RIKSREVISIONEN

Intervjuanteckningar KTH17

Datum: 2023-03-22

Plats: Teams

Deltagare

<u>KTH:</u>

- Forskare, skolan för Elektroteknik och datavetenskap,
- <u>RiR:</u> Sara Monaco, Ludvig Stendahl

Start telling us about yourself and your research

Bring researchers in the domain together, but also giving advice to the to the steering on what is needed, what needs to be done. It's a very agile entity, and four years ago I think it was four years. I found that KTH has not a good situation on data sharing, so the facilities how to share data there was at the same time a top down activity. And the platform was launching some bottom up activity, asking researchers what are the needs, what are the problems and bottlenecks and data sharing. Wasn't particularly about sensitive data. But KTH needs solutions for it. In this role then, I was also placed into the IT infrastructure group to identify what are the needs, what need to be pushed forward on KTH level for this purpose. And I was also, I'm in the KTH digital futures governing board. And there's also one person from RISE. He was also helping out identifying needs when it comes to data sharing. It often turned out that security questions are somehow a burden to enable the data sharing, obviously I mean so they pointed out that we need more routines and more possibilities.

When we discuss this in the IT infrastructure group, it was often the conclusion we have X solutions for specific researchers and environments, but if a specific researcher comes then it becomes too costly for a project. And then often the solution is to keep the data at the host and send the researcher to work there.

So I know about those who were working with health data, so they usually go to the company that has the data. It was always Karolinska or the health data center with the region. So it is that one goes and has some work in place within their IT infrastructure.

And when you talk about data storage, is that during the research process in itself or also when the project is finished so to speak?

I talked about data sharing.

OK, sorry.

Because it's slightly different because data sharing gives you managing the access rights and it's not telling you where the data is stored and if the data is stored at a good place and if the access rights are managed well, then there is no problem.

So we actually thought that data storage facilities would be the problem. And when the researchers were pointing on needs of data sharing. But on our access to data or data sharing. But most of the time it's actually the knowing where the data is and the access control so. So what this bottom up initiative was leading to was to launch on the KTH Library, the software Invenio, which is hosting or guiding researchers where some data is and who can have access to it.

But this software is not... I don't think that they are managing sensitive data, so this is just noncritical data, which is within this environment. So this was then taken or driven by Rosa. But the platform was helping in the years to identify who... we basically need these catalog function to know where the data is and to manage the access control.

But we still have some researchers who want to have the data work with the data here. So I can say maybe with something now for my own experience.

When I was working with region Stockholm on health data and I'm working on privacy and trying to identify... creating synthetic data based health records... usually I'm working theoretically. That means I develop algorithms and bounds and do not need to touch actually the data itself. It's all on the on the paper. And that's what also we learned later. Often we are... the researchers here are more often developing the methods and tools, and they maybe at the end want to test it. It's not really that we are working on the data is the research... it's working the algorithms. So I was testing some algorithms, I was doing that in a master thesis and this master thesis was then done by Health Data Bank, which is located at Stockholm University, so I was sending the student to a colleague at Stockholm University who has this data bank and this is a common model where people is sending the researcher to the place where the data is hopefully safely stored. Hopefully.

Would you say that this mindset of thinking of sensitive data, if the data is open or not... is that something that is your experience, that there is something that all your students and colleagues think about when they start a project or start a...?

Yeah, we pretty fast always... if you don't think about it, you're pretty fast get to know about it. I mean I work in privacy, so that's for sure that I know about it. And if I have other colleagues here who work with Karolinska, I mean all of them... We don't have a solution where the KTH... I don't think we have a good solution where the infrastructure is providing researchers data space for the individual problem. The solution that I always see is that they send their researchers to the host, so to Karolinska. And since, yeah, I experience all my researchers that they do it in this way.

And I would also say that my colleagues and also at KTH digital futures, I think the people are aware that we are not... our computers are standalone managed computers directly connected to the Internet. So we are aware that we are not having the security measures in place for, for having the data on the computer. Of course, there might be people who not, but... I haven't heard about somebody who is uh... Actively telling me or telling me directly I'm sitting on the database which is... Uh, critically in the office. So i I more often hear from the people that they don't even want to have the data. Not too happy with this, so it's also convenient not to have it.

I assume you have funds from VR... how do you make your research open access?

Another example I can describe... with VR it's a very theoretical research, so there is no data, it's formulas, so therefore I don't see a problem there.

But I have a collaborations with KTH Living Lab. They have the data in the basement, collected for me. Smart meta data. And the problem is on smart meta data privacy. So the people living in this environment use energy appliances and actually to create a research data set. To develop algorithms using this research data set, we also ask students, or we asked the people living in this environment to write down when they were using different appliances. So we have actually the smart meta data and we have from each (also) socket data. And the purpose or the goal of this is that we wish to publish this data as you just said, make it open data. And to do so, I - of course if you work on privacy - you know that this is not what you should straightforwardly do. So we made an ethical application. We sent in an ethical application. And the ethical application was rejected. Uh, and the ethical application was rejected because it was saying that they don't see any ethical problem there.

Ah, it was rejected because there wasn't weren't any ethical problems or OK?

Yes. So we actually contacted before and I contacted before an ethics professor at KTH, Sweden over. He was in the ethical board. He was telling me the way how you should do is you should submit these ethical application in the best case they reject it because then they have decided on it that there is no problem and then you can go ahead. And if somebody afterwards ask you, you are actually fine.

As a as a researcher, I actually think the ethical board is not correct on the assessment that there is no problem, because I actually think the smart meta data reveals lots of sensitive information.

Uh, so we still try to anonymize the people. We, I actually don't know who was living there. We had a person in between who was collecting the data. And we get the data and we will make the data available, but only for those who have a a research purpose, for research purposes. So we wrote everything how we are going to do it, but the ethical board was deciding that this is non critical.

Umm, where I actually think that they may have done a tool. They are not up to the date, I would say.

Interesting.

But yeah, so if you ask me about the question regarding open data and which of course in health and that's what the region Stockholm health data center has a problem, they want to make the data available, the politicians want to make the data available.

But the GDPR, of course against making everything available available. So there is a thin line where you need to identify the utility so make it as open as possible and as closed as necessary. Yeah, and I think there is a, it's a tradeoff. I would not. It's a contradiction there. It's a contradiction. If you believe it's open. Open.

Would you say that you are mean--- No, this is actually in your research. I mean, you're doing research on this, so you have a lot of competence and knowledge in this area. But if you compare to other research, would you say that the responsibility lies on the individual researcher to sort of know what the laws and regulations or... could you do? You get any support to sort of navigate in this area?

So they exist in other platform, which is the health technology platform. And the health technology platform has established now. Support where researchers can get support to apply for the ethical application. Because I think it's always... To sort out the what is possible and what is not possible and what measures are necessary and not necessary, I think it's good to have this ethical application. So with the platform, the health technology platform is offering this support to the researchers and they communicated in their newsletter. If this reaches all researchers, I don't know, but it's on the going. Our newsletters from the platforms go to all researchers. In principle, everybody can actually see it.

Uh. Rosa knows it. Since last year, we have a data steward, which the platform was pushing for some years before we had the of, of course, this GDPR data officer who was also at that time running events where people were informed about it.

I would say if one is not ignorant then people can know that they support function. And then...

The information is there.

Information is there, but then on the same time the solutions offered by the IT infrastructure might be sometimes not good for everybody. And I know from the IT infrastructure group that it's a resource question. Who should cover the cost for this specific solutions? Is it the researchers or should it be the university offering? And if the researcher not needs to cover it, then the project budget is going too high. Or it's going to take lots of costs in the project budget. And I don't know if they have found a solution, so priorities are to have first the data steward, to have the information distribution, to have these... It turned out in our investigations from the platform that most researchers need not necessarily sensitive data, so for this data sharing we didn't was our topic was not secret or security or sensitive data. It is about general data sharing for research purpose. And there it's actually to find it and then control the access. Who can have access to it? Not for the sake of making it open, but yeah.

Interesting. And data management plans? Is this something that you? I'm sure you come across them. And what? What do you think about them?

Yeah, they exist. Templates, in particular coming from, I think the GDPR... the representative at KTH was giving a whole session on data management plan and I think Rosa has it on her agenda. Umm, I did it for the KTH living lab.

What do I think about it? Yeah. It's an admin.

Are they helpful?

It's an administrative task.

Many researchers that have the same opinion when we talk about them.

It's perhaps good to write it.

Could they be helpful in a way? I mean it didn't.

I don't look at them afterwards. Maybe that's an answer. I do it, I have it, but I don't look at them afterwards. It may be good to go through and think about all the items once. I actually think the ethical application is much more. It was much more demanding to think about. How do we argue? Why is this a good way of keeping the data safe? So I think for the sensitive data, I think the ethical application is making you think more.

Internationalization. I mean, research is by default very global and international in most – perhaps not all research –but could you share your thoughts or if you have any challenges in particularly in these recent period when there's been a lot of writings in the media about collaboration with certain countries, and I know KTH also has a lot of stipends and so on. Has that affected how you work in any way?

And the students who working with their sensitive data are not from this areas in the world. So what is my thoughts on it in principle? I see the conflict between the different goals. Research should be open and should be international and should be on... and I see these political situation. I have no final thought on it. I know that the vice President for internationalization is stopping the collaboration with a CSC with those seven universities. So I happily follow the leadership there and don't want to put my... actually no, I don't know what to say. I don't. And I believe this is for common for a couple of colleagues that they don't want to make a... I also don't think that we should at a university, be on any such political decision that needs to be up.

This support you can get from KTH central level... you mentioned the Research Support Office... and are there other? I mean if you have a problem of any kind related to data, do you know where to find support?

Of course, it's Rosa. If others know it, they go to the IT service and they will pretty fast send them to IT infrastructure. And they will have this situation – who pays for the cost? The digitalization platform was announcing workshops where people can reach Rosa. Many reserachers in the domain was sent this newsletter, but if they all read it, I don't know. I think the information is there... the support from KTH is there, up till the point where the cost questions is not sorted out.

Do you have any other in the kth digitalization platform – do you have any contact with other higher education institutions in Sweden?

No. The platform is mostly an internal instrument for the I mean it has the external part, but it mostly is with respect to the local industry and collaborations with industry. It is not on internationalization or collaboration in the higher education.

You've been in Sweden for a long time.. do you have any reflection if you compare these questions to the same question in other countries... same challenges elsewhere?

Privacy is a topic in German very high up. Maybe I'm raised with a more sensitive behavior on this. I think in Germany it's driven too far. They are restricting too much. I think the objective to make data available publicly is good. I also think they are doing a good job of inviting regions and researchers. In Germany I think it's slower. The data is more locked in.

Risk analyses. Is that something that you've ever come across as a researcher, doing risk analysis when you start research project or during a research project, to map out different risks?

I don't know what kind of risk analyses. I mean, you do the privacy risk analysis if you want to open, make data open. If you want to make share data, you need to make an ethical application. If it is reasonable to do so. If you have enough argument. Do it the risk analysis if you want to have the data on your local computer and how to act on this. I think for most of my colleagues answered by we don't want to have it on our own computer. Because we don't have the facility. And then we also don't need to worry about so much. And if we want to have it, you will talk to your team. Then you are struggling to find a solution. What is OK or... If you can, you actually, and it's practical, you go to the host and do the implementation there. So it's a risk avoidance in general. We are trying to avoid to have the data and we try to... It might sometimes hinder us or doing uh, whatsoever research.

So for those researchers who are really working on the data because they want to figure out a specific item, then you find them inside life lab or they they exist and then there is a Med tech lab where actually these collaborations are set up for the sake of making it easier such that you can work on the data in the other environment.

And in your own research, what kind of data do you have? Like, what's the format of the data and what's the data about when you have sensitive data, if you, if you work with it, what kind of data is it?

It was the data from the patient data but that was the health data bank, Stockholm University. These are patient journals. They are anonymized, but it's only pseudo anonymized, so it's not something one should share with anybody. So it's patient records, patient journals and for the data which the ethical board does not consider to be sensitive, it's smart meta data. It's time series data of energy consumptions. Smart meta data shows you what appliances you are using. For example ovens. They can show if you for example are following Ramadan. It's sensitive. So you really see if a person follows religious trends. And by this moment it is actually sensitive data according to GDPR. Little information makes a whole data set to be sensitive.

Do you sometimes come across... you have a lot of different data from different sources... if you aggregate them, can they become sensitive, is that something you come across?

I have not worked with this data but with the region Stockholm and health data center, they were also producing synthetic data and I was guiding them on this research question. If you have a patient's journey through the health system and if you have a mobile phone positioning data, with a name. If you then know that a

person is as a certain time in a certain hospital and place and you see on the journal data a patient in this place and then you see it another one and you have these two signatures and you can decorrelate and then you know it.

We've talked a lot about you know personal data, privacy, and GDPR. But do you have any research that might be of interest for other state actors, or if it might be a national security or trade secrets?

(Shaking his head). No national interest. But I have a collaboration with (company) and other previous collaborations with industry, where we sign an NDA. But the data we use is public data. It's more that they share with us the technology.

What we write down in publications, we need to be careful and double check that. I think that's very often that researchers have collaborations with industry and we have an NDA on the knowledge to be shared. And also collaborations, they don't want us to have meetings on Zoom but instead Teams.

Is data management regulated in agreements or contracts?

The NDA is not having a regulation on the data, but the project which is on top of it. They point on where the data is located. Now in this project, is it located in Switzerland and it's research data and it's open?

It's again open, so again the the attempt of having the research on open data. Avoiding the conflict. But I'm pretty sure that other researchers have also data which is not open. And pretty sure that this we have as well.

Dual use... have you been informed about? I know KTH have a person that is guiding in this area?

I have not been informed, I of course know about it. I have fundamental research on biometrical identification. If one is evil one could actually use this research for surveillance.

I don't know about any guidance from the steering (management/administration). I mean I know about this Dagens Nyheter case in Lund, which is of course very questionable. And we have talked about it among the colleagues. Now have nothing which is so close. But we have not been given any direction from the steering. Do you think we should (be given steering)?

I don't know. I mean there are a lot of regulations that that you should adhere to so. But I mean, it's difficult for individual researchers to navigate and all these. So maybe... I mean we met coordinator... so he can actually, if you have a question, he can actually go through your research project and then give you advice on how to act.

Should they look through all research?

Or can they? I don't think they can.

Yeah, they cannot. So it means perhaps when the researcher feel says I need one can ask for the advice.

They could perhaps do a sort of a risk analysis and contact the researcher if there is sort of a suspect research that could be...

But could do a self-assessment. They could set up a self-assessment, when you end up in this vision, self-assessment is the wrong point, then maybe it tells you now you need to talk to...

What do you think about this? Do you think this should be like, a top down approach or bottom up approach that the researchers contact like central administration for help or vice versa. That central administration does risk assessments?

I know from the platform that the researchers... They had as you also said, there's just too many regulations, so you can provide and offer the support, but some might never listen and might not see the point that there is a need for it.

And I believe it needs to be both bottom up and top down. So from the top, maybe the prefects or the department heads need to point some researchers "here you are coming close to it. So please do an assessment." And then on the same time, there should be always the information and open support where you can go and ask for it. I think it needs to be. That needs to be both.

But dual use is not really... I think I know that we have these people, and then maybe I'm the person where this is not really on my agenda, and I haven't thought so much about it. And we have seminars where people come to inform us about sustainability and equality and all this from the top, and maybe it's an idea to also once in a lifetime, send somebody in a seminar and inform about this question.

Did we touch upon anything that you think it's sensitive during this interview?

No, it's all fine.