

コアファシリティにおける研究データ管理アーキテクチャの構築

Constructing Research Data Management Architecture in a Core Facility

松平拓也, 笠原禎也, 高田良宏, 濱貴幸, 長井圭治, 蟹屋敷祐介

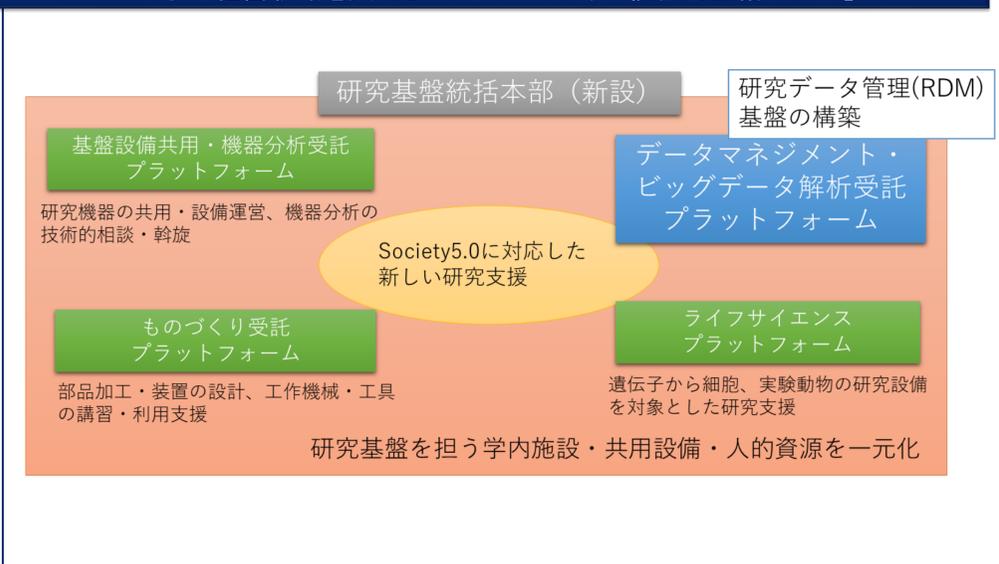
Takuya MATSUHIRA, Yoshiya KASAHARA, Yoshihiro TAKATA, Takayuki HAMA, Keiji NAGAI, Yusuke KANIYASHIKI

ku-rdm@ml.kanazawa-u.ac.jp

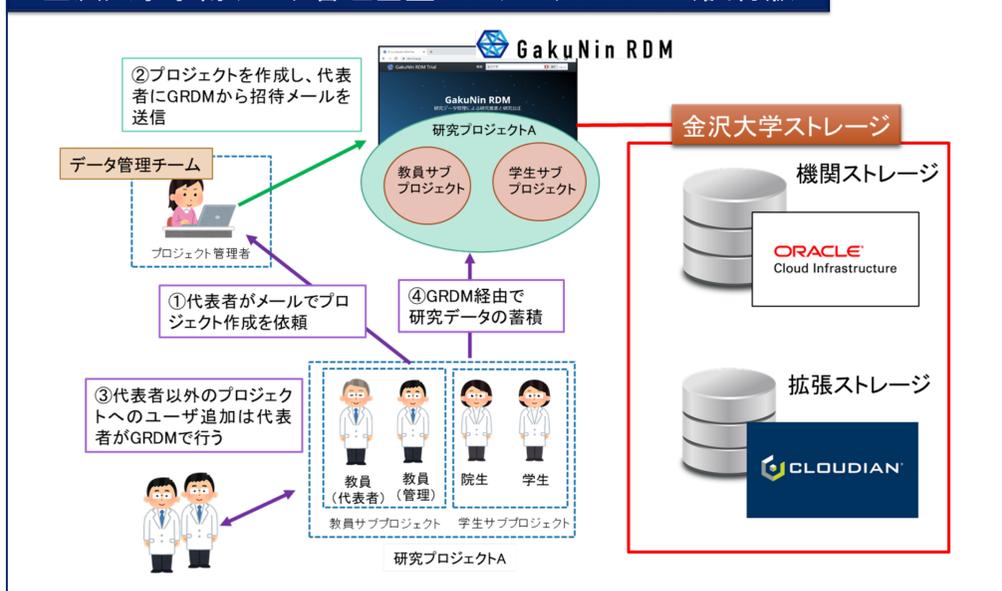
金沢大学は、令和2年度に「先端研究基盤共用促進事業(コアファシリティ構築支援プログラム)」の採択を受け、採択事業の柱の一つとして、研究データを適切に管理可能とする「金沢大学学術データ管理システム(ARCADE2)」の構築を進めている。特に本システムは「金沢大学学術データマネジメントポリシー」において、研究データに限ることなく、本学における教育に関する活動を通じて収集または生成されたデータを含む「学術データ」を管理対象と定めている。現在、本システムは試行中で、コアファシリティ環境においても適切に利用できるようにアーキテクチャの開発を進めているところである。

Keywords – 研究データ管理, GakuNin RDM, オブジェクトストレージ, 学認

先端研究基盤共用促進事業(コアファシリティ構築支援プログラム) 「大学の経営戦略を支えるコアファシリティの統合的整備モデル」



金沢大学学術データ管理基盤システム(ARCADE2)試行版



金沢大学における研究データ管理の課題

現在の研究データ管理基盤提供方式

- GakuNin RDMを利用したWebUI経由のオブジェクトストレージのみ
- ⇒ 研究データの管理は実施形態や規模により、様々なユースケースが考えられる
 - 個人の様々なデータの保管
 - 学内プロジェクトによる複数人での研究データ管理
 - コアファシリティのような、実験計測機器から生成される実験データ管理
 - 組織を横断した外部資金プロジェクトでの研究データ管理 などなど

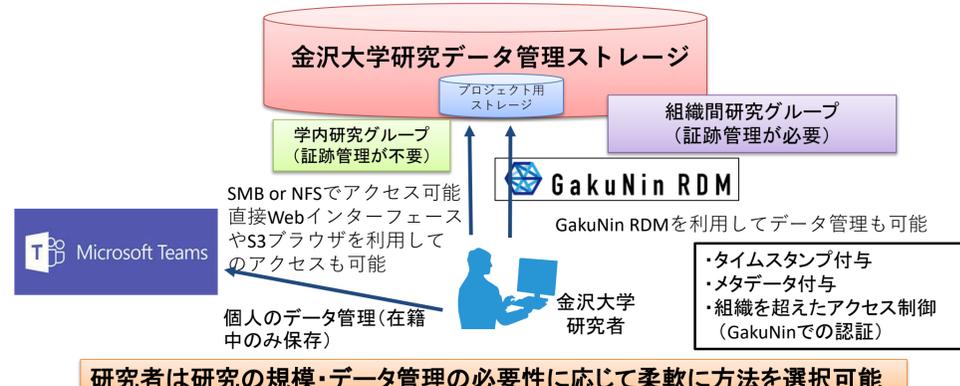
GakuNin RDMに限定することなく、様々なユースケースに対応した汎用的な研究データ管理基盤が必要

研究公正・オープンサイエンスの考えから、データを個々にローカルで保存するのではなく、オンラインストレージ上に管理しやすい環境構築を目指す

金沢大学研究データ管理の新たな考え方

- 中央に研究データ管理用のストレージを配置
- 研究プロジェクトに対してオンライン用ストレージを割り当て
 - S3だけではなく、SMB/NFSによるストレージアクセスを提供
 - GakuNin RDM以外の接続インターフェースの提供

ARCADE2の実運用版として2024年度中に供用開始予定



コアファシリティでの研究データ管理に向けて

コアファシリティに関連する研究者に対してヒアリングを実施

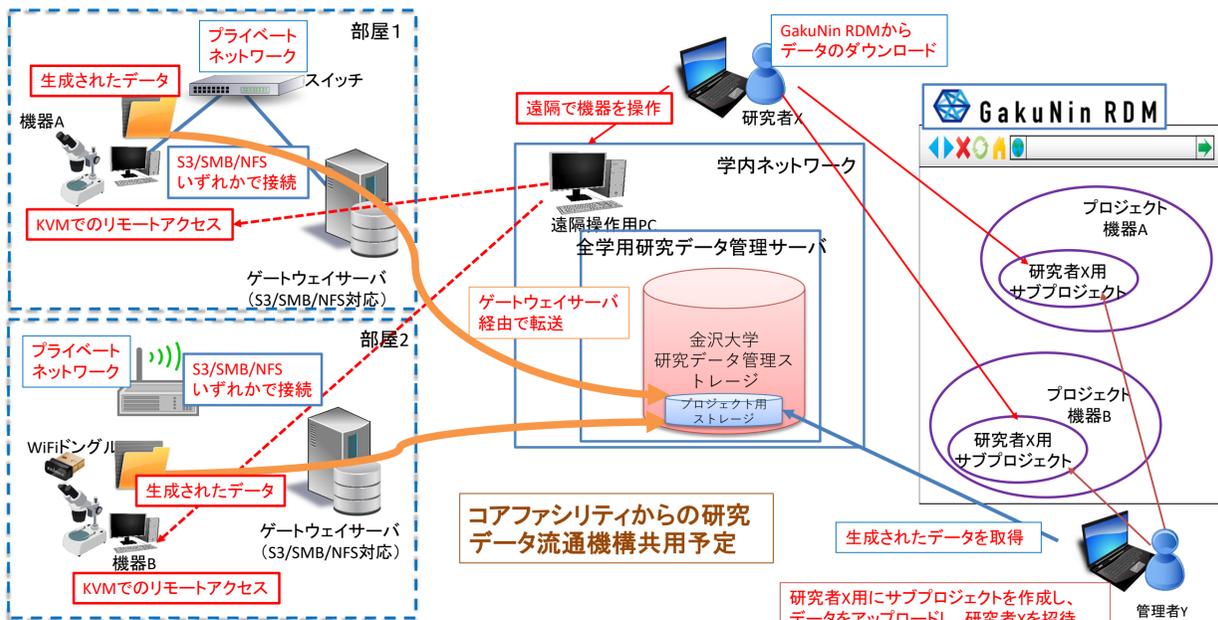
- 機器の環境について
 - 電子顕微鏡などの機器に接続されているPCはOSのバージョンが古く、ネットワーク未接続
 - データを取得する際はUSBメモリなど外部媒体経由で機器管理者のPCに移動後に利用者へ送付

⇒ USBを経由することなく、GakuNin RDM上にデータをアップロードできる機構が必要

- データ保管について
 - 分析機器から得られたデータはそれぞれの接続PC内に保存
 - バックアップにおける不安がある

⇒ 得られたデータを自動でGakuNin RDMに保存できる機構が必要

各機器から生成されたデータを自動的にオブジェクトストレージへ保存し、GakuNin RDM上から管理可能とするアーキテクチャを開発中



大容量研究データの扱いについて

研究データ管理におけるMinIO自動階層化の活用

- 大規模な予算の研究では、1回の実験で数百GBの生データが発生
- 生データはExcelや他の形式に変換され、変換後は参照頻度が低下
- 各研究者が数TB~数十TB規模のデータを所有し、長期間の保存が必要

MinIOの自動階層化機能を利用して、データのライフサイクルを効率的に管理

MinIOは、大規模なデータ管理に適したオープンソースのオブジェクトストレージシステム。S3互換API備え、効率的にデータを保存・共有でき、研究データのアーカイブや分散保存に最適。

MinIOの自動階層化を使った具体例

- 2つのストレージ用意
 - 高性能ストレージのMinIO: 初期段階での生データ保存に使用
 - 低性能・大容量ストレージのMinIO: 長期間保存が必要だが、参照頻度が低いデータ向け
- 自動データ移行
 - データは最初に高性能ストレージのMinIOに保存
 - ライフサイクルポリシー(例: 2年間アクセスなし)を設定し、古いデータは低性能MinIOに自動移行

GakuNin RDMは自動階層化に未対応なため、GakuNin RDMのWebインターフェースで自動階層化を管理可能な機能を新開発