

DataCite Annual Review: 2017



WHO WE ARE	1
MESSAGE FROM THE EXECUTIVE DIRECTOR, PATRICIA CRUSE	2
MESSAGE FROM THE EXECUTIVE BOARD PRESIDENT, JOHN CHODACKI	4
MESSAGE FROM THE EXECUTIVE BOARD TREASURER, MARCO MARSELLA	5
MEMBERSHIP OVERVIEW – GROWTH AND CHANGE	6
NEW SERVICES – MEETING THE NEEDS OF OUR MEMBERS	9
NEW PROJECTS AND FUNDED INITIATIVES – EXTENDING OUR SERVICES	13
COMMUNICATION, ENGAGEMENT, AND OUTREACH – EXTENDING OUR REACH	17
FINANCIAL OVERVIEW	20
EXECUTIVE BOARD ROSTER	21
STAFF ROSTER	22

Who We Are

DataCite is a leading global non-profit organization that provides persistent identifiers (DOIs) that enable researchers to identify, locate, access, connect, and cite research datasets with confidence.

OUR VISION: Connecting research, identifying knowledge.

OUR MISSION: DataCite's mission is to be the world's leading provider of persistent identifiers for research. Through our portfolio of services, we provide the means to create, find, cite, connect, and use research. We seek to create value and develop community-driven, innovative, open, integrated, useable, and sustainable services for research.

OUR GUIDING VALUES

- DataCite is trustworthy – we're fully dedicated to open research and all its related content.
- We're community owned and driven – founded by the research community, we're a democratic organization that engages with and listens to the broad research community to meet the needs of researchers.
- We provide timely support to our members – we respond quickly to the changing needs of our community, adding new services as needed.
- We actively involve our membership in the building of new services – we're open to building services for all types of technologies.
- We're a global leader – we have an international reputation for providing DOIs for traditional and non-traditional research outputs.
- We're transparent – we have a structure that allows us to be proactive, agile, and responsive to the needs of the research community.
- We advocate for the role of all research content in the research landscape – we engage in outreach that reflect the interests of our diverse community.

Message from the Executive Director, Patricia Cruse



This is DataCite's ninth year in operation and the third year since I joined as Executive Director. On behalf of our Executive Board and staff, I'd like to express our sincere gratitude for the continued support of our members during our first nine years, as well as your ongoing support as we look ahead to our tenth anniversary and beyond. DataCite has experienced significant growth over the years because of our members, particularly in our capacity to serve more organizations, which thereby furthers our data-sharing mission.

In 2017, DataCite welcomed 25 new members. And we are very pleased that so many new types of organizations (for-profits, research institutions, government agencies, and others) have joined us, which has helped to extend our reach and impact new communities and disciplines. Moreover, with our expanded membership and the recent change in our pricing structure, DataCite is building a roadmap for financial and organizational stability that will allow our organization to prosper for many years to come.

As we continue to mature as an organization, member satisfaction and reliable, effective services remain the hallmarks by which we measure our performance. During the past year, we made a positive difference with many of our services, including Fabrica – our new DOI service. Our goal is to make data sharing and citation easier by providing easy, understandable, predictable, and robust services for creating, managing, finding, using, and tracking DataCite DOIs. It's important to note that our activities have not only been limited to technical development but have also included support services that foster the successful use of all our services. We also continue to be involved in several grant-funded initiatives, including FREYA (a

Horizon 2020 project funded by the European Commission) and the Make Data Count project, both of which will continue to help build services that are important to our users.

One of our many important activities in 2017 was the development of a new vision statement (Connecting Research, Identifying Knowledge), which challenges each of us to think about new ways to connect with the research community. As we move forward together, I'd like to thank all our DataCite members for remaining connected and engaged in our activities and share my excitement for what's to come as we continue to grow our organization, services, and staff to support the research community.

Patricia Cruse

Message from the Executive Board President, John Chodacki



I want to thank our staff and all our members for your willingness to dedicate yourselves and also for your willingness to adapt to changing landscape in order for DataCite to flourish. We have used our collective creativity and insights to build new opportunities and strategies. We have a lot to be proud of.

I also want to specifically thank Adam Farquhar and Wilma van Wezenbeek for their service and leadership to the DataCite mission. I have big shoes to fill and look forward to the challenge.

I know in the coming years, we will see new challenges and find new ways of tackling them together. I am looking forward to working with all of you.

John Chodacki

Message from the Executive Board Treasurer, Marco Marsella



As DataCite's new treasurer I am pleased to report that DataCite's finances are in good health. The new fee structure, introduced in 2017, is beginning to bear fruit and is promoting a steady increase of DataCite's membership. In 2018 we are adding new members at a rate of more than one per week allowing us to predict that income from membership fees will grow 78% to 644,000€ -- the new fee structure also produced significant reductions in fee for most members.

At the end of 2017, we experienced a shortfall due to the delay in payment for the FREYA project. Despite the delay, grant funding is expected to provide in 2018 over 590,000€ or about 48% of total income.

2017 ended with a balanced budget despite the slow start in transitioning EZID/Purdue clients to DataCite membership and the aforementioned delayed project funding. In 2018 we predict a positive balance of 192,000€ after realizing the 50,000€ carryover, showing that DataCite is in good health and heading in the right direction.

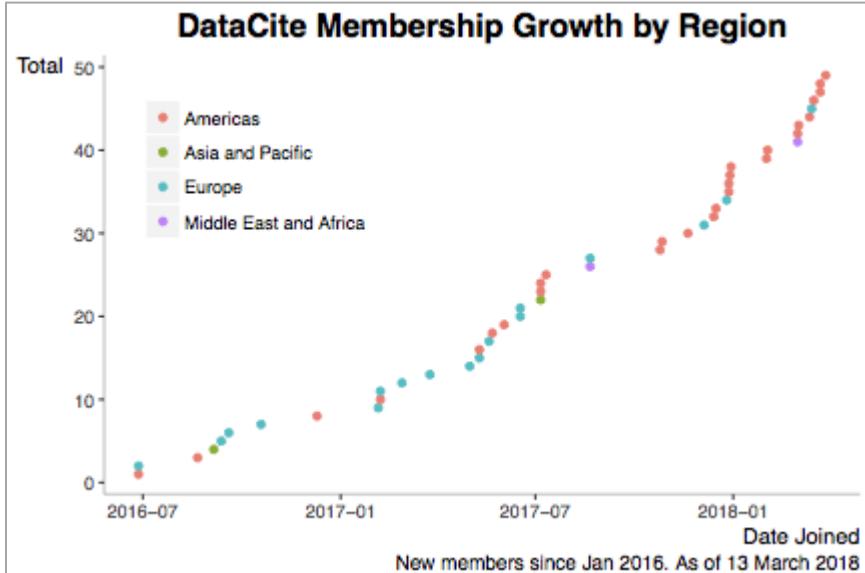
Our income still significantly depends on project funding; however, the new fee structure is expected to promote membership growth and potentially reduce this dependency thereby providing DataCite with a more predictable income stream.

Marco Marsella

Membership Overview – Growth and Change

At DataCite's General Assembly meeting in 2017, members took the bold step of changing our membership and DOI service fee structure. One of the primary goals of the new fee structure was to lower our fees, thereby opening DataCite up to a greater diversity of organizations – large and small. This move contributed significantly to the acquisition of 25 new members between May and December 2017, nearly doubling our total membership to 68 organizations.

DataCite membership also grew due to the California Digital Library (CDL) and Purdue University's new strategic direction for their EZID DOI services. The new direction was based on the desire to support DataCite's long-term sustainability. Over the course of the next two years, EZID DOI services will be phased out for clients outside of the University of California. DataCite, CDL, and



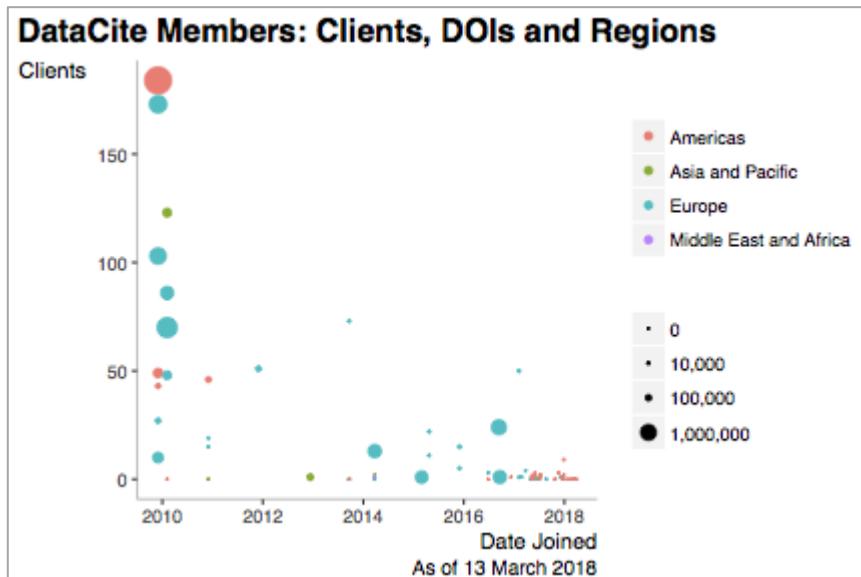
Date Joined	Americas	Asia and Pacific	Europe	Middle East and Africa	Total
2016-07	2	0	1	0	3
2016-10	3	2	0	0	5
2017-01	8	4	0	0	12
2017-04	10	6	2	0	18
2017-07	25	23	10	0	58
2017-10	28	26	15	0	69
2018-01	38	35	35	0	108
2018-03	48	42	45	0	135

Purdue are working together to support EZID users in joining DataCite and migrating their DOI service needs to our organization. If we look at membership growth by region, the outcome is more members from the U.S. (the Americas) joining DataCite. DataCite appreciates CDL and Purdue's efforts to support our membership and benefit the research community. Specifically, Joan Starr and Michael Witt are to be commended for their tireless work, expertise, and support of this initiative.

Welcome New 2017 DataCite Members!

<ol style="list-style-type: none"> 1. Brain Life 2. Center of Computational Science Research 3. Citrine 4. Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria (Madroño) 5. Data Archiving and Networked Services (DANS) 6. DataFirst, University of Cape Town 7. The Digital Collections of Colorado 8. Dryad 9. Eawag (Swiss Federal Institute of Aquatic Science and Technology) 10. Elsevier 11. The Environmental Data Initiative 12. Fondren Library, Rice University 	<ol style="list-style-type: none"> 13. Euro-Mediterranean Center on Climate Change 14. GeneLab (NASA Ames Research) 15. James Joyce Library, University College Dublin 16. National University Singapore 17. Orbis Cascade Alliance 18. Poznan Supercomputing and Networking Center 19. Publications Office of the European Union 20. ResearchWorkspace 21. Riga Technical University 22. The Smithsonian Institution 23. University of Southern California, Information Sciences Institute 24. Virtual Library of Virginia (VIVA) 25. University of Iowa Libraries
---	--

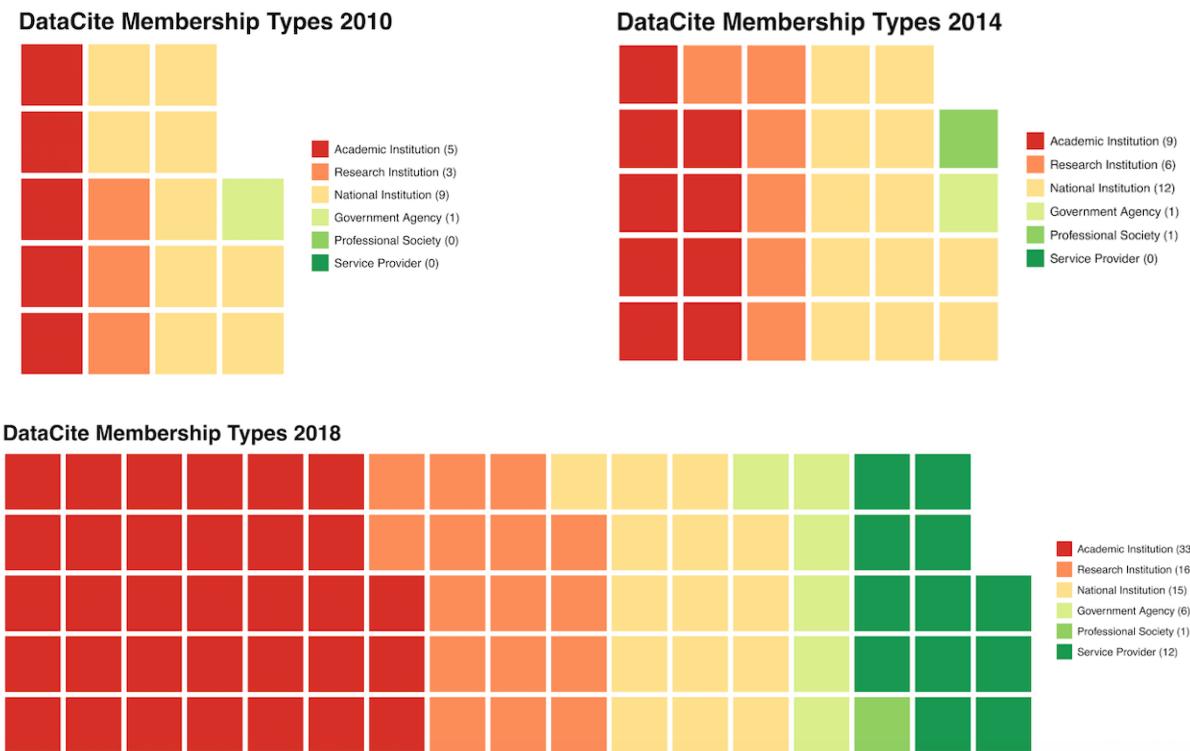
If we look at growth areas by region in relation to the creation of DOIs we see that members



providing DOI services that joined in 2009/2010 and have many clients are, as one would expect, very active in DOI generation. However, that is beginning to change as we see four

very active DOI service providers that joined after 2014 and two DOI service providers each with a single client.

Another goal of changing our fee structure was to diversify the types of organizations that are able to join DataCite. If we compare the acquisition of member by type in 2010, 2014 and 2018.through March 2018 we see a change over time. DataCite began with a strong emphasis on national institutions. Over the years we have seen strong growth from academic institutions and recently growth of research institutions. With the inclusion of for-profit organizations in 2016 we see growth in the service provider category



New Services – Meeting the Needs of Our Members

DOI Fabrica

In October 2017, we launched a partial release of DataCite's newest service, DOI Fabrica, that allows members and their clients to register and manage DOIs, prefixes, and accounts. We are currently in



the final stages of development and testing for the full 1.0 release in April 2018 that will feature all the functionality of the Metadata Store (MDS) user interface, including DOI and metadata registration. Designed with modularity, extendibility, and legacy service integration in mind, DOI Fabrica is our core platform for future member and client services.

In addition to focusing our development efforts on providing an excellent user interface for DOI Fabrica, we ensured that the new service is integrated with other aspects of DataCite infrastructure. For instance, we developed a new API, coming with the April release, that provides all the functionality of provider, client, prefix, and DOI management. We also worked on the backend database service that is shared with the MDS, performed significant work on authentication, and launched a new API compatible with the EZID service. Lastly, we developed robust support and training materials to help users get up and running with this important new service.

Highlights of DOI Fabrica's current functionality include:

- Web-based frontend with extensive input validations, documentation, and support for needed workflows.
- Client account management (add, delete, edit clients, transfer DOIs) via the web interface.
- Simplified prefix management so providers can self-assign new prefixes when needed.
- Inclusion of all URLs registered in the handle system in the DataCite search index.
- Initial integration of re3data into DataCite services.
- Registration of draft (reserved) DOIs that can be made available publicly or deleted.
- Registration of metadata in many formats (DataCite XML, Crossref Unixref, and schema.org).
- Basic statistics and visualizations illustrating the number of clients and DOIs.
- Standalone DOI Fabrica application (using the Electron platform).

We are very excited about the full release in April. We are also thrilled that our interaction with Fabrica users during development provided valuable insight that we will continue to use to expand and enhance DOI Fabrica after the 1.0 release.

Other Services

In addition to our development of DOI Fabrica, we introduced and improved other services in 2017. Please review our [Public Development Roadmap](#) (launched in June 2017) for a deeper look at our work going forward as well as a summary of our work completed in 2017, including the following projects.

Metadata

We rewrote DataCite's Content Negotiation service to include more metadata formats (such as schema.org) and provide additional functionality. In September 2017, we released DataCite Metadata Schema 4.1, which focuses on metadata for software.

Search

We migrated the Solr-based DataCite Search to a completely new Docker-based infrastructure. We launched a new search index, currently for providers and clients, based on Elasticsearch – and we see this as the future platform for DataCite Search. We also added new attributes to the DataCite Search index, including funding and geolocation information.

Data Citation

Together with Crossref, we launched a common platform (Event Data) to store links to DOIs, including cross-links between Crossref and DataCite. We also worked with the Research Data Alliance (RDA) community on the Scholix standard for exchanging literature-data links, and we facilitated integrations between data centers using DataCite DOIs, such as Cambridge Crystallographic Data Centre (CCDC) and Pangaea, publishers such as Elsevier, and integrators such as OpenAIRE. In addition, our ongoing work in the Make Data Count (MDC) project will bring all our data citation work together and make this information available to our members and users later in 2018.

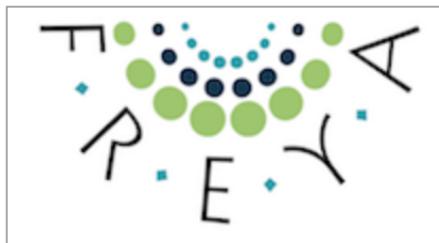
Infrastructure

Our new and existing services depend on a solid infrastructure. In 2017, we continued to make significant progress in providing reliable and scalable infrastructure across our service portfolio, as demonstrated by the following highlights.

- We continued to upgrade and replace existing infrastructure. In addition to the previously mentioned work done on the Solr Search and Content Negotiation services, we relaunched the OAI-PMH service.
- We deployed all our static websites (including homepage, blog, and schema documentation) in a content delivery network (CDN), which led to significantly faster access times to these services from users outside of Europe.
- We deployed most of our servers as auto-scaling services, allowing automatic adaption to respond to the increased demand and deployment of new releases.
- We improved the monitoring of our infrastructure and increased the visibility of this monitoring to our members and users. Our [status page](#) now includes automatic information updates in cases of service downtime. We also now display historical uptime information on this page.

New Projects and Funded Initiatives – Extending our Services

In 2017, DataCite engaged in a variety of collaborations and projects, but they all share a common theme of building services in support of our members and working with new communities. DataCite continues to be involved in many collaborative and funded initiatives – all with the goal of improving our services, changing communities of practices, and advocating for data sharing.



Project Title: [FREYA](#)

Funder: European Commission, Horizon2020 program

DataCite Funds: 877,500 €

Start Date and Duration: Dec. 1, 2017, 36 mo.

Partners: Science and Technology Facilities Council (STFC) (lead), European Organization for Nuclear Research (CERN), European Bioinformatics Institute (EMBL-EBI), The British Library, PANGAEA, Data Archiving and Networked Services (DANS), Australian National Data Service (ANDS), ORCID, Crossref, Hindawi, Public Library of Science (PLOS).

FREYA, the successor to THOR, is a 36-month project that started in December 2017. This is a significant project for DataCite and after a competitive process the team's top-notch proposal was awarded funding. The mission of FREYA is to foster a robust environment for a range of persistent identifiers as an essential component of the European Open Science Cloud (EOSC). FREYA partners are providing the essential building blocks for supporting changes in the way researchers work and the tools they use. This is all coming together in a new vision for how research is conducted, exploiting the full potential of Open Science and is a core to DataCite's strategic mission. DataCite is excited to work with new and our former THOR partners.



Project Title: [Make Data Count](#)
Funder: The Sloan Foundation
DataCite Funds: 171,500 €
Start Date & Duration: 05/01/17; 24 mos.
Partners: California Digital Library (lead), DataONE

Understanding the reach and use of research data is a top priority for DataCite. Currently, there are a lack of services and practices for measuring activities related to data sharing, use, reuse, and citation. With the Make Data Count project, DataCite, working with the California Digital Library (CDL) and the DataONE initiative, are developing and deploying the social and technical infrastructure that will elevate data to the level of first-class research content (alongside more traditional outputs such as publications). The Make Data Count project hit a major milestone in February 2018 with the first release of a code of practice for research data usage metrics.¹ This is an essential step towards realizing usage metrics as a critical component in our understanding of how publicly available research data are being reused.



Project Title: [THOR](#)
Funder: The Sloan Foundation
DataCite Funds: 490,109 €
Start Date & Duration: 06/01/2015; 30 mo.
Partners: The British Library (lead), European Organization for Nuclear Research (CERN), European Bioinformatics Institute (EMBL-EBI), PANGAEA, Australian National Data Service (ANDS), ORCID, Dryad, Public Library of Science (PLOS).

Project THOR came to a close in November 2017. From its inception, THOR has been a tremendous asset to DataCite, allowing us to hire key staff, build out our services, and reinvigorate our organization. The project had four concrete aims: (1) establish interoperability; (2) integrate services; (3) build capacity; and (4) achieve sustainability. At the end of the 30-

¹ Fenner M, Lowenberg D, Jones M, Needham P, Vieglais D, Abrams S, Cruse P, Chodacki J. (2018) Code of practice for research data usage metrics release 1. PeerJ Preprints 6:e26505v1 <https://doi.org/10.7287/peerj.preprints.26505v1>

month project, THOR has achieved its objectives. Research efforts have advanced the state of the art and are having an operational impact. Innovative service developments have been rapidly deployed, and partner services were integrated into essential production services in life sciences, high-energy physics, earth and environmental sciences, and social sciences and humanities. Outreach events have engaged thousands of people in Europe, North America, and Australasia, as well as touched many others across the globe.



Project Title: [Towards a FAIR Digital Ecosystem in the Cloud](#)

Funder: The National Institutes of Health (NIH)

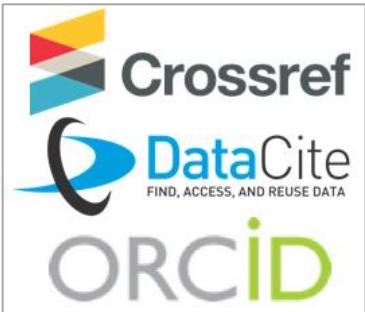
DataCite Funds: 34,300 €

Start Date & Duration: 12/01/17(?); 6 mo.

Partners: Harvard IQSS, Massachusetts General Hospital

NIH's Big Data to Knowledge (BD2K) Data Commons Pilot initiative will test the feasibility of (and develop best practices for) making NIH-funded datasets and computational tools available through communal, collaborative platforms on public clouds. The BD2K program is an NIH data science program that launched in 2013 to facilitate the broad use of biomedical big data.

DataCite is participating in the second phase of the project, started in December 2017, which involves working with partners to develop Global Unique Identifier (GUID) capability to provide a persistent, machine resolvable identifier platform for all FAIR objects in the NIH Data Commons that is fully aligned with community practices, recommendations, and metadata models. This project is a quick spring to produce an MVP in 6-mos time. After 6-mos, NIH will make additional awards based on the success of the MVP.



Organization identifiers are a key component of the research information infrastructure. In early 2017, DataCite, Crossref, and ORCID announced the formation of an [Organization Identifier Working Group](#) to refine the structure, principles, and technology specifications for an open, independent, non-profit organization

identifier registry to facilitate the disambiguation of researcher affiliations. After a stakeholder meeting in January 2018, the Executive Directors from DataCite, Crossref, and ORCID were charged as the interim executive committee to draft a formal proposal to launch the initiative.

Communication, Engagement, and Outreach – Extending Our Reach

As a member-led organization, DataCite continually seeks new ways to engage with our member community – our success depends on an informed and active membership. We engage with our members through in a variety of ways, including DataCite sponsored events, webinars, our blog, helpdesk, support site, and social media, all of which we use to share information with our members (while also listening to our members' needs). While communication with our members is a priority, several factors impacted our communication efforts in 2017, including the departure of our Communications Director, our focus on new member acquisition and onboarding, and Fabrica service development. Going forward, we will hire a new Communications and Member Relations Director and re-emphasize the importance of communication with our members and the broader DataCite community.

Listening to the community and sharing our development roadmap



As a geographically distributed organization, one of our biggest challenges is to understand our members' needs, collect their ideas, and share our progress. In 2017, we implemented [DataCite User Stories](#) and the [DataCite Development Roadmap](#). Our User Stories

collect descriptions of new features and requirements, while the Roadmap takes those User Stories and organizes them into projects and prioritizes our development. These tools allow the DataCite team to clearly understand what is important to our users.

Communication tools and Fabrica services

We use DataCite's blog, webinars, and support site to share information with the broader community and to gather community input. Blog topics include information about metadata schemas to support software citation, content negotiation tips, updates on community initiatives, such as the Organization Identifier (ORG ID) initiative, and, most importantly, information on our new Fabrica services. As we roll out new services such as Fabrica, webinars play an important role by providing a forum for user training and for gathering a better understanding of how we need to modify the service to meet our members' needs. [DataCite Support](#) provides another key avenue for sharing information on Fabrica, with the goal of helping Fabrica users be as successful as possible. Lastly, our dedicated [help desk](#) provides ongoing support for detailed questions from the DataCite community.

Conferences – sponsorship and participation



Once again, DataCite, together with the California Digital Library, Crossref, and ORCID, organized [PIDapalooza](#), a two-day celebration of persistent identifiers and networked research. On January 23-24, a sold-out crowd of 150+ PID people from 23 countries across six continents gathered in Girona, Spain for the second annual PIDapalooza. This year's festival featured updates and discussions, demos and use cases, brainstorming, and networking. In addition to the 40+ interactive half-hour parallel sessions, PIDapalooza featured plenaries from leading PID enthusiasts, including Geoffrey Bilder (Crossref), Jo McEntyre (EMBL/Europe PubMed Central), Melissa Haendel (OHSU Library), and Carly Strasser (Coko Foundation). Festival goers could access all the presentations on [figshare](#) – thank you figshare! We look forward to seeing even more members of the DataCite community at next year's festival.



DataCite's own Martin Fenner was the lead organizer behind Force2017, held in Berlin in October 2017. Force2017 brought together scholars, researchers, librarians, data managers, grant administrators, funders, publishers, editors, societies, and anyone else interested in scholarly communications. Year after year, Force proves to be an ideal conference for the DataCite community to share and learn about DOIs for the scholarly community.

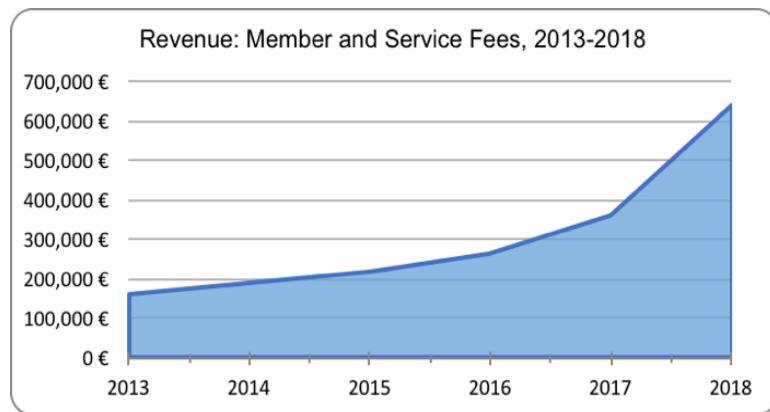


Research Data Alliance (RDA) plenaries are held every six months in different locations around the world. In 2017, we gathered in Barcelona (April) and Montreal (May). RDA has become the “go-to” meeting for DataCite staff to engage in numerous working groups and initiatives. In addition to being productive events that bring together a unique community of data scientists, librarians, computer scientists, and domain scientists, RDA is a great place to meet face to face with many of our members. These plenary meetings are special working events that help DataCite learn from the community, improve our services, and, ultimately, improve data sharing across disciplines, technologies, and countries.

Financial Overview

DataCite is a not-for-profit organization sustained by fees from our member organizations. Our primary objective is to achieve financial sustainability by controlling our expenses and increasing our membership. In line with this objective, DataCite continuously invests in a robust, scalable, and trusted infrastructure and services.

Solid, continuous, and predictable funding – via member and service fees – is essential for sustaining DataCite's core DOI services. In 2017, we saw a sharp increase in revenue from membership fees, moving us closer to a significant milestone of funding all our core activities from membership revenue. Our income from membership fees increased by nearly 40% in 2017, from 264,312€ in 2016 to 360,832€.



Grant funding is another important component of DataCite's financial structure. Grants allow DataCite to build and test new services, engage with new communities, and adapt emerging services. Moreover, the outcomes from our grant initiatives allow us to bring innovation directly to our members, which is a huge win for the DataCite community. It's important to note that the revenue from our member and service fees provide our organization with a sound infrastructure that serves as the foundation for our grant activities without disrupting our core DOI services.

Executive Board Roster

	<p>John Chodacki (President) is responsible for overseeing the strategic planning, development, and management of the California Digital Library's digital curation group, University of California Curation Center.</p>	 <p>Adam Farquhar (Deputy President) is Head of Digital Scholarship at the British Library, where he focuses on establishing services for scholars and researchers.</p>
	<p>Mike Frame is Chief Scientific Data Integration and Visualization at USGS. He participates in a number of interagency cyberinfrastructure, data integration, and scientific activities.</p>	 <p>Mark Hahnel is the founder of figshare, a repository where users can make all of their research outputs citable, shareable and discoverable. He has a PhD in stem cell biology at Imperial College London.</p>
	<p>Marco Marsella (Treasurer) is Senior Advisor Global Information System at FAO, where he contributes to the design of the system and the promotion of DOIs in the Plant Genetic Resources community.</p>	 <p>Salvatore Mele is head of Open Access at CERN, where he co-architected the SCOAP3 initiative. His team develops solutions for Open Data and runs INSPIRE, a global Open Access digital library for High-Energy Physics.</p>
	<p>Karen Morgenroth is Manager of Information Management Operations at the National Research Council Canada, where her teams focus on digital information management, repositories and open government initiatives.</p>	 <p>Irina Sens (ex-officio) is the Deputy Director at the German National Library of Science and Technology (TIB). She led the German project "Registration of Scientific primary data" and was responsible for the introduction of DOIs for data.</p>

Staff Roster

	
<p>Trisha Cruse, Executive Director, leads DataCite with a strong commitment and passion for data sharing. As Executive Director, her role is to advance DataCite's mission, build strategic partnerships and work with a diverse community of stakeholders.</p>	<p>Robin Dasler, Product Manager, is our newest. Robin oversees development of products. Before joining DataCite, Robin was a senior fellow in CERN's Scientific Information Service and lead of the sustainability segment of the THOR Project.</p>
	
<p>Britta Dreyer, Business Manager, provides member support, manages our finances and, and managing DataCite's day-to-day business are her main activities. She holds a BSM from the Pepperdine University and an MBA from the University of Applied Sciences and Arts in Hannover.</p>	<p>Martin Fenner, Technical Director envisions, develops, implements and manages a robust technical architecture for DataCite as well as DataCite's technical contributions for grant funded initiatives. Until 2015 he was technical lead for the PLOS Article-Level Metrics project.</p>
	
<p>Kristian Garza, Application Developer, contributes to the development and implementation of a robust technical architecture for DataCite as well as DataCite's technical role in the Make Data Count Project. Improving to the community-wide adoption of good data sharing practices is one of his main goals. Kristian holds a PhD in Computer Science from the University of Manchester.</p>	<p>Richard Hallett, Application Developer, joined DataCite in late 2017. Having worked in several for years in software development for a variety of different organizations and industries (from large multinational companies, to small web agencies, to not for profits and then back into the commercial sector with e-commerce), Richard is pleased to be working in an open-source environment.</p>