From JPDM to jPOST：データ論文を用いたデータ再解析の半自動化

From JPDM to jPOST: Semi-automated data reanalysis using data articles

吉沢明康[1]，高橋悠志[1]，小林大樹[2]，守屋勇樹[3]，幡野敦[2]，高見知代[2]，松本雅記[2]，荒木令江[4]，田畑剛[5,6,7]，岩崎未央[6]，杉山直幸[8]，小寺義男[9]，福島敦史[10]，田中聡[11]，五斗進[3]，河野信[3,12]，奥田修二郎[1]，石濱泰[7]

Akiyasu C. Yoshizawa[1], Yushi Takahashi[1], Daiki Kobayashi[2], Yuki Moriya[3], Atsushi Hatano[2], Tomoyo Takami[2], Masaki Matsumoto[2], Norie Araki[4], Tsuyoshi Tabata[5,6,7], Mio Iwasaki[6], Naoyuki Sugiyama[8], Yoshio Kodera[9], Atsushi Fukushima[10], Satoshi Tanaka[11], Susumu Goto[3], Shin Kawano[12], Shujiro Okuda[1], Yasushi Ishihama[7]

1 新潟大・医・メディカル AI セ，2 新潟大・医・システム生化学，3 情シス研・DS・DBCLS，4 熊本大・院生命，5 MassSoft，6 京大・CiRA，7 京大・院薬，8 国循・創薬オミックス，9 北里大・理，10 京都府大・院生命環境科学，11 Trans-IT，12 北里大・未来工

1 Med. AI Ctr., Sch. Med., Niigata Univ., 2 Dep. Omics and Systems Biol., Sch. Med., Niigata Univ., 3 DBCLS, DS, ROIS, 4 Grad. Sch. of Med. Sci., Kumamoto Univ., 5 MassSoft, 6 CiRA, Kyoto Univ., 7 Grad. Sch. of Pharma. Sci., Kyoto Univ., 8 Omics Res. Ctr., Nat. Cerebral and Cv. Ctr., 9 Sch. Sci., Kitasato Univ., 10 Grad. Sch. Life and Environ. Sci., Kyoto Pref. Univ., 11 Trans-IT, 12 Sch. Frontier E., Kitasato Univ.

Introduction

jPOST (Japan Proteome Standard Repository/Database) の開発は、今年 10 年目に突入した。jPOST は今まで、以下の 3 本柱で構成されてきた：

(1) **リポジトリ**（データ蓄積＆公開用サイト）…論文投稿時に公開が義務づけられている、（質量分析の）生データの寄託を受ける

(2) **再解析**…統一的な方法で生データを解析し直す

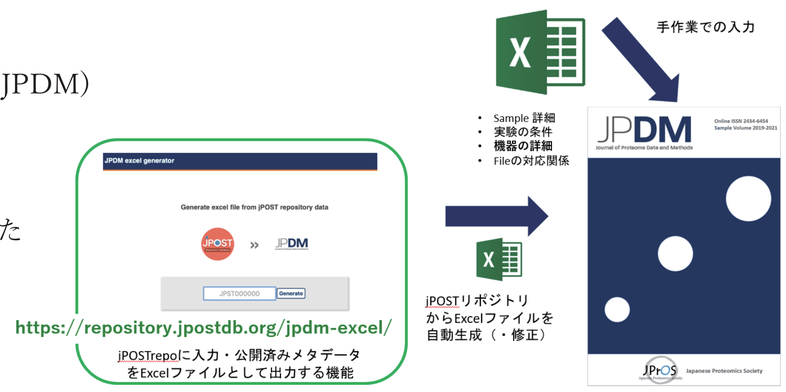
(3) **データベース**…再解析の結果を収録し多様な手法で結果を表示する

データの再利用、特に再解析にはメタデータが必須であること、現在のメタデータでは Sample and Data Relationship Format (SDRF) 情報が不充分であることから、データ論文ジャーナル *Journal of Proteome Data and Methods* (JPDM) の創刊の他、日本プロテオーム学会 (JPrOS) を通してコミュニティ・ベースでの論文調査などを進めてきた（トーゴーの日シンポジウム 2022・ポスター 30）。

本発表では、JPDM に掲載されたデータ論文を用いて再解析を半自動化する試みについて紹介する。

データ論文ジャーナル JPDM

- 電子ジャーナル *Journal of Proteome Data and Methods* (JPDM)
  …2019 年創刊（日本プロテオーム学会 (JPrOS)）
  - データ論文 (data descriptor) を掲載
  - 「データ論文」…研究論文ではなく、研究で得られたデータ（を再利用するため）の論文。
  - 再解析をスムーズに進めるために、より詳細なメタデータを取得することが目的。



- データ論文…寄託したリポジトリを問わず、公開済みのデータについての論文はすべて掲載対象。
- jPOST に寄託したデータの場合…jPOST に自分が入力したメタデータを、JPDM 形式で自動出力可能（加筆することでそのまま JPDM 原稿として利用できる）。

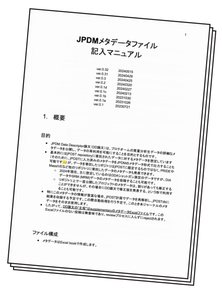
jPOST トップページ https://jpostdb.org/

JPDM サイト・トップページ
https://www.jhupo.org/jpdm/

JPDM メタデータ

- 2023 年 7 月に**書式アップデート**（旧書式での提出は、再提出が必要）
- 一般の（寄託されるデータの）メタデータと再解析用メタデータの違い
  - **一般**のメタデータは**一つの研究論文で使用した全データを 1 個のプロジェクトとして**登録されることが多い
  - **再解析**では、例えば
  - 1 個のプロジェクト内の全データは**メタデータが共通**
    - → 分割が必要なことも
  - biological replicate は同一プロジェクト内で別データセットに分割の必要あり
- これらの処理を自動で行う

第 4 の柱



この JPDM メタデータは、再解析をスムーズに進めるための「jPOST 第 4 の柱」と位置づけられ、近日中に入力マニュアルも公開される予定である。

今までの開発については トーゴーの日シンポジウム 2022・ポスター 30（の特に第 1 ページ）を参照されたい。

↑ コチラ

https://doi.org/10.18908/togo2022.p030

また jPOST repository については、本シンポジウム・**ポスター 25** を参照されたい。

再解析自動化に向けて——メタデータ変換

プロジェクト分割の流れ

全生データファイルについてメタデータの各項目を比較

→ メタデータが完全に同一になるもの毎に新プロジェクト

→ biological replicate は同一プロジェクト内・別データセット

- 実際に JPDM の（メタ）データから変換を行ったところ、今まで手作業で行ってきた処理を、概ね自動化できた。
- 但しエラー発生箇所もあり、それらはいずれも、メタデータの書式が想定から変更されていた箇所だった（右図）。
- JPDM 査読プロセスのこれ以上の強化は困難であるため、別途対策が必要であり、そのために「入力補助」ツールを開発中である。

エラーの発生したメタデータの例

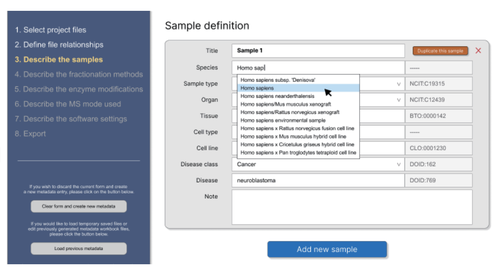
A	B	C	D	E
	Subcellular_1	Subcellular_1_Ontology	Subcellular Fraction_1	Subcellular Replicate_1
.raw				
.raw				
.raw				
.raw				
.raw				
.raw				
.raw				
0.raw				
1.raw				

正しくは「Subcellular_1_Fraction」「Subcellular_1_Replicate」の筈が、「Subcellular Fraction_1」「Subcellular Replicate_1」となっている

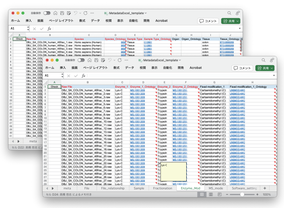
メタデータ入力の補助

2 種類の入力補助ツールを開発中である：

1. jPOST repository のメタデータプリセット入力画面と同様のインターフェースを持った、JPDM メタデータ入力用 web ツール（下図）
2. 入力内容チェックツール（項目名の誤りや、必須項目が抜けている場合にそれを指摘する。JPDM 論文の査読者側も利用できることを想定）



jPOST repository でのメタデータ入力画面と同様、ドロップダウンリスト・インクリメンタルサーチなどを備えたインターフェース



これらのツールは、今年度中にプロトタイプを動作させることを目標としている。

謝辞

jPOST の研究・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。

また JPDM の発行には、日本学術振興会の科学研究費補助金・研究成果公開促進費（国際情報発信強化 (B)）の支援を受けた（採択課題番号：21HP2004）。

Licensed under a Creative Commons 表示4.0国際ライセンス
©2024　吉沢 明康（新潟大学）