



CONNECTING RESEARCH,
ADVANCING KNOWLEDGE

The DataCite Journey

Reflecting on Progress and Envisioning the Future

Matt Buys

October 12, 2023

DataCite Annual Community Meeting



[@datacite](https://twitter.com/datacite)



[@datacite@openbiblio.social](https://openbiblio.social/@datacite)



What if...

...we could transform the landscape of research and knowledge sharing by seamlessly connecting every research output and resource, making them effortlessly discoverable, citable, and reusable on a global scale? Imagine the impact on academia, innovation, and the progress of human knowledge.



Photo by Thomas Baker from Noun Project

Our Foundation

Building on our Shared Strategy

Community at the Core



We are a global community that shares a common interest: to ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future.

As a community, we make research more effective with metadata that connects research outputs and resources—**from samples and images to data and preprints.** We enable the creation and management of persistent identifiers (PIDs), integrate services to improve research workflows, and facilitate the discovery and reuse of research outputs and resources.



2900+

Repositories



52+

Countries



55m+

DOIs

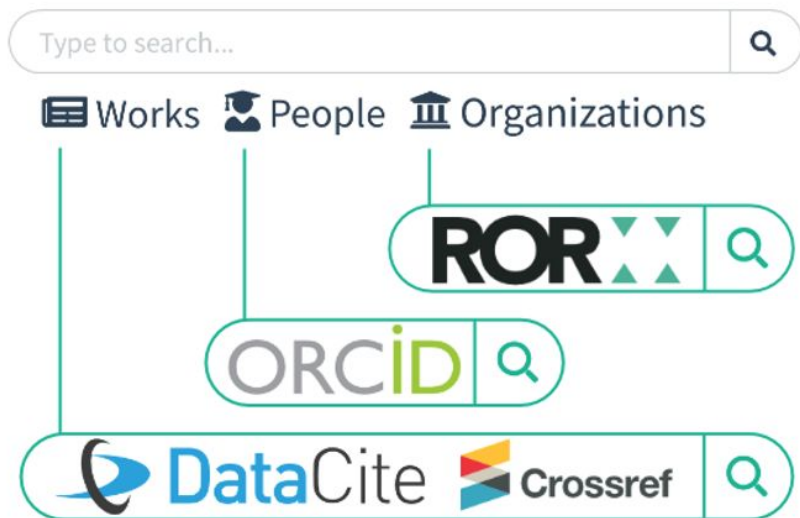


1350+

Organizations

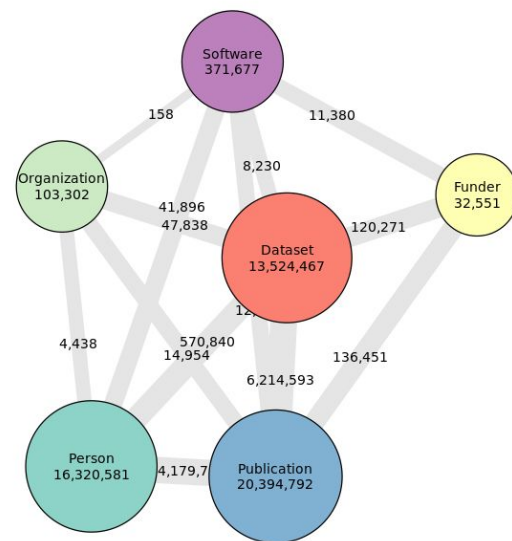
Find and connect research

Find Research with DataCite Commons

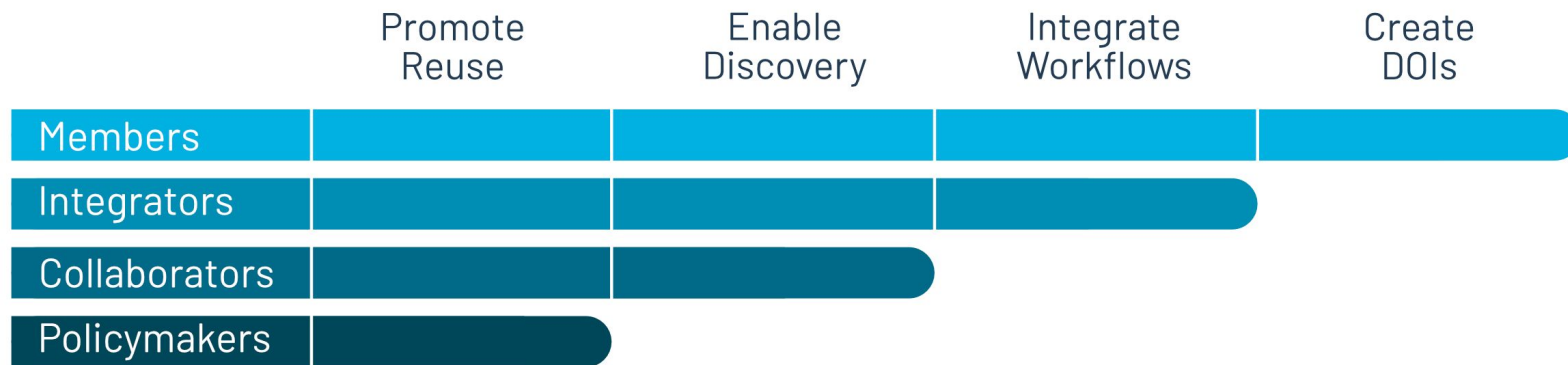


GraphQL API

Number of nodes and connections (March 2023)



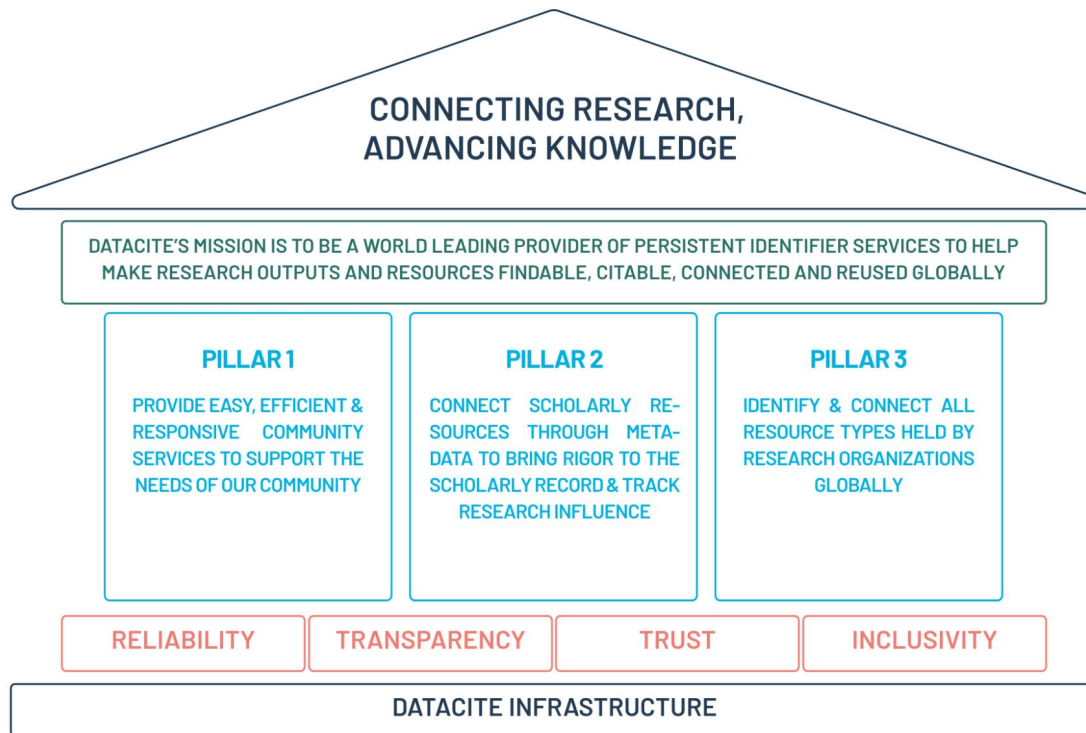
Community Coordination



Shared Infrastructure Services

	Create DOIs	Integrate Workflows	Enable Discovery	Promote Reuse
Why?	Make research outputs and resources discoverable and citable for the long term	Enhance research workflows and services through integration	Enable discovery of research outputs and resources and reporting with advanced tools and analytics	Promote reuse with flexible, state-of-the art tools and technology
What?	<ul style="list-style-type: none">• DOI & metadata management services• Link checker	<ul style="list-style-type: none">• APIs to automate system integration• Usage tracker	<ul style="list-style-type: none">• Metrics & relational metadata files• Search & analytics dashboards• APIs and UIs	<ul style="list-style-type: none">• Content negotiation & citation formatter• ORCID auto updates & claiming
How?	Register DOIs and provide rich metadata in standardized formats	Automate DOI registration and/or retrieve relevant metadata through system integration	Make outputs resolvable and discoverable through metadata including connections between outputs/resources and other entities	Enable reuse through good citation practices and embedding DOIs in scholarly infrastructure workflows

Vision to Reality



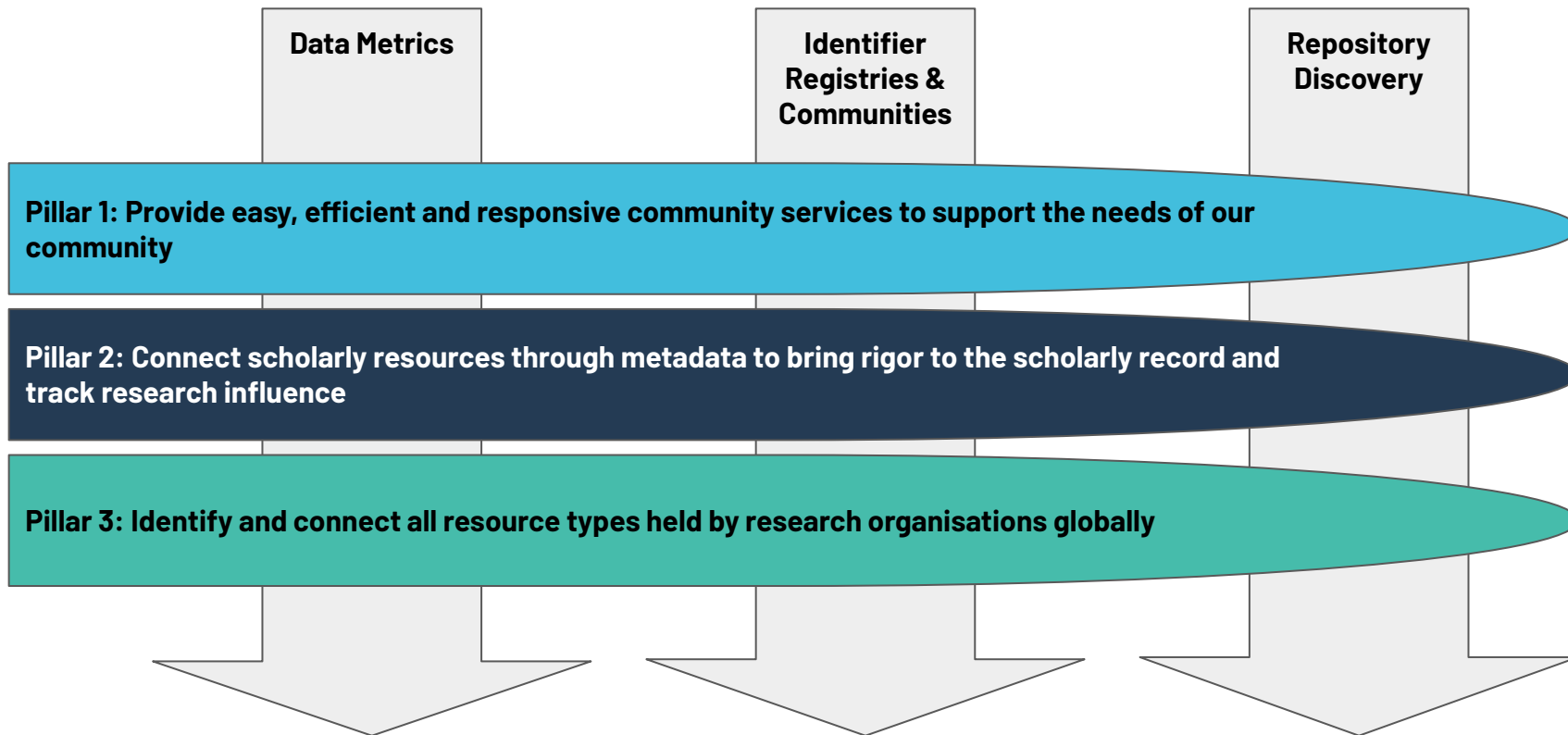
Strategic initiatives

In line with our mission and vision, DataCite actively participates and leads various initiatives through collaboration with stakeholders in the community to make open science a reality.

- **Data Metrics** – We help further the adoption and implementation of responsible data metrics with, for example, the Make Data Count initiative.
- **Identifier Registries & Communities** – We support community-led initiatives such as the Research Organization Registry (ROR), IGSN and the Global Access Program.
- **Repository Discovery** – We contribute to the development of repository discovery initiatives such as re3data.

Multi-Faceted Approach

Linking Initiatives & Strategy



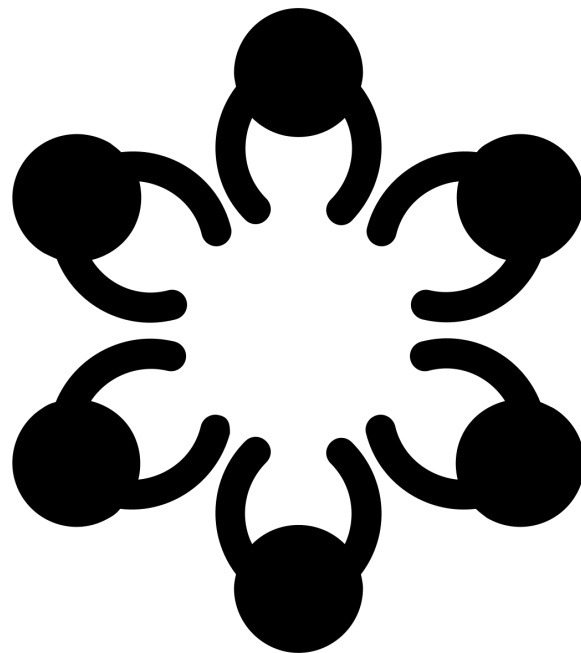
Our Progress

Highlights from 2023

Sharing A Few Highlights from 2023

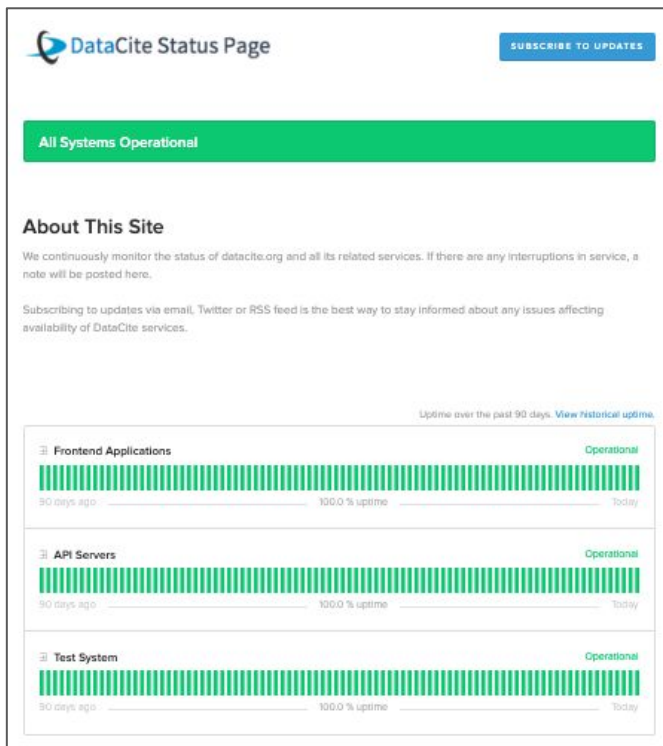
Increased Adoption & Usage

- We are proud that over the past year, DataCite has witnessed a substantial and noteworthy surge in the utilization and resolution of DOI metadata, signifying a growing recognition and reliance on our persistent identifier services within the scholarly community and beyond.
- 1,620 resolutions of DataCite DOIs every minute.
- 700,000 API requests every day.



Sharing A Few Highlights from 2023

Improved Services & Functionality

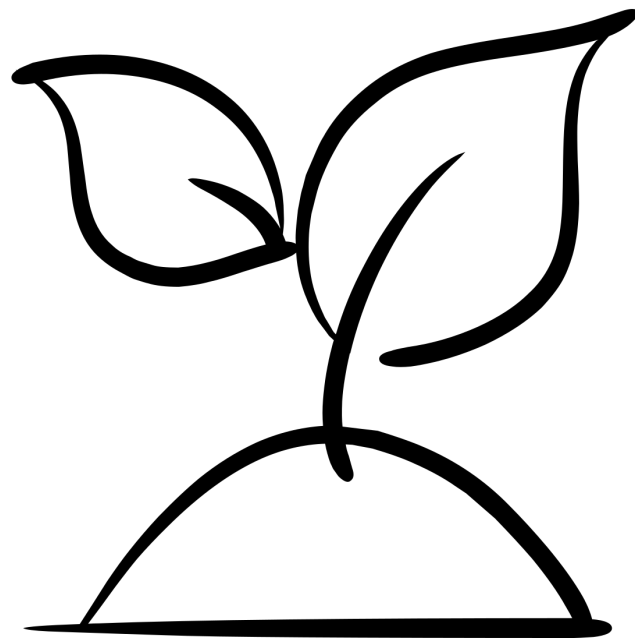


- Our dedication to the enhancement and refinement of our services has resulted in a continuous stream of improvements and added functionality. We prioritize the evolving needs of our community, ensuring that our offerings remain at the forefront of research data management. Including community driven functionality releases, improve stability in services, reduced bugs and metadata enhancements.
- This has been a collective effort across staff and governance groups.

Sharing A Few Highlights from 2023

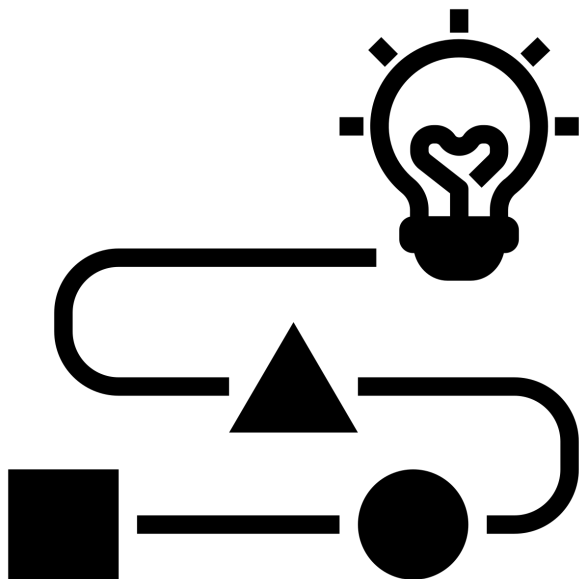
Remarkable Community Growth

- DataCite's community has experienced remarkable growth, both in terms of the number of members and the diversity of stakeholders engaged in our mission. This expansion underscores the resonance of our vision and the trust that the global research community places in our persistent identifier services.
- Over 300 repositories have joined DataCite during 2023 so far! This is more than one repository each day joining our community.
- 25% of our membership is represented with the GAP regions and this continues to grow.



Sharing A Few Highlights from 2023

Launching Strategic Initiatives

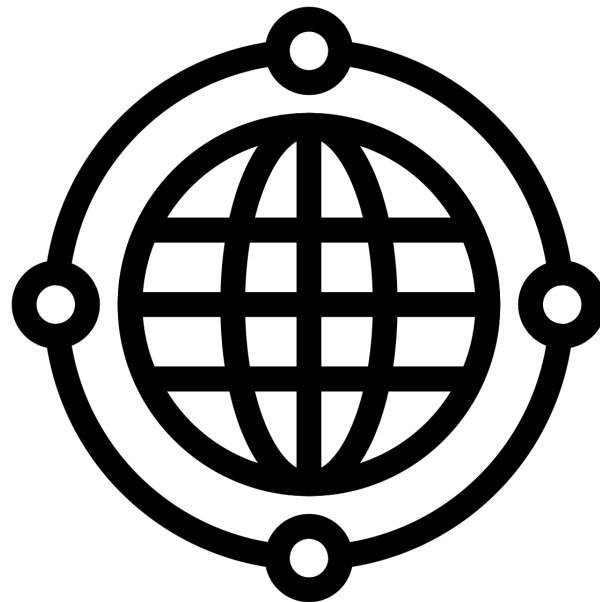


- In the past year, we have successfully launched a series of strategic initiatives that have not only met but exceeded our expectations. These initiatives, meticulously designed to align with our vision and mission, have already started making a substantial impact on the research landscape, fostering greater accessibility and connectivity for research outputs worldwide.

Sharing A Few Highlights from 2023

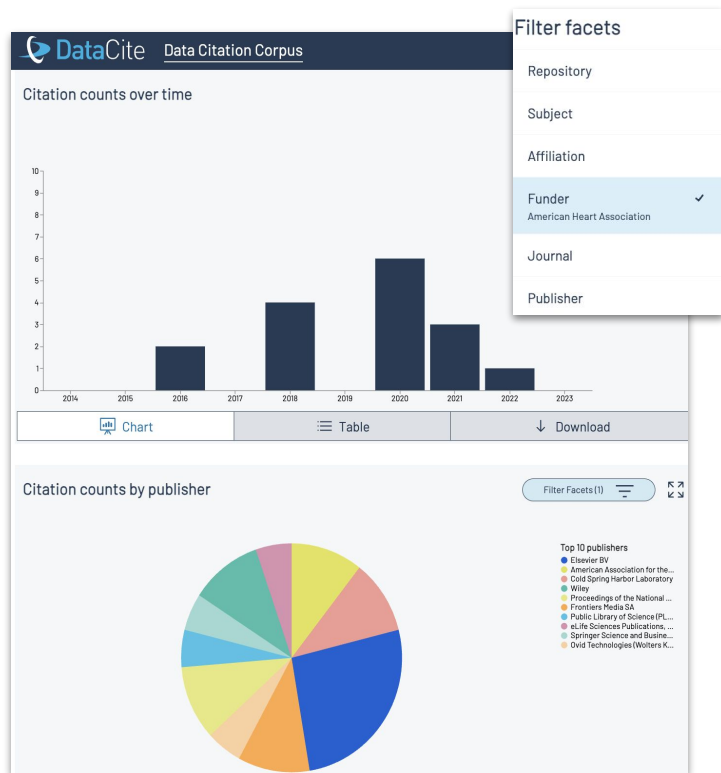
Global Access Program

- In 2023, DataCite achieved a significant milestone with the successful implementation of the Global Access Program (GAP). GAP is a transformative initiative aimed at enhancing access to our open infrastructure services, particularly in regions that have been traditionally underrepresented in research and knowledge sharing. This groundbreaking program was made possible, in part, by the generous grant 2022-316573 from the Chan Zuckerberg Initiative.



Sharing A Few Highlights from 2023

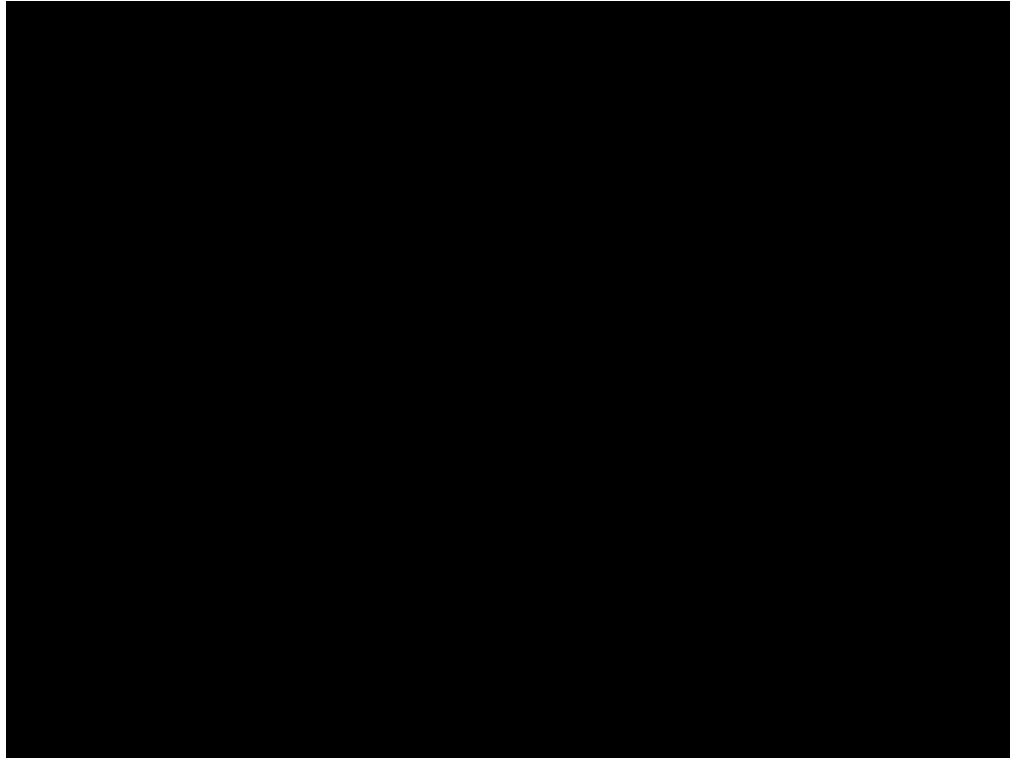
Global Data Citation Corpus



- In 2023, The Wellcome Trust funded the Global Data Citation Corpus, an initiative to enhance data reuse understanding in research. Led by DataCite, it compiles a trusted, central repository of research data citations spanning articles, preprints, government documents, and various research outputs, partnering with organizations like the Chan Zuckerberg Initiative and EMBL-EBI, aiming to revolutionize data citation practices. This central repository will promote transparency and collaboration among researchers worldwide.

Bringing It Together

Leveraging Technology



Envisioning the Future

Collection Efforts

Building Shared Infrastructure Trusted Partner

One of our primary strategic objectives in the coming year is to solidify and expand our position as the **trusted partner for persistent identifier (PID) services** within the global research community. This entails establishing deeper collaborations, delivering even higher standards of reliability and transparency, and **forging stronger bonds of trust with our members and community stakeholders worldwide.**

- Building on the Grant ID services pilot and scaling across the community.
- Supporting the needs of our community across resource and output types.
- Partnering with key communities, similar to our efforts with IGSN

Building Shared Infrastructure Metadata Completeness

To enhance the robustness and comprehensiveness of our metadata, we are embarking on an ambitious journey to **accelerate metadata completeness**. This initiative will involve the introduction of enhanced member services and data insights, **empowering our community members with tools and resources that facilitate the creation and enrichment of comprehensive metadata records.**

- Metadata insights
- Ingesting relational metadata from third parties
- Exploring enhanced services to support metadata completion e.g. subject metadata.
- Enhanced metadata schema evolution.

Building Shared Infrastructure Tracking Research Outputs

In line with our commitment to advancing research knowledge, we are focussing on developing and launching **improved visualizations and analytical services** tailored to the needs of our community. These services will provide valuable insights and tools for tracking research outputs, enabling the research community **to better understand the impact and influence of their scholarly work throughout the entire research lifecycle.**

- Enhanced visualizations and dashboards.
- Coordinated community efforts in regard to application and use of metrics.
- Partnership with policy stakeholders.

Building Shared Infrastructure Discovery

Our strategic focus for the upcoming year also encompasses the enhancement of **discoverability and reuse of our members' metadata**. We aim to achieve this by implementing enhanced harvesting services that will make it easier for researchers, institutions, and the broader community to **locate, access, and effectively utilize the wealth of research resources and outputs available** within our datafile.

- Launch an enhanced harvesting service.
- Annual data dumps.
- Ingesting metadata into the DataCite Commons.
- Aggregating metrics and addressing downstream discovery use cases.

Building Shared Infrastructure Global Access Program

As we look ahead to 2024, DataCite remains committed to the overarching goal of **empowering global communities through equitable access to research infrastructure.**

Building on the success of the Global Access Program (GAP), our primary focus for the coming year is to further **expand and deepen our support in lesser-represented regions**, ensuring that researchers and organizations from all corners of the world have the tools and resources they need to thrive.

- Continuation of the Global Access Fund
- Dedicated community engagement across regions.
- Outreach and partnership activities.

Building Shared Infrastructure Data Citation Corpus

In 2024, the Global Data Citation Corpus, supported by The Wellcome Trust, is set to prioritize key objectives. **It aims to integrate seamlessly into core DataCite technology, expand its coverage in terms of both depth and breadth, and address various downstream use cases.** These initiatives will further **solidify its role as a pivotal resource in the research community, enhancing data citation practices and supporting valuable research insights.**

- Scaling beyond the prototype developed.
- Integration into DataCite discovery services.
- Extend breadth and depth of coverage.



CONNECTING RESEARCH,
ADVANCING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
blog.datacite.org



support.datacite.org
support@datacite.org



[@datacite](https://twitter.com/datacite)



[DataCite](https://www.youtube.com/DataCite)



[@datacite](https://www.linkedin.com/company/datacite)