

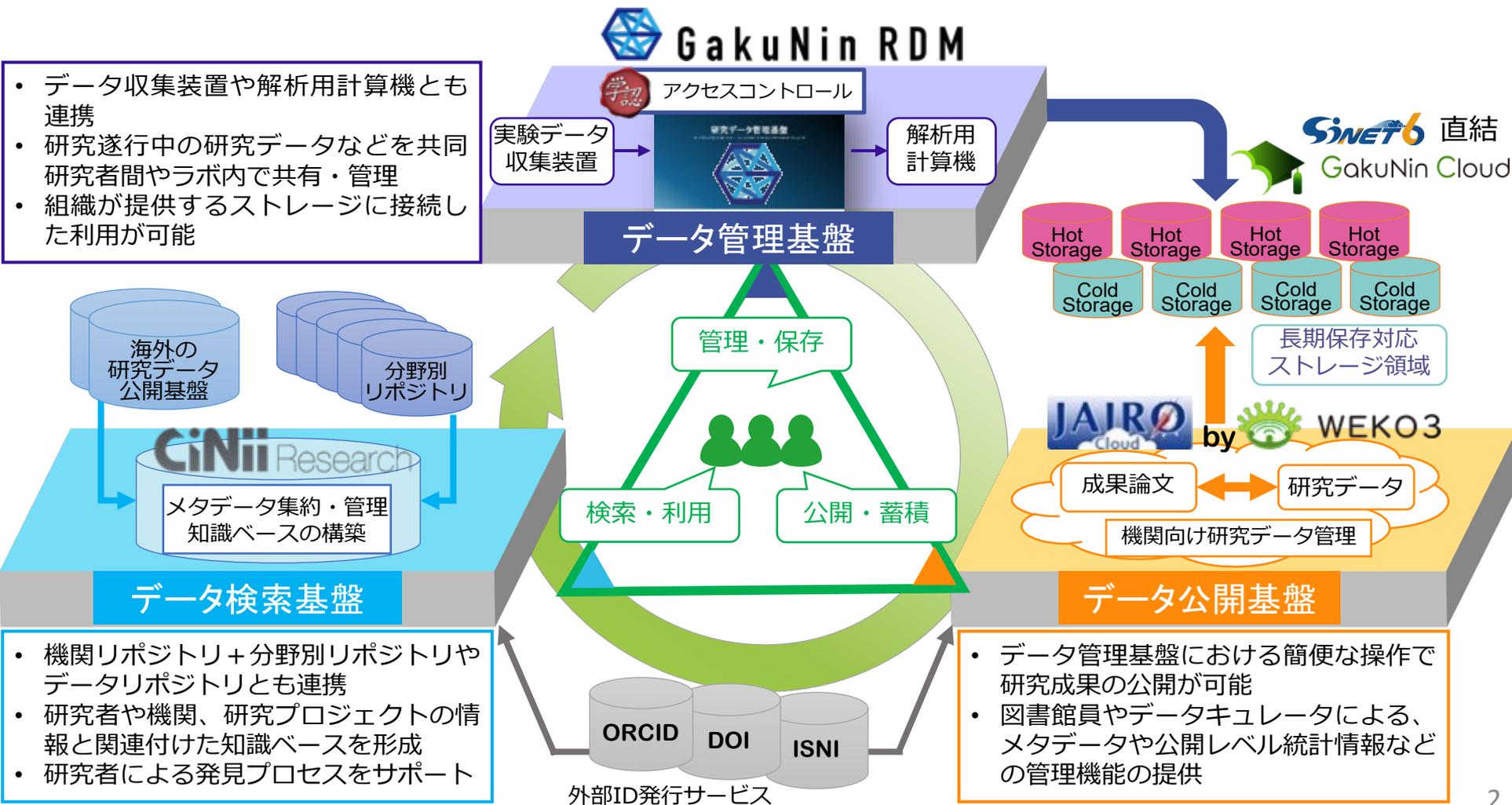
# 研究データ管理基盤GakuNin RDM の現状とユースケース

下山 武司

国立情報学研究所 オープンサイエンス基盤研究センター

# 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



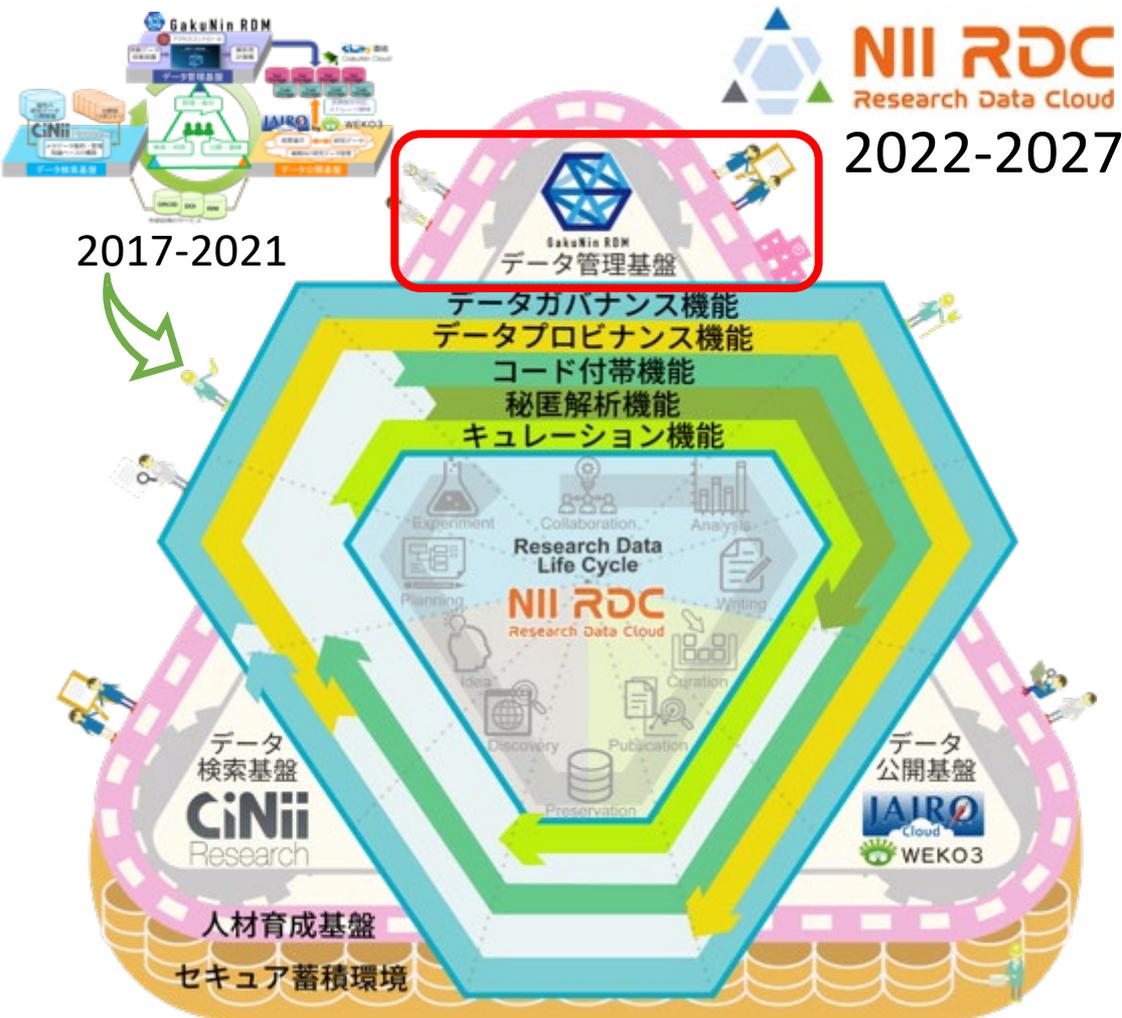
- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能

- 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

# 学術研究データプラットフォーム

既存の3基盤を有機的に繋ぐ先端機能を実現しオープンサイエンスの実践に不可欠な人材育成の仕組みを提供



- **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

- **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

- **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

- **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

- **キュレーション機能**

専門的なデータ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

- **セキュア蓄積環境**

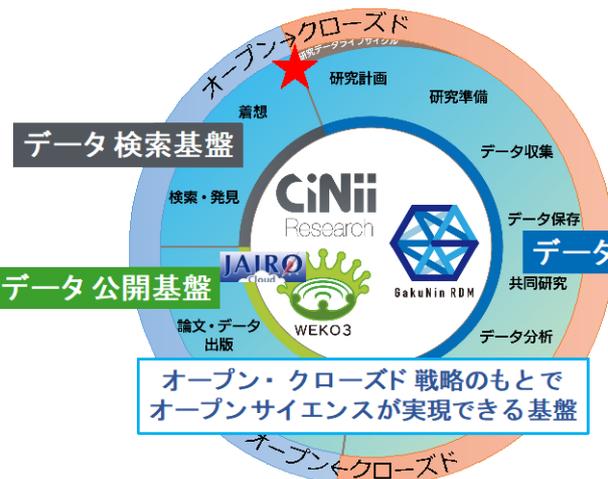
専用HWと高度な暗号化技術による超鉄壁ストレージを提供し、データの共有と保護の両立を実現

- **人材育成基盤**

研究データ管理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

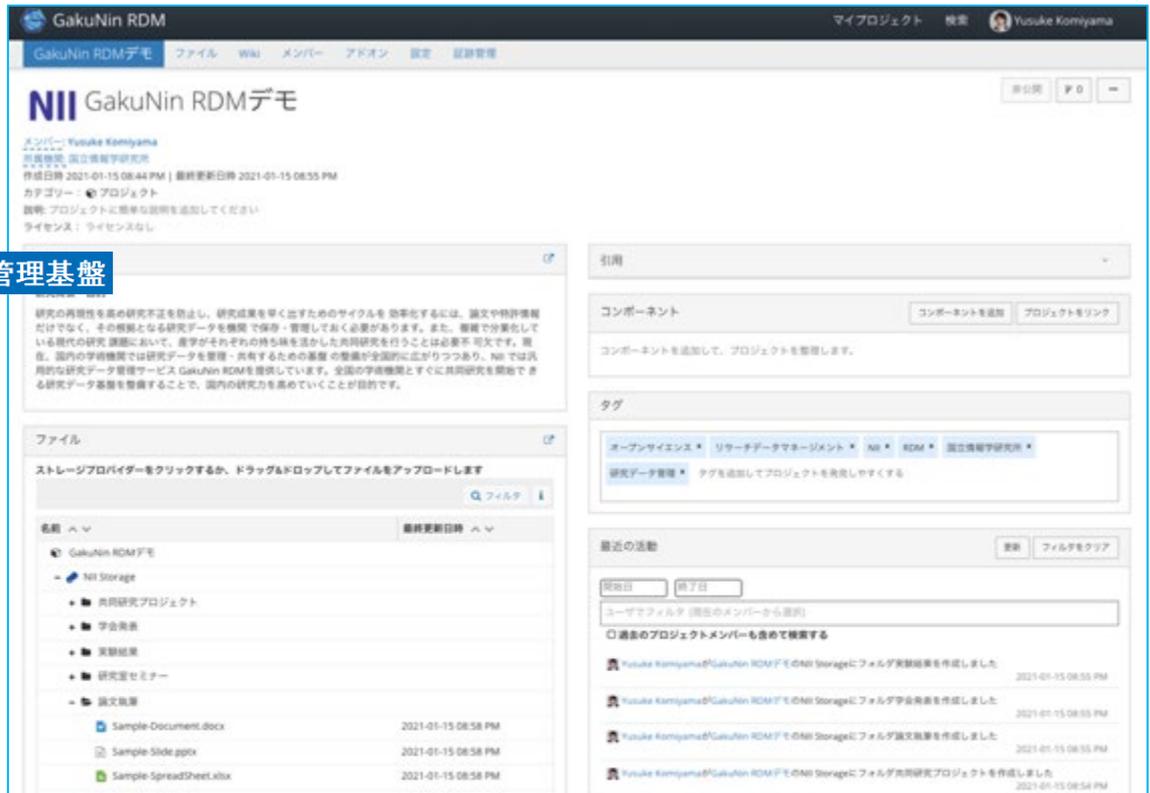
# データ管理基盤 GakuNin RDM

# 研究データ管理基盤 GakuNin RDM



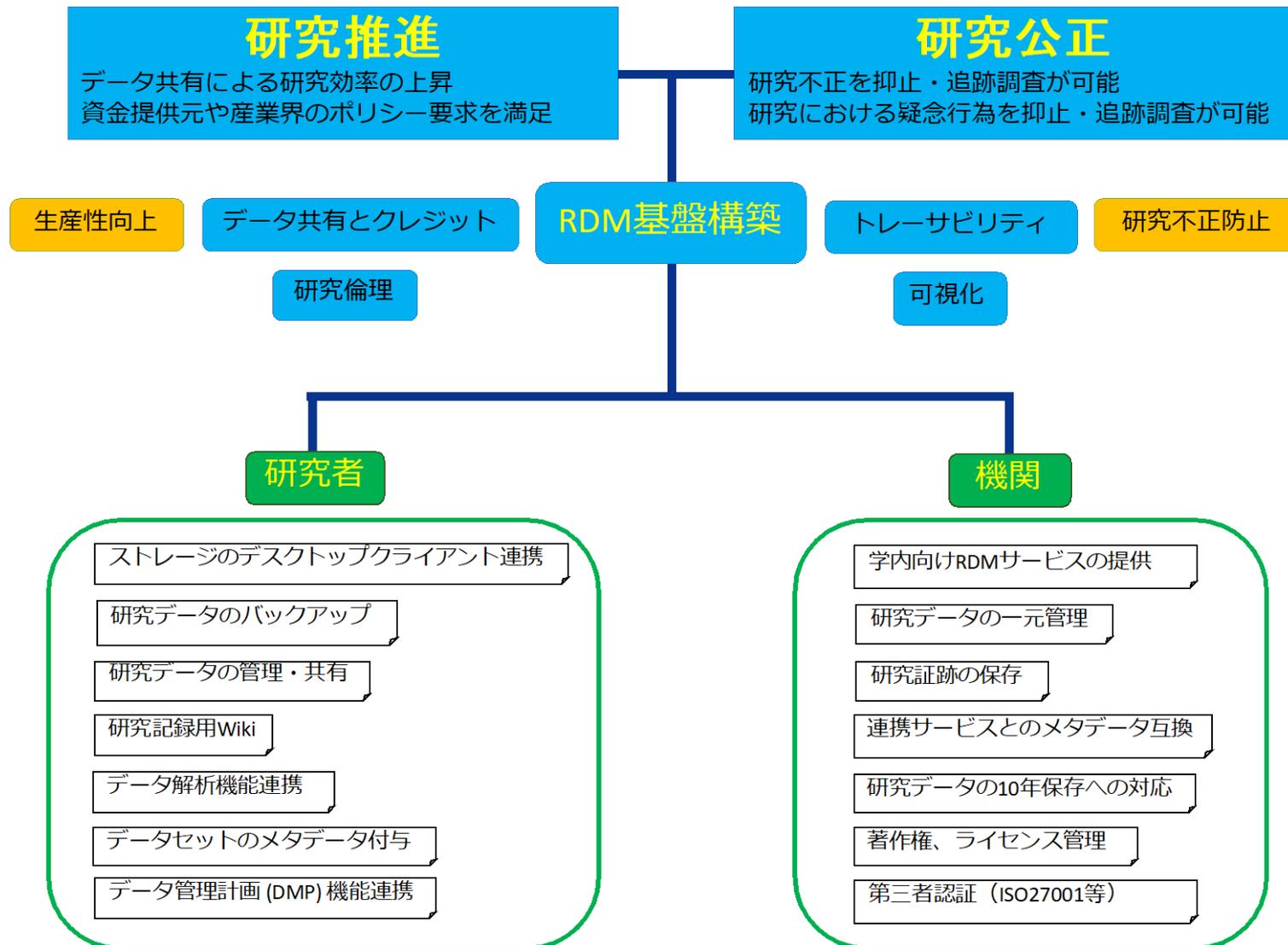
オープン・クローズド 戦略のもとで  
オープンサイエンスが実現できる基盤

研究データプロセスと  
研究データ基盤  
NII Research Cloud

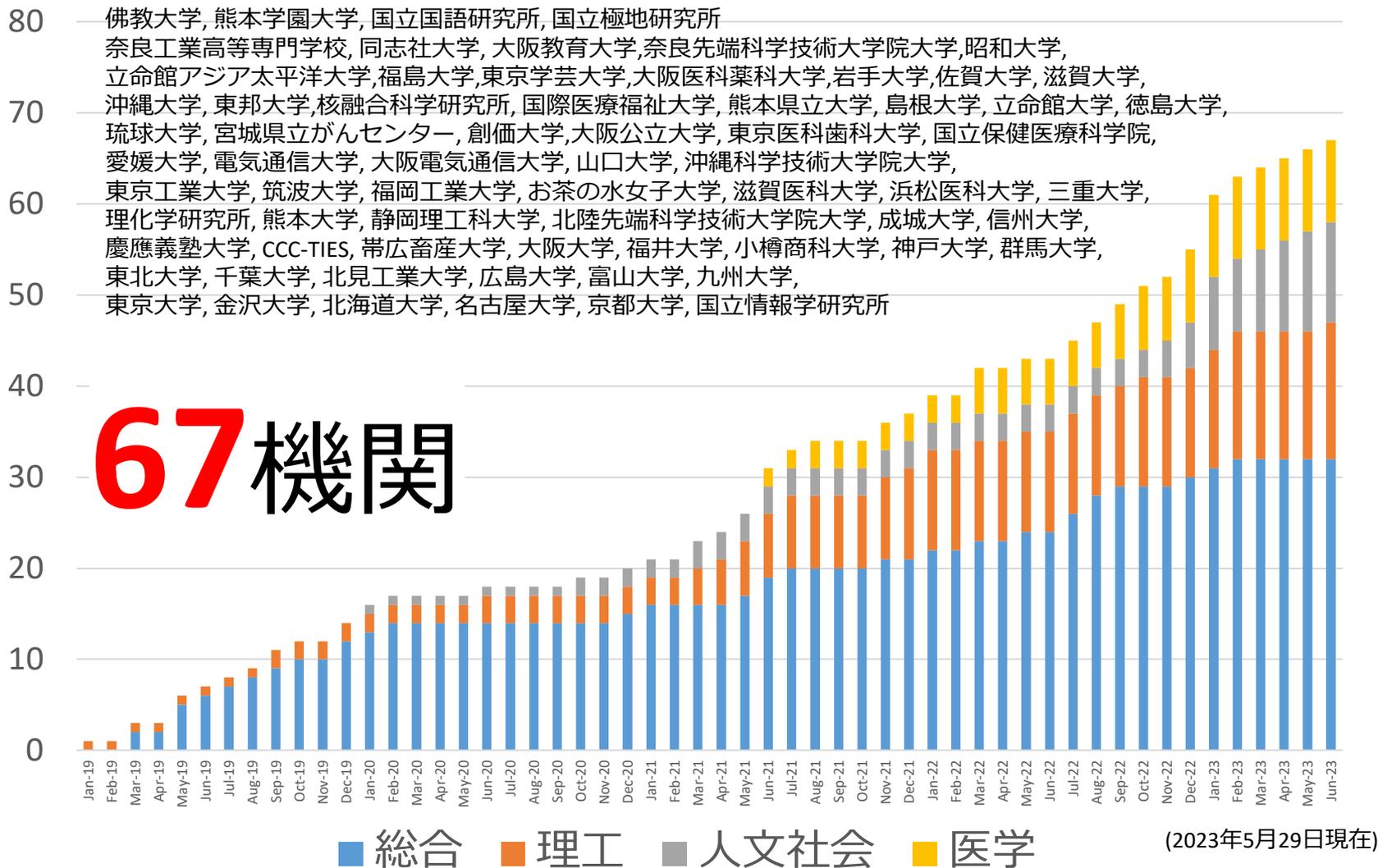


GakuNin RDM (GRDM) は、国立情報学研究所 (NII) で運用している研究データ基盤NII Research Cloudの一部で、研究中のクローズド (非公開、制限共有) なデータを取り扱う事に特化したRDMサービス。2021年4月からは24時間365日オペレーターを配置した運用体制で全国の学術機関にRDMサービスを提供。

# GakuNin RDMのサービス・ビジョン

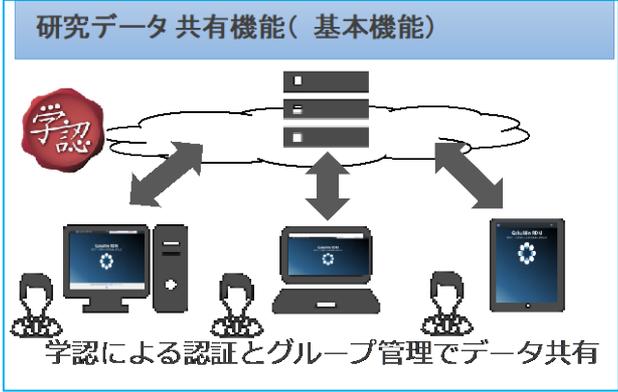


# GakuNin RDMの利用機関数の増加



— 佛教大学, 熊本学園大学, 国立国語研究所, 国立極地研究所  
 奈良工業高等専門学校, 同志社大学, 大阪教育大学, 奈良先端科学技術大学院大学, 昭和大学,  
 立命館アジア太平洋大学, 福島大学, 東京学芸大学, 大阪医科薬科大学, 岩手大学, 佐賀大学, 滋賀大学,  
 沖縄大学, 東邦大学, 核融合科学研究所, 国際医療福祉大学, 熊本県立大学, 島根大学, 立命館大学, 徳島大学,  
 琉球大学, 宮城県立がんセンター, 創価大学, 大阪公立大学, 東京医科歯科大学, 国立保健医療科学院,  
 愛媛大学, 電気通信大学, 大阪電気通信大学, 山口大学, 沖縄科学技術大学院大学,  
 東京工業大学, 筑波大学, 福岡工業大学, お茶の水女子大学, 滋賀医科大学, 浜松医科大学, 三重大学,  
 理化学研究所, 熊本大学, 静岡理工科大学, 北陸先端科学技術大学院大学, 成城大学, 信州大学,  
 慶應義塾大学, CCC-TIES, 帯広畜産大学, 大阪大学, 福井大学, 小樽商科大学, 神戸大学, 群馬大学,  
 東北大学, 千葉大学, 北見工業大学, 広島大学, 富山大学, 九州大学,  
 東京大学, 金沢大学, 北海道大学, 名古屋大学, 京都大学, 国立情報学研究所

# 学認フェデレーション参加のIdPと連携可能 調達が必要でシステム導入が容易



GakuNin RDMポータル  
プルダウンメニューから自機関を選択

GakuNin RDM  
研究データ管理による研究推進と研究公正

自機関の認証システム  
(IdP)でログイン

シングルサインオンで  
GakuNin RDMを利用

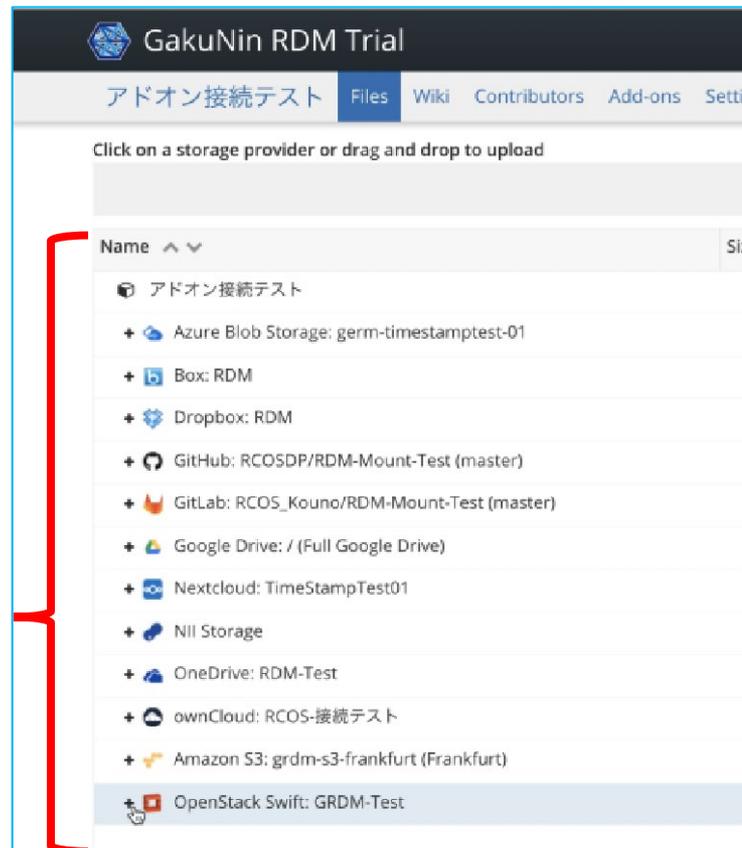
GRDMでは、学術認証（学認）フェデレーションにサービスプロバイダ（SP）として登録済みのため、学認に参加していればアイデンティティプロバイダ（IdP）連携のみで導入が可能

# GakuNin RDMとクラウドストレージや外部ツールとの連携で研究を促進

## NII研究データ 基盤と外部ツールとの連携



研究プロジェクトに様々な外部クラウドサービスを紐付けて管理



GRDMでは、NII提供分のストレージ以外にも、機関側で契約、所有しているパブリッククラウド、プライベートクラウドのストレージをAPIで接続して利用可能。

# 研究データ管理機能 プロジェクト画面 (1)

## 研究プロジェクトでのデータ共有・管理

## プレビュー機能

## プロジェクト用Wiki

## マルチクラウドファイルマネージャー (構成例)

-  Google Drive
-  Nextcloud
-  NII Storage
-  One Drive

Name	Size	Version	Download...	Modified
Demo project				
Google Drive: OSF_hoge				
RDM_Demo1				
RDM_Demo2				
RDM_Demo3				
Nextcloud: Documents				
About.odt	77.4 kB			2019-01-29 03:11 PM
About.txt	1.1 kB			2019-01-29 03:11 PM
Nextcloud Flyer.pdf	2.5 MB			2019-01-29 03:11 PM
User Data Manifesto.pdf	36.9 kB			2019-01-29 03:11 PM
NII Storage				
demo				
DEMO_CSV_FILES.csv	3.8 kB	2	0	2018-12-14 05:34 PM
RCOS_leafletA3_2018.pdf	803.3 kB	1	0	2018-12-14 05:34 PM
RCOS/パンフレット英文原稿_Paquette氏_提案原稿v2_GRDM.docx	22.4 kB	1	0	2018-12-14 05:34 PM
OneDrive: RDM_Sync				

## バージョン管理

## ディスカッション機能



# 今後拡張予定の機能

## ・機関管理者向けツール

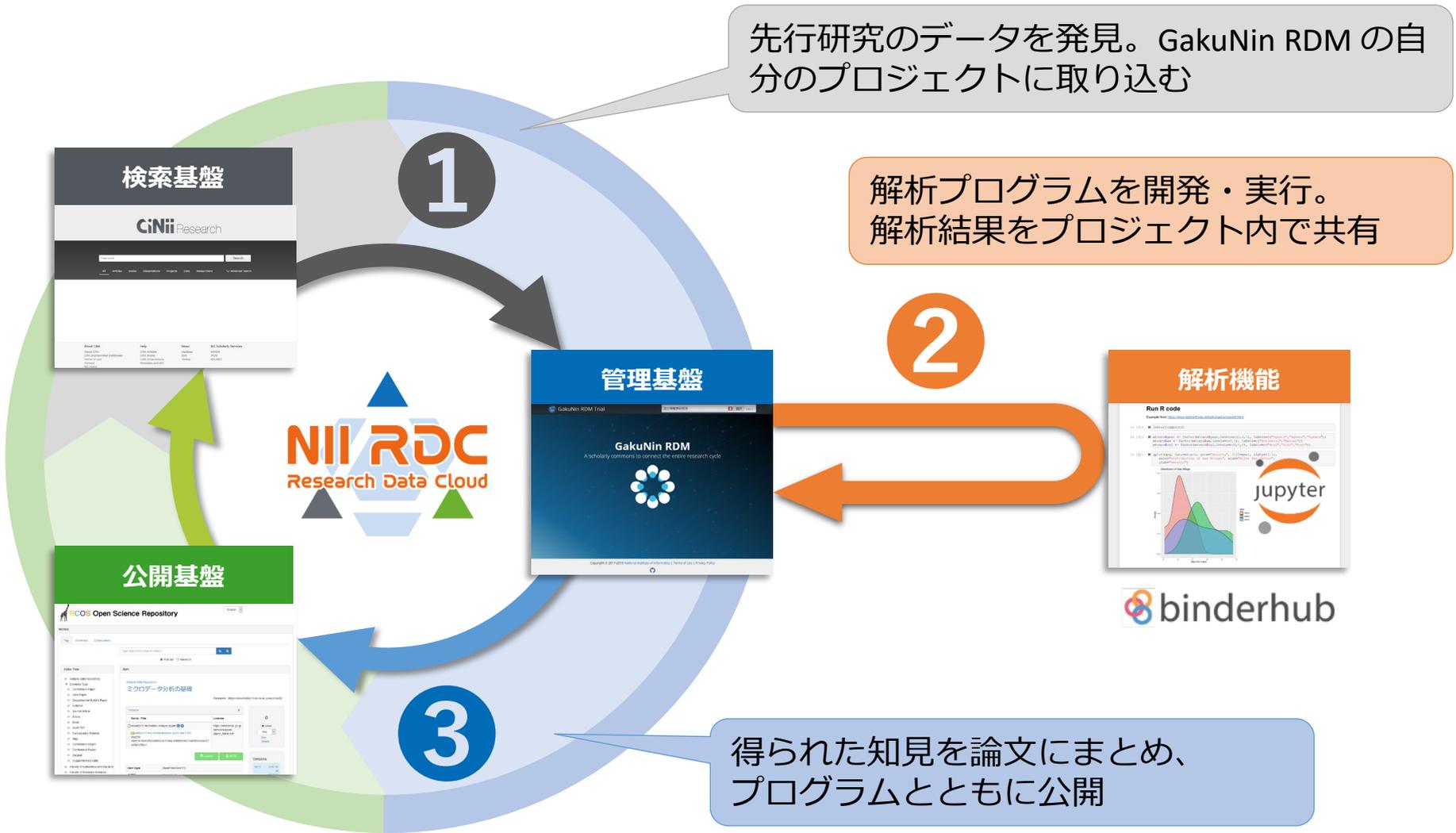
### ・機関ストレージ移行のための機能

- ・ 機関に属するプロジェクトに保存されている機関ストレージ内のデータを、新しい機関ストレージに移行するための機能
- ・ NII標準ストレージから新規に構築される機関ストレージ(※)への移行を想定
- ・ 3種類の機能(エクスポート機能, リストア機能, 検査機能)からなる(※)現時点ではS3, S3互換ストレージのみに対応, 今後拡張予定

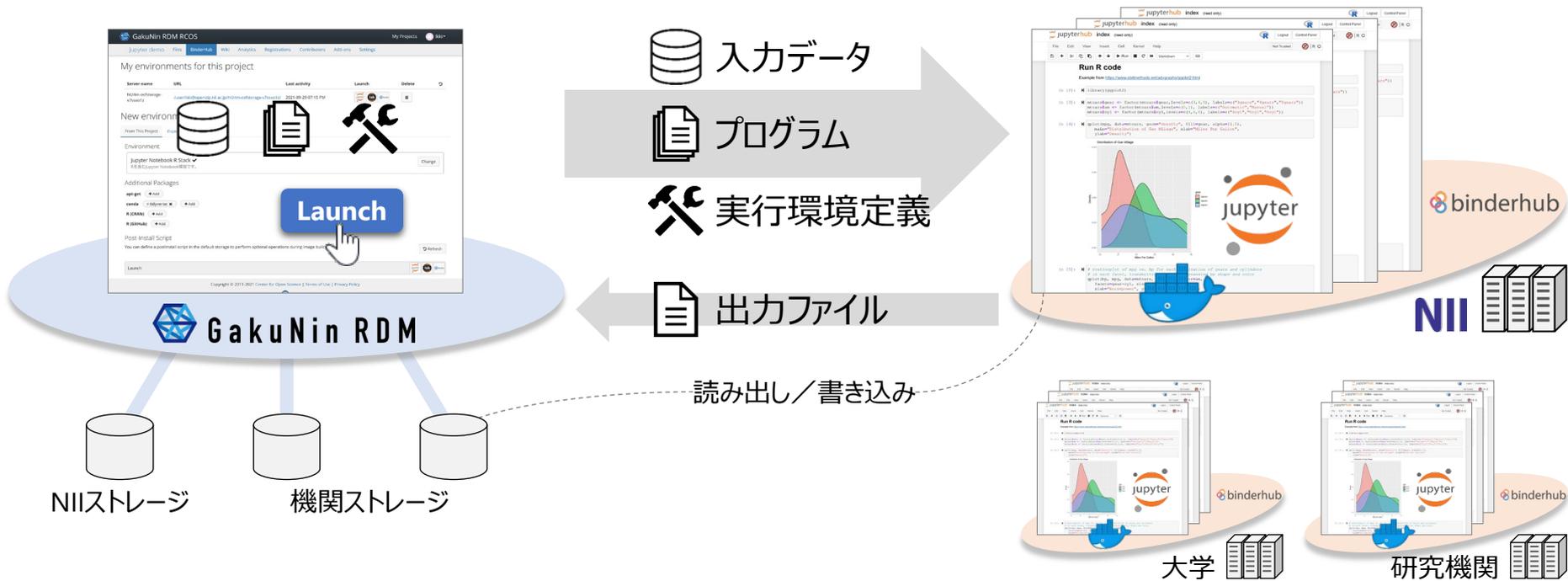


# コード付帯機能

# データとコードが循環する世界



# コード付帯機能の一部: GakuNin RDM データ解析機能



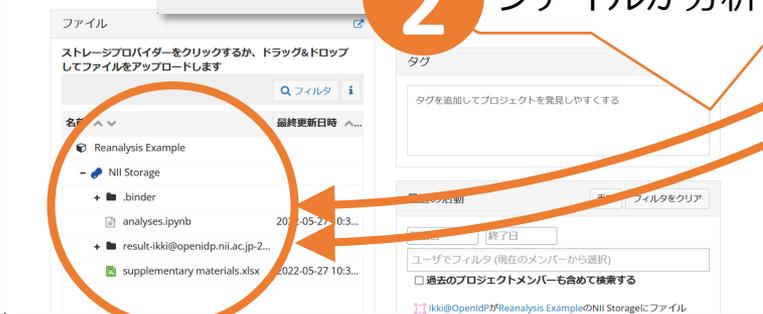
- JupyterHub がインストールされた計算機と連携し、データ解析環境をGakuNin RDMから1クリックで構築
- NII所有の計算機のほか、クラウド上のVMなど外部計算機とも連携可能

# 基本的な使い方

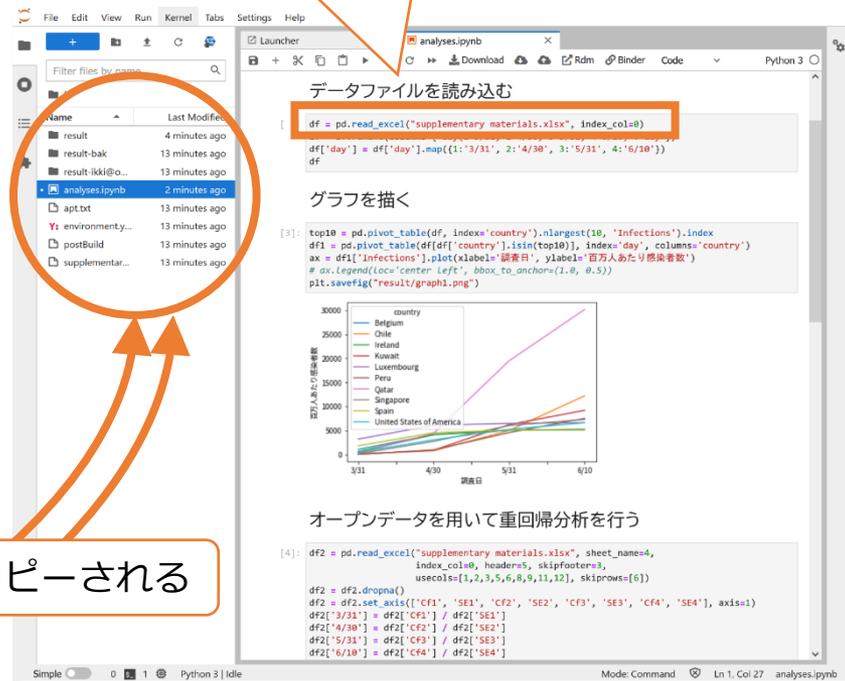
1 分析環境を選んで作成ボタンを押す



2 ファイルが分析サーバーにコピーされる



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



分析結果を管理基盤に書き戻す

4

# 応用例1: 社会学データの簡易分析アプリ

1 リポジトリで起動リンクをクリック

2. データの利用申請には、JAIRO Cloudへのユーザー登録が必要です。JGSSDDSアカウント

JGSSオンライン分析アプリケーション  
https://jgssdds.repo.nii.ac.jp/records/2000166

名前 / ファイル	ライセンス	アクション
JGSSオンライン分析アプリケーション (ペータ版) 初回起動リンク		Information

Item type: JDCat(1)  
公開日: 2022-05-23  
タイトル: JGSSオンライン分析アプリケーション  
言語: ja  
作成者: JGSS研究センター

概要: [JGSSオンライン分析アプリケーションの実行の手順]  
1. JGSSオンライン分析アプリケーションへのアクセス用リンク (https://binder.cs.rcos.nii.ac.jp/v2/gh/JGSSDDS/JGSS\_Analytics\_Application/HEAD) をクリックし、しばらく待ったJDCat分析ツールのログイン画面 (https://jupyter.cs.rcos.nii.ac.jp/) に遷移します。  
2. Openidp (https://openidp.nii.ac.jp/) のアカウントを使ってログインします。  
• ログイン画面 (https://jupyter.cs.rcos.nii.ac.jp/) の所属機関欄でOpenidpを選び、取得済みのアカウントでログインします。  
• Openidpアカウントを取得していない方は、「ac.jp」所属の大学・機関のドメインのメールアドレスでアカウントを作成してください。詳しくは、JDCat分析ツールのウェブサイト

2 NIIのサーバー上で簡易分析アプリが起動

JGSS オンラインデータ分析ツール

1. 変数の選択  
2. 分析の選択  
3. 演算結果

1. 変数の選択  
読み込んだデータセット:  
2)JGSS-20172018\_v1.0\_sav  
分析対象の変数を選択  
年齢 (ageb)

2. 分析の選択  
■ 変数の概要  
ヒストグラムのbin数を指定  
1 6 20  
■ 変数の代表値  
平均値  
中央値  
最大値  
最小値  
不偏分散  
標本分散  
不偏標準偏差

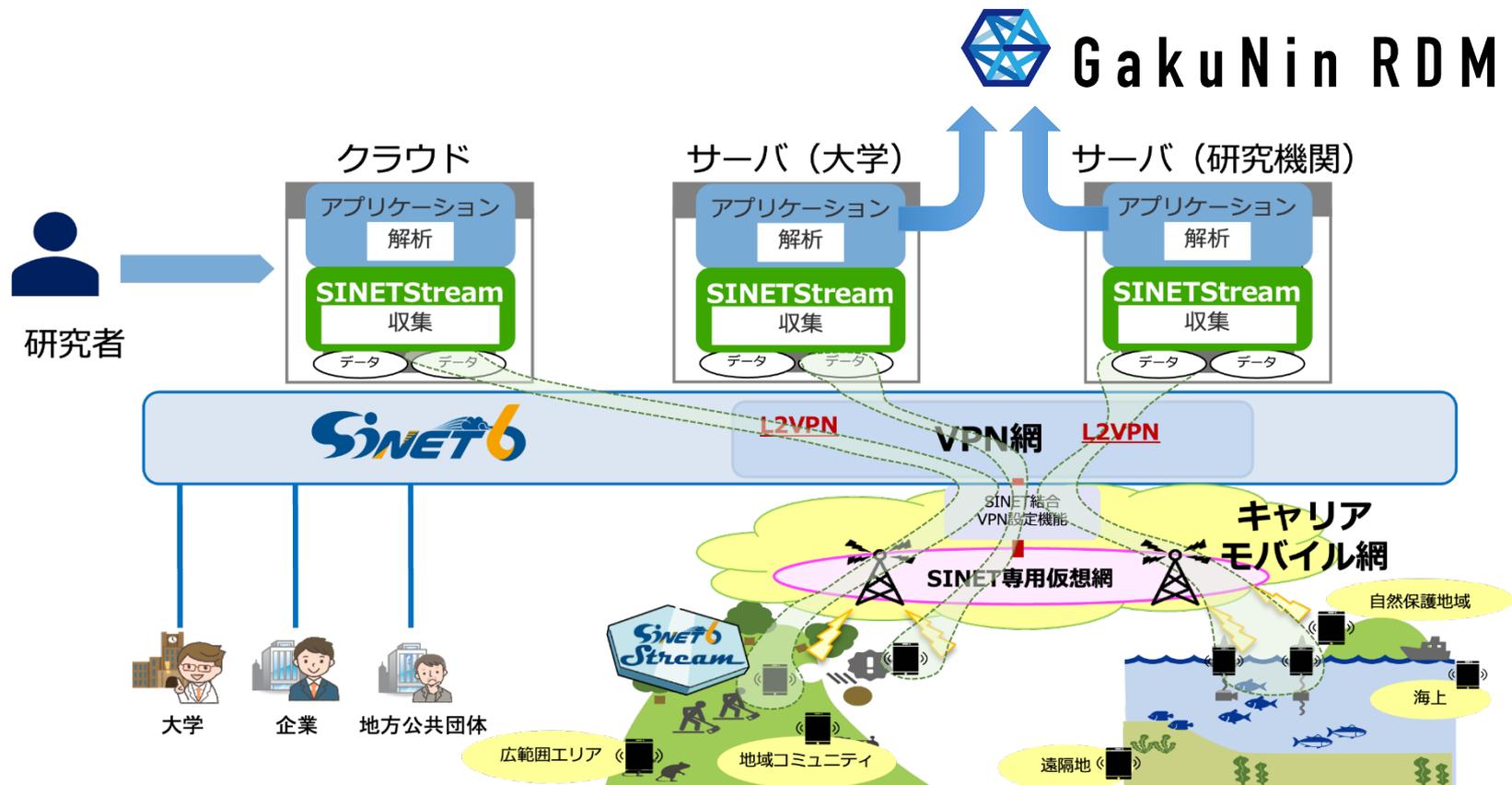
3. 演算結果  
ヒストグラム ( ageb )  
Frequency  
500  
400  
300  
200  
100

3 プログラムを書かずにデータの概要を把握

<https://jgssdds.repo.nii.ac.jp/records/2000721>

# 応用例2: IoTデータのリアルタイム分析

- 広域データ収集基盤 SINETStream と連携
- IoT機器などから流れてくるデータをリアルタイムに分析し、GakuNin RDM で共有



# 秘匿解析機能

# 秘匿解析機能の開発の狙い

---

- 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
- 進捗状況
  - 秘匿解析機能(v2022)を所内に構築し、内部で試験中
  - 一部をトライアルサービスとして大学向けに限定的に提供
- 各種の検討を行った上、2027年までにRDCのサービスとして提供予定

# 秘匿解析機能の実現の課題とアプローチ

- 秘匿解析機能の実現に向け、大きく5つの課題に対して検証中

## 課題

- ① 高度な秘匿解析機能の実現
- ② 暗号を意識させない利便性
- ③ 利用ルールの整備
- ④ 設計・開発・運用技術の検証
- ⑤ ユースケースの開拓

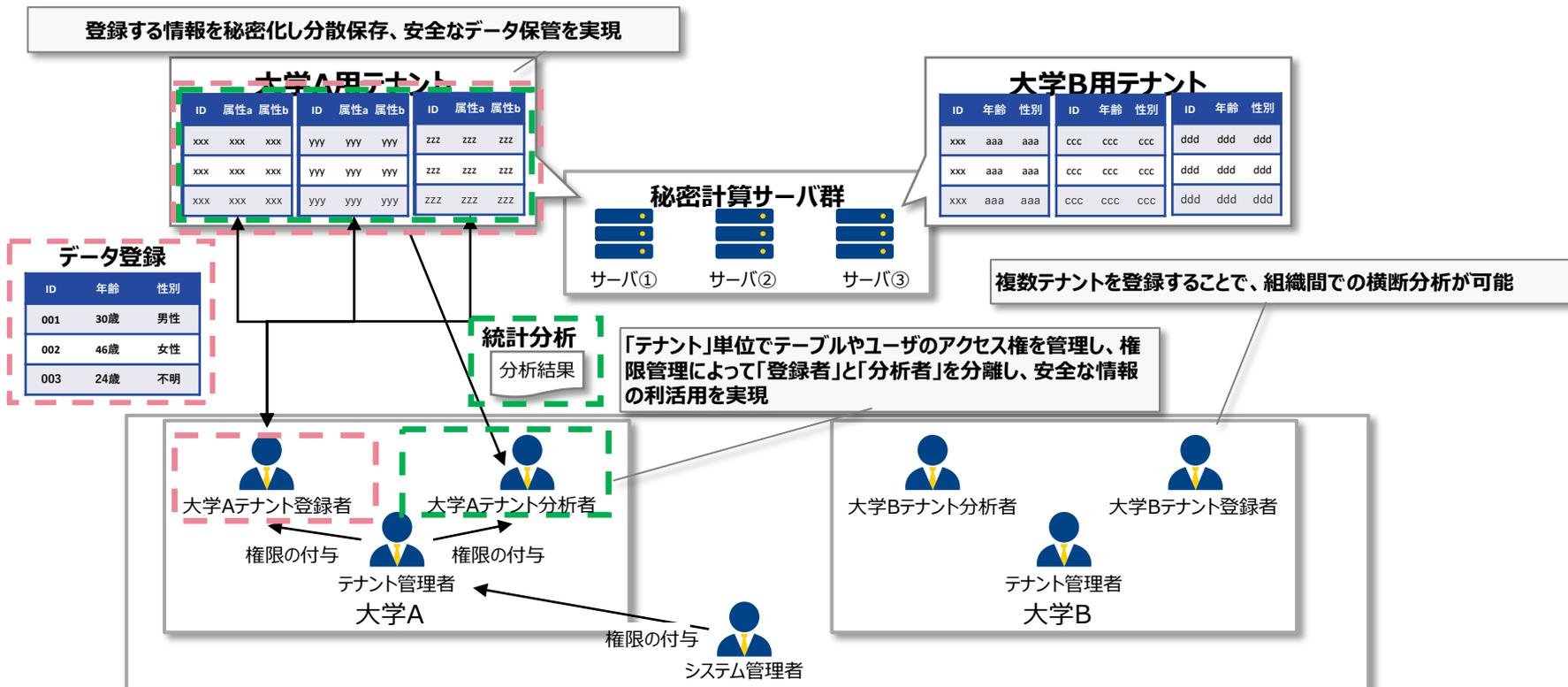


## アプローチ（現状）

- 外部から秘密計算技術を導入し評価・検証
- データ解析の標準的な環境へ対応（jupyterからの利用）
- 「データハンドブック」で個人情報保護との関係を解説
- NTTとの共同トライアルを実施中

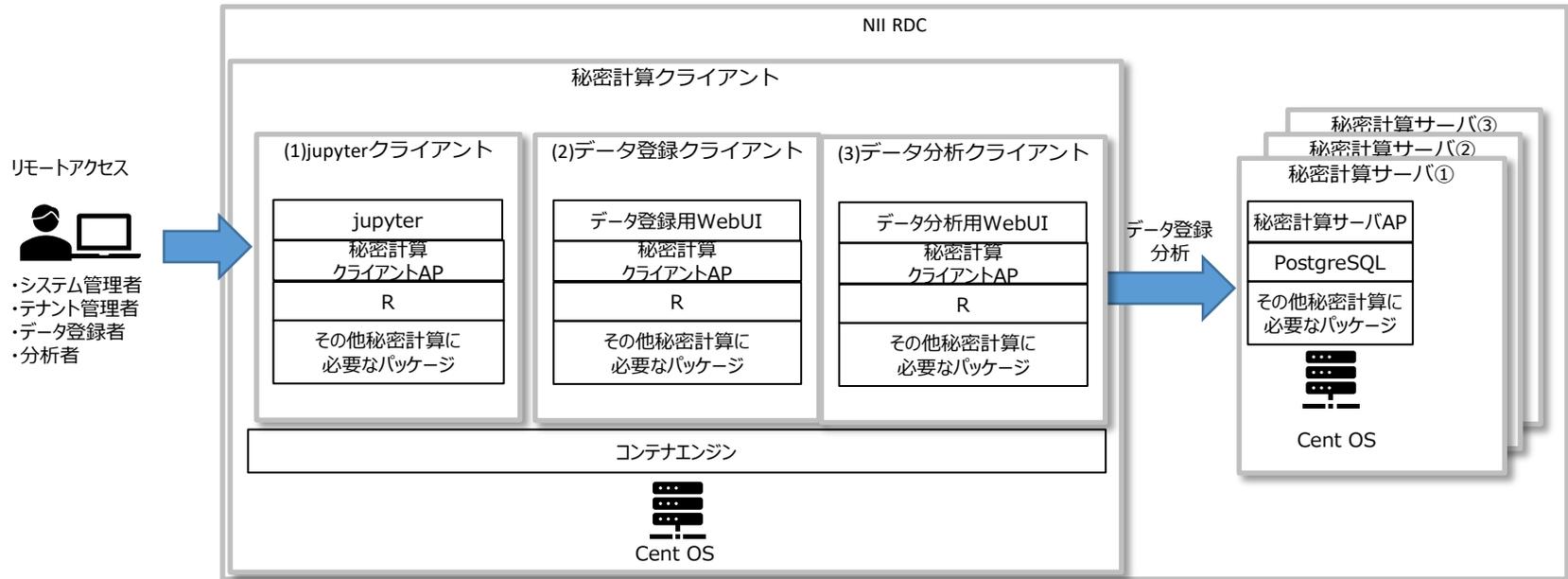
# ① 高度な秘匿解析機能の実現に向けて

- NTTが開発した秘密計算技術を試験導入し、NII RDC内に環境構築
  - 秘密分散ベースのマルチパーティ計算
  - 統計機能（総和、平均、分散、最大値、中央値、クロス集計、ロジスティック回帰等）、データ管理機能、データ操作機能等、一般的なデータベース相当の機能を具備
- 現在、内部でPoCを実施し、複数大学間データ共有分析の課題等を分析中



## ②暗号を意識させない利便性の実現に向けて

- データ解析の標準環境対応機能を開発
  - jupyter および Web インターフェイスを開発



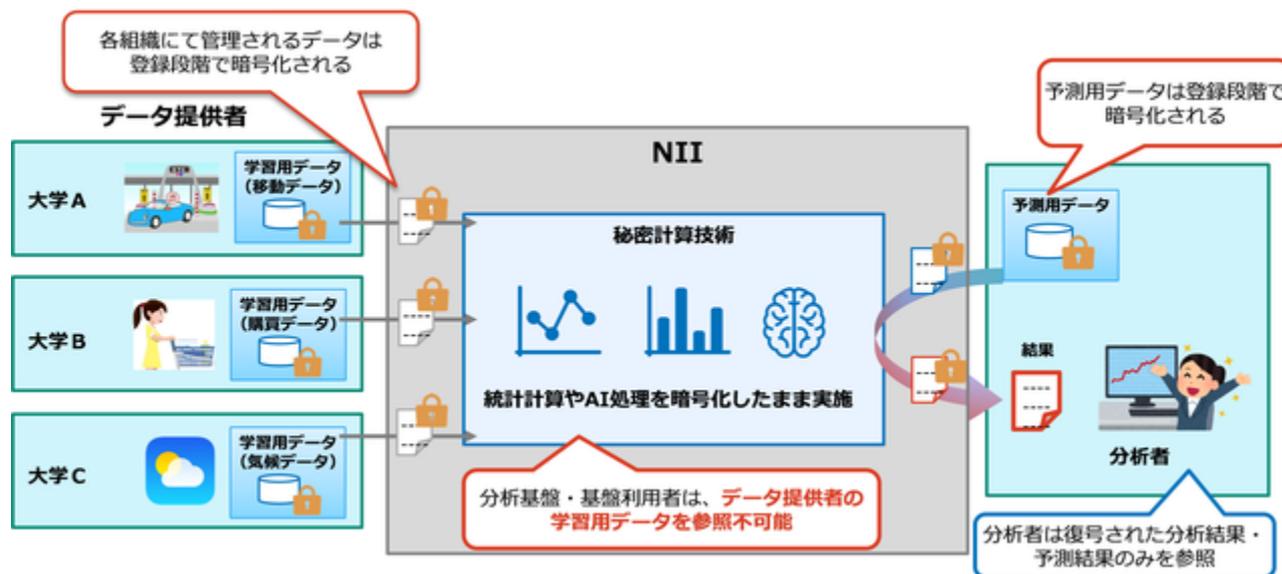
## ③利用ルールの整備に向けた取り組み

- NIIが「データハンドブック」を公表
  - オープンサイエンスのための データ管理基盤ハンドブック  
～学術研究者のための“個人情報”の取扱い方について～（第1版）
    - 2022年7月27日 国立情報学研究所  
オープンサイエンスのためのデータ管理基盤 ハンドブックにかかる検討会 編
    - <https://www.nii.ac.jp/service/handbook/>
  - データの取扱いにおいて配慮すべき事項を個人情報を軸に整理
- 秘匿解析機能に関する記述 (P.91)
  - データを暗号化した場合、データの漏えいが発生しても被害を最小限に抑えることができるため、安全管理措置としてとても意味があります。
  - また、漏えい等が発生し、又は発生したおそれがある個人データについて、高度な暗号化等の秘匿化がされている場合等、「高度な暗号化その他の個人の権利利益を保護するために必要な措置」が講じられている場合については、報告及び本人通知を要しないとされています。
  - さらに、暗号化はデータの消去にも役立ちます。データの消去は、物理的消去、電磁的消去によって行われてきましたが、これに加えて暗号化データの鍵を消去する論理的消去もデータ消去として認める動きがあります。
  - NII Research Data Cloudでは、このような考えのもとに、データの暗号化機能や秘匿解析機能をサポートしていく予定です。

## ④設計・開発・運用技術の検証

## ⑤ユースケースの開拓に向けて

- NIIとNTT、秘密計算システムの大学向けトライアルを開始  
 ～世界初の「AI 4大カテゴリの主要なアルゴリズムによる学習・推論  
 が可能な秘密計算AIソフトウェア」を提供～ 2023/01/23
  - <https://www.nii.ac.jp/news/release/2023/0123.html>
- 現在、下記ページで申込受付中
  - <https://rcos.nii.ac.jp/news/2023/01/20230123-0/>



# RCOS