

Interview 24

Interviewee	26-User-C
Interviewer	Ashraf Shaharudin (TU Delft)
Date	13 June 2023

Interviewer

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

Interviewee

Yes, I'm a <redacted>, I'm a <redacted>. I work at <redacted>. But my background is more in the <redacted>, <redacted>, in this <redacted>. And I'm doing <redacted> on how we can model or measure accessibility of people to certain places, for example <redacted>, <redacted> I'm part of <redacted> project and we are focusing on the built environment impacts on health.

Interviewer

And so do you use open data in your project?

Interviewee

Yes. Actually I'm trying to use as much open data and also open source software as I can. My project is a <redacted> project, so we're doing a lot of things for the <redacted> context, not just the <redacted> one. So I'm using a lot of OpenStreetMap for that, which is open data, and also, for example, <redacted> population statistics and also similar population statistics data from for example <redacted> and for <redacted>, we used at some point. Also the <redacted> data in [country C], the housing addresses. So things like that we try to use as much as we can, basically.

Interviewer

Yeah, and your project studies not only [country C], but also you mentioned <redacted>. And are there any other countries?

Interviewee

Yes. So I've run one study only for [country C], also we run a study for one city in [country C], for <redacted>. And for <redacted> and <redacted>. And now also working for <redacted>, and also <redacted>.

Interviewer

OK. And I think you also mentioned, not during this interview, but previously that you used ArcGIS software. Did you use it with open data as well?

Interviewee

A little bit, I think in the time when I used ArcGIS, it was while I was working at a company that did a lot of consultancy jobs for municipalities or water management companies, for example. But there I worked mostly with their own data. So the own municipal registers or the really more, like sensitive data almost, relating to the drinking water infrastructure for example. While in my studies and now in my research I do use a lot of open data, but then more with also like the open source, QGIS.

Interviewer

OK. And do you have any challenges or grievances or unmet needs in working with open data?

Interviewee

In general, right?

Interviewer

Yes.

Interviewee

Sometimes it's hard to know, to find it, to really find the way to download the data, because I really want not just to view the data, but to download it, to process it and create something new with it. So that is always a challenge actually, even with the data that I know and I've used before, it's still if you want to do it again, it remains a challenge often.

I've also worked a little bit with these population data from abroad and there are challenges also that all the documentation is usually in <redacted> for example, or in <redacted>, so there's another barrier that you can overcome, but it's even harder.

Interviewer

And once you download this data are they usually usable? Or do you have to then look for another data because you realize it's not actually usable for your purpose?

Interviewee

That's a good question because I'm now thinking of the data that I could actually use. I know sometimes the format is hard to -- it really need some preprocessing, so this I have done, for example, for the [country C] basic data sets. I've downloaded once and preprocessed once, and I'm using it ever since [and does] not [have] to go through that again, basically.

Interviewer

So a lot of processing that you have to do?

Interviewee

Yeah. Processing, exactly. Also, I'm using now a bit of satellite data, satellite imagery. So that also requires a lot of preprocessing because you can either view the view or process data online or you can

download the raw data, but it's hard to download exactly what you're looking for so you have to preprocess yourself.

Interviewer

Yeah, yeah. And are these data up to date?

Interviewee

Yes, I think for the data that I'm using, yes.

Interviewer

And they are mostly in machine readable format?

Interviewee

Yeah, exactly. For the ones that I used, of course, I cannot think of an example now where I couldn't. But I have for example, for the satellite image data that I'm working on now, which is also open data, but it was hard to obtain the actual machine readable data harder than for example finding a viewer for this data.

Interviewer

Harder in a sense that it's hard to look where to download in a machine readable format?

Interviewee

Yeah, yes.

Interviewer

Now I'm going to talk about open data ecosystem in general, which is defined as network of interdependent yet self-interested actors. So these are like data providers, data users, data intermediaries or data standards body. Of course they are interconnected, but they also have their own interest in participating in this ecosystem.

What is your perception of the health or sustainability of the current open data ecosystem? In other words, how do you see the future of open data in terms of its financial sustainability, in terms of its continuous impact in terms of the support from various sectors. Do you think that it's going in the right direction? Do you think that there are a lot of things that need to be resolved?

Interviewee

I think yeah, I'm using a lot of open data that comes from governments. So I think maybe the funding or the financial question is -- I mean it's a different question for -- so I'm using a lot of data from the [country C] government or from governments abroad or datasets from the European Commission, for example. And the other data that I use is OpenStreetMap for example. So this is more of a collaborative effort. So I don't really know how they would function financially. So yeah, that's hard for me to say.

I also think that most of the data that I use is [country C] data often or data that is available worldwide or European wide at least because I'm also doing my case studies in different European countries, and I'm trying to follow the same procedure for these different context. So that is another limitation that is sometimes really hard to find the like the same data or a comparable data source for different countries. I think it would also be hard to align all of that because they also lose some detail.

Interviewer

With open data that is sourced from public organizations like maybe from the [country C] government or from European Commission, for example, do you think that there will be continuous support in funding this open data. Your general perception, do you think that public sector at least is quite supportive of open data and will continue with open data policy?

Interviewee

Yeah, I think so. For the data that I'm using, so population data quite sure that they will keep it up to date every year. Also the ones abroad they were really regularly like yearly or every couple of years updated. European Commission data I'm using, this satellite imagery, I think that is also a continuous mission which I'm confident they will continue.

Interviewer

Is it from ESA?

Interviewee

Yes, Copernicus data. From that program. So, yeah. I think for the data that I'm using, I'm trying to use those data also because I want the recent data that is available in multiple countries, so I'm probably often ending up with the most reusable data sources? Because in the cases where it's only once collected, let's say once collected five years ago, or in one country, then I cannot use it often because of these the limitations of having to do it in multiple ways. Of course you encounter such data sets as well and those I cannot use.

Interviewer

And you mentioned about using data from OpenStreetMap. Do you face any issues with using data from OSM? For example, do you sometimes question the accuracy of those data or do you have any issues with regard to inconsistency of the data? Especially you are doing with multiple countries, yeah.

Interviewee

Yes, inconsistency, you see between cities, even if they are major European cities, you see differences definitely. Also sometimes you see that some tags are not used in one city, but only in the other. Because it's a different community, I think. And also you see differences in accuracies, how many tags are associated with specific objects basically. So this, you see. So that can also be hard sometimes because the workflow or the selections that you use for one city, maybe do not work with the other one. So you need to be a bit more general sometimes than you could be with the specific case.

And the other question you had was about?

Interviewer

The accuracy and the consistency.

Interviewee

Yes. Inconsistency, yes. And accuracy, I think it's often very good but yeah, it's not fully complete especially if you go into detailed filters, then you cannot get everything.

Interviewer

Yeah, but in general, you think that the data from OSM is quite reliable for you to do your research?

Interviewee

Yes. Yeah, I think so. Especially because I'm using major European cities and there, it really looks quite good. Yeah, yeah.

Interviewer

OK, so this is probably a good segue to the next topic which is on open data intermediaries because OSM can be one of the intermediaries because they facilitate the access and use of open data. So let's take OSM first as an example. Do you see that the future of OSM is quite sustainable in terms of, because it relies on crowdfunding as well as crowd sourcing data, so do you think that based on your experience and based on how you see data that is in OSM, whether the community is quite vibrant and you see that they're going to continue to flourish or do you think that at one point they're probably gonna die?

Interviewee

I hope they continue to flourish. I'm not really sure. I've never really contributed to the community. Only use the data. Maybe [contribute] once in my studies in the mapathon thing.

But I think so far, the quality is quite good and it increases still. And you see it being updated, also often when there are like construction work. So you see those changes also reflected later. So I think so far it's good. Because you get the completeness and the accuracy I think is still increasing, so I hope it will stay that good.

Interviewer

And so, among your colleagues, for example, do you often hear people using OpenStreetMap data?

Interviewee

Yes. So I have one close colleague that also works with spatial data. He also uses a lot of OpenStreetMap and also some other data sources, for example open data from the <redacted> and basic data. But also, I

think some other sources like Foursquare, things like that. But I'm not sure if it's really open data actually.

Interviewer

OK. And how do you think that perhaps OSM or other type of intermediaries can play a better role in the open data ecosystem?

Interviewee

A better role?

Interviewer

Maybe you can see in a way, like, what are the current shortcomings of pitfalls of OpenStreetMap that perhaps if they do it differently then they can actually be better.

Interviewee

Yeah. As I said earlier, I really always want to download the data so that I can process it myself and not just look at it. So I think their system is very nice, you can view the data, you can use the base map, for example, in a lot of products and this is all very easy or maybe not easy, but you can just look at the data. But to download it, I'm using their API, for example, and I think this is not accessible to everyone who would want to download, for example, all the OpenStreetMap buildings in <redacted>. So I think some way that makes it easier for people to get the data locally and before processing it, I think that would be useful.

Interviewer

So because in your work you retrieve the data from API, so once you download it, only then you can see them?

Interviewee

So I use an API to retrieve the data and then I store it in, usually a GeoJSON file thing or a shapefile, and store them locally. Yeah, but I'm also trying to make the data for my research open data again through like sharing it in <redacted>. Share it again.

Interviewer

And you were saying that the shortcomings are that people can't view the data or can't download the data first?

Interviewee

No, I think that you can have a very nice viewer for the data. So that's very useful. Also use that. But to download the data, I think, you need a bit more of technical skills.

Interviewer

OK. Now we're gonna talk about Esri as an open data intermediary, do you think that Esri plays a role in enhancing access and use of open data.

Interviewee

I think also for Esri, the Atlas, for example, like the viewers are very nice because there you have a lot of open data that you can view or see. But I think it's really hard then to -- or I I've never done it and never managed to download the data and to like have a local copy so that I can really like get my hands on it and process it again. So that's why sometimes I think I come across the Esri products, basically the online maps that they have, are a nice example to get a feeling of what is in the data. But then if you want to get the actual data, you still need to go to the source basically.

Interviewer

Yeah. So in a way, it only helps you to see what data exists and then you can go to the source?

Interviewee

Indeed. I think then often it's the [country C] data only, so it's also a bit harder to find equivalence in other countries.

Interviewer

Yeah. Is there any particular reason why when you see the data from ArcGIS Online, for example, instead of using it with ArcGIS platform with the ArcGIS software, you instead look to the source and get the data from the source?

Interviewee

Yeah, I think often because I want to calculate different things. I want to make changes to the data. And also maybe the another reason is that I'm using a lot of my data processing in Python code. Because for me that is easier to rerun my procedures compared to in GIS, both ArcGIS and QGIS because that's often more manual processing. Of course you can document the steps but it's not just like pushing one button to rerun again with the different parameters. So yeah, that's maybe also why I do not do the processing in GIS. Often I use it mostly to sometimes visualize or to explore like data quickly, check what is in this data set, and what kind of attributes there are.

Interviewer

OK. Because you find it a bit more trickier to use the ArcGIS platform to do all the processing instead of Python?

Interviewee

Yeah. Yeah, exactly.

Interviewer

OK. And do you think Esri plays a role in connecting other open data actors in the ecosystem?

Interviewee

Other actors, so other data providers for example?

Interviewer

Data providers with data users or among data providers or among data users themselves. Do you think Esri facilitate the interaction between these sectors?

Interviewee

I think maybe they sometimes facilitate between data providers and data users, for example, in one of the places where I worked at the time was a water company and they have a lot of Esri product system. And there we also used the ArcGIS Online to open up data, sometimes to the public, so everyone living in the city knowing where the water taps are, for example. And also sometimes more privately in the company, so restricted access with two colleagues from the different departments, because of the data being sometimes a bit sensitive. But then it would still be very nice to communicate with others because it's very easy to just open a website and look at the data, click on some things.

Interviewer

Yeah. OK. So you can use ArcGIS software to make data open, and also you can share it between in a closer environment?

Interviewee

Yeah, yeah. I think especially with people who do not -- were not able or do not want to download data and do programming and work with GIS. I think it's a very user friendly package, basically mostly the ArcGIS Online because it just works on your browser and it's quite intuitive.

Interviewer

And is this feature possible with open source software like QGIS?

Interviewee

I'm not fully sure. I think I heard that they have also an online version now, but I've never used it, so I'm not fully sure about that. I think someone once told me that. I think generally the ArcGIS software is often very good and stable, more than QGIS, which is, sometimes a bit more unstable or slower. So I can imagine that ArcGIS Online is still better, but I'm not fully sure.

Interviewer

And do you have any grievances about using open data related service from Esri?

Interviewee

So yeah, I think my main grievances or issue that I'm sometimes facing is that I cannot easily download the data so that even if I find something, for example, see something in their viewer, I need to go to the

actual data provider itself to download the raw data and do the processing that I want to do with it. So that's my main.

Interviewer

Yeah, because you can't immediately download it because they already process it in a certain way and you want to process it in a different way?

Interviewee

Yeah. Exactly. Yeah. I think often there's no real way to download it.

Interviewer

Yeah, unless you use the software?

Interviewee

Yeah, exactly. Unless you stay in ArcGIS platform. Yeah.

Interviewer

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

Interviewee

So negative impacts of Esri? Or potential?

Interviewer

Yes.

Interviewee

No, I think for open data I do not really see negative. Because I think their viewer is nice. Yeah. I don't know. Maybe potentially it could make governments more lazy in making something good themselves. Also to export their data. Because I think ideally you would want to have all the governmental data with nice viewer and downloadable, and all the documentation in one place, which is a bit harder I think if there are some parties in between. But I mean generally they do, I think, a good job with their viewers

But of course, their software is not open source, so this is quite expensive software that you can really only afford if as a big company usually. So that is kind of a barrier for maybe smaller companies or individuals to use this software or some bigger companies. It's just a lot of money. That's important.

Interviewer

Yeah. So even -- what about the data that they provide as open data in the Living Atlas where anyone without the software can use, do you think they are mostly usable?

Interviewee

Yeah, I think, mostly for just laymen people. I think it's a very nice way to see those maps and to have a feeling of what kind of data is out there and to look things up manually and stuff like that. I think that works really well.

Interviewer

Yeah, but mostly to just explore them and go to the data source to actually get the data?

Interviewee

Yeah, exactly. I think that is my main issue with how Esri shows the data or distributes the data.

Interviewer

And how does your experience in using QGIS differ from ArcGIS?

Interviewee

So when I use QGIS, it was really only the desktop software of QGIS, so I'm not sure if there is more, because as I said, I think someone once told me that they were also working on online stuff. What I like of QGIS is that it's open source. You can have any student download it and people play with it basically without the fee.

But it's I think a bit more unstable. So this is sometimes annoying, especially if you want to do bigger projects. So I do understand why companies choose ArcGIS, because it's more complete and well done I think. And I also think that the way in which [ArcGIS] have like the desktop versions and the online versions and also apps that you can build with their product is very nice. So especially for big companies, when there are a lot of people working with it, I totally get that they go for ArcGIS over QGIS, which is much more restrictive and more unstable.

Interviewer

Yeah, but for your purpose, you think that QGIS is good enough?

Interviewee

Yes. Yeah, mainly because I'm doing a lot of projects on my own or with my one or two colleagues. I'm not so much like trying -- I mean, we're sometimes considering also making data more publicly available and this could also be done better maybe in ArcGIS, because now we do that manually -- manually [as in] someone like the software developer goes building a web map application. Well, in ArcGIS, that can be very easy to just put some data online. So this is something that we are not doing yet or if we do it, we found a way around it. But it could be much easier with ArcGIS.

But apart from that, we are not really -- So for example in the previous job I had, we were also often making from the Geo department at the municipality or the water company, making data available to other colleagues in other departments, and that was really useful to do it with the ArcGIS Online mostly or sometimes also desktop, but mostly the online web. But now because I'm working more on my own, I do not need to do that per se.

Interviewer

Can you please briefly explain the process that you're using open data with QGIS. What do you need and how do you do?

Interviewee

What I often do is I really go to the source website from the <redacted> or the <redacted> to download the data, and then for example, load it into QGIS to mostly using QGIS to visualize the data and to also check my data, to explore if I can make changes, because you can really like zoom around and click on things to see if it's what you expected it to be, to validate if it did things correctly.

The processing, I'm doing a lot in Python, because that really allows me to automate it. I think you can also use Python in QGIS, also with ArcGIS actually. But it's a bit more inefficient in my experience, so I think it's easier for me to work with Python, with other like non GIS libraries.

Interviewer

So you don't use API when you use QGIS? So you don't need plugin?

Interviewee

Yeah, I do use plugins sometimes to load base maps for example. I think I have also sometimes explored plugins to download data like for OpenStreetMap, maybe also the <redacted> data from [country C]. But I think the plugins, I often or sometimes try them or explore them, but I do not really end up using them.

Interviewer

So typically you would download the data yourself and then bring it to QGIS?

Interviewee

Yeah, exactly. Yeah, yeah.

Interviewer

My last question, in an ideal world, do you have any suggestions or wishes of how Esri's way of doing things or business model should be?

Interviewee

So I think that would be very -- yeah, I think my main issue with Esri software it is a very expensive product sets, you get a lot, but it does cost a lot. So I think ideally, I, as a user, would like a free version or like an affordable version, or maybe a version that works for small companies or like smaller package of things, or more restricted use. Where you talk about the sustainability of the ecosystem that is of course not achievable.

Interviewer

OK, OK. That's all before we end the interview. Do you have anything that you haven't said that you think you want to say to share with me for my research?

Interviewee

I think maybe my last thing that maybe I've touched upon, but I'm using a lot of open data, but also tried then to use the open source software to process it, which Esri is not. So to me, it's like a mismatch. I'm trying to work with open data, but then if I use proprietary software, it's still not fully open and not reproducible, for example, for research purposes. Not for everyone. So this is a bit of my issue with ArcGIS.

Interviewer

And a lot of, like, especially <redacted> research projects, they do encourage using like, open data, open source.

Interviewee

Yeah. Yeah, exactly. We really tried to do that and also to open up our data again in the <redacted> repository. I think that it's really nice to use open data, but then also try to -- not always, because sometimes you need [to use] some proprietary software for specific step and that is just necessary -- but if it's not necessary, I try to use the open source. So that is the main reason I think why I do not use ArcGIS.

Interviewer

OK. Thank you so much.