

## Interview 1

|             |                              |
|-------------|------------------------------|
| Interviewee | 01-Esri-A                    |
| Interviewer | Ashraf Shaharudin (TU Delft) |
| Date        | 11 May 2023                  |

### Interviewer

Yeah. OK. So my first question. And my first questions are background questions. The first one is could you please describe your role in [Esri distributor in country A]?

### Interviewee

I'm <redacted>, and I am the <redacted> in [Esri distributor in country A]. What does that mean? I look after <redacted>. And they all fall under my umbrella. That means I have about <redacted> in the [country A] to look after, so it's quite a lot in comparison to some other countries.

### Interviewer

OK. And how long have you been working in this or similar role?

### Interviewee

So in this role, I am in my <redacted> year doing this this role in <redacted> in the [country A]. Immediately before that I worked in <redacted> for an organization called <redacted>, which is at the <redacted>, which is a long history of running geo-enablement services, and during that time open data really kicked off. We had open data portals before open data was a big thing and I was in that landscape for <redacted> years. Prior to that I was a <redacted>.

### Interviewer

OK. And how is your current role related to open data?

### Interviewee

It's not directly related I suppose, although we try and make sure that <redacted> have access to the data that they want to enable the teaching. And often the most difficult thing is finding that data. So it's signposting into existing open data. Sometimes, it's taking open data and turning it into actually usable data. You know, PDF is open but it's not usable, so we still do some open data creation within our team. But it's mainly signposting obviously, so awareness.

### Interviewer

I'm gonna talk about open data services by Esri now. Could you please describe the open data services offered by [Esri distributor in country A], including who are the services for?

### Interviewee

So through the Living Atlas, I suppose we've got tiered levels of encouragement and enablement, whether or not it's authoritative data. So, we encourage customers to make their data open where they can and to make that data accessible primarily to users of an Esri system but most Esri distributors would agree that if we can make the data available in our system, it does not exclude anybody who is not using Esri software. But we really have that customer-focus. So there's a little bit of the data team within [Esri distributor in country A] which is trying to surface open data that they find to make it more accessible, generally around projects that they might do, and definitely with our grads, we encourage our grads to do some data wrangling. You'd know from your studies, working with dirty data is really interesting to make it usable. But most of our focuses on customer engagement and encouraging customers to make their data truly accessible. And that is either through their own open data portal, which is stood up usually through ArcGIS Online probably in

portal or more laterally in an implementation of ArcGIS Hub. Something like that, but it's that enablement. Some customers do it because they think it's the right thing to do. Other companies and organizations have a legal responsibility to do that, so we help them comply with data laws in the [country A]. Our job is generally helping them make their data accessible and less of it -- think of an 8020 rule, 80% is focused on helping the customer and maybe 20% is focused on taking some data that nobody really owns and we'll just make it accessible and see who uses it.

**Interviewer**

OK. So in terms of encouraging, I see two components here. The first part is actually you provide the platform like the Living Atlas and ArcGIS Hub for them to publish open data. But you also mentioned about helping them to comply -- to facilitate them to publish open data. So is that more of like a consultancy services?

**Interviewee**

It can be. It depends on the technical knowledge within the organization. So it really is dependent on that. It's almost coming on to your next question about our history of things. In <redacted>, there was a big push from the [country A] government to go open. The reasons for that was that they were seemed to be an economic stimulus from making data available. So there was lots of money pumped in to try and make data available. But there is also a little bit of a backlash against companies like Esri, who charged money for software. It was conflated into open data and open source, and actually more laterally people have realized that you can have open data and you can be using a proprietary system. So, <redacted> kind of that all kicked off. And our job is really still hand holding.

So, coming back to your question, if you've got an organization with a remit to make data accessible or a desire, but they don't have the skills, what we'll do is help them build an open data portal. That's usually the most difficult thing, get the architecture right for their portal. A lot of it is now off the shelf, which is good. It's got easier. Our early adopters had more hassle, more problems, and now we're into a smoother transition. And then we will maybe just educate them on the quality of the data, the metadata, the boring stuff that you have to write, which is super useful otherwise you don't know what you're using and what you can do with it. And it's about knowledge transfer. So there may be some consultancy into that. It may just be part of the thing that we do. We have people who, you know, this is their job. So if they can help customers as part of their site license agreement without charging them any more money, they see more value in working with us. It's not just about the money you pay for the software, it's the support you get around it. Other organizations, you just have to give them the license codes and then you go off and they do it. They have experts.

**Interviewer**

And the organizations that you sort of have to handhold are mostly public organizations or private organizations?

**Interviewee**

It's a mix, really. There's a lot of third sector as well. If we're thinking now -- I don't know what you would call it in <redacted>, not-for-profit charity work, so third sector. There's a lot of organizations there that don't have as many staff and they have a lot of things to do. So we have a very small team in the third sector which will help those organizations as well. But it's a real mix from commercial organizations. More on the public sector government funded organizations and the third sector, but also the university sectors. If your research is funded by public money, then you should be making your data publicly accessible.

**Interviewer**

And apart from providing services to publish open data, do you also provide content of open data in the Living Atlas? Do [Esri distributor in country A] does a lot of that?

**Interviewee**

We used to. Our methods have really changed. In the beginning, we used to take data which was given to us by customers and we would host it in ArcGIS Online and we would clean it up and we would write the metadata and we would put it then into the Living Atlas and we had our own [Esri distributor in country A] open data portal. And really, why did we do that when it should be the data creators who are hosting the data? Well, sometimes you have to show them what it will look like for them to invest in the process of making that happen. And we've kind of run our open data portal down as their open data portal sort of come up, but the aggregator is the Living Atlas. Does that make sense?

**Interviewer**

Yes, it makes sense but it's an interesting point.

**Interviewee**

So sometimes you have to do the hard work at the beginning. For five years down the line when you can transfer that responsibility back to the big organizations and then you can help the smaller organizations who can now see the value of an open data portal. And that technology has matured and it become easier to use, easier to implement its scale. So these small organizations feel that they can now do it.

**Interviewer**

OK, so in a way you don't see that providing content is a main services related to open data that you should continue providing?

**Interviewee**

We're a software company. So largely like if you think about Esri holistically and you may get a different answer from different people, but we are a software company. We make money from selling software and solutions. We're not a data company, our customers are our data companies; they provide data, they make data. We don't make any data. We don't maintain any data. We're not the authoritative source on data. We'll take data that we think is useful but not usable and we'll surface it where we need to but we would rather the data owner takes responsibility of that data and they maintain it and update it. People ask us for data all the time, and it's like we're not -- we're the software, we're the software company. You're the data people.

**Interviewer**

And so what are the benefits of offering those open data services that you mentioned to [Esri distributor in country A] itself? So why do you do that from business point of view?

**Interviewee**

You can't do GIS without data. Fundamentally all of our analysis tools require data coming in. And problems, data-driven questions require data as well, so the more data that is out there, the more people can take data from different sources to combine it with their own asset information to answer their own questions. So if you take climate change as a really good example, we're working with the <redacted> and they're at the beginning of their open data journey. Well, actually there are not at the beginning of the open data journey, they're halfway through their open data journey, but they're suddenly realizing that their open data that they've had up to this point is not really suitable for the end user. So they've transformed their data into more accessible GIS formats to go forward.

How did they get there? We had to do some of that. We had to demonstrate why we thought what they had wasn't right because we knew what our customers wanted. So if you think about the ecosystem of customers, we've got the big guys at the top -- the Government, <redacted>, <redacted>, some of the big engineering firms, you've got utilities companies -- and then underneath you've got housing authorities, local councils, and transportation companies, water companies, and they're all asking and answering questions. But they need the data from these big people to flow down to them and it has to be reliable, up to date, it has to have metadata and it has to be almost plug and play, it has to be usable. And sometimes these big people don't get it right and where we can see that. So if you think of us in the middle, we're listening to our customers going our way: you know, we've got these problems, we need to look at the exposure to climate change, we need to be running microscale climate models against our buildings, to look at our assets and work out what we need to fix them, when we need to fix it, we need to know what's coming. And it's like, well, why don't you go and get the climate data? And they're like: have you ever tried to use it? And then you look at it and you go: oh my God, it's raw climate model output, this is terrible and there's so many models, which one do I pick? And then these scientists are going: but the data is available, we've done our job, we have ticked the box, our data is open. But it's not usable, it's not usable by the end users. All of these people are screaming out for usable climate data. That stuff requires one of them to be a climate scientist to understand it, and they're not, they're GISers in housing or transport, that's their interest. You're the climate scientist. So after a while you go, maybe it's our job to do this with the big companies and the big organizations whilst listening to the pleas of the general consumer. And once you make a couple then you change the mindset up here and you increase the value proposition coming in and then you link them and then you step back. Because we're supplying software up here and we're supplying software down here. And if these people are using Esri, great. If they're using QGIS, they can still use it. But they have the same problems and it's that bit in the middle. And often what we see is the bit in the middle. If I see the value of Esri beyond just getting software, it's that we speak to everybody. We know the big problems, we know the little problems. And if we can fix them because we can see the shape of the jigsaw piece in the middle, then we can do that. But then we want to step back because we're a software company, we're not a data company. We have consultancy, we have a big consultancy arm, so we often step in and we use our consultancy and we make a solution. And once it's up and running, we just stepped back and then it should just work.

**Interviewer**

OK, very interesting. And so what are the main technical and non-technical activities that you have to carry out in order to provide these services?

**Interviewee**

There's obviously a lot of data wrangling to make sure that the data is in the correct format. So on the technical side this data wrangling. There's less and less coding to implement solutions. If we go back five or ten years, then, you might have to tweak some code to make a solution work in the way that you want. But as our software stack evolves, there's less and less coding. So you wanna be able to do off-the-shelf and then you want to do a little bit of customized off-the-shelf where it's just configuration of the off-the-shelf tool. And you don't wanna code. Now the reasons for not coding are time and money. So the more customization you have, the more likely it is to break when there's an update to any part of your stack. The more standard your implementation, the easier, the more robust it is to be kept up to date by software updates. So the technical skills are on data wrangling, and then the customization of off-the-shelf tools where it's just customization configuration, less coding.

Non-technical skills is going back to -- I think the most important part is going back to listening. You listen to the problems from organizations. And you work out what their need is, and then you work out how you should expose that data. What format should it be? So if I go back to climate, the climate models were all dumped out in NetCDF. Have you heard of NetCDF?

**Interviewer**

No.

**Interviewee**

So NetCDF is one of the worst file formats that you would be unfortunate enough to work with in geospatial. It does quite a lot of really cool things. It's essentially a 4 dimensional. It's got well 3 dimensional, but you can use elevation or time as the extra dimension. And that's cool for climate data, but it's a pain to work with. So why not have it in a feature service that you can just use an ArcGIS Online straight away or in Pro or whatever, or you could then transform it from that feature service to anything that you want, to GML, and then use it in whatever system. So understanding the needs down here and understanding that delivery mechanism up here is not a technical thing, it's a listening thing. Listening is super important. How you then turn that data from NetCDF to a functional feature service isn't easy. It takes a thought process. And a little bit of experimentation to make sure that everything works. But when you get it right, it allows you to then enable your dashboards.

So have you heard of start with the end in mind? So starting with the end in mind for the climate data was that -- our little team in <redacted> actually took the lead on this because we could see what we wanted. We wanted our data-driven dashboard that a <redacted> could understand, that they could change climate models and they could change the years and they could have pictures changing, maps changing. So it needs one feature service, but you have 150 NetCDF files with multiple models, multiple scenarios for climate change. But you want one data-driven dashboard driven by one data set. So the technical exercise is then to get the data into the right format to allow you to serve the solution to your customers. In this case it was a <redacted>. So there's a lot listening, working out what their needs are, and then the technical exercise is implementing the solution to get the data usually in the right format, building the dashboard turns out to be really easy if the data is in the right format. When the dashboards are really hard to build, it's probably because your data is in the wrong format or not quite optimized. So the technical skills, data wrangling, a little bit of configuration of tools. Non-technical is mainly listening, understanding user needs.

**Interviewer**

Yeah, but I see the role of [Esri distributor in country A] is more of capacity building instead of like providing simply data to users. Then in terms of sustainability also then you also have to make sure that the these organizations that you have provided sort of like the first start should be able to continue to do these things, so do you also provide training to them?

**Interviewee**

Yeah. So in <redacted> example we took the lead in <redacted>. There was <redacted>, which was happening in <redacted> and -- when did it happen <redacted>? I can't remember. That was our delivery point into the <redacted>. So we had to make something, but we were working with the <redacted> and showing them what we were doing. And eventually the directors at <redacted> could see why what they were doing was not acceptable and how to change. The question then was how can you take what we've learned and turn it into a deliverable for them. So we built some online learning because actually I went to teach <redacted> how to take complex climate data, make data-driven dashboards. So we made an online learning object, we delivered that through <redacted>

team in structured training and we continued a dialogue for free with that team to make sure that they're happy and they know how to work with the data. So there's also been some consulting. The consulting arm of [Esri distributor in country A] built them, they're open data portal and help them stand that up and get some data into it and think about the structure and the metadata. And then they took the learning object that we had and they're now adding data in using that methodology to make it accessible to the end user. So the knowledge transfer is crucial and it can either be done through consulting, structured training. In our case it was semi altruistic; we had a desire to do more with climate, they're are big climate organization, actually we wanted them to cobrand our resources into <redacted>, so we asked them for help on the climate science and the validation, they asked for help transferring knowledge into their GIS team. So we got something out of it, they got something out of it. No money changed hands. It was just time. So they spent some time on our project, we gave some time back into their team. Everybody goes away happy and we got our data dashboards into <redacted>. That's how it works in <redacted>. When money flows, it's a little bit more complicated, but you know we've got that relationship as well. But you're right, you do a project, you then you have to transfer the knowledge and then hopefully every month or every year there's less contact with the customer on that thing. You're actually doing something else. It's like, what's the next problem you can help with?

So in a consulting model, you don't wanna go: "right, I've helped you, so I'm never gonna get any more money from you". You wanna say: "OK, we've solved this problem, we'll touch back in six months' time and we'll see if there's any more knowledge you need on that but what's your next big problem that you think you have? How can we help on the next thing?" Because our software, our entire industry is getting better. Our tools are getting more usable. The problems that we can solve are getting bigger. We have real time data. We have more AI than we've ever had before. How can we use it to solve problems. We could just say we can't solve this one at the moment, so we'll park it for five years and we'll come back to it. Which one of those part projects can be tackle next?

#### **Interviewer**

Yeah, but in this particular example that you gave on <redacted>, so you said, like, there's no money exchanged because it's more of like you both get something from this. But are there a lot of like other cases that actually they are proper consultancy service, that money is exchanged?

#### **Interviewee**

Yeah, yeah, definitely and there's no hard and fast rules on whether it goes down a consulting or if it is done as sort of pro bono stuff. Some organizations have service level agreements, so they maybe have 10 days of consulting time every single year, which is worked into their contract. So if they're coming through a year and they're like, "hey, we've got six days left and we don't really know what we want to do but we've always thought about doing this, but we need some time and help". And we're like: "cool, well, we'll see how far we can get in six days".

Does that make sense? So we do have this consulting arm and we have this -- I think all of the Esri distributors that you've mentioned usually have really good contacts with their customers. It's not: "Hey, Ashraf. How much money have you got? How much software do you wanna buy this year?" It's like: "hey, you bought your software, how do you wanna be more successful with that software? Tell me all of your problems and let's go through them and look for the easy ones, and then we'll move on to the medium ones. Some of them are critical. Let's move them around." The relationship between Esri and the customer: we should understand your problems, you should tell us your problems and we work out which ones are easy, which ones are hard to solve, and then we try and solve them together. That's how you build a long term relationship with good customers. We used to maybe be more: here's your box, give me the money, see you next year. But it's really changed. Any

business that wants a long term stable business model has to invest time in that customer relationship and often it means doing some stuff completely for free and then you know eventually they want the solution and it grows, and it requires more licenses, and it requires a little bit more support and a bit more training, so there's a little bit of revenue, but it should always be for the benefit of the customer.

**Interviewer**

And so how many staff are involved in this providing open data services?

**Interviewee**

Who knows? Who knows? So on our content team is maybe 15 people strong, but they provide both open data -- supporting customers with open data -- but they also provide paid for data, so premium services. Those services might be access to things like more <redacted> data or more <redacted> data or hydrographic data where these organizations give us the data so we can serve it directly to the customers and then we charge the customer, we take our bit of money for the hosting and the service and we pay the license fees back to the data provider. Does that make sense? So it flows through -- so the data provider just updates -- we update the data, we serve it to the customer. So that team as a whole is maybe 10-15 people big, but it has dual roles. But then the professional services are consulting wing. They could be working on a project with open data, they might not, and that team is 200 people strong, so there will be people in there working. So it's really difficult to give you an exact number.

**Interviewer**

And what are the skills of these staff? Because I imagine that you don't only need people from technical skills, but also, you know to do engagement, then you need people skills as well.

**Interviewee**

Yeah. And it's comes back to you've got some technical skills to stand up and implementation and understand the technology. So there's usually a senior solutions architect that looks at it from a broad level and then there's a solutions architect to implement the solution. And you'll have a customer success manager type person who understands the customer needs and understands the technology capabilities. So they're the ones, you know, that jigsaw piece that I talked about, they're the ones that should understand the problem and then they asked the senior solutions architect to make a solution to fit the problem and then somebody implements it and then somebody transfers the knowledge. But it's that team there would be kind of separate from our data team. They would maybe liaise with the data team. But they work autonomously, that 200 people. Sometimes they ask the <redacted> team as well, sometimes they don't.

**Interviewer**

And what are the cost components of providing these open data services?

**Interviewee**

Sorry, what's the cost?

**Interviewer**

So I don't need the figure of the cost, but more of like what do you need to deliver these services. So you already mentioned the human resources, but also perhaps I don't know, server or anything else, that sort of like formed the cost to provide open data service.

**Interviewee**

Well, so, so again we try and do as little as possible in terms of hosting data. The data we do host

tends to be some of the premium data which is not open, so it's maybe not your focus of your investigation. In the past, where we used to have more open data, we would run it in either portal, so ArcGIS Online but on premise, so that requires a server and hardware maintenance. The staff cost are IT team who would update and security patch our server. But we've moved to hosted environments and now our servers are in the cloud but we still require -- so the box, they're now maintained by a third party but our IT people maintain access to it and our consulting team then manage the data into it. But if we can, we will use ArcGIS Online and then we just pay for the ArcGIS Online account. Then all of the maintenance is taken care of by Microsoft for Azure, Amazon for AWS, and somebody at Esri Inc who looks after the Eagle <?> implementations and then we are just looking after the data. So from a cost delivery model, the most expensive is where you own the infrastructure and you maintain it because that infrastructure maybe has 5 to 10 years of shelf life before it's so slow it's laughable and you have to have an IT staff and security patches and all of the risk and you pay for the electricity and you need a building and you need air conditioning. And then the next cheapest is that cloud server environment. And the cheapest for us is ArcGIS Online. And each time you move towards cheaper, you have less humans involved. It's not replacing the humans, it's repurposing them to do more, so we can we can engage with more customers if we use a pure cloud environment because we don't have to deal with all of the infrastructure. So we can have more solutions for more people at scale if we use that. But most of that cost should be transferred to the customer if they are hosting their data as I go back to that point that we used to do more of this, but now we want our customers to do it.

#### **Interviewer**

Hmm. OK. What are the challenges encountered by [Esri distributor in country A] in providing open data services?

#### **Interviewee**

In the past, it would have been maintaining those data, making sure that they're up to date, and just up to date at the right time, if you've got a quarterly update or if you've got a 12 month update, just making sure that they're all up to date. Now, because we're trying to do less of it, it's actually reminding our customers and just helping them when they have problems. So if you think about that staffing thing that I've said, you actually grow the need for more support staff to support more customers during their own implementation. If you were doing the implementation, then you become the expert and you need maybe multiple bodies to manage the data, but you can shift those people over into support, to support a wider customer base. It's just little technical challenges, but it's where that effort is directed. It used to be directed towards our services, but now it's towards those customer services.

Other bits, some data changes that the data schema changes so then, longitudinal studies are more difficult because there's no harmonization. So that can either be that the data structure has changed or the geographic area has changed. I don't know what they do in <redacted>, but the [country A] has a super complicated geographic structure from tiny areas to big areas and some data is continuous across the country, some data has holes and you have to combine it with another data set to make it continuous but then you can't compare these; so complicated. So sometimes understanding the underlying geography is the most difficult thing and you need to help the customer understand the geography to do it. Other things might be around anonymizing the data, so data protection issues. Before you can make your data accessible. "You need to make your data accessible". "OK, but I have some individuals". And it's fine in the city but when you're in a rural area. You know I live in <redacted>, we have a very rural bit of <redacted>, we are predominantly white country and the further north you go, the whiter it gets. So then, if you have ethnic minorities in

areas, they can be identified because they stick out, there is only a handful of them, so how can you deal with these and things like crime and location. So often we're providing the techniques around those little things about how to make it anonymous and protect people's identities. And then we have data governance as well the GDPR rules which are across Europe <redacted>. But yeah the challenges are around those. I think the technical challenges largely have gone away. There used to be: How can you make this service fast? How can you get so much data into the cloud? How can you make it responsive? Those things largely have disappeared, and we do less coding. We do more customization, configuration. The tools have matured. So now you're getting into how can you make this data secure? How can we update it automatically? How can we compare different geographies? How can we protect individuals? Those are the really interesting questions around the ethics of some dataset.

**Interviewer**

Do you find it hard to convince public organizations or big companies to take on your open data service?

**Interviewee**

And it yeah, it depends on the organization. Some will be harder than others. It's really around -- often it's around convincing them that the data, the open data they already host is not fit for purpose. Now going back to that Jigsaw piece, it's like we're in the middle going: "yeah, yeah, we know that you say that you do open data, everybody wants to use your data, but they hate what you're producing. Can we help you help everybody else?" And sometimes that takes just one meeting and they go: "Yeah, we get it". And other times it takes a couple of years to change the mindset. So it depends on the organization. The cost has come down because go back to the previous question, you know the hardware versus the managed service versus pure ArcGIS Online; so, the cost is coming down so the financial burden of doing open data has come down massively in the last 10 years. So that argument is gone. But then those organizations may have less money than ever to actually do it. So the argument comes back up. But it's around: "What are you doing? What do you need to do? And are you doing it the best way? Is there a better way?"

**Interviewer**

But do you also try to convince organizations that haven't yet published open data to also make their data open?

**Interviewee**

Yeah, where they can, where they can, yeah. Even down to commercial drone capture organizations. Sometimes we work with them to make their data accessible, especially in the education spaces, training data. And some of them are really good. Some of them -- one company gifted us 10 flying days just to make education resources. They said every time we go to a site we'll fly the customer's data and then the next day we'll fly some data for you and for us, where we can use it in our demos and we can give it to you. And if the customer complains, we've got two different flight days, so two different data sets. One they pay for, one they give. So, some organizations see the value of making some data. Obviously they want people to see that they're doing stuff and then they go, "OK, drone data, that company seems good. I've seen their data before. Let's go and speak to them about flying some commercial stuff for us". There's always a sales pipeline that people think about.

**Interviewer**

OK. Do you have any example of projects or cases where Esri open data services, demonstrated impacts in terms of like big reception, in terms of like maybe media attention?

## **Interviewee**

Yeah, I mean the <redacted> data is living -- is accessible directly through our platform. So we can see the use metrics. On that, more customers will be using some of the open basemaps that they produce. And we can see how many people click on it. So, it's demonstrated by a huge utility.

But then if I go back to the <redacted>, that narrative with the <redacted> was really strong. So they couldn't see it before, we make some data available, they then make their own data portal, people start going: "this is exactly what we wanted" and that encourages them to put more data into that accessible portal. And then they feel like they're serving the right community now with the right data and it then grows. Yeah, that's probably my personal favourite. It's completely mindset change within that organization on how to deal with it.

But the <redacted> as well as the <redacted>, which is our regulator has made loads of data around flooding available and also water quality. Lots of people in the [country A] have started swimming in rivers in the last five years. Water quality in the river is not great. Making dashboards showing what a quality in real time live data, freely accessible, gets people to think about complaining about the water quality. So the regulator would love to make our water cleaner. How can we do that? Well, we need lots of people to be really angry about the water quality. Just publish the water quality data, make people angry, show them what they're swimming in and then they'll complain to you and you can complain to government, and then they can force industry to stop polluting the rivers. So it's a circular argument. You know, the government, they're meant to serve the public but if the public don't tell them what they're angry about, how do they know what to improve? And we have a role in this, as GIS experts, because we can surface the data in an accessible format to the public to change their mindset to force government to change legislation to improve the quality of the thing we're complaining about. But it starts with an organization making their data open but accessible, so the water quality data was made into a dashboard and it used big brown symbols. Brown, dirt, poo all of this symbology was tied to you know, the bluer, the cleaner it was. It got brown, the dirtier it was. Super simple. You could have a 5 year old understand where they should and shouldn't swim. And it changed people's attitude. They like: "I love swimming in the river. Oh my God. This river's filthy. Why is our river filthy?" Because here's your polluters. OK? What can we change? So those are a couple of examples where you can actually change policy by making data accessible and making people angry.

## **Interviewer**

OK, very interesting. From the business point of view, how do you think Esri can leverage open data even more to support the business of [Esri distributor in country A]?

## **Interviewee**

Well, it goes from the big guys down to the small guys really. We've got lots of customers that need to do more and they need data. Data is the fuel for the GIS. So the more data that's accessible, the more questions that can be answered, and the more people will be using GIS and data science. As you just have to keep -- But we're doing that: "Would you like to just make your data available?" "No, it's too expensive." "It's really not. So you've already got all of the tools. They're included in your license." "Well, we don't know how to do it." "Well, we'll help you." "You got any money? Yeah, well, a little bit of money we can give some days. And you can do it." "What's in it for us?" "Well, it's a good news story. And actually it will support all of these people that you should be supporting anyway." So we want to do that. It will just help us, partially through consultancy money, but mainly keeping our customers happy because they need the data to ask and answer their own questions through our software tools. If they can't do it, there's no need for GIS if nobody's got any data. So again, we're not the data company, we're the software company but we know the data is crucial, so

we try and encourage everybody to make their data available. I had lots of people want to ask and answer questions and they have to use our software.

**Interviewer**

Do you see for example events like -- I'm not sure how, what is it called in the [country A], was it Esri Tech conference or something like that as an avenue for you to showcase what you've done with open data for certain organizations and try to convince other organizations to perhaps engage you for similar services?

**Interviewee**

Yeah, definitely. I mean, our conference is next week, which is part of the reason why I'm a bit stressed this week. But the message definitely come across. So <redacted> we had the chief scientist for the <redacted> gave a keynote presentation for half an hour to 3,500 people talking about climate change and how organizations need to prepare for the climate change that we know is coming. And then everybody in society do as much as he can to reduce the climate change which is coming after 2050. So we can't change anything up to 2050, it's already locked in, but we have an opportunity to make life better or worse. And that rhetoric was aimed at everybody in the room. If you're a transportation company for a city council, what do you need to think about? Is it flooding? Is it heat? If you're a housing authority, what do you need to think about? We have no air conditioning, really in <redacted>, but we might need it in 2050. Our buildings are not suited to keep the heat out because we don't have much of it at the moment, but that will change. And he was saying, you all have knowledge of your local problems in your infrastructure, [and] what we have is the climate data, tell us what data you want, we will make it available. They already had our script on how to do that. That is the role of our technical conferences. So this was the big guy at the top of the science tree saying we're making more models, we're going to make sure that they gonna come to you in a better format, but please communicate with us what you want, please tell us what you need. And then they had some technical presentations on what they were doing. And actually we had some customers that were already thinking about climate change, giving presentations. So our conference last year <redacted>. So it worked quite well. But we use that as some customer presentations and then some technical Esri presentations, so it's like: here's a solution, here's how to build it. So it's the customer that's been often helped to build a solution presenting the problem and the solution and then Esri showing how you as another organization with a similar problem can use the tools and create what they've created for your question. And that's the role of our conference really.

**Interviewer**

OK, very good. Now I'm gonna talk about Esri in the open data ecosystem? Do you think [Esri distributor in country A] plays a role in enhancing access, supply, or flow of open data?

**Interviewee**

Yeah. And it comes in some ways. In 2011, Esri was not doing a huge amount for open data. So that's when the open data movement really kicked in. And if you look at the change: so back then, it was academic institutions who were leading in the supply of open data. I was <redacted> and we were making data as accessible as possible. And then governments started to think about whether they should have an open data portal. So there is data.gov.[country A] and there is a <redacted> data infrastructure project, which actually the <redacted> government asked the <redacted> to build their portal because they had no idea how to do it and we all had already built one, so we built them their portal and that became the <redacted> Government portal. But it was built by <redacted>.

And then it started to flow into -- so Esri is a juggernaut, an oil tanker. Sometimes it's right at the front and sometimes the solutions take a while to catch up, but then open data became part of Esri's

language, both at Esri Inc and [Esri distributor in country A] and all of the other distributors, and we were producing the portals. But as I've told you already, the next transition is there's an aggregator here called Living Atlas. If you make your data in your open data portal, you can feed it in and then anybody can discover it. You host it, you update it, it's yours, you own it, you'll never lose ownership, but it will be surfaced through metadata discovery here. And actually that's all the SDIs were doing back in 2012 when <redacted> them using Geo Network, one of the worst tools that you could ever hope to engineer with. Really complicated, never fully worked unless you coded against it. It was really hard and now it's super easy. I can stand up and open data portal and less than an hour probably. It took us flipping months to do it. But it's a really good transition. You have a government initiative, you have the early sort of like bleeding edge, then you have that transfer of responsibility into sort of government and then commercial organizations. We're not making a killing out of open data, but we have lots of customers and we have this model. So if you're part of our software ecosystem, we'll make it as easy as possible for you to use the licenses that you're already have to comply with your remit to have open data, and we won't exclude any other GIS providers from accessing that data and using it within their systems. But we'll make it really easy for our customers. You can discover the open data directly from Pro and Living Atlas, so just makes it easy.

**Interviewer**

Do you think that [Esri distributor in country A] also plays a role in connecting other open data actors?

**Interviewee**

Yeah, I mean, again, if you're an organization that have loads of open data you and then you're an Esri customer, the route is there. If you're not, then it's slightly more difficult. What I'd love to see is sort of a another metadata scraper which scrapes non-Esri portals and pulls them into the Living Atlas as a separate batch and says, well, we don't have any in our ecosystem, but we found some in another ecosystem but I'm not in charge of the development. It would be nice if we tentacled out to sort of scrape other things and surfaced it within that one. Because really, you just want to search one place and if you're in an Esri ecosystem searching hours, but maybe you could find stuff from outside. So we do help, but maybe there's more we could do.

**Interviewer**

Do you have any suggestions or wishes of how other actors could or should do to improve the current open data ecosystem?

**Interviewee**

Well, this is a historical rant. Data.gov.[country A] was really good when it started because there wasn't a huge amount of data in it. And then people realized that they could also just put metadata into that search engine, and then the reason they weren't connecting it to data was that there were some license restrictions on that data. So it was open, but only if you were a member of a certain club. And suddenly, you couldn't find the data for the metadata. So you would find exactly what you want and then find that you didn't have a license to use it. This isn't open.

So I think the Living Atlas works because anything that you find in it is accessible, and I think if you search for something you should be able to get that data instantly. You should also be able to explore the data before having a direct hook to it or downloading it. We also want to move away from downloading data. The people used to whenever something was made free the at the beginning of open data, they would go and grab it just in case tomorrow it wasn't free. And there are cases the famous case of the <redacted> data.gov portal lost its funding when the <redacted>, so the budgets didn't get allocated so they shut their data portal overnight. And people who had relied on that data

portal, lost access to the data. So people would start downloading everything. Then slowly, that data is out of date. But you forget that the data is updated on the 1st of May, it's the 1st of June you're using out-of-date data. If you can just connect directly to the data source so machine to machine and only use machine to machine requests. So you want Amsterdam, connect to that, get all of the building footprints from Amsterdam from the Council updated on the 1st of May. You go back into the project on the 1st of December, it's gonna get the stuff from the 1st of October for you automatically.

The other thing is making sure that where there is temporal change, the user has the ability to specify the time that they want. There's no point only having the most up-to-date data. We've had this conversation with data providers in the past where they go "hey, your data is out to date, delete it" and like "I want to keep it", "but it's out of date" and like, "yeah, but maybe somebody wants to look at 2012, 2013, 2014 and see how things are progressed", "Ohh, didn't think about that". So maintaining, you can do it really cleverly within one database if you have a change log. But you need to have that way back feature to be able to see temporal change.

So what would I change? Let's summarize this. Just the data. Get to the data. Make sure it's machine to machine so people don't download stuff. Make sure data doesn't go away. Make sure that you have the most up-to-date data through machine to machine, but allow users to specify if they want a certain time period, they want that snapshot. And then people should be happy. But critically, make it easy. Easy and intuitive to understand.

#### **Interviewer**

OK, my last question, what do you think about the emergence of open source software like QGIS and also like open database like OpenStreetMap. Do you think it would change the way that [Esri distributor in country A] does things?

#### **Interviewee**

Well, it's not emergent. It's definitely not emerging. QGIS has been around for 20 years, 15 years and it's really good. It's really powerful. OpenStreetMap, it's been around for 15 years as well. There are two different propositions. But let's deal with the data. OpenStreetMap is amazing. It doesn't erode the value of the national mapping agencies in any country. It enhances them. It's a different data set. It shows different information. It also has some faults. The tags are not consistent. And the quality of information varies spatially. If you are in a middle class area, lot of middle-aged white dudes, then the data tended to be a lot better. Projects like HOT OSM have really helped that, but again, sometimes it's only ambulance chasing that we do within HOT OSM. It's like looking for areas that have had earthquakes or volcanoes or there areas of conflict and then updating them. But I know that if I want to look at building footprints in <redacted> or in <redacted>, the <redacted> remit is that they have to provide the same quality of information regardless of the geographical location within the [country A]. And your <redacted> has exactly the same remit. But I go to OpenStreetMap if we want to find out if it's a bank, a supermarket, a bike shop, a pub or whatever, it has extra information that I can use, and I can combine them. So I think that it just enhances. There's more data, different quality, it has different use purposes. The critical thing is understanding what you should and what you should not do with that data and what the limitations of that data set are. And there are limitations of the <redacted> data and of the <redacted> data as well.

Second part, open source software. Competition is good. Competition keeps you honest. Competition drives you forward. QGIS is an amazing piece of software. And the open source community is amazing. Esri builds its software and it uses open source software in the background. It uses some of the libraries like the ogr2org libraries. <redacted> who builds those. He's an amazing

person. And we use those transformations, data transformations. So we use open source software. But ours is a system. And that system, when one component is updated, it shouldn't break any other ones. When I worked for <redacted>, we had a fairly open source stack, we had some proprietary stuff, FME for data transformation coming in, but it was generally open source. And it was great, we didn't have to pay license fees, but we had some really knowledgeable engineers to keep that service alive. So we had a Live A and a Live B in two different machine halls in two different buildings or two different electricity networks in <redacted>. Although we find out in one occasion both of the went down at the same time, bad times. Really, like there was a power outage in one, in the week before there had been a fire in the machine hall, so we were down to one machine hall, and then there was a power cut. And then our funders asked why did your service go down? You're like, oh, my God, it was the perfect storm in one week, two areas of the <redacted> were screwed. And we had a Dev A and a Dev B. So we had four licenses to use Oracle software for our databases. It was crushing financially because we had to pay for four different implementations. And the open source community kind of solved that, but we needed a team of engineers to keep it up and running. I reckon that you could serve something that was very similar to what we had back in 2012, 2014, completely cloud hosted in ArcGIS Online and have somebody -- one person -- sort of maintaining it now rather than five or six engineers and software developers maintaining it. So maybe 10 people keeping this thing alive. We've moved.

Open source is great, but in that stack, if one thing updated, you couldn't implement that until you had done an entire test on the entire system to see if it would fall over. And in one case there is a piece of software which hadn't updated and it was the instability, so we then had to swap out this one component for something else that did the same job and didn't cause a problem in the rest of the system. So you couldn't update the rest of the system until you had found this missing piece. That shouldn't happen in our software. It should all be stress tested and compatibility tested. So there are differences in it. But I think open source software does keep us honest. I have QGIS on my laptop and I still use it. I am an Esri employee and I am not ashamed to say that. I think if you understand your competition then I can have an honest discussion with my customers about the pros and the cons of using Esri software or QGIS. But yeah, I don't know. I mean, some Esri people think like that and other wouldn't ever use QGIS software.

**Interviewer**

Yeah, that's actually the end of my questions. Before we end the conversation, do you have anything that you want to add that you want to share with me that you think it's important?

**Interviewee**

No, I mean really good questions. Nice conversation. And you know it's one sided because you're asking me questions. But data -- we don't have a need to use GIS if we don't have data. It's like the fuel that you put in your car, you need it. And if we encourage that open data narrative, then there's just more data. And if there's more data, there's more questions that can be answered. And you know, we have a lot of questions that need answered both on the local scale and the global scale. You can't avoid climate change. We've got conflict in Europe again. We've got distribution of resources, the wheat crisis with Ukraine which could impact Africa because everybody's grain prices went up and they would maybe, think about selling theirs for more money and then leaving them nationally with less. And all of these problems can be addressed with bits of GIS. The World Bank has a huge GIS team and you know there are thinking about those global problems and it all requires data. And if you think about.

**Interviewee**

Have you read Factfulness by Hans Rosling?

**Interviewer**

I haven't, but I've heard of it.

**Interviewee**

I'll send you an e-mail with it. He was a doctor and he used to give statistics on infant mortality versus GDP. And people would say where are you getting this data? And he's like: the World Health Organization website, this data is open and accessible. And if you, in government, are not using this data to inform policy, you are asleep at the wheel. So data is the key. You could say that the WHP weren't necessarily doing the best job surfacing that information for politicians. Politicians don't understand shape files, SGML, geodatabases, whatever. They might understand a map. They might be able to use a dashboard. The difference is between data and information. Data in its raw form largely useless until you do something to it to make it usable and useful. Information is usable. You can combine multiple sources to extract knowledge about the thing that you're interested in, and with knowledge you can implement change, but it starts with data. But it doesn't end with data, it's the start. So that's kind of like my parting thing. I'll send you the link to Factfulness because it's a good read on data and thinking about -- if the people in power have no access to data or no knowledge of the data reflecting modern society. What that are they actually doing with their policies and what goes into. If their view of the world is imparted onto them when they're at school and it's 40 years since they were in school, they are 40 years out of date and their perception of the world is 40 years out of date. And a hell of lot has happened, largely good in the last 40 years in terms of infant mortality, education, certainly the education of girls and equality. We're a lot better off although in the developed world, we seem to be becoming less equal. But it's a good read and it starts with data. At its fundamental is sort of open data and accessible data. But not just data, information. GIS is that interface between data and information and if we get it right, as an industry, we can enable and empower people to make better informed decisions.

**Interviewer**

OK, thank you. I will stop the recording now.