

## Interview 25

Interviewee	27-User-D
Interviewer	Ashraf Shaharudin (TU Delft)
Date	28 June 2023

### Interviewer

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

### Interviewee

Yes, sure. So I'm a <redacted> in geoinformatics and I work with <redacted>, trying to find methods to identify <redacted>. And I also teach here. I teach like <redacted> that mainly also like spatial analysis, some remote sensing analysis or they are combination. So that's what I'm doing at the <redacted>.

### Interviewer

OK. Do you use open data in your work?

### Interviewee

Yes, so that was actually the main idea at the beginning. We realized that the types of green spaces are not all of them are in the existing land use and land cover maps. So we wanted to identify them and then create a green space map at the end. But we also want it to be -- the approach that we take with the workflows that we create -- to be reproducible so that other people also use it. In this sense, it's quite important that we use of course open data and that's why we actually try a lot to use only open data but that's not always possible, of course. But yeah, so using of open data is actually like crucial part of my work, my research work.

### Interviewer

So you take open data and then you combine it with other source of data and then you also intend to publish the data that you have value added as an open data?

### Interviewee

No, not really. So the workflow I developed to map certain type of green space, I tried to utilize only whatever is available, like whatever data sources available, and that of course suitable for this particular green space type. But what I mean that it's not always possible is that some datasets, like digital surface model that has a higher resolution is not available for free. And so we still do use one or two datasets that are that we buy or that we bought. But like around 90% of the datasets are something that we can download, so everyone can download from certain sources for free.

### Interviewer

And do you have any challenges or grievances or unmet needs with using open data?

### Interviewee

Yes, definitely. So the main issue was free data is either not high resolution or very precise or it's not well documented. Let's say Sentinel data, it's really available data and has some really good coverage, but if you look at very small green spaces, then its resolution is already too rough, so it's not very high, and then we cannot really use it for this purpose. But we can then use fully available digital orthophoto for instance, and that's provided by state of <redacted>, which is like 40

centimeter resolution. But we observed that in most cases those data sets are not well described. So, for instance, unclear when the image was taken, and if I talk about green spaces, the temporal component is of course important. So then it's quite difficult to describe or make assumptions based on the data that you use, because it's not really clear from which date the data comes from. So that I would say the main issues, how well it's described, how well it's documented and if it's resolution is good enough for my purpose, for my use case.

**Interviewer**

And so typically, where do you get the higher resolution data?

**Interviewee**

So for <redacted>, because my main study area is <redacted>, the state digitalization and geodata ministry or organization, the data provider. So usually this data, of course, you could buy only, it wasn't for free. But now due to the EU initiative of opening data, they started this January, 1st of January, making datasets available. Yesterday, they don't have all the datasets yet, but things like digital orthophoto or land use dataset or I don't know house data set that I also called <redacted>, they are now made available to download. But some other datasets like digital surface model or so, they are not yet. Hopefully they will be soon available. I mean, they supposed to, but it still takes long. So it's not like at once opening all the datasets. So that's mainly where I get it.

Or if I use Sentinel, of course, from the Sentinel hub or some other sources.

**Interviewer**

And apart from the lack of resolution, do you have any issues with using open data?

**Interviewee**

Well, I didn't really experience big troubles of using open data, so at least the sources or the data sets I use were pretty okay, so there wasn't really big issue with them.

**Interviewer**

And now we're going to talk about Esri as an open data intermediary. How do you use Esri products in your work? And for what purpose?

**Interviewee**

Yes, as I said, I actually did use previously for spatial analysis, but I recently started to use quite actively the deep learning packages that they have, which I find actually really useful and really cool because they have pre trained models and then you can just take it and apply on your data set. That's what I did until now.

Apart from that I use all of the other tools that are available within the Esri, like clipping, map algebra or another spatial analysis tools. I also use it actually for teaching in some courses and but also again mainly for machine learning workflows where we, for instance, look at forest fire where we actually take the workflow that's all embedded into the ArcGIS Pro from starting to the end, and then going through all the important steps of analysis. So I mostly use it for image analysis.

**Interviewer**

Do you also use it for visualization?

**Interviewee**

Yeah. So mainly for making final maps of the analysis. Actually, I also find it quite nice how it looks and how you actually interact with it, because it's quite straightforward and easy to create.

**Interviewer**

And are you familiar with Esri open data related services like Living Atlas, ArcGIS Hub, or ArcGIS Online?

**Interviewee**

So I am slightly familiar with Living Atlas. I've never used ArcGIS Hub actually. I did use living Atlas, something like couple of years ago for teaching purposes, looking for some forest or some locations of natural hazards, I guess it was landslide or something. So they had like nice coverage of worldwide coverage of points where the landslide occurred. So that was, I guess, one of few interactions I had with living Atlas.

**Interviewer**

And how do you find these services? Are they useful?

**Interviewee**

I found it quite nice because then you didn't have to go to, let's say NASA database and find the dataset and download it. But you could just search for the name in the Living Atlas and then take it. But what I experienced was Living Atlas at that time, at least, is that most of the data that you do have there, you cannot like export shapefile and then use it somehow differently, but rather you can use it as input for certain tools but not export it. But you can of course export the results that comes out. So while it's there available, you can still not take it and export and utilize it further as free data.

**Interviewer**

So you cannot take it and use it in different environment or different platform?

**Interviewee**

Yeah.

**Interviewer**

And any other grievances in using open data related services not? Not necessarily grievances, but something that you perhaps don't really fancy, don't really like.

**Interviewee**

Well, not really, but probably maybe, at one point is that, at that time at least, the spatial coverage, so global coverage wasn't so good, rather US datasets were like dominating the library. But if you searched for something in [country D], there wasn't much available, but also because in [country D], datasets are generally not available. Yep, so I would probably say the spatial coverage of the datasets.

**Interviewer**

And do you have any experience of using QGIS?

**Interviewee**

Yes, I actually use mostly QGIS and not QGIS.

**Interviewer**

And how is your experience using QGIS different than ArcGIS?

**Interviewee**

I can say positive and negative. From a positive perspective, I would say it's quite nice that it's open software, so you can actually do analysis and then describe it, and anyone else can repeat this process again because you can download the same version of QGIS and then repeat it.

But it also goes other way around, some workflows may be easier in ArcGIS because the tools for that are already there. If you want to do it in QGIS, then you have to you have a work around or the other way around some tools and QGIS are not directly one to one in actually -- so you need to find the kind of work around again.

But from negative perspective, I would say that Esri is, of course, a commercial and it's well maintained, and the tools that you have are always there and they always work the way you would expect it to work, which is usually not the case for QGIS. Sometimes [QGIS] are not well maintained, sometimes after updates, some plugins no longer exist or they don't work simply, or they change parameters within the tool that was there before and you cannot find it anymore. Or it just breaks sometimes and then there was no obvious reason why it didn't work.

Yeah, I do like using QGIS because it's open free, but sometimes because of this tool maintenance issues sometimes I do get the feeling that OK, I should go to ARcGIS and just run it there because faster and it will not break.

**Interviewer**

And what about integrating open data to the software? Do you find like integrating open data to ArcGIS much easier than QGIS or the other way around?

**Interviewee**

By integrating you mean using within the environment?

**Interviewer**

Yes.

**Interviewee**

I would say it's the same because depending on how you get the data. Yeah, you do have common formats that you get like, let's say TIFF or shapefile and then they would be equally possible. It would be equally possible to work in both of them. So I wouldn't say that one has obvious advantages and the other one disadvantages of integrating open data.

But rather, this advantages and disadvantages occur in the later steps of like working or like processing the data.

**Interviewer**

Do you have any issues with using data with OGC standard in ArcGIS environment? Have you ever encountered any problem with that?

**Interviewee**

Honestly, I have never had such issues.

**Interviewer**

And typically when you source open data from different sources, it's not through ArcGIS platform? It's typically from outside and then you bring it to ArcGIS software? Is that the case?

**Interviewee**

Yes, that's the typical workflow would be. So I get data somewhere else and then I work with this data set in ArcGIS.

**Interviewer**

In general, do you think Esri plays a role in enhancing access, supply, or flow of open data?

**Interviewee**

If you think about the Living Atlas, I would say yes, because certain datasets are now openly, freely usable through Living Atlas. Of course it makes a lot of work easier [because] you can directly search for it and use it. So it reduces the time spent on actually looking for data somewhere else and then figuring out how you should get it because not all the web pages have this download click that you can click, you have to look how you actually get this data set. So in this term, it definitely does enhance the usage of the free data, yeah.

**Interviewer**

And do you think they also enhance or play a role in connecting other actors in the open data ecosystem?

**Interviewee**

Like the actors with each other or rather actors to ArcGIS?

**Interviewer**

The actors with each other. For example, data providers with user or user with another user. Do you think they provide that kind of avenue for people to connect with each other?

**Interviewee**

I would say indirectly and not directly because probably if you know where the data comes from because it's described in Esri and you need something more, you could probably contact them and or get into contact. But I wouldn't say to me it's really the main aspect in using or integrating the open data in Esri.

**Interviewer**

And do you see any negative or less than ideal impacts of Esri towards other open data to actors?

**Interviewee**

I would say well, I'm not 100% sure of course about I don't know what the licensing of like taking dataset for example from NASA and then using it or making it available through Living Atlas. So what the agreement between these two stakeholders is. So, it could be that if it's the case that it's only available through Living Atlas and not otherwise. I guess [if that's the case] it's then of course not a good idea because Esri is a commercial software and not everyone has access to it. It cost a lot, the license. In this sense probably it's not so good.

But if the licensing issue, so the data that Esri take still available somewhere else and its access is not limited somewhere else, then I don't think it is somehow limited [i.e. a problem]. But yeah, it

probably depends on what kind of data that is and but also could be if it's the case that the data is only made available through Living Atlas or ArcGIS Hub, would it further be made available for everyone else who doesn't have ArcGIS.

Just wrapping it up, am I allowed to use the dataset that only within ArcGIS, that I need to have a license for ArcGIS, or can I still access it or is it still made accessible somewhere [else] so that I can access it and use it, for instance in QGIS?

**Interviewer**

So in terms of legal aspect, it's quite ambiguous. And also you mention in terms of technical aspect as well because it's hard to export this data from Living Atlas out of Esri environment, right? So that's also a technical impediment in in that sense?

**Interviewee**

So I guess if you are active Esri reuser, so you do have license to all different extensions and so on, and then it's actually quite nice to have the Living Atlas or open data within this environment. But then it's still like somehow a close capsule for only people who has the license. Not for everyone.

**Interviewer**

And like in an ideal world, do you have any suggestions or wishes of how Esri's way of doing things or business model could be changed or could be better?

**Interviewee**

So I guess I need to keep in mind that it's a commercial software and probably wishing that everything is open and everyone can easily get license, also for online, it's probably not possible. But I wish [based on] my experience, some things like datasets but also tools are not properly described or explained. I experienced recently with this deep learning models that it's quite nice trained. They referred to the paper and of the method they used, but then there is no further information on what are the parameters that they took or how they train it. So if you want to use it and then you still want to describe, it's not really possible because this information is not existing. So I actually asked them the help of Esri, what should I describe in this case, and they said, yeah, the original paper based on which the tool was created. But it's still not how the implementation actually is. So I wish some tools are better, like parameters, are better described or the datasets probably within Living Atlas are also better, the metadata. So that you can actually use it.

**Interviewer**

So they are like black box within the software?

**Interviewee**

Yes.

**Interviewer**

OK. And these training models that you mentioned, sorry, I'm not very familiar with deep learning, so do they also rely on training data sets?

**Interviewee**

Yes, exactly. That's actually why I like using it. I don't have to take training data set myself and train from scratch, but that's been done by Esri. So they do collect training data sets and then they train the model so I can then implement it. That saves a lot of time, of course, and obviously they do have

access to really good data sets to work with, which I wouldn't have otherwise. So that's actually really positive side of this whole AI tool box. But on the other hand, it's not documented well enough to know what they actually -- how they configured the classifier or the model.

**Interviewer**

But do they describe enough what is the training data set?

**Interviewee**

No, there is actually no information about the data set was to train. Once, they do say something like very general, the model was trained with orthophotos 30x40 centimeter resolution. But that's all you know. So you don't know what date or what kind of sensor it was.

**Interviewer**

Yeah, because this is your field, as a researcher, is it something that that you also want to know, like the training datasets what exactly they are?

**Interviewee**

Yeah, sure. Because depending on what you're trained on and then made available will affect on what you further use on. So if it's only trained on, like certain centimeter data set, and I only have 1 meter data set, and of course the results will not be the same as you would wish from the classifier. That's why it's quite important to know what the input data to the model was and what the accuracy was and how they trained it. It all affects your results.

**Interviewer**

OK, so my last question, what is your general view about how open data intermediaries not limited to Esri, can play a better role in the open data ecosystem?

**Interviewee**

I guess maybe by providing the place from where you can access data set, for instance, all possible free data sets, so that you don't have to visit every single a page of every single organization to get the data. A place that actually sums [i.e., combines] them together like database or at least gives the links to where you can access it, would be quite useful. I mean, it's very nice that open and free data exists and there are a lot of initiatives to make even more available, but maybe a place where you have somehow metadata, some links stored so that you can search and see where it is, and then you can maybe directly download or get some access link to somewhere else would be quiet beneficial I would say, as a researcher because you spend a lot of time searching for data and that would probably save a lot of time.

**Interviewer**

So that's all from me. Those are the questions. I'm gonna stop the recording now.