

WORKSHOP 3: Terms for data in contracts
9 February 2021

**Roope Pajasmaa, Platform of Trust: Experiences from data licencing terms
(translation of transcript to English)**

"In my opinion, the MyData entity and concept is something that I would encourage different operators to get familiar with, because it could give them the opportunity to licence the use of their own data and the governance of their own rights. Here's an example of it – as I happen to be familiar with it – the (-) project or Creative Passport project."

It's very nice and interesting to speak about this from a somewhat different point of view. As Anna said, my background is in the music copyright management organisation Teosto, and I see there are many familiar people and ex-colleagues participating in this call, so it will be nice to hear feedback from them as to whether I'm making any sense. My presentation is actually a continuation from the end of Olli's presentation.

So I will discuss a few practical use cases to describe how issues pertaining to fair data economy have been solved in the real-estate and construction field. My material is in English, and available on the Dialogue portal. As Anna already told you, the material is available afterwards, and in case any questions come to your mind as to how Vastuu Group and Platform of Trust have done things, it is – of course – allowed and desirable that you contact me. Briefly, I want to tell you about my background and how this is related to the IPR field. So I have a long experience from the OP-Pohjola Group where I did and managed development of digital services. Then I proceeded to become the development manager for the copyright management organisation Teosto in 2014. In this presentation, I'll try and link the known challenges related to music copyright with the opportunities that I actually already see existing in the world currently. And finally, before proceeding to Platform of Trust, I worked as the CEO and leader of a platform company founded by Teosto for a couple of years. So from that context, I'm trying to tell you the story here.

Data sharing is strongly linked with the music field. Everyone who's deeply competent in that know it is an extremely notable problem – how to share the data. The work that I did and that I got really into during my final years was live-music data sharing in particular. In fact, when I started solving the problems on a very detailed level, the reason why I decided to join Platform of Trust was that Platform of Trust could offer many key solutions to data governance in the music field as well. We were developing a service called concertify.io – a service that is still up and running. About Platform of Trust as a company – I have told my colleagues that we have successfully kept this as a secret. The first time I heard about Platform of Trust was only around October or November, the reason being that the project has been done in a very intensive manner, creating the technical acuirements that I will soon tell you about. But Platform of Trust belongs to Vastuu Group, the basic activities of

WORKSHOP 3: Terms for data in contracts 9 February 2021

which include supplying basic ICT services to the real-estate and construction field. One example is the Valtticard that is needed on construction sites – a card that must be possessed by all people who enter a worksite, on which the person's identity is authenticated, as well as the company that they represent. Vastuu Group supplies such services to the real-estate and construction field. Platform of Trust as a company, in fact, began as a project or movement, and its history goes way back in projects that started during Juha Sipilä's government, concerning fair data economy and so on, at which point operators in the real-estate and construction field ended up thinking about how digitalisation could be sped up in the real-estate and construction field. First, the project got started, a steering group was formed, after which they decided to corporatise via Vastuu Group. Since then, the company has been granted the growth engine funding, a 5-million-euro capital loan, the terms of which, in fact, have been practically fulfilled this year, in 2020. So the background is strongly linked with the real-estate and construction world, and since Olli talked about trust and how to build it, practically when it comes to this entity, it has been built via strong steering group activity in which different operators in the real-estate and construction field have come together and started solving the challenges. Through that, concrete use cases have been found and executed.

Olli told us about the networks and the rulebooks; Vastuu Group and Platform of Trust have actually followed and been involved with all of those. So we're involved in IDSA among other things. GAIA-X, too, and what is a particularly strong development area for us is the city domain, so we're also involved in a network called OASC – Open Agile Smart Cities – because especially in built environments and city environments, there are many opportunities to share data related to built environments.

We are also involved in MyData Global, so since we will soon start talking about how personal data is related to built environment, we are developing and, in fact, already have existing solutions for that. So about MyData, -, when we start talking about data sharing, we soon start thinking about how data sharing is related to personal data (regulations) and things like that, so Vastuu Group is indeed also a MyData operator and was practically chosen a couple of weeks ago as the MyData operator for the City of Helsinki – which, of course, is a notable milestone for us as a company but also for Finland per se: now that we're starting to build MyData entities and things are progressing in Europe, there is quite a notable and concrete example that can be shared in the networks.

Olli also told us about the data lakes and stuff and what the business problem with data sharing currently is. Data lies in different kinds of information systems, it lies in different generations of information systems – and there are many expenses from that. At some point, we calculated that an information system project in which you make some sort of an API or revise information retrieval can cost approximately 300,000 euro, in addition to which you have the upkeep expenses. That is what we at Platform of Trust and as Vastuu Group have started to change and what we want

WORKSHOP 3: Terms for data in contracts 9 February 2021

to change, precisely to lower the expenses, so that the data wouldn't be siloed anymore, and first and foremost, in order to build new future services.

So at Platform of Trust, we have built data-sharing capabilities. In Olli's presentation, there was a chart of the connectors, and what Platform of Trust is at its simplest is a set of connectors, procedures with which our platform can be connected with basically any source of information and information system, instead of having a need for doing a big, massive project. We have a monthly fee for this. It's the 1,400 euro mentioned on the slide; in my opinion the monthly price per se is not essential, what is essential instead is that there already exists a completely new way of retrieving data from the information systems and also share it and licence it. So to those who may work with bigger information system projects, treasurers in particular, I would say that the good news is that we will soon find fair and good ways of sharing data also in a cost-efficient way. So our current price estimate for making a so-called connector into a real-estate automation system, it costs around 5,000 euro, but I know that the price will go down so much that, in the future, around 1,000 euro is enough to make these so-called connectors.

And that will change the environment as a whole, in that the data and sharing data becomes the main thing and added value, when you get rid of the big IT projects and integration expenses in particular – you get to talk about the real deal. In the Platform of Trust concept, we're involved with built environment, and on the left-hand side of the chart there is a picture of a city or a part of a city where there are buildings. There may be, say, a stadium, there may be shopping centres, there may be residential real estate. All of those include, in one way or another, especially when it comes to newer buildings, different kinds of real-estate automation systems. In a way, real-estate automation systems are standards – there are strong operators in the market that one could say exist in most real estate. In our concept, we have with our own product development investment already created so-called connectors into these systems, or alternatively, we have created, according to our client needs, into systems that don't exist yet. But in the big picture, any real-estate or construction operator who wants to get data from their own systems or other systems, can connect to our platform where we currently have approximately 30-40 sources of information ready, and in the best case, sources of information are already used by them. We are carrying out the compatibility regarding data; we also have the trust capabilities – in the form of agreements, for example; we also have mechanisms and platform solutions for productising data of which I will soon give you an example, or I'll show you what it is.

On the left-hand side, you can see the data owners and data administrators, and on the right-hand side, in our network type of structure there are different companies as service providers who make data-based user interfaces, and then, for example, AI solution suppliers. And with our solution, we are connecting these operators with each other. At this point, I must stress that we do not store this data in our own systems, rather the data either stays in the client's own basic systems or, for

WORKSHOP 3: Terms for data in contracts

9 February 2021

example, in the client's data lakes which Olli discussed. And with the help of the connectors, you either retrieve the data or provide and also licence the data into the operating systems that want to use the data. Senate Properties is a state-owned real-estate holding company where there currently is a system where our platform helps to transfer this so-called... well, BIM stands for Building Information Model. The building information model data goes through us, different temperature data, carbon-dioxide data, service requests, and feedback requests. Currently, this data goes through us. At the bottom, you'll find some companies from whose systems the data may come from, or alternatively, since they take care of the Senate real estate, they can also make use of the data. Well, actually here, before we move on to the next picture, we have a practical visual example of how carbon-dioxide data or temperature data is utilised in the Senate solution. So there is a building blueprint like this from which the necessary maintenance personnel or the people in charge of real-estate automation can visually see, say, what the carbon-dioxide emissions of a single room are. This data comes from the real-estate automation systems that already collect this data. Therefore it is also possible that lots of wildly different data is collected from just one property, and based on that, they do even more accurate conclusions. For example, cleaning needs can be predicted with the help of motion detectors that detect motion or use of the space – whether anyone has entered the space – and then the exact same need for use of space can be modelled by adding, for example, data from a lighting automation system – whether the lights have been on. So there will be more and more data sources based on which one can predict the use of space, which – of course – will make it more accurate. Based on that, it is possible to develop different kinds of commercial services. In this case, for example, we may notice that a certain space has been used on a daily basis, in which case it can be cleaned every day, or in case we notice a space has not been used in a week, the cleaning cycle can be made more infrequent, say, once a week. They're business decisions, of course. A practical example is the Turku Civic Centre and also of how operators have licenced the data among each other. In this case, I have chosen an example that is perhaps a bit easier, so Senate Properties is the publisher of the data, but in fact, Senate Properties is also partially the utiliser of the data. RTK-Palvelu who maintains the real estate is the utiliser of the data.

But this is the network of operators who have agreed together on the use of data. The Granlund real-estate automation system is one of the basic systems through which data flows. Data flow in addition to technical data flow, an essential capability that enables everything that has been done in the real-estate and construction sector and that was obligatory to do is data harmonisation as well as data product solutions.

So in practice, we have a so-called harmonisation team, an ontology team – a group of people who, whenever a new source system is connected to our solution and they want the data to flow, they make a so-called harmonisation model, an ontology model, so that the data is compatible. So we have the ontology team who models the dataset in such a way that, say, in case we have five different systems with carbon-dioxide data, and in the information systems, the carbon-dioxide data can be

WORKSHOP 3: Terms for data in contracts 9 February 2021

presented in slightly different ways – it may be a thing as minuscule as a comma in a different place or a unit being used in a different way, so the data harmonisation team harmonises the data in such a way that the data from different data sources is machine-readable in a consistent manner in the utilisier system.

In case you're interested in our ontology model, at the bottom of the slide there is a link to it, and our ontology work is open in the sense that ontologies have been described exactly like the connector functionalities, based on open-source functionalities. So it's there, free for anyone to utilise.

Here is an example of our data sheet, how it's defined, so here's an example of what information a data sheet includes. There's correctness verification level, availability rate, update frequency, accuracy, precision, faultlessness rate, completeness rate, security level, harmonisation level, and what it is interoperable with. And an essential part – is data product licence – terms and conditions – in which it is defined where data can be used. So there's an agreement on and licencing of, for example, geographical area, purpose of use, whether it's for internal or external use, whether it can be reused, whether it can be resold, whether it can be combined with other data sources, how the data must be and can be stored, confidentiality, personal data, licencing, potential MyData consents, channels for licencing, and level of data sheet certification. So in practice, we have this sort of a data product sheet.

Then if we have a look at data licencing, here's a practical example of that. So we can simply search for open data by certain licence terms – buy or get weather data, for example. But when you start doing, well, multi-licencing or multi-permission, you must agree on the rights and means of retrieving data. In practice, licencing data – depending on the case – makes it more complex. So what does this type of philosophy enable? If data is licenced and shared as openly as possible, what will that enable business-wise? Firstly, it benefits data owners or data managers. Like I already mentioned, distributed data can be governed better, it can be done in a more cost-efficient way, and it can also be done in a more agile way.

By combining different data sources, one can bring about added-value services or build completely new types of services. I think that's the most interesting part – when you start talking about different types of practical solutions with fair data, by combining data in novel ways, you can build something completely new. As I said, in the real-estate and construction field, we currently do, model, and have already built completely new types of needs-based services, such as the cleaning solution service. Another thing we're currently building is different types of situation picture or situation room concepts as to how one can, by combining different data sources, get a situation picture of, say, what the current situation on a construction site is. Combining different data sources enables all that. I will tell you, and if you want to have a look, this is also found on our website. So we also have a readymade concept and licencing mechanisms, so that any data owner can also sell their own data on a marketplace. We have some examples of existing data sources there, and this is also

WORKSHOP 3: Terms for data in contracts 9 February 2021

an interesting novel business opening – into all markets, in fact. There already are some data marketplaces – here is one example – and there will be more and more. That will open up completely new business opportunities for one's own data. I picked a few of my own experiences as to how this philosophy and what Olli told you and what I told you about the ways in which you could share data (on) the connectors, approaching things, I guess, from a new perspective: what it could mean in the music field in particular, for that's what I'm familiar with. In my opinion, the MyData entity and concept is something that I would encourage different operators to get familiar with, because it could give them the opportunity to licence the use of their own data and the governance of their own rights. Here's an example of it – as I happen to be familiar with it – the (-) project or Creative Passport project. These projects seek to manage the identities of music authors, and there are quite many rights, but in order to be convincing and reliable enough, I think the projects should start applying the MyData principles that are being built in different networks – i.e. the rulebooks. Data sharing from the source systems in a novel way would also enable novel services. In my own work at Teosto back in the day, I spent some time thinking about, even though I worked with music copyright management, there are individual right holders and people or groups of people who have other rights besides music rights, and by sharing data or by retrieving data from different kinds of information systems, one could build these IPR stack services. As I mentioned, for example, music copyright management marketplaces are already plausible – there are solutions for that in other fields, in case there's need for that. But I guess my last point is related to my experiences from the work I did earlier, and now that I have worked at Platform of Trust and seen the work done around this, so when you're doing something new in a completely new way, some strongly motivated operator is needed for orchestrating this whole (and facilitators). So there's a need for someone who is brave enough to take a product-development risk, brave enough to build something new, - the technology side is relatively easy here – there are standards for that – but managing the change in order to develop new services that might even break the current procedures and models is the more challenging thing, and it takes some muscle. To wrap things up, I drew – just for exercise – a chart to make you think about what a MyIPR Finland concept could be and how one could build it. As an example, by retrieving some data from the CMO's Teosto, Kopiosto, Sanasto, and Gramex systems and combining that with the MyDataWallet philosophy, agreed-upon data harmonisation rules, and that there is an operator who orchestrates all of this, one could develop these kinds of MyIPR services and view relatively easily, in fact, right now. That concludes what I have to say in this forum, and I'm also glad to answer any questions that you may have.