

June_18_409_005

We're gonna talk today about transforming risk assessment by using open standards in data collaboration.

So we've been working on the risk data library, developing open standards for metadata to describe risk information. And we've been working on this several years. We've now got some traction with some users implementing the risk data library and we wanted to share the concept division of the risk data library with you, some of the details about how to use it and share with you some use cases, examples where it's been used at the moment.

And hopefully by the end, I encourage you to also explore the RDL and use it in your own work as well. So we've got four of us in the room and a recording from somebody remotely. So we first have Pierre Chonofsky and Pierre's DRM specialist at GFDR and runs the RDL program as part of the Digital Earth team at GFDR.

Online we're gonna have Rachel Vint, who's director of Open Data Services and they provided all of the best practice in open data standards to bring the risk data library to the point it is today. So Rachel will talk through some of the context of open standards and their importance beyond risk data as well as for the risk data library.

And then we have Daniela Zuluaga from Arup, who's a, Daniela's a senior risk and resilience engineer, brings a lot of expertise in risk and resilience assessment and developing climate resilience strategies.

So she'll talk about how Arup are using the RDL. And then Pradeep Mandapaka will talk from the perspective of JBA risk management using the standards as well. So he's a technical director, bringing hydrology and climate expertise into JBA's work on risk and resilience.

My name's Stuart Fraser. I'm a senior DRM consultant at GFDR. I've been working on not only risk assessments but also open standards and tools to share risk information more broadly for several years.

So with that, I will just give you the agenda. So Pierre's gonna take us through the vision for the risk data library, the background. I'll take you through some of the capabilities of the RDL standard.

And then we'll show Rachel's video on the RDLs in the context of open data standards more broadly. And then we'll have Daniela and Pradeep demonstrating the value of RDLs in practice. So Pierre, over to you.

Hello, everyone. My name is Pierre Schonesky. I'm a disaster management specialist at the World Bank, and I'm leading that project together with us, too. Thanks for coming. I consider you all geeks because coming to a session on open data standard at the end of a long day.

I appreciate. So I'm going to give you the why and the vision of the risk data library. So first of all, so we are part of GFDR, and as you might know, the global facility for DSR reduction and recovery.

We have been really, from the beginning, supporting the access to risk information. That's really always been a key objective alongside of supporting capacities for countries. So providing the possibility for anyone to access to risk data was one of our key objectives.

And so it's really also aligned with, of course, the Sendai framework that you probably know about. And the fact that, yes, disaster risk management, disaster risk assessments should be based on an understanding of disaster risk in all its dimensions.

And so when we say all the dimensions, it means many different information we have to combine together, right? So that's really the beginning of our journey towards thinking about the why we should have data standards for disaster risk information.

And so we are kind of a statu quo in terms of how we work on risk data. And that was funny, you know, I attended the session on Google speaking about artificial intelligence, but the reality of our day-to-day life is we are still emailing our colleagues, oh, hey, where can I find that data sets?

I saw that report, but I can't find the data behind it. So how do I do? And I'm sure that resonates in all of your work. So that really is a situation where we are still in. So an increase in terms of demand for risk information, because of also the mandate for no oil investment, for instance, to do a disaster risk cleaning and so on, but still difficulty to find that type of data.

And often, if you are lucky and you find a disaster set, you don't have the right description, you don't have the right way to find about what is in the content of that data set. And also what was a process through which we came up with that type of information.

So that's really what we have been trying to address with the risk data we work. And our vision is the following. So what if I could easily find information? What if I could have tools that can help me to search for, to access, to share with my colleagues, to share with other organizations risk information?

So that's really the driver for our project, and hopefully something that will be useful for you. And so for that, and also based on that, we really think about the open models. And so of course, there is the open data movement, but much more we have what we call the FAIR, so it's findable, accessible, interoperable, and reusable data.

So that's coming from the scientific community. That's really been also an influence for us on how we should think about the way we should be able to access risk information. And here is the result. So the risk data array, or let's say the risk data array standard, is about describing the different key categories of information that are needed for disaster risk assessment.

So it's about describing your exposure part, describing your hazard part, describing the vulnerability part that is linking the two elements together, and then describing the risk output, so the loss, but also the damage data, right?

So that's really that vision that we want to promote. But we are, of course, one organization, and for that standard to be adopted and for that vision to be achieved, we need to create what we call that network effect, right?

So we need a lot of organizations, we need a lot of early adopters to come with us and to show that actually we can share data with information in a better way. So I'm hoping that throughout that session, you will have a better understanding of what is a risk data library, and maybe you will want to test it.

Thank you. Oh, yeah, I forgot that slide, that we have many partners. So we have a steering committee with GEM Foundation, UCLA P Center, OASIS Modeling Framework, IDF, of course, 3S3, GBA, with Pradeep in the room, IFRC, the Red Cross, and I should also acknowledge the support from the 3S3 Foundation for the last part of the development that you are going to see.

So thanks again.

OK, so I'll just take you through some of the components of the RDLS. We won't go into too much detail. There's so much to cover. But there is lots of documentation on the website for you to explore each of these components in detail and get to grips with how to complete the metadata standard.

OK, so Pierre already mentioned the four components. So within the vulnerability data, we provide capability to describe vulnerability curves, fragility curves, damage to loss models, engineering demand parameters, social vulnerability indexes.

So it allows you to describe lots of different data in terms of vulnerability. Exposure data, we focus on describing the asset category, buildings, forestry, livestock, et cetera. The

taxonomy that's been used so people can immediately see what taxonomy has been used to describe the occupancy type, the construction type, or the asset type.

And we provide the quantity description. So within the metadata, you can search for exposure data that's providing replacement cost, or number of people, or kilometers of road. So you can immediately search by that quantity.

Within hazard data, we describe the hazard type. And within that, the hazard process. And we're in the process of aligning this with the UNDRR hazard information profiles, so trying to bring an existing standard to describe the hazard type and process.

So you can describe earthquake footprints, whether it's ground motion, liquefaction within the same data set. And you can then filter by those details as well. We provide details like units of intensity.

I've been looking for flood maps previously and had to dive into the data to figure out whether it's meters, centimeters, or decimeters in terms of flood depth. You shouldn't have to open the data set and interrogate it to find that basic information.

So we've added things like that in there, so you can immediately see what it's showing. And of course, it allows you to showcase current hazard data, hazard intensity under different climate change scenarios and describe the scenarios that have been used.

Further than that, we enable you to describe event sets for probabilistic models or single events for scenarios. So you can provide a single event or a set of a few events. Observed or simulated, and the metadata allows you to describe the origin of those footprints.

And it allows you to describe the occurrence probability, the return period associated with a particular hazard map or a scenario. It also allows you to describe multiple footprints so you can provide lots of uncertainty around the footprint data, provide the range of intensity measures within the same data set.

And we allow you to describe events, cascading events, so you can provide or you can identify the event that's triggered another and provide linked hazard maps that way. Within the exposure data, we support existing taxonomies, like the Jed for All taxonomy, various taxonomies used in the insurance industry.

So we're not proposing to describe exposure data in a new way, but we link back to these existing structures. We support data building footprint to aggregate data types. And you can describe multiple asset types, lines of business, including population, natural environment, as well as the built infrastructure.

Within vulnerability data, we classify each curve using the hazard type and the exposure category that have been defined also within the hazard and exposure components. So each component is using the same consistent code lists.

So when you're searching for residential buildings, you can pull up the residential exposure data, but also the residential vulnerability curves that relate to residential data. So it's all consistent across each of the components.

And it describes in the metadata the impact type metrics, direct -indirect impact calculation methods that have been used. Also links out to source data that's been used, source information, relevant technical reports, peer -reviewed papers, et cetera.

And then with the loss data, we can provide labeling to define whether it's describing direct loss, indirect losses, loss of damage to structures, contents, production, impact on people, whether it's economic or insured, and the quantity, whether it's the number of buildings, number of people, or length of infrastructure, for example.

And again, describe the loss calculation approach. And the loss data provides an explicit link back to each of the other components. So if you find a set of risk estimates described in the risk data library standard, within that metadata, you'll find an ID that links back to the exposure data set, the vulnerability data, and the hazard data that have been used to create those risk estimates.

So in that way, we're providing a complete standard that will allow you to describe a full package of risk information. So you can describe a single data set or a full package from a risk assessment.

We just wanted to show you very quickly, so that's the homepage, but what this looks like in practice in a data catalog. And the standard is set up as a JSON file, so it can be loaded into lots of different data catalogs.

Our vision is that this would be used in different data catalogs, and this is our example catalog that, and we also have data on the World Bank data catalog as well using the RDLS. So here we've built a tool that will allow you to filter by the risk data category, so you can immediately filter your loss data from the others.

You can filter by country, the usual filter. Geographic scale, hazard type, licenses, and things like project name as well. So we've loaded up here lots of data from an Afghanistan risk assessment done in 2018 by GFDRR.

It's all openly licensed. And just looking at the exposure data here, you can find all of the resources, so lots of TIF files describing the building's infrastructure, agriculture developed under that exposure data.

And some of the categories and the dimensions and the quantities that are included in that data set. And then some additional information around the purpose who created the data, and again, the license.

So with the metadata, we're able to provide this filtering. We're uploading more and more data to this new portal so that it becomes the home for data stored by the risk data library. So yeah, one example, but we're working on this being able to be imported to not only J CAN but C CAN, and we can work on importing this to other types of catalogs as well.

So I just want to talk very quickly about some of our early adopters before we have two of our early adopters in the room come up on stage. So I've been working with the OASIS Loss Modeling Framework for a number of years.

It's a modeling framework maintained and built within the insurance industry, increasingly being applicable to development sector, risk financing and insurance. They've worked hard to develop some open data standards for exposure data and results data, so the losses, but they've never worked on describing metadata.

So it's a real gap in how they're providing information. So they've done a study to look at how they might use the RDLS, and they've concluded it's very compatible. And the work that's ongoing this year is to integrate the RDLS into the OASIS ecosystem and provide some online examples of models stored with the RDLS information.

We also have the European Commission Risk Data Hub managed at the JRC, who have integrated the RDLS into their risk data hub platform. So some of their feedback on the left-hand side there, which persuaded them to use this and implement it.

Actually, without support from us, they did this entirely from the documentation online. So hopefully that shows that that's useful and it's all very workable. But they were very pleased that it was data tailored for disaster risk management and providing greater interoperability and integration of their data that's going on that platform.

And then an academic example, we have some of the Tomorrow Cities team in the room. So this is a recent conversation and I'm pleased that Tomorrow Cities have decided to start to use the RDLS to describe their data from this Urban Disaster Risk Hub.

So, yeah, an interdisciplinary research hub focusing on urban scenarios, future visioning, and impact assessment at the urban scale. And they've been working in nine cities. I think I've got that right.

So that project's coming to an end and the team are looking to host the data somewhere and making it accessible. We're pleased that they're going to start using the RDLS to share that on their Tomorrow Cities data catalog.

So a few early adopters there. We're going to just break to listen to a video from Rachel Vint. As I said, she's from Open Data Services. They really brought the best practice from developing lots of different open data standards into this project to make it as useful and easy to implement as it is today.

So, yeah, we'll play this video.

Hi everyone, my name is Rachel Vint and I'm a director with Open Data Services. We're a workers cooperative and international experts in data standards. So our core business is researching, building, maintaining and promoting open data standards to solve global challenges.

So we've worked with the risk data library team recently to help develop the risk data library data standard within current best practices. And today I'm going to briefly look at why data standards are needed, what they provide, and finally why you should care about them.

So what's wrong with data anyway? Why do we need data standards in the first place? So we all know that working with data can be incredibly messy and time consuming, requiring cleaning and wrangling before being used.

Data can also be really difficult to find. It can be inconsistent. It can be siloed and inoperable and not machine readable and generally just not helpful sometimes. An example of why data standards are important is to think about all the different ways there are to write the date.

So without choosing and agreeing a standard way of representing the date in data, it's really impossible to know whether the date on the slide is the 10th of November, 2012, the 11th of October, 2012 or the 12th of November, 2010.

So data standards themselves are the rules that describe how to record and publish data. And they're supported by an ecosystem of software tools, documentation and guidance and communities of practice.

Open standards are freely available and developed using collaborative processes. So they function as public goods that anyone can benefit from and that anyone can contribute towards their development.

Data standards also help to solve complex real life problems, problems such as the lack of timely and forward looking humanitarian aid data. And so by using the data standard, which is the International Aid Transparency Initiative, this allows for better decision making and more cooperation so the aid is more effective.

Another problem is how to tackle global corporate corruption and financial crime. And again, by using the beneficial ownership data standard, this allows for better corporate transparency and accountability.

And finally, what we're here to talk about today with the increasing demand of data for climate adaption and disaster risk reduction, this is helped by the risk data library standard, which makes it easier to find, publish and share the relevant information.

So to summarize what an open data standard is, they lay out how to format and structure the data, but what makes it open is that the data is available publicly. The decision making is open and transparent and it builds upon and aligns with other standards.

And this final point is important to avoid the new data standard pitfall. So this is possibly the best cartoon I've seen about open data standards, or although I'll admit there aren't many. So for those that can't read the slide, it says the situation.

There are 14 competing standards, 14 ridiculous. We need to develop one universal standard that covers everyone's use cases. Later, there are 15 competing standards. And that's why the risk data library standard builds upon other established standards, such as GED4ALL, GEM Building Taxonomy, UNDRR, et cetera, without duplicating what's already been established.

So along with not duplicating what's already out there, there are some other best practices in open data standards. These are things like using open formats, having clear documentation, using version control, having machine readable data, working with key stakeholders, linking with other standards, as we've mentioned, building a community around the standard and having transparent governance.

There are also three main components to an open data standard and practice, not including its governance. So these are having schema and codelists. So these are the things that define the structure and format of the data, such as the meaning of each field.

You also have documentation, so guidance and reference documentation so that you know how to publish and use the data. And then you have open source tools, so things like tools for converting, validating and exploring the data.

And as it is an open data standard, so you can contribute to the ongoing development of it by getting involved, such as through the public GitHub repository. And so finally, why should you care? Because when done right, open data standards make everyone's work easier, whether that's analyzing or producing risk data.

The data is then accessible, interoperable, and of high value. The data can then be used to make and measure change. And as a result, the data has real world impact. Thank you.

Okay, so if you weren't convinced before, hopefully that explains why we spent so much time working on the RISC data library. And some of the workflow, the documentation that Rachel showed there, the conversion tool at the end, that's all available on the website and makes it much easier to create the metadata in JSON format, so there's a workflow laid out that's all documented in the user guide.

I'm Daniela.

Thank you very much. Thank you. Hi, everyone. Good afternoon. Hopefully, this will be interesting still. But I wanted to just share my experience or the experience that my team has had with the RDS. Just to introduce myself and my company, in case you don't know us, I work at Arup.

We're a global engineering firm. And we, my team, focuses on leveraging our multidisciplinary teams who evaluate risks to develop environments and people. And because we are risk practitioners, we are in a day -to -day job is collecting data, working with data, analyzing data, to build risk assessments to support our clients in decision making.

And so I just wanted to start with a little bit of background. And so I know Stu mentioned the status quo right now, which is it's hard to access data. So I just wanted to highlight that for us. This is one of our key pain points.

So again, we feel this every day. When we try to find data, and it's really hard. And even when we get data, it's hard to understand and use it. And this has also led us to develop our own tools internally, just to try to help us or ourselves organize.

So we have processes, protocols, guidelines, databases that we have worked. So I just want to mention three really quick. The first one is universal taxonomy for risk assessment. So I had a session this morning where we presented this also in collaboration with GFDRR.

And again, this is just tying into the theme of why we need to at least try to standardize basic frameworks, again, not to be prescriptive, but just to have a little bit of organization. Then we have also developed what we call HAPI, which is our hazard API.

Again, trying to facilitate that extraction of hazard data in an automated way. So we built this API that we use ourselves for day -to -day work. We connect to data sources from partners that we work with for the data that we develop ourselves.

And we have also on our own internal library of vulnerability functions for agility functions. So again, these are all pieces that we have to work with on a day -to -day basis. And now in terms of our experience, so we have seen that metadata is really key.

We have seen that if we don't have that, there's a lot of regeneration of data and tools are probably already exist somewhere else, but then this needs to just constant mushrooming of new things like the cartoon that we saw before.

So I just wanted to bring one example that I saw firsthand. A couple of years ago, we were working on a project with the World Bank. So the World Bank put out this RFP to try to estimate the investment needs on adaptation in Columbia's road network.

And of course, to do that, you needed to do a risk assessment. I knew that the country already had done a really good full detailed probabilistic risk assessment multi -hazard for the multiple sectors, including roads.

And so we wanted to propose like, hey, we don't need to develop the model again. Why? We just need the data. If you can get us the data, we can skip that step and focus on the important part, which is the second step of getting to the measures of the adaptation financing, which is what they wanted.

And so we were able to do that. It was a happy case because we were able to get the data, but it wasn't as smooth because there wasn't a standard. So we had to ask the World Bank to ask the Ministry of Environment to get the data.

I drove physically and got in a hard drive, then had to dig into what the data sets were, what it meant, but we were able to use it. So again, that's just one case where it is valuable to have the data available and usable.

The other thing that's key for us is that it has, the RDLS has GitHub or like the digital components, which is key for us. So again, we also have our own GitHub repository. So this is just a screenshot from our own.

So we have a bunch of different repos on GitHub. This is one that we use in our hazard API and happy to process Fathom data, which is one of our partners. So again, this is what we do on our daily jobs.

And so if I have this access, our functionality to use GitHub, it's just perfect for us as well. And then lastly, I just wanted to share three instances where we tried it out, we tested it out to see how valuable it is.

So first, which is what I'm showing here, we're working on a project in Madagascar, also with the World Bank, to do multi-hazard risk assessment for the transportation sector in the country. And so I tested it out because we have the results now.

So we have, I use the RISK result or the RISK shapefile. And there's a, what you see kind of on the top screen is you have a template that's a spreadsheet. And so you start filling out the fields which are intuitive for, if you work with this data, it's easy for us to fill out.

Like Stu was saying, it's fields that we already know and use, but then when you use the tools that are online, you can convert that spreadsheet into the JSON file, which is what creates the metadata.

And so it was a really simple process to use and follow. And that just, again, leverages. The example that he used it is on the shapefile of the airports that we assessed. And so that would just really enable the delivery of the results.

The second case that I tried is, so we had a project that we concluded before for the Cook Islands with the Asian Development Bank. And we had a deliverable, which is

again, the GIS files. So before we had this, we just delivered it in the standard, let's say ISO standards, GIS standards, all good.

But then I said, what if I just test this re-actively? So I took one of our files as an example, again, just the rest result file. Because I already had that, I simply exported the JSON file. I used the online tool converter to change it back to the spreadsheet so I could complete the fields and then re-export it with the completed fields.

And that worked. One thing that I wanted to point out that is really helpful is that when you use the tool, it gives you flags like that error. And that's really good because it lets you know what you're missing.

So it's like, hey, you forgot to fill out this field, which is critical. So I go back and fill it and then I update it. So just because I was trying to use it different in different ways. And the last way I tested it was not on a risk layer or data set, but on a resource, which is something that I found really useful.

So if you go into the tools and the worksheets, you see that, yes, you can use it for risk, hazard data, exposure vulnerability risk data, but also for just general resources. So I use the taxonomy that we published, which is a PDF report, but it's a useful resource that you could cite or use in any of my projects.

And there's another tab that allows you to do that. So again, it's the same process, the same thing, easy export, it flags any errors that you might've made in the process. And that way, because you have the unique ID, you can then package it together.

This is my one project. I have all of these layers. I have this resource that I used, which is not a GIS file, but it's just a report that's still usable. So with that, I think that's it, what I had in terms of my experience.

Oh yeah, that's just that Jason, that resulted from that. But to conclude, I just wanted to share a little bit of why as from the predictioners point of view, this is really useful and I

think would facilitate our lives internally and then externally delivering, working, receiving data from others as well.

Thank you.

Good afternoon, everyone. I'm Pradeep Mandapaka from JBI Risk Management. We at JBI Risk Management, we are a team of scientists, engineers, and mathematicians. And we develop this high resolution flood mapping and flood modeling at a global scale to be mainly used by insurance and reinsurance and banking sectors.

In addition to catering to these sectors, we also work closely with the International Financial Institutions, government agencies, and humanitarian agencies for flood risk assessment projects as well as disaster risk financing and disaster risk reduction projects.

And in these projects, we use sometimes our own data sets, but we also develop custom hazard and probabilistic models to serve these projects. So in order to maintain these projects, we need a standard metadata standard.

So we are an early adopter for RDLs because we saw great value in using this standard for a variety of data sets that we developed as part of these international development projects. So today I'm going to share our experience generating the metadata workflow following the Risk Data Library online converter tool.

We used this workflow generation for two of our recently finished projects and where we developed hazard and climate change event sets. We populated the spreadsheet templates and we uploaded the data into the online converter to obtain the JSON file.

The feedback is that the workflow, in general, in terms of populating the Excel files, it's very simple and understandable and the templates are quite useful. And there is a lot of supporting information that is available online.

And the validation feature which is present in the online converter tool is quite helpful as Daniel just demonstrated. It throws very helpful error messages. And we are actually applying this to all our new data sets that we are going to develop in our ongoing projects as well as our upcoming projects.

And we also saw great value in this RDL is that we are trying to implement it internally to version control our own data sets as well as version control our documentation for the data. So I'm going to stop here now and thank you.

OK, thank you very much. A few more details on some of the use cases that have come about from the RDLS. So yeah, we're excited to see some traction and people taking up the RDLS and starting to use it.

It's an ongoing project. We're learning. And every time somebody uses it, there's some feedback. We're already working on the next version. And we do really want everybody to get involved and provide feedback via the GitHub and help improve it as a community.

It's something that will improve over time and will become more useful as more people use it. So hopefully that's given a window into some of the capabilities, some of the examples of how it's being used.

And we've got 20 minutes left, so we'll use the last 24 questions and take any questions, a bit of discussion. So yeah, please come up to the stores. And while we discuss, we'll also leave this information on the screen so you can access all the documentation, examples, workflow from the riskdatalibrary .org and directly at the RDLS standard.

For any support, you can contact. Use the email at the bottom there. OK, can I open it up to any questions from the audience?

I can feel a low energy, but I'm sure we can manage to finish that there with, yeah, please.

Etsuko Yasu from Brando University, Canada. Basically, the things are a little bit over my head, to be honest, it's a little bit too technical, but my question is from the social vulnerability point of view, the challenge for us is having updated data, even just the census data, they update almost literally every day, right?

So, I wonder how you can make your data current and accurate, like just constantly, I have to check back again and again, or is there any way that you can achieve that?

So, the data would be stored with the risk data library when it's created as part of a project, a risk assessment or other project. And all of the sources for that data set would be stored within the metadata.

There would be a version or a year clearly identified in the metadata. So if somebody came to that, they could see that it was created in 2021, say, look at the data that contributed and see whether that was up to date or not, and then take a decision to use it directly or build on it or start from scratch.

So we wouldn't be updating the data outside of the projects, but we'd be able to store all the information for you to make that decision.

Thank you for this presentation and for this effort, I think it's very good to make data more accessible. Do you have a question? In disastrous management to do inclusive decision making, we need more disaggregated data.

We also know that in a lot of data there's a lot of bias. What are you doing to, I think it's very good to make data, for example, machine readable. But we also know that making data machine readable will often even establish or how do you say it, make the bias even bigger.

Because people will say, well, you know, we've did machine learning on it and it proves that this is true, but the data that went into it was already not inclusive and was not biased. So aren't you afraid that this might propagate the bias in data and hamper inclusive decision making?

And what are you doing to counteract it?

Can I take that one? Yeah, thanks. Okay, so we used to say garbage in, garbage out, right? So I think we are talking about kind of a value chain of like a process with different elements, different steps, different people who are going to work on the information and are going to pass that to each other.

I think what is helpful with having a standard is that upfront, you can already establish kind of a baseline of a framework of the type of data that you want to collect. And you know that that framework, you can, you will use it and you know that other colleagues or other partners will use the same.

So that's already a baseline of, okay, we agree that for instance, and we included socioeconomic indicators in the standard that having disaggregated gender information is important for vulnerability purposes, right?

And that's why also it's important to have open processes on and discussion on what is important actually. And so the fact that we know include gender information has been based on an open process and like the UR community, we had discussion about that.

And based on this conversation, it's now translated into standards and then tools. And I know, I mean, a standard is not sexy, but in the end, it's really support all the tools and the systems that we are using on a day to day basis, right?

Yes, and I think it's a good answer, but I think it only covers a part. Because it's, for example, a lot of lost data, for example, is not gender disaggregated yet, right? Which means that we don't know if gender plays a role in being more affected, or we do from the little that is disaggregated.

And I think if you don't disaggregate that data, you're also gonna find risk reduction measures that are not the most effective to actually target that disbalance. So yes, it's

important to include gender information, but we also need to, as a DRM community, become better at collecting disaggregated data, and not just on the socioeconomics or on the vulnerability side, but also on who's more impacted.

And it's not just gender, it's also disabled people, it's also ethnicity background. And I think the danger is, and I'm not against anything that you said, and I know it starts with opening up data, so I'm not saying that this isn't a step in the right direction, but I think it starts with being more aware of these things and the danger of making this biased data sets more accessible is that we then are gonna do analysis and say,

see, there is no bigger impact on, for example, women than men by disasters, why we do know they are, but then the data doesn't show it because it's not disaggregated. So I think we need to, besides the standards, also maybe adopt standards on what type of data we collect how.

Yeah, totally agree, but look, like for instance on gender, is it about, the question is it about us being responsible to include gender data into disaster risk management, or is it about the question on having gender data in census, for instance, and so then you're starting also to look at who is responsible for what, and so we can solve all the problems, but at least we are trying to support tools in an open manner that can help to address those problems.

But they can also help take them, and that's why we want. That's a risk, I agree, that's a risk, yeah, yeah.

I think this is an example of something we don't flag easily within the data sets and within the metadata. And things like this, we can think about adding a flag to say this is gender disaggregated or a filter at least so people can see how many, see the proportion of data sets that are and aren't and just raise a bit of, yeah, make it more visible at least.

Any questions, yes, Carol.

So I think it's a great initiative, and so you get access to many data.

data sets, and when it will be very successful, there will be similar data sets of the same kind in the database, I would assume. So, and I can imagine that it's not your objective then to say, well, this is better than the other.

So how are we going to do that? Because, like I said, great initiative, and if it's successful, we have to make choices later on what to use. So you probably thought about it.

The question has come up before yeah, and Yeah, rating the quality of data sets isn't isn't within the scope of creating a standard I think that the decision on quality has to come down to the risk analysts and What we're trying to do is open up the information so they can make that decision on the basis of seeing the data record in the catalog and and Hopefully they can filter and make that make that or get closer to that decision Before having to open up every data set and interrogate it.

I don't know Daniella and Pradeep, but maybe want to come in

I think this morning's session where Daniel presented taxonomy of risk assessment datasets where each dataset can be associated with a particular flag or a level. So that could help in deciding which dataset can be used for a particular purpose.

Yeah, I think it's, of course, me as a practitioner, I have to decide. I think it's kind of the risk. So if I, for example, would upload a dataset, then you don't know exactly what the quality is. Nobody knows exactly.

So sort of that, yeah, we need to sort of share the information, what is good to use or what not. But again, that's a secondary problem. I think this is already a great initiative.

Well, it's kind of the same question as for like a market, you know, a market, you need for a market to function, you need transparency. And so it's all about having transparency in what people are providing.

And so we don't, we don't want to, to stop competitions. Uh, of course, no, no, we, we want to promote innovation. We want to promote the fact that people will be able to share what they have done and explain how they have done it.

And, uh, and then we'll be able to compare and to, to select. Yeah. Perfect. Thanks.

It's also worth mentioning we're not trying to host lots of data in one place, so we're not inviting lots of uploaded data. So on the World Bank data catalog it would be data that's been signed off in the normal way through a World Bank project and deemed quality enough to finish the project and sign off that project.

The same thing with data in Arup goes through QA and JBA and I'm sure at DHV as well. In a way, it's dodging the question a little bit, but in a way the data that goes on to a company or organization's website or data catalog, it becomes their responsibility to have signed that off and make sure it's good quality.

And hopefully they're using the risk data library so we can search all of those different catalogs in the end.

And maybe can I just add one thing, because when we were looking at this, we're also trying to understand. So one thing is the standard, and another one is the library. So I could also use the standard for just internal purposes and not publish, if I can't, because it's, I have an NDA, it's confidential information, I have to publish, but I could still use it to organize my data sets and everything that I need to do.

So I think if you decouple the two, then there's two different uses.

Yeah, just have a question exactly on the balance between, you know, with the standards of the data scheme. We were discussing this today and the metadata part, which I think, I mean, is the new philosophical trends that we are proposing to go a little bit more into the metadata part and a little bit less, I mean, to the data scheme.

So I would like to understand from your point of view is that there's going to be like a still like a hybrid approach, I mean, for the future. Do you think that we will abandon the data schema, which would be a pity, I mean, for me, and go just for the metadata parts because you think, I mean, it's more effective, I don't know, just, I mean, what do you think is going to happen?

I mean, yeah, we can see that you are part of the history. Of the old school part, yeah. So just on that, I mean, I don't think there is any difference between the standard and the schema. I mean, the schema becomes a standard when it is used by different people, right?

I think maybe the shift that we have done is that we really want different organizations to, we don't want to centralize everything. That's really not the vision. We want people to use our tools, our standards, for their own purposes, and then if they share, that's great.

If they don't want to share for some reasons, and that's also great because they are using tools that are compatible with others. So that's the vision. But of course, to achieve that vision, you also need some key players, champions.

And we, oh, I mean, the bank, of course, is going to play that role. We already have made all risk information managed by the bank, collected by the bank, will be compliant with the standard, but we need also other adopters to show that it's useful and to create, again, that network effect.

Yeah, if I can just add on that, maybe I didn't explain myself, I mean quite a bit. So the standard, I mean it's okay, but I mean I understood you want to go more on the metadata component board and into the data schema itself, so not transforming the data, but I mean invest a little bit more on the metadata description, okay, in a sense.

Is that true or I mean are we still aiming at having I mean the standard I mean in front?

Yeah, so we have a focus on the metadata part, but we are still, so the standard, you can standardize everything, right? You can standardize your content, you can standardize the way you describe your content, right?

So we have a standard for the metadata, but for some part, you also have a standard for the content. The thing is, for instance, for the exposure, we already have a lot of standards for the content. We have JAM, we have JED4ALL, and so on.

So there is no reason for us to create another one, as we saw with the racial examples. So that's also the work we have been doing, where we need standardization, where there is a gap in terms of we don't have that common vocabulary.

So just some background to where the Wrist Data Library's standard started. It was started with some funding from UK DFID, now FCDO, to address problems in data access. So we had three challenge funds.

And we created a database schema for exposure data, which is JED4ALL, for vulnerability curves, which is called MOVER, and was developed by UCL. JEM led the development of JED4ALL. And we also developed a hazard data base.

And MOVER is still being used by UCL, and JED4ALL is still being used, I believe, by schema as well as by JEM. So these do exist, and they provide structures for the data. I think we realized at some point that to drive adoption, the metadata was the thing that people can pick up and describe their data and allow people to access it more easily without shifting their workflows to a whole new data standard.

For example, OASIS, I showed the OASIS example earlier. They have defined their open exposure data standard, but we want to be able to apply the metadata on top of that to make it more powerful together.

So that was what the shift is. But the database schema do still exist, but our focus is on the metadata at the moment. Any other questions?

We are going quick, Technica, for all of us. We've got a few minutes left.

for half past six, we were actually beyond half six, so obviously. Any other... Any other questions? Well, I think, can we wrap up, I'd like to just ask Daniela and Pradeep, from your use of the data, what have you found most valuable in being able to pick up this data standard and use it, get to grips with it, decide that it's valuable enough to continue and persevere with?

So I think there's two things. The first thing that I love, the word the error screens, because that's just easy to identify. But then the second is that it may seem silly because we, if we work on this every day, we know that yes, I need to know what the unit is, feet or meters, but if it's not there, it's just a headache.

And so what I would, like we would do before is you have your files, you hand them over and then have like a read me text file saying, this layer is this, this layer is this. This is the unit, this is my assumption.

So just finding that or having that in one unique place that it's tied together, all my data sets as a package, I think that's what makes it easier. And again, it's nothing, right? It's nothing new that we hadn't, that we're discovering something new, which is like, I mean, yes, why, why didn't we think of this before?

It's as simple as adding one field. What's your unit? What's your return period? Is it PGA? Is it, what is it? Because it's the information that we already do and use. So I think it's done. It's just that it's simple and it fits in with what we're already doing.

Yeah, just to add to it, also in terms of JSON file creation, one can create JSON file, but to create it from spreadsheet, it's really, we found it very helpful because spreadsheets are something we are very familiar with, so we can just populate the spreadsheets easily and use the online converter to convert the spreadsheets into the JSON files, and in the process, the validation feature is also very helpful with all its error messages and also the documentation,

the wealth of documentation that is available online. Although it did take us a bit of time to go through all that document, all that information, but it was not a very steep learning curve. It was quite, we see great value in using our DLS even internally within JVA.

Thank you very much, and I've had conversations today about the RDL and one thing that came up was having some example spreadsheets would help people get to grips with using it and how to fill it out, so there are things we can still improve that we've not thought of yet, we've not done yet, so we are still looking for feedback to improve this, I think every time there's a project or a question there's maybe a tweak to some code lists or an extra object or field that we could consider adding or a use case that maybe isn't quite completely dealt with,

so we are very open to any feedback via the GitHub, via the email, we are actively developing the next version of it, so we would welcome you having a look at the website, having a look at the documentation, giving it a try and let us know how you get on and hopefully we can build a bit more traction and have more and more data sets on these catalogs and a more effective sharing of risk data overall.

Thank you for coming, for your interest. I was just thinking, you know, we are talking about data standards and we are here in Japan and for many of us, you know, it's a completely different language, different alphabet, and but still we are talking together about, you know, disaster risk and disaster risk management.

So it's just to say that we need those common structure for us to be able to exchange, to exchange data. And so I feel, okay, maybe it's not the most sexiest thing, but I believe sometimes the less sexiest things are maybe the most important when we are talking about resilience, and I do believe data standards are important.