

# Open Science Talk No. 56 (2024): IOI and Infra Finder : a computer-generated transcript <sup>1</sup>

00:00:10 Per Pippin Aspaas

Open Science Talks, the podcast about open science. My name is Per Pippin Aspaas, and today I'm joined by Emmy Tsang from IOI, or Invest in Open Infrastructure. So welcome to the podcast, Emmy!

00:00:31 Emmy Tsang

Thank you very much for having me!

00:00:34 PPA

So what is this IOI – Invest in Open Infrastructure?

00:00:42 ET

Yeah. So Invest in Open Infrastructure, which is quite a long name – that's why we call it IOI – it's a non-profit initiative. Our mission is to increase the investment in and the adoption of open infrastructure to further equitable access to and participation research.

00:01:01 PPA

So this is your mission. And and what is your particular role there? Are you many people involved in IOI, and who are you in this in this picture?

00:01:12 ET

Yeah. So I am the engagement lead at Invest in Open Infrastructure – I lead the engagement team, which is 3 people at the moment, and the whole team at IOI is, I want to say, eight people full time at the moment, although we're growing quite quickly. We are an international team. Some of my colleagues are based in the US, some in Europe, including myself. And then we also have a colleague based in Kenya, in Nairobi. So this enables us to have a relatively global view on the issue and topics and developments in the open infrastructure space. And we're always looking to engage more players and make sure that we are taking into account and listening to and learning from the perspectives from regions all around the world.

00:02:06 PPA

Yeah. What I could learn from your web page is that you specialise in, like, developing web tools and services to help people identify infrastructure that is tailor made for open science. And do you also hand out money yourself? I mean, seed money or stipends or such? Could you tell us about that?

00:02:29 ET

Yeah, absolutely. So, yes, as you correctly pointed out, a big part of our work is to do research, but not just publish papers and reports. That is an important part of what we do – we do believe that, you know, in order to make sure that the investments into open infrastructures are made in an informed, coordinated and strategic manner you need that information and that evidence and data. So we do have a research team who specialises in, and has deep and broad expertise in, the open

---

<sup>1</sup> This is a computer-generated transcript of the podcast episode Open Science Talk No. 56 (2024): <https://doi.org/10.7557/19.7551>. The automated transcript has been proofread by Per Pippin Aspaas and is included here for the sake of Universal Design and improved discoverability by full-text search engines.

infrastructure space very broadly, and do deep investigations and, you know, do publish some reports. But more importantly, I think, we are always thinking about how to make sure that our work is actionable. And so people who are reading and interacting with that research and data are not only learning more about the landscape and developing that understanding for themselves, but also thinking about: OK, how do I use this information to now change how I make decisions or how I'm discussing with my colleagues? You know: how do we as an institution or a library, for example, use this information to help, you know, understand the implications of our decisions and the different considerations and the pros and cons and everything else? And so that is also something that we are focusing a lot on and also learning how we can make this research useful for our stakeholders. And that includes, you know, institutions but also consortia and other networks, funders in the space, and the open infrastructures themselves amongst other people. Your question to whether or not we hand out seed money or other things – we do, but in a very intentional way. So one of our other sort of core programmes of work, if you will, apart from the research and the actionable tools and recommendations, is what we're calling funding pilots. So this is, again, really taking that evidence and research that we've been doing and putting it into experiments into practise to see how we can capitalise additional investment into the space. And we have with every single part of money that we try to gather, we try to hand out – we have very specific hypothesis as to what we want to test in terms of, you know, it's not just about handing out money. That is important – it's, you know, we all know that, I think, the open infrastructure space have some very deep and entrenched issues when it comes to sustainable funding, which I can go forever into – but in order to find solutions collectively as a community, we need to think outside of the box and we need to test outside of that box. We need to think about how – you know, for example, we can have more people at the funding decision making table. We need to think about how to decrease our reliance on grant funding that are usually two to three years and you know, infrastructures are left hunting for grant after grant and and always having, sort of, that fear and that worry of not being able to sustain beyond that grand period, right, in order to make fundamental changes to this sort of deep issues that are well recognised in this space, we also need to design experiments, intentional experiments to understand, for example, how do we engage the people, the players outside of the usual suspects circle into funding open infrastructure, for example. So some of our funding pilots are designed with that intention in mind. Specifically, I can speak to the IOI Fund for Network Adoption, which we have launched in 2022 and are now currently actively developing. And that really is designed with two very specific hypotheses in mind. One is to see if we can engage, for example, commercial players in this space to incentivize them to contribute and invest back into community driven infrastructure and then the other one being – the other design hypothesis here being – trying to see how we can incentivize and support the adoption of open infrastructure at a network or consortial level. So thinking about that sort of economy of scale, as we look at the adoption of open infrastructure, how do we work closer together with these networks and consortia who already are well equipped and has wealth experience in offering, sort of, consortial level services that enables members to more easily adopt infrastructure and also thinking together about how to develop additional services that would would address the members' needs.

**00:07:34 PPA**

So how, in turn, are you financed – I mean, eight full time positions by now and doing some research, handing out some seed money. But how is this non-profit IOI funded, currently?

**00:07:51 ET**

Yeah. So we are supported through a mix of philanthropic support and financial commitments from institutions, consortia based organisations and others. So that means practically, you know, a mix of

grants, but also other service revenues. We have – I didn't mention this before – we also have additional ongoing partnerships within our strategic support core programmes with a few open infrastructure service providers. So you will be familiar with them – arXiv, for example, the preprint service – and, similarly, bio- and medRxiv, as well as 2i2c, who is a cloud infrastructure service provider in this space. And we're working with these initiatives specifically to look at how we can help them co-create and craft a strategic blueprint to improve their governance and sustainability. And that work is in large part – in a lot of ways – supported by the funders of these open infrastructure services, and that's also a revenue stream for us. In terms of operationally at the moment, we are supported by a 3.47 million U.S. dollar grant from the Arcadia Fund that's going to end this year. So we are also looking for additional support and also diversifying that support. So putting some of the – again, the sort of theory around financial sustainability that we have been building and learning about into practise ourselves to ensure that we do have that financial sustainability moving forward.

**00:09:34 PPA**

And one thing you would always want to do in order to acquire more funding, of course, is to showcase what you have done by now – that the money that you already have had has been spent on good things. And one thing is actually launched this month, April 2024, the Infra Finder. So could you briefly explain to our listeners – what is the infra Finder?

**00:10:04 ET**

Absolutely! We're very excited to be launching Infra Finder and I would also acknowledge upfront that this – a large part of this next phase of development work on this tool – has been supported by the Mellon Foundation in the US so and also our sustaining circle, members that have been supporting our work for the last couple of years in an operational manner, so Infra Finder is part of what I was describing before – in terms of a tool to turn the research and evidence that we've been building into actionable guidance and recommendations. So it is a tool that is aimed at making the process of discovering and evaluating open infrastructure solutions easier. At launch in April, it will contain information about 57 open infrastructure solutions and standards in the research and scholarship space. We have worked very closely with the service providers in order to gather this information to ensure that it is up to date and validated. And we also – one of the things that we have heard a lot in developing Infra Finder from community of – our target users, who are basically librarians, for example, who may have a set of technologies in their minds – so, let's say, if these are folks who are choosing between different data repositories to implement at their institution, the need for a single source of reliable information to help them make that decision and to communicate that decision making, and that pitch, almost, to their colleagues and their leadership, for example. So we understand – in building Infra Finder we are trying to really build the design and build the process to make sure that the information that we are displaying within the tool is up to date and validated, that it is all in a single place and one of the other things that we are testing out and building out and will be available at launch is a comparison view. And so you can compare up to four solutions side by side. And that will be a table with, you know, again, all the aspects of different information that we're collecting within Infra Finder in a very easily viewable and comparable format. And in terms of the information that we are collecting – this includes information on the open infrastructure services, technical attributes, for example, whether or not they have an open code repository or an open API, but also on their community engagement mechanisms, their governance, their key policies, etcetera, etcetera. And that list of information, again, came from working very closely in focus groups and in user testing with librarians and other target users in the space where we really work with them to understand what kind of information they need in order to make these decisions. Where are they finding challenges in trying to understand or compare some of this information, what aspects are

important. And that's how we ended up with these first lists of information that we're pulling together.

**00:14:01 PPA**

So infra Finder actually stands for Infrastructure Finder. Is that correct?

**00:14:08 ET**

Yeah, that's the full form of it, if you will. But we do use Infra Finder because it's – it's just easier to pronounce and hopefully, folks will be able to remember it easier, yes.

**00:14:20 PPA**

Sure. And, like, you mentioned the needs of librarians. I can really relate to that, working in a university library myself. So one personal concern that I have is that it seems that many idealistic, co-created, scholar-led infrastructure initiatives actually manage quite well in developing tailor made open source software for research. However, they they don't – they generally don't have marketing specialists, they don't have a sales department, no PR strategy – they are researchers, that's it. So by contrast, you have big commercial players that manage all that very, very professionally, of course. So is there a risk that you through Infra Finder end up promoting the big commercial ones and the other ones will not be as visible, or is that precisely what you want? To avoid that you can, sort of, have commercial solutions for open science side by side with with non-commercial ones. Did I pronounce my question understandably?

**00:15:34 ET**

Yeah, I think so. There are a couple of aspects within that question that I think is really, really interesting. First of all, you know, there's always – different institutions have different needs and priorities. That's something that we knew very well heading into the development process, but also we were able to verify, as we were doing the focus-group and end-user interviews. And so, you know, there are well-resourced institutions who are able to say: OK, you know, openness and alignment to our institutional value in terms of, let's say, diversity, equity and inclusion or Open Access etcetera are the most important things when it comes to choosing solutions for us, right? And so in looking at, for example, some of the decision-making rubrics – for example, the University of California Consortium of Libraries, you know, has a very extensive rubric on how they're assessing different solutions based on their financial sustainability, governance, commitment to open, etcetera. And, again, for well-resourced institutions being, sort of, on the more idealistic side if you will, it is not an issue. But for, let's say, a little bit less well-resourced institutions – or, I would say, the majority of institutions that we've been speaking with – usually this is not a liberty that folks have. Folks are most concerned about the costs of maintenance. For example: how much is it going to cost me to maintain this instance, this open source instance of this infrastructure in the long run? Do I have the capacity to do that? And then, also, the other question that we've heard a lot of is on interoperability: does this solution fit and work well with the other solutions and infrastructures that I currently have in place? And so these are very pragmatic concerns and they're not wrong, right? Sometimes, it will mean that a commercial provider, regardless of their marketing ability, will be able to provide a solution that actually works better for those folks. And so, what we're seeing is that – I think, something that's very elegant and we're very interested to hear the community's feedback on as we launch Infra Finder is: OK, the Infra Funder is a platform – we collected basically the same information from all of the providers regardless of whether or not they are commercial, non-commercial, for-profit, not-for-profit – whether or not they, you know, are open source or not open source, for example – it's displayed in the same way, the same format, the same website, the same interface. And so, in a sense, like you are kind of – we are trying to make some of those solutions that are perhaps less discoverable because of

that lack of marketing capacity to be more visible, to be equally visible, in the same platform, right? How does that affect the decision making? And then the other aspect is: in the end, we know that different institutions have different requirements and different priorities and that's not to be penalised. You know, we don't want to say: OK, because you can adopt open, you are doing better than other institutions. It's simply that we do all have different constraints and we do really want to recognise that. And, equally, for the infrastructures – again, for not having, for example, open governance minutes, let's say – it's not something that we want to penalise the infrastructure for. It's sometimes, you know – sometimes it is that they don't value these things and that is not important in their development, but I would say that's fair. Actually, most of the time, it's because they don't have the resources and time to be able to do these things and they are considering it and they're gradually moving towards being able to be more open, if you will. And so – one of the things that we have thought a lot about and have put into our design and you will see that in the launch of Infra Finder is for example, the indication where things like the Code of Conduct is implemented versus in progress, right? And we do want to recognise progress towards that value alignment as opposed to just saying: these solutions are open, these solutions are commercial. Because it is really – the open infrastructure world is not a binary. It's rather a spectrum, and that works, actually, very well for the spectrum of needs that we see in both the user space and also the provider space.

00:20:52 PPA

One very basic, or, to say, philosophical question, even, is: what is an infrastructure and what is a service? Like, to give you one concrete example – we had a previous podcast episode about DataverseNO. Some colleagues of mine were interviewed there and they took an open source software that had been developed initially at Harvard University – the Dataverse software – and they set it up for Norwegian purposes, in the Norwegian context. So there are now many, many partner institutions across Norway that use the same software. It's run by one institution, UiT, but what they are very, very concerned with is this definition of infrastructure. DataverseNO is an infrastructure, it's not just a software and a website. If you will, it is something run by people. So there are people managing all their research datasets that are submitted to DataverseNO before they get online, before they get published I mean. So, as they define it, it's the people – the support people behind the infrastructure are part of the infrastructure. But I don't know if this definition would apply to the Infra Finder. Do you have some thoughts on that?

00:22:17 ET

Yeah. I mean, we have – we have a lot of thoughts and it's tricky, it's definitely a challenging question. Similarly to the sort of spectrum of openness, we believe that infrastructure is a spectrum. And so, if you go back to the classic analogies for thinking about infrastructure – things like plumbing and pipes for water, or bridges and roads, right? – you know, there's an element of dependency where there is a community or group of users who are dependent on that infrastructure or that system in order to perform their day-to-day work, for example. And we've explored this further in, actually, a literature review that we've published last year. But this doesn't really help us in differentiating between what you're saying, what you're calling software, and what we're calling infrastructure, right? Software is also being dependent on and indeed it is incredibly contextual and that's the other thing that the other main point that came out from our literature review. The example that I would give here is disciplinary-specific data repositories, for example, where, you know, you have a very, perhaps a small community that is doing research in a relatively niche area. They have a data sharing infrastructure that the community uses, but it's a small community – so it's also not just about how many people are using it or how many people are dependent on it, because that infrastructure is critical for that particular community. And so it really, like – in engaging with these different communities and doing

that research, we have gained a very deep and empathetic, if you will, understanding of the fact that many times you can't draw a straight line between these technologies and systems and say: these are infrastructures and these are not. It is really dependent on, you know, the communities and the communities that have built them and what they think is important and what they are using day-to-day. And so, you know, in listening to your case around the the Norwegian community, you know, the questions that I may ask in that case is: why is it important for you to answer this question and what are you trying to achieve in seeking an answer, right? Why is this answer important for the community, to understand what is infrastructure and exactly what is not, and how does that help us make better choices in terms of our investment into the technology itself?

00:25:06 PPA

Yeah, you could say that – if a municipality needs to take care of the water supply, for instance, of course they need to set up lots of pipes and things, but it doesn't really work all by itself. You need service personnel that can fix things, repair things and expand things as more houses are built and so on. It is a little bit similar here – that we don't want them to just think: OK, so this is the software, now we have the solution, it will run by itself. Actually trying to make visible the amount of personal costs that that we need to do, we need to have those fixers and those maintenance people, the service personnel, so to speak, to have the the pipes working in the long run.

00:26:00 ET

Absolutely. I cannot agree more to that point and it's – something that, you know, we see as one of the biggest issues in the infrastructure world is the fact that first of all, people don't usually notice infrastructure until they break. And so the labour behind the maintenance and upkeep of the infrastructure is often invisible until the moment that the infrastructure breaks and you realise that you actually need these people. And so an investment into the operations, the maintenance, the addressing of technical debt, for example – for digital infrastructure in particular – is incredibly important. And it's also something that is relatively underfunded because of how, for example, grant funding works, right? Grant funding is typically, unfortunately, very much oriented towards new, novel solutions and it incentivizes – it builds a culture that incentivizes the creation of new things as opposed to the maintenance of existing which actually works. And so it's something that we have been in conversation with a lot of funders about and we are hoping – you know, we're also in connection with a lot of other initiatives who are understanding the importance of investment into maintenance and into operations, to try and see how we can help move that needle into generate more awareness about the importance of maintaining infrastructure. There is of course a lot that we can also say about the work that we've been doing into understanding things like: what sort of financing models would be more suitable towards maintenance. For example, what sources of funding – what source of financing mechanisms – should be used for what, and I think this also builds on existing work in the community. A lot of people have been looking at this and we, sort of, take all that knowledge and try and learn and build on that. The other interesting aspect of this, I would say – and something that we are paying increasing attention to – is something that we're calling temporarily calling graceful transitions. It's about understanding, you know, there are – not everything needs to be maintained and not everything needs to be sustained into perpetuity, as much as not everything needs to be grown infinitely, right? And so how do we normalise and make more visible the transitions in the space that are actually good for the overall ecosystem and that, sort of, natural life cycle aspect of it? How do we talk more openly about these things that are not usually easy to talk about – sun setting, departures of key personnel, for example, which a lot of people, like, in the gut feeling considers, sort of, not success, definitely – but we really want to understand how we can change the narrative around that. And that's something that we actually are

writing about at the moment and which will become part of our the first iteration of our report, which is going to come up annually, called The State of Open Infrastructure Report, which will be released, first iteration, in May.

#### 00:29:48 PPA

Yeah, the Infra Finder is not – the sun is rising on that one, it's not a sunset thing because it's brand new. But, ideally, how would you want to maintain it and perhaps expand it in the years to come? Do you have some thoughts about the sustainability and how to make sure that the information that is there at any given time is as up to date as possible?

#### 00:30:16 ET

Absolutely, it's something that we are – I'll start with saying that it's something that we are thinking very actively about and trying to learn from, again, other initiatives in this space who are maintaining, sort of, databases like this. So here's a call to anyone who's listening: if you have additional ideas, please reach out to us and we'd love to learn from you. The other thing I would say is that with the development of Infra Finder and its previous prototype, we have always taken an iterative approach to development, meaning that we want to continue to make sure that the tool, the product is useful for the community. And the usefulness changes because the community and their needs change. And so what we are hoping actually to achieve with this launch is to really get users' feedback as much as possible. We encourage you to use it, explore it and tell us what you think about it. Is it useful? Do we need more services? Do we need more information? Do we need to rephrase things in a different way so that it can be more useful for you? Do we need to develop new features? Etcetera, etcetera. And these are all – there are no wrong answers here. All feedback is useful in helping us learn how to develop, how to further expand and develop the Infra Finder. And then, in terms of ensuring that the data is up-to-date and maintained, we are extremely grateful in this first release to be working with these 57 service providers who have been just phenomenal in working with us and suggesting additional – you know, giving us feedback on how, for example, we can improve our data collection instrument that we've been using, what they would like to tell their potential users and adopters, etcetera. How do the data fields work for the different solutions? Again, this is all incredibly useful feedback that they've given us, and we've tried to incorporate that in the way – and with that process we've built a close relationship with these folks and we intend to keep doing that. So at launch, there will also be the launch of an Expression of Interest Form for additional infrastructures to be added to Infra Finder. And if you are developing or maintaining an infrastructure, I would strongly encourage you to consider whether or not you want to express your interest in being part of this. It will be a wonderful opportunity not only to showcase some of the work that you are doing, but also to – yeah, really work with us and help us understand how to tell the value and the story of your work better and to showcase that to the users. And, if you are a librarian, if you are a target user or anyone who has come across Infra Finder and found it useful, there are ways that you could support the development of this tool. It's been a very long journey of developing and understanding and designing this with the community and we do really want to be able to continue this work if it's valuable for you. So on the website there will be additional information about how you can support the development of Infra Finder, financially and otherwise, and we really would love for you to check that out and to let us know, even if it means – giving us feedback is definitely a way to contribute to its ongoing development. But if you can tell your colleagues, for example, tell your peers about Infra Finder and encourage folks to use it, that would also be phenomenal.

#### 00:34:14 PPA

Open Science Talk is produced by the University Library of UiT the Arctic University of Norway. Thanks for listening.