

WORKSHOP 3: Terms for data in contracts **9 February 2021**

Olli Pitkänen, 1001 Lakes Ltd: Rulebooks and terms and conditions of data sharing **(translation of transcript to English)**

"For each data set, different terms of use can be created. For example, whether it is allowed to use the data commercially, whether it is allowed to share the data with outsiders, and so on. And there may also be data-specific responsibilities regarding the data's correctness, quality, and so on. There's also a governance model which means the governance of the network: who makes decisions for the network, in which ways, and how new members are accepted in, and so on."

Right, so earlier, and still today in many organisations, the situation is such that there are many separate information systems. There may be one information system for HR, one information system for client information, one for ERP, and so on. And within a single information system, one can do searches and analyse data, but data does not flow very well between the different information systems. It's a situation that many organisations currently desire, and many of them are quite close to that already, so they are building big data storages that can be called either data lakes or data hubs or data warehouses, but the idea behind all of those is that there is a single big information system, data warehouse, that combines data from different information systems, and thus one can do searches there and make analyses in a much, much richer way than in just these separate silos. And that's when data becomes much more valuable. It's a one plus one equals more than two type of situation, in which the value of data grows fundamentally when it gets combined. This is something of a future ideal. This is what they're already aiming at in many places, but not much progress has been made as of yet.

So they're trying to combine data also between organisations. In principle, it could be possible to build a shared data pool, data lake for several organisations, but a somewhat more modest goal is to build channels at least between the organisation-specific data warehouses, in which case they could at least do some searches and analyses based on the data of different organisations. In that case, of course, the amount of data grows again to a great extent, and it becomes significantly more valuable. But this is only a future ideal. On the next slide, I have listed some problems related to data sharing between organisations and why it is not yet really happening today. First of all, of course, there are technical problems. Data does not flow between organisations, because information systems are not compatible with each other, the data is in different formats, and so on, data security issues remain unsolved. But technical problems can typically be solved by investing money, time and work in it. But there are also issues that have more to do with business and judicial things. Organisations often fear losing valuable data, losing control over – for example – trade secrets, in case they share it. Vice versa, there is the risk of infringing the rights of others. Like Anna said in the beginning, there can be lots of different types of rights targeted towards data. These

WORKSHOP 3: Terms for data in contracts 9 February 2021

days, data protection is probably the one that gets the most attention. The concept of personal data is so broad that almost all existing information and data can potentially be related to personal data, in which case you have to always consider at least the requirements of the data protection legislation. Similarly, copyright and related rights, industrial property rights, agreement-based rights – all of those have to be taken into account, and it is possible for one to infringe the rights of the third parties when sharing data. Another type of problem is that companies and other organisations do not necessarily fully understand yet why they should share data; they cannot see the business opportunity and the opportunity to make money with the data or improve their own operations, and so on. That's why they rather choose not to share it than take the risks. A central problem in the area is the lack of trust or downright mistrust; organisations have learnt to be careful with these matters by trial and error. As to how these problems could be solved, one central means is having a shared rulebook.

I will show you how Sitra, in their fair data economy project, have built a model rulebook for sharing data. I was involved with coordinating this work, but there were experts involved with that from several companies and organisations; there was a big group of people working on it. It's been in progress for quite a while; we have now published the version 1.2 in English, and the Finnish translation will be published soon. It's quite a heavy and big package that comments on data sharing in organisations' networks from various viewpoints. **Its purpose is to cover not only judicial issues but also business, technical, and ethical viewpoints.** The model rulebook by Sitra consists of several parts: first of all, there's a general part that gives instructions, describes what the rulebook is about, how to use it. What is very crucial to it is these inspection lists, with the help of which participating organisations can assess their own capability and what things they should pay attention to when sharing data. Based on the inspection list, we make the agreements concerning the rulebook, which – of course – are a very central part of the rulebook. I will come back to them soon. There is also the ethical code of conduct and a common glossary for the activities. At this point, I think I should stress that the model rulebook by Sitra is very generic. In fact, it does not comment on what data is transferred nor on what kinds of organisations participate in it. In a judicial sense, I think it's interesting that the entire rulebook is built as an agreement between several parties in which all parties are quite equal in principle. It is not assumed that there would necessarily be some kind of a central, dominating operator – although it is naturally understandable that that is often the case in practice. The starting point is that the different parties can act in wildly different roles; their roles may vary; the number of parties may vary; new ones come in, old ones leave; and the data transferred there can vary a lot. What this means is that, in some sense, the model rulebook is applicable to many uses and situations. But conversely, it may always require some work to make it applicable to the specific purpose that it's planned for. Next about the agreement structure. In the model rulebook by Sitra, there are several model agreement terms and conditions. First of all, there are general terms and conditions that comment on intellectual property rights, data protection, responsibilities, auditing, confidentiality, termination of agreement, and so on, and the idea is, of course, that these would remain quite unaltered in all networks. For each network, they make a

WORKSHOP 3: Terms for data in contracts

9 February 2021

founding agreement in which they can deviate from the general terms and conditions if necessary. In the founding agreement, the founding members can define to a large extent, for example, whether other organisations are allowed to join the network, what kind of data is shared there, and on what terms. If new parties are allowed to join, that will take place via an Accession Agreement. That's when the new member commits to the founding agreement and general terms and conditions.

For each data set, different terms of use can be created. For example, whether it is allowed to use the data commercially, whether it is allowed to share the data with outsiders, and so on. And there may also be data-specific responsibilities regarding the data's correctness, quality, and so on. There's also a governance model which means the governance of the network: who makes decisions for the network, in which ways, and how new members are accepted in, and so on. As an attachment, there is also a description of the network that is the outcome of going over the inspection lists. Then if we move on, the Sitra model is – one could say – quite progressive, advanced, but far from the only one in the field. And I must mention as an example the model terms and conditions by the Technology Industries of Finland, which come very handy when two companies want to share data between each other, so in such smaller-scale circumstances: data sharing between two companies. In those cases, it is useless to have such a big rulebook as the Sitra model, for it is much easier to use the model terms and conditions by the Technology Industries of Finland.

As for international references, iShare is a Dutch organisation that has built a rulebook especially for the logistics sector. The Sitra model is very generic and is not dependent on the field, whereas iShare has been designed for a specific field – that of logistics. And it is in use, for example, in the port of Rotterdam in such a way that there are different operators related to transportation of cargo – there are shipping companies, forwarding companies, trucking companies, and so on, and they share data between each other regarding the cargo they transport in accordance with the iShare model. What is quite big at the moment is International Data Spaces Association – IDSA. It originates from Germany, it started in the research world, in the Fraunhofer Institute, and has now spread to become a pan-European project, the original idea of which was to develop technological standards for sharing data. But then they noticed that it was not enough to be competent with the technology and that they also needed something in the legal side, and thus they started developing their own model rulebook. But when at IDSA they heard about the completed rulebooks by Sitra and iShare, they decided to put them to use pretty much unmodified, so they now have something of a shell model for their own rulebook that contains many references to the Sitra and iShare rulebooks.

That's how the Sitra model rulebook is now spreading to become a pan-European model rulebook. Then I will superficially and briefly demonstrate the technical side of this, so regarding IDSA, in this model, they share data between organisations. A central concept is connector, which is actually a computer programme that gets added on an information system. And the connectors can communicate with each other via the data network and transfer data from one organisation to another. There can be several kinds of

WORKSHOP 3: Terms for data in contracts 9 February 2021

connectors, but typically they make use of metadata so that they tell each other what data is available, what is wanted, and then, they are often related to rights management and access management, so the connector monitors that only the right data is transferred to the right party – nothing more than that. In the chart, the solid arrows describe transfer of data, whereas the dashed lines describe transfer of metadata, so as you can see, metadata moves in all directions, it plays an important role, whereas the data itself actually only moves between the main operators.

The GAIA-X project that is a huge European project for defining data-sharing standards and infrastructure. In Finland, Sitra and VTT run the Finnish hub, and there are already quite many organisations involved in Finland as well, and more and more keep joining up; it's becoming a big thing – how to get data moving in Europe. Technically speaking, GAIA-X largely makes use of the IDSA model described earlier, and therefore, the Sitra model rulebook is apparently going to be included in GAIA-X as the GAIA-X model rulebook. So from a European standpoint, we are in quite a central position in defining the practical rules as to how to share data between organisations in the future. And the metadata in GAIA-X, too, is very similar to the metadata in IDSA, making use of the IDSA solutions. You can also send me questions afterwards via email if something comes to your mind.

All right, thank you very much, Olli. It is so great to see, first of all, that Finland is one of the top countries in this sort of development; at the same time we talk about how the EU is lagging behind the rest of the bigger economies. So the EU is generally lagging behind when it comes to data sharing or benefitting from data, but on the other hand, the starting point is the ethical approach, an even-handed approach that benefits all parties, human-centric, thriving and balanced data economy – after all, that's a goal set by the EU. So, in that sense, it is really great that we are one of the top countries in such novel approach which actually benefits everyone in an even-handed manner. I will now give our listeners an opportunity to ask questions but I may have a couple of questions in mind, but, so please use the raise hand functionality or ask for permission to speak in some other way. Does anyone have questions to Olli at this point? It doesn't seem that way – I guess no questions have sprung to anyone's mind yet. But as I said earlier, we can certainly come back to these matters at the end – especially after the next presentation which is even more concrete, a practical example in fact.