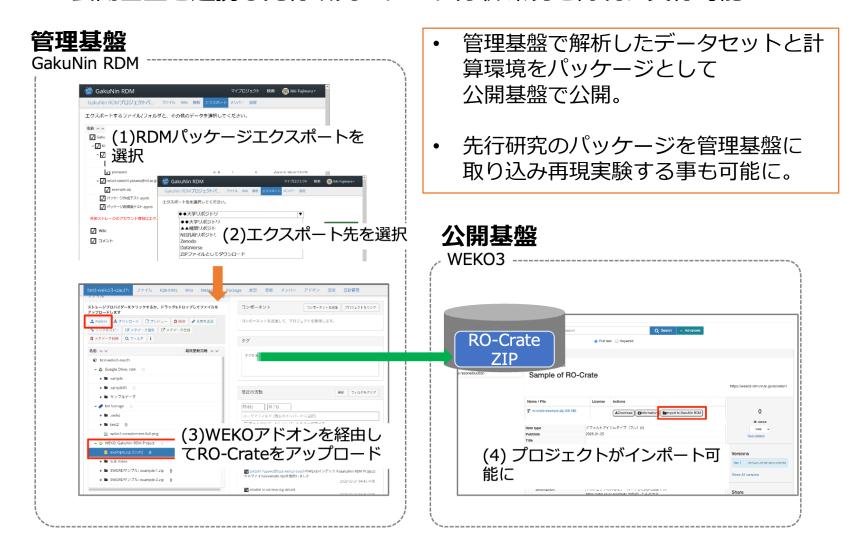
コード付帯機能は管理基盤と連携し、データ解析環境を簡単に構築可能





# 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





# 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

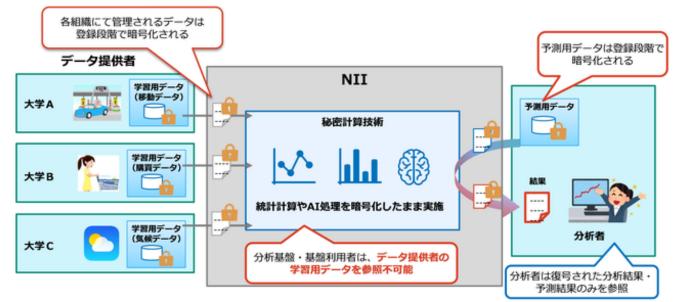
#### 狙い

- 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。



#### 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

#### ●教材構成

マイクロコンテンツ教材(合成音声)

+確認テスト

#### プラットフォーム連携チーム



### 理化学研究所

- 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の設計と実証
- 関連する高度化機能との仕様調整と共同開



#### 研究データ基盤高度化チーム

#### ルール・ガイドライン整備チーム



- 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の設計
- 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まとめ
- 学内整備のための事例形成

#### NII RDCと分野PFの連携



分 野

データプロビナンス機能

コード付帯機能

セキュア蓄積環境

NII RDC利用のプラクティス

# 検索基盤 CINH Research

データ管理公開の組織的役割

キュレーション機能

データガバナンス格

秘匿解析機能

組

怕

人材育成基盤

データ管理公開の組織的支援

### 融合・活用開拓チーム



東京大学 THE UNIVERSITY OF TOKYO

- 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



ボトムアップ

#### 人材育成チーム



- 人材育成を主とした研究データ管理体制の構 築を推し進め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



# 研究データエコシステム構築と機関連携



### 研究データエコシステム東海コンソーシアム (2023/10月~)

### 【目的】 東海地域の学術機関が

連携し、研究データエコシステムの整備、普及、利用促進

### 【会員】

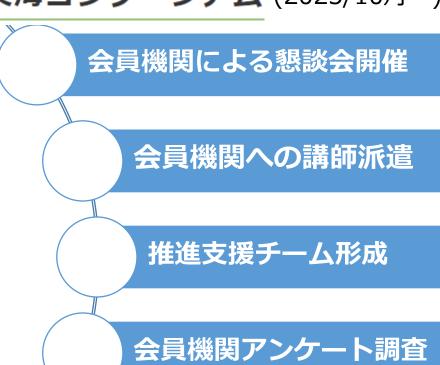
- ・正会員:大学,高専,研究所等の 学術機関(東海地域に限定しない)
- ・準会員:学術機関に属する部署

### 【会費】 不要

(研究データエコシステム構築事業 の名古屋大学への配分予算で運営)

### 【活動期間】

• 2023年度~2026年度



北陸地域との合同企画



# 採択21件中医学医療系ユースケース

### 1. 喜屋武亨、琉球大学、医学部保健学科

擬似人流データを用いた身体活動量の推定と地理的・社会経済的 環境から見た地理的地域特性の解明:地域住民の健康増進に向け た活動量シュミレーションシステムの開発

2. 吉岡京子、東京大学、大学院医学系研究科健康科学・看 護学専攻地域看護学分野

乳児の股関節脱臼の見落としゼロを目指す異常判別AIとコミュニ ティスクリーニングシステムの開発

- 3. 木村映善、愛媛大学、医学部 国際的なRWD 研究を実現する医療情報分析基盤の検討
- 林美加子、大阪大学、歯学部附属病院 健康医療データの質を保証する安全安心な医療AIサービスプラッ トフォームの構築
- 岡田知久、京都大学、大学院医学研究科 マルチスピーシーズ霊長類脳画像の知識型データ共有エコシステ ムの構築:データ駆動型脳科学を推進する共創型国際研究基盤



### 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 ユースケース公募受付中!

### 背景・目的

本事業では、研究データ基盤を活用し、異なる分野間での データ連携を前提としたAI・データ駆動型研究のシーズ・ ユースケースの創出を目指しています。

### 対象者

全国の大学、高等専門学校、大学共同利用機関法人、 独立行政法人の研究者

### 【2024年 公募スケジュール】

| 申請期間           | ~2/15 | 2/16~6/15 | 6/16~8/15 | 8/16~10/15 | 10/16~12/15 |
|----------------|-------|-----------|-----------|------------|-------------|
| 審査結果通知<br>(予定) | 3 月末  | 7 月末      | 9月末       | 11月末       | 1月末         |
| 予算執行可能日        | 4/1   |           | 審査結果通知日   |            | 2025/4/1    |

詳細はこちら





# RCOS rdm\_support@nii.ac.jp