

DataCite Annual Review 2019



Index

Who We Are	3
Message from the Executive Director, Matt Buys	4
Message from the Executive Board President, John Chodacki	6
Message from the Executive Board Treasurer, Marco Marsella	7
DataCite Community	8
DataCite Services and Infrastructure	12
Projects and Funded Initiatives	17
Financial Overview	19
Executive Board Roster	21
Staff Roster	24

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,
Matt Buys

Since our founding in 2009, DataCite has collaborated closely with the research community to expand our services, increase our membership, and promote an open culture of data sharing. As a global leader in the field of research data, we embrace the opportunity to further develop and promote the research infrastructure of the future by engaging with our members, taking a multidisciplinary approach to research data, and cultivating the technical expertise to deliver services that benefit the entire research community!



2019 was an exciting year for me personally, as I joined DataCite in October and I have really enjoyed getting to know our community. We are a small passionate team of staff that are dedicated to delivering value to our members and broader community. DataCite really embraces the ethos of a community of practice, where our community stakeholders work together with a common mission in providing the means to create, find, cite, connect, and use research data.

In 2019, we said farewell to Trisha Cruse who has been involved with DataCite since our launch in 2009; and recently led the DataCite staff as the Executive Director since 2015. Trisha's leadership of DataCite has made a significant impact and she will be missed by us all. She left DataCite a much stronger organization, and I am very grateful for her support and guidance during the transition period.

Looking ahead to 2020, we will be focusing on four strategic priorities:

2020

..... **Formalizing the new business model.**

As a membership organization, members are at the core of everything we do and it is important that we are collectively able to work on DataCite's strategy and long term sustainability.

..... **Optimizing internal operations, systems and processes.**

Improving our internal systems to ensure that we continue to scale optimally and can support our growing community.

..... **Consolidation of services and infrastructure.**

Make our services easier to use for members, and easier to maintain for staff, by continuing the consolidation towards a single foundational API, and a single web interface.

..... **Continuation of member driven product development.**

As a community driven organization, we continue to improve our processes and decisions. This means that all decisions we make about our services and infrastructure are driven by community input and validation.

.....
Together with the DataCite team, I look forward to continuing to build our community, delivering high-quality services, and ensuring our organization's future sustainability.

Executive Board President, John Chodacki



2019 was an important year for DataCite. DataCite celebrated its 10th anniversary and looked back at impressive growth, in terms of both membership and services. In its 10th year, DataCite also welcomed a new Executive Director, Matthew Buys. Matt came to us from ORCID bringing a wealth of experience in scaling Persistent Identifier infrastructure. We said goodbye to Patricia Cruse, who made an invaluable contribution to DataCite during her years on the Board and as Executive Director.

DataCite is a robust community of libraries, research institutions, and data centers that house data repositories. As a community organization, it is crucial that all members have a voice in DataCite decisions. To ensure this was the case, 2019 was set aside as a time for us to clarify our path forward as a sustainable community organization and involve the entire membership in determining our path forward. We are confident that together we are poised for 2020 to be the year that our organization moves to a sustainable member model that allows all of us to move forward.

DataCite continues to be an important stakeholder in the PID and data communities. Through our core infrastructure and our continuing project work including the Make Data Count and FREYA projects, we are able to provide the community with new value-added services. In 2019, we also joined likeminded organizations in launching ROR, the Research Organization Registry. The quick uptake of ROR by both our members and the larger community demonstrates that there was a real need for an open, community-owned identifier and for DataCite to remain active in community projects.

Looking forward, I am excited to have Matt Buys as our new Executive Director and, under his leadership, know that DataCite will continue to work closely with our community to ensure we add value to your research and workflows. I look forward to continuing the path we are on together.

Executive Board Treasurer, Marco Marsella



DataCite's finances are in good shape. Expanding membership as well as participation in a number of high-profile projects produced a steadily increasing income that more than compensated the corresponding increase in expenditures. During 2019, our project-related income represented 40% of the total, reducing DataCite's dependency on projects. This allows us to focus on the best proposals while ensuring that core services remain sustainable.

Measures have been taken to rationalize expenditures leading to lower costs despite significant one-time setup fees for new, improved management tools. I feel confident that DataCite is financially in a good position and will continue on the path of decreasing the dependency on grant funding.

DataCite Community

We welcomed many new members to our community this year and we are proud to continue to provide services to our growing community:

1900+
Repositories

189
Members

41
Countries

19m +
DOIs

New members

In 2019, DataCite saw substantial member growth: 46 new members joined DataCite, from 13 different countries. This shows that we are a truly global organization and that the importance of research data and the persistent identification or research outputs is increasingly recognized around the world. Welcome to these new members!

NEW MEMBERS

- | | |
|---------------------------------------------------------|----------------------------------------------------------------------------------|
| | |
| 1. Michigan State University Libraries | 26. National Institute of Allergy and Infectious Diseases (NIAID) |
| 2. Blackfynn Inc. | 27. Rochester Institute of Technology - Research Computing |
| 3. Kudos Innovations Ltd | 28. FSC Global Development GmbH |
| 4. Center for Severe Weather Research | 29. Deutsches Zentrum für Luft- & Raumfahrt e.V. (DLR) |
| 5. HEAL-LINK
(Hellenic Academic Libraries Link) | 30. Temple University, of the Commonwealth System of Higher Education |
| 6. The University of Chicago - Globus | 31. Telkom University |
| 7. Carnegie Mellon University | 32. Knowledge Inc. |
| 8. Cambridge Crystallographic Data Centre | 33. Elettra Sincrotrone Trieste S.C.p.A. |
| 9. RoyaltyStat | 34. Canadian Research Knowledge Network (CRKN) |
| 10. Broad Institute | 35. Idiap Research Institute |
| 11. CyArk | 36. Universidad del Rosario |
| 12. Clarivate Analytics | 37. European Multidisciplinary Seafloor and water column Observatory (EMSO ERIC) |
| 13. Aridhia Informatics Ltd. | 38. Spallation Neutron Source Science Center |
| 14. Group on Earth Observations | 39. London Business School |
| 15. Ludwig-Maximilians-Universität München (LMU) | 40. Centre National pour la Recherche Scientifique et Technique (CNRST) |
| 16. University Library System, University of Pittsburgh | 41. Deutsches Archäologisches Institut |
| 17. Universidad de Chile, Library System | 42. Open Society Institute |
| 18. Northwestern University Libraries | 43. University of Vienna |
| 19. PhysioNet | 44. Open Imagery Network, Inc. (DBA Radiant Earth Foundation) |
| 20. Gigantum Inc. | 45. Yale NUS College |
| 21. Forschungszentrum Juelich GmbH | 46. Deutsches Klimarechenzentrum (DKRZ) |
| 22. TU Wien Bibliothek | |
| 23. CLARIN ERIC | |
| 24. Princeton University Library | |
| 25. Kaggle | |
| | |

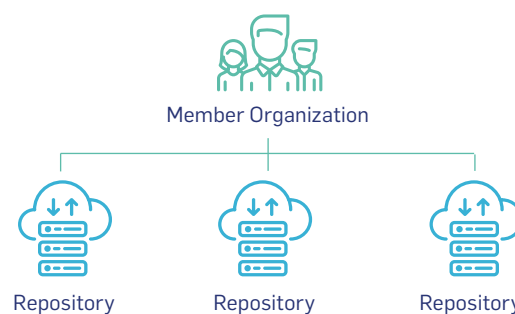
The member model

2018 saw some changes to the member model that were further developed and discussed throughout 2019. We introduced the concept of repositories, which we define as a service operated by research organizations, where research materials are stored, managed and made accessible. A repository is a single unit and DataCite links the repository to information in re3data, where additional repository meta-data are available. You can see all repositories you are working with in DataCite Fabrica.

Another change we made is that we now distinguish between providing services to internal and external repositories. If members only work with repositories within their own organization, they can join as a Direct Member. If they provide DOI services to external repositories, they can form a consortium together with the organizations they are providing the services to. Several of our existing members started the process of establishing consortia, something that will continue throughout 2020.

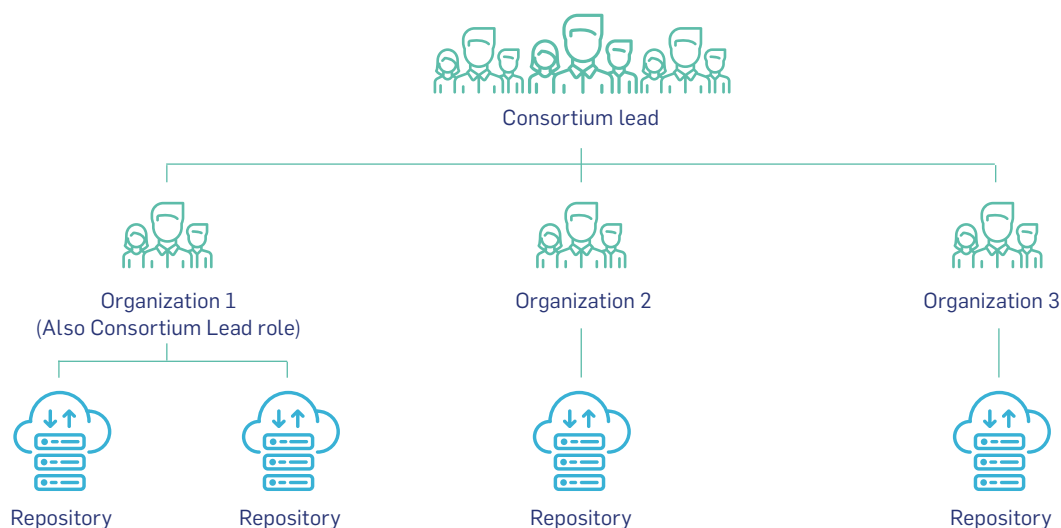
DIRECT MEMBER

This type of member is a consumer of DataCite's DOI services and consists of an organization that works with one or more repositories within their immediate organization. The repositories are under the same administrative structure as the organization.



CONSORTIUM MEMBER

A consortium is a group of like-minded organizations that have come together to collectively participate in DataCite's community and governance activities and use DataCite's DOI services. A consortium is composed of five or more non-profit organizations that are under different administrative structures. Consortia are generally located in a single country or subject based. Organizations within a consortium can work with one or more repositories.



MEMBER CONSULTATION

The change to the member model led to some questions about the DataCite fee structure. Members expressed concerns that fees might be too high for small organizations registering low numbers of DOIs. Therefore, at the end of 2019, we embarked on a member consultation to ask members for their input on the fee structure. An advisory group is looking at the outcomes and developing a proposal that will be voted on at the next General Assembly business meeting. Both the updated member model and fee structure will be implemented throughout 2020.

DataCite Services and Infrastructure

Overview of DataCite

The DataCite technical infrastructure has been hosted with Amazon Web Services (AWS) since 2011. Some services are global, all infrastructure hosted in a particular region is in AWS EU-West-1 (Ireland). DataCite uses a standard architecture with the following major components:

ARCHITECTURE COMPONENTS

.....

1. Virtual Private Cloud (VPC) with a public and private subnet.
.....

2. All application servers, databases and file storage are hosted in the private subnet, not directly accessible from the internet.
.....

3. The public subnet connects the private subnet to the internet via application load balancers (ALB), a content delivery network (CDN, Cloudfront) and bastion host (SSH access for staff) for incoming are only accessible via https. On top of these services sits a web application firewall (WAF) to protect our VPC from unreasonable requests, currently focusing on rate-limiting overly high traffic that would otherwise overload our system.
.....

.....

4. Our VPC hosts a production and test system. The test system is an exact replica of the production system (but using smaller resources, e.g. the database) used by staff and members. Every resource in the production system is duplicated for high availability, in different AWS zones if possible (e.g. for databases).

.....

5. The AWS infrastructure is managed by the [Terraform](#) service, which treats infrastructure as code. All configurations are stored in a (public) GitHub repository, and configuration changes in that repository trigger automatically trigger the intended changes in our infrastructure.

.....

6. For service monitoring, we mainly rely on AWS Cloudwatch but export the data (e.g. application server logs) to the [DataDog](#) service for better accessibility.

.....

7. The databases (MySQL and Redis) and search engine (Elastic-search) run as services managed by AWS.

.....

8. All compute resources run as Docker containers, with three small virtual machines (EC2) left for special purposes (e.g. bastion host). The Docker containers are orchestrated (managed) by the AWS ECS Fargate service.

.....

9. All static websites (e.g. homepage, blog) are stored in the Amazon file service (S3) and accessed via a content delivery network (CDN) with multiple edge locations worldwide.

.....

10. All data are stored in file storage or databases, and the latter are backed-up every 24 hours.

.....

Highlights of 2019

Services and Infrastructure Work

In 2019, DataCite made many improvements to existing services, and set ourselves on a path toward further streamlining and consolidation of products and services.

REPLACE OUTDATED SOLR WITH ELASTICSEARCH

Solr and Elasticsearch are the two most commonly used search engines. DataCite started using Solr in 2012, but in 2018 we faced two major challenges: a) the Solr version (4.0 from 2012) we were using had become obsolete and the upgrade to a recent Solr version would have been a major undertaking, and b) hosting Solr in our new Docker-based infrastructure was a challenge, and we wanted to migrate to a managed service. For the latter reason, we switched to Elasticsearch (where AWS is providing a hosted service similar to their database services) instead of upgrading our Solr service.

Like all major changes in the DataCite infrastructure, the main challenge was a smooth transition from Solr to Elasticsearch that didn't interrupt any user-facing services and kept backward compatibility where possible, e.g. in the OAI-PMH service. This is the main reason that the transition took 18 months, with Solr and Elasticsearch running in parallel until retiring Solr in December 2019.

The biggest visible change to users of this transition to Elasticsearch is that indexing of new or updated DOIs now takes a few minutes instead of the 24-hour interval with Solr.

DATAcite GraphQL API

One of the major contributions of DataCite to the EC-funded FREYA project is the PID Graph, allowing users to explore the graph of scholarly resources connected via persistent identifiers. The use cases for the PID Graph have been explored in the initial phase of the FREYA project, in the spring of 2019 we picked GraphQL as the underlying technology that powers the PID Graph. [GraphQL](#) addresses the needs identified in the PID Graph use cases we had collected, enabling complex queries across multiple resources. What made GraphQL particularly attractive is the flexibility of backend services, allowing us to integrate existing DataCite infrastructure that uses relational databases and the Lucene-based search indexes. In addition, GraphQL is a mature, widely adopted open source technology. The DataCite GraphQL API using [Apollo Federation](#), integrating data from external PID services such as Crossref Funder ID, Research Organization Registry (ROR) and re3data.

To simplify working with the DataCite GraphQL API, we started to use Jupyter notebooks, which make it very easy to fetch PID Graph data and visualize them. In the FREYA project, we have done two workshops/hackathons using Jupyter notebooks for PID Graph in 2019, and will do more in 2020. We have also used the Research Data Alliance as a platform, starting an [Open Science Graphs for FAIR Data Interest Group](#).

IMPROVEMENTS TO FABRICA

We had previously launched the Fabrica service in 2018 as a new web frontend for members to interact with DataCite services, replacing the MDS web interface that was retired in November 2018. The focus of Fabrica is DOI registration and management, but it is also used to manage DOI prefixes and accounts for members and repositories.

In 2019 we spent a lot of time working on the Fabrica service, for example by improving the DOI registration form, and by adding support for the new consortia member model. We made good progress in improving the underlying technology that included many bug fixes and small improvements.

DISPLAY OF CITATIONS AND USAGE STATS

While DataCite has for many years been working on infrastructure to collect and expose information about the reuse of research data, we in 2019 worked hard to make this information also directly available to users via the DataCite Search web service. We [launched this important new functionality in January 2020](#).

For this work we pulled together information from the Event Data Service, provided by DataCite and Crossref. One difficulty was deciding how to deal with the more than 20 relation types the DataCite metadata schema makes available to describe links between content with a DataCite DOI and other resources. The implementation that was launched in January 2020 is based on the Scholix schema and distinguishes between citations (incoming links), references (outgoing links), and relations (other kinds of links to scholarly resources, e.g. versioning) as major categories, grouping all possible relation types into one of these categories.

Going forward, we will work on improving the performance of this new data citation service, will integrate it into Fabrica, and will work with DataCite members and the broader community to iterate how this information is presented.

UPDATING AND REALIGNING ANCILLARY SERVICES

We made significant changes to how members interact with the test environment in 2018, and 2019 saw us finishing out those changes by retiring the test prefix. The test prefix was used for several years in our production environment, and its removal helped us to create a harder line between production services that result in lasting, permanent DOIs and test services that do not.

We also updated our content negotiation service this year, removing support for custom content types. This will help us better align the service with its original purpose of providing multiple standard formats for metadata.

As part of our work on the FREYA project, we introduced a new /activities endpoint for our REST API that exposes DOI metadata provenance. This allows our members to see when and by whom changes to the DOI record were made.

The DataCite Profiles service is what enables users to login to public-facing DataCite services in order to, among other things, claim DOIs to their ORCID record. In 2019, we adopted authentication services by Globus, a non-profit initiative at the University of Chicago, so that we could enable our users to login with their institutional accounts in addition to the ORCID and Google logins previously supported.

UPDATES TO THE METADATA SCHEMA

In 2019, our Metadata Working Group released two minor versions of the DataCite Metadata Schema. Version 4.2 added new descriptive properties for DOIs, most notably sub-properties for Rights to allow you to specify a rights identifier. Version 4.3 added support for ROR IDs as part of the long-awaited addition of creator affiliations.

Projects and Funded Initiatives



maDMP

Project Title: maDMP

Funder: National Science Foundation, EAGER Grants

DataCite Funds: € 92272.69

Duration and Start Date: 24 mo., January 1, 2019

Partners: California Digital Library and DataCite

Objectives: maDMP focuses on creating metadata and DOIs for DMPs so they can be connected to other components of research.



FAIRsFAIR

Project Title: FAIRsFAIR

Funder: European Commission, Horizon2020 program

DataCite Funds: € 242500

Duration and Start Date: 36 mo., January 1, 2019

Partners: The project is a pan-European collaboration of 22 partners with six core partners: DANS in the Netherlands (lead), CSC in Finland, the DCC and the STFC in the UK, Trust-IT in Italy and the European University Association (EUA) in Belgium.

Objectives: This project supplies practical solutions for the use of FAIR data principles throughout the research data life cycle. DataCite will collaborate with Re3data to make FAIR repositories discoverable.



FREYA

Project Title: FREYA

Funder: European Commission, Horizon2020 program

DataCite Funds: € 877500

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

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).

Objectives: FREYA, the successor to THOR, got underway in 2018. 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.

Project Title: PARSEC: Building New Tools for Data Sharing and Reuse through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas

Funder/ Sponsors: French National Research Agency, France; National Science Foundation, USA; São Paulo Research Foundation, Brazil; Japan Science and Technology Agency, Japan

DataCite Funds: € 177070

Duration and Start Date: 24 mo., May 1, 2019

Partners: American Geophysical Union, Ecological Society of America, European Geosciences Union, Japan Geoscience Union, Earth Science Information Partners, RDA, etc.), ORCID, DataCite, Research Data Alliance.

Objectives: PARSEC will address the following two issues: 1) large amounts of research data related to the Earth and its ecosystems are either not well preserved or preserved at all. 2) there is also limited information on how diverse data are re-used for research and quantifying the value of curated data for such purposes, and how the quality of data preservation affects these outcomes. DataCite will therefore provide a widget and researcher profile displaying data usage statistics and citations.

COREF

Project Title: re3data COREF

Partners: DataCite, Karlsruhe Institute of Technology (KIT), the Berlin School of Library and Information Science at the Humboldt-Universität zu Berlin and the Helmholtz Open Science Office at the German Research Centre for Geosciences (GFZ)

DataCite Funds: € 130406

Start Date: January 1, 2020

Objectives: re3data COREF aims to connect re3data as the reference for research data repositories with other services and infrastructures. By providing customizable and extendable core repository descriptions that are persistently identifiable and can be referred and cited in an appropriate manner, re3data shall be advanced to facilitate the reuse of reliable and trustworthy information on research data repositories.

Project Title: Research Organization Registry (ROR)

Partners: Crossref, DataCite, Digital Science

Start Date: July 1, 2018

Objectives: ROR was launched in January 2019 and provides organization identifiers that are globally unique, stable, discoverable, and resolvable. In addition, ROR will develop an appropriate metadata schema for organizations and explore interoperability with other identifiers through relationship metadata.

ORCID-DE 2

Project Title: ORCID-DE 2

Partners: DataCite, Helmholtz Open Science Office at the German Research Centre for Geosciences GFZ, the German National Library, Bielefeld University Library and the German National Library of Science and Technology (TIB) Hannover

DataCite Funds: € 74940

Start Date: January 1, 2020

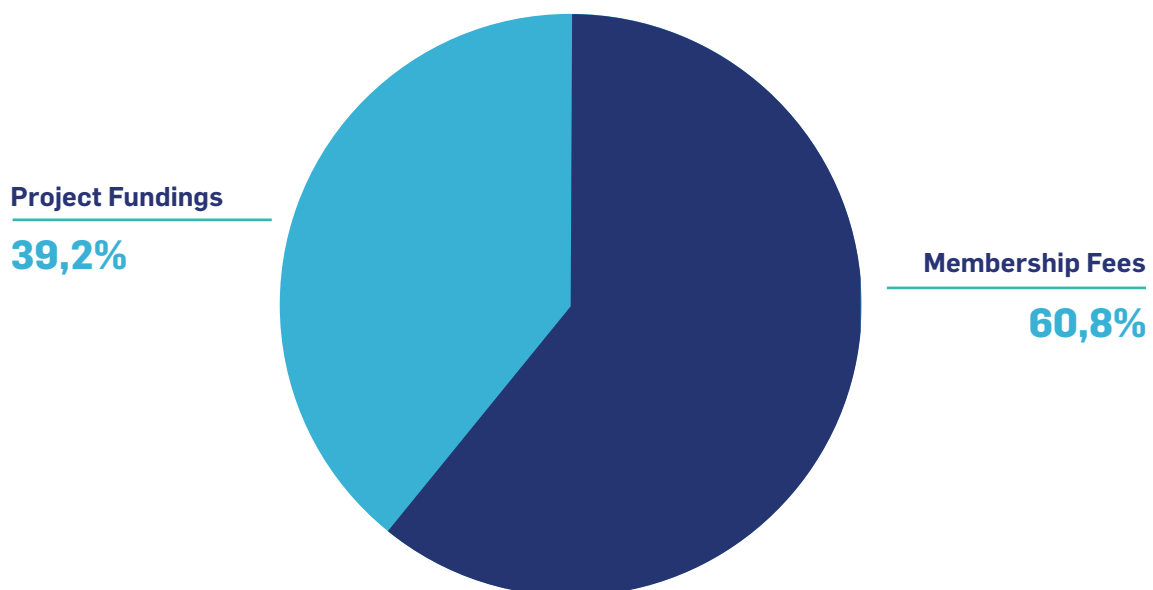
Objectives: During its 30-month project period ORCID DE 2 seeks to expand and consolidate the existing network of scientific institutions that have already integrated ORCID in their infrastructures. One major goal is the increase of the support for institutions and target groups interested in and using ORCID. Furthermore, in the course of ORCID DE 2, a survey about the need of an identification system for organizations will be conducted and implementation explored.

Financial Overview

Based on pre-audited financials our 2019 net operating income was € 36K over budget (39.36%), resulting in € 135K surplus. This can be attributed to continued membership growth and continued cost controls. The surplus will support the expected 2020 deficit as we transition to the new membership model. Our 2019 cash balance is compliant with § 62 Abs. 1 Nr. 3 AO (10%) considered as free reserves and compliant with § 62 Abs. 1 Nr. 1 AO as the remaining reserves are used for pre-financing.

Revenue

Revenue continues to track positively with 46 new members during 2019. Membership revenue was €848K (6.95% above budget); and project funding €548K (2.66% under budget). Total revenue was €1.396K

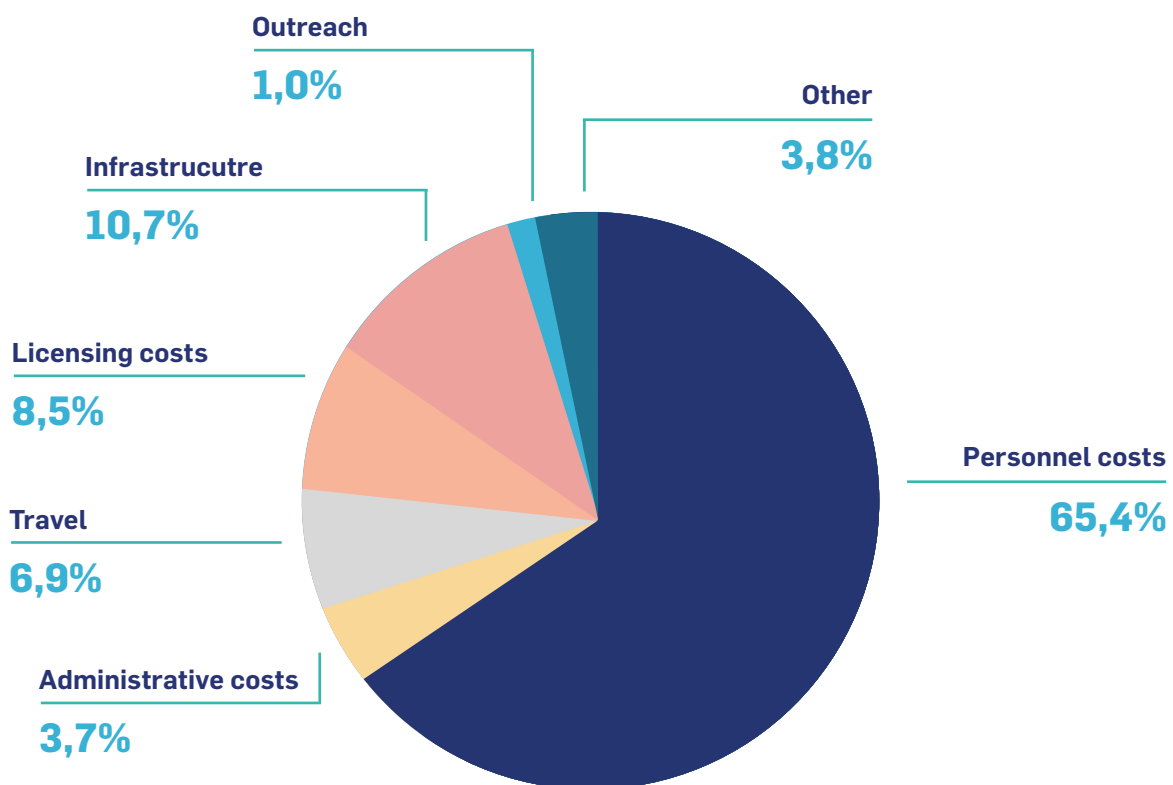


Expenses

Our expenses during 2019 were under budget by € 4K under budget (0.33%). The following can be noted: Our administrative costs were higher than budgeted due to one-time configuration costs (€ 21K) for the new CRM (Salesforce) and billing system (JustOn). We have budgeted the recurring costs (€ 5K) for the billing system in 2020. The updated system ensures we can scale efficiently, provide accurate reports and comply with German tax requirements (DATEV export).

Other licensing covers various tools across the organization. Datadog has been the main cost driver in this line item. We have created Datadog dashboards for monitoring performance and usage of services. This allows us to track the number of requests, response time, CPU and memory utilization for the REST API in real time.

Our infrastructure costs were higher than budgeted as we were retiring Solr and transitioning to Elastic-search, a change that also improved indexing and search functions. Last month we purchased reserved instances for Elastic Container Service, ElasticSearch Service and Relational Database Service. The other services are using on demand pricing.



Executive Board Roster



JOHN CHODACKI

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.



ADRIAN BURTON

Dr. Adrian Burton is Director of Data Policy and Services at the [Australian Research Data Commons](#) which is a government-funded research infrastructure for data, cloud, and tools.



MOHAMED A. BA-ESSA

Mohamed is Manager of Preservation and Digital Services at [King Abdullah University of Science and Technology \(KAUST\)](#), where he is responsible for Research Repository and Research Data Services, University Records Management and Archive, and Library Systems and Digital services.



MARK HAHNEL

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.



VIV HUTCHISON

Viv Hutchison is the Science Data Management Chief for the [US Geological Survey](#). Her team develops and maintains a suite of science data management applications and leads a prominent open forum called the Community for Data Integration.



MARCO MARSELLA

Marco Marsella 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.



SALVATORE MELE

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.



TORSTEN REIMER

Torsten Reimer is Head of Research Services at the British Library. His teams are developing the collection and services supporting researchers and research organizations, onsite and online, including digital preservation, repository, data and identifier services.



REBECCA ROSS

Rebecca Ross is the Director of Strategy and Engagement at the Canadian Research Knowledge Network (CRKN), Canada's national collective negotiating body for academic libraries in Canada. Rebecca is responsible for overseeing strategic planning, stakeholder engagement, and marketing and communications of CRKN's diverse portfolio of licensing and digitization, preservation and access.



IRINA SENS

Irina Sens is the Deputy Director at the [German National Library of Science and Technology](#) (TIB). She led the German project "Registration of Scientific primary data" which was responsible for the introduction of DOIs for research data.



HERBERT VAN DE SOMPEL

Herbert Van de Sompel is Chief Innovation Officer at [Data Archiving and Networked Services](#) (DANS) in The Netherlands. He has a long-term interest in infrastructure to support scholarly communication in the digital era.



JAMIE WITTENBERG

Jamie Wittenberg is Head of Scholarly Communication at [Indiana University Libraries](#), where she oversees research data, open access, and publishing services. Her team collaborates with faculty and students to curate, preserve, and share their work.

Staff Roster



MATTHEW BUYS

Matthew Buys, Executive Director, leads a passionate and committed team at DataCite who provide the means to create, find, cite, connect, and use research globally. Based in Amsterdam, Matthew focusses on building DataCite into a sustainable global community. Previously Matthew was the Director of Engagement at ORCID where he played an important role in growing the community into an international-scale research effort with over 1,000 member organizations and 7 million users. He completed a BA (Psychology) and Post-Graduate Diploma in Management at the University of the Witwatersrand. In addition, Matthew has also completed courses in Java, Flash, XML, html, Perl, and SQL.



ROBIN DASLER

Robin Dasler, Product Manager, oversees development of products and contributes to the EU-funded FREYA project. Before joining DataCite, Robin was a senior fellow in CERN's Scientific Information Service and lead of the sustainability segment of the THOR Project.



HELENA COUSIJN

Helena Cousijn, Community Engagement Director, is responsible for all DataCite's membership and community activities. She's committed to DataCite's mission of enabling data sharing and reuse and is especially passionate about data citation. It's important to her to communicate in a way that makes DataCite's services accessible to everyone. Before joining DataCite, Helena worked as Senior Product Manager for Research Data Management Solutions at Elsevier. She holds a DPhil in Neuroscience from the University of Oxford.



BRITTA DREYER

Britta Dreyer, Business Manager, manages the DataCite business office. Member support, accounting, and managing DataCite's day-to-day business are her main activities. She holds a BSM from the Pepperdine University in Malibu and an MBA from the University of Applied Sciences and Arts in Hannover. Before coming to DataCite, Britta managed the family business. Her fields of work include human resources, marketing, and process optimization. She is



MARTIN FENNER

Martin Fenner, Technical Director, envisions, develops, implements and manages a robust technical architecture for DataCite. Until 2015 he was technical lead for the PLOS Article-Level Metrics project. Martin has a medical degree from the Free University of Berlin and is a Board-certified medical oncologist.



RICHARD HALLET

Richard Hallet, Application Developer, joined DataCite in late 2017. He contributes to the DataCite technical infrastructure and the EU-funded FREYA project. Having worked 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.

passionate about business planning, project management, communication psychology, and her three children.



KRISTIAN GARZA

Kristian Garza, Application Developer, contributes to the development and implementation of a robust technical architecture for DataCite as well as DataCite's role in the Make Data Count Project. Improving to the community-wide adoption of good data sharing practices is one of his main goals. Previously he held different research positions within the CLUSTER II Mission at the MSSL and at the ELIXIR Interoperability platform. Kristian has a PhD degree in Computer Science from the University of Manchester.



MARY HIRSCH

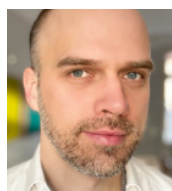
Mary Hirsch, Member Support Manager, takes care of making sure DataCite provides exceptional support for its community. Mary worked as a technical analyst at a global information company before transitioning to work as a documentalist at a research institute in Barcelona. She is passionate about Open Science.



SARALA WIMALARATNE

.....

Sarala Wimalaratne, Head of Infrastructure Services, is responsible for the robust technical infrastructure services provided by DataCite. She has been working with the Open Science and PID communities for many years. Before joining DataCite, Sarala spent 10 years at the EMBL-EBI leading multiple data integration projects including the Identifiers.org resource. She holds a PhD in Bioengineering and a BE in Software Engineering from the University of Auckland.



PAUL VIERKANT

.....

Paul Vierkant, Outreach Manager, contributes to the DFG-funded projects re3data COREF and ORCID DE. Prior to DataCite Paul worked for the Helmholtz Open Science Office and different universities where he was involved in building publication and data repositories. As a dedicated Open Science advocate Paul strives to spread the idea of openness in scholarly communication. Paul loves beautiful infographics and his crazy family even more.



DataCite Annual Review: 2019

Identify, cite, and connect research
with confidence.

www.datacite.org

