Workshop 2: Data protection, incentives and dimensions of the copyright infrastructure, 7 December 2020

Sami Niinimäki, Ministry of Education and Culture: research, development and innovation activities
(Translation of transcript into English)

> *"At this point, we need the fastest supercomputers in the world, and EuroHPC will provide us with that. As an example of that, we can agree on IPR cooperation and smoothly between research and company operators, and also utilise a platform, such as EuroHPC, where there is also competence sharing and data services that this entity requires."*

Actually, what I can add to the conversation is examples from the university and science policy side of things – examples of infrastructure supporting research, development and innovation activities, which is funded. Mainly they're nationwide, broad, even Europe-wide infrastructures where they create a minimum framework for cooperation, I guess you could say, and for the services that support whichever whole is in question.

In fact, everything we've heard so far today corroborates what I have in mind: in actual fact, we largely struggle with the same problems. To some extent, it also corroborates the notion that it would most certainly be a good idea to cooperate on an even broader scale, and kind of…

So actually, the approach here is metadata in research and studying. And as was mentioned, we have nationwide services: we have different kinds of data pools and service entities with different angles from which to approach data, which is ultimately descriptive data. Research data pool, for example, and a website called tiedejatutkimus.fi well demonstrate how there is data collected from research organisations, there is data collected from research funders, and I guess in the future, we will get different kinds of descriptive from the research infrastructure as well. All of these, as a centralised whole, create added value that can be sort of used in steering, but it can also be used in developing the organisations themselves, and building services on top of that which we don't necessarily even cannot fully understand.

So research data pool, and the now-launched website is called tiedejatutkimus.fi, so with the help of that, we can get what the funding is targeted towards, and what level of research is done with that, and the descriptive metadata on that, gathered and transmitted by this one and only website. I guess what's good to mention in this context is the "Koski" data pool, which is more descriptive of data on study and degree registers. In that sense, also this service that extensively utilises data. In that sense, you can build a lot of services on top of that, maybe… if I were to innovate a bit here, say, you could have right to study and related benefits and check-ups, for example, so we could consider doing something like that through it. In modern times, such data pools often require separate laws, in order to get personal data

processing rights and registrar position-related things in order as smoothly as possible. That way you may be able to avoid transaction expenses – mentioned earlier – and thus, making their position official and take care of the juridical side. The legal side often raises questions – especially when it comes to research data with sensitive dimensions, legal questions often surface rapidly, creating complications, so the usability of data is weakened or slowed down at least. And let's put it like this: the maximal potential is never reached. This is still unsolved in a broad manner, really getting into the data, so we have many strategies – as has been mentioned today. Let's say, after that, in order to create something new, we must have our infrastructures in order, but we also must have the legal side in order, so that things would be as smooth as possible, so that data actually becomes oil. Whether data is actually oil – oil, after all, leads to something of an ecocatastrophe, and we most likely don't want data to lead to something of a catastrophe. So the metaphor is bad. However, open science is a basic principle that we're promoting, and in that sense, open data materials, open science publications, open research data are the starting point. To a very large extent, research funders – both private and public – require this. There are all kinds of reasoning for that; quality is one thing that should be mentioned more often, because quality – in fact – is ultimately everything in order for us to be able to do anything. Openness ultimately leads to improved quality. If we think about incentives, in open science, of course, that open materials get more citations, and in that sense, if citations are the researchers' "currency", I'm sure it's a good idea to operate openly and publish openly and share data openly. But of course, it is the nature of open science that openness is a continuum – at one end, there are materials that cannot be shared; sometimes you may want to patent things – that's strategical openness. And that is kind of everyone's own choice at that stage, but the research funder's demands must be taken into account, too. When it comes to public funding, at least when it comes to publications, and data as open as possible, as closed as necessary – to mention such general quotes.

The third point is FAIRDATA service entity, and I guess that's a bigger toolbox in which researchers are provided with different kinds of descriptive tool opportunities, data saving opportunities, as well as long-term data storage opportunities. These are basic functions that any researcher should utilise. I'm sure there are other services as well, but the idea here is that these are at least nationally available through one window, through one service, the purpose being that Finnish researchers' day-to-day work would be easier. Now that we're discussing long-term storage and long-term availability what arises is the question of IPRs, and if the research publication were a basic work for a researcher whose name is clearly mentioned and it gives them citations with time, and in that sense, it is identified with the researcher, the thing with that is that – where it's published, the publisher may want the rights to themself, and in that sense, the IPR is merchandise in a way. But things are advancing in such a way that people are talking about a (green OA right) to be included in the copyright law, in which the research would retain a copyright with the research publication, which would enable the researcher to (--) in

the future as well, so they will not lose their work beyond paywalls or into cyberspace. In the worst case, if a research publication is not long-term-stored, there have been actual cases in which they've been lost, and that's quite a waste in an economic sense. Similarly, there is a need for long-term storage services for datas, and that's when I think the IPR is distributed, at first, I'm sure, to the research group. But when it comes to long-term storage, the only solution that is actually sustainable is that the research organisation takes responsibility for the research datas, therefore being responsible for taking care of them and for curating them and such dimensions. A service entity serves the research world. In parallel with that, there is a Finnish service called Finna that also brings forth research work – it is a metadata catalogue, roughly speaking, but a very clean version, a very smooth service. On Finna, you can find different cultural materials, and their metadata in particular. But you can also find open study materials there these days. So it's a very comprehensive service; it's not on the list here, but mentally speaking, it belongs with the FAIRDATA service entity. So thus far, we have discussed different kinds of data pools and service entities.

Next, as an infrastructure – **there is this EuroHPC thing getting launched in Kajaani, a super computing ecosystem.** Why I want to bring this up is because it's a significant research, development and innovation hub and a demonstration of the cooperation done in RDI - research, development and innovation. The Finnish funding share is half funded by the Ministry of Education and Culture and the Ministry of Economic Affairs and Employment, and the rest is paid by the consortium and the Commission. It's a, let's say, 200-million-euro investment. We definitely want to advance cooperation between companies and research organisations through that, but enabling it in a new way, creating trust between different operators, and in a way, providing services that it requires. So whenever we have data, there is ultimately a need for computing in order to get anything out of it when it comes to research, development and innovation. At this point, we need the fastest supercomputers in the world, and EuroHPC will provide us with that. As an example of that, we can agree on IPR cooperation and smoothly between research and company operators, and also utilise a platform, such as EuroHPC, where there is also competence sharing and data services that this entity requires. These are the examples I wanted to bring up, and I'm sure there are things to learn from these when it comes to copyright, and in that sense, one doesn't have to reinvent the wheel, instead one can perhaps even get involved with the practices and, let's put it like this, get to the top of development faster. I don't think I have anything else to say at this point. These were the examples, and I'm sure there are others – this is not a comprehensive list. Thank you.

Thank you very much, Sami. We're in adjacent departments, and we have been in contact – usually in research-related matters and in matters related to research and copyright. But as you said, there is a synergy to be achieved from the copyright infrastructure point of view. So primarily, there is not necessarily a place for new legislation for copyright, at least not separately, for if it's there, it's legislation

pertaining to data infrastructure framework which follows the EU data strategy – in fact, the first initiative concerning the EU Data Governance Act was published a few weeks ago. So that includes descriptions of the different dimensions as to how to advance data sharing, and that will be scrupulously followed, because that is, in my opinion, very much related to the next subject: what will happen with the public-sector open data when data must be shared, while on the other hand, on the Data Governance Act, on the draft, it is said that the IP rights must be taken into account.