

## 第3回 SPARC Japan セミナー2020

「初めての研究データ」

# 全国規模な研究データ管理サービスの 提供者視点でのベストプラクティス

込山 悠介

(国立情報学研究所)

### 講演要旨



研究データを研究者が独自に管理すると、貴重な研究データが研究室で紛失・消失、また気密性の高い研究データが外部に流失してしまう恐れがあります。また、研究者の所属機関が組織的な研究データ管理を行っていない場合は、研究不正の温床になる可能性があります。また、最近では競争的資金による研究プロジェクトでは、データ管理計画に基づく研究データ管理が要求されていることも増えています。国立情報学研究所では2021年から国内の学術機関に向けて、研究データ管理サービス GakuNin RDM の本運用を開始しました。本発表では、全国規模の研究データ管理サービスの提供者の目線から、実際の研究データ管理の事例を踏まえて、現実的なベストプラクティスについて述べます。

### 込山 悠介

国立情報学研究所コンテンツ科学研究系助教。博士（農学）。2014年3月、東京大学大学院農学生命科学研究科博士課程を修了。日本学術振興会特別研究員（DC2）として生物情報工学の研究に従事。2014年4月、東京大学医科学研究所ヒトゲノム解析センター特任研究員としてスーパーコンピュータ運営や生命科学データベース開発のプロジェクトに従事。2016年4月より現職。2017年4月より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の発足に伴い、研究データ管理サービス GakuNin RDM の研究開発、運営および国際連携に従事。



### 研究データ管理（RDM）サービスとは？

国立情報学研究所（NII）の研究データ管理サービスの背景についてお話しする前に、まず研究データ管理についてお話しします。研究データ管理サービスは、英語では Research Data Management の頭文字を取って RDM サービスといいます。RDM は、ある研究プロジェクトにおいて使用された、または生成された情報をどのように組織化・構造化していくかということを指す言葉です。

そして、「RDM サービス」にも広義と狭義とがあります。広義の RDM サービスに当たるのは、本日のメインテーマにもなっているような、研究データのライ

フサイクル（研究データの計画・管理・共有・解析・公開・検索）の全てのプロセスに関わる業務をサポートしていくための IT インフラ、あるいはその研究支援体制（意思決定、予算、人的リソースの配置、教育プログラム）の提供です。本講演内においては狭義の RDM サービスを、研究者が発表前の研究データを管理・共有することに特化した IT インフラとして紹介させていただきます。

### 国立情報学研究所（NII）の学術情報基盤の全体像

今回は、国立情報学研究所をご存じない方もいらっ

しゃいますので、弊所の学術基盤サービスを簡単に紹介します（図 1）。参加者の皆様は図書館に所属している方が多いと思いますので、日本語論文検索の CiNii や機関リポジトリのサービスである JAIRO Cloud 等をご存じだと思いますが、その他にも NII の事業サービスとしては、全国の都道府県に張り巡らされている学術情報用ネットワークの SINET や、学術認証フェデレーションといて、所属機関の全学共通のログインシステムで、さまざまな Web サービスヘューザーがログインして使っていただける認証フェデレーションを提供しています。他にも無線 LAN のサービスである eduroam 等も提供しています。また、RDM サービスとも関連しますが、クラウドサービスに関する導入支援等を行うサービスも行っています。

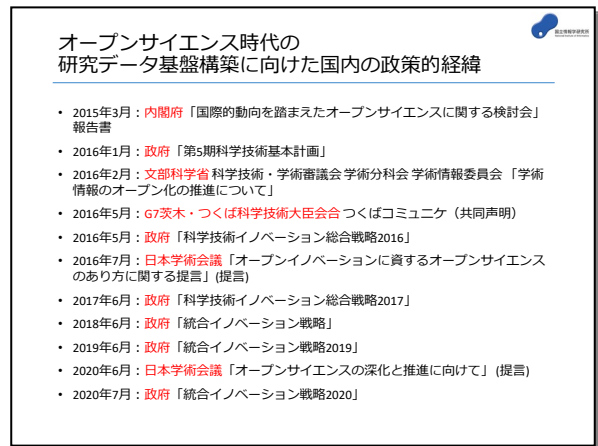
### RDM サービス提供の背景

NII が RDM サービスを作ってきた背景として、2015 年ごろから政府のオープンサイエンス政策の流れがあり、その中で研究データ基盤を構築していくことになりました（図 2）。

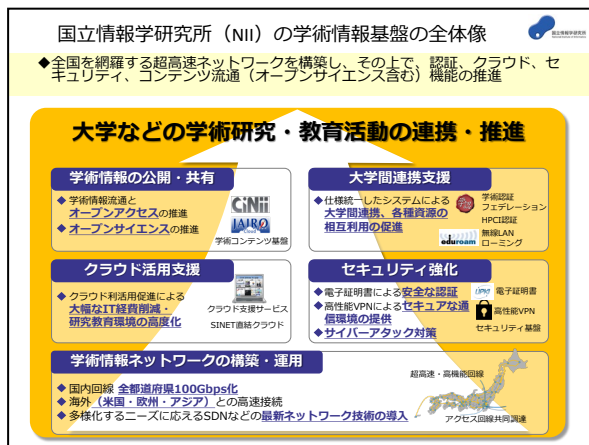
直近では、「統合イノベーション戦略 2019」の中に、「NII を中心とした研究データ基盤・リポジトリの整備、研究データの管理・利活用方針」といったことが盛り込まれています。また、「統合イノベーション戦略 2020」の中には、NII の研究データ基盤システムを用いて先進的な研究データ、研究マネジメントを推進するということが書かれています。それがさらに進み、

内閣府ムーンショット型研究開発制度の運用・評価指針の中にも、NII の研究データ基盤 NII Research Data Cloud (NII RDC) 等を使って先進的な研究データをマネジメントしていく事を推奨するということが盛り込まれています。NII の GakuNin RDM は、ムーンショット型研究開発制度で先行利用の形で、お使いいただいていることもあり、一部、研究代表者の先生方への説明会等も始まっています。

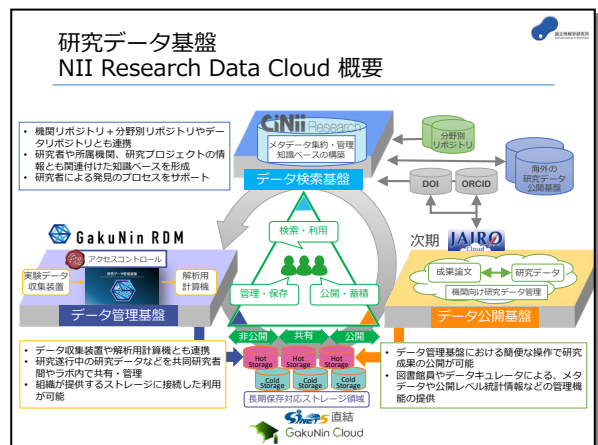
GakuNin RDM は、少し広い視点で見ると検索・公開・管理の3つの基盤からなる NII RDC のうちの一部となっています（図 3）。データの公開基盤である JAIRO Cloud と、データを検索するための基盤である CiNii Research と連動する形で、公開前の研究過程のデータを取り扱う管理基盤 GakuNin RDM を新たに提供していく位置づけになります。



(図 2)



(図 1)



(図 3)

## 研究データ管理サービス GakuNin RDM

既に GakuNin RDM を実証実験でご利用いただいている機関の方もいらっしゃいますが、改めて説明させていただきます（図 4）。研究データライフサイクルにおける研究計画の段階から、データを収集し、データ分析等を行い、論文を書いてパブリッシュする研究者の日常を支える基盤で、Web ブラウザ上で研究データを保存・管理することができるようになっています。2021 年 4 月からは 24 時間 365 日の運用体制で全国の学術機関に提供していく予定になっています。2021 年 2 月現在、夜間帯はありませんが 365 日の有人の運用体制を取っています。

GakuNin RDM については、研究を推進していくための機能と、研究公正（研究不正の防止）という二つの大きな柱で研究データ基盤を構築してきました（図 5）。ターゲットはもちろん研究者が主役ですが、それ

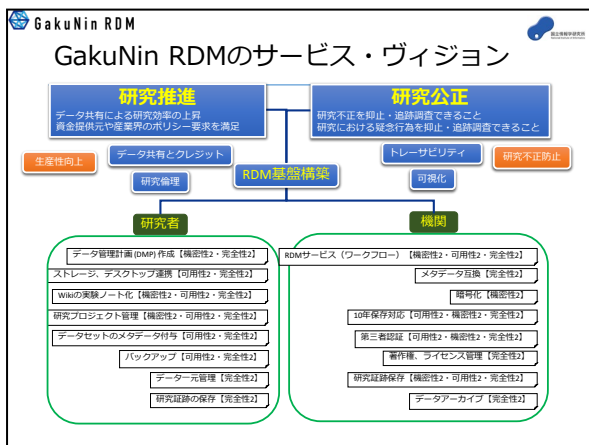
と同じくらい研究機関にとっても重要な機能を盛り込んできました。開発黎明期は、米国の Center for Open Science が開発した Open Science Framework (OSF) をベースに国内の需要を盛り込んで機能拡張を行っていました。初期は本当に国内では私のコンピューターの中だけで動いていたようなときもありましたが、2017 年 4 月にはオープンサイエンス基盤研究センターが発足し、本格的な開発がスタートしました。2019 年 4 月から約 1 年半の実証実験を全国で実施し、この 2021 年 2 月 15 日から本運用サービスの提供を開始しました。

## GakuNin RDM の機能

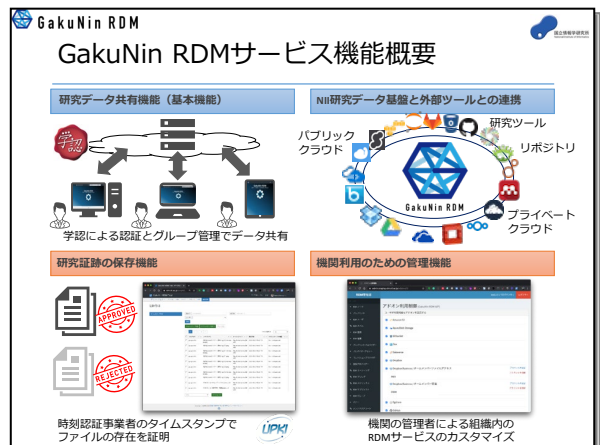
GakuNin RDM では、所属機関が学術認証フェデレーション（学認）に参加していれば、いつも利用されているアカウントでログインすると、利用機関同士であれば、すぐに機関を超えてデータの共有（制限共有、限定共有）を開始することができます（図 6）。もちろん研究者 1 人のワークスペースとしてもご利用いただけます。そういった基本的なデータ保存という機能があります。しかし、それだけでは研究データ管理のサービスとは言えないので、研究者が普段使うソースコードを保存するリポジトリや、引用文献を入れるような追加機能、さらに、研究プロジェクト等より大容量のクラウドストレージを連携できる機能を持っています。他にも基本機能として、システム中に保存したファイルに対して研究証跡の保存機能が働き、NII



(図 4)



(図 5)



(図 6)

以外の時刻認証事業者（第三者機関）のタイムスタンプによって特定時刻のファイル存在を証明するという機能が動作しています。また、機関のシステム管理者に向けて、機関内での RDM サービス提供を実施しやすくするための機能を充実させるような開発もおこなってきました。

GakuNin RDM のログイン画面に行くとプルダウンメニューに参加機関のリストが並んでいます。自機関の名称を選択すると、自機関で普段使っている認証システムを経由し、ダッシュボードにログインできるという仕組みになっています。また、研究室内や共同研究者間での階層的な研究プロジェクトの研究データ管理も行えるようになっていきます。階層自体は無制限に作る事ができるので、うまく活用すると、より大規模な研究プロジェクトでも使っていただくと考えています。

さらに、図7のように、さまざまな研究ツールや外部のクラウドストレージにも接続することができます。右側の図は極端な例ですが、一つの研究プロジェクトで十数種のクラウドを連携させて使うといったことも可能です。

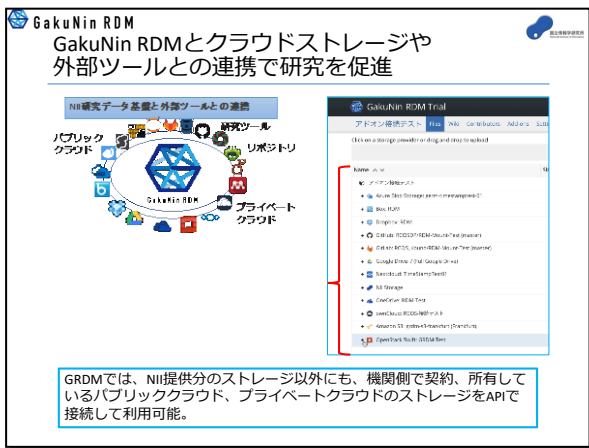
それから、研究データの証跡管理を行うことができます。一度システム中にファイルをアップロードして、もしシステムの裏側でファイルが削除されたり、書き換えられたりすることがあった場合は、それを検出する機構が入っています。それにより、プロジェクトを管理する研究主宰者は、プロジェクトに所属する研究

者の研究データへの不正操作を検知できることで、研究不正を抑止することができます。

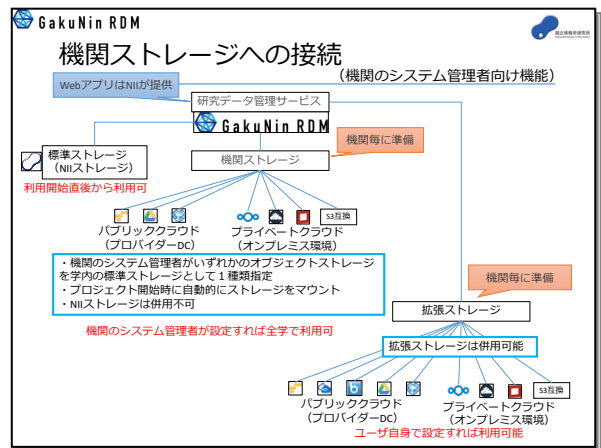
機関のシステム管理者へ向けて開発した機能として、GakuNin RDM には機関がお持ちのストレージを接続して利用していただける機能もあります（図8）。機関がエンタープライズ契約している商業用のクラウドサービスや、機関の中にストレージサーバーをお持ちであれば、GakuNin RDM へ接続して使っていただくことができます。もちろん、GakuNin RDM をご契約いただくと、標準ストレージ（NII ストレージ）として、利用開始直後から使用可能な領域も一定容量あります。加えて、ストレージを準備していただく必要がありますが、機関ごとまたは、プロジェクトあるいは部局単位で切り分けてストレージを設定したいという要望にもお応えできる形になっています。

ここで簡単に、RDM サービスのおすすめの構成を大学の規模別に紹介します（図9）。実証実験の段階では大規模大学と中規模大学の事例が多かったのですが、大規模大学は、想定される取り扱い研究データが大きい分野・部局があり、あるいは部局ごとの自治があり、全学的なシステム統一が困難、あるいは外部との大規模な共同研究のプロジェクトが多いという場合には、標準的に準備されているストレージと拡張ストレージの併用を検討されているところが多かったです。

中規模な大学では、全部局で統一した RDM サービスが導入可能、あるいは地域の研究所や企業との共同研究プロジェクトがあるということで、学内にサーバ



(図7)



(図8)

一があるストレージを接続して、標準ストレージと置換えて使っている事例がみられました。それだけでは不十分ということで、さらに外部機関とも連携するためのストレージを拡張ストレージへ割り当てているという事例もあります。

小規模な大学では、情報科の教員や情報システム部門の職員が学内 IT システムの面倒を見ている場合があります。そのような場合に、標準ストレージだけを使うのであれば、ご契約いただいた直後から全学的な RDM サービスをご利用いただけます。

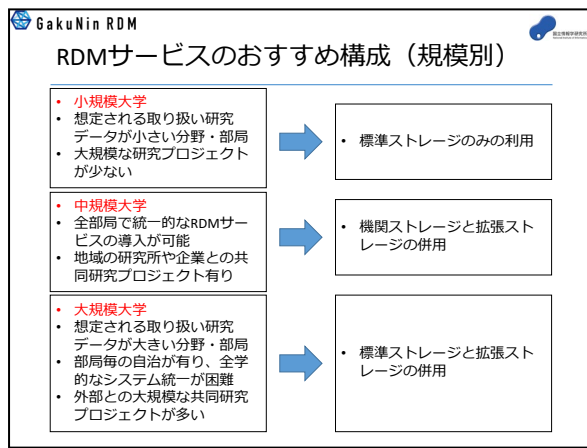
機関ストレージの具体的な例をお示しします。図 10 は、GakuNin RDM の裏側に、機関で利用する契約をしている Dropbox Business のストレージを割り当てた図ですが、例えば既に学内で使われているデータ共有のストレージがある場合は、そのデスクトップクライアントと RDM をうまく連動させて、なおかつ研究

証跡機能も働くということを一部のクラウドストレージでは実現しています。これを使うと、エンドユーザーにとっては、何も考えずに普段の Windows や Mac でファイル管理を行えば、それが自動的にクラウドに同期されると同時に研究データ管理が行われます。

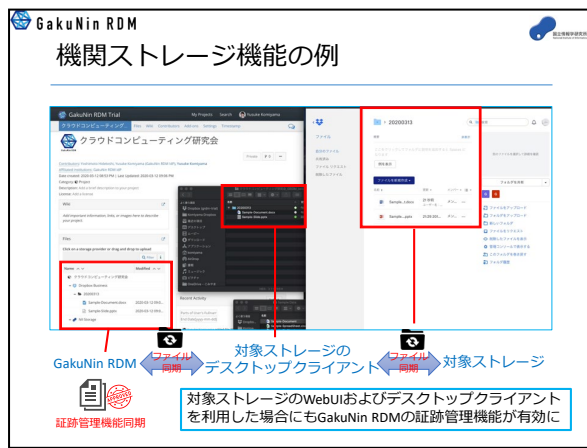
また、システム中に保存されたデータで、かつ自分にアクセス権があるデータのみを検索も実現しています(図 11)。システム中に入ったファイルに関して、ファイル名やアノテーション、ディスカッションのコメント、ID を検索できる機能も拡張しています。

また、留学生や海外の研究者も含めて利用したいということで、ユーザインタフェースやメッセージは日本語と英語に対応しています。Web ブラウザの言語設定を切り替えれば、それだけで言語が切り替わるという仕様になっています。

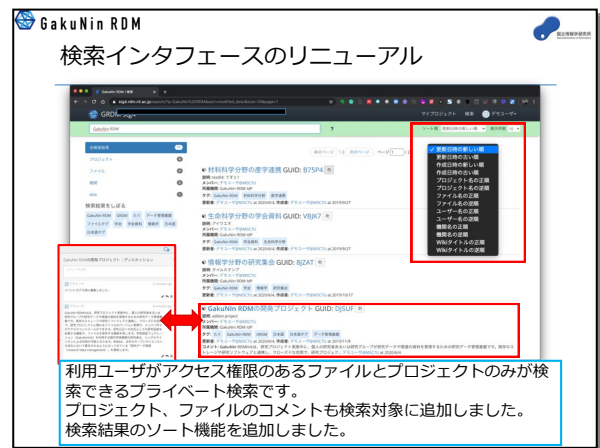
GakuNin RDM は、2021 年 2 月 15 日から本運用でサービスを提供しています(図 12)。プレスリリースと同時に利用申請の Web サイトもオープンしました。機関のシステム導入の担当者様にお申し込みいただくことになっています。現在既にお申し込みの受付が始まっています。これまで先行利用していただいていた機関につきましては、データ移行することなくそのままご利用いただけます。冒頭でも申し上げましたが、4 月から 24 時間 365 日のフルスペックでの有人オペレーションを始めていくということで、2020 年度内はその準備として、夜間を除く 365 日の有人オペレーション体制を取っています。



(図 9)



(図 10)



(図 11)

当面は、アプリケーションの利用料は無料としています。また、大学からご要望が多い NII クラウド導入支援サービスのクラウドチェックリストにも、サービス開始と同時に掲載させていただきました。ポリシー面の詳細につきましてはクラウドチェックリストを参照していただけます。2021 年度はサービス運営として ISO27001 (ISMS: 情報セキュリティマネジメントシステム) の認証取得を進めていく予定です。

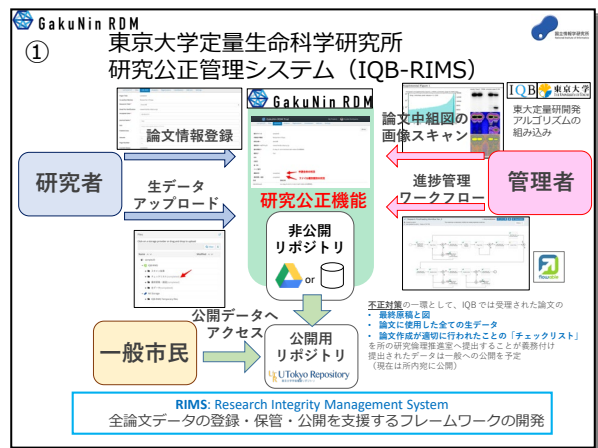
### GakuNin RDM の利用事例

ここから GakuNin RDM のユースケースの説明に移ります。昨年度、この SPARC Japan セミナーに登壇させていただいた際、GakuNin RDM の実証実験では利用機関数は 18 機関でしたが、現在では 22 機関に増えています。最近ではムーンショット型研究開発制度の採択大学のプロジェクトマネージャー (領域代表) からもお問い合わせが来ています。また、これまでは国立大学が中心でしたが、私立大学や NPO も参加していただいております、学術機関の規模や種別も多様化しています。

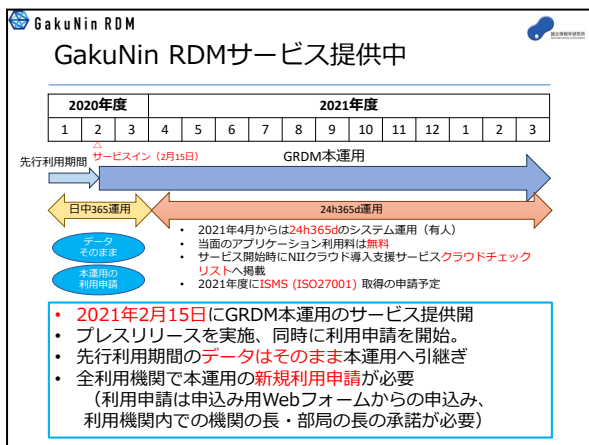
利用事例を紹介します。一つ目は、東京大学定量生命科学研究所 (IQB: Institute for Quantitative Biosciences) と共同で開発している研究公正管理システム (IQB-RIMS) です (図 13)。これは、生命科学分野の論文に含まれている画像を、部局に所属する全ての研究者の論文を対象として出版前に倫理部門でチェックしたいというユーザーシナリオがありました。これまでは、

メールアドレスのやりとりをされていて、スプレッドシートでデータを管理しリマインドもメールで行い、論文中の画像データの解析もかなりの部分が手動で行われていました。その半自動化という目的を中心に GakuNin RDM を入れていただき、機能拡張することで研究公正機能を実現されています。これには一部、業務プロセスを管理するワークフローエンジンを導入・連携しています。2019 年度の紹介時は開発段階でしたが、2020 年度は実証実験に進んでおり、今後、大学の公開用リポジトリに、全ての論文にひも付くデータを公開する機構を入れようとされています。

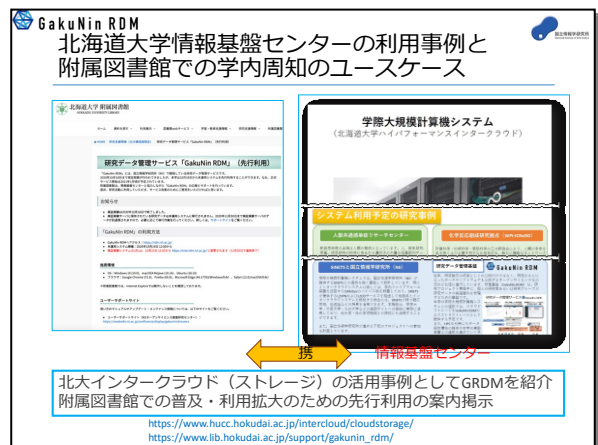
二つ目は北海道大学の事例です (図 14)。情報基盤センターのスーパーコンピュータのユーザーが利用できるストレージを GakuNin RDM に割り当てることができるといことで、情報基盤センターのパフレット等に掲載していただいております。一方で、附属図書館



(図 13)



(図 12)



(図 14)

の方では、利用方法や必要な情報等を掲載していただき、学内研究者への普及をしていただいております。こういった図書館と情報基盤センターの連携が RDM サービスの利用者を実際に広げていく上で非常に大切なことだと思っています。そして、URA・研究推進部門から出されている北大コアファシリティ構想の中で、学内に点在している実験装置や遠隔地の実験装置をリモートで操作するプロジェクトのデータの受け皿として GakuNin RDM を検討していただいております。

三つ目は名古屋大学の事例です(図 15)。学内で提供されているクラウドストレージを GakuNin RDM に連携できるということで、情報戦略部門でかなり丁寧なマニュアルを作って掲示していただいております。これも情報基盤センターと情報戦略部門との連携の事例と考えています。

また、その他の事例として、2019 年度はテキストのみの紹介でしたが、2020 年度は大型研究プロジェクトのデモムービーも作成しました(図 16)。京都大学の医師と北海道大学の数理学者が連携してデータを取得し、分析を行い、最後は医師がアノテーションするという用途で利用していただいております。

最後の事例は、大学経営統合での IT インフラ統一のユースケースです(図 17)。こちらは北海道国立大学機構が三つの大学を統合するときに、IT インフラを統合したいということで GakuNin RDM の導入をご検討いただき、学内に設置したクラウドストレージ、学内にあるサーバーのストレージと GakuNin RDM を

連携させるということで既に動き出しています。

## まとめ

以上、全国規模で提供している研究データマネジメントサービス GakuNin RDM の紹介と、その機能概要について話題提供させていただきました。2021 年 2 月 15 日から GakuNin RDM の本運用サービスが動いており、2021 年 2 月現在の利用機関は 22 機関となっています。今後も受付を行い、さらに関係各所と連携しながら、各大学で使っていただけるように進めていきたいと考えています。

- 八塚 既に幾つかのご質問を頂いています。まず、「個人情報を含む研究データのように秘匿性の高いも

(図 16)

(図 15)

(図 17)

のに関しては、GakuNin RDM で管理できますか」という質問です。

●**込山** 現状、システム側では暗号化等を行っていないので、秘匿性の高いデータを入れる場合は、ファイルをユーザー自身が暗号化して入れる必要があります。

●**八塚** コミュニティのポリシーやユーザー側の対策が必要ということですね。

●**込山** ユーザー側の機関ポリシーに従って使ってよいかということ判断していただく必要があると思います。ただ、ファイル単位の暗号化がユーザーからご要望の多い機能だということは認識しています。

●**八塚** 次に、「証跡管理機能として、リポジトリに関与する全ユーザーの編集履歴などをログファイルで取得することは可能でしょうか」という質問です。

●**込山** GakuNin RDM の証跡管理機能では、プロジェクト（ご質問ではリポジトリ）の管理者はプロジェクトに関する全ユーザーの編集履歴などを、ログファイルで取得することが可能です。また、学内の調査委員会などから指示を受けた機関のシステム管理者が、プロジェクトごとに全ユーザーの編集履歴をログファイルとして取得することも可能です。

（注：当日の回答が質問内容に沿ったものではございませんでしたので、正しい回答に修正いたしました）

●**八塚** 次の質問です。「利用している研究者が他の大学等に移った場合はどうなるのでしょうか」。

●**込山** 現状では、所属機関ごとにアカウントが作成されるので、他大学に移った場合は別人という形になります。初期の頃はこれを統合する仕組みを考えていましたが、機関を超えて研究者が移籍する場合に、研究者が勝手にデータを持って異動してしまうことはい

かがなものであるという意見もあり、現在は別のユーザーとして新たにアカウントを作ってログインしていただく形になっています。