

GakuNin RDMを用いた 研究データ管理のはじめ方

GakuNin RDM担当
国立情報学研究所

於

大阪会場 2023/10/17

福岡会場 2023/11/2

東京会場 2023/11/21

第6期科学技術・イノベーション基本計画 (2021年)

「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。

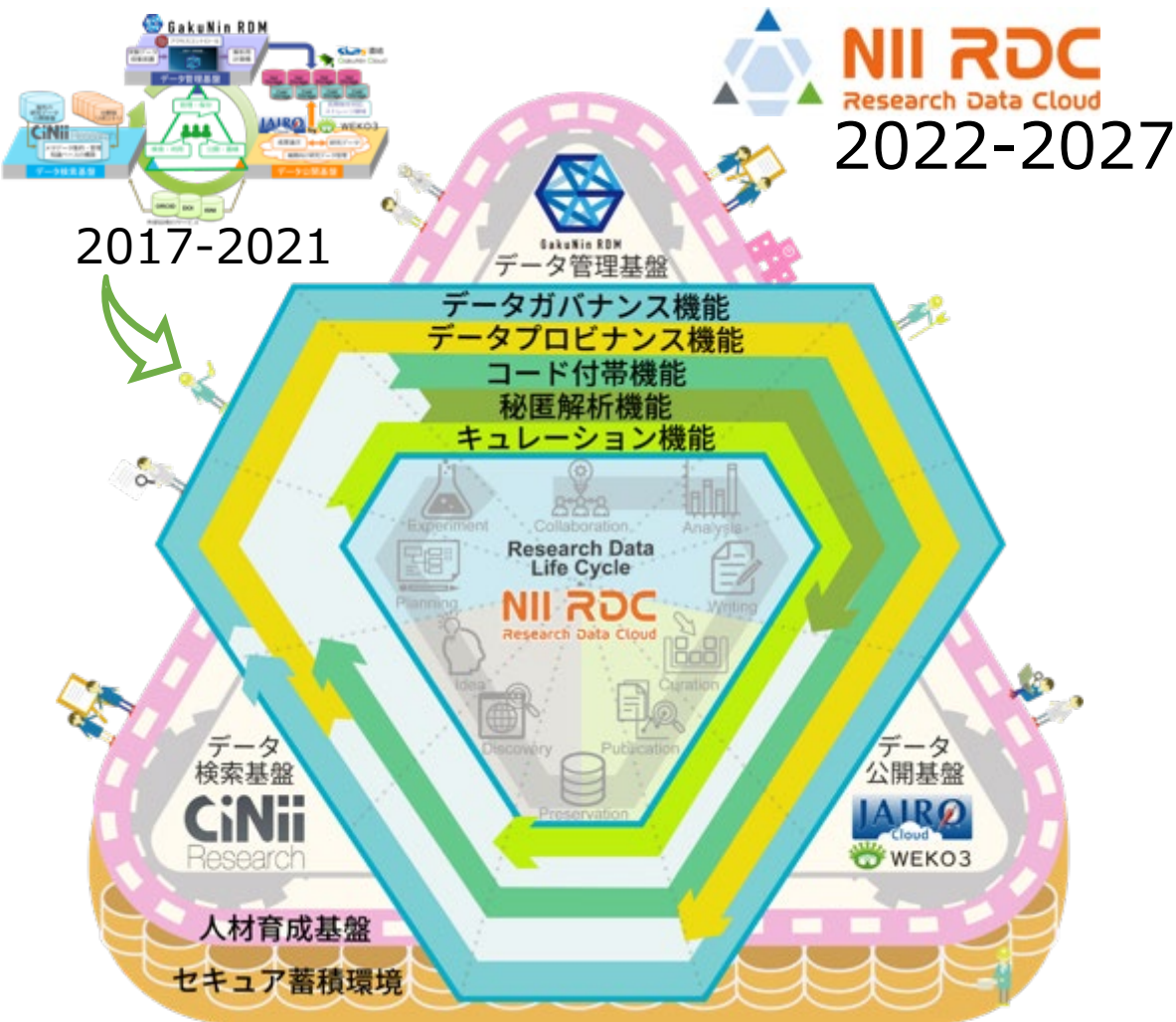
【科学技術・イノベーション政策において目指す主要な数値目標】

(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**。

学術研究データプラットフォーム NII Research Data Cloud (NII RDC)

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



機能評価試験中

● **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

一部機能提供中

● **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

実証実験中

● **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

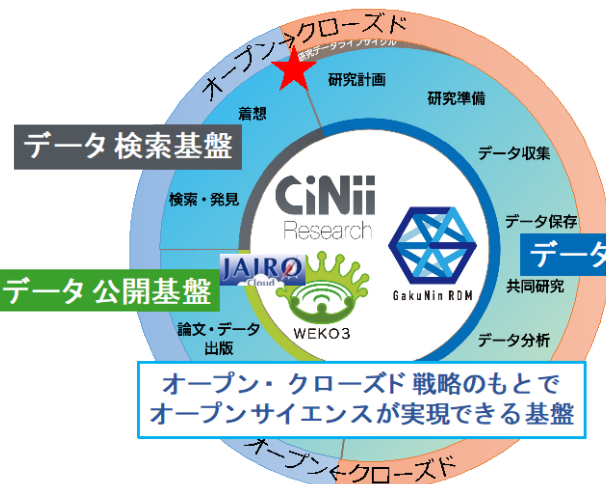
● **キュレーション機能**
専門的なデータキュレーションを実践できるエコシステムを構築し、データ再利用の促進に寄与

● **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

学認LMS

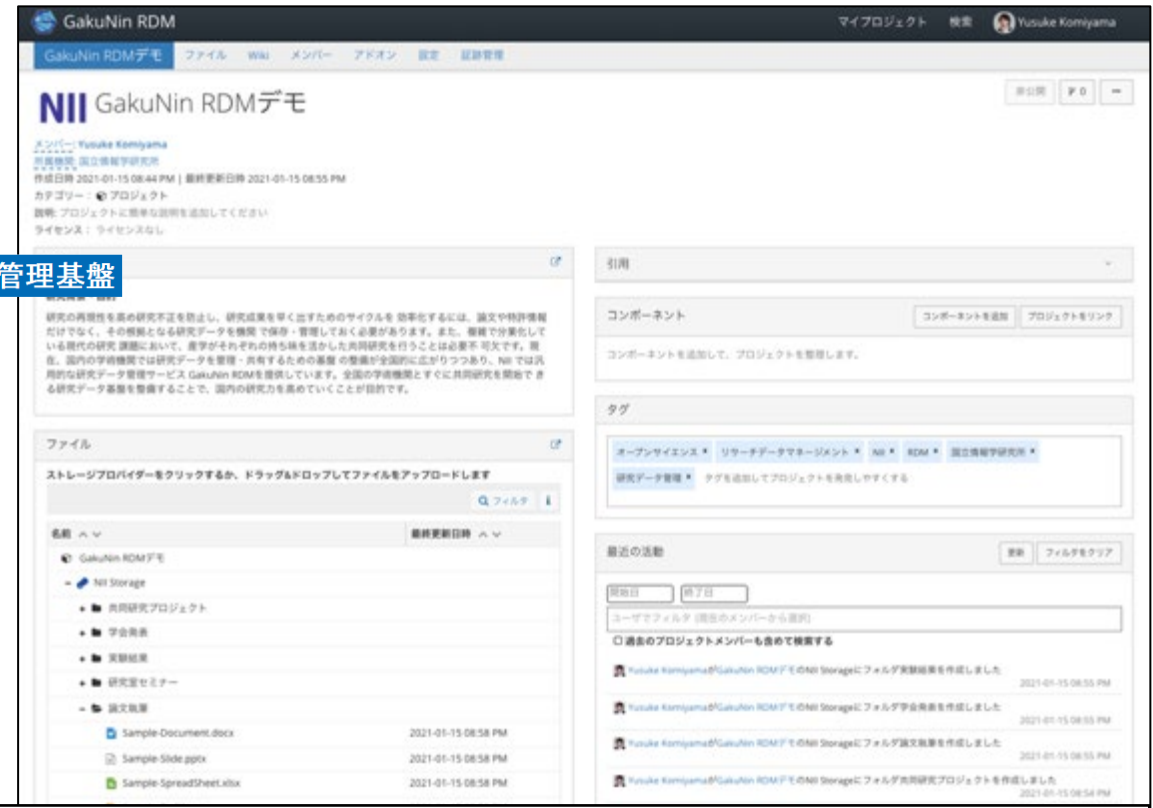
● **人材育成基盤**
研究データ管理に必要なスキルを学ぶ環境を提供し、全ての研究者を新しい科学の実践者へと育成

研究データ管理基盤 GakuNin RDM



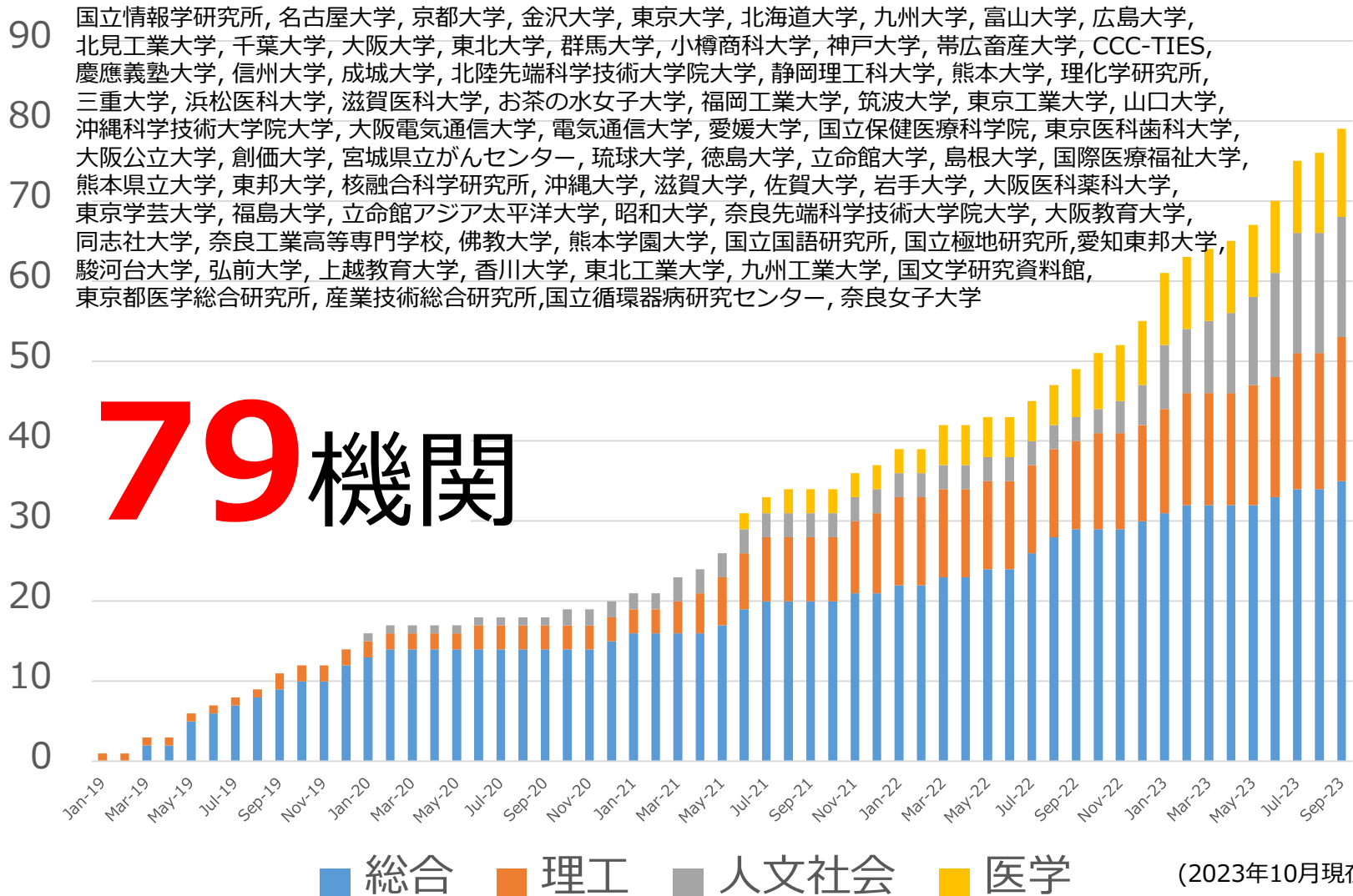
オープン・クローズド戦略のもとで
オープンサイエンスが実現できる基盤

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。

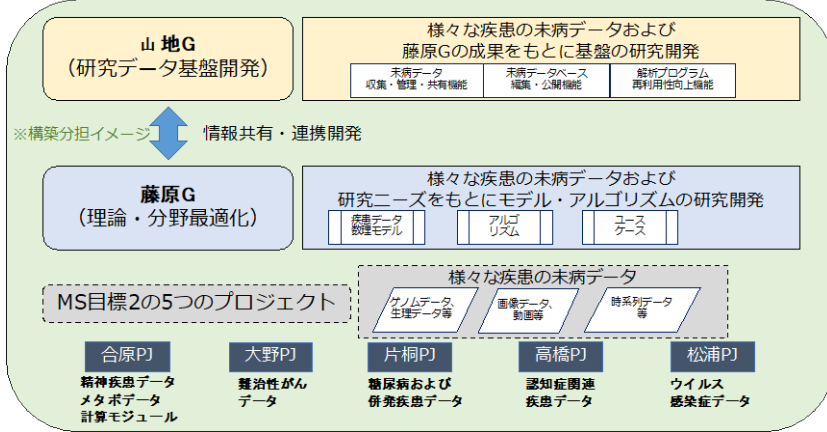
GakuNin RDMの利用機関数の増加



GakuNin RDMのユースケース

ムーンショット目標2未病DBへの環境提供

大規模未病データベースシステム



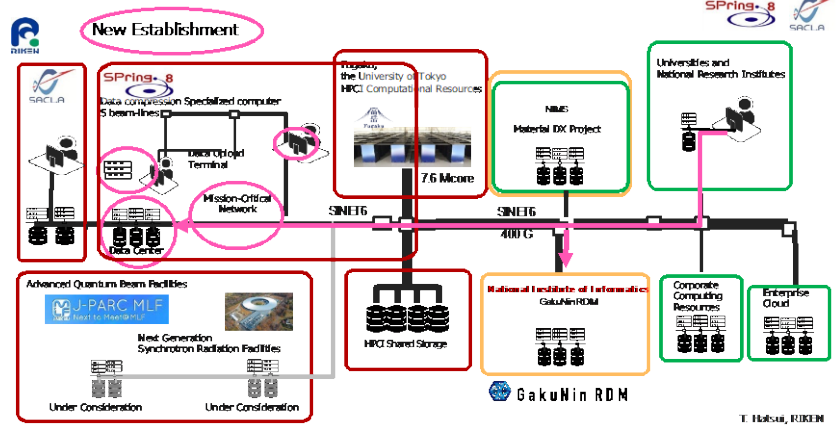
mdxと管理基盤・解析機能の連携

Gakunin RDMとmdxの接続

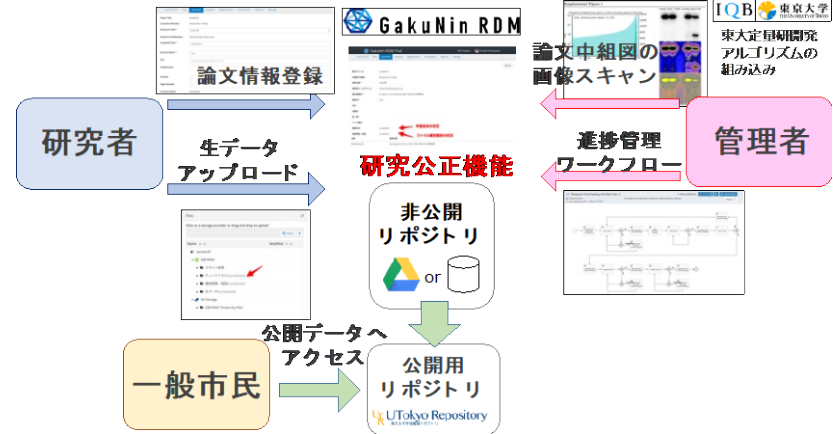
- mdxの共有オブジェクトストレージ (S3互換)をGakunin RDMと接続
- つまり, ストレージ+アクセス性能のよい高性能計算機をmdxが提供



理研SPring-8新設データセンターと管理基盤のストレージ連携



東京大学定量生命科学研究所の研究公正管理システムとの連携



研究データ管理基盤 GakuNin RDM

研究推進: GakuNin RDMは、研究者が自身の研究データを、データポリシーやデータマネジメント計画に沿って管理するためのWebサービスである。多様なクラウドストレージや研究用ソフトウェアを接続・利用することが可能。

(1) 学認フェデレーションへの対応

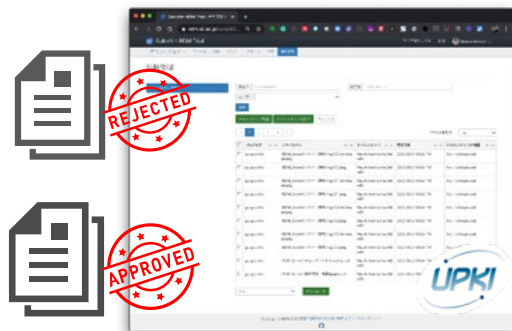


(2) NII RDCや外部研究ツールとの連携

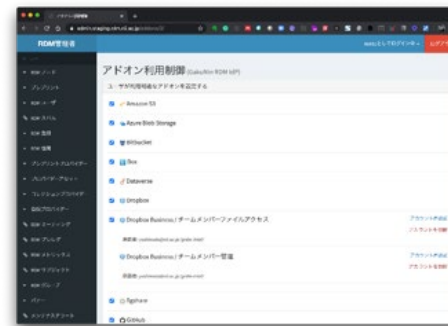


研究公正: GakuNin RDMは、大学や研究機関がガバナンスのために所属する研究者の研究データを管理するためのサービスでもある。研究機関はストレージをカスタマイズすることができる。

(3) 研究データの証跡保存



(4) 機関のシステム管理者向けの機能



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

この画面は、研究プロジェクトのメタデータとファイル管理を扱うインターフェースです。左側のメニューには、タイトル、共著者、日付、ライセンス、要旨 (Wiki)、ファイル管理が並び、それぞれが画面内の対応する項目にリンクしています。右側のメニューには、引用、関連プロジェクト、検索用タグ、操作ログが並び、それぞれが画面内の対応する項目にリンクしています。

プレビュー機能

この画面は、RCOS_leafletA3_2018.pdf (Version: 1) のプレビュー画面です。左側にはファイルのリストが、右側にはPDFのプレビューが表示されています。また、プロジェクト用のWikiページも表示されています。

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

この画面は、マルチクラウドファイルマネージャーの構成例を示しています。ファイルとフォルダのリストが表示されており、サイズ、バージョン、ダウンロードリンク、および最終更新日時が示されています。

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

この画面は、プロジェクト内のディスカッション機能を示しています。コメントの追加や返信が行えるようになっています。

研究データ管理機能 プロジェクト画面 (2)

プロジェクトの階層化

プロジェクトの階層化機能により、プロジェクトを階層的に管理できます。例えば、教員プロジェクト、学生Aのプロジェクト、学生Bのプロジェクトなど、異なるユーザーや役割に基づいてプロジェクトを分類できます。

Name	Size	Version	Downlo...	Modified
教員プロジェクト				
Google Drive: GakuNin RDMデモ				
NII Storage				
Sample-Document.docx	6.9 MB	1	0	2019-10-23 09:53 PM
Sample-Slide.pptx	7.8 MB	1	0	2019-10-23 09:49 PM
Sample-SpreadSheet.xlsx	1.2 MB	2	0	2019-10-23 09:27 PM
Sample-ZipFile.zip	7.4 MB	1	0	2019-10-23 09:15 PM
学生Aのプロジェクト				
NII Storage				
Sample-Document.docx	6.9 MB	1	0	2019-10-23 10:32 PM
学生Bのプロジェクト				
NII Storage				
Sample-SpreadSheet.xlsx	1.2 MB	1	0	2019-10-23 10:32 PM

研究活動の履歴

最近の活動履歴を確認できます。例えば、Yusuke Komiyamaがタイトルを研究データ管理基盤GakuNin RDMからGakuNin RDMデモプロジェクトに変更しました、2022-05-16 12:56 AM。

フォルダアップロード機能

フォルダアップロード機能により、プロジェクトにフォルダをアップロードできます。フォルダ選択画面で容量事前計算・可否確認が行われます。

**フォルダ選択画面
容量事前計算・可否確認**

公的研究資金に基づく研究データに対するメタデータ登録機能

メタデータ登録機能により、研究データのメタデータを登録できます。登録することで資金配分機関へデータ提出するフォーマットでダウンロードできます。

メタデータ編集

プロジェクト/ファイル/フォルダ

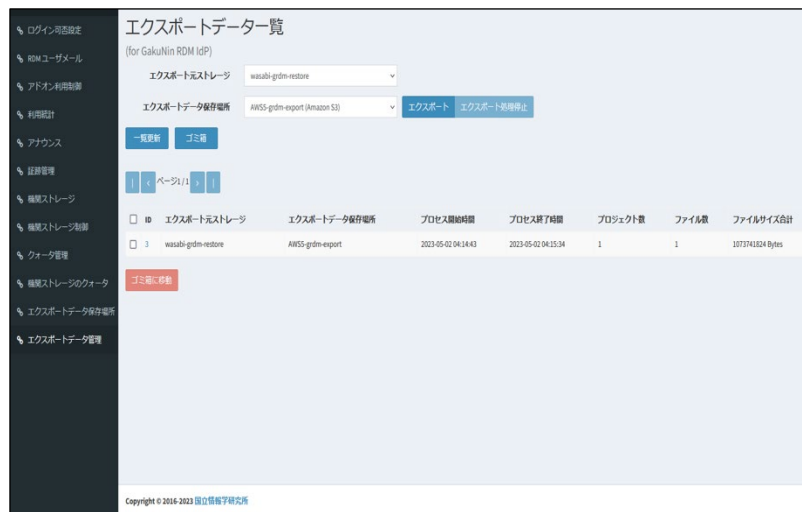
CSVへのエクスポート

機能の詳細については、「NIIサービス説明会2022」をご参照ください
<https://www.youtube.com/watch?v=QFLLbzuSXDw>

GakuNin RDM周辺機能

今後予定している機関管理者向け機能

- 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



エクスポート管理画面



リストア画面(含内容確認)

GakuNin RDMを操作する コマンドラインツール

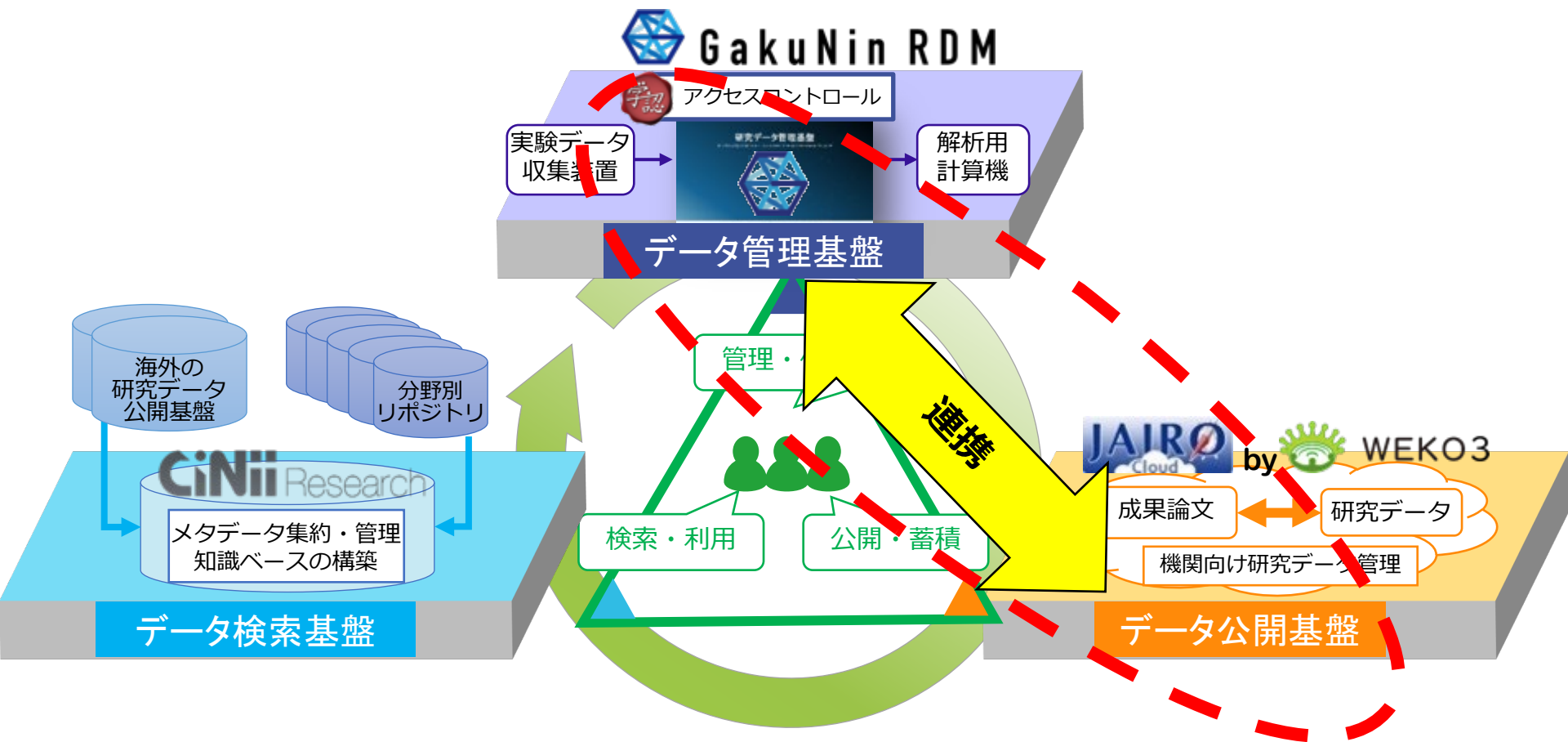
- コマンドラインツールのPythonコードをGitHubに於いて公開中
 - ファイル・フォルダのリスト・アップロード・ダウンロード操作
 - <https://github.com/RCOSDP/rdmclient>
- 利用方法
 - ユーザ情報設定画面で生成される「パーソナルアクセストークン」と、GakuNin RDMプロジェクトの「GUID」を利用
 - 「パーソナルアクセストークン」はパスワード等と同様、セキュアな管理が必須
 - 詳細な利用方法については、今後マニュアル等を整備していく予定
 - 今後、コマンドライン機能を充実させ、プロジェクト作成・ユーザ設定用のコマンドラインツールも提供予定

実行例：ファイルリスト抽出

```
% export OSF_TOKEN=「パーソナルアクセストークン」  
% osf --base-url https://api.rdm.nii.ac.jp/v2/ -p 「GUID」 list -l  
2023-09-29 21:26:59 57 osfstorage/sample.txt  
2023-09-29 19:39:01 9 osfstorage/テストファイル.txt
```

公開基盤連携機能 (コンセプト)

研究データ基盤 NII Research Data Cloud



事前準備： WEKO3でのアプリケーションの作成

WEKO

Language: English

wekosoftware@nii.ac.jp

Home / Account / Applications / New

Settings

- Profile
- Change password
- Security
- Applications**
- Groups
- Session
- Administration

New OAuth Application

Name

Name of application (displayed to users).

Description

Optional. Description of the application (displayed to users).

Website URL

URL of your application (displayed to users).

Redirect URIs (one per line)

One redirect URI per line. This is your application's authorization callback URLs. HTTPS must be used for all hosts except localhost (for testing purposes).

Client type

Confidential

Select confidential if your application is capable of keeping the issued client secret confidential (e.g. a web application), select public if your application cannot (e.g. a browser-based JavaScript application). If you select public, your application MUST validate the redirect URI.

Redirect URIs: (GRDM Web Pod URL)/oauth/callback/weko/weko.sample.ac.jp.(Client ID先頭4文字)/

Client type: Confidential

WEKO3の Application 画面を開き、OAuth2アプリケーションを作成。同一のWEKO3とGRDMを接続することを可能にするため、Redirect URIにClient IDの文字列を含む仕様としている。一度アプリケーションを登録してClient IDを確定させてから、RedirectURI を変更。

GRDM機関管理者画面での WEKO3アプリケーションの登録

Configure a WEKO3 application

WEKO3 Display Name *WEKO3 Display Name: (好きな名前)*

WEKO3 URL *WEKO3 URL: https://(WEKO3のURL)*

WEKO3 OAuth Client ID *WEKO3 OAuth Client ID: (Client ID)*

WEKO3 OAuth Client Secret *WEKO3 OAuth Client Secret: (Client Secret)*

キャンセル 保存

追加

Zotero

Copyright © 2016-2023 国立情報学研究所

WEKO3に作成したOAuth2アプリケーションを
GRDMの機関管理者のアドオン利用制御画面から登録。

GRDMエンドユーザ画面での プロジェクトダッシュボード

The screenshot displays the GakuNin RDM project dashboard for 'WEKO3-2'. The top navigation bar includes 'WEKO3-2', 'ファイル', 'IQB-RIMS', 'Wiki', 'メタデータ', 'Export', 'メンバー', 'アドオン', '設定', and '証跡管理'. The user profile 'Satoshi Yazawa' is visible in the top right. The main content area is divided into three sections:

- File List (Left):** A table with columns '名前' and '最終更新日時'. It lists the project structure: 'WEKO3-2', 'NII Storage', '.weko', 'WEKO: GakuNin RDM Project' (highlighted with a red box), 'Sub Index', 'WEKO3連携', 'rdm-project.zip', '公開したいデータ' (highlighted with a blue box), and 'DATA-TO-BE-PUBLISHED.dat'.
- Tags (Top Right):** A section titled 'タグ' with a text box for adding tags to projects.
- Recent Activity (Bottom Right):** A section titled '最近の活動' with a red border. It includes filters for '開始日' and '終了日', a search box for users, and a checkbox for '過去のプロジェクトメンバーも含めて検索する'. The activity log shows:
 - Satoshi Yazawa deleted a draft file (test) from the WEKO index (2023-06-28 10:17 PM).
 - Satoshi Yazawa linked the WEKO index to WEKO3-2 (2023-06-28 10:16 PM).
 - Satoshi Yazawa linked the WEKO index (Index for Test) to WEKO3-2 (2023-06-28 10:15 PM).
 - Satoshi Yazawa approved the WEKO3-2 WEKO add-on (2023-06-28 10:15 PM).
 - Satoshi Yazawa removed the approval for the WEKO3-2 WEKO add-on (2023-06-28 10:00 PM).

WEKO3アドオンはストレージアドオンであるため、プロジェクトダッシュボードのファイル一覧内で確認することが可能。GRDM経由でWEKO3アドオンに対して行った変更操作は最近の活動に記録。

GRDMエンドユーザ画面での WEKO3のメタデータ編集ダイアログ

The screenshot shows the 'ファイルメタデータの編集' (Edit File Metadata) dialog box in the GakuNin RDM system. The dialog is overlaid on a background showing a file browser with a tree view of folders and files. The dialog contains the following fields and options:

- メタデータ様式:** WEKO3 デフォルトアイテムタイプ (dropdown menu)
- 公開日*:** 2023-06-28 (text input)
- タイトル (日本語):** データセット (text input)
- Title (English):** DATASET (text input)
- その他のタイトル (日本語):** (empty text input)
- Alternative Title (English):** (empty text input)
- 説明 - アブストラクト (日本語):** テスト用のデータセットです。 (text input)
- Description - Abstract (English):** This is a dataset for testing. (text input)

There is a button labeled 'クリップボードから貼り付け' (Paste from clipboard) located to the right of the '公開日*' field.

ドラフトファイルのメタデータは編集することが可能。
このファイルメタデータはアイテム登録に利用される。

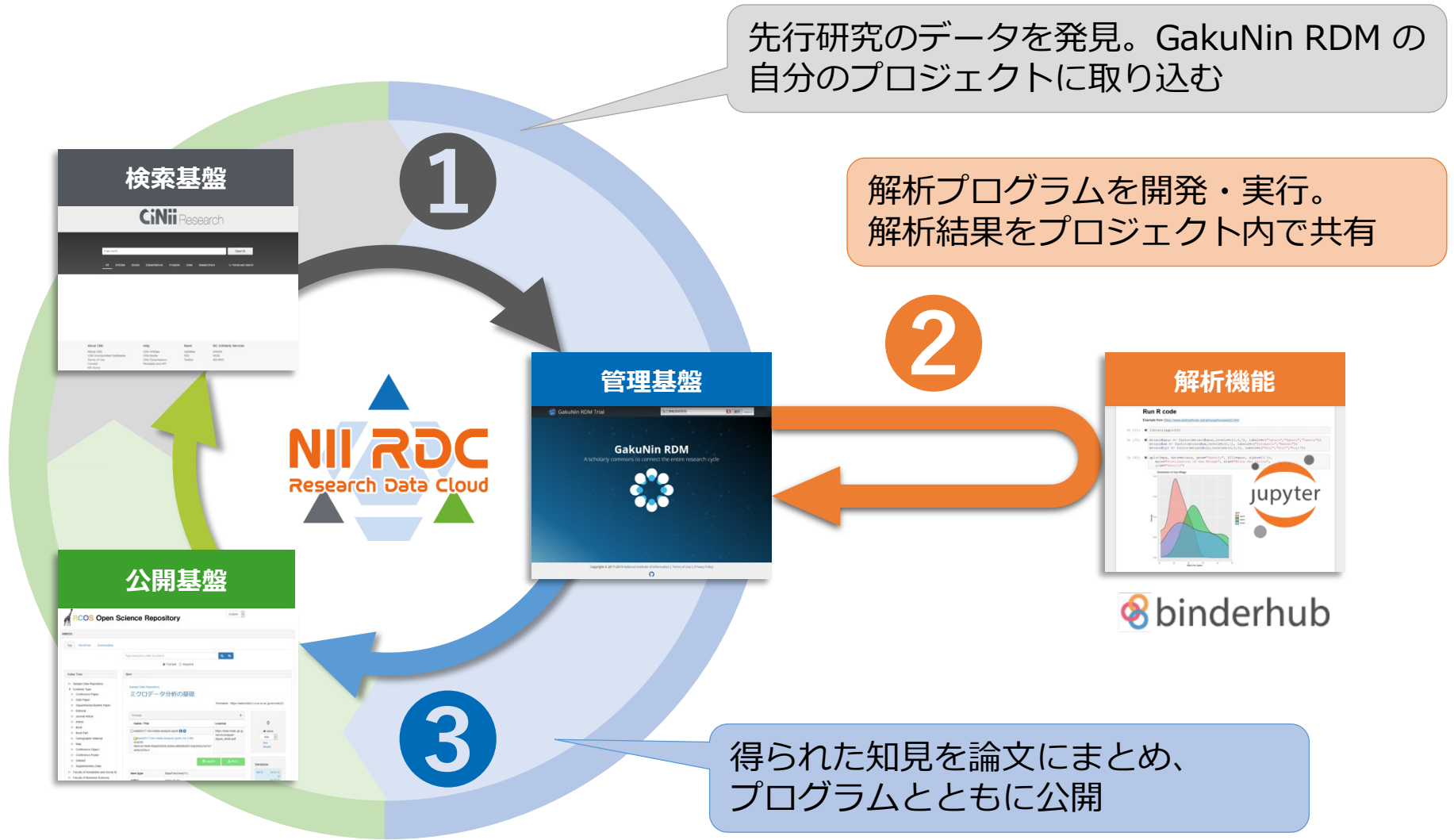
GRDMエンドユーザ画面での WEKOアイテム登録ダイアログ

The screenshot displays the GakuNin RDM WEKO3-2 interface. A modal dialog box titled "ファイルの登録" (File Registration) is open, asking for confirmation to register a file named "DATASET.dat". The dialog text states: "ファイル(フォルダ) 'DATASET.dat' を WEKOに登録しますか? この操作は元に戻せません。" (Do you want to register the file (folder) 'DATASET.dat' in WEKO? This operation is irreversible.) The dialog has "キャンセル" (Cancel) and "OK" buttons. The background interface shows a file list on the left, a toolbar with options like "登録" (Register), "ダウンロード" (Download), and "プレビュー" (Preview), and a main content area with sections for "引用" (Citation), "コンポーネント" (Component), "タグ" (Tag), and "最近の活動" (Recent Activity).

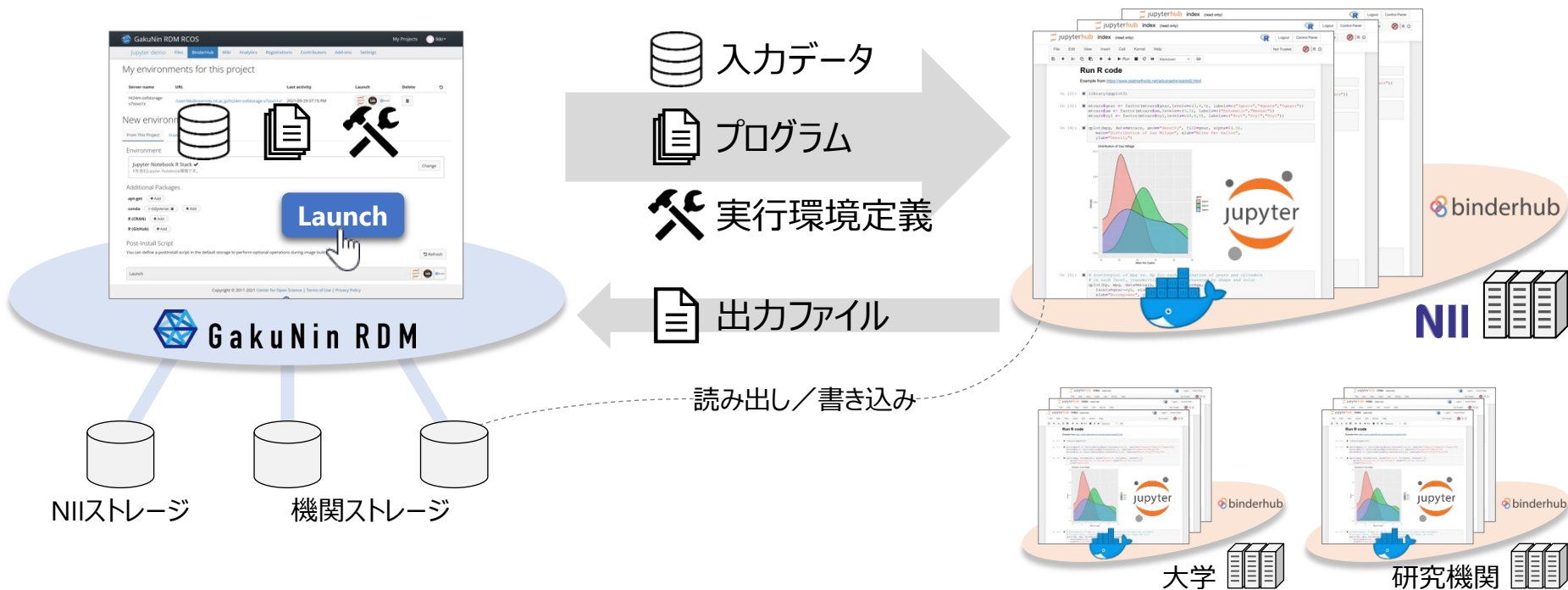
WEKO3アドオンでドラフト（編集中）ファイルを選択し、ツールバー「登録」をクリックすると、そのファイルを登録するかを問い合わせるダイアログが表示。初期メタデータとして、ドラフトファイルに指定されたファイルメタデータの内容を使用。アイテムの登録に成功するとドラフトファイルは削除。

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

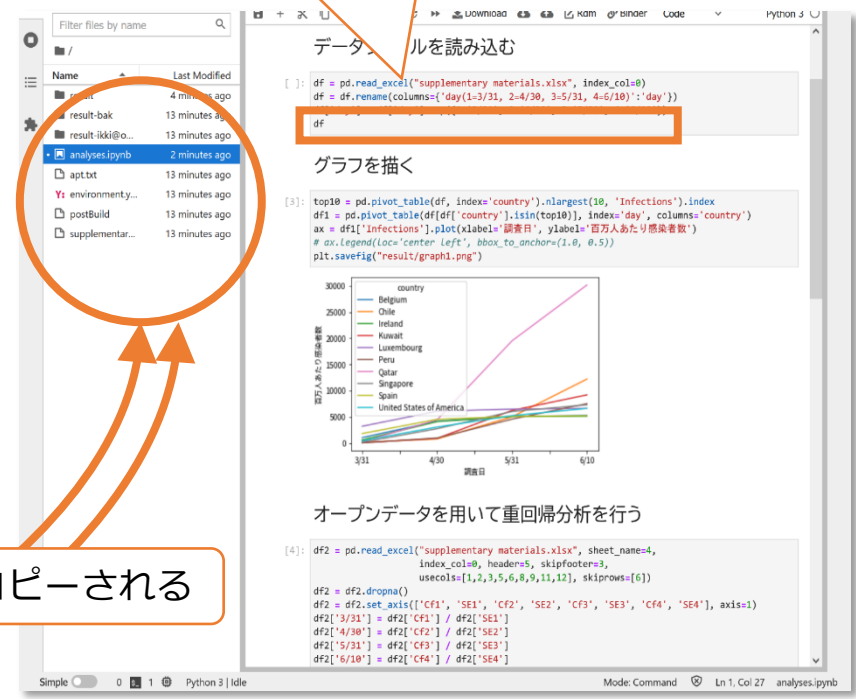
データ解析機能の基本的な使い方

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**

秘匿解析機能

秘匿解析機能

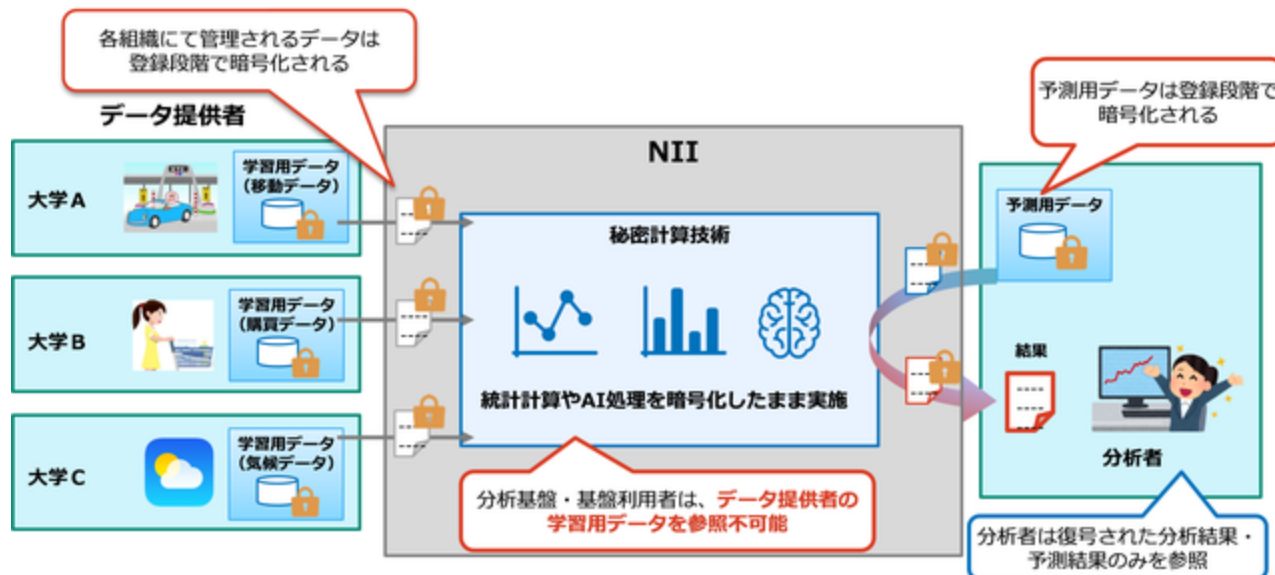
秘匿解析機能とは：データを秘匿（暗号化）したまま解析する機能

● 目的：

- 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

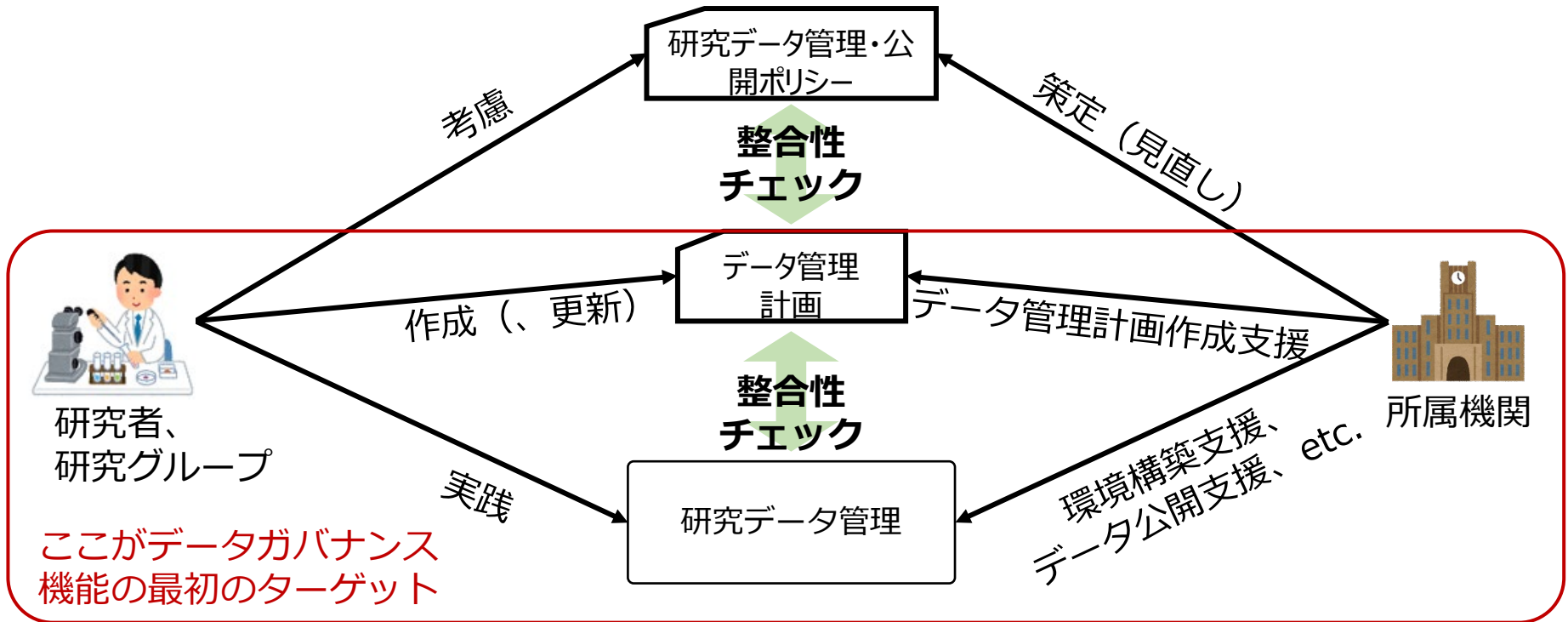


秘匿計算システムトライアル

- **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>
- NIIオープンサイエンス基盤研究センター (RCOS) は、NTTとの共同研究の一環として、研究者が**秘密計算システムを実験的に利用できるトライアル**を実施。
 - 内容：NIIの計算機上で、NTTの秘密計算システムを無償で試用できます。
 - 対象者：国内の大学・研究機関に所属する教職員・研究者の方
 - 受付期間：2023年1月23日から2024年1月31日まで（予定）
 - 提供期間：開始日（ご相談）から2024年3月31日まで（予定）
 - 利用申込： sc-trial2023@nii.ac.jp
 - 研究者の氏名、所属、連絡先メールアドレス、研究内容の概略、秘密計算を利用するメリットを記載の上、上記メールアドレスへメール送付
 - 担当者：藤原一毅
 - 詳細案内：<https://rcos.nii.ac.jp/news/2023/01/20230123-0/>

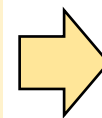
データガバナンス機能

研究データガバナンスとは



研究データ管理の品質を守る・向上するための規律に従う行為。

- ポリシーを策定し、運用する。(研究機関)
- ポリシーおよび管理計画に従う。(研究者)
- ポリシーに応じて環境を整備/利用する。(研究機関/研究者)



研究活動の促進、
組織・研究者としての
信用の向上

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)

実験実行環境用の手順書

メタデータ検証用の手順書

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 Noteb

リサーチ

「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

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
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
つ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。

```

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  
display(HTML(f"<button onclick='window.open(\"{url}\")>実行環境を作成する</butt
    
```

新規実験用の実行環境を作成します。
以下のボタンをクリックしてください。新規タブで開きます。

実行環境を作成する

1. 研究リポジトリのメタデータを用意する

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

データガバナンス機能の機能評価試験版サービスの利用案内

GakuNin RDM におけるデータガバナンス機能の機能評価試験版サービスを提供しております。先行ユーザーからのフィードバックを受けてデータガバナンス機能の改善を実施し、実証実験レベルへのブラッシュアップを実施する計画です。詳細については、GakuNin RDM サポートポータル「お知らせ」にてアナウンスしております。

「データガバナンス機能の機能評価試験開始のご案内」

- 提供予定期間： 2023/6/19～2024/3/31
- 問い合わせ先： データガバナンス機能サポート
dg_support(at)nii.ac.jp
- 担当者：平木俊幸
- 問い合わせ時に必要な情報：
氏名、所属、連絡用メールアドレス、利用希望者リスト、参加希望理由を記載の上、上記までメール送信ください。

まとめ

- 「第6期科学技術基本計画」の閣議決定を受けて、大学において組織的な研究データ管理やデータガバナンスの必要性が高まってきている背景を述べた。
- GakuNin RDMとは、どのようなサービスか概要を紹介し、さらに公開基盤との連携機能のコンセプトについて紹介した。
- NII RDCで取り組んでいるコード付帯機能、秘匿解析機能、ガバナンス機能について紹介した。

GakuNin RDMサポートポータル

検索 サイト内検索

ENGLISH NII 国立情報学研究所WEBサイト

GakuNin RDM サポートポータル

はじめての方 ご利用中の方 機関管理者の方 ログイン

研究チームのデータ管理を GakuNin RDMに統合しよう。

GakuNin RDMは、チームを率いる研究者の悩みを解決します。

お知らせ

一覧へ >

2023.02.03 15:00

メンテナンス

連携タイムスタンプサービス(UPKI)のメンテナンス作業実施のお知らせ (2023年2月19日)

GakuNin RDMと連携するタイムスタンプサービスメンテナンスのため、次の時間帯において若干の通信遅延または1秒程度の通信断が複数回発生する可能性があります。

<https://support.rdm.nii.ac.jp/>

RCOS
rcos-ext@nii.ac.jp