

研究データのフロー自動化の検討

～ NextCloud における解析ソフトウェアの入出力データの収集・管理 ～

林 寛生¹⁾, 實本 英之¹⁾, 菊地 伸治¹⁾,
西村 将臣²⁾, 道川 隆士²⁾, 野田 茂穂²⁾, 横田 秀夫²⁾

1) 理化学研究所 情報統合本部

2) 理化学研究所 光量子工学研究センター

hiroo.hayashi@riken.jp

A Study of Automation of Research Data Flow - Management of Input/Output Data of an Analysis Software on the NextCloud -

Hiroo Hayashi¹⁾, Hideyuki Jitsumoto¹⁾, Shinji Kikuchi¹⁾,
Masaomi Nishimura²⁾, Takashi Michikawa²⁾, Shigeho Noda²⁾, Hideo Yokota²⁾

1) RIKEN Information R&D and Strategy Headquarters

2) RIKEN Center for Advanced Photonics

概要

研究データの利活用促進のための方策として、データガバナンスやメタデータ付与を無意識的に実行するため、筆者らはデータフローの自動化に関して検討している。一例として、クラウドストレージの NextCloud 上で、画像解析ソフトウェアの入出力データを収集・管理するケースにおいて、NextCloud Flow を利用したデータフロー自動化を検討および試験実装した。本稿では、その結果について述べる。

1 はじめに

実験や観測、またそれらの解析によって得られる研究データは、分野の枠を越えた様々な研究者との共創を目的に、ロケーションや組織を跨いで研究者間で共有することによって、さらなる利活用を推進することが求められている。しかしながら、研究データの効率的な共有は容易ではなく、他のロケーションや組織が所有する研究データの発見やアクセスにはまだ課題が多い。これらの課題の解決には、一定のルールや方針に基づくデータの収集・管理（データガバナンス）やデータの効率的な検索・取得に有用なメタデータの付与などが必要である。

筆者らは、理化学研究所における研究データの利活用促進のための方策として、データガバナンスやメタデータ付与を無意識的に実行するための、データフロー自動化について検討している。まず、その一例として、画像解析ソフトウェアの ICP/VCAT5 の入出力データをクラウドストレージである NextCloud 上で収集・管理するケースにおいて、NextCloud Flow を利用し

たデータフローの自動化について検討、および試験実装を行った。

本稿では、今回の検討・実装の結果について述べる。

2 環境について

本章では、本検討の対象とした環境について、その概要を記す。

2.1 NextCloud

NextCloud [1] は、クラウドストレージを実現するためのオープンソースソフトウェアである。理化学研究所では、研究データの長期保管と公開のためのサービス群である理研研究情報管理システム(R2DMS: RIKEN Research Data and copyrighted work Management System)を研究者に提供しているが、研究データの格納および複数の利用者間での共有を目的としたオンラインストレージサービスとして、NextCloud を採用している。

2.2 NextCloud Flow

NextCloud Flow は NextCloud のワークフロー

エンジンとして提供され、データファイルの操作（移動や新規作成、削除など）をトリガーとして、一連の処理を自動的に実行する機能である。処理の実行には、任意のシェルスクリプトを起動することが可能であるため、NextCloud が載るサーバにインストールされたソフトウェアや Linux コマンドなどを利用することができる。ただし、使用するスクリプトやトリガー条件などについては、あらかじめ NextCloud の管理ユーザが設定しておく必要がある。

2.3 ICP/VCAT5

VCAT5 [2][3] は、理化学研究所で開発され、生物学、金属組織学を対象とした光学顕微鏡、医学工学の X 線 CT、MRI などのボクセルデータを対象とした画像処理に広く使われている画像解析ソフトウェアである。ICP (Image Communication Platform) [4][5][6][7] は VCAT5 をクラウド上に組み込んだシステムで、VCAT5 の入出力となるデータの収集・管理やユーザ間のデータ共有などを実現し、効率的なデータ利用のためのインターフェースを提供する。

理化学研究所において、これまで ICP/VCAT5 は独自にサーバを構築して運用されてきたが、研究者間のデータの共有・利活用をさらに深化させるため、また、システムの維持管理コストを抑えることを目的に、研究所の共通基盤への移行が検討されている。

3 研究データのフロー自動化の検討

本章では、R2DMS における研究データフローの自動化を検討する上でのシステムのアーキテクチャと、本検討の具体的な実装について述べる。

3.1 アーキテクチャ

図 1 に、R2DMS のシステムのアーキテクチャを示す。

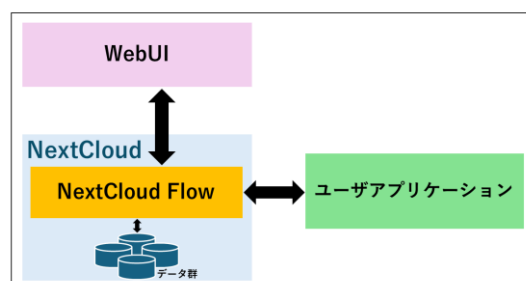


図 1 R2DMS のシステムのアーキテクチャ

WebUI は、ユーザとシステムとを繋ぐインターフェースの役割を担う。ユーザからのリクエスト受け、それに対応する通信を NextCloud Flow に対して実行する。

ユーザアプリケーションには任意のアプリケーションが想定される。WebUI からのリクエストを元に、NextCloud Flow を介して、指定されたデータに対してユーザが意図した処理を実行する。

NextCloud Flow は、データを一元管理するレイヤーである NextCloud の一機能（ワークフローエンジン）として組み込まれ、NextCloud 上のデータファイルの操作をトリガーに様々なデータ処理や計算を自動的に行う機能を提供し、ユーザアプリケーションや WebUI との間でデータを受け渡す役割を担う。

3.2 実装

3.1 節で説明したアーキテクチャの構成要素に対して、それぞれどのような実装としたかについて以下に記す。

1) WebUI

本検討では、NextCloud が提供する GUI (Web ブラウザでアクセスした際の UI) をそのまま利用することとした。

2) ユーザアプリケーション

前述した ICP/VCAT5 を実装例として採用した。このソフトウェアは解析部分とデータ管理部分に分かれており、解析部分は従来機能のままで、データの入出力先として NextCloud のデータストレージを参照するように設定し、管理部分については、その機能を NextCloud Flow へ移植することとした。

3) NextCloud Flow

ICP/VCAT5 から移植したデータ管理の機能

は、大きく分けてデータ登録、サムネイル作成、メタデータ作成、メタデータ修正、来歴管理、エクスポート、フォーマット変換である。それぞれに対し、一連の処理を実行するためのシェルスクリプトを作成し、スクリプトを起動するトリガーとともにフローとして NextCloud に設定した。以下に、それぞれの機能の実装（フローのトリガー設定とスクリプトの内容）について概要を記す。

①データ登録

・フローのトリガー：

データファイルを特定のフォルダへ格納（=新規作成）すること。

・スクリプトの概要：

新規作成されたデータファイルに対して NextCloud が自動的に割り振るユニークな ID を参照し、その ID を付したフォルダを作成する。トリガー条件を満たしたデータファイルをそのフォルダに移動することで NextCloud への登録とする。

②サムネイル作成

・フローのトリガー：

データファイルを特定のフォルダへ格納（=新規作成）すること。（※ただし、データ登録(①)のフロー内で連動して実行される）

・スクリプトの概要：

NextCloud に登録されたデータに対し、多種多様な画像フォーマットの読み書きに対応した Bio-Formats [8] のコマンドラインツールを利用して PNG 形式のサムネイルを生成する。サムネイルファイルはユーザが参照できるフォルダに配置する。

③メタデータ作成

・フローのトリガー：

データファイルを特定のフォルダへ格納（=新規作成）すること。（※ただし、データ登録(①)のフロー内で連動して実行される）

・スクリプトの概要：

NextCloud に登録されたデータに対し、Bio-Formats のコマンドラインツールを利用して画像に関する詳細情報を抽出し、データ固有の ID やファイルの最終更新時刻等のデータ登録時に得られる情報と合わせて Markdown 形式のファイルに出力する。このファイル（メタデ

ータファイル）はユーザが参照できるフォルダに配置する。

④メタデータ修正

・フローのトリガー：

メタデータ編集ファイルを特定のフォルダへ移動すること。

・スクリプトの概要：

トリガー条件を満たしたメタデータ編集ファイルを読み込んで、既存のメタデータファイルの一部を上書きする。なお、メタデータ編集ファイルは、メタデータ作成(③)においてメタデータファイルを作成する際に、編集可能なメタデータの要素と現状の値を出力したテキスト形式のファイルであり、フローのトリガーとなるファイル操作の前に、必要に応じてユーザが編集することを想定する。

⑤来歴管理

・フローのトリガー：

ICP/VCAT5 の処理履歴ファイルを特定のフォルダへ格納（=新規作成）すること。

・スクリプトの概要：

VCAT5 の終了時に出力される一連の画像処理の履歴（ICP/VCAT5 独自フォーマットの XML）ファイルから、各処理における入出力データと処理詳細の情報を抽出し、それぞれのデータのメタデータファイルに来歴情報（親子関係）を追記する。また、任意のグラフを可視化するオープンソースのソフトウェア Graphviz [9] を利用して、一連の画像処理を可視化したファイル（PDF 形式）を作成する。

⑥エクスポート

・フローのトリガー：

サムネイルファイルを特定のフォルダへ移動すること。

・スクリプトの概要：

トリガー条件を満たしたサムネイルファイルに対応するデータファイルを VCAT5 が参照するフォルダにコピーする。また、ICP/VCAT5 独自フォーマットのテキストファイル（入力ファイル群のパスを示すリスト）を作成する。

⑦フォーマット変換

・フローのトリガー：

フォーマット変換パラメタファイルを特定

のフォルダへ移動すること。

・スクリプトの概要：

トリガー条件を満たしたフォーマット変換パラメタファイルを読み込んで、ICP/VCAT5 が独自に開発した変換ツールを用いてデータファイルのフォーマットを変換する。なお、フォーマット変換パラメタファイルは、メタデータ作成 (③) において抽出した画像の詳細情報を基に作成されるテキスト形式のファイルであり、トリガーとなるファイル操作の前に、必要に応じてユーザが編集することを想定する。

3.3 利用者からの評価と今後への検討

3.2 節で述べた実装について、ICP/VCAT5 の開発者グループで試験利用を行っている。

現時点において、使い勝手が悪いとの意見が上がっている。原因の1つは操作が煩雑であることである。NextCloud Flow で実装した機能を使うためにはトリガー条件を満たすファイル操作が必要であるが、機能によって扱うファイルや格納・移動先のフォルダが異なるため、直感的な操作が難しくマニュアルが手放せないとの評価である。

操作ミスによるシステム誤作動の懸念も挙げられている。上述のように操作の煩雑さから扱うファイルや格納・移動先のフォルダを誤り、ユーザの意図しないフローが起動してしまうことが容易に想像できる。また、NextCloud の GUI 上にてドラッグアンドドロップでファイルをフォルダに移動させると、移動の失敗に気づかず、意図したフローが起動しないケースがあることも報告されている。

これらの使い勝手に関する課題については、NextCloud の外部に専用の WebUI を導入して対応することとなった。NextCloud では WebDAV プロトコルに対応したファイル操作のための API が提供されており、WebUI と NextCloud 間で行うファイル操作において再利用可能な部分を API としてまとめることで、ユーザにとって使い勝手の良い WebUI を構築しやすい環境を整える。

この独立した WebUI はまだ設計段階であるが、操作ミスの防止やユーザビリティの確保のため、「①NextCloud に登録されるデータのサム

ネイルを WebDAV API でダウンロードして一覧表示 → ②ユーザが興味のあるサムネイルを選択 → ③ユーザが特定の機能に対応するボタンをクリック → ④当該機能のトリガー条件を満たす WebDAV API を呼び出し」という方式を基本とすることを想定している。

4 まとめ

理化学研究所における研究データのフロー自動化検討の一例として、研究所の共通基盤上のストレージである NextCloud において、幅広い分野で利用される画像解析ソフトウェア ICP/VCAT5 の入出力データを収集・管理するケースを取り上げ、データフロー自動化を実現する手段として NextCloud Flow を利用し、試験実装を行った。

NextCloud Flow では、NextCloud 上の単純なファイル操作をトリガーに任意のシェルスクリプトを起動できることから、ICP/VCAT5 のデータ管理部分の主な機能を移植することができ、ICP/VCAT5 の入出力データの収集・管理に係る様々な処理が自動化できることを確認した。ただし、ユーザの利便性には課題があるため、NextCloud の外部に独立した WebUI を導入する方針である。

本検討において NextCloud Flow で試験実装した機能の一部(データ登録、サムネイル作成、メタデータ作成、メタデータ修正)は、一般的な Linux コマンドや多様なデータフォーマットに適用できる汎用ツールなどで実現しているため、共通機能として他のユーザアプリケーションでも利用できる見込みである。しかし、ICP/VCAT5 に特有のフォーマットやツールに依存して共通化が難しい機能もあり(エクスポート、来歴管理、フォーマット変換)、今後他のユーザアプリケーションも含めて一般化する場合には、利用するアプリケーションごとにある程度の独自開発が必要になると考える。

2.2 節でも述べたように、NextCloud Flow は、NextCloud の管理ユーザがあらかじめ起動するシェルスクリプトやトリガー条件を設定しておかなければならないという制約がある。多様な研究データ管理を想定する上では、一般のユー

ザが任意のタイミングで任意のフローを定義する
ようなケースも考えられるため、Globus [10]
および Globus Flow を利用したデータフローの
自動化についても今後検討を進める。

参考文献

- [1] Nextcloud GmbH, "Nextcloud - Open source content collaboration platform", Nextcloud, <https://nextcloud.com/>, (参照 2025-09-19)
- [2] 横田 秀夫, "生物画像処理システム", 映像情報メディア学会誌, 67 巻, 9 号, pp. 784-788, 2013.
- [3] 画像情報処理研究チーム, "VCAT5", 理化学研究所, <https://logistics.riken.jp/vcat/vcat/ja>, (参照 2025-09-19)
- [4] M. Morita, T. Tawara, M. Nishimura, S. Yoshizawa, B. Chou, I. Kuroki, T. Ijiri, Y. Tsujimura, R. Himeno, and H. Yokota, "Communication Platform for Image Analysis and Sharing in Biology", International Journal of Networking and Computing, Vol. 4, No. 2, pp. 369-391, 2014.
- [5] 森田 正彦, 吉澤 信, 井尻 敬, 俵 丈展, 西村 将臣, 趙 武魁, 黒木 一平, 舛本 現, 辻村 有紀, 姫野 龍太郎, 横田 秀夫, "生物画像処理のためのクラウド型システム", Medical Imaging Technology, Vol. 33(3), pp. 112-117, 2015.
- [6] 理化学研究所, "クラウドを利用した画像データの
新基盤システム「ICP」を開発", 2013 年 12 月 5 日,
https://www.riken.jp/press/2013/20131205_2/
- [7] 理化学研究所, "材料組織の画像解析クラウドシステムを開発", 2016 年 11 月 2 日,
https://www.riken.jp/press/2016/20161102_5/
- [8] University of Dundee & Open Microscopy Environment, "Bio-Formats", Open Microscopy Environment,
<https://www.openmicroscopy.org/bio-formats/>,
(参照 2025-09-24)
- [9] The Graphviz Authors, "Graphviz", Graphviz,
<https://graphviz.org/>, (参照 2025-09-24)
- [10] The University of Chicago, "Globus", Globus,
<https://www.globus.org/>, (参照 2025-09-19)