NII RCOS
Research Center for Open Science and Data Platform

GakuNin RDM 上でのデータガバナンスの実現に向けて

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
データガバナンス機能担当
平木俊幸、横山重俊

2023年5月29日

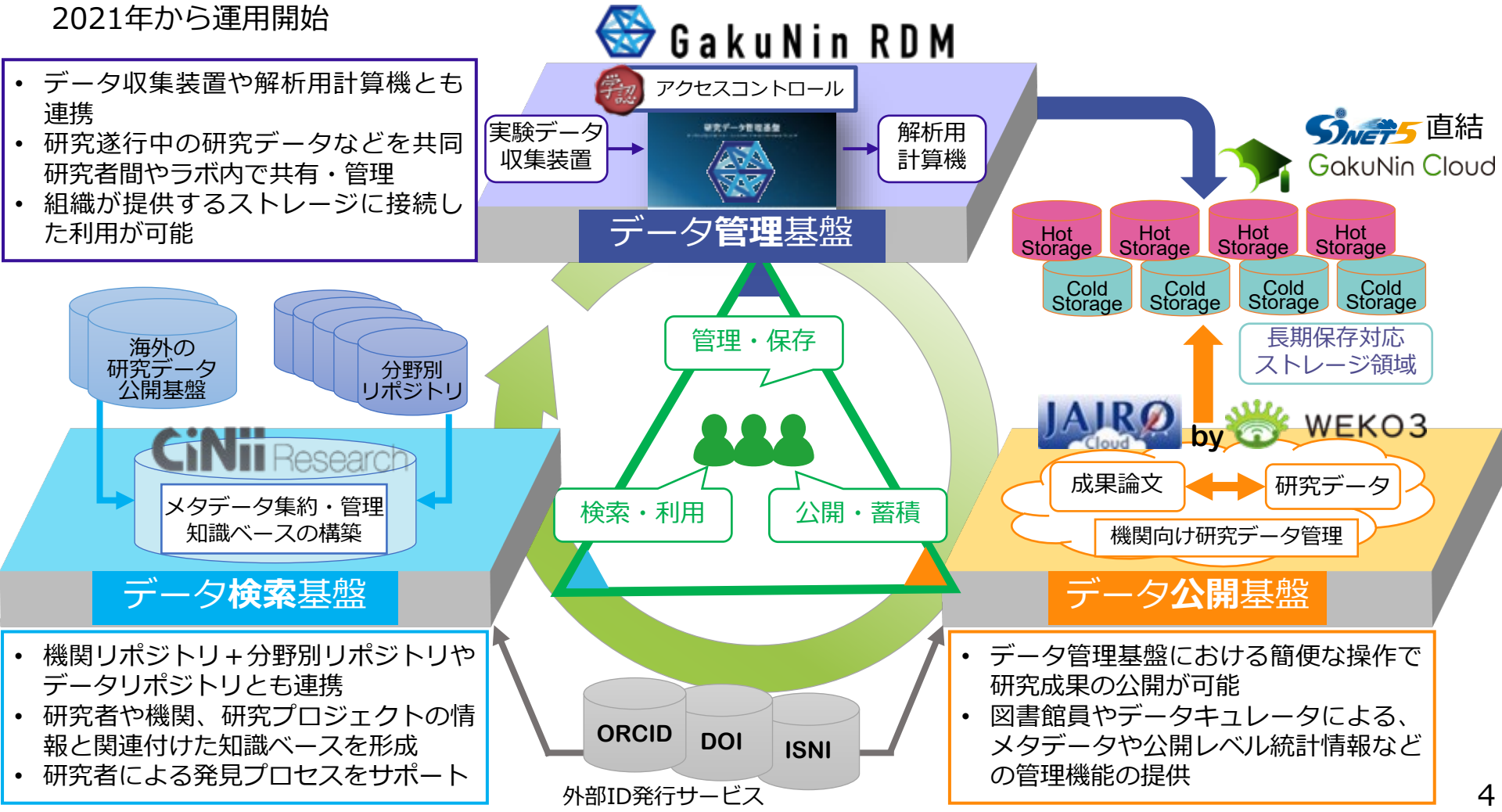
Table of contents

1. データガバナンス機能とは

研究データ基盤 NII Research Data Cloud (NII RDC) の 3 つの基盤

Open science と研究公正を支え、データ駆動型研究を推進する情報基盤

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

NII RDC における 7 機能とデータガバナンス機能の位置づけ

2017-2021

NII RDC を 7 つの側面から機能拡張

2022-2027

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

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

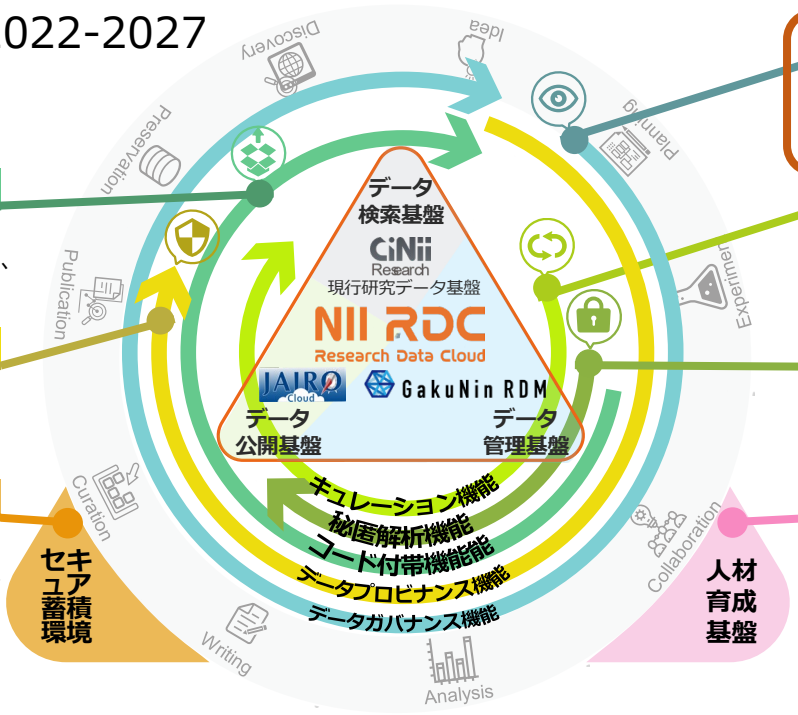
キュレーション機能 **流通**
専門的なキュレーションを実践できるエコシステムを構築し、データ再利用の促進に寄与

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

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

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

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

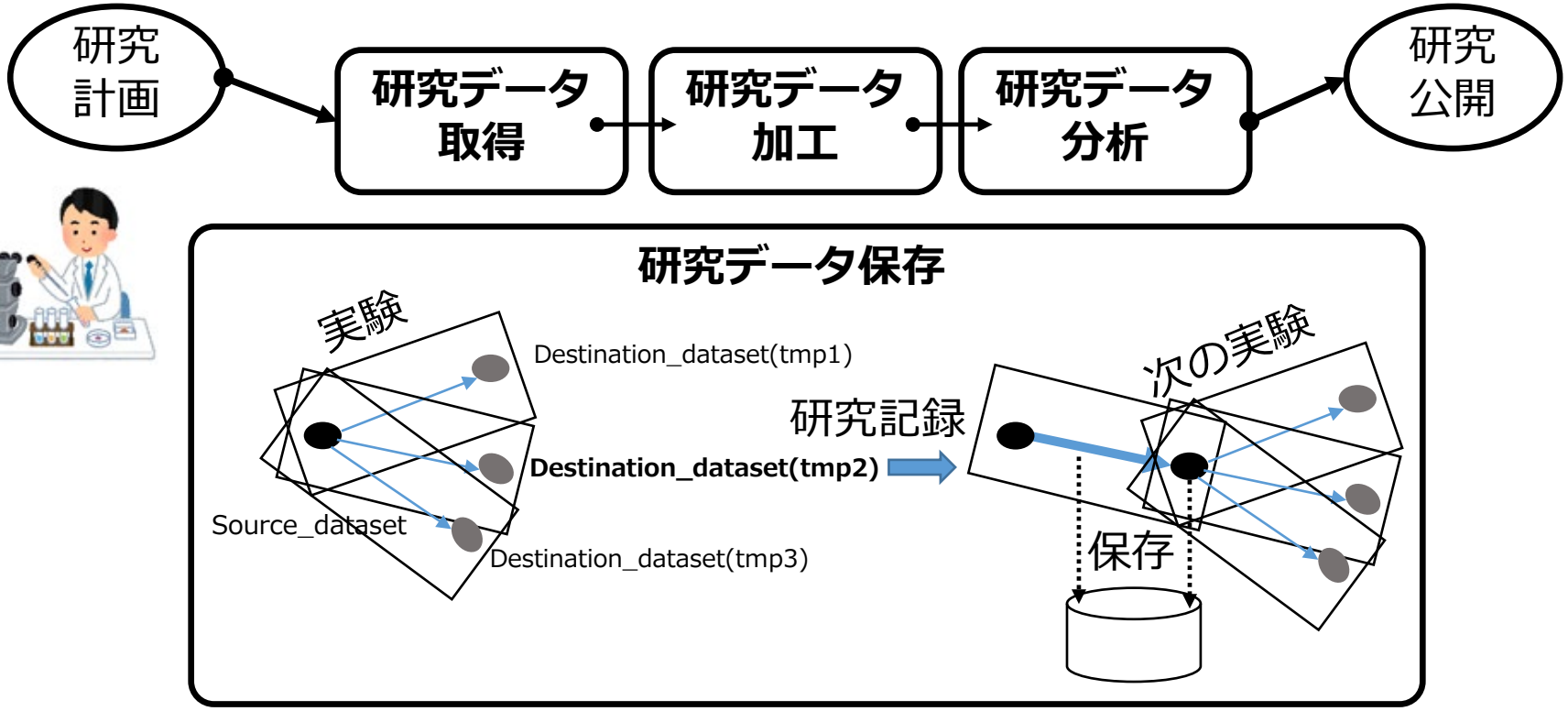


データガバナンス機能は研究活動全体にわたって研究データ管理の品質の向上をサポートする。

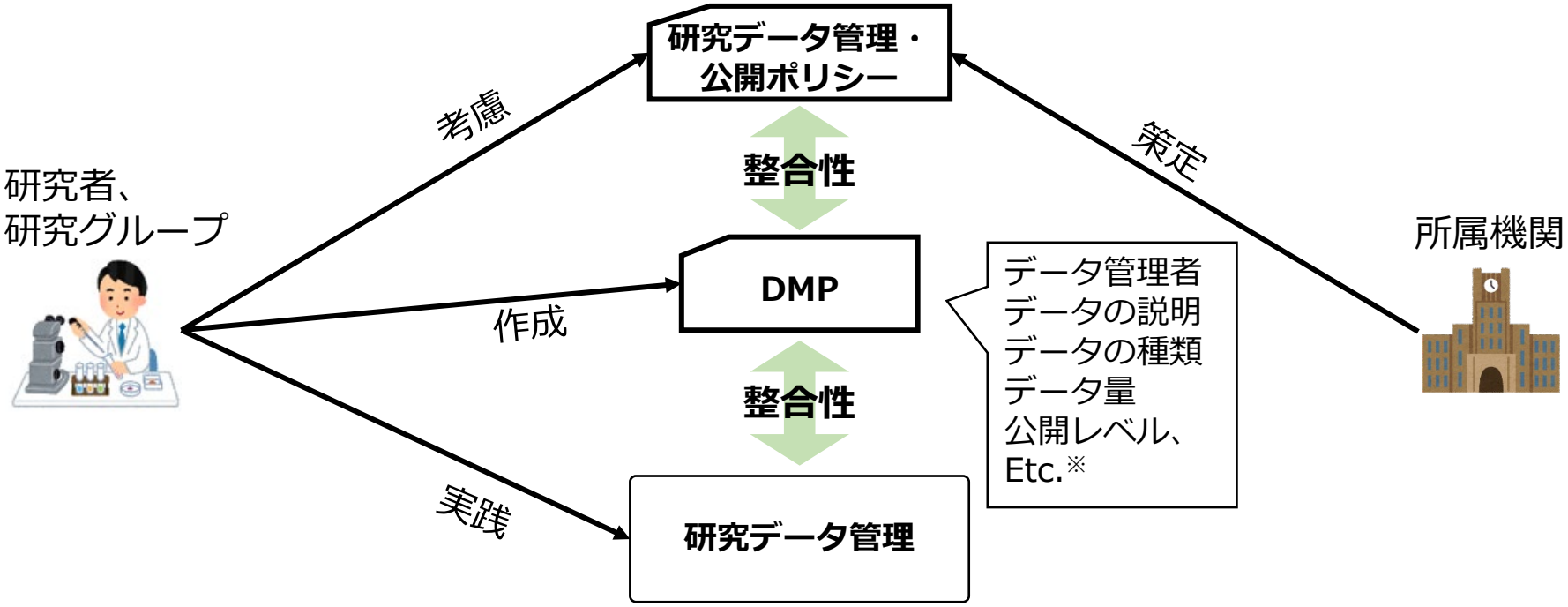
研究データ管理とは

研究データ※の取得・加工・分析・保存・公開などの行為

※研究データ：研究者が研究利用等の対象としたデータ。



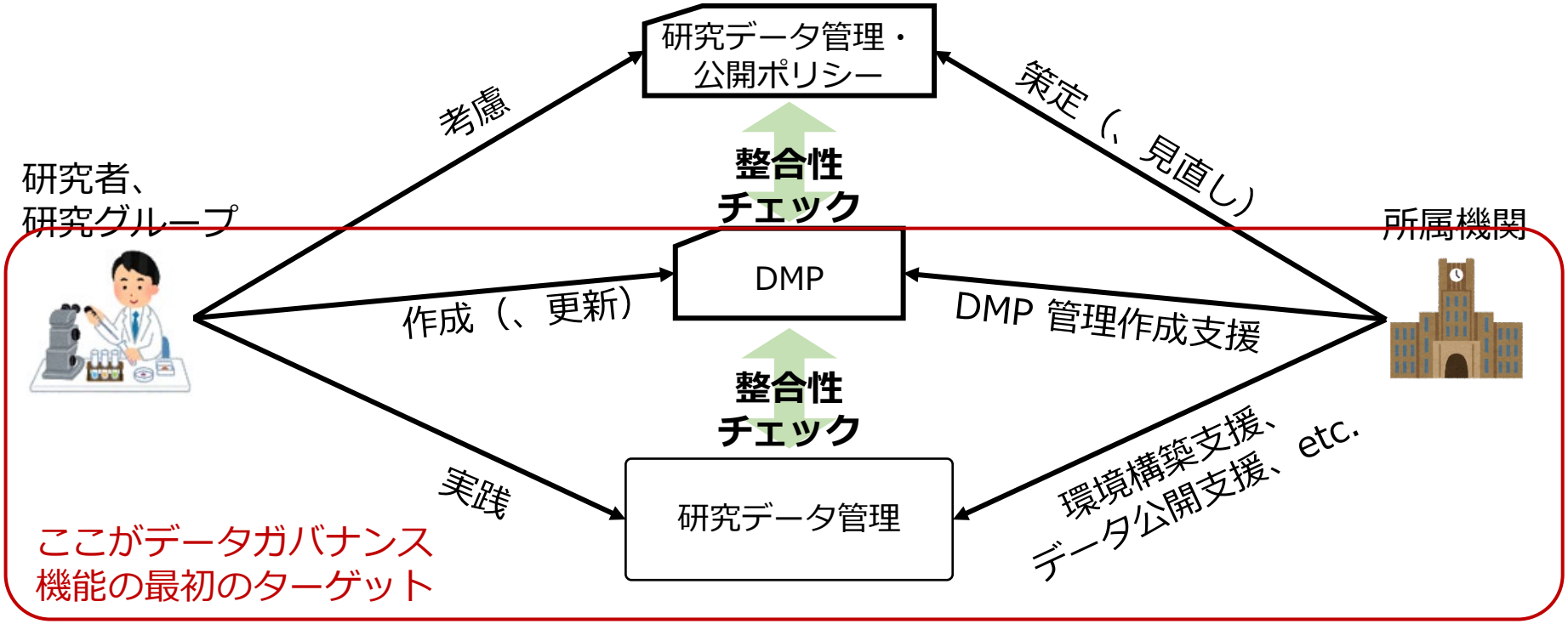
データ管理計画（DMP）とは



研究のために収集・作成する研究データの取扱いや整備・保存・公開についての計画を定めた文書

※DMPの記述内容は提出先の助成機関によって異なる。

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

GakuNin RDM におけるデータガバナンス機能の課題設定

NII RCOS
Research Center for Open Science and Data Platform

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…

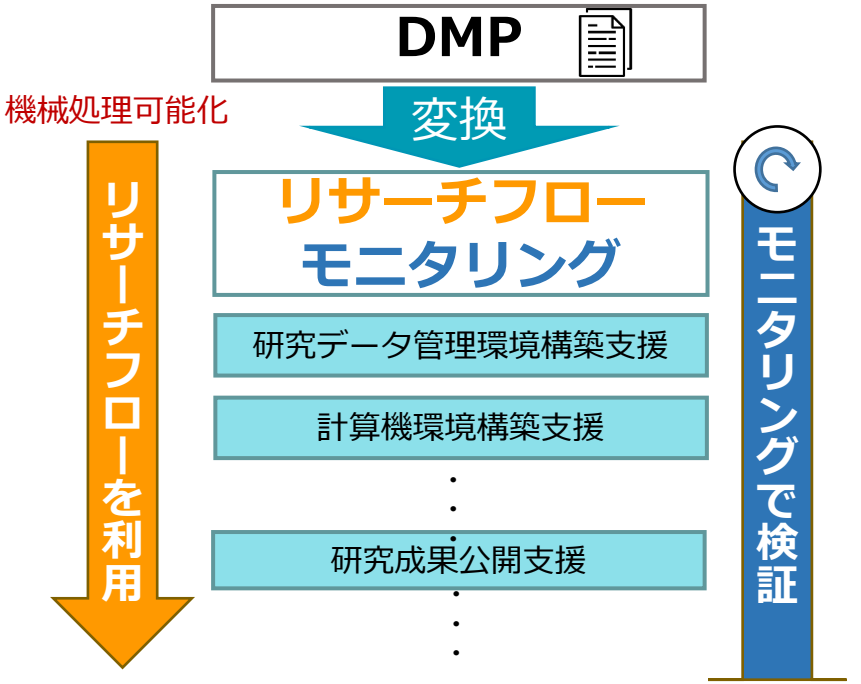
No.	データ名	データの説明	管理者	分類	公開レベル	秘匿理由
1	○○実証においてセンサより撮像したデータ及び関連データ	○○実証においてセンサより撮像したデータであり、道路の画像データ	○○研究所	委託者指定データ	レベル４（広範な提供・利活用）	秘匿しない

提出

それで終わり

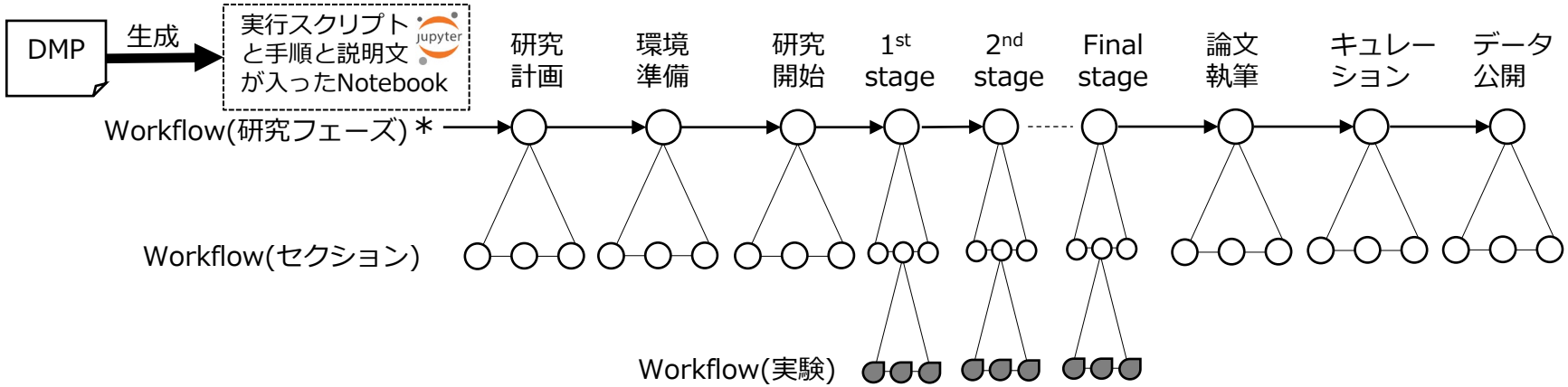
解決案

- 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。

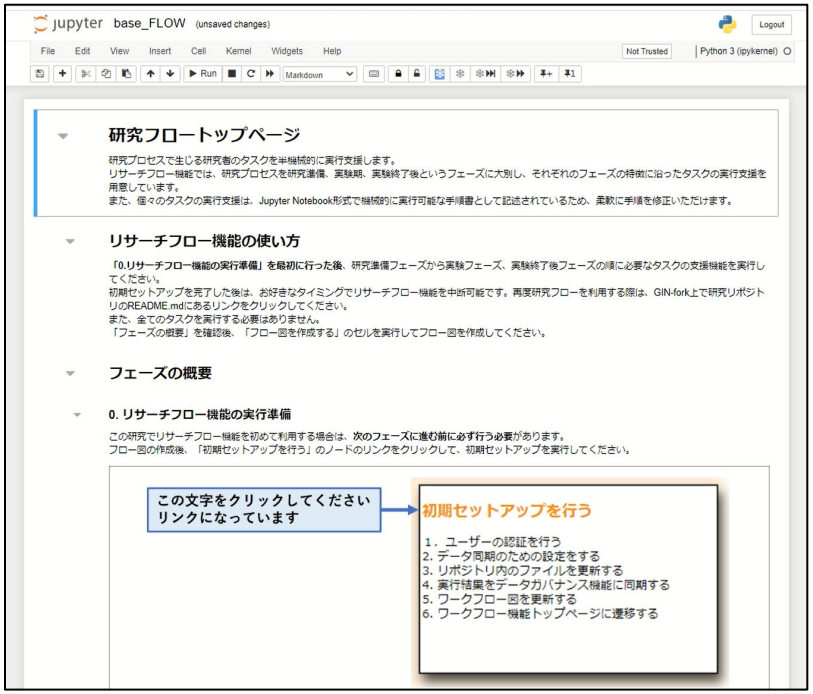


リサーチフロー

DMP によって表現された計画に基づき生成された、実行可能な手順書



リサーチフロー
（Notebook）の例



ポリシー → プラン → データ管理実施

モニタリング

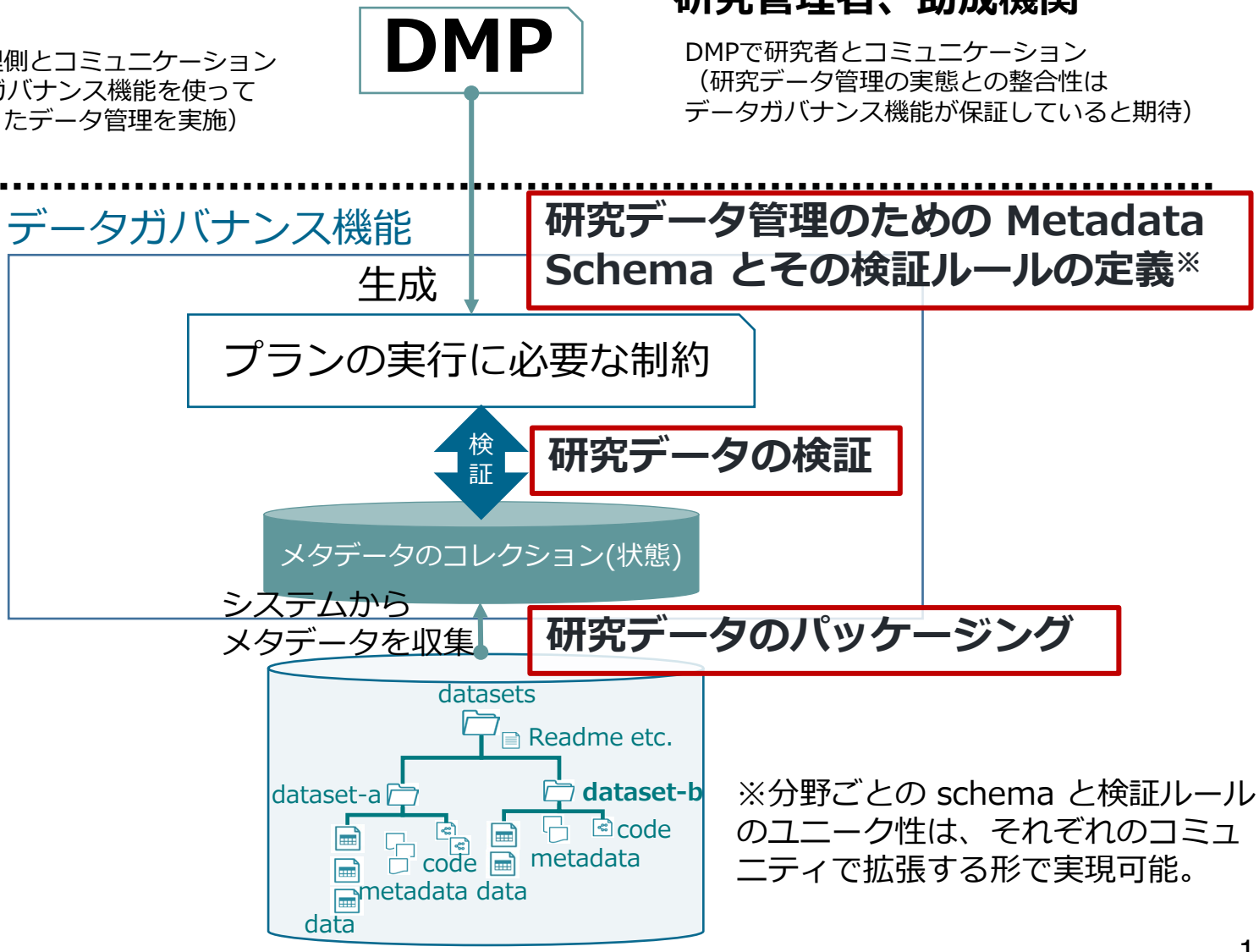
研究データの状態が DMP 等による制約を満たしているかどうか検証する

研究者

DMPで管理側とコミュニケーション
（ データガバナンス機能を使って
DMPに沿ったデータ管理を実施）

研究管理者、助成機関

DMPで研究者とコミュニケーション
（研究データ管理の実態との整合性は
データガバナンス機能が保証していると期待）



※分野ごとの schema と検証ルールのユニーク性は、それぞれのコミュニティで拡張する形で実現可能。

2. データガバナンス機能のデモ

デモの内容

研究データ管理状態をモニタリング機能を用いて検証し、NG であればその原因を把握でき、解決する。

【想定シーン】

論文査読に向けて、論文の根拠となる研究データを査読者が再現可能な状態を作り出すための準備を進めている。

査読者へ出す前に、現時点での研究データ管理状態が以下の二つのレベルで問題ないかどうか検証する。

①関連ファイルの存在性
②再現性

【デモの流れ】

論文の根拠となる研究データの再現性を、モニタリング機能を用いて検証したところ、**NG**が返ってきた。

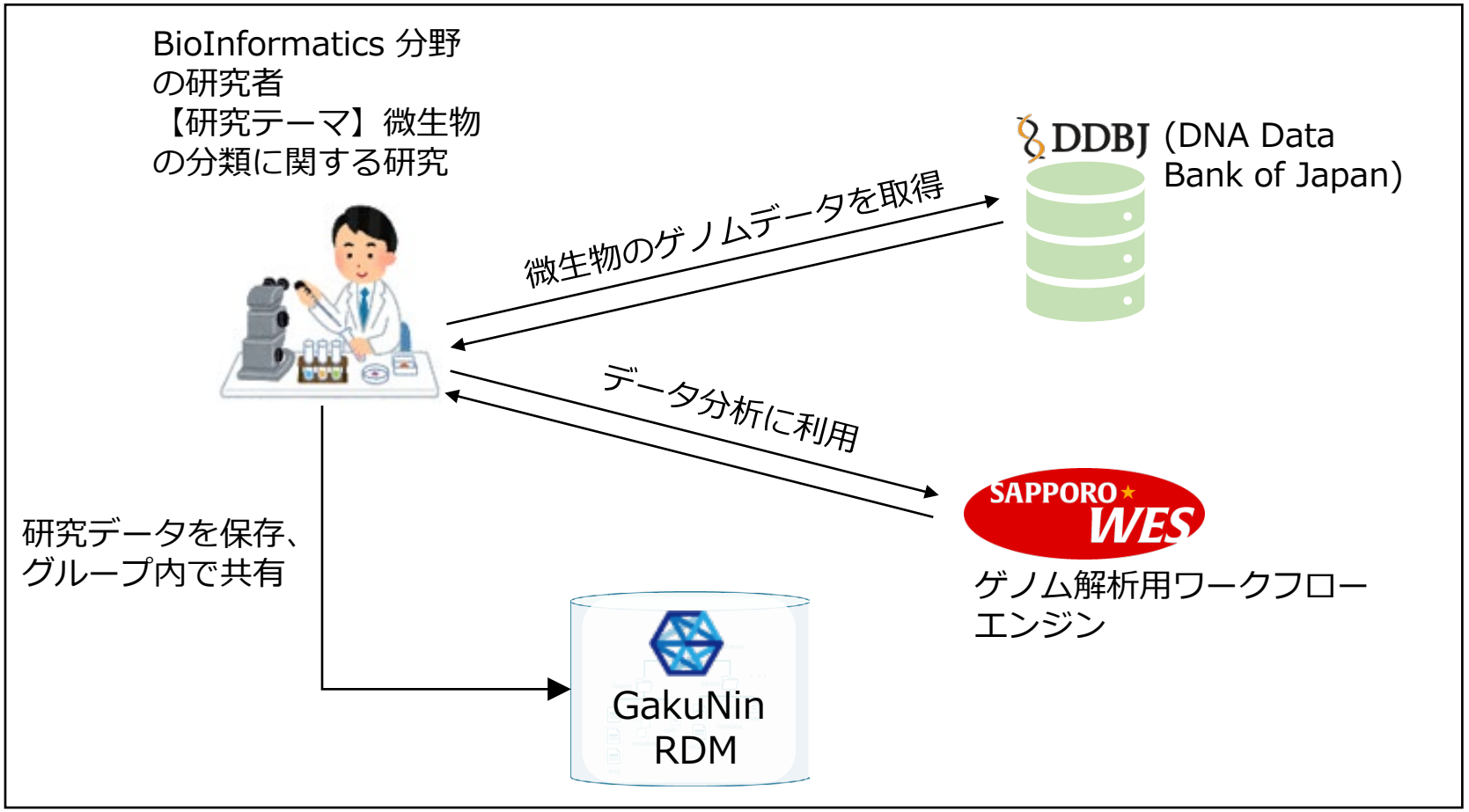
検証結果の記録にあるNGの理由が分かるメッセージを確認し、問題個所を探す。

問題個所が分かったのでそれを解消する。

モニタリング機能を用いて再現性を再検証する。

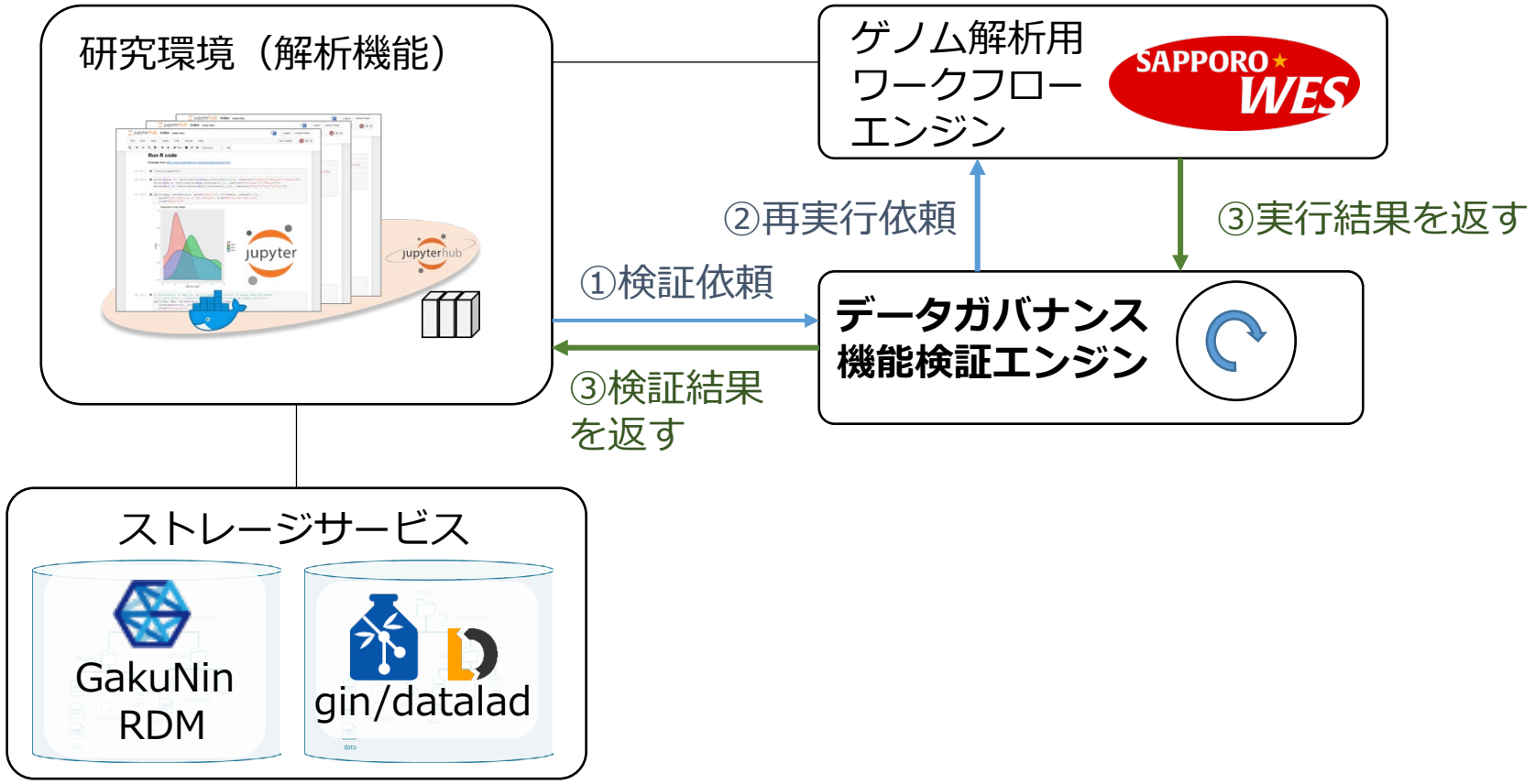
検証の結果、**再現性が確認された**。

デモで想定しているユースケース



↑この研究活動のうち、今回は分析結果に関連するファイルの存在性と分析結果の再現性を検証する。

デモのシステム構成



今回のデモでは gin/datalad を
用いて研究データを管理

デモでの DMP の一部と Metadata schema + 検証ルール

Moohshot (MS) 型研究開発制度におけるメタデータの一例※

リサーチフローの構造および GIN リポジトリのサイズを決めるメタデータ

項目名	項目の説明
資金分配機関情報	公募型の研究資金を配分した資金配分機関（府省含む）の英語略称
プロジェクト名	プロジェクトの研究代表者が統括する研究開発の範囲の名称
データの名称	管理対象データの特徴を示す名称を入力
概略データ量	管理対象データの概ねのデータ容量
ワークフロー識別子	使用するリサーチフローの種別
想定利用最大データ量	リポジトリに格納するデータの想定利用最大値
データセット構造種別	実験データのディレクトリ構造の種別

Metadata schema + 検証ルール

Moonshot 用

プロジェクト用メタデータの検証で利用

GIN 用

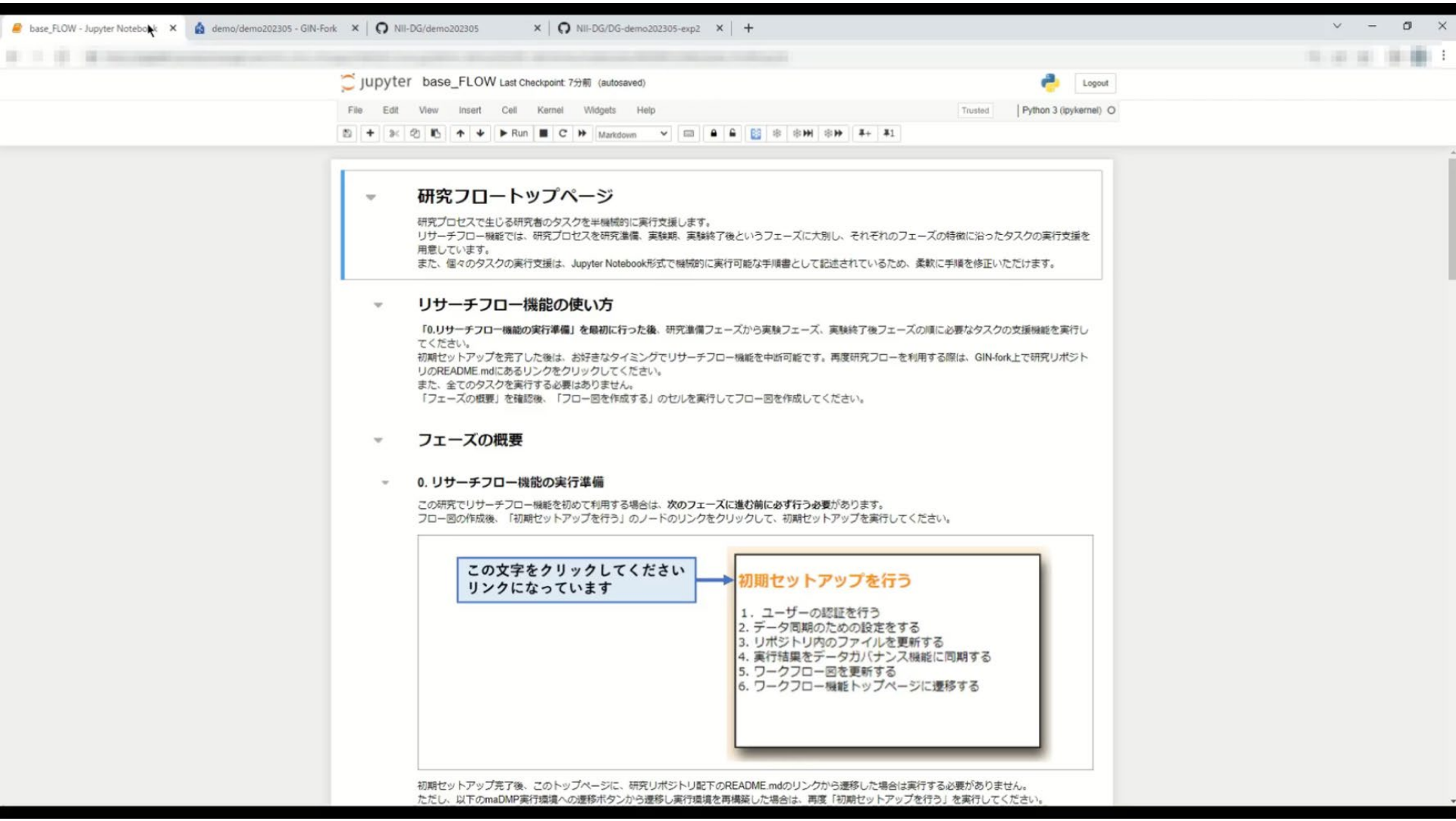
リサーチフロー生成で利用

Sapporo 用

再現性検証で利用

※詳細説明は「ムーンショット型研究開発制度におけるメタデータ説明書（第2版）」(https://web.archive.org/web/20230309075934/https:/www8.cao.go.jp/cstp/ms_metadatainstructions.pdf)を参照。

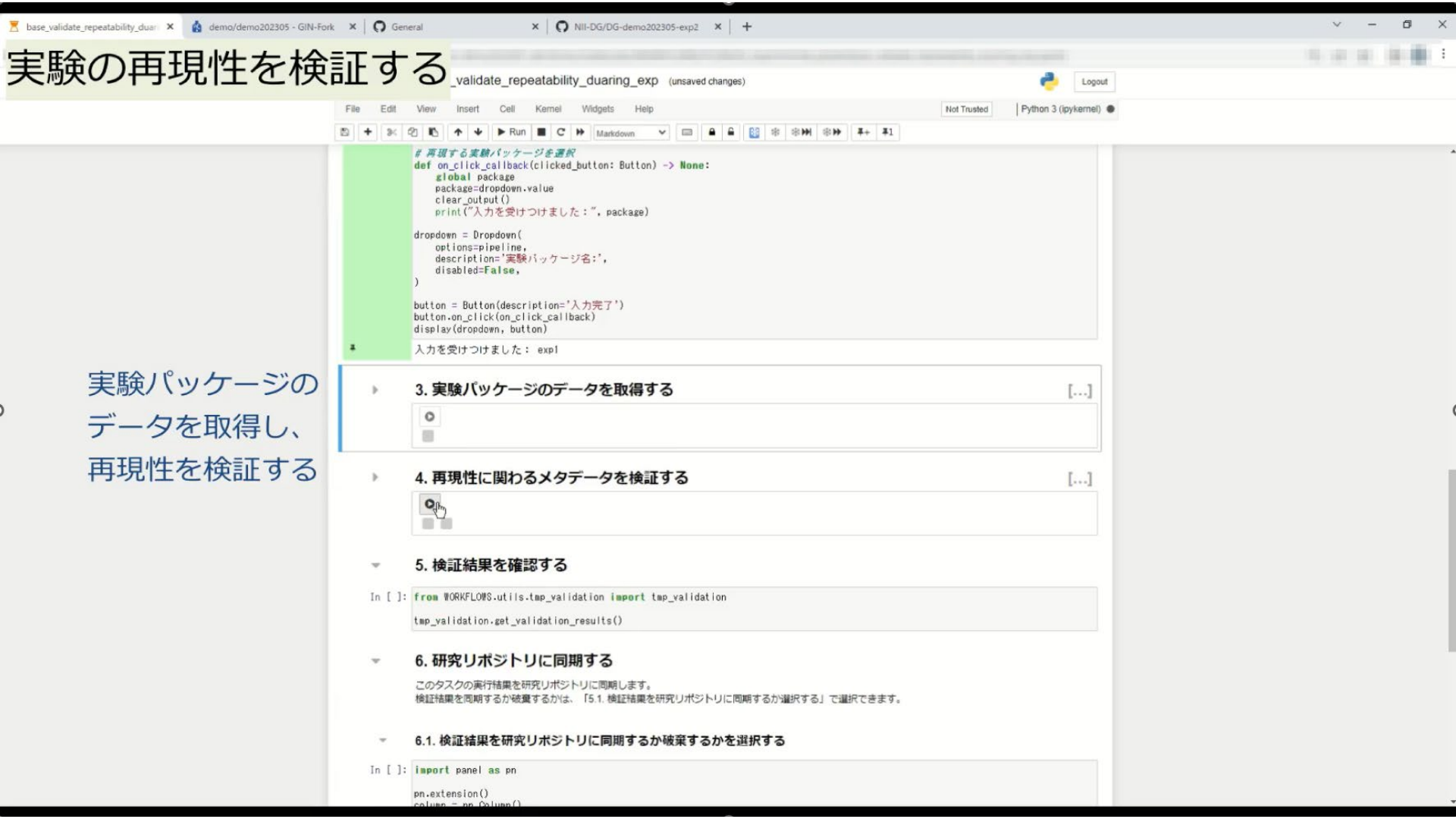
検証実施の様子（再現性検証ページ）



検証実施の様子（検証実施）

実験の再現性を検証する

実験パッケージのデータを取得し、再現性を検証する

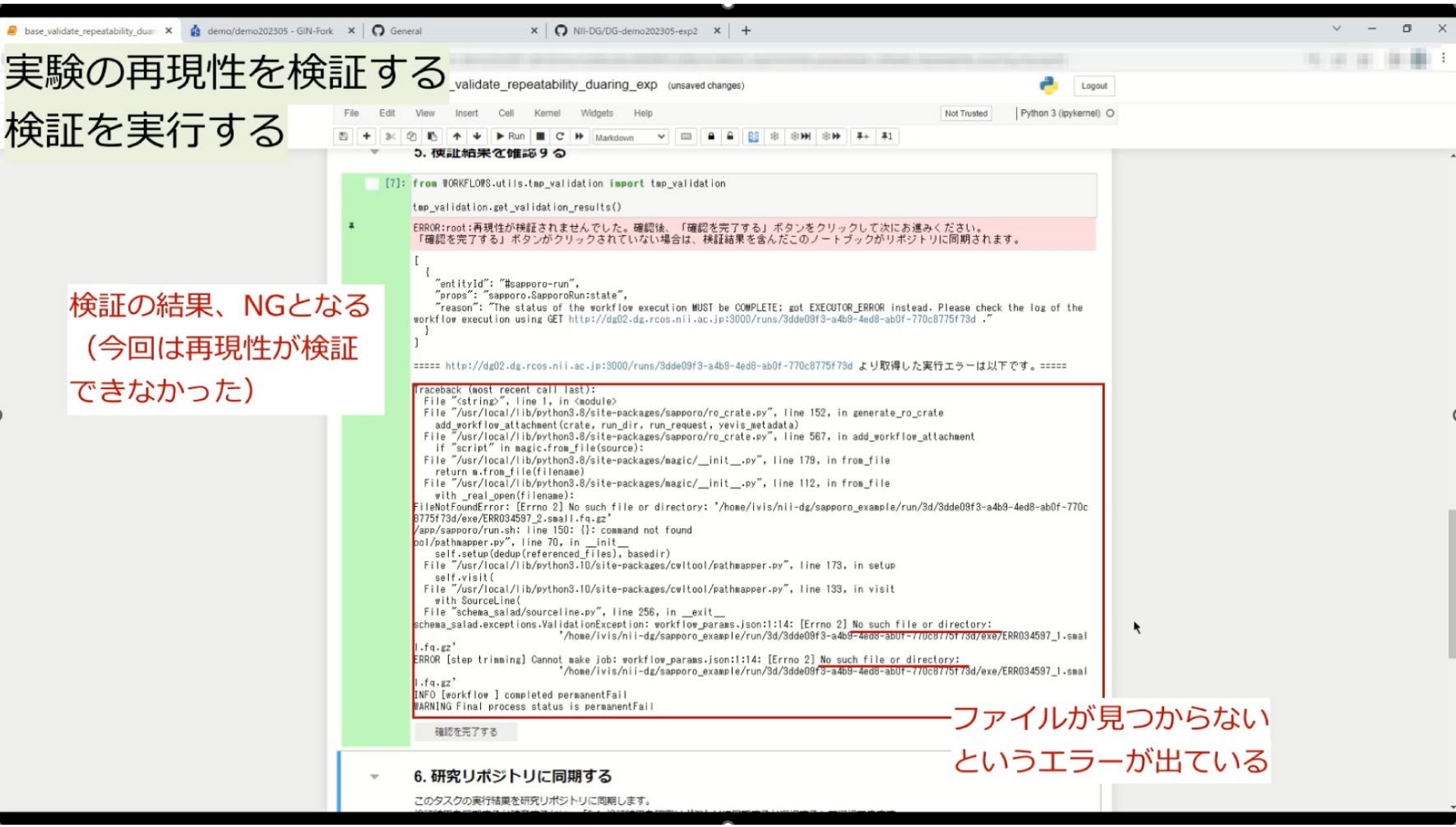


検証実施の様子（検証結果がNG）

実験の再現性を検証する

検証を実行する

検証の結果、NGとなる（今回は再現性が検証できなかった）



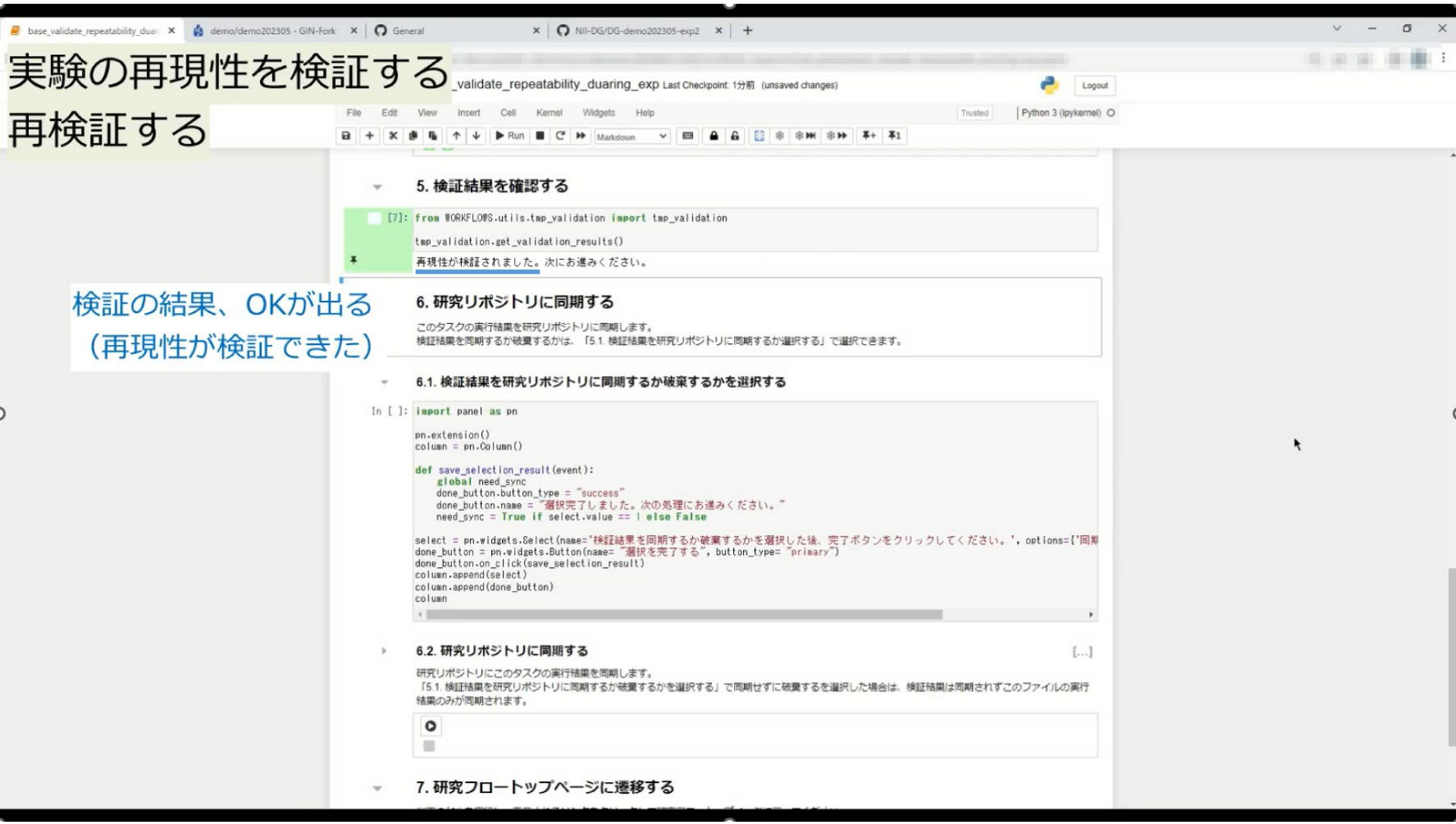
ファイルが見つからないというエラーが出ている

検証実施の様子（問題点を解決して再検証）

実験の再現性を検証する

再検証する

検証の結果、OKが出る
（再現性が検証できた）



3. 今後の進め方

取り組み方針

1. 検証機能の拡充
   - DMP との連携の強化
   - FAIR 原則[1]に基づくデータ管理状況の検証

2. リサーチフローのブラッシュアップ
   - データガバナンス機能の機能評価試験版サービス利用者からのフィードバックへの対応など

3. 他基盤との連携実施

4. 研究機関側に必要な機能の検討
   - 研究データ管理・公開ポリシーへの対応等

[1] https://force11.org/info/the-fair-data-principles/

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

NII RCOS
Research Center for Open Science and Data Platform

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の提供を予定しております。
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計画しております。

データガバナンス機能の機能評価試験版サービスの提供準備が完了次第、GakuNin RDM のサポートポータル（https://support.rdm.nii.ac.jp/）の「お知らせ」にてアナウンスいたします。

- 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
- 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