

Interview 27

| | |
|-------------|------------------------------|
| Interviewee | 29-User-C |
| Interviewer | Ashraf Shaharudin (TU Delft) |
| Date | 19 April 2023 |

Interviewer

Let's go with the first question. Could you please describe your role at <redacted>?

Interviewee

My current role?

Interviewer

Yes.

Interviewee

So, my role is quite complex. I have <redacted> roles and <redacted>.

Interviewer

How long have you been working in this or similar role that you have mentioned?

Interviewee

So the <redacted>.

Interviewer

Can you mention a bit what's the function of the <redacted> center? What is the objective of the center?

Interviewee

The objective of the center is to enhance the use of data and also data related technologies, so related to basically to artificial intelligence, but not only. Also big data, so, also for example building the data infrastructure but also good data management within <redacted>. So basically, we are not let's say, like <redacted>; we do not have a specific computer science program. We have shared groups but not education programs. But we see that data and the use of data is really increasing in all domains of various disciplines, so food, health, environmental sciences. We use more and more data. So the idea was to enhance the use of data within those domains. Very shortly, this competence center has five programs and I'm managing one program which is data science education. It's on me to develop and implement strategy how to include more data science education in the university, which is actually not primary business on data science and AI, but it's increasing. So we have more and more people and professors and teachers teaching about it and also students who are interested in those topics.

Interviewer

OK, that's nice.

Interviewee

So just to finish, one example activity that I'm busy with now is setting up a <redacted>, which will focus on data science and or artificial intelligence for <redacted>. So applied in <redacted>, but it's a data science program.

Interviewer

So it's gonna be a specialization or it's a program?

Interviewee

No, this will be a <redacted>.

Interviewer

OK, that's nice.

Interviewee

We already have a <redacted>, so they have been developed in the past years, but now we are just we want an additional <redacted>.

Interviewer

OK, alright. How does your role and your professional background relate to open data?

Interviewee

I did <redacted> in the topic of spatial data infrastructures, especially the assessment of spatial data infrastructures, and I remember at that time <redacted>, most of the data were not open yet. And then in the meantime, they became open, so also INSPIRE directive, well, those kind of initiatives. So that's also when my interest went a little bit to open data, but it was not a primary interest. That's one.

I was also involved in a project <redacted>. It was still at the time when I was working at <redacted> and at that time also <redacted>, he was involved in open data project and then they started experimenting with open data. And there was a project to measure the impact of open data and what we did, we took the <redacted>, and we basically measured, how this data set is being used? What is the difference actually in the usage of the this data before and after it became open in 2012? We had some data from before 2012 and we collected data after 2012, so it was 2013, 2014, 2015 mainly with organizations using these datasets, so for example, <redacted> but also some private companies and also Esri.

Interviewer

Around what year was this project?

Interviewee

It was 2013, 2014 and 2015, some somewhere there.

Interviewer

OK, alright. Now I want to move more on the topic of open data. What do you think is the value of open data to society?

Interviewee

I think the value is actually present in different -- well, I think it's quite complex here because the value -- I mean the problem with value is that it's difficult to measure this value. That's actually what we see. But I think the value is there. So in monetary term, it's difficult to express but the value is definitely present. In the sense that, OK, I think it's provides much more transparency to the citizens about, OK, what government is doing, what data they are collecting. The transparency, I think that's one.

Another value is, I think, related to research and that research can make use of this data and we see it here in the university, for example, we had <redacted> before the data were open, for example, that the <redacted>, and we as a <redacted> had to pay quite a lot for the use of this data. So basically, <redacted> were working on a small study area because you paid per area. Nowadays, the data is completely open, so you can study much more much bigger area. I think the research has

improved in the sense that more <redacted> can use more data, but also they can do studies on much bigger areas described by those open data because it's free. You can study basically countries or Europe or whatever. So I think that's another value, for research and education.

And also for commercial use. I'm completely supporting the idea that private companies use open data for their commercial purposes. We also have quite a lot of discussion with <redacted> but also within universities and also within the governmental organizations: Is it OK that the companies take free data from governments and then they make money on top of it? Is it OK? I think it's OK. I think that's correct. There's nothing wrong with that because they apparently they add value to this as open data intermediary. There is a value. So the customers, they buy their products with this data. So I think that's completely OK.

Well, that's a little bit my -- societal, transparency. For companies, more possibilities to make business, let's say to make more products. And for research education.

Interviewer

I know you mentioned that perhaps in your current role, you don't use open data that much, but I wonder if you have any challenges or grievances with regard to open data now that perhaps your <redacted> are facing or you yourself have experienced in the past?

Interviewee

So, my critical look at what's going on with open data is that we are still in the very early phase of opening data to the society. And that means that many organizations, they just open raw data. They just put it on the website in a non-user friendly format, I would say, or in a format which is only understandable to professionals, to people with high IT knowledge and that's it. And then they think, we [already] make our data open and technically it's true but practically to reach the higher goals of open data – as I said, transparency, the higher goals to the society, to education, to research -- I think we need to do a little bit more. So, there is I think still a gap that a lot of data is open but it's still difficult to use and if the data are not easy or not user friendly or the way how they are exposed is not user friendly, I think the usage is still limited. It can be better. So I think that's my critique.

Interviewer

OK, now, I'm moving on to the next topic, which is open data ecosystem, which is defined as network of interdependent yet self-interested open data actors. What is your perception of the health or sustainability of the current open data ecosystem? Do you think that they are sustainable now?

Interviewee

I think if you look at the especially on the geodata in [country C], I think the community understands and also profit from open data already, and I think it's a kind of -- I think people within the geoinformation science take it already for granted that data within [country C], geodata, are by definition open except data that contains some privacy sensitive information. So I think there is a kind of critical mass nowadays within the geoinformation science community and the ecosystem, which actually makes it kind of something that is granted already that data are open. In that sense, I think it's sustainable because I think it would be difficult to change it. If now a minister would come and say, OK, you need to pay for data or we are now against open data, I think the critical mass is too big to change it, so I think in that way it is sustainable.

Interviewer

OK. In terms of the way that different open data actors interact and influence each other, for example, how whatever data providers or the intermediaries or the users are doing with open data.

Do you think that we are heading towards a more reuse of open data, for example, towards a more inclusive open data ecosystem? Do you think we are in a way moving towards a better direction or we just gonna continue to be in the state that we are now?

Interviewee

I think we are definitely moving into this [i.e., better reuse of open data?] direction, the question is how fast are we moving. If I look at the situation in 2011 and now I think it's much, much, much better. I also do every 2 times per year, a kind of exercise with <redacted> where they have to search for open data on different topics and I see the huge difference how easy it is now to find and use open data. So I think we are moving into this direction.

But I'm a little bit worried if we can really reach the whole society with this open data, because what I said earlier, lots of open data is being actually made available as open data, but in a very raw format which is not usable for most of the users. And to make it usable for those users you need those intermediaries. For example, what Esri does, I think it's great, but what they do is basically limited to the users of ArcGIS platform. And so if you are a user of ArcGIS platform, you have a full access to user friendly data which basically can reach the society. But that's what Esri is doing within the Esri ecosystem and only within the Esri platform. And of course, for example, if we can say what we see is that Esri platform is being used by government and so government basically is reusing this data in a user friendly format to create, for example, viewers, which are open to the society, then in indirect way this open data are reaching basically the whole society. But Esri is only one company doing this and it is only geographic data. I'm wondering if this kind of things also happen with non-geographic data and how the situation is there. I don't know. I'm not sure about it.

I think there is there is a need for this intermediaries and the question is how many of those intermediaries are there and are they also going to now, of course, if they will reuse the open data and provide services on top of this, they will ask for money and Esri does it indirectly because that's how the clients pay for the Esri platform and with this they also get this data, but they pay indirectly for this data. And as far as its platform is used by the governmental organizations and the data is well used by the society, that's OK. But there is role of intermediaries and I'm just thinking how open can they be to the society with the services and to what extent they need to ask for money to provide those services on open data. I don't know if I'm clear about it.

Interviewer

Yes, yes, that's a good point. Do you think that the ecosystem now is relying too much on government as the sole data provider, meaning that other organizations like businesses are not yet releasing a lot of open data? So do you think that the ecosystem right now is still quite unidirectional instead of more circular ecosystem, do you think?

Interviewee

Yeah. Yeah, I think it is like this. I think that's definitely the case that it's mainly the government providing open data. And I don't see actually -- I think we would be too optimistic if you would wait that private companies will also provide their data as open data in the same way as government does it, because I think that they do not immediately have this incentives. So they have costs of providing their data to the society, the monetary cost to provide it, but also nonmonetary costs or risks. Yeah, that they say what will happen with our data. So I think the incentives are not very big for them and the benefits, if you are only providing your own data, the benefit is somewhere else. And the government has much more incentives and also not only the incentives but how to call it?

Interviewer

Responsibility?

Interviewee

Yeah, they also have these mainly the top down incentives. So the minister in 2012 the Minister had to say OK, let's make the data open because we need to be more transparent, we need to provide this data also to the company so they will generate income and then also taxes. And also the clear incentive from the society is that whatever the government does, so also collection of data, it's already paid by the tax money. So there's also this kind of basic incentive that OK, it's already paid, so I have to give it back to the society. But companies do not have those incentives. So I don't believe that companies will ever provide their data as open data to the same extent. I think that's maybe utopian.

Interviewer

And what do you think about feedback loop, meaning that government provides data and then intermediaries or end-users that use this data and then they realize they are mistakes, they are things that should be corrected. Do you think there exists a lot of like, this feedback process telling the data provider back that you have to correct this data? Do you think the ecosystem has that kind of mechanism right now?

Interviewee

It's difficult for me to, let's say, to generalize because the question is on which level. So is it on the country level or in general or it's about my feeling?

Interviewer

Perhaps in the domain of geospatial information, perhaps that small ecosystem of geodata.

Interviewee

OK. So, definitely I would say this feedback loop works very well. At least from what I know, the cases which are the closest to me is actually the loop of – OK, what Esri does with governmental data, mainly the basic registers. Most of the basic registers are provided by <redacted> and I know also <redacted>, <redacted> and also one person involved in <redacted>, which I'm working on, he's also working [there], and he's the product owner of the <redacted>, improve the map application. So what I see is that <redacted> as a governmental organization, they are also providing tools to receive feedback from open data users. So they are actively asking for feedback in a way that they develop platforms in which feedback can be given. I think that feedback loop is working perfectly.

And also from my experience from Esri, I also know that Esri content team, I don't know if it's called still like this, but people who are busy with open data there, they also have very close contact lines with, for example <redacted>, but also other governmental organizations providing data. They can report any mistakes or also ideas about how to improve the exposure of open data to the ministries and ministries are very happy that Esri actually actively reusing their data and providing the data further to the society. I think that works fine in my view.

Interviewer

Alright, I'm now moving on to the next topic, which is open data intermediaries. You already mentioned quite a bit about the role of open data intermediaries, so I'm gonna skip to the next question. How do you think open data intermediaries can play a better role in the open data ecosystem? You mentioned that perhaps in the geospatial ecosystem you can see that Esri is playing a prominent role and in others you don't know if they are Esri in different ecosystems. But apart

from the number of intermediaries, do you think that Esri can do better, for example, or are there intermediaries that can do different ways to contribute to the ecosystem?

Interviewee

So I don't know if Esri, for example can do better or not because that's a little more judgmental. I don't want to -- well, it's their own ideas how they will go on with this open data. But what they do now is basically they have this, let's say the software approach to open data that they say, OK, the main core business for Esri is software, so they are actually focusing on selling software to different organizations. So the software can solve some problems, societal problems. But of course the software needs data and they basically put this data for free to the software users and they invest quite a lot in that directly. But indirectly, probably this software is much better because of this ready to use data within the software. So that's what Esri does. With open data, they basically do not ask for extra money. I think that's maybe changing now a bit, I read somewhere that they also can provide some extra services on data, but that's paid, which is also logical because they invest the money in that.

Also other companies, I think they can provide services on this data and so that's maybe let's say space where improvement can be seen I think. But I think it really depends on the business model of the company. The business model of Esri is software basically, not open data and just putting investing in putting open data in their software, it improves actually the chances that the software will better fit the demand from the client. Other companies, they might focus on, let's say, making business on top of open data. And I think that's also perfectly fine. If they can create value and ask revenues from this added value, that's perfectly fine.

Interviewer

OK. I'm gonna move to the next section which is on [Esri distributor in country C] specifically as an open data intermediary. I think you mentioned that they do play a role in enhancing access to open data, but do you think that they also play a role in connecting other actors in the open data ecosystem?

Interviewee

What do you mean with connecting?

Interviewer

For example, do they facilitate the end users communication with data providers? Or do they facilitate the communication between end users with another end users?

Interviewee

I think yes, they do it. Maybe not explicit. Maybe implicit. They do it not explicitly because for example if they deliver software, which contains lots of open data and the software comes from one Esri, but it goes to very broad number of organizations and all organizations using the same software. They use the same data. I think in that way they connect very different end users of open data. In that way I think that's happening.

And I'm not sure if they are playing really a role in open data ecosystem as an organization which would for example, organize the conference on open data and how to use open data. I think that, they don't, except that during the Esri conferences they have one session about content. So just to show what kind of data, what kind of content do they share but they do connect other actors within the ecosystem. I think they do it but it's implicitly, so it's not an explicit task of Esri to connect the actors, but implicitly by having large client group using ArcGIS platform with open data. I think people are connected and especially, maybe also the connection happens in ArcGIS Online. ArcGIS

Online has also this Living Atlas and content. I think that's actually the platform where all open data that is coming via the Esri platform comes together. In that sense, different actors can use or search for open data there. And this also actually what happens with <redacted>. So for example, in different <redacted> that I'm participating in, we ask <redacted> to search for data sets in official SDIs, so <redacted>. But I also asked <redacted> sometimes to look at the ArcGIS Online platform, to the Living Atlas and also search for specific data sets that they need for different projects. I think Esri, with this Living Atlas and ArcGIS Online platform is becoming also a kind of open data portal but mainly for ArcGIS users. But that's where they are connected basically into one shop center.

Interviewer

Do you or <redacted> use mostly ArcGIS or QGIS?

Interviewee

Mostly ArcGIS.

Interviewer

OK. And do you have <redacted> that perhaps use QGIS and they say that there certain features are better than ArcGIS?

Interviewee

No, I don't see it. I think it's just different and I think we are also, as a university, we also have to give <redacted> a freedom to choose software. Some of them use QGIS. Also, in some courses, we use other software than Esri. And also from Google Earth engine. So there is also a large repository of global data in Google engine which is also, by the way an open data set. But most of <redacted>, they use Esri because we have a campus license. Basically we teach GIS concepts and not software development. So for us Esri provides a good platform to teach students the GIS concept.

Interviewer

And so do you have any grievances with using ArcGIS platform and open data related services provided by Esri?

Interviewee

So you mean maybe some more critical look?

Interviewer

Yes.

Interviewee

Well, to be honest, I'm not sure now. Sometimes I'm a little bit lost in what they offer. So you have [Esri distributor in country C], they have this content and I think it's advertised on the website in a quite nice way. And then if you use ArcGIS Online especially, then you have this Living Atlas. And for me as a user, I'm not really sure who is providing content to this Living Atlas, and if it's also content coming from the <redacted> or is it global one?

There was also another website from Esri content from Esri Netherland so sometimes it's a little bit unclear, what kind of data from which sources are sitting there? And this is quite important because the datasets which are coming from the content team of Esri, those data sets are considered as trusted data. So of course it's an open data coming from the government, but it's authorized by the content team. So it can be trusted, it's documented, it's fine. But then you have this Living Atlas, and then there is even more data and then it's quite difficult to choose. OK, but which data I can trust? Because basically also other users of Esri platform can also put their own data into the content of the

ArcGIS platform. For some users it's quite important to be a little bit sure about the quality of the data. So that's a little bit sometimes a little bit too much to choose from.

Interviewer

Do <redacted> ever come to you and say I couldn't find the data that I want in the Living Atlas and I have to look for in other sources? Do they ever come to you and say that?

Interviewee

Yes. Definitely, definitely. And sometimes it's about a very specific data sets that they are using or they are searching for. And especially this question, yeah, so they do a project that they need very specific data set from a specific region of the world, from some specific country. And then what happens is that they find something on a ArcGIS Online provided by someone and then they asked me, OK, can we use it? Can we trust it because we found it in ArcGIS platform. So it seems that it's authorized, it's OK, it's a nice source, a good source, good authorized data. But sometimes they also find datasets which are quite doubtful if the data has anyone checked. There is no documentation.

So the thing is that there is a platform provided by Esri, but there is no control on what's there except for data that is provided by the content team. For example, the content team from [Esri distributor in country C] is quite well organized, but you have also other resellers of Esri in different countries and if you search for that, they are not so well organized. So you'll also read, OK, it's data set provided by, for example, Esri <redacted> or Esri <redacted> or Esri <redacted>, but you look at the data and it's not as good as good quality as the data from coming from the another resellers like the [country C] one.

Interviewer

So in a way, even though we are in [country C], but sometimes we are doing analysis of other countries. So in a way then even [Esri distributor in country C] give a lot of data, but then it's probably not gonna be relevant to us because we need other data from other countries, right?.

Interviewee

Yeah, exactly. Very often they do some small project or research on other countries. If we only study [country C], then it's OK. But if you go a little bit outside, then it's becoming a little bit mess.

Interviewer

That's a very interesting point because, yeah.

Interviewee

Which is also, I think difficult for Esri to make this choice because they could say, for example, they could say, we make a second <redacted>. So <redacted> is completely controlled by <redacted>, by the government and everything there is controlled. But it costs a lot of money to actually provide this control and to guarantee the quality of this data there. And it's also closed to other data providers, so not everyone can put data there. Esri is providing actually an open ecosystem for open data and so all users of ArcGIS platform can provide and share their data there. But the end result is that you end up with a good data and data that is a little bit -- sometimes you see that specific dataset is of a very poor quality, but some dataset is somewhere in between, so you don't know if it's curated or not. Or partly. That's a bit problematic. I don't have a solution for that, but that's observation.

Interviewer

Other than the data quality do you see any negative or less than ideal impacts of open data services that Esri provided?

Interviewee

Again, what's your question?

Interviewer

Perhaps do you see any negative or less than ideal impacts of Esri on other actors in the open data ecosystem?

Interviewee

No, no, I do not see that that Esri has negative impact on others. I think it's more positive. It's an example of how you can play this intermediary role within the open data ecosystem. So I do not immediately see the negative impacts. Yeah, maybe a little bit in a terms of exclusion. So, you only have access to a good quality open data and services on this open data if you are a user of Esri platform, which is OK. But look at students -- sometimes because the students -- [while] you are a student, you have access to a license of Esri via your university, and then after the student is finishing the university, this access becomes limited, unless you work for a government which also has a license. It is completely logical. But the thing is that people who, for some reasons can be financial reasons, but maybe it's also not only people, but maybe nations do not have access to Esri platform. They also do not have access to good quality open data services. They still have access to this raw data but not all people can use this data, have knowledge to use this data. So they need this intermediaries, to better understand the datasets.

This exclusion, it can happen that OK, that if you don't have access to the platform which is paid platform, you are basically excluded from this access. We might think, is it OK or not? I mean, that's how the world works. So as Esri ask, of course, for money, for the value of their platform. But some people cannot pay for it and then they are excluded from this excellent services.

Interviewer

Do you have any suggestions or ideas of how Esri perhaps can alter their business model to be more inclusive? Or do you think that other actors actually should come in and provide the gap? I mean, fill in the gap.

Interviewee

I don't know. I think I don't have a solution immediately because that's actually how the system works. So they are companies and of course they need to earn money. They also provide taxes with this money, so that's positive.

They are actually reusing open data, in providing open data in such a way that is very understandable to the society. But this information, easy to digest information, is not accessible to all people. So some parts of the society, of nations, are excluded from this access. And how Esri can improve it? I don't know, to be honest. I don't know. Maybe to provide some part of Living Atlas, which is completely open to everyone as a kind of showcase. So you can have two instances of ArcGIS Online Living Atlas that you need to log in, so you need to have an account. I could think about something that they also provide the kind of Living Atlas with some key data sets, which are important for big societal questions and challenges like climate change just to showcase, OK, we also have this data as Esri and that's also what we can share. Maybe something like this? I'm not sure about it.

Interviewer

OK. So that's the end of my questions.