

機関リポジトリ担当者はAIと仲良くなれるのか

今満 亨崇

2026年3月

「研究成果のダウンロード数が爆増している……一体何があった？」

2024年1月のことである。アジア経済研究所の機関リポジトリ (Academic Research Repository at the Institute of Developing Economies。以下、「ARRIDE」という。)の利用統計を確認していた筆者は頭を抱えることになる。2023年12月のコンテンツダウンロード数は、それまでのダウンロード数と比べて約3~4倍となっていた。またこの傾向は1ヶ月で終わるものではなく、現在も増加したままの水準が維持されている。

当時は何か新しいボットが登場し、ARRIDEからデータ収集をしているのだろうと推測した。ボットとは自動処理をするプログラムの総称であるが、本記事ではweb上の情報を自動で収集する「クローラー」と同義で扱う。ボットは例えば、個々のwebページがGoogleやBingといった検索サービスでヒットするようになったり、機関リポジトリに登録された論文が大手文献検索サービスで検索できるようになったりといった場面で役に立っている。

さて、ARRIDEの急なダウンロード数増加の原因がボットであると考えた根拠は次のとおりである。

①ダウンロード数の詳細を確認すると、全てのコンテンツのダウンロード数がおおよそ一律に増えていた。いずれかの研究成果が注目されたのであれば特定のコンテンツのダウンロード数のみが増えるはずである。

②ARRIDEは日本の学術機関が広く導入しているJAIRO Cloudという機関リポジトリ構築クラウドサービスを利用している。これが2023年秋に大規模アップデートされた際に、「研究成果の可視性を向上」するための機能強化がなされている(国立情報学研究所オープンサイエンス基盤研究センター n.d.)。

③ JAIRO Cloudの利用統計はボットからのアクセスを除外して集計する仕組みとなっているが、新しいボットへの対応は遅い¹。

サーバのアクセスログを確認できないので断言はできないが、状況証拠としては十分だろう。ボットとはいえダウンロード数が増加すること自体は歓迎できる。どこかの検索サービスでヒットするようになるのはもちろんのこと、可視化やキュレーション等のwebサービスで利用されたり、統計的な分析の元データとして活用されたりして、最終的に研究成果の広報につながる可能性があるためである。

「そういえば最近システムが安定してないな」

ところで時期を同じくして、ARRIDEにアクセスしても画面が表示されなかったり、エラー画面が表示されたり、コンテンツの新規登録に失敗したり、といったことが増え始め、現在も継続している。JAIRO Coud導入機関が参加するメーリングリストを見ても、似たような報告・対応依頼が多く寄せられており、2023年秋のアップデートが良くなかったのではないかとする空気感がそこにはあったと思う。筆者も「あれだけ大きな変更をしたのだから仕方ない。システムの様子を観察しつつトラブルの生じない運用(コンテンツの登録や設定変更)をしよう」と考え、実施している²。

「AIボット、お前だったのか」

当初は上記2つの事象を別のものとして捉えていたが、2025年6月以降に公表された複数のレポート等を読んで認識を新たにした。ボットはwebサービスの運用に悪影響を与えない頻度で情報を取得するのが基本的なマナーであるが、両方の事象はそれを無視し、サーバが処理できなくなるほどの高頻度でアクセスしてくるAIボットが主な原因であると強く疑っている³。

COAR(2025)はCOAR(Confederation of Open Access Repositories)がメンバー機関に対して実施したアンケートの調査結果であり、世界中の66機関が回答している。AIボットの影響で多くの機関リポジトリに、週1回以上の頻度で速度低下やサービス停止などの不具合を生じていることが明らかになっている。

田辺(2025)は上記のCOAR(2025)の結果を示しつつ、ボットの仕組みやその一般的な対策、またJAIRO Coudのボット対策の現状を、非システム担当者にも分かりやすく解説している。また、単純にボットのアクセスを制限するのは、オープンアクセス及びCC(Creative Commons)ライセンスの観点から難しいとの考えを示している。

Jason(2025)らはノースカロライナ大学チャペルヒル校図書館が経験した、ボットによる高頻度のアクセス⁴とその対応事例を報告している。OPACから始まり、様々なデジタルコレクションが断続的なサービス停止に追い込まれたとのことである。また対策の中で得られた知見の一つとして、図書館や文書館が管理するシステムは高度な検索機能やデータ構造を有しており、コレクションとしての価値が高く、さらにオープンアクセスを重視する姿勢により、高頻度にアクセスしてくるボットに対して特に脆弱である可能性を示している。

COAR(2026)は上記のような状況を踏まえて、高頻度にアクセスしてくるボットへ対応しつつ、正当なユーザー（これには人間だけでなく常識的なアクセス頻度のボットも含む）のアクセスを妨げないためのアドバイスとサポート情報をまとめている。ただし、完全な対策は存在しない([there is no “silver bullet” solution to this problem.])事を強調しており、この事象への対策の難しさを物語っている。

「ところでLibrahack事件と似ているな」

ボットの大量アクセスにより図書館のwebサービスがダメージを受けるという単純化した構図で見ると、Librahack事件と同じといって良いだろう。Librahack事件は2010年に日本で発生した事件で、図書館のOPACから情報収集するボットを作成した利用者が逮捕・勾留された事件である。ボットによりOPACがサービス継続不可になる状況は確かに発生したが、最終的には常識的な頻度のアクセスを処理しきれないOPAC側のシステムに不備があったことが明らかになっている(LIBRAHACK 2011)。この事件はマスコミ報道されたこともあり図書館業界の内外から、当事者の図書館、システムベンダー、警察それぞれの対応の不備が指摘されている。例えば図書館に対しては、システムの事前対策が不十分な状態で警察に相談していることや(上原 n.d.)、ICTやwebへの無理解とシステムベンダーへの依存(新 2011)が課題として指摘されている。このような図書館への指摘は、今回の問題への対応を考えるにあたり参考になる。

おわりに

AIは非常に便利である。AIサービスを構築する者も、大企業から個人開発まで様々である。機関リポジトリ担当者としては、研究成果の活用を促すツールとして発展を期待しており、その全てとぜひ仲良くしていきたい。

一方で、AIはその学習に大量のデータを必要とする。データ収集のために見境なく、高頻度にアクセスをしてくるAIボットは、機関リポジトリの利用統計のノイズとなり、またwebサービスを維持できないほどの過負荷をサーバへ与えていると推測される。このようなボットに対して図書館は、技術的に妥当な対策を実施し続け、仲良くする努力をしていくことになるだろう。当館の場合はARRIDEをクラウドサービスで構築している都合上、サービス提供元に対してCOAR(2026)で挙げられている対策の実施を要望することが最初にできる対応である。

ただ、技術的に妥当な対策を全て実施し、また専門家等の第三者から見てもそれ以上の対策がない状況になってしまったら、公的権力の活用も考慮しなければならなくなる。安直に頼るべきものではないし、そのような事態になる可能性も限りなく低いが、そこまで険悪な仲にならないことを切に願っている。

著者プロフィール

今満亨崇（いまみつみちたか） アジア経済研究所学術情報センター図書館情報課。担当は図書館の管理するシステム全般。最近の活動に日本図書館協会開催の研修講師や「情報技術の発展と図書館の情報サービスの安全性、利便性 ——アジア経済研究所図書館の対応——」(情報の科学と技術 75巻10号)などがある。

参考文献

- 新出 2011. 「Librahack事件と図書館」. Internet Archive.
<https://web.archive.org/web/20250521024056/http://www.nal->

lib.jp/events/reikai/2010/277_atarashi.pdf, (参照 2026-02-11).

- 上原哲太郎 n.d. 「JANOG的にLibrahackを解説する」 .
https://www.janog.gr.jp/meeting/janog27.5/doc/janog27_5-librahack-pub-uehara.pdf,
(参照 2026-02-11).
- 国立情報学研究所オープンサイエンス基盤研究センター n.d. 「JAIRO Cloud (公開基盤)」 .
<https://rcos.nii.ac.jp/service/weko3/>, (参照 2026-02-11).
- 田辺浩介 2025. 「AIロボットと機関リポジトリCOAR Annual Conference 2025での発表と議論」 .
<https://doi.org/10.34477/0002000664>, (参照 2026-02-11).
- COAR 2025. "Open repositories are being profoundly impacted by AI bots and other crawlers: Report from a COAR Survey". <https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>, (参照 2026-02-11).
- COAR 2026. "Mitigating the impact of AI bots". <https://coar-repositories.org/news-updates/mitigating-the-impact-of-ai-bots/>, (参照 2026-02-11).
- cont_system NA 2025. 「jairo-crawler-list」 .
『Bitbucket』 . <https://bitbucket.org/niijp/jairo-crawler-list/src/master/>, (参照 2026-02-11).
- Jason, Casden. et al. 2025. "Mitigating Aggressive Crawler Traffic in the Age of Generative AI: A Collaborative Approach from the University of North Carolina at Chapel Hill Libraries". *code4lib JOURNAL* 61, <https://journal.code4lib.org/articles/18489>, (参照 2026-02-11).
- LIBRAHACK 2011. 「Librahack : 容疑者から見た岡崎図書館事件」 . <http://librahack.jp/>,
(参照 2026-02-11).

注

1. 除外対象のリストは2015年来同じものを利用し続けている。なお2025年には「更新頻度は年1回を予定」とアナウンスされている。
2. 例えばARRIDEへのコンテンツ登録には2つの方法がある。画面の指示に従い必要なデータを入力・設定していく方法と、必要なデータをzipファイルでアップロードする方法である。後者の方がシステムへのデータ処理回数が少ないのでトラブルも少なくなると考え、また経験則とも合致するので、そちらの方式を主に使用している。
3. ここで挙げるレポート類でも、AIロボットが原因のように扱ってはいるが、それを断言することは避けられているように思う。特にCOAR(2025)では、"While there is no way to be 100% certain of the purpose of these bots, the assumption in the community is that they are AI bots gathering data for generative AI training"([ロボットの目的を100%確実に把握する方

法は無いが、コミュニティでは、AI訓練用のデータを収集するAIロボットであると想定している])と記されている。

4. 事実上DDOS攻撃(分散型サービス拒否攻撃：複数拠点からの大量アクセスにより、webサービスが正常に動作しなくなるようにする攻撃)であると記述されている。