

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, 蟹屋敷祐介

Takuya MATSUHIRA, Yoshiya KASAHARA, Yoshihiro TAKATA, Takayuki HAMA, 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

研究データ管理の背景

1. 研究公正 (Research Integrity)

- 文部科学省「研究活動における不正行為への対応等に関するガイドライン」
- 日本学術会議「科学研究における健全性の向上について」
⇒ 研究機関は研究者に対して、当該論文等の発表後原則10年間研究データを保存することを義務付ける規定の整備・運用を要求

2. 研究推進 (Open Science)

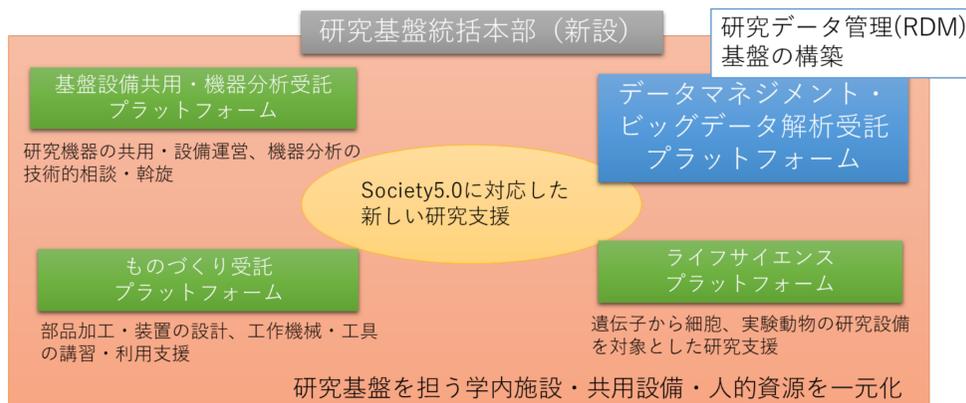
各種研究データを誰でも利用できるように一般に公開し、研究の効率化を図る動き

- 内閣府「統合イノベーション戦略2023」

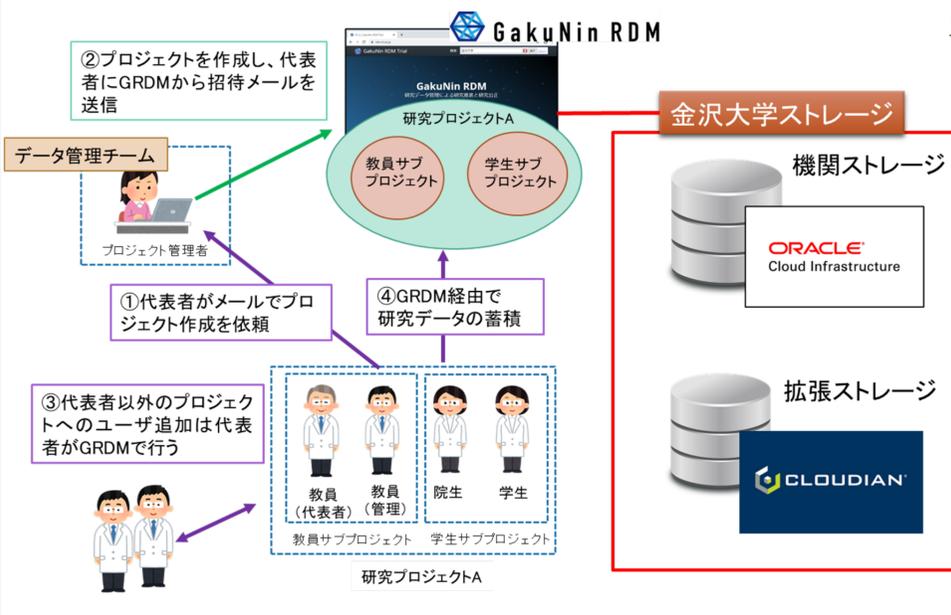
RDM (Research Data Management)

ある研究プロジェクトにおいて使用された、または生成された研究データを、明示的に組織化・構造化した上で保管・管理すること

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)



■ コアファシリティ内に用途を限定するのではなく、全学での利用を想定した汎用的な研究データ管理(RDM)基盤を構築

- 研究プロジェクトの多くは様々な研究機関の研究者が集まって実施
⇒ 本学で全ての「ITインフラ」を調達・運用することは、負担が大きく非効率的
- RDM基盤に求められる条件は日々変化
⇒ 常に動向を把握し、毎回システムに反映を行うのは非常に困難

GakuNin RDMの利用

GakuNin RDM

国立情報学研究所(NII)が運用する異なる組織の研究者間でクラウドに研究データの管理・保存が可能な研究データ管理サービス

- 金沢大学学術データマネジメントポリシーの制定
各研究域から1名推薦、研究推進部、図書館など、全関係部署から構成される「研究データマネジメントポリシー策定WG」を組織し、ポリシーを策定 (令和4年3月11日制定)
- ポリシーの対象データ範囲
研究だけではなく、教育に関する活動を通じて収集又は生成されたデータ「学術データ」を範囲

コアファシリティでのARCADE2利用に向けて

■ 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

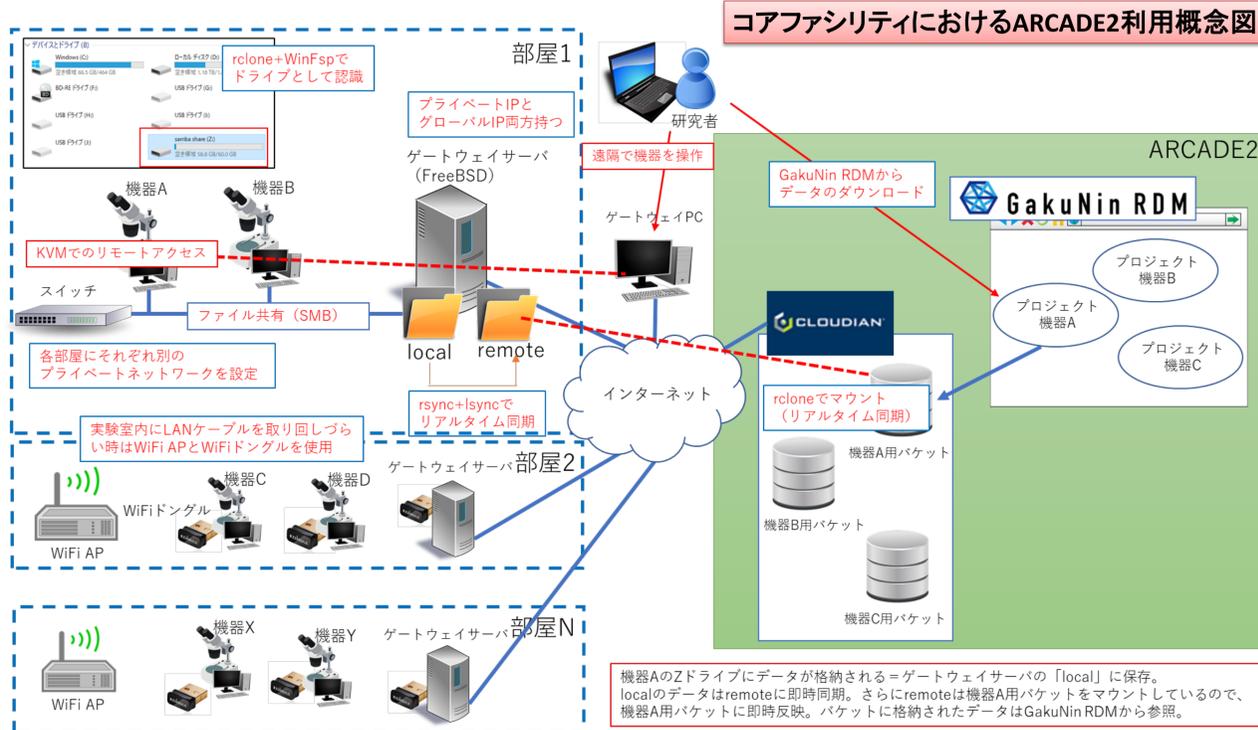
⇒ USBを経由することなく、GakuNin RDM上にデータをアップロードできる機構が必要

・ 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ためGakuNin RDM上に保存したいが、機器がGakuNin RDMにログインできる仕組みが確立されていない

⇒ 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ARCADE2が持つオブジェクトストレージへ保存し、データをGakuNin RDM上から管理可能とするアーキテクチャを開発中



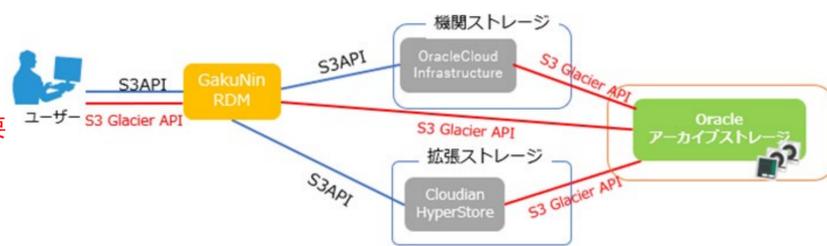
大容量研究データの扱いについて

大きい予算の研究の場合、1回の実験で数百GB程度の生データが発生
生データはexcelなどのデータに変換後は参照頻度は高くないが、保存が必要(数TB~数十TB所有/1人当たり)
⇒ 参照頻度の低い生データはOracle Cloud Infrastructureの「アーカイブストレージ」に保存できるようにする必要

■ アーカイブストレージ

- 格納されたデータを取得するのに1時間程度を要するもの、オブジェクトストレージの10分の1の価格
- GakuNin RDMでアーカイブストレージにデータを格納するためにはAmazon S3 Glacier APIを使う必要がある

GakuNin RDMはこのAPIに未対応のため、Amazon S3 Glacier APIを利用可能とし、研究者は状況に応じてアーカイブストレージにデータを格納し、必要の際はオブジェクトストレージにリストアップ可能な機能を新開発



- 研究者自身でオブジェクトストレージからアーカイブエリアへの移動を可能とする
- 一定期間アクセスのないデータは自動でアーカイブストレージに移動される
- データがオブジェクトストレージ/アーカイブストレージのどちらにあるか分かるようにアイコンで示す