Full application

Completed

Discretionary Award

Application title

Enriching OpenAlex with Comprehensive Grant Metadata to Enable Open Funding Intelligence

Lead applicant

Dr Kyle Demes

Application reference

323416/Z/24/Z

Administering organisation

ImpactStory (doing business as OurResearch)

Last edited

12 September 2025, 15:17

Contents

Application summary

Application title

Enriching OpenAlex with Comprehensive Grant Metadata to Enable Open Funding Intelligence

Proposed length of funding (months)

36

Proposed start date

26/05/2025

Research subject area

Data Sciences, Tools and Technology

Administering organisation

ImpactStory (doing business as OurResearch)
500 Westover Dr. #8234
Sanford
North Carolina
27330
United States

Proposal summary

Proposal summary

Our goal is to build a sustainably open research funding and impacts database that completely displaces the need for proprietary solutions by exceeding their coverage and metadata quality. We'll do this using a multi-pronged approach leveraging existing open grant databases, ingesting grant and report data directly from funders globally, and developing new tools that integrate grant and funding impact reporting automatically into open research infrastructures. We'll deliver a fully open funding intelligence database that includes (i) text mining and funder matching of 100M+ full text works (ii) newly minted or enhanced metadata records for 10M+ grants from funders around the world (iii) linkages between grants and publications derived from trusted sources beyond acknowledgements sections, and (iv) AI-guided tools that make it easier for funders and researchers to report on funding impacts. We'll do this by leveraging the infrastructure we built to run OpenAlex, the world's most comprehensive and completely open SKG, scaling our data sources, and directly collaborating with the global community of funders to make open the default for research funding metadata.

The proposal

Proposal

- **File name** Wellcome Proposal Main Text.pdf
- **File size** 281.6KB

The uploaded file is included in this PDF starting on the next page.

Background

Research funding shapes the entire research ecosystem by empowering discovery and driving research activities intended to solve challenges and benefit societies around the world. And yet while information on research funding and its impacts is critical for being able to develop and evaluate funding strategies, it is surprisingly difficult to obtain. The problem here is that research is published in venues that aren't accountable to researchers, funders, or institutions. These publishers have no incentive to give institutions an inventory of the research published by their researchers, no incentive to verify that researchers are appropriately acknowledging their funding, and no incentive to provide funders a list of the research outputs they funded. In fact, there is a strong financial incentive for the publishers to keep that data closed so that they can sell it back to funders and institutions, creating a market where value is derived from access to information, instead of value-add products and services.

OpenAlex, the world's first comprehensive and completely open index of the world's research, solves part of this problem. For the first time researchers, academic institutions, and funders have access to a well-structured inventory of research outputs. Less than three years since its start, OpenAlex is already disrupting the academic publishing system by displacing proprietary products, enabling new innovations, and shifting the research metadata market from one based on limiting access of metadata to one based on value-add services and products on top of freely available metadata. However, OpenAlex's core content does not yet extend to funding metadata or linkages between funding and research outputs.

In its current iteration, OpenAlex only indexes grants as objects when funders have minted DOIs for their grants through Crossref– that's only 132k grants vs. 7M in Dimensions and 8M in SciVal. These grants are exclusively biased towards the handful of funders who mint DOIs through Crossref, but are also lacking most of the metadata needed for these records to be meaningfully integrated into an open knowledge graph. For instance, only 260 of those grant records (0.2%) currently have metadata on the institutions who were awarded the grant in OpenAlex, and they are all from the Austrian Science Fund. Importantly, researchers usually do not cite grant DOIs in acknowledgements sections of their papers or send DOIs of their grants to publishers when publishing, keeping these grant objects from getting linked to the research outputs they fund. The only metadata on funders and grant IDs linked to OpenAlex publications come from the creation of these linkages by upstream data providers like PubMed and Crossref. Publishers are increasingly requiring this metadata upstream but it only exists for 9.5M records (3.6%) in OpenAlex and is incomplete when it does exist.

Text mining acknowledgement sections will help, but it isn't the solution

The current standard for matching research outputs to the grants and funders who supported them is to get data mining access to fulltext from publishers and to parse the Acknowledgements sections of those publications to identify text which can be matched to known funder entities. This method works reasonably well for proprietary research analytics databases that are owned by commercial publishers (SciVal, Dimensions) because they can

restrict access to those texts for toll-access publications and even impose tight restrictions on reuse of that metadata when an organization pays the prohibitively high fees to get text data mining access to their works. This has solidified these databases as the gold standard to which other funding metadata initiatives are compared.

To address the lack of data available to them, many funders now require grant awardees to report on research outputs that were supported by specific grants. However, this reporting adds administrative burden to awardees who submit incomplete and unstructured lists of publications and typically only up to a year after the end of the grant, while publications from grant-funded projects may continue years after a grant period ends.

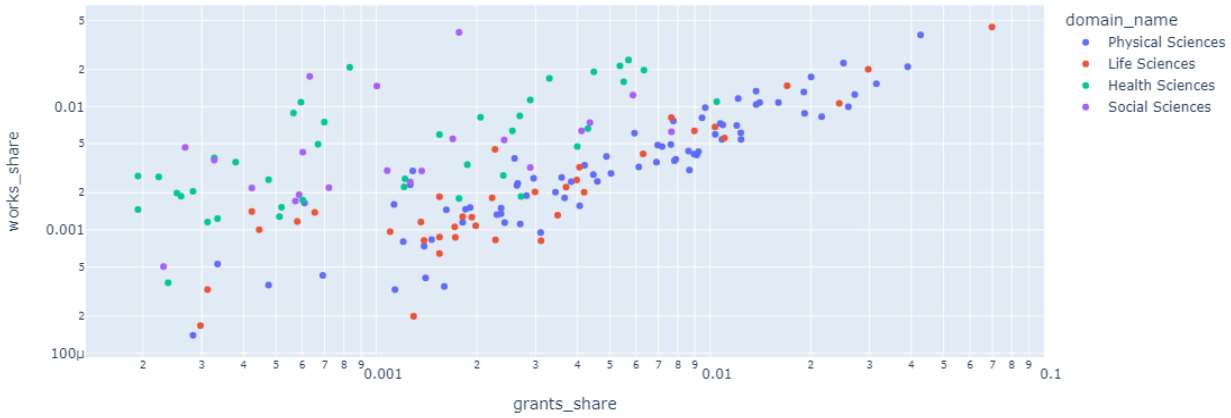
When funders compare their internal grant reports to the industry standard databases, they find that these databases are still far from meeting their needs of comprehensive impact reporting. For instance, the Canadian Institutes for Health Research (CIHR, Canada's major health research funding body) performed such an analysis using Dimensions and found that only 60% of publications had an acknowledgements section referencing them (many papers don't even have these sections); of those publications, only 30% had a grant ID number (even though they are required to include them); and only 80% of those were valid ID numbers. The conclusion being that text data mining of acknowledgements sections, even with access to all open and toll-access publications, is not the solution needed– it's just currently the only one available.

Our Pilot Study

To scope out a solution for better open funding metadata, we undertook pilot collaborations with four funders: Canada's Natural Sciences & Engineering Research Council (NSERC), the Austrian Science Fund (FWF), the Dutch Research Council (NWO), and the French National Research Agency (ANR). Each funder is committed to increasing the FAIRness of their research funding metadata and at very different places in that journey, ranging from recently implementing requirements on grant ID reporting to having already minted DOIs for all their grants. With each funder, we mapped out the status of their grant metadata, the status of their reporting metadata (e.g., which works were funded by which grants), and their needs for open funding metadata. We piloted the ingestion of their grant and reporting metadata to fully scope the work needed to clean, process, and integrate these data into the OpenAlex graph.

Grant object metadata was straightforward to ingest into OpenAlex– its structure and contents largely resembled that of publications and, in some ways, was more standardized. For instance, metadata on the researcher(s) and institution(s) awarded each grant were subsets of OpenAlex authors and affiliations in more consistent forms. And so these fields had higher disambiguation matching rates and accuracy for grant objects than they do with publications and datasets. Minimal manual curation was needed, but we have already built pipelines for curation of those fields. And because our *aboutness* (topics, fields, SDGs, etc) classifiers were built off title and abstract metadata from papers, the Project Title and Summary fields from grants also resulted in high match rates: 87% of grants were assigned to at least one Topic with Confidence and 64% of grants were confidently matched to at least one SDG (14% higher than publication average).

We then visually explored the results of these classifications by aggregating across research topics the number of publications in OpenAlex in each topic and the number of grants ingested in that same topic. We assumed a roughly positive relationship between the two, and so, empirical evidence of that would suggest our classifiers worked similarly on grants vs. publications. For instance, when looking at the share of Canadian publications in each topic plotted against the share of NSERC grants in each topic (link to analysis), we observed a tight correlation between grants and publications for topics in the Physical and Life Sciences (under the remit of NSERC, red and blue dots below) and not for Health and Social Sciences (both funded by different organizations in Canada, green and purple dots below), confirming that our topic classifiers worked well for grant metadata.



Overall, funding report metadata (e.g., research outputs funded by a grant), was more complex and less well-structured than grant metadata. This metadata is often collected from awardees copy and pasting from CVs and other sources. For each report, we parsed these unstructured texts and then tried matching text to works that are indexed in OpenAlex. Leveraging the pipelines we have developed for processing publication reference metadata proved very effective in dealing with these reports and we were able to confidently establish new linkages between grants and funded outputs. For instance, with the 15k grants for which NWO shared outputs data, we retrieved 45k works. We matched 93% of works with a DOI to OpenAlex, adding grant information to 7k publications that previously lacked this information (note: recent upgrades to our reference matching algorithm for works without DOIs and inclusion of works from more DOI registration agencies mean we'll be able to match even more moving forward).

But even for funders who already minted grant DOIs and cleaned report data to establish links between works and grants, we were still able to demonstrate additional value. FWF has minted DOIs for their grants, but (i) they have additional metadata on those grants beyond the scope of their Crossref records that they want to make open (ii) they won't mint DOIs for grants before 2015, but want to make the historical records discoverable and (iii) they have the same output tracking needs as NSERC and NWO because minting DOIs for grants doesn't necessarily make them easier to match to funded outputs. ANR still struggles with situations where incorrect grant IDs are included in acknowledgements sections of papers, faculty forget to disclose some works in their reports and those works don't have acknowledgements sections available to mine, and

when new works supported by a grant are published after the final grant report. By ingesting their metadata into OpenAlex, we will be able to provide them a list a potential match errors (same grant ID reported, but different topics, institutions, or authors) and true matches that were missed by their reporting limitations (e.g., grant periods and PI errors).

Planned activities

We're proposing a comprehensive multi-pronged approach to make OpenAlex the best source for funding intelligence, including: leveraging existing open grant initiatives, helping funders make their datasets openly available, and creating open infrastructures that make these data linkages the default moving forward. Here we present an overview of the planned activities followed by a timetable and expected dates of completion for major project milestones.

We'll ingest data from **existing open grant initiatives**, like Open Grants, EuropePMC grants, 360 Giving, and DataCite Awards to collectively add over 1M grant records and 2M linkages between outputs and grants. As part of these previous initiatives, many funders have already curated comprehensive lists of the many variations of names that researchers use to refer to funders in their publications: since those are publicly available, we'll be able to use those to enhance the existing OpenAlex string matching to make sure that funders don't have to duplicate work they've already done.

We'll then set up automatic **ingest pipelines for funders** around the world who have already made their grant databases openly available and reusable. We'll start immediately with those profiled in Elsevier's funding intelligence products, targeting 8M grant records, to help users of proprietary products transition to open metadata more quickly. But we will expand to include countries and funders underrepresented in existing databases and solicit additional sources from our user communities to rectify the North American biases in current solutions. In the first year, we will prioritize well-structured databases with many grants first and then tackle the long tail of less well-structured databases with fewer grants in subsequent years. After the first year of ingesting these databases, we'll develop a second version of these pipelines that can scale globally based on the different formats we discover from our development partner funders. For funders who have not yet started making their funding data available but wish to, we will provide a list of best practices to get them started. We will work closely with DOI registration agencies (especially Crossref and DataCite who already mint objects for grants) to make sure that all funders understand the benefits of minting DOIs for their grant records. We will encourage and facilitate the DOI-minting process, but will also continue to support funders who decide not to mint DOIs for the grant IDs (e.g., ones who use RAIDs or their own internal IDs) by ensuring that all grant objects have a unique OpenAlex ID. This mirrors our approach to publications where an OpenAlex work ID might be linked to one, multiple, or no DOIs.

Alongside the initial work to ingest grant databases, we will simultaneously start to **mine publication records for funding metadata**. We will start to parse fulltext of open access publications (including preprints later linked to publications) in OpenAlex to tag the funding and/or acknowledgements sections when they exist. For toll access publications, we will also search for this information at scale on journal landing pages and within the 30,000 unique

metadata repositories that we currently harvest for Unpaywall and OpenAlex. With that information parsed, we will set up new processes to **match funding metadata to existing funders** (via ROR records, adapted from our affiliation matching processes) and then match to existing grant IDs that already exist in OpenAlex (adapted from our citation matching processes). We expect the initial round of funder text matching to add funder linkages to at least 20M publications but will likely only match grant IDs for 1M publications, since at this point in the project we won't yet have a large corpus of grant objects to match to. Still, these pipelines will serve as the foundation for the project and we will re-run at least twice each year as we continue to add metadata for millions of grants.

While the OpenAlex technical team (Jason Priem, Casey Meyer, and the newly hired Senior Data Engineer) are working on these initial parts of the project, Kyle Demes and the newly hired Funding Community Lead will initiate funder **partner research** with our first cohort of Funders. We are aiming for 30 different funders in the first year from different countries and funder type and size to serve as development partners on the project (13 have already confirmed interest, and we are already in discussions with another 12). With each funder, we will confirm their interest in the project, signing MOUs when required (2/4 of our initial funders required this). We'll (i) scope out the current state of their funding metadata (how many grants they want to make open, what metadata they can make open, what metadata they cannot make open, the current format of their metadata), (ii) review and enhance their ROR records with accurate information on name variants to facilitate more accurate text matching, (iii) learn about their specific needs for funding and impact intelligence. We'll synthesize this information to present to funders at a **collaborative workshop** with the initial cohort of funder partners. At the workshop, we'll then collaboratively design a full spec for the project and develop a set of best practices that can be scaled in the following years as new funder partners come on board.

In the second year, we'll expand the activities of the first year: ingesting more distributed and smaller open grant databases and expanding our outreach on ingesting grant and report metadata to more funders globally. We will continue to parse metadata on outputs and continue linking those to grants indexed in OpenAlex. With our core group of funder collaborators (which may change in composition as the initiative grows), we will start to focus on **understanding their reporting needs**. Specifically, we'll fully spec out the funding intelligence and impact report analyses that they want to see automated in OpenAlex. Our development team will prototype these analyses and a custom dashboard for funding intelligence in the openalex.org user interface that meets their needs. We'll also work with funders to collaboratively develop a sustainability model with paid offerings for services that would add value to the database to help ensure the database can operate open indefinitely, like custom bibliometric analyses, new tools that facilitate reporting, and custom curation of award-output linkages years after end of grant periods. In the final year of the grant, we'll ship and deploy pipelines that automate **all of these activities.** We'll continue adding new, smaller grant databases to test these pipelines and put out a call for any funders globally to submit their grant and report metadata to these pipelines, troubleshooting until they work for all use cases.

Timeline

Year 1 (starting May 26, 2025)

- Begin hiring for all three positions [completed by Sept 1, 2025]
- Mine publication existing records for funding metadata [completed by Nov 1, 2025]
- Funder partner initial research [Completed by Nov 1, 2025]
- Funder collaborative in-person workshop 1 [mid November 2025]
- Ingest existing open grant initiatives into OpenAlex [completed by Dec 31, 2025]
- Match funding metadata to funders and grant IDs [completed by Feb 1, 2026]
- Ingest initial grant and report metadata from first partners [completed by May 25, 2026]

Year 2 (starting May 26, 2026)

- Virtual workshop with funder partners to report back on progress and validate initial findings [July 2026]
- Recruit additional funders to expand grant and report ingestion [completed Sept 1, 2026]
- Ingest grant metadata from 2nd round of open databases [completed by Nov 1, 2026]
- Funder collaborative in-person workshop 2 [mid Nov 2026]
- Complete metadata ingest directly from second round of funders [March 2027]
- Launch MVP of open funding intelligence reporting templates for funders [April 2027]

Year 3 (starting May 26, 2027)

- Collect feedback from funders on MVP of reporting templates [completed 1 Jun 2027]
- Launch automated pipelines for adding new grant databases [completed Oct 1, 2027]
- Launch automated pipelines for funders and researchers to submit unstructured report metadata for linking to grant objects [completed Dec 31, 2027]
- Launch of v2 of funding intelligence reporting templates [completed by Feb 1 2028]
- Final virtual workshop with funders to share results on implementation of reporting tools [April 1, 2027]

Why us?

OurResearch is uniquely positioned to successfully deliver this project: we have an established track record delivering similar projects, OpenAlex is the ideal home for connected and open research funding metadata, the team has demonstrated the technical expertise needed for each of this project's components, and importantly a self-organizing global community of institutions, governments, and funders are keen to collaborate with us on this project.

Co-founded by Jason Priem and Heather Piwowar, OurResearch is a non-profit that has been advancing open science through sustainably open technical solutions for the last 13 years. For instance, our product *Unpaywall* mapped copies of publications in distributed repositories against publisher records to (i) make open access works more discoverable and (ii) to allow researchers and organizations around the world to analyze trends in open access publishing and design and evaluate their OA strategies. Unpaywall is used in every major library discovery platform (e.g., Alma, Summan, WorldCat, Web of Science, Scopus, Dimensions, EBSCO EDS) and in major national OA assessments (e.g., the UK REF, the European Open Science Monitor, the French Open Access Monitor, COKI). OpenAlex, our latest solution, is the world's first comprehensive and completely open map of the world's research ecosystem. We built

OpenAlex by ingesting many disconnected open datasets, combining and enriching metadata on research outputs, and developing algorithmic approaches to connect publications to other nodes in the graph, like institutions, authors, cited papers, and much more (more on that here). Less than three years since its creation, OpenAlex has already been referenced or cited by 540+ publications, underlies three major rankings bodies, CWTS Open Leiden Rankings, Financial Times Business School Research Ranking and THE Interdisciplinary Science Rankings, serves 18M API calls a day, has 100k web users each month, and is displacing proprietary databases.

Specifically, (i) we have already demonstrated an ability to harvest grant metadata from diverse sources by adapting ingestion pipelines that currently pull metadata from 30k institutional repositories which vary in metadata quality, (2) we have already demonstrated our ability to match grant records to acknowledgements sections through our solutions that parse OpenAlex's open access PDFs to find and match variably structured reference and affiliations metadata, and (3) we have already demonstrated an ability to do this all completely openly, making 100% of the underlying data and source code available under MIT and CC0 licenses. Through our pilot studies, we have further demonstrated that each activity we're proposing will work. And finally, we have already confirmed broader interest among other funders globally.

Evaluation and Impact

The creation of a completely open research funding intelligence database will have far-reaching impacts throughout the research ecosystem. All funders and research institutions need access to this data, but it is currently only available behind paywalls that prohibit access to many and drain funds from others. After this project, all academic institutions and funders will have access to an open funding intelligence database. The open nature of this data will also enable large-scale analyses that are not currently possible using closed systems, like analyses on funding trends, gaps, and impacts across the entire globe. Perhaps most importantly, funding impact analyses using OpenAlex will allow funders to glimpse a much broader picture of their impacts than possible in existing proprietary solutions. That's because compared to the comparator proprietary solutions, OpenAlex has broader coverage of output types, open access journals, works in non-English languages, and works from all areas of scholarly activity.

By the end of this grant, we expect OpenAlex to be the gold standard for funding intelligence databases, completely displacing the need for commercial databases and including metadata from funders underrepresented in current databases. The bibliometric community is already performing structured independent research comparing the content and metadata quality of OpenAlex against proprietary solutions, but we'll put out a call to the community asking for such analyses throughout and at the end of this grant to formally evaluate the project (and have included budget for this research in this grant proposal). We will also work with funders throughout the project to characterize the impacts of their funding that are only captured in OpenAlex to help bring even more funders to our open solution. We'll also measure the success of this project by the number of funders who cancel their costly subscriptions to commercial databases (and total $ saved) in favour of using OpenAlex for their funding intelligence and impact reporting.

Sustainability

OurResearch has demonstrated over the last 13 years that we can sustain open science solutions by making our data free and charging for services that bring value-add to clients. Once the major work building and automating our products is complete, the operational costs are very low and can be sustained in perpetuity through smaller service contracts. Because 100% of the data and source code are completely open, anyone else in the world could maintain and build off this open grant metadata solution. Throughout the course of the grant, we'll work closely with funders to understand what services they would pay for that could sustain our ability to offer the data free indefinitely. The funders in our initial pilot project have paid for such services already and given us many additional ideas. For instance, funders have almost universally expressed an interest in the commission of custom bibliometric analyses and the creation of reporting widgets that make awardee reports easier and more accurate. Imagine a PI signing into either their OpenAlex author profile or their Funder Portal and receiving a list of suggestions of outputs to confirm or deny having been funded by a specific grant based on AI embeddings of text. We'll work with funders to make sure that such solutions could be implemented in their local systems via OpenAlex's free API or directly submittable by grant PIs and then pushed to funders. We'll continue to iterate on related service offerings with funders to find a market fit that keeps the data open freely for everyone around the world. At the end of the grant, all new hires will stay on the OpenAlex team, transitioning to similar roles on other projects.

References

References are included in main text as hyperlinks.

Does your proposal involve human participants or human biological material?

No - does not involve human participants or human biological material

Describe how you will undertake your proposed research in an environmentally sustainable way.

OpenAlex is openly available for anyone in the world to use freely. This requires us to operate lean and cost effective servers that are inherently energy efficient as well. We offer users the ability to access our data through a user interface, an API, and a bulk snapshot download. Users with small specific needs can cheaply access that information through elasticsearch servers while users with computationally heavy needs can download the full dataset from an AWS bucket much more energy efficiently than hitting our API servers for large volumes of data. Users with medium-volume data needs can use our API. We are currently developing a new API offering that will work more energy-efficiently for medium-large scale data request (columnar Redshift-based API that retrieves only the information needed by the user). In terms of our non-computational operations, our team works remotely and only travels when absolutely necessary. For the work we're proposing in this grant, the vast majority of proposed meetings and workshops will be held virtually. But two conferences (one in each of Year 1 and 2) will be in person for our team of core funder partners and these workshops must be in person. We have included in the budget, funds for carbon offsets for required travel.

Are you applying with coapplicants?

Yes

Team composition and management

The team will be led by Kyle Demes (Main applicant) who is the current Chief Operating Officer of OurResearch, spending at least 20% of his time on this project. Chief Executive Office, Jason Priem, will spend at least 10% of his time overseeing the integration of this project into OpenAlex, advising on the project, and overseeing the 3 newly hired staff positions working on this project. All three new hires will be integrated directly into the OpenAlex team, attending daily stand-up and weekly all-hands meetings. The main applicant will also lead weekly team meetings specifically for this project and weekly one-on-one meetings with project staff to mentor them and ensure work meets project deliverables, troubleshooting as needed. Chief Technology Officer, Casey Meyer, will work closely with the newly hired Senior Data Engineer to ensure tight technical integration of the project and its outputs into OpenAlex and training the Senior Data Engineer to adapt the existing data pipelines that he created as the foundation of OpenAlex. Kyle Demes will work more closely with the Metadata Curation Specialist and Funder Community Lead in their daily work and will coach all three new staff in the preparation of final outputs and presentations at conferences. The project team will provide quarterly updates to include in the public OpenAlex townhalls and will report semi-annually to the group of funder collaborators on progress, challenges, and next steps, including a formal written report annually.

Outputs management and sharing

Provide an outputs management plan

The output of this project will be the addition of millions of grant objects into the OpenAlex dataset and 10s to 100s of millions of new connections between grants and their funded research outputs. Grant objects will be classified within the existing OpenAlex aboutness taxonomies (https://shorturl.at/jngZn), awardee metadata will be linked to OpenAlex Authors and Institutions, and funder metadata will be attributed to grant objects (e.g., Funder, program, grant ID, other persistent identifiers, funding amount, award years). These outputs will be added in an on-going basis with new metadata available in OpenAlex between 3 hours and 3 days after it is added to our system. The bulk of these outputs will be available before the end of the final grant year, but with the new pipelines we build, they will continue to grow for years after the project's end date. We ensure open documentation of all of the metadata in OpenAlex (https://docs.openalex.org), how they are derived (e.g., https://shorturl.at/sspOV), and a knowledge base with common FAQ on how to use and interpret (https://help.openalex.org). We keep these documentations updated after new releases on an on-going basis (< 3 months after feature launch)-- most will be in place halfway through this project. 100% of our source code is made available under an MIT license in a public github repository (https://github.com/ourresearch/OpenAlex). 100% of our data is made available freely under a CC0 license, through a public snapshot that is updated quarterly, an open API that permits 100k calls per day freely, and an open user interface. The data and source code for this project will be made freely available under MIT and CC0 licenses through the same mechanisms, since they will be integrated into the broader OpenAlex knowledge graph. The data is stored on AWS servers, where it is also freely available for bulk download. We store previous version of the snapshot for historical records– additional backups of the dataset are available through research groups around the world that provide access to local versions of the snapshot as well as other Zenodo (https://zenodo.org/records/13941458). OpenAlex has exponential growth in users from around the world, currently with 18M API calls a day and 100k unique web users each month and we'll update our users (https://groups.google.com/g/openalex-users) and followers with progress on this project regularly. We will also give webinars on new features as they launch, present at the conferences we attend (library conferences, research administration conference, open research information conferences, and scientometrics conferences), and put out whitepapers on the future of open funding intelligence to expand awareness. We are also planning joint press releases with our funding collaborators at the beginning of the project as well as with each major milestone.

Select the approach you will use to maximise the impact of your significant research outputs to improve health and benefit the wider research community. If an outputs management plan is not needed, select 'Not applicable'.

Make research outputs available for access and re-use

Collaborations

Are any collaborations essential for this proposal?

No

Location of activity

Will the funded activity take place at more than one location?

No

Wellcome Trust supported facilities

No

Will you need funds to be awarded directly to more than one location?

No

Research involving animals

Does your proposal involve the use of animals or animal tissue?

No - does not involve animals or animal tissue

Risks of research misuse

Confirm that you have considered whether your proposed research could generate outcomes that could be misused for harmful purposes.

I confirm

Have you identified any tangible risks of this type?

No

Freedom to operate and conflicts of interest

Describe any freedom to operate or other intellectual property related issues that might affect your ability to do the proposed research or to use, share or commercialise the research outputs. Explain how you will address these.

Not applicable.

As an open science infrastructure, 100% of our data and source code are available under CC0 and MIT licenses. That also means that we cannot collect and share personal information or data from any collaborators that cannot be made available under a CC0 license. Each of our collaborating funders will have a different policy specifying which of their metadata can be made available under these terms and that will limit their ability to share with us, but even the most restrictive policies will enable data sharing that can achieve the purposes of the grant.

Describe any conflicts of interest which might affect your ability to do the proposed research or to share or commercialise the research outputs.

not applicable

Lead applicant details

Full name

Dr Kyle Demes

Current role

Job title	Start date	End date – if known	Department	Organisation	Main employer
Chief Operating Officer,	01/01/2024		OurResearch	ImpactStory (doing business as OurResearch)	Yes

Previous roles

Role	Department	Organisation Name	Country	Start date	End date
Research Intelligence Consultant	Research Intelligence	Elsevier		01/11/2020	01/04/2023
Director	Institutional Strategic Awards	Simon Fraser University		01/09/2018	01/10/2020
Senior Advisor, Strategic Initiatives	Vice-President of Research & Innovation	The University of British Columbia		01/02/2016	01/08/2018
Strategic Partnerships & Research Administration	Hakai Institute	Tula Foundation		01/05/2015	01/02/2016
Postdoctoral Fellow & Instructor	Resource & Environmental Management	Simon Fraser University		01/05/2013	01/05/2015

Have you taken any career breaks?

No

Are you currently receiving any funding towards your salary?

No

Add your current and previous education and any relevant training

School	Country	Degree or qualification	Subject	Start date	End date (or expected)
The University of British Columbia	Canada	Doctor of Philosophy (PhD;Dphil)	Marine Ecology	01/09/2009	01/05/2013
California State University, Monterey Bay	United States	Master of Science (MSc)	Marine Science	01/09/2007	01/07/2009
University of South Florida	United States	Bachelor of Science (BSc)	Biology (marine)	01/09/2004	01/08/2007

Are you a healthcare professional?

No

Do you want to do this grant part time?

No

Career contributions

After a successful start to a research career in marine ecology (including the British Ecological Society's Haldane Prize), I began collaborating with governments and First Nations to design a coast-wide kelp forest monitoring program and served as an ecologist on a special report that changed the policy trajectory of the most controversial resource projects in contemporary Canadian history. Inspired by the impact of this work, I left an academic path to build bridges between academia and other sectors to catalyze more real-world impact from research. I transitioned a funder from giving grants to Universities to funding programs in communities that attracted academics to solve problems for communities. I joined the Vice-President of Research at The University of British Columbia to develop data-driven approaches enabling research excellence. Using bibliometric and research funding data, I identified barriers to collaboration and designed and implemented programs that incentivized collaboration on grand challenges. This work was awarded the 2017 Best Paper of the Year Award from Society of Research Administrators International because funders and universities around the world were looking for solutions to the ever-growing isolation of academic knowledge (https://files.eric.ed.gov/fulltext/EJ1213139.pdf). The impacts of this work became clear quickly and the University established an on-going $6M/year internal granting program based on it. I then became the inaugural Director of Institutional and Strategic Awards at Simon Fraser University, leading a team of 11 highly qualified staff to develop and implement data-driven research development programs for the University. I moved to Elsevier's Research Intelligence portfolio, consulting for 50+ North American Universities in the use of research metadata in driving institutional strategy, winning the 2021 Innovation Award from the Canadian Association of Research Administrators for this work. I am now COO for OurResearch, and proudest of my efforts to build OpenAlex, the world's first completely open and inclusive bibliometric database, ensuring everyone around the world can benefit from the insights of open research metadata. A list of my academic contributions can be found here: https://openalex.org/works?&filter=authorships.author.id:a5086928770

Describe your approach to developing and supporting a positive and inclusive research culture, including examples from previous and current groups.

Interdisciplinary collaboration has been a core focus throughout my career. This approach is inherently inclusive as it requires and places significant value on the diverse perspectives of others to drive impact. As a researcher, I bridged disciplines and then sectors learning from experts in disparate fields (law, policy, economics) to drive impact of my work in marine ecosystems.

As a research manager, I institutionalized policies and programs that incentivizes this approach throughout Universities. And then as a consultant, I listened to perspectives of diverse clients to support solutions for their needs. When EDI became federally required for institutions and researchers in Canada, it was my responsibility at both UBC and SFU to develop programs and policies that addressed systemic barriers in research. I helped research groups prepare custom-tailored EDI action plans for their specific research groups and spoke at national conferences on data-driven approaches to identifying and rectifying systemic barriers to participation in science. My approach was again driven by a commitment to collaboration with folks with diverse perspectives.

As a people manager, I continued this approach, ensuring that my research trainees and professional staff understood the value of collaborating with others and gave them the space to develop those collaborations. I also make sure to understand the individual aspirations of trainees to be able to provide them with motivating opportunities for growth towards their goals and also creating new opportunities for my team that leverage the talents of the team members. Whether in academic or professional settings, I always strive to share with trainees the challenges or projects that I'm leading to get them exposure to how I solve problems while allowing them opportunities to learn the value of their unique contributions to my work while also mentoring them through their own projects. This approach has proven successful through the success of my former students (who have won academic awards) and trainees (who have won professional recognition from their peers, including a national award for best practices in EDI in research support). Mentorship takes time and effort, but ultimately always pays off and one of the proudest moments of my professional career was receiving the 2020 Excellence in Management Award from the Canadian Association of Research Administration after my team staff nominated me.

Now at OurResearch, it is central to my mission to create a bibliometric database that is inclusive of all the diverse scholarly activity that is underrepresented in existing proprietary solutions (e.g., social sciences, humanities, arts, business, law, works in other languages, works from the global south) and to ensure that all researchers, funders, and research organizations around the world have equal access to the insights that such a database can provide.

Current and recent research funding (including Wellcome grants)

In November 2024, OurResearch was awarded a grant entitled "User Interfaces to Accelerate Adoption of OpenAlex" from The Navigation Fund for $688k over two years to enhance the Openalex User Interface. Jason Priem (OurResearch CEO) was the main applicant on the proposal and I am the administrator of the funds, spending less than 5% of my time on the project. 100% of the grant funds went to the salary of a new hire working on user interfaces.

In February 2024, OurResearch was awarded a grant from Arcadia Foundation for $7.5M over 5 years to build OpenAlex, a sustainably open index of all the world's research. Jason Priem (OurResearch CEO) was the main applicant on the proposal and I was a co-applicant. I currently spend 80% of my time on this grant.

As the Director, Institutional Strategic Awards at Simon Fraser University, I was responsible for overseeing pre- and post-award processes for all institutional research grants (grants awarded to institution and not individual researchers). In that capacity I administered research funding on behalf of the institution. I left that position, and therefore connection to those grants, in September 2020.

Describe how the currently active grants listed above relate to this application. If you hold grants related to the topic of this application, explain how these differ and confirm there is no overlap in funding.

The grants from The Navigation Fund and Arcadia are both closely related to this application. The Arcadia grant gives OpenAlex seed funding to develop and maintain an open index of the world's research. For that grant, we're working on mapping research outputs (publications, books, dissertation, datasets, etc.) to important nodes like authors, institutions, journals, etc. The Navigation Fund grant is specific to designing user interfaces that make it easier for non-technical users to interact with and benefit from the index of the world's research. Grant metadata is outside of the scope of both grants, but there are obvious synergies where the project we're proposing here would not be possible without the Arcadia grant to build OpenAlex and the UI work the UI designer from the Navigation Grant will lay the groundwork to further enhance the value of the products of this grant.

What percentage of your research time will you spend on this project?

20

Coapplicant details: Jason Priem

Full name

Mr Jason Priem

Current role

Job title	Start date	End date – if known	Department	Organisation	Main employer
Chief Executive Officer	01/01/2012		OurResearch	ImpactStory (doing business as OurResearch)	Yes

Have you taken any career breaks?

No

Are you currently receiving any funding towards your salary?

No

Add your current and previous education and any relevant training

School	Country	Degree or qualification	Subject	Start date	End date (or expected)
University of Florida	United States	Master of Arts (MA)	Social Sciences Edducation	01/05/2001	01/08/2002
University of Florida	United States	Bachelor of Arts (BA)	History	09/01/1997	01/05/2001

Are you a healthcare professional?

No

Do you want to do this grant part time?

No

Career contributions

Jason Priem has significantly reshaped scholarly communication by pioneering altmetrics—a field that seeks to capture the broader impact of research through online attention and engagement, rather than relying solely on traditional citation counts. Early in his career, his influential writings and the widely circulated "Altmetrics Manifesto" spurred a reevaluation of how academic influence is measured. He then co-founded the nonprofit, OurResearch, providing researchers and institutions with practical tools to monitor and share the diverse impacts of their work. While he's developed and launched many solutions that have advanced open science, Unpaywall is perhaps his most impactful project since it was the first tool that classified research outputs in open access categories and linked works to open versions in institutional repositories. It's still used by 700k users a month to find open access research and is used as the base in every major open access tracking initiative on the planet. His next product, Unsub, has helped libraries around the world save millions of dollars a year by providing the data necessary to unsubscribe from costly for-profit journals not being used by their researchers. His latest open science solution, OpenAlex, provides researchers and organizations around the world with the full insights of access to a global map of the world's research ecosystem for free. This product is disrupting the reliance of organizations on proprietary databases but also enabling innovation not possible with closed databases (e.g., analyses mapping global flow of knowledge; new products creating value on top of the open data; open and transparent rankings). Through these efforts, he has not only expanded the criteria for assessing scholarly contributions but also fostered a more inclusive and dynamic ecosystem for understanding research impact in the digital age.

A full list of Jason's academic outputs can be viewed here: https://openalex.org/works?filter=authorships.author.id:a5023888391

What percentage of your research time will you spend on this project?

10

Describe your approach to developing and supporting a positive and inclusive research culture, including examples from previous and current groups.

Jason has spent his career making the global research ecosystem more inclusive. He pioneered altmetrics to ensure that researchers were being recognized for contributions to work outside academia, revolutionizing research assessment. He created Unpaywall to help ensure that open access content in institutional repositories was discoverable globally, increasing access to research to users around the world. Most recently, OpenAlex aims to map all of the world's research knowledge, in comparison to existing proprietary databases that are too expensive for some to afford and intentionally exclude research outputs from the global south, in non-English languages, diverse types of research outputs (e.g., preprints, datasets, books), and research in particular fields (e.g., social sciences, humanities, arts, business, law, and computer science). This project ensures that everyone in the world has free access to a complete picture of the world's academic knowledge. As a people manager, Jason firmly believes in hiring people who are committed to similar ideals who are self-motived to achieve their own goals and gives them the space and opportunities to achieve those goals.

Currency requested

Select the currency in which you want to apply.

USD

Is this your local currency?

Yes

Costs requested and justification

Are you asking for staff?

Yes

Staff costs

Cost type	Number of staff asked for	Role	Name (if known)	Basic annual starting salary	Salary grade or scale	Months on project	% time	Cost requested from Wellcome
Salary	1	Senior Data Engineer	TBD	$275,000.00	Senior	36	100	$1,004,150.00
Salary	1	Lead, Funder Community	TBD	$90,000.00	Early Career	36	100	$325,472.00
Salary	1	Metadata Curation Specialist	TBD	$70,000.00	Early Career	36	50	$126,572.00
Salary	1	Project Lead & COO	Kyle Demes	$185,000.00	Senior	36	20	$133,805.00
Salary	1	CEO	Jason Priem	$215,000.00	Senior	30	10	$64,408.00
Salary	1	CTO	Casey Meyer	$260,000.00	Senior	24	15	$101,616.00
								Total: $1,756,023.00

Justification for staff

Three existing staff will work on this project: our COO ($185k + $31.4k fringe) will spend 20% of his time in all three years leading the project; our CTO ($265k + $45k fringe) will spend 15%, then 20%, then 10% of his time overseeing the technical implementation of new data processes and pipelines for grant metadata in OpenAlex; and our CEO ($215k + $36.5k fringe) will spend 10% of his time in all three years ensuring overseeing staff, advising on project implementation and with high-level funder relations. Three new positions will be hired with this grant. The *Senior Data Engineer* ($275k + $46.7k fringe) will be responsible for the ingestion and processing of funding and report metadata as well as the creation of new pipelines that automate these processes as scale. The *Lead, Funder Community* ($90k + $15.3k fringe) will be responsible for developing and maintaining relationships with Funders throughout the course of the project, including documenting their needs, working on MOUs/licenses as appropriate, day-to-day coordination of the project, planning the funder workshop, and serving as a key liaison between data teams at OpenAlex and within each funder. The *Metadata Curation Specialist* ($70k/year + $11.9k fringe) will be a half-time role supporting the Senior Data Engineer with high-touch data curation needs, including cleaning and pre-processing of funder metadata, developing name variant lists for funder entity text matching, QAQC of funding metadata in OpenAlex, and the review of community curation requests pertaining to funding metadata. OpenAlex will top up this position to hire a full-time person with their remaining time spent on additional community metadata curation projects (e.g., www.cometdata.org).

Are you requesting adjustment support?

No

Are you requesting training and continuing professional development?

No

Are you requesting materials and consumables?

No

Are you requesting animals?

No

Are you requesting equipment?

Yes

Equipment costs

Type	Type of equipment	Number of items	Cost for each item	Cost of maintenance contract	Contribution from other sources	Cost requested from Wellcome
Computer equipment	Standard laptop	2	$1,800.00	$0.00	$0.00	$3,600.00
Computer equipment	Performance computer	1	$2,500.00	$0.00	$0.00	$2,500.00
						Total: $6,100.00

Justification for equipment

The work we're proposing is to find data across distribution sources, combine and clean it, and integrate it into our existing open knowledge graph infrastructure. The only equipment we need to complete this work includes computers for the three new hires. The Metadata Curation Specialist and Lead, Funder Community do not require high performance computers and so we've budgeted $1,800 for each of them, but we have budgeted $2,500 for the purchase of a computer for the Senior Data Engineer who will need a machine with higher end storage and computing capacity. We use virtual servers for our work and have included server costs associated with this project in Other Costs.

Are you requesting a piece of equipment with a list price of £100,000 or more?

No

Are you requesting access charges?

No

Are you requesting overheads?

No

Are you requesting travel and subsistence?

Yes

Travel and subsistence costs

Type	Description	How much carbon will this offset (in tonnes)?	Cost requested from Wellcome
Conference attendance	Funding for each newly hired staff to present work at one conference	3.9	$15,000.00
Collaborative travel	Funds for project team to meet in person annually to review work and plan following year	22	$45,000.00
			Total: $60,000.00

Justification for travel and subsistence costs

We include two line items for travel cost: conferences for staff ($15k) and project team meetings ($45k). We budget $5,000 for each of the newly hired staff to travel to present their work at one conference each. We budgeted $15k each year for the 6 project staff (three new hires, and three existing staff) to gather in-person to review and plan collaborative work, troubleshoot challenges as they arise, and team-build. We will try to align these trips with open research information conferences to allow staff to participate in those in addition to their individual conference travel. In years 1 and 2, we will host collaborative workshops with funders that will include some travel costs for funders– those costs are included in the Other costs section.

Are you requesting overseas allowances?

No

Are you requesting fieldwork expenses?

No

Are you requesting clinical research?

No

Are you requesting public engagement and patient involvement?

No

Are you requesting contract research organisations?

No

Are you requesting other costs?

Yes

Other costs

Type	Description	Cost requested from Wellcome
Conference/seminar hosting	Workshops with funder partners (one in Year 1 and Year 2)	$430,000.00
Computing	Server costs to ingest, process, integrate, and distribute funding metadata	$1,150,000.00
Consultancy fees	Funding for researcher evaluations of project outcomes	$180,000.00
		Total: $1,760,000.00

Justification for other costs

We budget $1.76M in Other costs split across: server fees ($1.15M), funder collaborative workshops ($430k), and funding for researchers to evaluate the grants database ($180k). Servers and staff are the major expenditures for this project and we have budgeted $550k, $450k, $150k in Years 1-3 for server costs to ingest, process, clean, and integrate the new grant and report metadata sources into OpenAlex. We plan for higher costs in the first two years when the highest volume of new sources will be added and lower costs in the third year when these processes are optimized and the largest sources have been completed. Server costs include text data mining and parsing of tens of millions of PDFs, classifying and disambiguating text for millions of grants, and server access costs for users around the world to access the processed data freely. We budgeted $215k in each of Year 1 and 2 for collaborative workshops with funders to include venue and travel costs for delegates of our funding partners. We'll host virtual workshops annually reporting on progress, but in-person workshops are especially critical in the first two years as we learn from diverse funders (the state of their data, on-going work, aspirations, and unique challenges), agree on project specifications, and collaboratively set direction for the work to ensure it will be the solution needed by funders. Funding for travel is essential for some funders with our goal of rectifying coverage bias in existing systems. We will put out calls to the research community to evaluate coverage and metadata quality of our grant metadata throughout the project with $30k, $60k, $90k in Years 1-3 respectively. Many funders require a third-party independent assessment of OpenAlex before adoption and so these funds are needed for the research community to be able to undergo this important work.

Summary of costs requested

Total funding requested in this application

$3,582,123.00

Full economic costing

Is your organisation based in the UK?

No