



# 10 Key Interviews

INSIGHTS INTO THE SUSTAINABILITY OF  
OPEN INFRASTRUCTURE SERVICES

**01**

An Invest in Open Infrastructure project  
*More info: [www.sparceurope.org/ioiinterviews](http://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](http://www.sparceurope.org/ioiinterviews)*



# WHO WAS INTERVIEWED?

## Overview

**4TU.Research.Data**

*Featured in this document.*

Code Ocean

Dryad

Mendeley

F1000 Research

Figshare

Our Research

arXiv

Redalyc

EDP Sciences



Interview: 4TU.ResearchData

**"There are different layers to sustainability, with the community's trust and strategic, financial and technical networks being important"**

Marta Teperek (Head, 4TU.ResearchData)

Alastair Dunning (Head of Research Services,  
TU Delft Library), the Netherlands

An Invest in Open Infrastructure project  
More info: [www.sparceurope.org/ioiinterviews](http://www.sparceurope.org/ioiinterviews)

**04**

# 4TU. RESEARCH DATA

Marta Teperek & Alastair Dunning



## At a Glance

### Type of activities:

Data repository, with related services and training for all researchers in science, engineering and design

### Life-cycle stage:

Founded in 2010, with an established business model and governance based on key members of technical universities in the Netherlands. Now planning different tiers of membership for other potential members

### Current legal structure and funding model:

Consortium of three Dutch universities (no separate legal entity), with staff employed by the universities and the universities contributing to the (direct) cost via membership fees, allowing the services to be offered for free for researchers

### Technology:

In the process of moving a significant part of their technical infrastructure to Figshare; more specialist services will continue to be developed in-house

## Sustainability

*'Sustainability is keeping the service running. You could look at different layers of that. Keeping the technical service running, having a communication service on top of that, and turning it into a product, all while preserving the actual content and the data with identifiers. Sustaining the community's trust is also an important element.'*

## Piece of Advice

*'Be as honest and transparent as possible about your decisions and considerations. Involve your stakeholders; take them along when you make decisions.'*

*'Sustainability requires the community's trust as well as solid strategic, financial and technical networks. For small things this may not be needed, but for anything significant to have an impact, you need that.'*

# 4TU.RESEARCHDATA

4TU.ResearchData is a research data repository run by three Dutch universities and it is not a separate legal entity. It is currently in the process of having its core task of running the repository supported by Figshare, while keeping options open for the future.

## Original vision

Head of 4TU.ResearchData, Marta Teperek, and Head of Research Services at Delft University of Technology (TU Delft), Alastair Dunning, explain that the initiative began 10 years ago following an overarching concern with the lack of places to store and publish researchers' data. 'Two impulses played a role in the repository's creation: first of all, there were a few large-scale data projects at TU Delft. Researchers had no place to store and publish their data, and these were large projects collecting serial datasets on weather conditions, land use, etc. Second, there was the "3TU Federatie" (an alliance of three technical universities in the Netherlands: Twente, Eindhoven and Delft). They were looking for problems/challenges to solve together. One of the challenges they identified was how to publish research data. Back in 2010, there was no Figshare, no Zenodo, and I don't think there was Dryad; there was no real way to do this; and there was nothing to fulfil the need. At the same time, 3TU was closely

involved in setting up DataCite and establishing a DOI infrastructure for research data as well.'

## Growth and sustainability challenges

In the last couple of years, 4TU.ResearchData has faced challenges related to costs and external factors, leading to the decision to move to a commercial infrastructure. 'Our main issue has been that in the past three to four years, the cost of maintaining and expanding a data repository and developing new functionality has increased, due to the increased focus on, and importance of, research data. It also put demands on our software developers and our business model. It was getting prohibitively expensive, and it was difficult to form a bigger consortium with more universities in a short period of time to pay for it all or, alternatively, to find more cost-effective ways to do the development. Our funding structure and the possibility of changing that quickly meant that we had to look for alternative solutions.'

That's why in the last two or three years we've undergone a change of technical infrastructure. We've started to talk about having additional members and are looking into other types of membership models. Most relevant here is that we changed from an almost entirely open source infrastructure to one that is significantly based on a proprietary solution from a commercial provider.'

The technical overhaul was needed in order to meet researchers' demands for statistics on their data in the repository and demands to provide restricted access to the data for reasons of confidentiality. 'When we started, there wasn't much; we had the archive to provide the functionality of publishing research data with DOIs, which was quite innovative at the time and something that researchers very much needed. But we found that there were growing demands as researchers wanted to get good statistics, easily visible for them, on who accessed their data, how frequently, how many times the dataset was cited, etc.'

# BE AS HONEST AND TRANSPARENT AS POSSIBLE ABOUT YOUR DECISIONS AND CONSIDERATIONS. INVOLVE YOUR STAKEHOLDERS; TAKE THEM ALONG WHEN YOU MAKE DECISIONS.

'Another big demand from our community was for us to provide some restricted access to research data. Publishers wanted to have reviewers access the datasets underlying the research covered in an article. Confidentiality is important so you don't reveal the names of people within the research and also that you maintain the same DOI, even though the data may still change or evolve. Additionally, we have a lot of commercial research and research involving personal data. Sometimes, removing personal or commercial data will make the data useless for any other research, as you need some kind of connection or identifier to, for instance, combine one dataset with the other. Therefore, you need some kind of functionality to enable restricted sharing of datasets.'

## **Opportunities, considerations and choices**

The demand for restricted access and statistics, among other things, is why, after much deliberation, 4TU.ResearchData decided to go with a commercial infrastructure. 'Those

demands from our users were growing and we were unable to continually meet them with the same ability in terms of resources and technical development capacity.'

In order to manage the costs, 'the time came to make a hard decision: How can we sustain the repository so it continues to be an attractive offering to our end users. Equally, how can we continue to innovate while not increasing our fees to the partner institutions? In the end, going for a commercial repository solution was a better solution because we don't have to maintain the pool of developers in-house. Developers tend to be quite expensive, especially the senior ones, the ones who can drive development forward. And sometimes, in academia, salary scales mean that it is quite difficult to attract and sustain that kind of potential, that kind of people.'

'By outsourcing the platform, first of all you have a pool of people who the commercial provider is able to hire and sustain, but also have access to experts in database development. Therefore, our decision to

outsource the platform, the technical solution running the repository, might be a better option for us. With the move to Figshare, it will be a lot cheaper to run the service than doing it all in-house.'

According to Dunning, 'there are a couple of other things to add to that story: we realised in 2014/2015 (before I joined) that being unable to continually meet the growing demands from our users was a problem. We actually also looked into whether we could outsource the development work to developers in India, and have it continue to be open source, but that didn't work out. There was also discussion on whether we could increase the size of the consortium, expanding nationally and internationally, particularly in working with other technical universities. If we want collaboratively-developed open source tools, we need big open source consortia, and these take time and effort to create, particularly at an international level when you're dealing with linguistic and cultural differences. Those are the big issues that need to be tackled. But that was just a minefield; doing that would have

taken so much time because we'd have to reach out to other universities, build a good business model, and build up the trust and a new governance model. We didn't have that time. So that's why we went down the route that Marta mentioned.'

Maintaining an open source model was thus not viable. 'When we started, it was the only way to go. While we still think it's important, but with time, we couldn't find a good way of making open source work. We weren't fully open source anyway... our code wasn't all documented or publicly available (e.g. on Github). Open source is important, but there are even higher values, in terms of publishing data and looking after researchers' needs that we had to meet.'

'We also looked at open source and had a conversation with the people behind Zenodo. They started a project aiming to deliver Invenio, an open source product which is also used to run Zenodo, as a 'turn-key research data repository'. However, that project was still too early in its development. We're interested in where that is going, and what might be feasible, but a year ago, when we had to make decisions, it wasn't at the stage where we could simply deploy it for our data repository. We also had conversations with Dryad and they were enthusiastic, but their service offer wasn't quite right for us, given our particular requirements, which didn't match with what

Dryad was offering. So that's another one we'll keep an eye on, and see where that goes.'

Hence, in order to maintain sustainability and users' trust in the service, outsourcing was the only solution. 'There was no viable alternative; we had to sustain the user community. In the beginning, we were offering researchers the publication platform that they very much needed. So we had users who wanted to continue working with us, but they missed elements and needed to meet publishers' requirements, and we needed to continue to be the trusted partner of the community. Sustaining the community's trust is also an important element.'

At the end, this is what sustainability is about, says Dunning. 'Sustainability is keeping the service running. You could look at different layers of that: keeping the technical service running, having a communication service on top of that, and turning it into a product: all while preserving the actual content and data with identifiers. But, if your task is running a service, you're trying to maintain that idea of sustaining that as much as sustaining the data and the content, and the service level as well.'

They also point to the importance of networks. 'Sustainability requires the community's trust as well as solid strategic, financial and technical networks. For small things this may not be needed, but for anything significant to

have an impact, you need that.'

Open source is still considered important. 'I want to get across the point that we haven't given up on open source. We still do and support various open source initiatives, e.g. run an instance of Dataverse.org. In this debate about doing stuff commercially or open science, there is perhaps a balance between the heavy lifting on core stuff done by commercial organisations, and then some of the lifting for the more specialist stuff that needs to be done by the universities. That's not clear-cut, and we also might want to join up at a data level to again do some of the core stuff ourselves. But when we had to make that choice, it was quite clear we had to go in this direction.'

### **Consequences of current funding model**

Currently, the 4TU consortium consists of four universities. Wageningen University & Research recently joined 4TU, but has not signed up to join the data repository yet. It is not a separate legal entity. For 4TU.ResearchData, governance and management oversight falls to the three head librarians of Twente, Eindhoven and Delft. Above them are the three university rectors. TU Delft hosts the service. 'It was agreed that TU Delft would take the lead because they had the biggest challenge in terms of research data publication. The data repository was



established in TU Delft. That means that when a user uploads data, they have to agree to Terms & Conditions for depositing data. This is a contract between two parties: the researcher and TU Delft (since it has to be a legal entity). It also means that the budget is held by TU Delft.'

A consortium agreement between the original three partners covers responsibilities at a broad level, deliberately not going into details. 'We just re-signed this last year. What is covered in the consortium agreement is all very high-level, deliberately, to allow freedom for the management. It doesn't go into details on what we do; it describes (high-level) how we work together and what the governance is. It does mention the payments and how the structure of money works, but it doesn't go into detail on what we are doing next year. It also covers the exit strategy, and what happens if something goes wrong. A lot of the responsibility then falls back to TU Delft.'

The work is organised in a front and a back office. 'The back office is basically the people that run the data repository and related services, and they are all employed at TU Delft. The front office consists of "ambassadors" at each university for the use of the archive, and they are paid by their respective universities under a separate budget. For instance, there are a couple of people in Eindhoven whose job

it is to get the message out for use of the data archive at Eindhoven University. People in the front office help inform people about 4TU.ResearchData, help to upload data, explain about the licenses and so on. They also do some other tasks related to research data support and thus split their time between 4TU.ResearchData and other tasks.'

Consortium members pay a membership fee that includes not only the data repository service but also training. 'We don't just provide the data repository as a service, we also provide training on research data. We recently joined the Carpentries organisation in America, so we do a lot of training on software carpentries for our universities. Another training is "Essentials for Data Support": an introduction to research data management for research data supporters. In addition to running the repository, we also do a review service when the data is uploaded: a review of the metadata that describes the dataset.'

'Hence, we offer value-added things, so we're more than just a technical service. The membership fee doesn't just pay for the access to the data repository; it pays for these things as well. The three members (Delft, Twente, Eindhoven) pay a fixed membership based on their relative size. Part of that pays the Figshare contract and the services provided by the team - researchers and

students get the services and support for free.'

The outsourcing to Figshare was done through a public tender. 'As mentioned, the fit with Zenodo and Dryad was not quite right, so we put out a public tender, published our requirements and, after a marking process, Figshare was the winner. We are now going through the process of moving our technical infrastructure from our in-house code to using Figshare's platform. And it very much is a platform, so we remain a repository with our own functionality and our own processes and procedures. All the data is still stored on servers at TU Delft.'

'During the tender process, we really wanted to avoid lock-in, so there are a few things that we did, such as making sure Figshare shared an exit strategy with us; Figshare commits to certain actions if we want out; that's part of the contract. We also ensured all metadata is issued under a CC0 licence, so we can always access it, download it and use it forever. None of the data will ever be owned by Figshare; it's all owned by the universities or the researchers. If Figshare ever goes bust, or they can't run their service anymore, their code will automatically become open source (they now have it in escrow) and then we have various things rules concerning formats and open standards and exports. There is always some sort of lock-in with a commercial partner, but

there are things you can do to protect yourself, and I think we've done quite well. This morning, I uploaded our call for tender; I was now allowed to make that public.'

However, the move from open source to a commercial model was not popular among everyone in the academic community. 'There is a strong belief within TU Delft, and also within 4TU and our user community, in the value of open source. Some colleagues in our university, the federation, and some of our end users wonder, "Why are you now going to Figshare?" Especially as there are discussions about publishers gaining a stronger hold on the market, and the data market, they wonder: "Why are you doing that? Why are you doing that now?" That is a negative side effect in the short term. We have to explain ourselves as to why we've made that decision.'

'Also, depending on the commercial provider, with a proprietary tool, it is not easy to co-develop certain solutions or features that we, and our community, are interested in. One recent example that may be a hurdle for us: Figshare has no integration with our TU Delft version of Gitlab, the non-commercial solution for version control for software management, and our researchers use that and would want that integration. The question now is: where and when will that be on the roadmap of a commercial provider? Or, can we do it through their open APIs? These kind of considerations

are now on the table.'

'Moving to Figshare frees up time and money, as we don't have to do the core task of running the repository. And now we can concentrate a bit more on what makes us unique, and leave the core stuff to commercial organisations. We can concentrate on the discipline-specific functionalities we offer, such as dedicated support for researchers working with netCDF datasets, the manual curation of all incoming datasets, or discipline-specific metadata enrichment, and it also frees up time to focus on the training and other aspects.'

### **Future vision for sustainability**

In the future, Teperek and Dunning envision a need for more development resources. 'We will need to rethink our cost/membership model. We increasingly receive requests from research communities to do projects or build upon the existing services that the repository provides. For example, one of the more special things we do is that we have a server which is called "OPeNDAP," which supports the sharing and publishing of data(sets) in NetCDF, a specific format. This allows other researchers to query, analyse and directly work on data without having to download everything. Some researcher communities, also outside the partner universities, became increasingly interested in working with us to

develop some additional services based on that (e.g. a tool to transfer other dataset types into NetCDF in order to benefit from the option to directly work on the data in the repository).'

These things offer a great opportunity to better understand what these communities want and need, but also to build a relationship with these communities and disciplines. This is something we need to think about in the long term because that obviously means we need to have some development resources to support such requests coming in. We're currently debating whether we should have a fee or should their institution become a member since we need to make sure we're sustainable long term and, if we add these more expensive developer costs, we need to find a way to cover them.'

Becoming open source is still part of 4TU.ResearchData's vision. 'We think about the importance of open source in the long term. We hope to engage in some partnerships in the future to develop sustainable alternatives to commercial providers for our repositories. We are interested in such possibilities, and in exploring them.'

'At TU Delft, we have this strategic programme for open science, with principles across all sorts of initiatives and programmes we run. We still want to participate in open source

tools, ideally not by ourselves, but rather as part of a collaboration and network. That could be a combination of universities and commercial organisations. And we also hope to make it possible to make the repository community-driven and have our own researchers contribute to its development.'

It remains to be seen whether this ideal is attainable, however. 'We now have a difficult strategic dilemma, because the more successful we are in setting up this Figshare partnership and making it work, the more difficult it will be to leave in the long run.'

### **Advice for peers**

'Research data is only one part of scholarly communications, and new things happen all the time. So, there is a broader component, not just about research data, but also in the continuous expansion of the research infrastructure ecosystem. Having people contribute is very valuable.'

'What would also be welcomed, especially across other types of research outputs and infrastructure, is clarity and transparency about the cost of running open source infrastructures and the cost of supporting commercial infrastructures. Ideally, projects share with each other what the cost (components) are of running open source

infrastructure, and also, in case of going through a (commercial) provider, what the (licence and other) cost are. Our community desperately needs clarity and transparency to be able to compare such cost and make informed decisions.'

'Finally, Dunning and Teperek urge their peers to communicate in an honest and transparent way. 'Be as honest and transparent as possible about your decisions and considerations and deliberations. Involve your stakeholders; take them along when you make decisions. End users and board members: Communicate!'