O-P Rissanen, Ministry of Finance: open data in public administration (Translation of transcript into English)

> *"So when we're (defining) the copyright [infra] system, we should examine it also from another perspective – staying away from the world that defines it currently, so-called creative fields, because many other types of data as well which does not originate from the so-called creative fields [are needed] … As for what we do here and what we aim to do in the copyright system, I don't see a contradiction that these would go in opposite directions…"*

The original title was open data – I'm expanding this beyond just open data, because the situation is such that – in my opinion – open data can include no copyrights or anything like that. And now that I listened to this, I was left wondering if I have the right idea, so I looked up the Finnish-language definition for open data on Wikipedia, as well as the English-language one. The English one said that no copyright can be related to open data – copyright was the word used. But in my opinion, it means any rights. So in a way, that would have been my presentation: I would have stated that when the data is public and open, it has nothing to do with copyrights – except that it must be ensured that it's not copyrighted. But I did a brave thing and expanded on it, so we're looking into this on a broader level from the public-sector data perspective, and I guess currently the best frame of reference for that is the TiHA project that we have. So we have a project on utilisation and opening of data that carries out the government programme. Here are some of the work packages, and I will briefly go over this, what the idea is, and after that, the big question is: how are copyrights related to this? And that is a question which I cannot answer straightaway. So that's the idea of my presentation in a nutshell.

The first work package is called strategic goals, the idea of which is that when we did the data policy report a couple of years ago, and it was a policymaker-level paper – very academic in nature. What we're doing now is we want to make it more down-to-earth, closer to more concrete goals – what it could mean. We just published a data-policy guide last Friday. We will inform on its publication today. Anne Kauhanen-Simanainen was one of the authors. With her, we did the first papers on data policy. Now we're trying out several methods to make sense as to what our goals really are regarding data utilisation, how we can make them applicable in decision-making, for example. We are not making a strategy, rather we're coming up with individual goals – which is what we'd planned in advance. The second work package is the availability of data. Strategic goals are at a stage where we want to get the first version ready before Christmas, and then, after Christmas holidays, we'll have the top people in the ministry look into them, and then, we intend to get statements on them – probably in late February. The preparations have been open: they were open for comments on Google Docs, and so on. But that's one context, and I don't think copyrights are mentioned there. And I guess that's what we have to look into here.

Availability of data is the second package, which will launch at a somewhat slower pace. It was our purpose that we'd connect it especially with this high-value datasets, which will come about through the new open data directive. But the Commission's schedule is late, and someone might even say it's one year late: more detailed descriptions as to what it concerns should have arrived last spring. And the Data Governance Act, to which Anna also referred, is one thing that has surpassed this. But we're promoting this, and we want to consider different methods to improve data availability.

The third work package is one that arose when we arranged a series of round-table discussions for the biggest government agencies in the autumn of 2019. Biggest in the sense that they have big data pools. With them, we went over the situation, and surprisingly, it was brought up that the quality of data is a central factor in which we should invest. Statistics Finland is now doing this. Here, there are obvious things related to data quality, such as correctness, and things like that, but more complex theme – I have learnt a lot once we started working on this. It's in good progress now. So I think we'll get to hear some of their thoughts before the end of this year. But one central factor related to data quality and modern times is, of course, the fact that they make better use of it in new technologies and in new uses. It's related to AI, it's related to other algorithmic methods of analysis, and things like that. And in some ways, perhaps also to blockchains, where reliability is achieved in a different way. But that work package is very interesting in terms of our future, and it doesn't mean that they just give us the framework, rather that it's piloted and tried out, and the aim is that there will be basic principles concerning the ways in which we invest in data in the public sector.

The fourth and final work package is the closest to technology and interface, but it's particularly linked with semantic interoperability. We always talk about APIs. So we talk about a new way of making interfaces. In API, the letter P – Application Programming Interface – stands for Programme, so it means the interfaces are programmable. We have had interfaces ever since the 1960s – we have transferred data between data systems, and new technologies have appeared regularly, and every time they have changed the business model or operational model to some extent, so this will change the operational model a great deal. It is not just a technological solution – what it's about is that whenever someone searches for data from a service through that interface, it is very different from a traditional interface where you just retrieve the lump of data after which it's processed. In this one, you can – for example – define what data to retrieve. What gives a good picture of the magnitude of the change is, for example, that once Osuuspankki has embraced this model, (the interfaces are owned by the businesses). So it is so strongly linked with doing business that defines it – it is not the IT department that defines those things, instead it's the business. And the same thing will happen in the public sector. And the same thing will be done on the EU level, too. But this is the package we have. If you look at the whole, we're creating an apparatus and there are certain strategic goals that we try to reach, we get to make better use of the data. Related to that,

there are methods advancing the availability. We examine the quality of data, and we also examine the (semantic interoperability in the interface). These make up some sort of a… the way I've imagined it, we're building a step-by-step model that allows us to make use of data in a better and more efficient manner – in the public administration and in the society at large. Data Governance Act is related to that – its significance must not be underestimated, for it will be very significant, because it will likely concern all authorities. So we have to re-examine the models according to which we have given up data, and that will be quite a big thing. In any case, we are creating operational models here, too, which we will combine with European developments. So that is the work we're doing here. That smoothly brings us to the question: if we have a system that processes copyrights, how does it fit this apparatus? In my opinion, it ultimately fits it quite well, as long as we remember to make it fit – and vice versa. So when we're (defining) the copyright [infra] system, we should examine it also from another perspective – straying away from the world that defines it currently, so-called creative fields, because many other types of data as well which does not originate from the so-called creative fields. As for what we do here and what we aim to do in the copyright system, I don't see a contradiction that these would go in opposite directions. It's just that there has been… I have worked with these matters, data-related matters, since the 1990s, and what is happening now that the copyright issue is no more unrelated to other development. Anna has done a great job at keeping us awake and aware when it comes to the EU preparation and other things as well, as to what is happening around these issues. In my opinion, these things should converge and become a part of the whole. To conclude… I didn't want to ask for a permission to speak earlier when EU-level development projects was briefly mentioned, because I don't have an answer to that either yet…the initiative by France and Germany called GAIA-X which is related to federated data governance and federated cloud services. There are many components to it, it is a very big and extensive project. In addition to the French and German governments' commitment to it, there are also lots of very notable European industrial companies involved. Big, massive companies: BMW, Commerce Bank, EDF, and so on. (--) I've tried to make sense of GAIA-X during the past six months, trying to understand its significance to Finland, from different perspectives. And I would say that, in the context of GAIA-X, at some point we have to do some sort of an exercise. Especially if we start building such a system for copyrights on a European level, we should look into how it fits GAIA-X. Because the data that's there must at least include industrial rights, because I don't think the big operators will give up their data within the system without any compensation. I have not familiarised myself with it in terms of that; I cannot say what is actually happening there at this point. But that's one defining factor. As for the Commission's commitment to GAIA-X, it's kind of two-fold, but the basic message from commissioner Breton, for example, was – a couple of weeks ago when there was a big seminar – that… he didn't say that the Commission is directly committed to it, but he said that it fulfils all the goals set by the Commission, and in practice, two billion euro. And I think these two things signal commitment to some extent. But that's kind of off-topic. On national level, we'll look into this, and on the EU level, I think that GAIA-X would be good to look into.