



AI 駆動大規模データ解析時代を見据えた質量分析データリポジトリの新展開

○高橋悠志¹⁾, 吉沢明康¹⁾, 松田史生²⁾, 木下聖子³⁾, 石濱泰⁴⁾, 奥田修二郎¹⁾

1) 新潟大学大学院医歯学総合研究科, 2) 大阪大学大学院情報科学研究科, 3) 創価大学大学院理工学研究科, 4) 京都大学大学院薬学研究科

PROTEOMICS



jPOST repository

<https://repository.jpostdb.org>

jPOST (Japan ProteOme Standard Repository / Database) はリポジトリ・再解析・データベースから成るプロテオーム統合データベースであり、その中でも **jPOST リポジトリ** は FAIR 原則に従って質量分析データを寄託し、これを蓄積するための情報基盤である。ウェブブラウザのみで完結する直感的なユーザインタフェースを提供しており、エンバーゴ機能や大容量ファイルの高速分割アップロードが可能などの特徴を持つ。2025 年 9 月現在、約 3,600 プロジェクトが登録済みであり、総データ量は 130 TB を超える。

近年では、JPDM (Journal of Proteome Data and Methods) 誌にデータディスクリプト論文と共に投稿されたメタデータを活用することでプロテオームデータの再解析を加速させる試みが開始されている。



GlycoPOST

<https://glycopost.glycosmos.org>

jPOST リポジトリの設計を踏襲して作られた **GlycoPOST** はグライコームにおける質量分析の生データリポジトリであり、GlyCosmos プロジェクトで管理されている。GlycoPOST のメタデータは **MIRAGE** (Minimum Information Required for A Glycomics Experiment) ガイドラインに準拠しており、現在 Sample Preparation、LC および Glycomics MS ガイドラインに用意されたメタデータが入力できる。

近年、UniCarb-DR (グライコミックスの質量分析のアノテーション済み同定結果のリポジトリ) との連携も開始され、アノテーションソフトウェアである GlycoWorkbench で作成した同定結果ファイルが GlycoPOST に投稿された場合にはプロジェクト公開後に自動的に UniCarb-DR にも登録されるようになった。

GLYCOMICS



MB-POST

<https://repository.massbank.jp>

日本発の化合物リファレンス・マスペクトル・データベースである MassBank に高品質なスペクトルデータを大幅に追加するため、メタボロミクス領域では人間による curation のみに頼らずに大量のリファレンス・スペクトルを収集するパイプラインの構築を目指して 2023 年度より Shin-MassBank プロジェクトが開始された。

MB-POST (MassBank-POST) は jPOST リポジトリや GlycoPOST と同様の設計を採用しており、化合物を質量分析で測定した生データとその測定条件を記述するメタデータを格納するリポジトリである。Shin-MassBank プロジェクトのパイプライン中で MB-POST は投稿される測定データを一挙に引き受け、後ろに控える MassBank Human や MassBank *in silico*、および MassBank Links へ受け渡す。

METABOLOMICS

各リポジトリへのデータ投稿の流れ

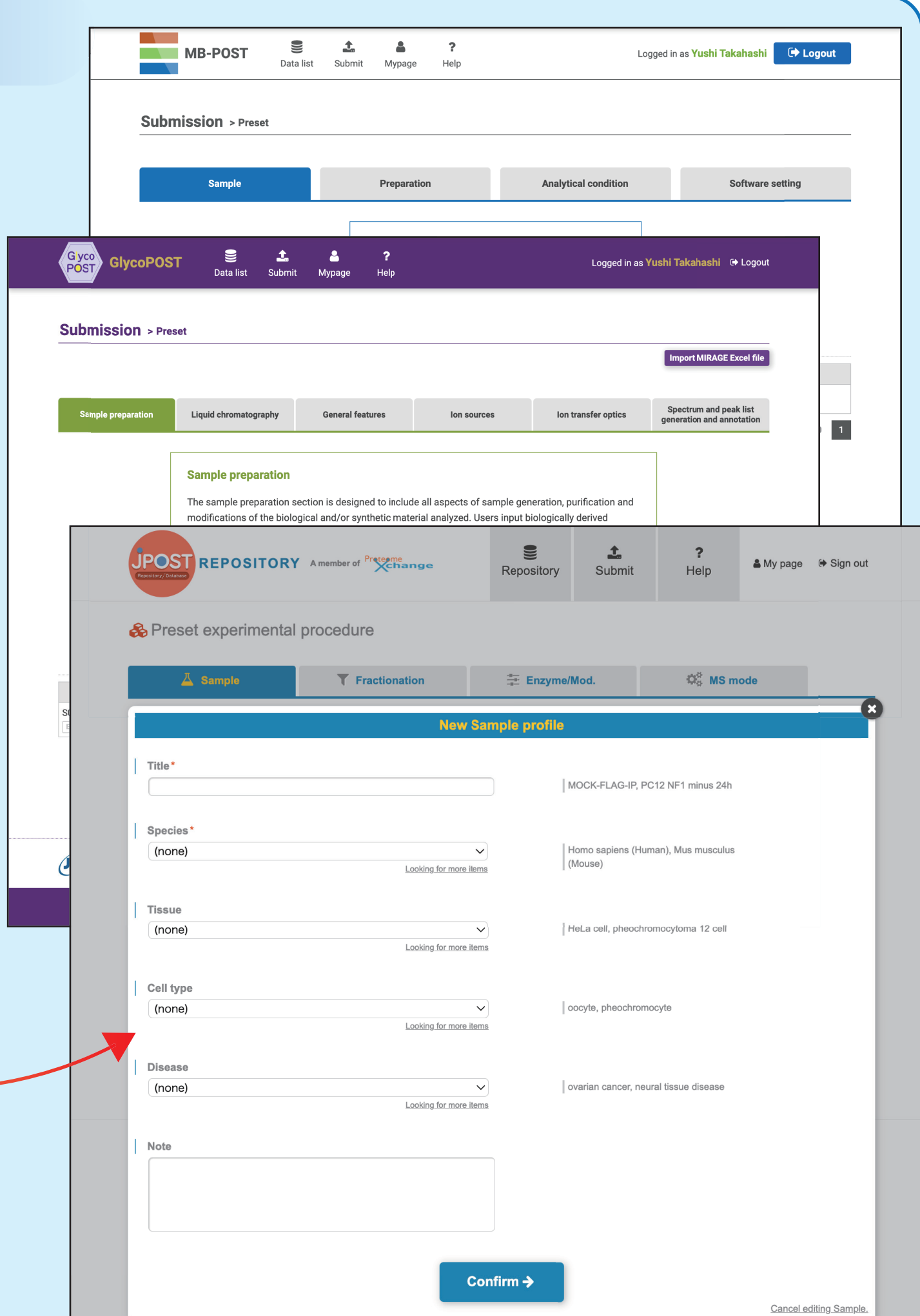
①プリセットの準備

- データの寄託を行うためには、最初に実験に用いた試料や実験条件などの情報をメタデータとして登録する。

- 各種メタデータは「**プリセット**」というひとまとまりの単位で登録する。過去に登録されたプリセットは同じ条件で行った別の実験データを投稿する際に再利用できる。

- 現在、各リポジトリには以下のプリセットのカテゴリが用意されている。

- **jPOST リポジトリ**
Sample / Fractionation / EnzymeMod / MS mode
- **GlycoPOST**
Sample preparation / Liquid chromatography / General features / Ion sources / Ion transfer optics / Spectrum and peak list generation and annotation
- **MB-POST**
Sample / Preparation / Analytical condition / Software setting

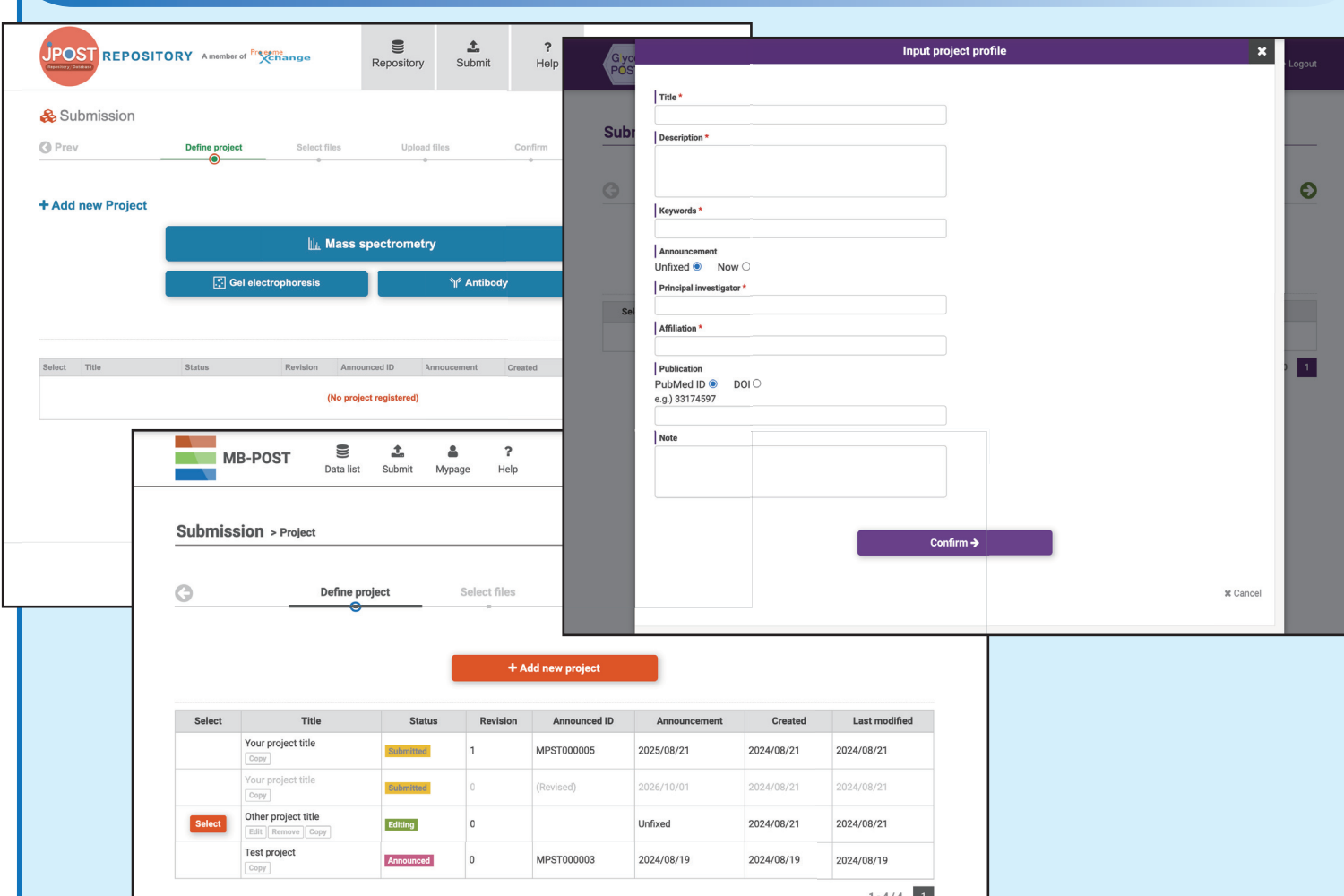


各リポジトリのプリセット一覧画面および jPOST リポジトリの Sample プリセット登録画面

②プロジェクトの作成

- 「**プロジェクト**」は投稿ファイルをまとめる単位であり、「論文を書くために必要なデータ一式」(Raw データ、ピークリスト、同定結果) が格納される。

- 試料の生物種が異なる場合や実験条件が大きく異なるなど、**メタデータ**が大きく異なる場合にはそれぞれを別個のプロジェクトとしても良い。

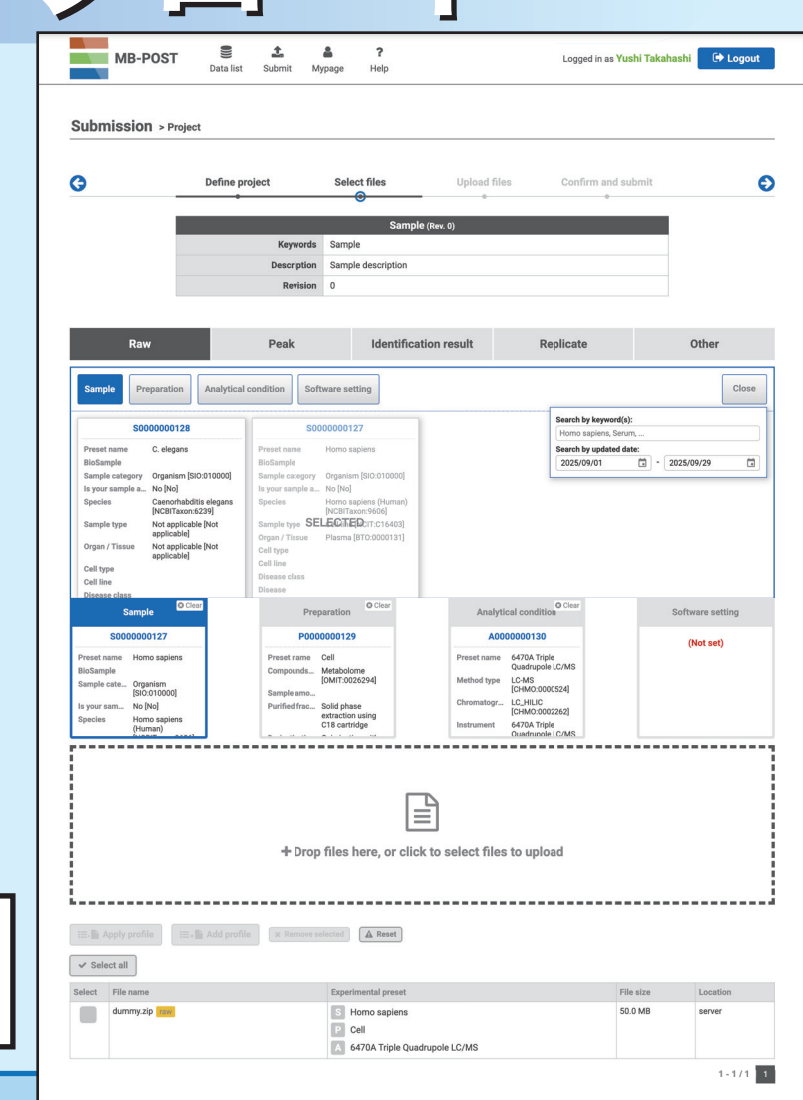


各リポジトリのプロジェクト一覧・登録画面

③ファイルへのメタデータ登録とアップロード

- プロジェクトに投稿ファイルを登録する際には、予め①で定義しておいたプリセット(メタデータ)をドラッグ&ドロップ主体の直感的なインタフェースを利用して各ファイルに対応付けることができる。

- アップロード時には各ファイルは自動的に細かいチャックに分割され、並列にアップロードされることでサーバーに高速転送される。



MB-POST のファイル選択・プリセット選択画面

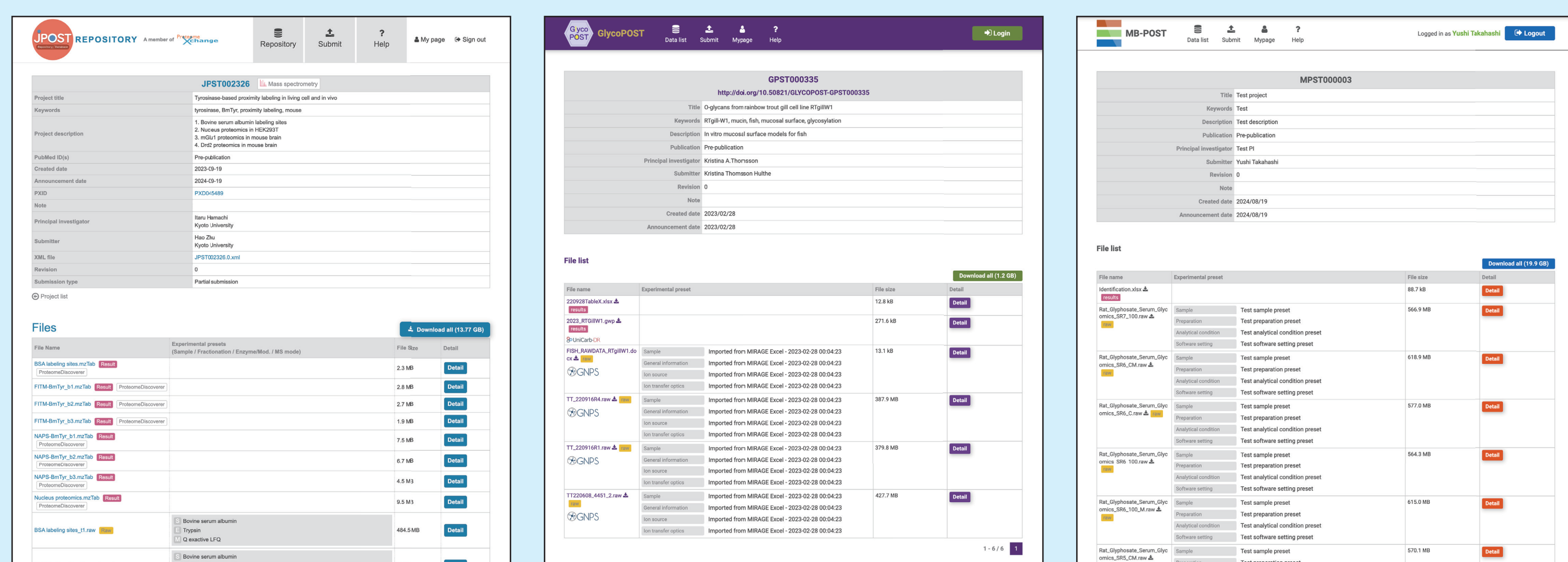
④プロジェクトのリバイス・プレビュー・公開

- 投稿されたプロジェクトを更新(ファイルの追加・削除、メタデータ変更)したい場合、各リポジトリの Mypage 画面で公開前のプロジェクトの**リバイス**が可能。

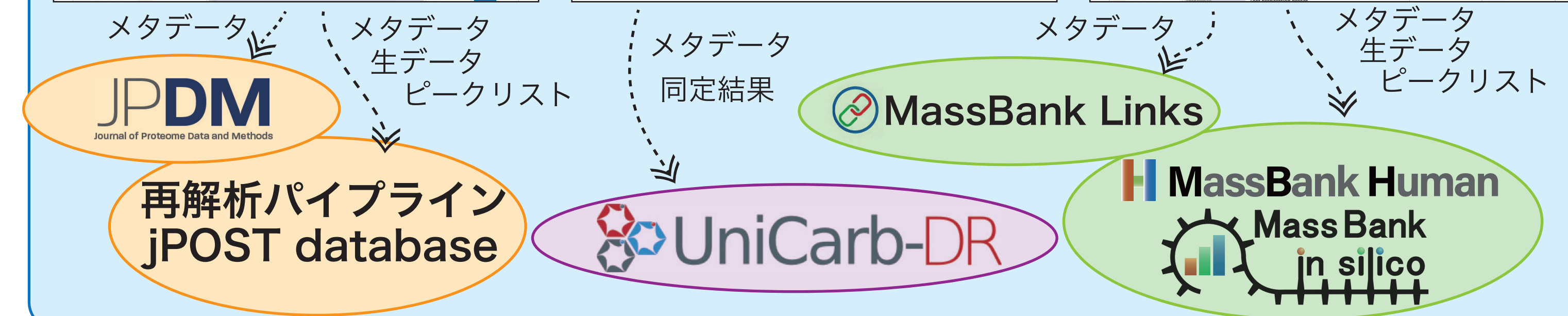
- Mypage ではジャーナルの査読用に公開前のプロジェクトのプレビュー用 URL も発行できる。

- 各リポジトリは「**エンバーゴ (embargo)**」機能に対応しており、ジャーナルのレビュー終了後にプロジェクトを公開するように公開日時を Mypage 上で自由に設定可能。

- 公開されたプロジェクト情報は再解析などの目的などでも利用されている。



jPOST の Mypage 画面

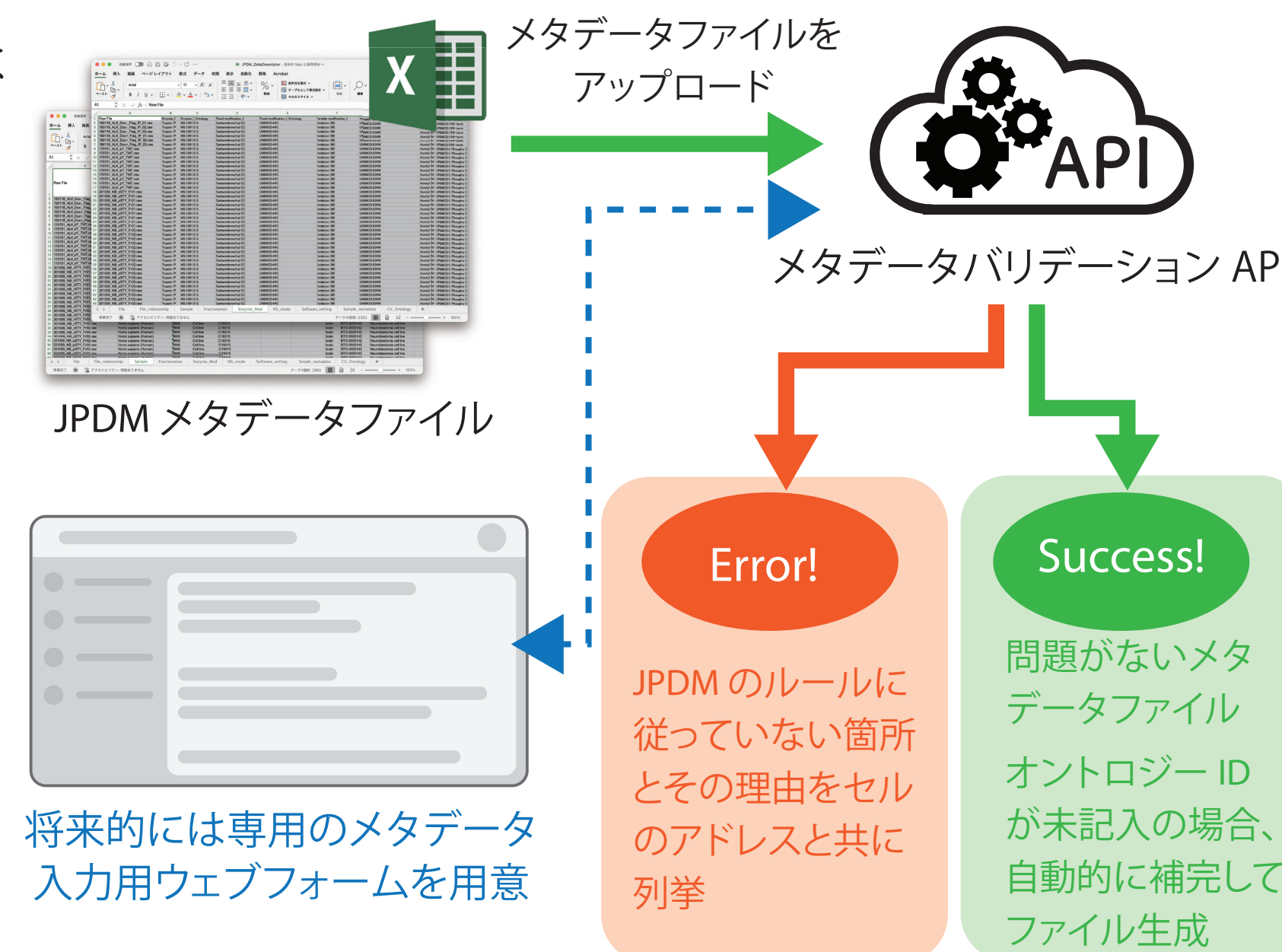


AI 駆動データ解析時代に備えて：メタデータの品質管理

- JPDM メタデータを活用した再解析では手作業でのメタデータ確認・微調整が必要な場合が多いことが判明

- AI 技術を活用した大規模解析でも実験データへの適切なメタデータ付与が要であるため、メタデータの品質管理が重要

- そのため、メタデータのバリデーションシステムの開発が現在進められている



謝辞 本研究・開発は科学技術振興機構 (JST)・NBDC 事業推進室による統合化推進プログラム予算によって実施した。