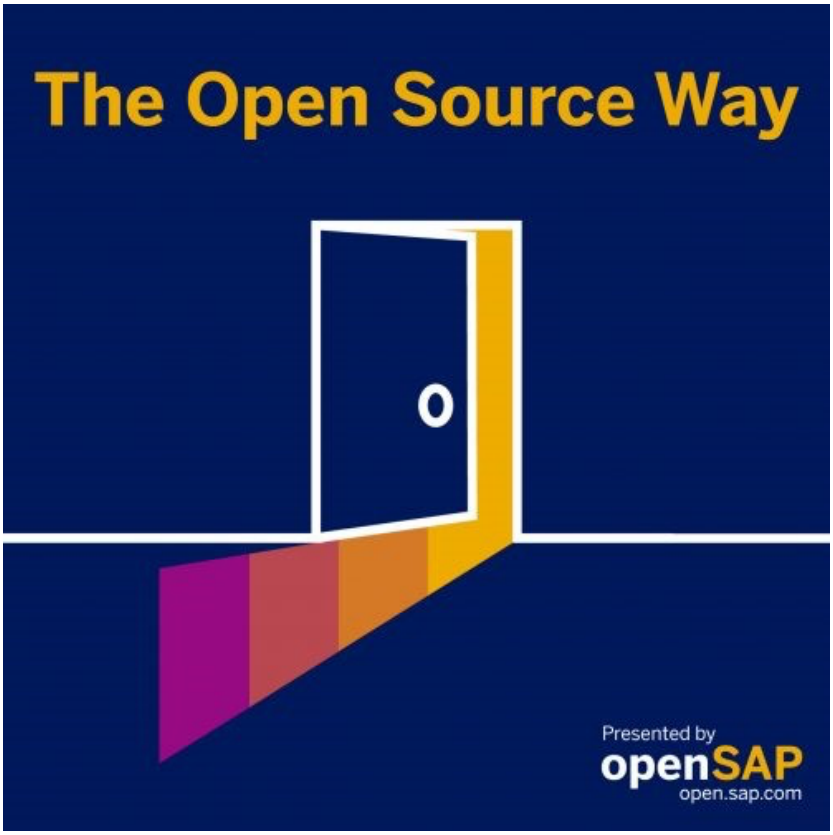
The Open Source Way

Episode 20: Gaia-X, Catena-X, International Data Spaces – Three Initiatives, One Goal



Transcript

Karsten: Welcome to the Open Source Way. This is our podcast series, SAP's podcast series about the difference that Open Source can be, and in each episode, we will talk with experts about Open Source and about why they do it the open source way. I'm your host Karsten Hohage, and in this episode, I'm going to talk to Nemrude Verzano about Data Space, Catena-X, Gaia-X – in general the cooperation efforts between the Fraunhofer Institute and some of our large customers at SAP. So, Rudi, nice to have you here.

Rudi: Yeah, hi, Karsten, thank you for having me today.

Karsten: You are more than welcome. Great that you could make it. Nemrude Verzano, called Rudi, is head of Industry 4.0 and digital supply chain innovation at SAP. And before that, he was been a consultant at SAP and also in between, I think, a freelance consultant, mostly focusing on integration and interoperability challenges in different industries, really. So, you have gone from being a consultant at SAP, a freelance consultant to now heading the activities for collaboration and interoperability. Is that basically how you, let's say, continue to live the customer focus that you've, acquired as a consultant?

Rudi: Yeah, in my case, indeed, I took this part because in the past I was involved mainly in all kinds of projects that include all kind of integration and, you know, interactions between different even legacy systems, for example, to get them somehow running. And therefore, I realized very quickly and easily that there must be something more than just, you know, doing the things over and over again and also nowadays, especially if things are getting quicker to be connected. It was obvious that I was somehow capable to solve and to cope with those challenges. And therefore, I went into this direction and also with the digital transformation that took place the last decade, it was it was somehow my path. It was always a good occasion for me to apply, of course, my insights, my learnings from different projects and therefore, for example, the Industry 4.0 Activities that our former CEO, Henning Kagermann, initiated also helped me a lot to get into this topic. So yes, the activities in the past in my career led to this way and

ended up also in the activities of Gaia-X and all kind of interoperability challenges we have in different industries. Now it's Catena-X that I'm currently working on.

Karsten: Ok, great. You mentioned Catena-X. Now, there are a couple of terms kind of in that term cloud. That's IDS, I think stands for International Data Space, Catena-X, Gaia-X. Can you just give us a brief explanation of what they are and how they're related?

Rudi: Actually, this is also related to the activities regarding Industry 4.0. Many mainly hardware-driven companies have challenges to connect their devices to business applications, for example, but also to connect them among each other. Ok, so therefore, all these things were initiated, I think, about seven or eight years back now. And it also turned out that the core components were usually always data, and data is something that is very important to everyone, but in the end, also a business relevant topic. And in Europe and especially in Germany when it comes to data, there is, you know, there is always something that makes people reluctant to share data somehow. So, and therefore the IDS initiative and the Gaia-X initiative are targeting these challenges to be able to share data accordingly, so that you are also able to monitor and control who is using your data and out of the Industry 4.0 Initiative, the idea was to be able to apply the European data protection policies to to any business, and therefore, one initiative that fit into the activity or into this initiative is the International Data Spaces and therefore everything comes together, then in the end, in the Gaia-X Initiative being initiated by Germany and France. But in the end, it's a European, I would say, approach to manage data accordingly. Ok, so in short, the IDS is the technical framework, Gaia-X is more the political and legal framework that they are trying also to manage, not everything, but at least also to initiate several things. But in the end, everything is related to the digital transformation so that you can indeed make use more of the data and control how data is being processed in the end.

Karsten: Just one thing, so we don't forget: This is an Open Source podcast. Are all of these being run as community projects, or which parts of them are open?

Rudi: All of them in the end, okay? So, this is also one thing that many companies, but also, of course, governments have learned, I would say, in the last 10 years, that most of the problems can be addressed together. Not only in working groups when it comes

to standardization, but also to work on the solutions together. And I think most of us have noticed that the Open Source community is getting bigger and bigger and not only limited to Kubernetes or other interesting application solutions out there. But in the end, we have accomplished a lot and therefore, this approach, Open Source community approach, has also influenced many of the discussions in, I would say, almost every industry. And what we are now observing is that many companies, are open to collaborate also with others, even with competitors. And in the automotive industry, where we are at the moment regarding Catena-X, we even have competitors, and not only in the in the supplier area, but also in the in the OEM space, being interested to join forces and use the outcomes of the community activities then in their separate environments in the end. Okay, so that's, I would say, a learning and a development that took place, especially in the last five, six years.

Karsten: Okay, just to make sure that I get the right understanding now: IDS, the International Data Space is really the core technology concept or concept we're talking about, right? That's the "How do we just share data among different organizations?" And then Gaia-X is basically "Add some, I don't know, legal stuff and business agreements and whatever on a European general level." And Catena-X is then even more specific to the automotive industry. Is that right?

Rudi: Yeah, you're right. IDS is indeed the technical foundation we started with, but it's of course not limited to it. And Gaia-X, I would say, is the group that is bringing different companies and governments together, or initiatives, I would even say, because we have all kind of working groups all over the place in Europe, for example, and with Gaia-X, there is indeed a group trying to connect the dots in the end. So, therefore, also when it comes to regulations and some policies, they want also to address, they start the discussions there, but bring those discussions then indeed to the dedicated governments or even the responsible ministries. So, that's their role. And Catena-X is a lighthouse project that will address the Gaia-X approaches and will apply the IDS framework to the automotive industry. That's how everything comes together.

Karsten: And that then, as this is more specific than Gaia-X, is limited to the European automotive industry?

Rudi: No, definitely not. Especially for us, as SAP, being active all over the place in the world, it was just always an intention to be able to get this approach to anywhere where it is needed, and therefore, we started with things that are essential for us in Germany, in Europe, because we have the respective working groups and also the governing bodies in our places. So, therefore, especially Germany and France. And, therefore, we have the backing there and we started there, and then we apply these approaches to North America, South America, and Asia, especially. We don't really have a concrete sequence how to apply this approach later, but definitely global oriented.

Karsten: Makes sense for a community-run Open Source project to not limit it to regional, of course. Now, we have mostly mentioned that SAP is involved, that automotive industry is involved on the European level, that government agencies are involved. But I think who you are mostly working with is also the Fraunhofer Institute, right? Who is another player in this? And then we have the Hyperscalers in that as well, right?

Rudi: Yeah.

Karsten: Both correct. And so, what's the direct relationship with the Fraunhofer Institute? What's their role again?

Rudi: Fraunhofer, especially Boris Otto, initiated the International Data Spaces idea years back. I remember several discussions in the past with him out of the Fraunhofer Institute. I would say Fraunhofer invented several things and brought things to the market also. But now, the players in the market are picking up that topic. The Hyperscalers, like Azure and Amazon, in our case, but of course, also the automotive makers, so to say, because they are interested in that approach to be applied in their supply chain. So, the idea that Fraunhofer initiated fits the needs when it comes to data sharing, data governance, that is currently not really in place, and therefore, it is a topic that we are jointly trying to solve. Also, again, in a jointly set approach like, for example, in the Eclipse Data Space...um...the Eclipse Foundation, we call it the Eclipse data space connector currently, we are using to be able to enable everyone to connect to OData space, for example.

Karsten: Okay, so here we're getting a little bit into the technical background. You said the eclipse data space connector is one part. Can you explain a little bit more about that?

Rudi: Yeah. You can compare it to your phone. Right, so I guess almost everyone has a phone now, and if you want to talk to anyone in the world, you need your phone to interact with him or with her, right? So that is something that you also need to be able to connect to our data space. And we call it connector, in our case, when it comes to the IDS names or topics. It's somewhat similar to an SAP connector in the past, where you want to connect to an SAP back end, for example. Ok, something like that. But this connector is just the so-called entry point to be able to connect to server components like, for example, a broker or even clearinghouse data, where commercial and legal regulations are maybe even reflected in so-called rules or policies, and those policies are being enforced in the end in connectors. You, Karsten, you can, for example, decide, who is going to be able to process any of the data and for how long. And if you want that data being deleted with your, yeah, with your decision, that data is being deleted wherever the data is being used. So, that's the idea. So, the connector is essential, but the further data governance relevant components are also coming into the game, but that is something that's going to be applied later.

Karsten: So, this would all be concepts that we need when we're building marketplaces that run across organizations, layered applications where the services are owned by different providers. Both of these and anything else?

Rudi: The idea is indeed to be able to set up your data space wherever you want as a kind of a decentralized approach; that you can share and manage data however you want to, so that you can, for example, create your own ecosystem in your environment, in your local environment or even in your industry, or even with your suppliers. So, it's up to you how we are addressing this in Catena-X, for example, is that we are able to manage this data sharing, data governance in the supplier environment where the OEMs and the suppliers can interact with each other without having a centralized component. So, that also means that they are able then to share data in a different manner than in the past, that even Tiers five or six can interact with the OEM. While it was not really possible in the past, because with these mechanisms that we are going to put in place, you are capable to manage with whom to share the data with, so that

you can indeed even organize accordingly how to expand the data access to your data that you are not maybe even able to do today. So, that's also one intention to be able to have more transparency along the supply chain and to be able to make more from the data.

Karsten: That just makes me wonder one thing, because for these kinds of collaborative or data exchange scenarios, of course, you need more than just the data connectivity; you need a central identity or some identity and authority services and so on. Is that all also built in the project or are you making use of existing things?

Rudi: It's a mix of things. I think we need to be frank here, that server things are evolving, especially when it comes to SSI or, I would say, most of us are somewhat familiar with SSO as "single sign-on" And when it comes to, decentralized IDs, the good thing is that a lot of technologies are evolving at the moment that we can adopt and apply to this. I would say hybrid setup, so it's not limited to IDS, which we started with a couple of years back. But in the end, it's also a transition to, I would say, our mix and hybrid set up anyway in the near future. So, including blockchain or blockchain components that are also evolving nowadays. So, it's a mix of things and therefore, in the end, to be able to address the challenges, it depends, of course, on also on the scenarios. Not everything fits to every scenario. There's not really a one size fits all approach. The core components are somehow the same, and in several areas, we will be extending the services or applications accordingly. But also need to consider that the technologies are evolving and SSI, for example, to be able to identify someone more easily in the future is something that is on the list.

Karsten: And in Catena-X, it's usually rather well specified scenarios you're working towards or is it more on a generic level, very simply put: "Here's some data that I don't have to keep a secret and then someone else might want to use, so I'll just kind of connect it to Catena-X", or is it very specific?

Rudi: I think we can say that currently, we are focusing on scenarios that are close to things that most of the participants are familiar with, including transparency or traceability, for example. But there's definitely some scenarios that we are, I would say, we are able to share data better than in the past, and to address them like demand and capacity planning. That is including definitely some sensitive data, for example, because

not everyone wants to share their data with others along the supply chain in order to be, you know, to be too transparent. In the end, for example, to be able to share with everyone along the supply chain, you can decide what to share, and you can also more easily support scenarios like, for example, circular economy, where it is almost impossible today to get data from the end of the supply chain back to the first participants of the supply chain, like the engineers or even the initial or the original vendors of server components. And with that, you have more information, and you can apply more machine learning and all relevant components to the data to create more contextual insights. So, IDS and Gaia-X approaches allow us to think about more opportunities, but the processes themselves will not really be changing, you know, compared to former processes. So, therefore, we stick with existing ones, but we also apply new business scenarios along the supply chain.

Karsten: I understand that something is a little bit of a contradiction at first: On the one hand, we're enabling more collaboration on data, like, for instance, for modern concepts like circular economy sharing, sustainability information and so on, right? On the other hand, the whole thing is about remaining in control of who handles the data, right? As in: This is a European project that kind of tries to make sure that we go by all the European data privacy and everything standards. Right?

Rudi: Yeah, indeed, the thing is, I think if you compare it to nowadays, if you are sharing data, you give it a full blown approach or there's a cumbersome task to pick the data that you want to share with whoever you want to share the data with. So, and in the future, the thing is that it's not only about the control who is going to process your data, but also to be able maybe to make more out of the data. Data is getting more and more meaningful, not only the privacy relevant data, but also all kind of data that is being generated by customers, by consumers, by others along the supply chain. So, the combination of that data also to monazite that data, whatever it is, even if it's just, you know, a bunch of data you want to provide to enrich, you know, machine learning models, to improve them, that is something that has not really been addressed accordingly so far. And with these technologies that we are working on and that we want to apply to Catena-X, we have more options and more governance in place also to be able to apply those approaches to business relevant scenarios.

Karsten: Would I be right if I said that IDS and Catena-X are basically trying to take care of that? You don't have to think through the "how do I provide data", and "what are my control mechanisms, how this data can be used and by whom?" I don't have to do this case by case anymore, but there is simply a framework that helps me provide it in the exact way that I want to provide it. Is that would that be correct?

Rudi: Yes, indeed. You can compare it to your profile that some of our, whatever, you know, internet data collectors, whoever it is, right. I think you don't really have to control how that data is being used, right? You'll see it with advertisements coming into your inbox or, you know, to your home by mail. So therefore, I think that can be controlled better in the near future with those mechanisms that are going to put in place. And I think this also allows to create more contextual information beyond maybe the domain that you are applying the data to. Therefore, I think, it gives us more options to take data, maybe, to another industry, to another partner or whatever data can be used for.

Karsten: Okay, makes sense to me. But one of the terms that is also about in this context is Gaia-X, of course. And I just seem to remember, wasn't Gaia-X, in the beginning, kind of thinking even bigger than that, and was thinking to replace the very well known, globally accepted search engine with a European version, basically?

Rudi: Yeah, indeed, that was the initial thought. I was also part of that discussion and therefore to cope with the well-established Hyperscalers, it was, I would say, even at that time, almost impossible to think of. But I think we learned a lot during this journey. And now this approach is not really on the list anymore. Okay, so, we removed it from the agenda. But still, the approach applies, though, to this overall set up and the Hyperscalers are also welcome in this in this journey. But the European idea to have the privacy policy in place, being able to control the data, accordingly, is still made in Europe.

Karsten: Now that you've said it, that that global idea or that large idea is off the table, we can name it. So, there is not going to be a European Google in the scope of Gaia-X, but rather some more specific business oriented offspring from it, like Catena-X, for instance, which is the most practically relevant these days, right?

Rudi: Yes, indeed, and there is already several other industries going into this, and I saw it at the last summit a couple of weeks back. So therefore, it's not limited to automotive. Our automotive project Catena-X is definitely the biggest. And if we are going to deliver something, something meaningful very soon, it can be adopted by other industries, also. So, that's a goal.

Karsten: When we talked, I wondered, okay, so these are open source projects now. We have different players, we have the large, currently European, German automotive companies and we have Fraunhofer in it, we have SAP in it, we have some of the Hyperscalers in it. So, when it comes to open source repositories and open source communities; who runs those? Are these representatives of the industry players or of the Hyperscalers or of SAP? Who is community manager for the open source communities?

Rudi: Let's maybe take the Eclipse Data Space connector as an example, because that is something that is being run outside of the Catena-X program but contributed by people within the Catena-X initiative. So, therefore, it is an open source project currently being led by Fraunhofer colleague. And, therefore, I would say, in a neutral manner. Okay, let's put it this way. And there's also some other players involved in this eclipse data space connectivity that are not part of Catena-X. So, this is already an open source initiative. Even if there is, I would say, 60-70 percent of contributors coming from Catena-X. But the goal with the Eclipse Data Space connector is to be able to use this Eclipse Data Space connector also in other industries, it's not limited to automotive. We consider it a real open source project that can be applied to any industry. And, as I said, not limited to Catena-X. And the other project that we are going to start also out of Catena-X will have, of course, a branding of Catena-X, to some extent, I would say. But in the end, it is not limited to Catena-X.

Karsten: And is that a role that the Fraunhofer Institute takes quite frequently in all this, that they are basically the mediator between the different specific profit oriented organizations?

Rudi: Yeah. I think you can put it that way, because, I would say, most of the concepts are coming from Fraunhofer Institute anyway. And therefore, those respective institutes are involved in the different projects and so, therefore, adapting a big role.

Karsten: How about SAP then? What again is our role and how are we doing? I mean, usually in SAP's old world, basically, we're used to being the star of our planetary system and here we're like an equal among equals, I assume. Or maybe sometimes even we're the planet and somebody else is the sun.

Rudi: Yeah, I totally agree. I would say this journey has definitely been interesting. If I remember my time in consulting where it was not even allowed sometimes to make use of open source software or third party software sometimes, depending on the customer and or project, this has definitely changed. And I would say that SAP is one of the biggest contributors in the open source world anyway. But with regards to Catena-X, the thing is that it is also a new, I would say, a new approach that we are targeting anyway, and we need to address this topic, these approaches and implement them for our applications, for our solutions. So, therefore, it made sense to join forces with the others that are facing the same challenges. And we were asked anyway, even if we had already started working on these topics, also the others. Therefore, it made sense to combine our activities to, yeah, to put together our learnings and to get the most and, I would say, our best outcome out of all those things that we are bringing in. Because it's not limited to cloud native approaches. It's definitely a win-win-situation for everyone who is part of this journey.

Karsten: I guess that's usually the idea of open source projects, right?

Rudi: Yeah, I totally agree, but not every open source project that you can find on GitHub or anywhere else is always fulfilling everything. Especially when it comes to enterprise readiness. You definitely usually have to put some data on top of it, and you don't really know if it's two days or two hundred days. And now, you know, being part of the approach right from the beginning, we can also apply our quality and product standards that we need to be able to run the solutions or the services accordingly in our environment.

Karsten: Might even become an interesting quality attribute, that if you're somehow involved or used by Catena-X, then you probably fulfil all enterprise readiness standards, because if you're being used by the automotive industry and SAP and all the Hyperscalers, and Fraunhofer Institute, who can agree on your service, then you've

probably done that decently. But something else: We're approaching our time that we're headed for. Who should be interested in this and where do they find more?

Rudi: So, I think when it comes to the overall topic, when it comes to data, I call it the data sovereignty in the end. You can find a lot about this on Gaia-X's home page or even the German ministry. And when it comes to Catena-X, the landing page is where you can find a lot of our ideas that we are trying to solve and to work on in the upcoming years. And yeah, in the end, many things were also released as blogs or open source. I think you just need to look it up or use your usual search engine to figure out what we are working on regarding data sovereignty, data governance, I would say, in a more decentralized manner. And I think there is more to come very soon. But I mean, to be frank, this journey has just started, and the group of participants is, I would say, not that big. We are I think now 300 or 400 companies, but this group is definitely going to grow, and we can expect more to come.

Karsten: All right, so basically, apart from the links that we always provide with the link to the podcast, just use your search engine and try Catena-X, try IDS, or try International Data Space because IDS is probably not a completely disambiguated acronym. To sum it all up, what would be the two or three key takeaways that you would want everyone to remember from this podcast?

Rudi: I think what is definitely worth taking away from this podcast is that the interesting thing is the different kind of challenges are being addressed by people that have been competing with each other in the past. They are now working together somehow. It's not limited to the automotive industry. I had discussions with all kinds of customers from different industries in the past that are having somehow similar issues, but, in the end, are interested in joining forces with others. And one reason is definitely related to interoperability and integration into all kinds of scenarios or businesses, or even to be able to be more agile or resilient in the near future. So, therefore, many things need to be addressed together because the challenges they have are somehow similar regardless of what business you are in. What has also been observed in the last couple of years is that it is getting more and more decentralized and, of course, applications need to adjust. Lighthouse projects like Catena-X that are deriving ideas from Gaia-X, and IDS, for example, are working on open source projects that can be adopted and can be used by other industries in the near future. That means all these challenges I

have mentioned can be solved with solutions coming from Catena-X, for example. But there's definitely coming more from the Gaia-X initiative, pushing for solutions to be provided also by other industries in the very near future. And one more thing: You'll find those projects, of course, on GitHub or if you use your search engine, again, I think, you should be able to find them.

Karsten: So, for all developers who we got interested, do check it out, do get involved. Thank you very much, Rudi, for today, for being our guest. It was nice to have you here. Let's say goodbye now. Bye.

Rudi: Yeah, thank you.

Karsten: And thank you all for listening to the Open Source Way. If you enjoyed this episode, please share it. Don't miss the next one. Actually, it's slightly related because in the next one we plan to be talking to Mercedes. And it will be published just as all other episodes on the last Wednesday of the month. You'll find it on openSAP and in all other places where you usually find your podcasts, like Apple, Spotify, and so on. Thanks all for being here and hear you next time. Bye.