

## 2023年7月度（第404回）ライフサイエンス分科会

開催日時：2023年7月6日(木) 13:20～13:45

開催場所：Web開催、LiveViewing

参加人数：約160名

内容：第20回情報プロフェッショナルシンポジウム INFOPRO2023 内特別企画  
医学分野のプレプリントの文献データベース収録

発表者：旭化成メディカル株式会社 廣谷 映子 氏

記入者：(一財)国際医学情報センター 土屋 志保

### 1. 内容

#### ●OUG ライフサイエンス分科会についての説明

#### ●背景、目的：

文献データベースへのプレプリント収録状況を調査・検証し今後の文献データベースの活用に繋げる

#### ●プレプリントとは：

学術雑誌に投稿し査読を受ける前の論文原稿で査読前論文とも呼ばれる。

#### ●調査内容：

プレプリントを収録している文献 DB である JDreamPREPRINT①、J-GLOBAL②、EmbasePreprints③、MEDLINE④の4つで収録状況を比較する。

#### ●調査対象となる文献 DB へのプレプリントの収録：

プレプリントサーバーの bioRxiv、medRxiv はいずれの DB も収録対象としている。③以外は arXiv も対象としていて、④では ResearchSquare などその他のプレプリントサーバーも対象としている。

#### ●結果：

- ・収録件数にはばらつきがあり、COVID-19 の調査では、①251件、②254件、③108件、④43件、hypertension では、①11件、②11件、③2件、④7件であった。
- ・①と②では、ハイフン「-」の検索仕様の違いによる検索結果の違いを認めた。
- ・①と②は同じデータと思われたが、両データベースに更新タイミングの差があることを確認した。
- ・データベースへの収録にタイムラグのあることを確認した。
- ・データベースによって「出版年」や「タイトル」が異なるものがあった。
- ・①と②で収録対象だが時間が経過しても収録されないデータが存在した。
- ・④に収録されないプレプリントは、収録要件である NIH の研究助成されていないものと思われた。

## ●考察

- ・ハイフン有無などの検索仕様に対する注意が必要なことを確認できた。
- ・データベースごとに収録タイミングの違いのあることを認識する必要がある。
- ・プレプリントのバージョン情報、また、掲載や更新時期を認識できる仕組み作りなど利用面での工夫がデータベース側にあるとよいと考えた。
- ・データベース構築におけるデータの内容の提示（透明性の向上）が望ましいと考えた。

## 2, 質疑応答

- ・重複確認の方法は doi か

→doi のほか、タイトルの一致 (JDreamPREPRINT と J-GLOBAL は一致する) などから確認した

- ・MEDLINE について：使用 DB は何か

→Dialog を使用した

- ・DB 側への要望は

→利用者向けに収録方法やデータ構成などの情報開示・利用しやすくなるとよい

- ・年など今後の DB への対応について提案は

→例えば ISSN に種類があるように年にも種類を持たせるなど

- ・年違いの具体的な年は？

→収録そのものが 2020 年以降などのため偏りがあるが、2022 や 2021、など。EmbbasePreprints は遡及収録があるため 2020 などもある。

- ・タイムラグの大きい DB はどれか (先行データ調査の参考にしたい)

→明言できないが今回の結果では Embase は遅れても比較的しっかり収録され、JDream と J-GLOBAL に遅いものがある印象。MEDLINE には追加収録がなく NIH 研究助成が無いことから収録されないものはされず、情報源として少ないが収録そのものが早いように思われた。

- ・doi が識別子として存在するがプレプリントにおいてはどのように考えるか？

→プレプリントはバージョン変遷を考慮した doi、別識別子が検討されていると聞いている

3, 次回以降の予定

8月：休み

9月21日(木)：未定

以上