Interview 26

Interviewee	28-User-C
Interviewer	Ashraf Shaharudin (TU Delft)
Date	26 April 2023

Interviewer

My first question to you, could you please describe your role at <redacted>?

Interviewee

In <redacted> it's called <redacted>, it's more like <redacted> actually, for GIS cluster at <redacted>. We have covered, well, as far as we know, quite a strange construction within <redacted>. We have with the central <redacted>, for instance, the GIS applications stuff and also <inaudible>. So, they have a part of the main research. And we are just the GIS cluster for the whole <redacted> instead of all <redacted> have got geospatial crew somewhere, we are a central geospatial crew.

Interviewer

So it's like a <redacted> for <redacted>?

Interviewee

Support service, sometimes it's talking with them how to use these applications or the whole architecture that they have. I think that's 60% to 40% is mainly, where can we find our data, where is it, where should it be, and the structure of it? That's the other half. That's kind of my job.

Interviewer

OK. And how long have you been working on this job or a similar role?

Interviewee

Well, this job for a year. Before that, I worked in a similar role at <redacted>. Partly the same role, and partly I was also a software developer for GIS services, so that was kind of busy. Before that I worked as an information analyst at <redacted>, which you probably know. And before that, I was a data manager. Well, now, in about <redacted> years now that I'm in similar kind of roles.

Interviewer

Yeah. And in your current role, how is it related to open data?

Interviewee

That's a very good question. My current role is mainly we are giving people advice on how to use these open data sources. But there are a lot of data sources. I've made a small inventory of all the services that were around for all the top 10,000 online servers and I think there were 1 ½ million different services

actually were available. So to give you an idea there can be a lot. And it's mainly how we use these open data sources and it's how to use them in their data collection to *average* <?> their data instead of just using it and you're saying, well, this must be this or this must be this. Especially because the [country C] system, it's quite a lot.

Interviewer

So in general, you compile all this data and then you help researchers in using this data?

Interviewee

Yeah, well, sometimes you don't even have to compile it, it's just in the server somewhere. You just have to tell them where it is and how to use it and how the definitions are made because sometimes the definitions are not very clear especially in the <redacted>. Other registrations have the <inaudible> registrations, and then you've got the <inaudible>, which are mainly the <inaudible> actually. Once you know how that works, then you can use a lot of data, but it's not available for everything everywhere.

Interviewer

What do you think is the value of open data to society in general?

Interviewee

Depends. What's open data?

Interviewer

Data that can be used by anyone at anytime, from anywhere.

Interviewee

Because now a lot of organizations actually don't have their data published. They mainly have apps or services or somewhere which actually is not usable data. So in fact yes, it's very useful if you know what you're doing.

Well, with all due respect, I hear a lot of people who want to use the data but don't know how to use the data because they've never used it. And then they get annoyed by the fact that the data is not good. Mainly I'm then thinking, well, OK, thank you for your opinion, but is your data not good or is it not good for the research that you're doing because you expected it to be something different than it is? This is mainly what's happening.

Interviewer

Do you have any grievances with regard to using open data in general?

Interviewee

No, no, no. But everybody should use it, I guess. It's available. You should know the value of it and that's where sometimes goes wrong. People just buy this data, so it's good, but that's not the case.

Interviewer

Do you face any challenges with using open data because you also mentioned that you have to compile this data in the server for example?

Interviewee

Well, yes, of course. The [country C] data for instance, you've got a lot of people who have open data initiatives, open software initiatives. They just sit around and just putting all kinds of formats. For example, <redacted> have five different formats for six different <redacted>. That doesn't work. If you ask me, then. Then you should just use one system. If you've got a system, make it a working system, actually so people can use it as data.

Because with all due respect, if you put a GML somewhere on the website and you say, well, good luck with the people, there are not many people who are going to use that knowing what they're doing. They need every year, their own ETL systems to extract the data from these GMLS to load it to a format that they know and they can push out. But they've got a complete system of registrations, actually.

Interviewer

Yeah. So you think it's a mess right now because there's so many different systems?

Interviewee

Maybe that's a little bit strong, but I think some organizations should be more strict in how they actually publish the data because they don't, it is actually publishing rubbish.

Interviewer

Yeah. OK. So what is your perception with regard to other actors in the open data ecosystem? For example, you mentioned some data users don't know how to use data. You mentioned also data providers should publish data in a standard format. What are the issues with other actors that you think can be improved?

Interviewee

Well, that's a very good question. I think the organizations behind the publishing have to change actually. With all due respect, it is now just being published by people who don't know what they're doing and pushing [i.e., promoting] it as being the best of the best. For instance, there are three systems to have all the water organizations with the models not cooperating with each other, using the same data model but in a completely different way, as a base registration within their organization. Every <redacted>, for instance, has their own registration. And some things are comparable and some things are not. And then they just compile that to a central data system. Actually they thought they were compiling it to a central data system and mapping all the data to some kind of the right place and then they publish it as a geo package only. The organization that's behind the publishing of the data, <redacted> <?>, wants extra money to get it published as WFS, so they only put a lot of extract every week or month on the website, saying good luck with it.

Well, the organization is actually publishing it, for instance, has their own ArcGIS online service place and they could just publish the service there, but they're saying they can't because the person who is behind the contract of ArcGIS online is only available for one of the other products and not product <inaudible> which actually want to publish.

I personally disagree on that. I don't think that's true, but I think they're not the right people actually on board to publish these data sets in a correct manner. You also see that some of these vector datasets, they've got a lot of attributes. No one knows what these attributes are. Sometimes they have the old SQL database name. And the explanation behind all these data models, which are huge, are not completely published or they're only published in a PDF somewhere on the website. But that also doesn't help in using the data if you want to use it. That's mainly on the publisher side.

On the user side, I guess ground awareness of what data is and how it can be used that should be between the ears. I don't know the correct <redacted> sayings. But you know what I'm saying? It should be, the data is good, be critical with what you use. Sometimes it's usable, sometimes it isn't, and it's not published for your research. I guess people should be aware of that, that this is published for a certain reason, which is not the reason why you're looking at the data. Of course you can connect them to some extent so that they are on the same location. So for this location, there could be a connector in a way. But you can't really claim quality from publishers in that way because the reason why people release these things are not the reasons why you want to use it somewhere else.

Interviewer

What do you think is the role of open data intermediaries? So these are actors that enhance the access and use of open data. You mentioned <redacted> just now. So for example <redacted> is, I would consider one of the intermediaries. What do you think their role and if they are playing their role properly right now?

Interviewee

Well, personally I wouldn't know what <redacted> does on this? What's the <redacted> of <redacted>?

Interviewer

I'm not sure.

Interviewee

Make that the same question for <redacted> of <redacted>. So are they doing their role correctly? I think they're very focused on publishing it for the country themselves. But do people in the country actually know what's in them? They should know a little bit when they're working with data. Especially when you're crossing a border, then it's gone, the knowledge.

Interviewer

So what about intermediaries like Esri? Because Esri also publish open data on their platform.

Interviewee

True. But then which language does as Esri publish?

Interviewer

I'm not sure. I guess <redacted> for <redacted> and <redacted> for <redacted>.

Interviewee

Yeah, well, this is a big problem because it can be in <redacted>, it can be in <redacted>, in <redacted>, it could be <redacted>. I was looking through 1 and 1/2 million services and every language is coming by and you're looking at, OK, I've got no clue what I'm looking at. So I need some kind of a Google Translate app to actually know what's in there. And to know what the texts are, because there are no standard tags and there are no standard summaries. You can't just compare the datasets with each other. But what you see is that people tend to use when they're working internationally, mainly tend to use these satellite image data because that's general for the whole region. If you look at something like the <redacted>, which is quite a specific [country C] height model, you don't see them using it for instance, just around the border of <redacted> or of <redacted> because they actually couldn't make use of it. You see the border effects in that, especially in the language used for the services.

Interviewer

That's a very interesting point. So you're saying that it's important for intermediaries not only to ensure machine readable, but also human readable, meaning that human in different places can interpret the data?

Interviewee

You can make it machine readable and then make it accessible, for instance, for the website which usually somehow translates because it's same for services for instance. You've got metadata. You've got tags. In these text also, I guess it would be easier to have some standard tags for some data sets, because with all due respect, not everybody knows what tags are. You could I guess introduce a way of basic tags, for instance, the Cadastral registration or something to do with the <inaudible> that you can actually compare these datasets, so you can actually get them in using the right tags. A basic list of tags that are available, kind of standardized from some kind of an ISO structure, because I know that there are some of these ISO structures. I know that they are not always as usable but you can use them as background and then introduce more additional standards to make the data a little bit more accessible than it is.

Interviewer

What do you think intermediaries like <redacted> and Esri can do that they haven't done it yet?

Interviewee

Exactly that.

Interviewer

OK. But apart from like processing this data, do they have other roles? Do you think do they have? For example, maybe they can connect different stakeholders open data?

Interviewee

I don't think there's structure behind it to actually answer the questions that are needed. I think they have to think how to publish that data. Maybe Esri have no idea about Esri Switzerland to say something. How many people are in there? But I guess the easiest way to actually deal with that is look at which services are actually your main service that you know of. All of them or some of them, especially in [country C], they've got some kind of the concept group. Let them look at the way that they're using the tags, translating the tags. For instance, to make the text translatable for a translator machine actually. Do you do you understand what I'm saying?

Interviewer

I don't know what translator machine is.

Interviewee

Like you can hear <redacted>, you can use Python scripts to actually get to.

Interviewer

Yeah, like Google Translate kind of machine?

Interviewee

Yeah. Actually using Google Translate to translate the text just fine, unless you're <redacted>, but for the rest of the world is fine. You can then just translate it and make it available for people in their own language, because with all due respect, internationalisation is always well, everybody can speak English, but as you hear, my English is very specific for my region. There is always a discrepancies in how you interpret it in <redacted>. There was a quite a big discrepancy in the whole interpretation of metadata within the <redacted>, for example. I thought that they were making these data open available, but actually they were saying no, we wanted to have it close in every line and every way. Because they misread the English. Those things happen. They don't happen now, it was like ten years ago. Those things can happen because people just don't know how to look at it. Actually, what is good data and what is not good. Personally, I think it would be good for <redacted> to actually, to some extent, uplift the data model with the service. Because sometimes there are domains within these datasets, there are links, there are websites, whatever. But they all don't link with each other.

Interviewer

So you think they can actually make the data model transparent to people?

Interviewee

Yeah. I've done it within <redacted>, at the end it's actually creating a list of every field that was there with the domains that were hanging beneath it. So you can just actually make a website of everything that's in there, you can just click through it, you can just make an HTML of the list from the database and push it to everyone to use it.

Interviewer

So essentially you've done something that <redacted> should have done?

Interviewee

Yeah, well, we've done more than what <redacted> has done. We've actually created the complete system of a <redacted> within a database, all the interconnection should be there in one format so everybody can actually use it within the organization. Of course, they need some knowledge of how to use data, but that makes it a lot more accessible than just pushing it to websites. Because with all due respect, people working at bottom box are not used to extracting data from a website to help to see something.

Interviewer

But whatever you have done, is it available to everyone or people in <redacted> only?

Interviewee

That, what I was saying is actually only available to people in <redacted>. The list of all the services is going to be available for everybody in <redacted>, for now, to see whether it works and whether it helps people in getting used to finding data from <redacted> or from other websites. Trying to see whether there are tags to try to translate them so they can actually just push it into just download button or whatever. Haven't really thought about that one. But then you can just create a service actually to once -- it's kind of something called… there's one very big website in America that's never made it to anywhere else. Can't really find it now. They have a big website with all these raster for government organizations in America.

Interviewer

OK. But it's only for government or for the public?

Interviewee

Well, it's actually public data in all these online datasets. You can use the system within Esri, the whole search for data in the online environment thingy, then you get everything. Also you get everything that you can't read. So then you just think, OK, but which one is the right one and you just got a list of things. You need to know exactly how it's named, because otherwise you just won't get it.

Interviewer

Interesting. OK, since you mentioned ArcGIS, so how does ArcGIS help or support you in using open data?

Interviewee

Well, we personally don't. But I hope it helps the researchers finding data.

Interviewer

For example, the geoportal is <redacted> is based on ArcGIS Online?

Interviewee

Yeah.

Interviewer

Yeah. Do you think it's useful?

Interviewee

Yes, it is useful, but you also see people just mainly use it as publishing tool. The user updated it to actually just publish things like, for instance, story maps or whatever research. Some of them have, but most of them don't have the experience with data management to actually see what they're publishing. So there are a lot of surveys with Field 1, Field 2, Field 3, Field 4. Yeah, I would <inaudible> these people, but because there's a manager. They can be student researchers or whatever. And I normally tell them when they come by that they should fill in their metadata, that they should use tags, that they should do these things because otherwise it won't be findable. They are publishing blah blah blah general global ID number, whatever. I don't know what this is, so they won't know what this is.

Interviewer

Yeah. Do you think there is any other alternative to ArcGIS Online for this sort of geoportal?

Interviewee

Depends. If you're just publishing data, then I guess that geo server applications would be just fine, but there are some other extras. <Inaudible>. It would be nice if there were more, let's put it that way.

Interviewer

OK. Is there anything that you don't like about using Esri products, either ArcGIS or ArcGIS Online or their Living Atlas or anything?

Interviewee

Well, actually 2 main things. One thing is the fact that they always push beta products as end products. I was looking at something, well, this is not finished yet. Please just first finish your product and then publish it to me because I have to explain it to other people why it doesn't work and I don't like to explain to people why it doesn't work. Create something that works. So I can actually tell them what they can do with it. You see, for instance, with the whole experience building from Map Builder to Experience Builder, but there are a lot of functionalities are not there yet and then they're being built

somewhere in time. No, you are not pushing this product to me when it's not working correctly, I don't need that. Just like windows. I don't like when Windows 98 came out because Windows 98 was not finished yet and within a year the Windows XP. It was kind of a similar way. OK, Thank you very much for this product. Please first finish and then give it to me because I'm a paying customer. In the <redacted>, we're not a big payer, but in <redacted>, they were actually a big payer of the product. So yeah, you just don't want these half products. You want the full product that works completely, unless there is a good explanation of why some things are not there yet. That's one.

And of course, the incredibly difficult licensing system within Esri

Interviewer

Could you explain?

Interviewee

Yes, they've got a license for everything that you want to do as an analyst. All these different things are all licensed based on people that could or could not have these licenses. I'm always like OK, thank you very much, just give me a full on price and everything and I will choose what not to use because I don't feel like using defense mapping at all, but make that option available and I can just not use it instead of make it pay. Personally, I think it's not very handy for Esri as an organization because then, for instance, all the analysts are going to use GDL to actually do all this 3D analysis.

Interviewer

So if I understand you correctly, meaning that the licenses are not clear, so sometimes you don't know whether you can use it or not?

Interviewee

You know you cannot use them. That's the problem. The thing is that that you have to pay extra for every license and it's it seems very strange. I don't think this is the most handy way to actually get money from people.

Interviewer

And this includes data that is originally open data.

Interviewee

This can also be premium data, as they call it. Premium published things. Of course you could work from it, right? Make your products more expensive or whatever. But don't do it like this.

Interviewer

OK. Anything else that you don't like.

Interviewee

I think those are those are the main two critiques that I have. The licensing system and the and the way to push new applications too early.

Interviewer

And the premium data that they usually give for fee, is it built on open data? Do you know?

Interviewee

Sometimes it is.

Interviewer

Ah, OK, so they take open data and then they package it and they sell it?

Interviewee

Yeah. Google does the same, so yeah, and most of organization actually do that. For instance, the whole system that I was talking about, the <redacted>, they're publishing the cadastral areas, the place where buildings are standing up actually having a format out and they actually have a list of organizations that can reconstruct this to the right format for you on the website, instead of just publishing data in the correct manner.

So I don't want the list of software parties that could possibly do this or that. I've got two jobs, so this doesn't help, by the way. I'm still working <redacted>. Sometimes, this is not helping getting open data used.

Interviewer

I have a last question. Do you have any suggestions or wishes of how Esri should change their way of doing things or business model. You mentioned a bit, but do you have any other suggestions or wishes?

Interviewee

So I guess not on top of the things that I already said. No clue. But I suppose they can give organizations the <inaudible> of how things are organized. With all due respect, sometimes you just have to ask yourself, OK, who the hell invented this? This and this person there and they said that this would be a good idea. OK. Thank you very much for this. But has this person checked anything with anyone ever? No, no.

<inaudible>

Typical idea of how these open data models actually get to existence is: push everything then when you fill everything in then probably will fit and then the story will be clear. I don't know, but it is something like alphabets, you've got a lot of symbols but you don't have a book. Just random symbols.

I guess the main thing not to underestimate the work that needs to be done to get these data services correct.

Interviewer

So sometimes too much data is not necessarily good?

Interviewee

Yeah. Well, if it's all good data, then that's no problem. But if you don't know how to interpret it, then you need time to interpret every data. Then you need time to interpret all the data sets that are in there, and if you don't have the time to actually interpret it, then you could make very, very strange conclusions.

Interviewer

So before we end, do you have any other remarks or things that you haven't mentioned that you want to share?

Interviewee

No, probably, but when I when I have some more then I will just mail. Then I guess that's easy.

Interviewer

Yeah, yeah, for sure. I'm going to stop recording now.