



10 Key Interviews

INSIGHTS INTO THE SUSTAINABILITY OF
OPEN INFRASTRUCTURE SERVICES

01

An Invest in Open Infrastructure project
More info: www.sparceurope.org/ioiinterviews



ABOUT THE PROJECT

Gleaning Insights

Research and scholarship is underpinned by a variety of tools, technologies and services ranging from for-profit commercial solutions and offerings from vendors to community-owned, open technologies and infrastructure. We often hear about the challenges for open infrastructure tools and services to scale, maintain, and compete in the broader market.

The 10 interviews comprised in this project highlight some of the key decision-making points, funding mechanisms and models, and other learnings from a series of commonly used services and technologies used to support research and scholarship. These include both for-profit and not-for-profit services, highlighting perspectives on sustainability across the sector.

This work is supported by Open Society Foundations and SPARC Europe, in collaboration with Invest in Open Infrastructure.

02 An Invest in Open Infrastructure project
More info: www.sparceurope.org/ioiinterviews



WHO WAS INTERVIEWED?

Overview

Dryad *Featured in this document.*

Code Ocean

Figshare

EDP Sciences

F1000 Research

Mendeley

Our Research

arXiv

Redalyc

4TU.Research.Data



Interview: Dryad

**"Pick your partners
carefully, a more open
future will only happen if
we choose partners
based on our values and
our mission"**

Tracy Teal, Executive Director, Dryad
John Chodacki, Director, University of California
Curation Center (USA)

DRYAD

Tracy Teal and John Chodacki



At a Glance

Type of activities:

An open and curated data-publishing platform for researchers to share and publish their data

Life-cycle stage:

Started as a grant-funded project in 2007 and matured over time. Since 2018 in a strategic partnership with California Digital Library (CDL). Relies on a diverse set of revenues -- a balance of data publishing charges from individual researchers and membership fees from institutions, publishers, and funders.

Current legal structure and funding model:

Independent 501(c)(3) not-for-profit (without shareholders), with a business model of organizational membership, data publishing charge, and grants.

Technology:

Open source

Sustainability

'Sustainability is the certainty that you have the resources to pay your bills and invest in your future. It is also the ability to look to the horizon and have a clear, values-driven plan for the long term.'

Piece of Advice

'Pick your partners carefully. A more open future will only happen if we choose partners based on our values and our mission.'

'Non-profit, values-driven organizations are really the most sustainable of all projects.'

DRYAD

Dryad is an open and curated data-publishing platform where researchers can publish and share their data. It started with a National Science Foundation (NSF) grant awarded to a project at an interdisciplinary centre, The National Evolutionary Synthesis Center, NESCent, focused on biology based out of Duke University in 2007. It transitioned to an independent 501(c)(3) not-for-profit company in 2010. Since 2018, it has been in a strategic partnership with California Digital Library (CDL).

Original vision

We spoke to Dryad's Executive Director, Tracy Teal, and John Chodacki, Director of the University of California Curation Center (UC3, which is the research data management and digital preservation program of CDL). They explain that Dryad was founded by the researcher community in the ecology and evolution fields, who sought better support for data sharing. 'Dryad is unique because it was founded by researchers who were trying to find a solution to the issues they faced with data sharing.'

'Dryad has always had a very close partnership with research and research institutions. It understands the research process of the institution and the journey of a research project, including the publishing and preservation workflows that researchers work with. What is core to Dryad's work is to be very mindful of these things and to make sure to complement them as much as possible.'

'Researchers were the real pioneers,' says Chodacki. 'When Dryad was first founded, researchers were building something to meet their own needs because nothing else was available. There still isn't anything else like Dryad. Dryad was founded to fit within this specific use case: Researchers work within their own research project, develop an idea and determine if it is successful or not, and along the way their research outputs are created. The different touchpoints in that process are pretty predictable, and one of them is publishing. Researchers are constantly working towards that point where they can speak with their communities about what they learned. Dryad is a solution that researchers themselves developed to fit into the way researchers work, centred on the idea that research data is a key part of the research workflow. Hence, when it started, those initial researchers got together and charted the future of data publishing.' With grant funding and in-kind support from the host institution, the initiative built from its core strengths. 'Over

time, the project matured and evolved just as did data publishing. One constant is the relationships Dryad has built with researchers and with institutions. Support from that community is a very good thing, and very often naysayers disregard it or think of it as a crutch. It is actually a legitimate and viable approach to sustainability. And *non-profit, values-driven organizations are really the most sustainable of all projects.*'

Growth and sustainability challenges

Dryad was initially funded by a grant from the NSF in 2007. Principal Investigator-ship was at Duke University and UNC. In 2010, it became a 501(c)(3). 'We started our history within the research triangle of North Carolina, so we have a proud history of working with research institutions. The NSF grant was awarded to the National Evolutionary Synthesis Center (NESCent), an interdisciplinary centre focused around biology, based out of Duke, but bringing together people in that region.'

NON-PROFIT, VALUES-DRIVEN ORGANIZATIONS ARE REALLY THE MOST SUSTAINABLE OF ALL PROJECTS

There was a lot of collaboration between Duke, UNC and Dryad, the team. As time went on, there was also the need to think about what the best model was, the best way of managing the project, and how to deal with Dryad as a separate legal entity. It was part of the original idea to evaluate what the best model would be. When Dryad was founded around 2007, it was founded as a project, not just to be the best data sharing project for researchers, but also for it to find its own best way of managing itself. During these two-three years, Dryad had guidance from researchers who were advising and actively working on the project. They were also planning for a sustainable set-up. They made the decision around 2010 to become a 501(c)(3), which was formally established in 2012 when all the paperwork was done. One of the goals was maintaining the connection to academic institutions, and a 501(c)(3) allows for more possibilities for funding in that respect, as it allows for benefiting from grants as well as for recovering the costs of offering services.'

'Starting as a project and taking the time to decide on the structural set-up, I think, is a great approach, because it gives you the time

to get started and figure out what the best model is as you go; not having to do it all at the same time. If at the beginning of a project, you realise it is not the right fit, it's harder to change. You give yourself time to intentionally make those kinds of decisions and plan, considering: What kind of hiring do we need to start doing? How are we going to get revenue in a structured way? I think it was a very thoughtful process. This deliberate process also separates Dryad from commercial projects that are forced to bet it all on speculative ventures or to look for investment from partners that are not mission-aligned. We have been very thoughtful about how to best support the Open Science future we are all striving for.'

CDL and Dryad formed an alliance in 2018. 'The UC3 team works on research data management projects within CDL and we work across the entire University of California system on topics like data publishing and data sharing. We had a project when we started: a data repository called "Dash" that we wanted to improve. We sought grants, hired a product manager to think about strategy and entertained working with different open

source communities around different platforms. We set goals but we failed to achieve them, including one that aimed to acquire hundreds of thousands of datasets only resulting in 100 that year. We also strove to integrate into publisher workflows like PLOS, Elsevier and Springer Nature, but they did not want to do this for just one institution. We came to a point where we were really looking for ways to break out of this situation to meet our original goals. We asked ourselves: Why do we as institutions invest resources into single institutional approaches that don't meet the set goals? So instead of trying to bring researchers to a bespoke UC solution, we flipped our strategy and decided to invest in the solution that the researchers were already using: Dryad. As a result, we started discussions with Dryad about putting our resources into their project.'

The partnership came at a good time for Dryad as well, says Teal. 'Dryad was interested in the CDL partnership because it offered an ability to jump-start updates to our technical platform. Since we were working at cost, this meant that we couldn't make certain technical investments. Many institutions have to work

on shoestrings, which can often turn into dead ends. However, through partnerships with like-minded organizations, we can expand our capacity, reach and advance our cause.'

Teal and Chodacki are of the opinion that partnerships can guarantee long-term sustainability. Chodacki: 'Sustainability is two things: one is being able to pay the bills, have cashflow and have the ability to not live in a sense of urgency or to be in fear of short-term or long-term crises. Second, there needs to be a plan, and that plan, the horizon, needs to be a far-reaching one to ensure that there is stewardship in the long run, especially in the space of data stewardship: looking at a 20–50–100-year horizon. Commercial products can't offer that security. It's really only in partnerships like the Dryad-CDL one where the organizations are focused on the horizon and not on short term sales strategies and profiteering that you can see the true sense of the word: sustainability.'

Opportunities, considerations and choices

Dryad offers membership to institutions, publishers, and funders. All members are able to cover the charges for their affiliated researchers. Members are Dryad's governing group as a non-profit. The strategic partnership between Dryad and CDL <https://www.eurekalert.org/pub_releases/2018-05/cdl-adp052918.php> is a

memorandum of agreement between the two parties around resourcing. It provided a solution for the challenges both organisations faced: 'CDL offers in-kind support of two full-time equivalents (FTEs) to Dryad, and in response, all UC campuses are institutional members and, with that, the cost per dataset for UC's researchers is waived. Dryad remains its own organisation, its own 501(c)(3). From a governance perspective, CDL is just one of the dozens of members. Other institutions join Dryad and their membership is not affected by CDL. Through CDL's resource sharing, it offers the entire Dryad community an additional anchor of stability that other members respect.'

'We're investing our research data stewardship resources where they will have most impact, and that is with Dryad.' This doesn't mean that CDL has any preferential treatment on governance. This means that Dryad is available for UC researchers at no cost to them and we see UC data publishing rates going way up. Going back to UC's initial goals, we're much more successful in driving adoption on the core goal: publishing datasets Teal continues: 'The partnership agreement is actually pretty standard, but its strength is that it is really a true partnership. We talk about strategic plans together going forward, and the fact that CDL is a library is also very important for Dryad. We think about staying connected to our communities, planning for the long term, with sustainability being a sort of ecosystem

together rather than a siloed project that doesn't engage actively with other groups. All great partnerships allow room for others to join and benefit.'

Dryad has dozens of institutional members and membership continues to grow. The Dryad network can also help on a practical level in curation: the core Dryad service. It does this by assigning DOIs to a dataset, which can be done by many groups. Dryad is aiming to be the best data curator, so our partnerships with such in-kind contributions and the network of institutional members we are building is really powerful.'

During its founding, Dryad required all datasets to be associated with a published paper. Over time, the requirement to have datasets associated with a paper was dropped. 'Dryad was founded as part of societies and editors from journals from societies, with this idea of complementing the journal publishing world. We haven't lost that as a core activity, but Dryad is evolving as the research community changes its understanding of how to support data as a first-class research output. Having the journal article as this anchor to every dataset was getting in the way of us promoting the data itself, and that also was getting in the way of thinking about the research institution, and what the research institution was looking for.'

Consequences of current funding model

'As a non-profit, we focus on covering our costs; that's all we need to do, and that's how we do our pricing, which consists of individual fees for authors (currently almost 50% of Dryad's income) and membership fees. We focus on our core operations and have a lean team that works on maintenance and meeting our commitments around curation. What is also important is looking for ways to ensure we have funds for broader innovations.'

Dryad see grants as key funding sources for innovation. Chodacki says: 'People sometimes look at projects that receive grant funds as being less secure. One of the messages that I like to bring out when I talk about Dryad is that it is a sustainable model; this may be the most sustainable model there is. There is no shame in getting grants. They have been part of Dryad's past and will be part of Dryad's future. Dryad is innovative; it needs innovation money. Maybe in the commercial world, companies would prefer to look for funding from investors (with all those strings attached) in order to focus on innovation. In our world, we focus on partnering with like-minded organizations and often that means grant projects. Our researchers know this works because their communities often run off grants, too. For us to step away from grants would be silly. It would be against the way our community invests in the future.'

Teal adds that 'another view is to see grants as a form of doing business, but on the innovation side. You can be sustainable for your core operations, based on the revenue that you bring in, and grants can be an opportunity to try something new. Being sustainable without grants is something we definitely all strive for and we should all think of as a success metric. However, being involved in grants is also a success, because those are awarded to researchers by governments, institutions and foundations. In many ways, it's an endorsement of the project itself. As Executive Director, I would be failing if I wasn't trying to go after those grants because the field is changing, and grants are really a way of connecting with the community around what's needed; I think of them as an important way to stay connected to the community.'

'When it comes to publishing, and data publishing and data stewardship, we're talking about the long haul. When we tell stories of the beginnings of these projects, saying, "Oh, had a grant here, and we had in-kind contributions from there," very often we say that with a hint of shame, or we talk about how it is something we had to overcome to get to sustainability. We act like diverse funding streams should end at some point, and that you're only successful when you stop receiving external funding. Grants are another shaming metric in the world of sustainability within our larger scholarly communications community.'

Commercial companies have sales. Why are we shaming them for their methods?'

Sustainability is not just about being financially sustainable, but also about being around in the long term. 'Especially in Silicon Valley, the long term can be seen as a one-to-three-year horizon; that's the dot.com, start-up mentality. More and more people look to that when they're trying to deal with all the uncertainty in publishing, with big giants like Elsevier or Springer Nature. They say, "Everything is changing, so we can really only plan for three to four years." But in reality, sustainability for projects that include the preservation of research outputs needs to plan using a 20-50-year horizon. For data publishing and stewardship, this requires us to look at experts in the field of long-term stewardship, which are libraries and institutions. They are designed to be memory institutions. I would say that the sustainability of any organisation in the research data space that isn't planning longer term at a very structural level should be questioned for not meeting the needs of the community.'

Another element of sustainability is how Dryad's governance structure is organised. 'Sustainability is also dependent on the governance model of the organisation, and the ability to have succession plans. Dryad has had different leadership, but the governance structure is set in place. It really ties the

priorities of the organisation to the community, and to these institutions.'

Hence, Teal and Chodacki urge others not to underestimate the importance of in-kind support from the host organisations. 'When people, especially in academia, look at vendors or solutions, they tend to look at those that are for-profit as being "sustainable," or they have some sort of understanding that they are more reliable. But one of the beauties of Dryad is that it is a great example of a collaborative relationship between a fully autonomous non-profit organisation and an academic institution, which actually leads to more reliability and sustainability than corporate initiatives, because universities are here for the long run but start-ups come and go.'

Teal cautions to be aware of the administrative consequences and overhead of the type of structure chosen: 'In terms of the organisational structure of the organisation: as a 501(c)(3), we do have administrative overhead. I want to emphasise that when you're setting up as an organisation, you need to think about the administrative structure and how you are going to manage that. Always plan ahead and envision what you want to look like, what needs to get done and by whom. It is running a business!'

Future vision for sustainability

'Sustainability also means having a plan for the long term and being mission-driven. Strategically, we are always evolving, from a "where we want to go in the world" perspective. Dryad's goal is to always focus on researcher needs. What we're really thinking about right now is shifting from this mindset of "store your data somewhere" to "data re-use,". We have exemplars in the community who are taking Dryad data and creating new knowledge from what we offer in the repository. We want authors to get better at sharing data, sharing metadata, and have better structured data, because they can see the impact of re-use. We are mission-oriented; we want better data in the world, making a positive impact on science and society. Right now, with COVID we see how valuable data is in its own right and how important data sharing is to quickly advance progress. To sum up, strategically, we want to make a cultural shift with data re-use.'

When we curate a dataset, this means actually having a dialogue with the author to help them improve the dataset. We are always educating ourselves and the community. That's a big kind of impact that we have: those relationships with the authors and trying to build a more data-aware research community.'

In terms of financial sustainability, Dryad is, however, looking to move away from individual fees for authors for data publishing and more towards an institutional and publisher membership model. 'Dryad does curation, so we review every single dataset that comes into the repository. That has a cost associated with it so we have a data-publishing charge. But we want to move away from individuals paying that fee to more of a membership model with institutions, publishers, and funders covering the costs. Institutions are getting more invested in supporting the data publishing and data sharing for their institutions. Similarly, some publishers are interested in sponsoring the fees for their authors, as a part of journal promotion. Things are changing and all of them are good for the author.'

'So, we're trying to shift more towards the model of membership or sponsorship as we move forward. The fee has been there since the beginning; it was part of the original modelling. It is never easy to get people to pay, of course, but with all developments, like Plan S, there is this larger eco-system question about who pays (author, institution, funder, library, etc.). We don't know how that will end up, so we're not taking sides. We recently actually had our first funder who became a member. So, we're working at covering the costs of curation with a diverse set of organizations: anyone that has a stake in the process.'

What is Infrastructure?

Teal and Chodacki would like to see more funding for the ‘plumbing’ of data repositories and scholarly communications in general. ‘I don’t believe that things like Dryad are necessarily infrastructure. I think they’re important, but they’re not the plumbing. We all rely on metadata, identifiers, the networking of the internet, and that core stuff is at risk. It is important for that to be funded. The concept of infrastructure gets passed around a lot. I think of it as “what plumbing do we all rely on”? What pieces are crucial, what other stakeholders do you rely on? How can we, as projects, contribute to those to make sure that they continue? If they fail, and you rely on them, you’d have to do them yourself.’

‘If I were to wave my magic wand, I would want IOI and SCOSS to be talking with Crossref, DataCite, ORCID, and ROR, these kinds of identifier systems. How do we get them all into shareable metadata stores, and how can we actually change scholarly communications through truly open and sustainable infrastructure? Imagine all the new projects that could start or existing projects that would thrive if that was facilitated.’

Teal and Chodacki are interested in providing more guidance for emerging projects and encouraging collaboration. ‘Investing in shared

infrastructure should be about investing in the people who run it. For instance, why is each leader of each of these projects trying to figure stuff out on their own? Is there a way to give them a guide or process to develop a custom business model? All these projects don’t have access to the same levels of expertise that a start-up would. We need to support open projects so there are well-developed open alternatives to commercial interests. That is a place where putting some investment could really help the ecosystem as a whole.’

‘Also, around collaborations: How can we all be complementary and help each other? IOI/SCOSS could (even informally) facilitate that. Without forcing anything, let’s see who has shared values, who is thinking about things in a similar way. Let them build alliances that will last and be able to drive adoption.’

Advice for peers

One piece of advice Teal and Chodacki offer new projects is to not go too big in the beginning. ‘There should be an initial in-between phase. With Dryad, that lasted two to three years. There was an initial grant and during that time Dryad was figuring out how to build something on its own. Regardless of whether it’s grant funded or not, there is value in that solidifying phase. There was a lot of

learning there. One thing that is important for sustainability and for any organisation is: don’t go too big. It’s better to be smaller and strapped than it is to be bigger. And don’t be impressed by people who do go big. It’s not about quantity; it’s about quality. Having six great employees is a much better situation to be in. That journey, of figuring out what the right size is for you, is really important. Constraints can be good. They really do make you think about that and be innovative.’

Chodacki furthermore advises, ‘Pick your partners based on your values and your mission. Very often you will hear about fancy business opportunities or sales opportunities, but sustainable alliances are about something more emotional and mission-driven than they are financial. So, first you have to find the right partners and, second, you have to make sure you keep those alliances strong.’

Teal adds, ‘Understand how the budget breaks down between core and non-core functions. That’s guidance I would offer to anybody who is starting or who is trying to evaluate their sustainability model; consider what unique value you actually offer. It is very foundational when you talk about grant-funding, about right-sizing, about your business model, about everything; what is it that you’re actually doing, what is the core thing that you’re offering?’

Teal's advice returns to the core idea of long-term sustainability: 'Steer the organization based on your mission. Also make sure to plan as if you personally won't be there in the future. That's good sustainability planning. Ask, "If I'm not there, will my organization be able to carry on? Will they understand my decisions?" Staying true to the organization's mission and values will always help others understand the context and direction you were headed.'