10 Key Interviews

INSIGHTS INTO THE SUSTAINABILITY OF OPEN INFRASTRUCTURE SERVICES

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.

WHO WAS INTERVIEWED?

Overview

Code Ocean *Featured in this document.*
Dryad
Figshare
EDP Sciences
F1000 Research
Mendeley
Our Research
arXiv
Redalyc
4TU.Research.Data

"Sustainability means a competitive market, an ecosystem of both not-for-profit and for-profit players"

Simon Adar, Founder, Israel/US

CODE OCEAN

Simon Adar, Founder

At a Glance

Type of activities:
Cloud-based computational reproducibility platform, research collaboration platform, professional tools for computational researchers

Life-cycle stage:
Founded in 2015, now a small-medium enterprise, sustainable in a financial sense and customer retention (charging customers for the services provided), growth/scaling phase

Current legal structure and funding model:
Equity in hands of professional venture capital investors, (co-)founders and minority investors

Technology:
Platform is proprietary; all content on the platform is open source, exportable and interoperable

Sustainability

'Sustainability is first and foremost, not having to rely on external money for the continuation of the company. This means providing enough value, so the community is willing and wanting to pay for the service. The most important thing is not whether you are for-profit or not-for-profit; but that you follow the FAIR principles and allow artefacts created on any platform, regardless of what kind of scholarly platform it is, to be exported outside of where they were created — basically the "I" in the FAIR principles: interoperable.'

Piece of Advice

'When thinking about sustainability, we should strive for a competitive ecosystem of players that provide the best value and products to the community.'

'What we do impacts scientific discovery, you have to be very passionate about what you're doing. Otherwise it is going to be very difficult. The passion, the mission and the vision are going to be very important for success.'

CODE OCEAN

Code Ocean is a for-profit, cloud-based computational reproducibility and research collaboration platform that also offers professional tools for researchers. Simon Adar founded the platform in 2015 when he was a postdoctoral researcher, having also worked in industry.

Original vision

When Simon Adar started Code Ocean, he had experienced the same problem in industry and academia, building, collaborating and reusing scientific code. When starting a new project or building upon an existing code project you wasted precious time dealing with infrastructure and environment setup, configuration, data management and other DevOps tasks. There was no platform specifically designed for computational science to use that had all the tools a researcher would want to use, without having deep IT expertise to write code to link them together. Such a tool would make scientists more productive, their research reproducible while enabling them to share their work with others. Adar wanted to focus on his research and not with IT infrastructure and cloud-computing set-up, security groups, and all of that. Adar's vision was to speed up the pace of discovery and increase researcher productivity.

'In addition, harnessing the power of the community, allowing for continuous improvement through seamless collaboration, and sharing, would multiply the impact of research. When you're creating your research, it's primarily meant to change the world and advance science. We need to do more than just publish articles or the narrative of science. We must include the different artefacts in a way that facilitates reuse. This is what will lead to faster discoveries and more impactful science.

'So, from a vision point of view, Code Ocean is a productivity tool designed for computational science, providing the tools (any open source programming language: Python, R, C++, etc, and proprietary languages like MATLAB along with a package and environment editor) in one integrated platform. It also curates and preserves your work in an executable and reproducible way so others can immediately start building upon it with confidence. You don't need to waste time trying to get someone else's code to run.

Growth and sustainability challenges

Adar chose to raise money from professional investors and venture capital firms such as Battery Ventures, Digitalis Ventures and a strategic investor such as EBSCO. The other approach is to get non-dilutive money, like grants, which he chose not to do. Choosing the right investors that understand and share your vision is important as you will need to work with them (and your board), overcoming the many challenges that will come up along the way.' Persistence is key, and despite all the ups and downs, you can take comfort in the fact that your vision materializes and you are able to provide real value to your users and customers.'

'EBSCO, is also our reseller partner with a significant footprint selling in the academic market, which allows us to get into almost every university in the world. With their

SUSTAINABILITY IS FIRST AND FOREMOST NOT HAVING TO RELY ON EXTERNAL MONEY FOR THE CONTINUATION OF THE COMPANY

capabilities in the academic market, we can focus on building the technology and direct sales into the corporate market.'

Opportunities, considerations and choices

Adar consciously chose to incorporate as a for-profit company because he did not want to be dependent on grant funding and other short-term financing. 'As an academic, as a postdoc, I was exploring different ways to fund Code Ocean. There were opportunities to fund this through organizations like the NSF (National Science Foundation), the NIH (National Institutes of Health), and with cyber-infrastructure grants. I decided, however, not to go this route because I'd seen so many academic projects get funding for the first three to four years and then the funder refocuses and those projects run into serious trouble and sometimes die. It's a shame. I wanted to create something which is sustainable, and I thought that the only way to do that is to fund it with a business model that will be sustainable for generations to come and will not just fade away after a few years when the funding dries up.'

Adar created a C corporation in the US (the most prevalent legal structure for a corporation) whose structure remains the same today. He went through Cornell Tech's Runway Startup Postdoc Program at the Jacobs Technion-Cornell Institute. 'The Jacobs Institute, which is a partnership between the Technion and Cornell University, was the first investor in the company and hold minority equity in the company.'

'It really allows you to start small, raise money, and grow the company in a traditional way, so to speak, with initial seed funding, then angel investors and venture capital money later on.'

This structure enabled him to hire the best professionals at competitive market rates. 'This structure allows you to scale and hire the best people in their roles.

'A highly competitive market with different players produces the best products. As a researcher myself, I would be worried if we're exclusively working with tools that are only built by not-for-profits. I would like to use the best platform out there and to see competition between difference service providers such as Google Cloud, AWS and Azure; this will drive costs down. I would like to be able to move between the different clouds if I'm not happy. As a researcher, I don't want to be left with only not-for-profit cloud computing. I don't want to be limited. I am a firm believer that competing commercial tools (for example, GitHub and Bitbucket) drive innovation and productivity. It's my philosophy, choosing the best tools possible for my research.'

We see Code Ocean as a platform that serves all publishers and all researchers in all fields of science. We follow the FAIR principles (Findable, Accessible, Interoperable, Reusable). All of the projects that are available on the platform, can be exported out. When thinking about sustainability, the most important thing is not whether you are for-profit or not-for-profit; but that you follow the FAIR principles and allow artefacts created on any platform, regardless of what kind of scholarly platform it is, to be exported outside of where they were created — basically the "I" in the FAIR principles: interoperable.'

Adar explains that Code Ocean resembles Github. 'Github as a platform is not open source, but the content on it is open. Code Ocean is the same: everything is out in the open, all research projects are Open Access. In addition, authors assign the licence for their code and the data. Code Ocean doesn't own the rights of the artefacts. We recommend using permissive licences like MIT, CC-0, CC-BY but provide many different licensing options. An author can even upload their own custom licence if they wish.'

Consequences of current funding model

The for-profit model meant that Code Ocean did not have to rely on grants for the continuation of the service. Instead, the idea was to provide enough value, so the community is willing and wanting to pay for the service. This also meant that Adar could monetize the service to corporations, which in turn enables him to offer services at a lower rate to the academic community. This allows us now and in the future to provide services, either at a lower rate or for free for the academic community.'

'For example, publishing in Code Ocean is free today, and we're providing free compute and verification for academics. The business model distinguishes between what is public, and hence free, and what is private, and thus paid. Our business model follows very much the Github model, where the things that are public are free since you're contributing to the community. For private use of the tool you can choose one of the paid plans.'

Adar sees his current business/funding model in a positive light. 'The positives are of course that we have a sustainable model. So many have tried to do what we're doing using grants and have lost important research, because it wasn't preserved.'

The only negative he can think of is that being a for-profit is not always seen in a positive light in the academic community. 'The academic community will often perceive for-profit companies as being overly aggressive or wanting to monetize academic IP.' In response to such criticisms, Adar emphasises Code Ocean's open character. 'We're a company that is extremely mission-driven, principled, and we've done a lot of work to make sure that there is no walled garden in Code Ocean. Anyone can download code outside of Code Ocean, and even provide researcher instructions to reproduce the analysis locally on their own machines. In addition, we've made an agreement with CLOCKKS and LOCKKS for preservation.'

Furthermore, Adar employs four strategies to mitigate any negative perceptions. 'First, give a lot to the community, in terms of resources such as compute and storage; second, put together a responsible and transparent preservation plan; third, make sure that people understand that this is not a walled garden and that anyone can, export the contents outside of Code Ocean; and fourth, make it clear that we don't own the content but that the authors (the academics) do.'

Future vision for sustainability

Adar envisions a future where research projects will be fully interoperable and reproducible, where Code Ocean continues to serve the academic community as a private company, but remains and grows in the corporate sphere. 'I see the different verticals feeding each other, with a network effect of researchers sharing reproducible code, interoperable code, and using reproducible code and data privately. On Code Ocean, Researchers are leveraging our reproducibility capabilities for their private use. You start projects that are reproducible from day one, and you are able to share this within your lab, or within your company, so you improve productivity, collaboration and get research results faster. You are more confident of your results because they are reproducible, properly curated and preserved leading to a higher quality research product.'

‘From a vision perspective, five years from now, I see significant acceleration in the pace of scientific research both in academia as well as in the corporate world. I see instant reuse, collaboration and scientists across the globe building upon each others’ work.

Advice for peers

‘If you’re taking on the challenge of building something, you have to be very passionate about what you’re doing,’ Adar emphasizes. ‘Otherwise, it is going to be very difficult. The passion, the mission, and the vision are going to be very important for success.’