

第1回北陸地区 学術データ基盤セミナー

～コアファシリティ連携から研究データエコシステム構築を目指して～
(兼「AI等の活用を推進する研究データエコシステム構築事業」成果報告会)

2024/2/20

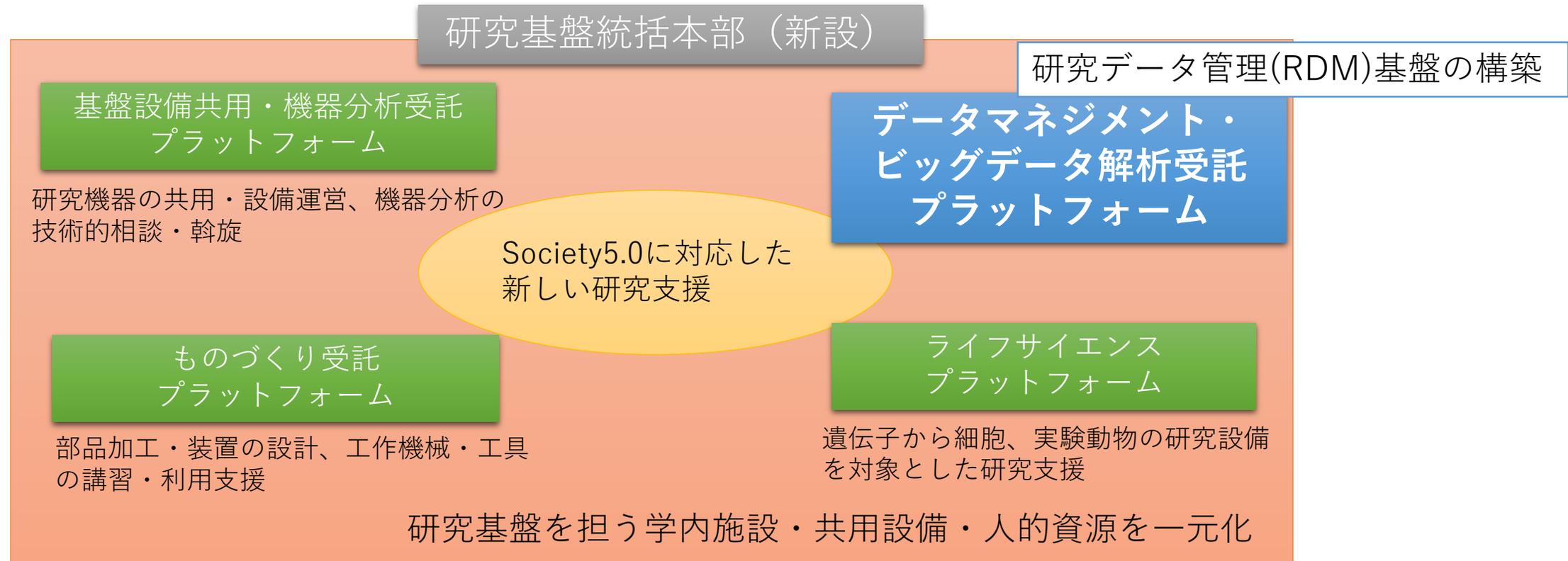
金沢大学における研究データ管理基盤取り組み状況について

金沢大学 学術メディア創成センター RDM構築WG

○高田良宏、松平拓也、笠原禎也、濱貴幸、蟹屋敷祐介

研究データ管理基盤システム構築の背景

- 令和2(2020)年度「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）」に採択
 - 研究基盤統括本部を中心とした研究設備の全学共用体制への再編



※令和2年度「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）」の採択機関の決定について
https://www.mext.go.jp/b_menu/boshu/detail/mext_00066.html

これまでの経緯（主に基盤系） その2

- 2013/03： **学術組織間デジタル資料分散共有システム（ARCADE）**
⇒ GakuNinのSPとして公開：金沢大学における初代のRDM基盤？
- 2018/05： **NII研究データ管理基盤『GakuNin RDM』の機能評価試験に参加**
- 2019/??：先端研究基盤共用促進事業（研究機器相互利用ネットワーク導入実証プログラム（SHARE））で『GakuNin RDM』と連携を言及 ⇒ 不採択

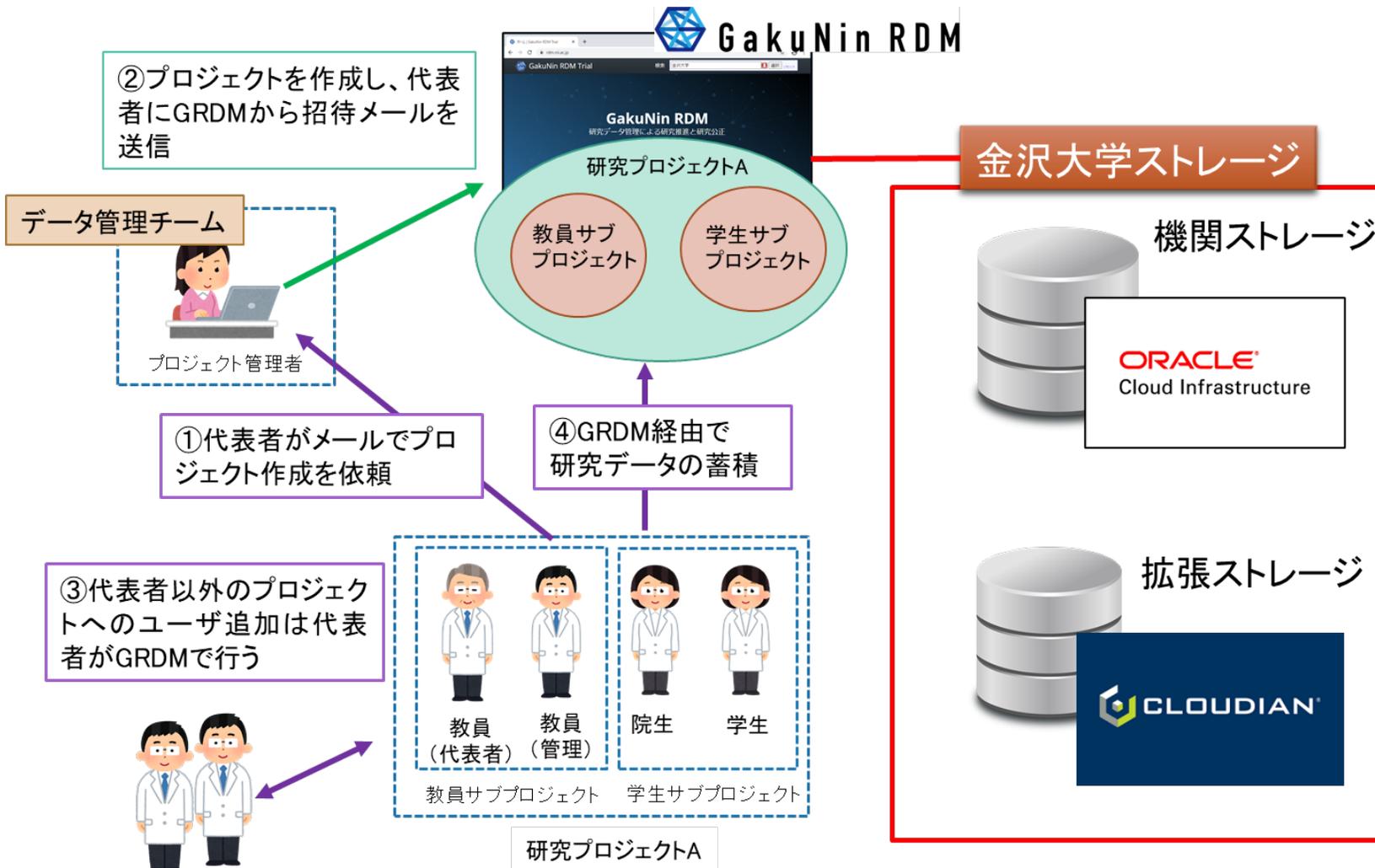
- 2020/04： **先端研究基盤共用促進事業（コアファシリティ構築支援プログラム）**
⇒ **データマネジメント基盤の構築を柱に挙げる ⇒ 採択**
- 2020/06： **RDM構築に関する調査・検討を開始**
⇒ 総合技術部と学術メディア創成センターの有志で
- 2020/12： **RDM構築WG発足(学術メディア創成センター内の組織)**
⇒ 総合技術部と学術メディア創成センターの双方からメンバーで構成
- 2020/12： **金沢大学学術データ管理システム（ARCADE2）試行版が完成**
⇒ 関係者で試行開始

これまでの経緯（主に基盤系） その2

- 2021/08：研究データポリシー策定に向けて検討が始まる
- 2021/10：研究データマネジメントポリシー策定WG発足
- 2022/02：システム試行説明会（総合技術部）
- 2022/03：システム試行説明会（理工系教員）
- 2022/03：金沢大学学術データマネジメントポリシー制定
- 2022/03：ARCADE2を全学で試行開始
- 2023/01：システム試行説明会（コアファシリティ関係者）
 - ⇒併せてコアファシリティ関係者にヒアリングを実施
- 2023/08：「A I等の活用を推進する研究データエコシステム構築事業」の「研究データ管理スタートアップ支援事業」開始
- 2024/04以降：ARCADE2の実運用版供用開始予定
 - ：コアファシリティからの研究データ流通機構供用開始予定

金沢大学学術データ管理基盤システム（ARCADE2） 試行版

- 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



■ 機関ストレージ

機関のシステム管理者が、学内提供のストレージをデフォルトストレージとして1種類指定可能
(本学全研究者が利用可・設定不要)

■ 拡張ストレージ

利用者が、GakuNin RDMが提供する接続インターフェースに該当するストレージを指定可能
(本学申請者のみ利用可・設定必要)

■ 金沢大学ストレージ導入指針

外部クラウドと学内ストレージのハイブリッド

研究データは機微な情報のため、一部の研究者は外部クラウドでのデータ保管を躊躇する可能性を考慮

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規模や形態により、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、実験観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



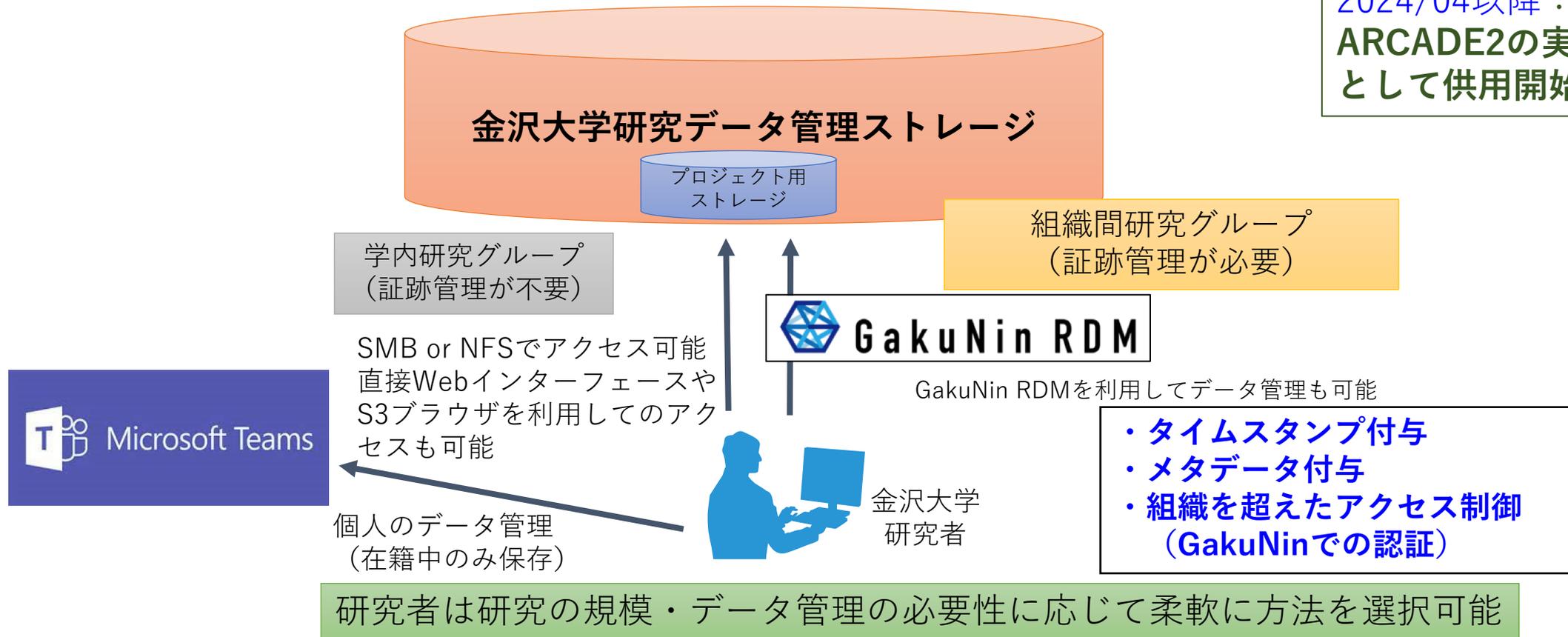
GakuNin RDMに限定することなく、様々なユースケースに対応した汎用的な研究データ管理基盤が必要

- 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

2024/04以降：
ARCADE2の実運用版
として供用開始予定

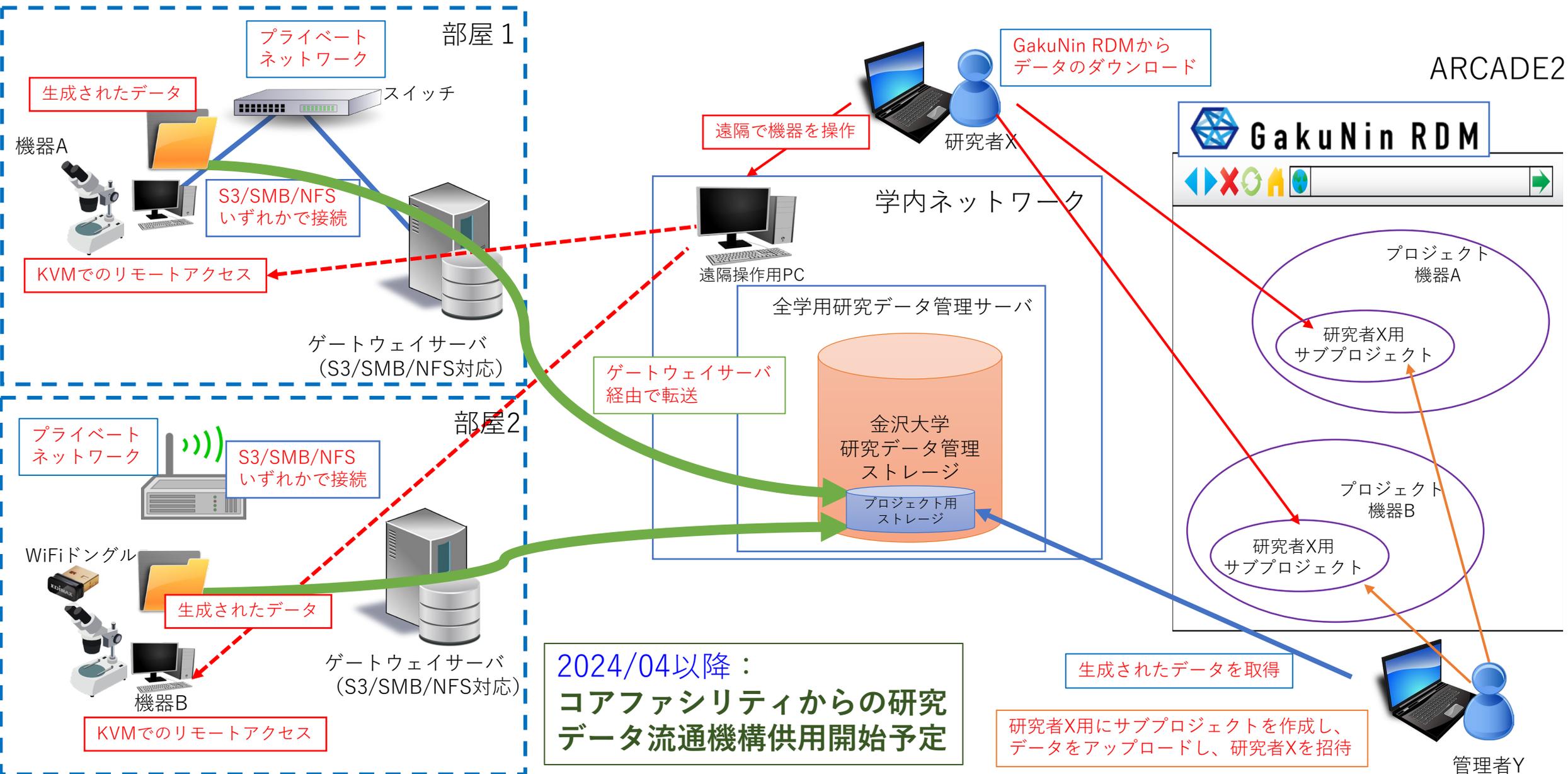


- タイムスタンプ付与
- メタデータ付与
- 組織を超えたアクセス制御 (GakuNinでの認証)

コアファシリティでの研究データ管理基盤利用について

- 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に移動後に利用者に送付
⇒ **USBを経由することなく、データを配布できる機構が必要**
 - 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
 - バックアップなどにおける不安がある
⇒ **得られたデータを自動でバックアップできる機構が必要**

コアファシリティにおける検証環境

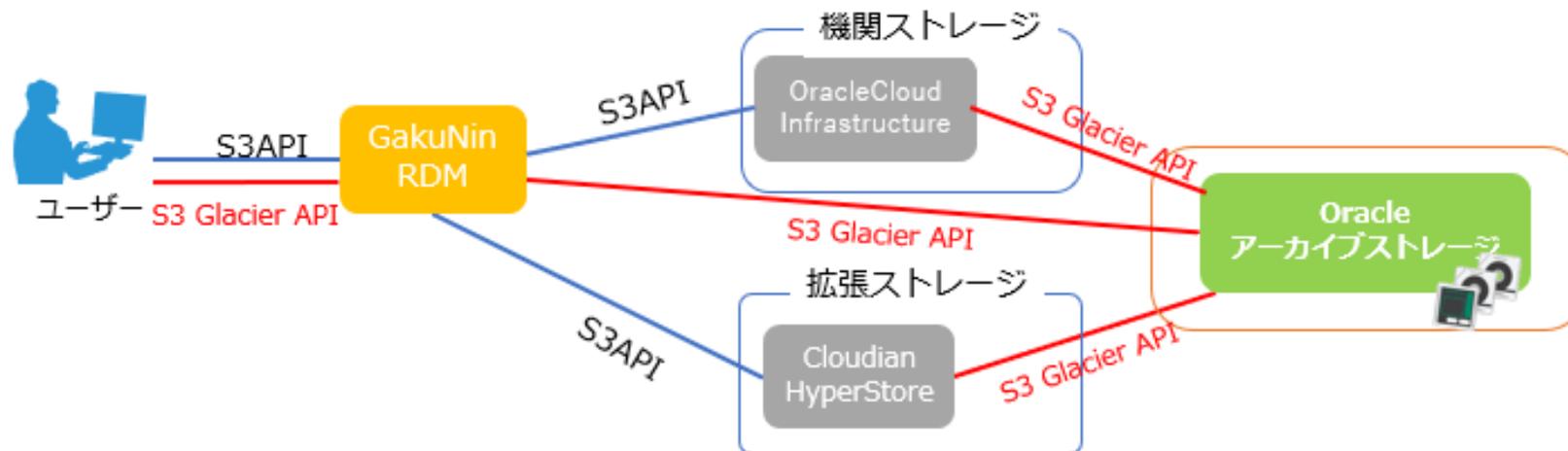


大容量研究データの扱いについて（今後に向けて）

- 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



GakuNin RDMはこのAPIに未対応なため、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、必要な際はオブジェクトストレージにリストア可能な機能を新開発