［オープンサイエンスの動向と情報科学の役割］

④ 学術機関向け全国的な研究データ管理サービス―情報学によるオープンサイエンスの実現に向けて―

込山悠介 国立情報学研究所

オープンサイエンスと全国的な研究データ統合基盤

国立情報学研究所（NII：National Institute of Informatics）では2017年より，全国の学術機関（大学および公的機関の研究所）でのオープンサイエンス推進を支援するための研究データ統合基盤NII Research Data Cloud（RDC）の開発を行っている．オープンサイエンスとは本特集の趣旨である，これまで研究室内で一子相伝的であった研究データや研究技法を公開し，利活用しやすい学術情報流通の枠組みを再整備することで，学術研究を効率的に進めようというパラダイムである．この潮流を支援する基盤開発のためにNIIではオープンサイエンス基盤研究センター（Research Center for Open Science and Data Platform，英語略称はRCOSの4文字）が設立され，研究データ基盤の研究開発や国際標準化，国内外の学術機関へのサービス提供を行っている．NII RDCには，研究データ検索CiNii Research[1]，研究データ公開WEKO[2]，研究データ管理GakuNin RDM[3),4)]のサービスが含まれている．NIIでは全国の学術機関に対して，学術情報ネットワークSINET，学術認証フェデレーション（学認），eduroam[☆1]，UPKI証明書発行，クラウド導入コンサルティングサービスである学認クラウド，論文検索サービスCiNii，機関リポジトリ[☆2]のクラウドサービスJAIRO Cloudなどの学術情報基盤を提供してきた．NII RDCはこれらのITインフラストラクチャと参加機関コミュニティのネットワークを背景に構築されている．本稿では，執筆時点でまだ日本で普及していない，研究データ管理（RDM：Research Data Management）業務をサポートする目的で開発された研究データ管理サービスGakuNin RDM[☆3]を取り上げた．GakuNin RDMサービス概要，システム設計，実証実験レポートとユースケース紹介，研究開発にかかわる国内外の学術機関との連携等について述べる．

研究データ管理サービスGakuNin RDM

学術機関における研究データ管理業務は主に，研究不正防止のための証跡管理など研究公正と，共同研究者間でのデータ共有や研究成果の公開など研究推進の2つの観点で必要とされている．2017年頃から資金配分機関[☆4]が指定する研究費申請書において，データ管理計画（DMP：Data Management

☆1 eduroam：education roaming，国際的に相互利用が可能な学術機関向けの無線LANローミングサービス．

☆2 大学における紀要や博士論文を学術機関内で図書館員が登録・公開できるデータベース．

☆3 GakuNinは学術認証フェデレーションの略である学認の英字であり，学認を使ってシステムにログインできる研究データ管理サービスという意味．

☆4 ファンディング・エージェンシー．科学技術振興機構や日本医療研究開発機構など．

plan）の提出を義務（推奨）化し始めたことがある．研究者は研究プロジェクト申請または採択時に，論文に紐付く研究途中および研究成果のデータを原則10年間保存することが要求される．研究者は，研究データの管理・公開の手法について計画を明記して資金配分機関に提出する必要がある．各学術機関では研究データ管理・公開のための情報インフラの整備と，研究者と研究支援者向けの研究データ管理のためのトレーニングプログラムの需要が高まっている．これまで学術機関における研究データ管理は，特定分野の機関，プロジェクトあるいは研究者個人の自助努力として取り組まれていたケースは見受けられる．しかしながら，全国的に一律に実施していくためには，担当部署の整理や研究支援者の教育などソフトウェア面の課題もある．一般に大学では研究データ管理のためのストレージは情報基盤センターが学内サービスとして運用しており，研究成果公開のための機関リポジトリは図書館のシステム部門が管轄している．また，研究者への普及・啓蒙などはURA[☆5]などの研究戦略部門が担当しているが，大学の規模や慣習に応じ横連携の有無や強弱は異なる．以上のような理由から研究データ管理業務は研究者，研究支援者と所属機関が共同で進めていく業務ワークフロー[☆6]となる[5]．

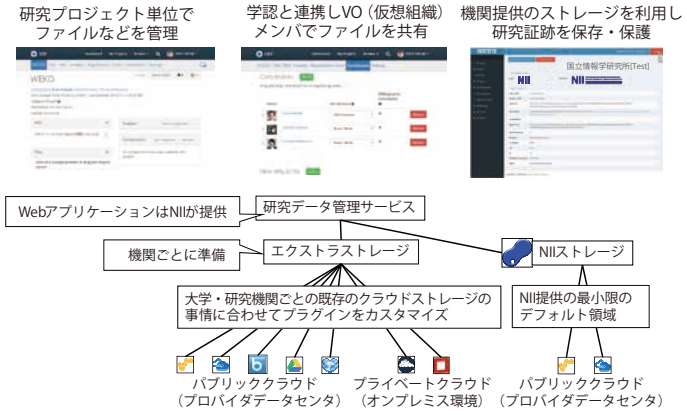

図-1　研究データ管理サービス GakuNin RDM

図-2　研究プロジェクトの管理ホーム画面

NIIが運営する研究データ管理サービスGakuNin RDMは，Webブラウザベースで利用するソフトウェアであり，特に国内の学術機関の組織内での研究データ管理のシステム導入や運用コスト負担を軽減するためのSaaS[☆7]としてユーザ機関にサービス提供されている．

GakuNin RDMのフロントエンドでは，サービス利用者は研究プロジェクトごとに研究データ管理用の入力用画面を立ち上げ，デスクトップのファイルをドラッグ・アンド・ドロップし，Webブラウザ上でファイルマネージャ[☆8]のように操作することができる．

GakuNin RDMではプロジェクトごとに標準でNII Storageという名称のストレージが最低限の容量分提供される．加えて，各学術機関が個別に契約し，学内提供しているエンタープライズ[☆9]レベルのオンラインストレージをGakuNin RDMの追加機能（アドオン）経由で接続することができる．代表的なクラウドサービス事業者のストレージのほか，オンプレミス[☆10]のサーバ上のオンラインストレージも利用できる．図-1にGakuNin RDMアプリケーションの機能とストレージ接続の概要を示す．図-2は拡大したGakuNin RDM中の研究プロジェクトの管理ホーム画面である．

☆5 URA：University Research Administrator
☆6 反復可能な業務処理のパターンを明示化したもの．
☆7 Software as a Service．一般に，特定の目的のソフトウェアをオンラインで提供したサービスを指す．
☆8 Windowsの ExploreやMacの Finderなどを指す．
☆9 企業や官公庁向けで無料サービス等ではなく，セキュリティが高いITサービスの契約形態．
☆10 On-premise．機関内の計算機リソースで情報システムの自社運営を行う方式のこと．

GakuNin RDMの試作版では外部サービスとの連携機能として，クラウドストレージのほかにNIIの研究データ公開基盤WEKO，研究データ解析基盤のJupyterHub☆11とシステム連携が実現されている．また，科学計算用ワークフローエンジンGalaxy☆12やビジネスプロセスモデル用ワークフローエンジンFlowable☆13とも試験的に連携している．さらに，NIIはオンプレミス用のクラウドストレージNextcloudのデスクトップクライアントツールをベースに，PCのファイルマネージャとの同期ツールも開発した．今後は試作した拡張機能をGakuNin RDMと連携する周辺サービスとして提供するために，規模の拡大（スケールアウト）を検討していく．

ここで，GakuNin RDM普及のためのターゲットユーザについて述べる．イノベータ理論☆14でいう初期採用（アーリーアダプタ）層として期待されるのは，学術機関の情報基盤センターのシステム導入担当部署である．次に前記追随（アーリーマジョリティ）層としては図書館職員が挙げられ，研究者が登録した研究データのメタデータを最適化するキュレータの役割を担う．また，全学的に研究データ管理業務を普及させるためのサービス説明会やセミナーの開催は図書館に強みがあり，これまでの図書サービスの講習会などのノウハウの活用が期待される．あるいは学生に対しては研究データ管理の教育プログラムが正規コースの講座として取り入れられていくことが考えられる．実際には，情報基盤センターと図書館だけでは組織を超えた連携が困難なケースも多いため，研究データ管理業務の実施に向けて研究推進・研究支援・研究倫理などの部門が橋渡しとなるケースが増加している．

システム設計

GakuNin RDMは米国NPO法人のCenter for Open Science（COS）が開発した研究データ管理のためのOSS（Open Source Software：ソースコードが開示されているソフトウェア），Open Science Framework（OSF）[6]のソースコードを分岐（フォーク）して，日本国内の実情に合わせた拡張開発を行っている．オリジナルのOSFにはない機能として，たとえば，外部クラウドストレージ（Microsoft Azure Blob Storage，OpenStack Swiftなど）用の追加機能，信頼のおける第三者機関のタイムスタンプ局（TSA：Time-Stamping Authority）を用いた研究証跡管理（研究データをいつ，誰が，どのように操作したかログを保存し調査できるように，保全すること）機能，導入機関の管理者向け機能などが追加で実装されている点が特徴である．図-3にGakuNin RDMの研究証跡保存機能の概要図を示す．

GakuNin RDMはシステムの機能拡張性を重視しており，RESTful APIによるサービス間の連携が可能なように，マイクロサービスアーキテクチャで構成されている．マイクロサービスとは，複数の小規模なサービスを疎結合し1つのWebアプリケーションを構成するソフトウェア開発手法である．その構成はWebアプリケーションのフロントエンド

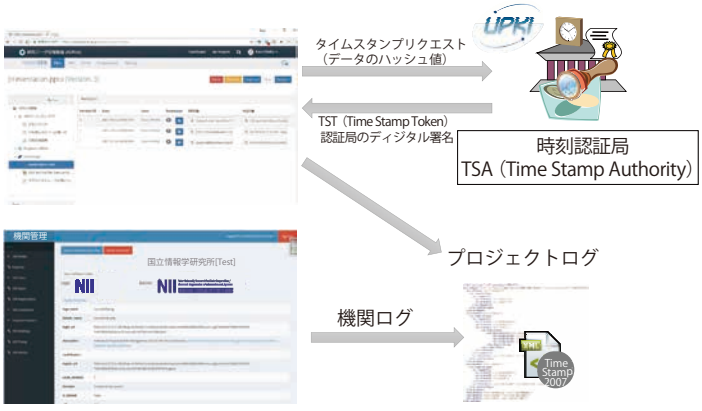

図-3　GakuNin RDMの研究証跡保存機能

☆11 Jupyter Notebookとは Webブラウザ上で Pythonプログラムを実行できる環境，プログラミングの過程で再利用性が高いだけでなく行間にコメントも残せるため教育目的でも使われる．iPythonの後継プロジェクト．JupyterHubは複数ユーザで Jupyter Notebookを利用するための管理システム．

☆12 生命科学分野で開発された専門的なソフトウェアやデータベースを連鎖させながら連続処理するためのツール．

☆13 ビジネスプロセスモデリング表記法のBPMN（Business Process Model and Notation）2.0に準拠した，Java言語で書かれた業務ワークフロー実行エンジン．豊富な入出力を定義したRESTful APIを持つ．

☆14 Everett M. Rogersが提唱したマーケティング用語．

サーバを中心に，管理用サーバ，ファイルストレージ中継サーバ，API サーバ，認証サーバ，データベースサーバ，検索用サーバ等があり，協調動作が可能なように設計されている．具体的にマイクロサービスの各コンポーネントを見ると，Web アプリケーションの中核部分は Python 言語ベースで開発されており，Python の Web フレームワーク[☆15]は Django を基本としている．管理用機能等のレガシーコード[☆16]では一部 Python の Web フレームワーク Flask が使われている．新しい機能については JavaScript の Web フレームワーク Ember.js で実装されている部分もある．マイクロサービスのコンポーネントにより新旧のフレームワークを混在させながら，機能拡張と再設計を反復的に行っている．GakuNin RDM の関係データベースマネジメントシステムには PostgreSQL，全部検索エンジンには Elasticsearch 等を用いている．図 -4 では GakuNin RDM のマイクロサービスアーキテクチャを示した．

開発と運用連携の改善

GakuNin RDM のソフトウェア開発はソースコードリポジトリ[☆17]GitHub を介して，OSS として公開で開発されている．ソフトウェアとして独自環境にインストールしたい場合や，機能拡張のための開発を行いたい場合は，ソースコードリポジトリを分岐して自由に開発することができる．多様なコントリビューターによる共同開発でも品質を保ち，工期を短縮するために，GakuNin RDM の GitHub リポジトリは継続的インテグレーション（CI：Continuous Integration）を実施している．代表的な CI ツールである Travis CI と連動しており，テストに通過したソースコードについて，機能単位でのプルリクエスト[☆18]を受け付けている．また，検証済みのソースコー

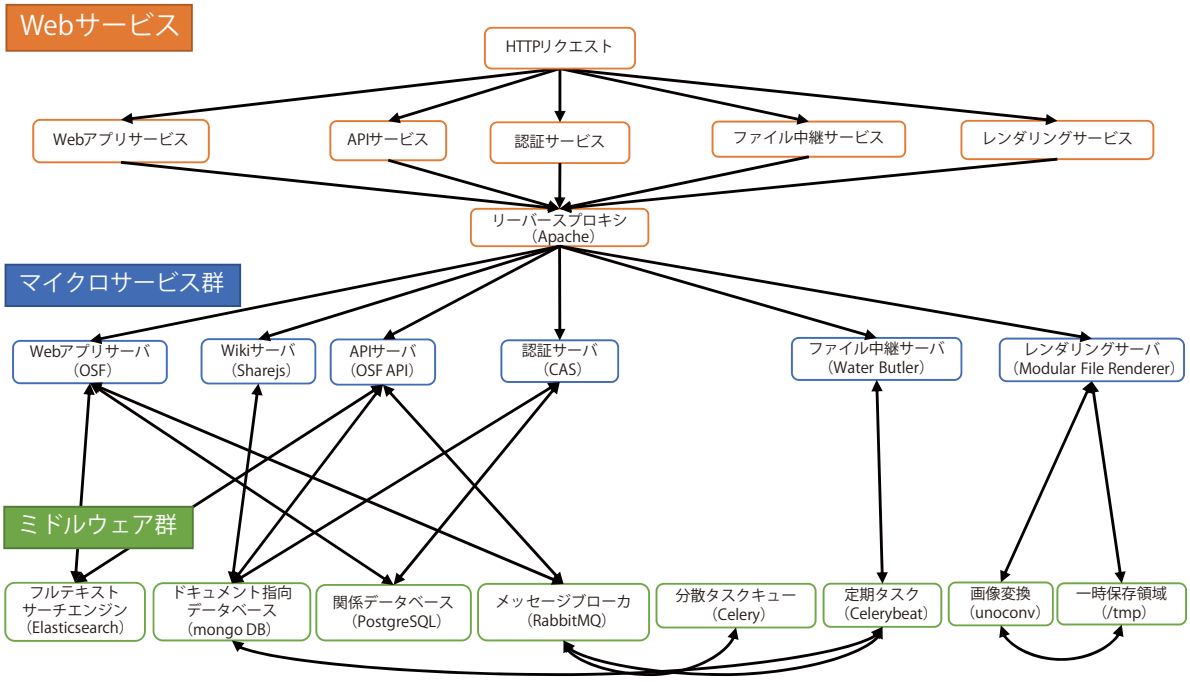

図 -4　GakuNin RDM のマイクロサービス構造

☆15 Web アプリケーションのモデルの雛形の規定およびその生成ツール．
☆16 現在では陳腐化したまたは廃止予定のソースコード．
☆17 複数の開発者でプログラミングを行う際に衝突が起こらないように，バージョン管理，マージ処理等を制御できる機能を持ったソースコード開発用の管理システム．
☆18 ある単位の開発済みソースコードについて，開発環境から本番環境用のソースコード管理者に受け入れ依頼を申請すること．

ドが即時に本番環境へ反映できるように継続的デリバリー（CD : Continuous Delivery）にも取り組んでいる．GakuNin RDM の GitHub リポジトリ中でマージされたコードは，CI/CD ツールである Jenkins を用いて固有パラメータが入力され，コンテナ型仮想化環境 Docker のコンテナにビルド[☆19]され，コンテナオーケストレーション環境である Kubernetes 上へデプロイ[☆20]される．

実証実験の報告

GakuNin RDM では 2017 年から 2018 年にかけて，開発ソフトウェアの機能を評価しフィードバックを得る目的で，国内学術機関を対象に 3 回の機能評価試験を開催した．計 24 機関の約 110 名の IT システム導入担当部署の教職員や研究者が参加し，テスト後にアンケート調査を通じてアイディアや批評が寄せられた．得られたフィードバックや知見を基に部分改修した．2019 年 4 月からは，実際に学術機関の現場での運用するための課題やノウハウを蓄積することを目的として，約 1 年半程度の中長期の実証実験を実施しており，この実験では学認を用いた認証連携を含め，ソフトウェア連携など利便性や，大規模ユーザでの同時利用を想定した段階的な負荷試験も行っていく．また，同時に研究者が扱う研究データの性質やプライバシーレベルによりどのようにストレージを使い分けるべきかを導入機関と調整しながら議論していく．

ユースケースの紹介

情報基盤センターストレージ連携

本節では，大学情報基盤センターが学内向けに提供しているプライベートクラウドと GakuNin RDM との連携の事例を紹介する．GakuNin RDM では外部クラウドストレージを API 経由で接続し，プロジェクト管理用の Web ページで一元的に管理できる．名古屋大学，北海道大学や京都大学の情報基盤センターでは，この機能を利用して，学内のオンプレミスサーバ上の Nextcloud を GakuNin RDM へ接続し，研究者に学内システムとして提供するサービスの実証実験が始まっている．Nextcloud ではユーザごとにデバイス単位で外部アプリケーションに接続するための，アカウントとパスワードをユーザが生成する機能がある．これを用いることで GakuNin RDM では，Nextcloud のメインアカウントを入力することなく，ストレージを接続して利用することができる．しかしながら，これを有効化していない大学の例もあるため別の連携方法の考案も今後の課題となっている．Nextcloud 以外の外部クラウドサービス環境でも，サービスごとにアクセスキー，シークレットキー，パーソナルトークンなどを GakuNin RDM 用に生成しておくことで同様の運用が可能である．いずれの場合でも，API 連携用の認証情報は GakuNin RDM のデータベース中では暗号化されて保持される．

図書館機関リポジトリ連携

本節では，大型研究プロジェクトにおける小グループ内での研究中データの限定共有と，研究成果データの限定公開を目的とした GakuNin RDM と NII 開発のリポジトリソフトウェア WEKO の連携の事例について述べる．研究者が論文を投稿し出版する段階では，その証拠となる研究データ（エビデンスデータ）を再利用可能な形で公開する必要がある．また，大型の研究プロジェクトでは，研究成果のデータをデータリポジトリでインターネット上に公開する必要がある．現在でも NII から WEKO2 をベースに，JAIRO Cloud という機関リポジトリサービスが提供されている．この，JAIRO Cloud には各大学の学術論文，紀要，症例報告書や博士論文が収録されている．開発中の WEKO3 では

☆19 個別ソースコードを実行可能な形式にコンパイルすること．
☆20 サーバを利用可能にすること．

文書だけでなく研究データの両方が取り扱えるように，ソフトウェアが再設計される見通しである．WEKO3 では欧州原子核研究機構（CERN）☆21 が提供する，リポジトリソフトウェアの開発用ライブラリ群である Invenio が利用されており，これはリポジトリの国際標準プロトコルである SWORD☆22 に対応している．GakuNin RDM ではほかのクラウドストレージ追加機能と同じように，WEKO 追加機能を実装しており，ワンストップで GakuNin RDM のストレージから，公開基盤のストレージへデータを転送できるようになっている．特にオシロロジー分野☆23 での，プロジェクト成果公開に試験的に活用され始めている．

研究不正防止の審査ワークフロー支援（生命科学分野応用）

本節では，NII と東京大学定量生命科学研究所（定量研）における，出版社でのアクセプト済み投稿論文を，出版前に学内の研究倫理部門で画像不正検査を行う目的で，GakuNin RDM とワークフローエンジン Flowable を連携する事例について紹介する．図 -5 は GakuNin RDM と研究倫理審査用ワークフローのシステム連携を図示したものである．GakuNin RDM では Flowable を追加機能として API 連携させるための開発を行っている．GakuNin RDM 中でワークフローが起動すると研究者が論文を登録するフォームが開くので，そこへ，タイトル，受理日，雑誌名などの最低限のメタデータを入力．ドラッグ・アンド・ドロップ操作でデスクトップから論文の最終原稿・組図，生データ，インデックス，チェックリストをアップロードして提出完了ボタンを押すと，論文中の画像データが抽出されて画像検査処理が実行される．提出画像データの検査結果は研究倫理推進室側に Flowable から通知が届き，担当者が該当画像をチェックし問題がなければ投稿を承認する．研究者から提出される論文の証拠となるデータがすべて揃い，研究倫理推進室が承認した場合は GakuNin RDM のプロジェクトが凍結され，以後は改変できなくなる．GakuNin RDM の研究証跡保存機能と合わせることで，ファイルがアップロードされた時間などが正確に証明することができるようになる．もし，研究不正が発生した場合にもファイルを操作した証跡が，GakuNin RDM 中に残るため追跡調査しやすくなり，そのため不正そのものに対する抑止効果も期待できる．

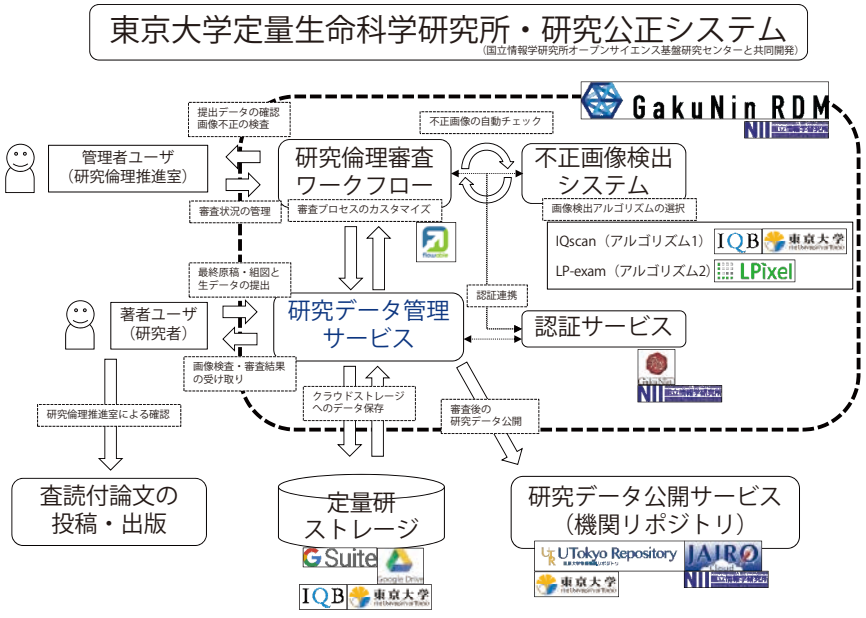

図 -5　GakuNin RDM と研究倫理審査用ワークフローの連携

☆21 CERN：European Organization for Nuclear Research.
☆22 SWORD : Simple Web-service Offering Repository Deposit.
☆23 神経科学，数理科学，臨床医学の融合した新学術領域．

国内外の学術機関との連携

GakuNin RDM の開発は前述の米国ヴァージニア州のシャーロッ

ツヴィルにあるCOSとNIIが技術提携しながら進めている．COSのセンター長はヴァージニア大学心理学部教授のBrian Nosek博士である．COSとNII間では研究者・技術者が相互に訪問し，日常的にメッセージツールでの情報交換を頻繁に行っている．一方で，国内連携としては，大学情報基盤センターのコンソーシアムである大学ICT推進協議会（AXIES☆24）研究のデータマネージメント部会があり，AXIESでは「学術機関における研究データ管理に関する提言」の策定が進められている．また，図書館リポジトリのコンソーシアムであるオープンアクセスリポジトリ推進協会（JPCOAR☆25）の研究データタスクフォース等からの意見を受けながら，システム開発を進めている．JPCOARではNIIと共同で研究支援者向けの研究データ管理に関する教材「RDMトレーニングツール」などが作成されており，NIIではトレーニングツールをベースとしたオンライン映像教材などが第2段まで開発されて提供されている．

今後の課題と展望

研究データ管理サービスGakuNin RDMの本格運用は2020年度の後半を予定しており，2019年現在はリリースに向けた実証実験を通じてユーザ数や利用機関数を段階的に拡大していく段階にある．トップダウンとボトムアップ両面からの提言書やガイドラインから，研究データ管理に求められる機能の拡張を進めていく．学術機関のガバナンス強化という意味では，研究不正防止や実験の証跡管理など研究公正は重要な一面ではあるが，それだけでは研究データ管理業務が研究者に普及しないことが懸念される．GakuNin RDMでは学際的な共同研究や，ロングテールデータ☆26に研究推進の支援を行うことで，情報学によるオープンサイエンスの推進を行っていく．今後の展望として，情報学の若手研究者や技術者と異分野のデータ専門家が議論・研究発表できるようなオープンサイエンスと研究データ管理の研究会が本会に必要と考える．なお，GakuNin RDMのソースコードはOSSとして以下のURL☆27で配布している．

参考文献

1) Kato, F., Kanazawa, T., Kurakawa K. and Ohmukai, I. : CiNii Research : A Prototype of Japanese Research Data Discovery, in eResearch Australasia 2018 (2018).
2) Yamaji, K., Aoyama, T., Furukawa, M. and Yamada, T. : Development and Deployment of the Open Access Repository and Its Application to the Open Educational Recourses, Springer, Cham, pp.395-403 (2016).
3) Komiyama, Y. and Yamaji, K. : Nationwide Research Data Management Service of Japan in the Open Science Era, in 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp.129-133 (2017).
4) Komiyama, Y. and Yamaji, K. : Interdisciplinary Research Data Management Service for the whole Universities and Research Institutions in Japan that Emphasizes Research Integrity, in Digital Infrastructure for Research 2018, No.6, p.164 (2018).
5) Funamori, M., Hayashi, M., Komiyama, Y., Tsuchiya, M. and Yamaji, K. : Requirements Analysis of System for Research Data Management to Prevent Scientific Misconduct, in 7th IIAI International Conference on Advanced Applied Informatics (IIAI AAI 2018) (2018).
6) Foster, E. D. and Deardorff, A. : Open Science Framework (OSF), J. Med. Libr. Assoc., Vol.105, No.2, p.38 (Apr. 2017).

(2019年2月1日受付)

込山悠介（正会員） komiyama@nii.ac.jp
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年東京大学大学院農学生命科学研究科博士課程修了．2014年東京大学医科学研究所特任研究員．2016年より現職．

☆24 AXIES : Academic eXchange for Information Environment and Strategy.
☆25 JPCOAR : Japan Consortium for Open Access Repository.
☆26 これまで収集困難であった，一つひとつのファイルサイズは小さいが，有益な情報を持つ多様な研究データのこと．積分するとビッグデータに匹敵する可能性がある．
☆27 https://doi.org/10.5281/zenodo.2544682