

Open Access Basics: Was ist OpenAlex?

Philipp Falkenburg

Veröffentlicht August 2022

Zitiervorschlag

Falkenburg, P. (n.d.). In *Open Access Brandenburg*. Open Access Brandenburg. <https://doi.org/10.59350/qs73-k2s86>

Schlüsselwörter

OA Basics, Nachweissysteme, Open Access, Open Data, OpenAlex

Urheberrecht

Copyright © Philipp Falkenburg 2022. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

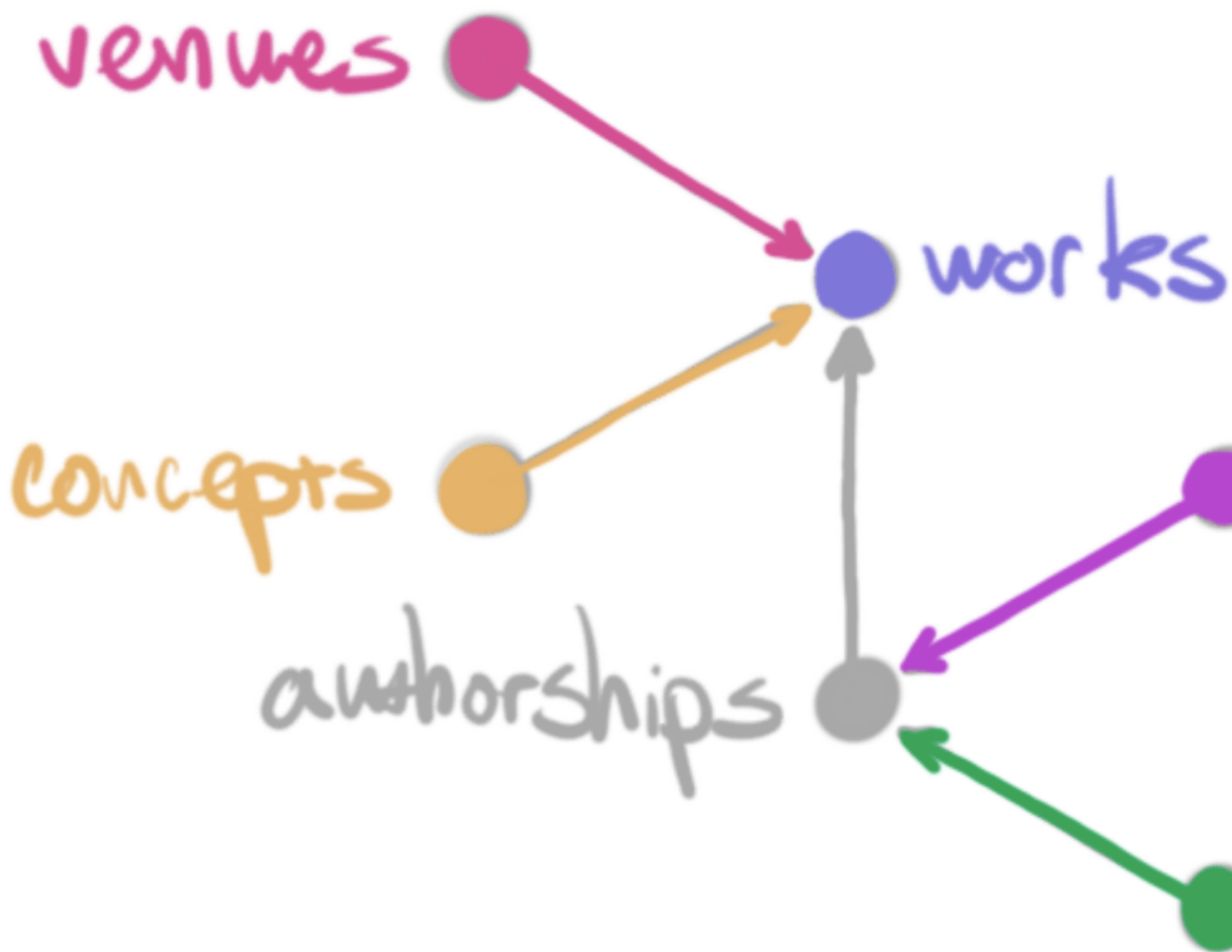
Dieser Blogpost behandelt die Grundlagen des zum Anfang des Jahres gestarteten Dienstes [OpenAlex](#). Er beantwortet die Frage, was OpenAlex überhaupt ist, beleuchtet die Hintergründe der Entstehung und versucht eine erste Einordnung für den aktuellen Ausbaustatus. Ein weiterer Fokus bildet das zugrunde liegende Modell globaler Wissenschaft und dessen mit Metadaten beschriebene Bestandteile als so genannte Entitätstypen.

Von Microsoft Academic zu OpenAlex

Als Microsoft im Februar 2016 mit Microsoft Academic eine Suchinfrastruktur für wissenschaftliche Literatur freischaltete, waren die Erwartungen riesig. Ein zeitgemäßes, ambitioniertes Angebot, das noch dazu im Gegensatz zu Google Scholar seine Datenbasis, den Microsoft Academic Graph (MAG), öffnete, bot das Potenzial, digitale Nachweisstrukturen für die Wissenschaft noch einmal neu aufzustellen. Besonders relevant war die Umsetzung des "Open Data"-Prinzips. Auch wenn Microsoft Academic nicht alles einlöste, was man sich 2016 versprach, wurden die APIs stark genutzt. Umso ernüchterter war die Community, als im vergangenen Jahr durchsickerte, dass Microsoft den Dienst einstellen und die APIs zum 31. Dezember 2021 deaktivieren würde. [Damit fiel der MAG zwar mit Ansage, aber doch über Nacht weg](#). Ein Begleiteffekt war und ist die neu entflammte Diskussion über die Abhängigkeit der Bibliotheks- und Wissenschaftslandschaft von kommerziell betriebenen Dienstleistungsangeboten. Konzeptionell ist diese Abhängigkeit gar nicht nötig, denn die benötigten Werkzeuge, Methodologien und grundlegenden Paradigmen sind der wissenschaftlichen Community lange bekannt: Open-Source-Software*, bibliometrische Verfahren und das Paradigma der Openness bezüglich (Meta-)Daten. Woran es aber im öffentlichen Sektor oftmals mangelt, ist die nötige langfristige finanzielle Ausstattung, um einen solchen Dienst skalierbar und verlässlich zu betreiben. Dazu addieren sich aufwendige Wege durch Abstimmungs- und Antragsverfahren zur Realisierung eines nachhaltigen, globalen Informationsangebots, einschließlich der dafür nötigen Organisationsstrukturen und Personalressourcen. Möglicherweise fehlt es schließlich auch am Willen, eine solche Vision Realität werden zu lassen. Der findet sich an anderer Stelle: In die Lücke sprang die im Bereich der Infrastruktur für Wissenschaft und Wissenschaftsunterstützung tätige Non-Profit-Organisation [OurResearch](#), ermöglicht durch eine [Anschubfinanzierung von der in London ansässigen Arcadia Foundation in Höhe von 4,5 Millionen US-Dollar](#). OurResearch wurde durch Angebote wie [Unpaywall](#) bekannt. Nun übernimmt das Unternehmen den Aufbau eines Dienstes mit dem Namen [OpenAlex](#). Dieser startete fast nahtlos zur Einstellung des MAG im Januar 2022. Erwähnenswert ist, dass OurResearch angibt, [nach den "Principles of Open Scholarly Infrastructures" \(POSI\) zu arbeiten](#). Mit einem Commitment zu diesen Prinzipien ist verbunden, dass alle Entwicklungen offen, ausschließlich an den Interessen der Wissenschaftscommunity und der orientiert, nachhaltig und transparent erfolgen, was für große, zentrale Datendienste von besonders großer Bedeutung ist.

Was ist OpenAlex?

Aber was ist OpenAlex und wie unterscheidet es sich von klassischen Bibliothekskatalogen oder Discovery-Systemen? Der Name leitet sich von der Bibliothek von Alexandria ab und unterstreicht damit die Ambition, das Wissen der Zeit zu sammeln und verfügbar zu machen, ergänzt um das Prinzip der Openness. OpenAlex ist dabei kein Bibliothekskatalog, denn der Fokus auf die Wissenschaftslandschaft ist viel breiter als lediglich eine Betrachtung des Publikationsoutputs. Das zugrundeliegende Modell der globalen Wissenschaft differenziert die Entitätstypen ("types of entities") *Works, Authors, Venues, Institutions und Concepts*. Die Stärke dieses Modells sind die Verknüpfungen der einzelnen Entitäten untereinander. Durch sie entsteht ein Netzwerk aus hunderten Millionen Knoten (Entitäten) und Milliarden Kanten (Verbindungen), auf dem wiederum szientometrische Operationen ausgeführt werden können.



Open Access Brandenburg

graph of data model of OpenAlex, showing the so-called types of entity and their interconnection: Die der Da Entitätstypen.

Die Entität *Works* beinhaltet Publikationen aller Art, wie Zeitschriftenartikel, Bücher, aber auch Forschungsdaten und weitere Formen wissenschaftlichen Outputs. Eine weitere Ebene der Verknüpfungen innerhalb dieses Entitätstyps stellen die nachgehaltenen Zitationen dar, womit OpenAlex ein mächtiges Instrument für bibliometrische Analysen zu werden verspricht.

Authors beinhaltet Personen, welche an der Erstellung von *Works* beteiligt sind. Hier kündigt OpenAlex die Entwicklung eines Algorithmus für die Disambiguierung an, was ebenfalls den Nutzen im Vergleich zu bestehenden Angeboten steigern dürfte.

Venues beinhaltet die Publikationsplattformen von *Works*, also die Zeitschriften oder Repositorien.

Institutions erfasst mit *Works* affilierte Organisationen, also beispielsweise Universitäten oder andere Forschungseinrichtungen. Die Beziehung wird über die *Authors* hergestellt. Auf dieser Datengrundlage wäre also eine einrichtungsbezogene Wissenschaftsmessung möglich, was möglicherweise die oft kritisierten Fallstricke eine betont quantitativen Wissenschaftsmessung verstärken könnte (siehe dazu z.B. [hier](#)).

Concepts beinhaltet hierarchisch angeordnete, abstrakte Konzepte zur Verschlagwortung von *Works*. Damit wird der (bereits beim MAG als Gamechanger erhoffte) Schritt zu einer, semantische Beziehungen besser nachweisbar machenden, Erschließungsstruktur vorbereitet. Auch hier gibt es naturgemäß erhebliche Herausforderungen, allen voran die semantische Schärfe und Passgenauigkeit.

Allen Entitäten sind wiederum definierte Sets an Metadatenelementen zugeordnet. [Diese können in der ausführlichen Dokumentation nachgelesen werden.](#)

OpenAlex und Open Access

Aus Sicht von Open Access und des Open-Access-Monitorings ist relevant, dass die Metadatenstruktur von OpenAlex auch den Open-Access-Status der *Works* nachweist. Wo es möglich ist, werden die Volltexte verlinkt und die *Venues* – sofern zutreffend – mit der Eigenschaft "is_oa: true" ausgewiesen. Das Schema dafür ist (beispielhaft) "oa_status: "gold"". OpenAlex erfasst also, ob eine Publikation Open Access ist, welcher Art Open Access sie ist und unter welcher URL der Volltext abrufbar ist.

Die Datengrundlage

Als Datenbasis für den Start des Angebots dient der letzte veröffentlichte Snapshot der Daten des MAG. Diese wird, den Entitäten entsprechend, um Daten weiterer Quellen, allen voran [Crossref](#). Aber auch auf [ORCID \(Open Research and Contributor ID\)](#), [ROR \(Research](#)

Organization Registry), DOAJ (Directory of Open Access Journals), Wikidata u. a. wird zur Vervollständigung des Datenangebots und Verbesserung der Datenqualität zurückgegriffen. Neue Publikationen werden über Crossref anhand der DOIs identifiziert und in den Datensatz eingespeist. Dieser umfasst nach Angaben der Betreibenden von OpenAlex [aktuell über 200 Millionen Works](#) und wächst täglich um etwa 50.000.

Aktueller Stand

OpenAlex ging bereits Anfang 2022 in den öffentlichen Betrieb über, allerdings zunächst im Minimalausbau und nur via API abrufbar. Die Entwicklung eines öffentlich zugänglichen Webinterface läuft aktuell noch (einen ersten Eindruck bietet dieses ["sneak preview"](#)). Gleiches gilt für andere Bausteine. Ganz im Einklang mit dem Anspruch der "Openness" kommuniziert OpenAlex seine Entwicklungsschritte und berücksichtigt Feedback aus der Community. Bis das Webinterface bereitsteht, wird diese Community eher technisch versierte Anwender*innen umfassen. Für viele potenzielle Anwender*innen aus der bibliothekarischen Praxis dürfte die Beschränkung des Zugriffs auf die Datenbasis auf API und Download eine entscheidende Hürde darstellen. Wenn die Weboberfläche im September 2022 wie geplant bereitgestellt wird, dürfte sich die Aufmerksamkeit auf und Nutzung von OpenAlex deutlich erhöhen.

Fazit

OpenAlex verspricht, was lange Zeit schon Not tut: ein global angelegtes, offenes und nachhaltiges System zur Abbildung von Forschungsprozessen anhand der über die Entitätentypen abgebildeten Dimensionen. Dieses Versprechen stand freilich bereits 2016 für den MAG im Raum. Wirklich beurteilen lässt sich auch OpenAlex und seine Leistungsfähigkeit in der aktuellen Fassung naturgemäß noch nicht.

Dafür, dass der Start des Dienstes erst ein halbes Jahr zurückliegt und das Projektteam eine beachtliche Geschwindigkeit vorlegt, ist etwas Geduld sicher angezeigt. Angesichts der beachtlichen Datenmengen und -komplexität stehen ohnehin große Herausforderungen hinsichtlich Konzeption und Betrieb im Raum. Es lohnt sich daher auf jeden Fall, Augen und Ohren hinsichtlich OpenAlex offenzuhalten. Denn möglicherweise – beziehungsweise hoffentlich – wird OpenAlex für die Community bald eine offene Alternative zu großen kommerziellen Playern darstellen, die zugleich problematischen Entwicklungen im kommerziellen Sektor, wie etwa der des [Wissenschaftstrackings](#), etwas entgegensetzen kann.

Weiterführende Informationen

- Mailinglist für den Websitelaunch von OpenAlex: <https://ourresearch.us4.list-manage.com/subscribe?u=26fcc4e14b24319755845d9e0&id=3e44ebb4c3>
- OpenAlex Tip of the Day bei Twitter: https://twitter.com/search?q=%22OpenAlex%20Tip%20of%20the%20Day%22&src=typed_query&f=live
- Factsheet zu OpenAlex von Helmholtz Open Science: Piwowar, H., Ferguson, L. M., Schrader, A. C., Weisweiler, N. L. (2022): Open Science Factsheet No. 9 based on the 64th

Open Access Brandenburg

online seminar: OpenAlex, Potsdam : Helmholtz Open Science Office. <https://doi.org/10.48440/os.helmholtz.046>

Zitierhinweis:

Falkenburg, Philipp (2022): "Open Access Basics: Was ist OpenAlex?." DOI: [10.59350/qsb73-k2s86](https://doi.org/10.59350/qsb73-k2s86)