

OpenAlexを使ってみた —NIMSの場合—

物質・材料研究機構

National Institute for Materials Science

小野寺千栄

きっかけ

- フランスのソルボンヌ大学では、Web of Scienceを中止し、文献情報ツールとしてOpenAlexを利用
- 9月に開催されたScholAgoraのOpenAlexワークショップで、AIを使った分析が紹介され、多くの可能性を感じた



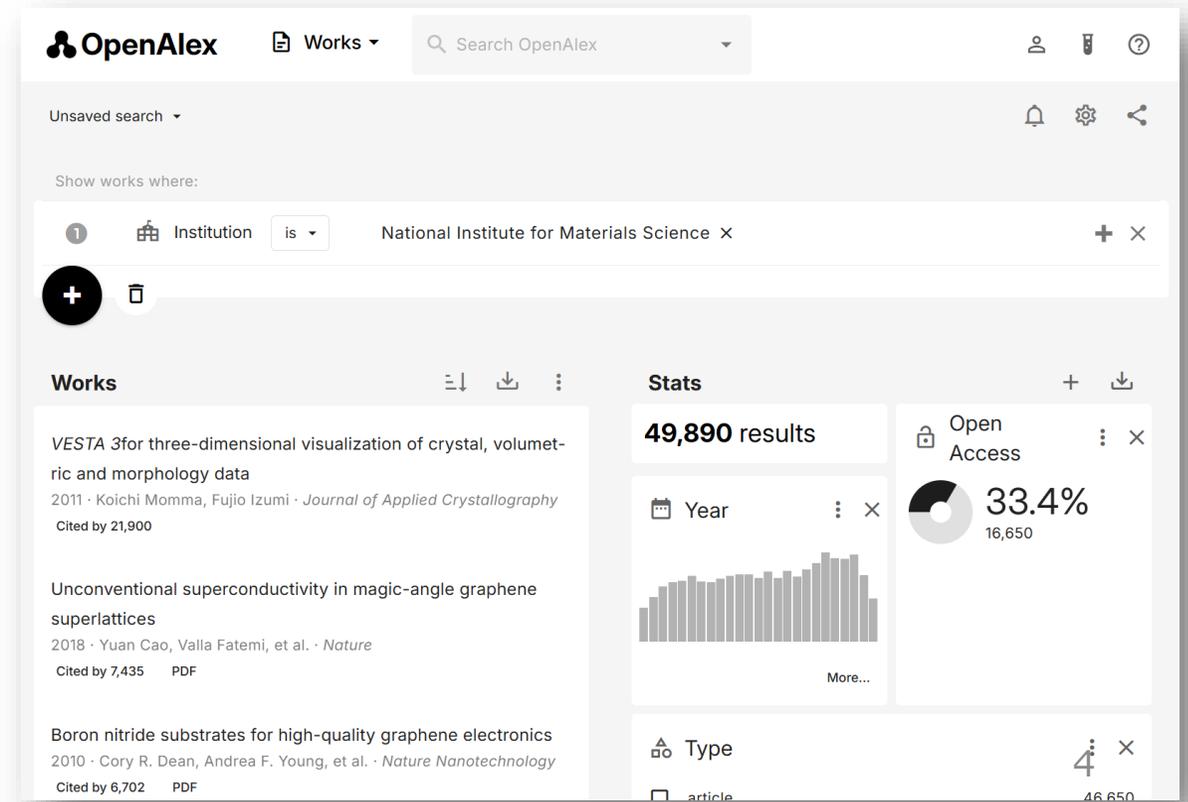
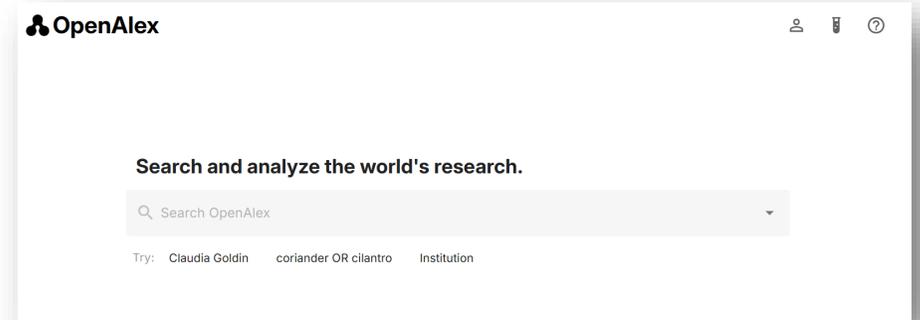
機関契約のオンラインリソースについて、一次情報（電子ジャーナル）にはILLやPay Per View等、入手の代替手段があるため、購読キャンセルの検討対象としてきた。二次情報（文献データベース）には有効な代替手段がなく、購読キャンセルは難しいと思っていたが、OpenAlexで代替できるかも？

NIMSで利用している文献データベースとの比較

| | 文献DB A | 文献DB B | Google Scholar | OpenAlex |
|-----------------|------------------|-----------------|-------------------------|-------------------|
| 価格 | △ | × | ◎ 無料 | ◎ 無料 |
| 検索条件の 保存 | ○ | ○ | × (文献の保存、 アラートのみ) | ○ |
| 検索結果の ビジュアル化 | ○ | ○ | × | ○ |
| ブラウザでの データ抽出 | ○ (最大20,000件) | △ (最大1,000件) | × | ◎ (最大100,000件) |
| API | ○ | × (別料金) | ○ | ○ |

さっそくOpenAlexを使ってみた

- 検索画面はとてもシンプル
 - トップ画面での検索条件の設定は限定的？
 - NIMSの研究成果は49,890件収録。
全件のダウンロードに十数分。
データの項目数は183！
ファイルサイズはTSVで514MB,
Excel (xlsx)でも124MB。
- ✓ ChatGPTにPythonスクリプトを作ってもらい、必要な項目に限定してAPIでデータを取得。



OpenAlexは10月に大規模な更新を実施中

- OpenAlex、Unpaywall、Unsubのデータ統合
(今後はOpenAlexのデータをもとに、Unpaywall、Unsubが動く)
- 10月中は、現在のOpenAlexと新しいOpenAlex (コードネーム: Walden) の両方を利用することができる

✓ NIMS成果の収録状況の新旧比較

| Type | 現行版 | Walden |
|--------------|--------|------------|
| article | 46,657 | 47,412 |
| book-chapter | 1,091 | 1,096 |
| preprint | 762 | 732 |
| review | 480 | 473 |
| other | 244 | 447 |
| dataset | 189 | 765 |

datasetが大幅増

The screenshot shows the OpenAlex website interface. At the top, there is a search bar with the text 'Search OpenAlex' and a user profile icon labeled 'Walden'. Below the search bar, there is a section for 'Show works where:' with a filter set to 'Institution is National Institute for Materials Science'. The main content area displays a list of works, including 'VESTA 3for three-dimensional visualization of crystal, volumetric and morphology data' and 'Unconventional superconductivity in magic-angle graphene superlattices'. On the right side, there is a 'Stats' panel showing '51,400 results' and a 'Year' bar chart. Below the stats, there is a 'Type' filter showing 'article' (47,410) and 'book-chapter' (1,096).

OREO OpenAlex Home

Works Tests
Total pass rates of key tests across the full sample set

Goodness ▾ Category ▾

60 tests Sort: Size ▾

| | | | |
|--|-------|---|---|
| | 40% 😊 | New Works other | Works that are in Walden but not in prod |
| | 15% 😊 | OA Status Changed open access | The <code>open_access. oa_status</code> field changed to a value other than <code>gold</code> |
| | 15% 😊 | FWCI Changed (10%) citations | The <code>fwci</code> field changed by more than 10% |
| | 14% 😊 | Locations Count Increased locations | The <code>locations_count</code> field increased |
| | 14% 😊 | Referenced Works Count Increased citations | The <code>referenced_works_count</code> field increased |
| | 12% 😞 | Primary Source Changed sources | The <code>primary_location.source.id</code> field had a value and changed |
| | 8% 😞 | Primary Source Lost sources | The <code>primary_location.source.id</code> field had a value but now is null or missing |
| | 8% 😊 | is_oa Added open access | The <code>open_access.is_oa</code> field became <code>true</code> |
| | 7% 😞 | Citations Count Decreased citations | The <code>cited_by_count</code> field decreased |
| | 6% 😊 | Primary Source Added sources | The <code>primary_location.source.id</code> field was missing but now has a value |
| | 6% 😞 | Locations Count Decreased locations | The <code>locations_count</code> field decreased |
| | 6% 😊 | Citations Count Increased citations | The <code>cited_by_count</code> field increased |
| | 5% 😊 | PDF URL Added open access | The <code>best_oa_location.pdf_url</code> field was missing but now has a value |

有料文献データベースとの比較：NIMS論文

● NIMS論文数

| | 2020年 | 2021年 | 2022年 | 2023年 | 2024年 | 2025年 |
|--------|-------|-------|-------|-------|-------|-------|
| クラシック | 2,406 | 2,242 | 2,242 | 2,348 | 1,796 | 1,143 |
| Walden | 2,277 | 2,097 | 2,331 | 2,345 | 1,762 | 1,135 |
| 文献DB A | 2,092 | 2,010 | 1,968 | 1,890 | 1,814 | 1,524 |

● 資料種別

- OpenAlex : [article](#), [book](#), [book-chapter](#), dataset, [editorial](#), [erratum](#), [letter](#), other, paratext, preprint, report, peer-review, [retraction](#), [review](#)
- 文献DB A : [Article](#), [Book](#), [Book chapter](#), Conference paper, Data paper, [Editorial](#), [Erratum](#), [Letter](#), Note, [Retracted](#), [Review](#), Short survey

有料文献データベースとの比較：NIMS CA論文

- Corresponding Author論文数

| | 2020年 | 2021年 | 2022年 | 2023年 | 2024年 | 2025年 |
|--------|-------|-------|-------|-------|-------|-------|
| クラシック | 1,622 | 1,527 | 1,445 | 1,432 | 943 | 293 |
| Walden | 1,568 | 1,460 | 1,443 | 1,432 | 929 | 293 |
| 文献DB A | 670 | 698 | 671 | 633 | 627 | 519 |

- 資料種別

- OpenAlex : article, book, book-chapter, dataset, editorial, erratum, letter, other, paratext, preprint, report, peer-review, retraction, review
- 文献DB A : Article, Book, Book chapter, Conference paper, Data paper, Editorial, Erratum, Letter, Note, Retracted, Review, Short survey

有料データベースとの比較：キーワード検索

- NIMSが公開しているデータベースPoLyInfoに言及している論文等
 - OpenAlex
 - ブラウザでは、検索対象項目は“Title & Abstract”に限定して検索：39件ヒット
 - APIで検索：134件
 - 文献DB A
 - タイトルと抄録を対象に検索：18件
 - 全項目を対象に検索：279件
(うち50件はOpenAlexのAPI検索でもヒット)

OpenAlexの183項目を活用する：APC

- NIMSのCorresponding Author論文のうち、Gold OAのために支払われたAPCは？

| | CA論文数 | Gold OA論文数 | APC支払い総額 (USD) |
|-------|-------|------------|----------------|
| 2020年 | 1,622 | 300 | 606,532 |
| 2021年 | 1,527 | 328 | 700,579 |
| 2022年 | 1,445 | 305 | 644,434 |
| 2023年 | 1,432 | 264 | 588,964 |
| 2024年 | 943 | 236 | 571,464 |
| 2025年 | 294 | 89 | 228,790 |

OpenAlexの183項目を活用する：出版社

- NIMS論文はどの出版社での出版が多いのか？
(文献データベースでは、出版社名は表記ゆれが多い)

| | 出版社 | 出版論文数 (2020年～2025年) |
|----|--|------------------------|
| 1 | Elsevier BV | 1,732 |
| 2 | American Chemical Society | 1,693 |
| 3 | Wiley | 1,139 |
| 4 | | 1,120 |
| 5 | Nature Portfolio | 953 |
| 6 | American Physical Society | 800 |
| 7 | Royal Society of Chemistry | 711 |
| 8 | American Institute of Physics | 465 |
| 9 | Multidisciplinary Digital Publishing Institute | 392 |
| 10 | Institute of Physics | 360 |

OpenAlexの183項目を活用する：CA論文×出版社

- NIMSの研究者（Corresponding Author）はどの出版社で出版しているのか？

| | 出版社 | 出版論文数 (2020年～2025年) |
|----|--|------------------------|
| 1 | Elsevier BV | 1,613 |
| 2 | American Chemical Society | 1,434 |
| 3 | Wiley | 890 |
| 4 | Royal Society of Chemistry | 541 |
| 5 | | 424 |
| 6 | Nature Portfolio | 382 |
| 7 | Multidisciplinary Digital Publishing Institute | 362 |
| 8 | Taylor & Francis | 249 |
| 9 | American Institute of Physics | 186 |
| 10 | Institute of Physics | 158 |

OpenAlexの183項目を活用する：CA論文×ジャーナル

- NIMSの研究者（Corresponding Author）はどのジャーナルで出版しているのか？

| | ジャーナル | 出版社 | 出版論文数 (2020年～2025年) |
|----|--|----------------------------|------------------------|
| 1 | Nano Letters | American Chemical Society | 263 |
| 2 | ACS Applied Materials & Interfaces | American Chemical Society | 161 |
| 3 | ACS Nano | American Chemical Society | 140 |
| 4 | | | 129 |
| 5 | Nature Communications | Nature Portfolio | 128 |
| 6 | The Journal of Physical Chemistry C | American Chemical Society | 110 |
| 7 | Science and Technology of Advanced Materials | Taylor & Francis | 97 |
| 8 | Journal of Alloys and Compounds | Elsevier BV | 97 |
| 9 | Acta Materialia | Elsevier BV | 88 |
| 10 | Journal of Materials Chemistry A | Royal Society of Chemistry | 87 |

OpenAlexの183項目を活用する：Has Repository Fulltext

- NIMS論文がフルテキストで公開されているリポジトリはどこ？
(NIMSのリポジトリ MDR: Materials Data Repositoryが出てきてほしい)

| | リポジトリ | NIMS論文数 (2020年～2025年) |
|----|---|--------------------------|
| 1 | arXiv (Cornell University) | 1,522 |
| 2 | PubMed | 1,162 |
| 3 | PubMed Central | 698 |
| 4 | HAL (Le Centre pour la Communication Scientifique Directe) | 508 |
| 5 | DOAJ (DOAJ: Directory of Open Access Journals) | 387 |
| 6 | OSTI OAI (U.S. Department of Energy Office of Scientific and Technical Information) | 309 |
| 7 | Research Square (Research Square) | 145 |
| 8 | Europe PMC (PubMed Central) | 54 |
| 9 | eScholarship (California Digital Library) | 46 |
| 10 | ePrints@IISc (Indian Institute of Science) | 34 |

mdr.nims.go.jp を見つけたい

- 最初にブラウザでNIMSの研究成果49,890件をダウンロードしており、その中には確かに mdr.nims.go.jp という値がある。
(項目名 : locations.landing_page_url, open_access.oa_url, best_oa_location.landing_page_url, best_oa_location.pdf_url, locations.pdf_url)
- ブラウザのLocationsタブ、APIデータには入っていることを確認。
 - 論文A
 - ブラウザ : <https://openalex.org/works/W4391341654>
 - API : <https://api.openalex.org/w4391341654>
 - 論文B
 - ブラウザ : <https://openalex.org/works/W4393030802>
 - API : <https://api.openalex.org/w4393030802>
- いろいろな条件でスクリプトを作ってみたが、素人×ChatGPTによる即席スクリプトでは mdr.nims.go.jp を含むデータは取得できず 🙄

感想

- + 有料文献データベースには収録されていない情報も多い
- + リポジトリ由来のデータが増えることを期待

- ブラウザではなく、APIによるデータ取得のほうが良さそう
- APIを利用するためのハードル
- 論文情報が収録されるまでには少し時間がかかる？