Interviewer: Yeah, so maybe we can go to figure number one. Maybe you can keep this figure in your mind throughout this interview where you can see, like... different levels of e-infrastructures of the research data. In level one you can see the IT-infrastructure which [is] needed for data storage backup network. In level two, you can see the research data creation process and some list of research infrastructure. Yeah, and you can see in level three, more general, or general category about the research data management. And the last level, it's... a kind of a guide and information for you about the research data services. Maybe we can move to the next slide and... So here, actually, we will... start by looking at this wheel. As you can see, these are the different six phases of the research data management, and we will look into these capabilities in the grey boxes, so you can tell us more about each phase separately. Maybe we can start shortly without taking any notes with the... You can maybe tell us briefly about your research with the focus on the data you collect or generate.

Participant: Yeah, so as you know I am [Name]. So, starting with that I think, I... could start with what I do. What I did in, at least the last five years, is that I am kind of railway dynamic engineer. We are modeling and also measure some train track interactions. Basically, we are designing the railways, how they look like and how we could increase the performance of railways. In specifically my case, we look at transitions: when you go from a bridge to normal tracks. So, that transition, how it works, is a bit challenging. We were looking at that, what are the challenges, what are the problems, how we could increase and improve the performance of that specific section. So, it could be from (--) You go to a tunnel, you go to a bridge, you change the type of the track, you go to Germany and then suddenly you see you have different things. In general, we are looking at that specific section, how we could improve it. Throughout my project we had two-three phases: one was initial modeling and then afterwards we did the measurement, which I think today we are going to talk about that section, and then we use those data to calibrate our models and we use them for later designs.

Interviewer: Okay. Nice. Maybe it will be easier for both of us to talk about recent project or section, as you mentioned. So, you can remember everything about your data and the plan is to start with these different phases. So maybe you can start with the plan.

We have also different things we want to know about it. Like, if you get any support or if there is any cost for this phase, it will be nice if you mention for us.

Participant: Okay, so I will start with the data. I will move to the phase of what data I collected. Of course, we planned the(--) We had a bunch of sensors in the field. First, we planned, “okay, we need to measure only one train passage”, which I think was in order of one minute of data with sampling frequency of, I think, it was 2000 hertz. So, you can imagine a couple of megabytes for each train passage. But the problem was that in the end, we couldn’t only record one train, because we couldn’t add a trigger level. When, when the train was coming, we had to have some kind of trigger point and then afterwards check, “okay, there’s a train, start recording”, but our system couldn’t work so we had to record continuously 24 hours.

Interviewer: Sorry, may I ask... So, you are now speaking about the Collection? Collect the data?

Participant: Yeah, I’m just talking about the planning, how we planned it and make it work, and then I will go back to the data. Basically, we couldn’t make them(--) Like, the system, I think, are the providers of the sensors, they couldn’t prove, couldn’t help us out to(--) Help us to only record the train passage which (..) One or two minutes each train passage. Then we had to record all the data for future measurements. Then (.) What we did, first we planned, “okay, how many sensors do we need? We, I think, had in total 34 sensors in the north of Sweden, which was really cold and harsh environment. So, we had to have a remote system, which I could do all these measurements through my office here at [university]. What we did, we managed to have a system, a computer there, a cabine [sic!] there, we put everything in that. And then also… From here I could measure what I was doing. And also, another thing that we had as a headache was the... Internet. We didn’t have proper Internet there, we had I think 3G, not 4G, not 5G, we had only 3G Internet. Then we managed somehow. I was able to log in and do some minor works and do some uploading and downloading through this system even though it was a bit difficult and time consuming. (.) Anyway, so yeah! That was the main thing that we had so... I think I should... Is it correct now? [inaudible]

Interviewer: Actually, I just wanted to ask one more question regarding the planning. Do you think (.) the planning phase, it's working well for you or… you face more issues than what you mentioned...? [speaks slowly]

Participant: We needed more help because at the end, you know, in this kind of project you have a limited budget and limited time. So, I think a bit more support and help, how we could manage this system with, through the [university] or somebody else, it would have been so nice. Because, as I said, we were recording every single second and that was, like, too much of data. I think I have four (.) hard disks full of data each. I don't know, many terabytes here with me... That's why we (--) If we would have (--) If we could have received some help that managed that only recorded train passage, we could have had all the data in one or two hard disks. So, that was one of the things that in the planning, we couldn't manage that, make that work. Of course, the planning was really important. Also, another thing was that (.) we weren't sure that we can mana-(--) we can do this task at the end. Because, you know, many delays because of traffic and so on, so forth. We couldn't manage. But if we just look at the [university] point of view, I think a bit of help (.) from the storage data, that how we could handle it, that was one thing that we didn't know what to do with it. We didn't have any help, so that's why we went with the hard disk. We bought some hard disk data and stored the data there.

Interviewer: Yeah, maybe this is about the Preserve. Later (.) we can talk about it. Before we move to the Collect phase, maybe I have another question about the Planning. Do you have any legal restrictions here with the planning phase or considerations?

Participant: We have, as far as I understand. I think we have a bit because the project is related to *Trafikverket*, and so we are not allowed to share some information in total. Probably we could share some, but not all. And that needs to be asked if we want to share anything. [speaks slowly] (.) It's a bit important, sensitive data… they call it.

Interviewer: Yeah, sure. Then if you think that you mentioned everything about this phase, if you like, we can move to the next phase, the Collect.

Participant: So, yes, as you wrote there, yes, I was(--) So, we had 34 sensors, and we were recording continuously for eight months, and sometimes I was just stopping it because I didn't have enough storage (.) space, so I was just stopping it and then start recording again after one or two days. The idea was to measure the long-term effect of the settlement, so we needed all the data from day one to day end. Then afterwards, that was the main thing that we wanted to look.

So, another issue was that since we had the sensors there and it was winter, so sometimes the sensors were broken, and we weren't able to know if the problem is in the sensors. or in the recording or in the software. I had to travel there, of course [soft chuckle], to check if there is a mechanical issue or something. So, that was a bit expensive. I think… our system could have been smarter than that to find it out but, of course, we didn't have that, at that time. But that was one thing that I'm(--) I can say about the (.) collecting the data. Then I was, as I said, we were recording the data in the field through a computer there, then I was uploading the data to cloud and then I was downloading it here to my hard disk. This process was always working. Record, upload, download in my [inaudible] and I was doing analysis-

Interviewer: Actually, this… This will be in the next. If... It will be great if you tell us about this collect phase, if you get any kind of support. Because you said you had to go by yourself and check the sensors and if they are broken and you collect some information and you check if the data that you are collecting is fine. You did not, like, think about getting any support in this phase?

Participant: Support from [university] do you mean, or...?

Interviewer: Like, even if it is outside [university] or inside [university]. I don't know if you try to get some support.

Participant: I think we had(--) So, basically, we had the sensor providers; they were trying to help us. But of course, it was really hard to answer some questions, whether the problem is because of the sensors or others. At the end, I had to make the final decision by traveling there and just checking and observing what's the problem. Sometimes, of course, there were mechanical issues because somebody (.) walked

there and put (.) their feet on the sensors and they broke the sensors. That was one or twice that we realized that the issues, yeah, random things, happened. So, no, I would say some support, but not all.

Interviewer: And is there, like, an extra cost for this phase or is it included in the project in total? In general?

Participant: Actually, it was a bit extra. Because we had to control that the data is correct. We had to do parallel measurements. So, we had to do some complementary measurements, which then added some cost to us. And the number was high, actually. It wasn't a small number.

Interviewer: Okay. So, if you feel satisfied about this phase, maybe we can go next to the Analysis. Like, how you do analyze your data, what kind of tools are you using to analyze them?

Participant: So, starting with the software or language that I used, I used MATLAB and Python. Of course, the data that I recorded, they were binary data. I was converting them to physical data, like acceleration, displacement, so on and so forth. So, I was converting that first. And then I was doing some filtering, signal processing, and so on, so forth to extract what I needed and also remove the (.) data that they weren't useful, like there were no train passes. I had to find the trains and then just extract the train that I want, do the signal processing, look into the (.) interesting data that I wanted to see, and then record those that I wanted at the end as clean data for next analysis. So, that was the thing that I was mostly doing as analysis.

Interviewer: And what kind of issues you faced during... analyzing the data?

Participant: So, one thing that it was(--) I think a big one, was that I was expecting to be able to convert acceleration to displacement, which is just a double integration. But at some point, we realized that the ratio between the noise and the signal was high. I couldn't get rid of the noise in the accelerometers. Basically, those data that I recorded in the accelerometer, they weren't useful, but the rest were. The main thing was that I wasn't able to, or nobody was able to extract the data from the accelerometer. Because the accelerometers were, I think, not those that we wanted. And normally, it's the case

that accelerometers are tricky. So, that was the main thing, I think, in the analysis I realized.

Interviewer: Okay...

Participant: I couldn't do that in a pre-phase. I couldn't check it before. Because before I installed them, they were working as I wanted. But... You know, sometimes you have noise because of this electromagnet from the rail and stuff like that. So, there is some white noise, which comes from the electricity and so on, so forth. That was the thing that we couldn't get rid of.

Interviewer: Okay. And during this phase, did you get any support, or you did collaboration with others? Let's say with [university research data support] [titters]. You know, they have this kind of support.

Participant: To be honest, I didn't get any help. No, I think I was single person on that.

Interviewer: Since you said that you faced some issues while analyzing your data, so what did you do after you faced these issues? You tried another way, or you never thought about getting support?

Participant: Mostly I was (.) looking at books, papers and literature, how they do, how they deal with it. And what are the tools available in the, like, yeah, the Internet, of course, and the people, my supervisor, they knew. So, in that level. We didn't ask any externals for help.

Interviewer: Okay, I understand. Then! Make it available for the others… What did you do in this case? And what kind of data are you making available to the others? Is it raw data or after processing it, or only results, or... What kind of...? [speaks slowly]

Participant: So, basically there are two kinds of data that are available. Some are published like final, final analysis. It's just a curve that people, they can use it. It's published and it's out, it's in a paper. Just one needs to (.) ask I will share that, it's like a couple of values. And I think that's open, that I could share it but still, I need permission which I think I asked for permission once, I shared it with one person and that person used it for a paper. Also, there are some train passages that I extracted myself, which I

think in [university] at least some master students and one or two PhD and postdocs have used the data. But of course, again I asked for permission and I shared it individually, so it wasn't like online somewhere, I haven't uploaded them anywhere.

Interviewer: So, it's up to requests for the students?

Participant: Yes, exactly.

Interviewer: Yeah, so these kinds of data, like, in the level of results after processing the data you are sharing it, not the raw data.

Participant: It's just clean, clean result at the end. Just need final responses. [inaudible]

Interviewer: Okay. Did you face any issues with this phase, like legal issues or...?

Participant: No, as far as I think the source was *Trafikverket* so I could ask "okay, you paid for this data" and the data is this which I have already published. So, it's out. "I'm going to share this figure with someone, is this okay" and they said "yes, it's okay".

Interviewer: Okay, are you sharing that only with students from [university] or are you also able to share it with people outside (.) the university?

Participant: There is one from [European university] but that was a shared collaboration between myself and that person. I think it's only, mostly [university] or(--) Because this project was part of a European project, so if anybody from that European project asks for the data, probably yes, we could share it with those people. But outside of that project, I'm not sure if we can.

Interviewer: Okay, thank you for the information. Then if you don't have any other thing to add here, maybe we can move to the next phase. It's Preserve. How you preserve your data. And again, it would be nice if you tell us, like, what kind of level of the data management you are reserving? Is it, like, the raw data or through the processing or only the results again?

Participant: So, I think(--) One, we were thinking about it two ways. One was that to save them in the cloud or somewhere – I think you were thinking about that – but at the

end we realized it's a bit expensive for the project. So... At the end, we came up with the (.) hard disks which... I just have them here; there is one here.

Interviewer: Okay.

Participant: Always next to me, there are four more here, so I have a bunch of them. (.) We recorded in them all the raw signals and also those that I extracted myself. So, if someone is, let's say after me, is interested in the data and they want to go back to the raw signal, still the data is there and they can use it, of course. But also, the extracted data is also there, so both are available and saved here, but they are in the hard disk. And [inaudible]

Interviewer: But are you putting any documentations about (.) this data somewhere?

Participant: So far, I hadn't any time to do it. Mostly if someone needs, they need to ask me so that the documentation is actually with me.

Interviewer: Okay, so in some years maybe you will forget about that. [laughs]

Participant: Probably [laughs]. I guess I will forget about it, but I hope that I have some time after my PhD to spend a few days and give some guidelines for that.

Interviewer: Okay. But you did not get any support from your department. You know there is, like, a different (.) storage and locations from, for example [university research data support unit], there is small storage for free. You did not think about that?

Participant: I think that was too small, the storage space that we had. I think it's in order of gigabytes, but we were thinking about terabytes here. So… That's why it was... I think it wasn't enough for me. I think the only thing that I use, I use the OneDrive, my own. I disconnected my laptops and everything from the OneDrive. I think I just connected the OneDrive there from the [inaudible] and the OneDrive from my personal laptop to just make it free for the data. The data was coming to my OneDrive and I was downloading it to the disks. That was all I did and I didn't get any help for that, because as I said we discussed that and then they came up with that "it's a bit expensive, maybe you should find another way". Then we came up with the hard disks, which was much cheaper.

Interviewer: Okay, so (.) as I understood you did not get any support in this phase as well.

Participant: No.

Interviewer: Yeah. Any other information or issues you are facing in this phase?

Participant: No, only as I said I wish(--) Because I see that there are smarter ways to save the data and record, and I think preserve them, but unfortunately, we didn't do that. Or maybe we could have done it with some help and support.

Interviewer: Okay, thank you. Maybe now we can speak about Reporting and… it's different than make it available. Like, how you publish the information about your data.

Participant: So, as I said, the papers are open access based on the European project. So, everything should be public, open access. So, we... published the final results in papers. And those are journal papers and my thesis, of course.

Interviewer: Okay. And it's also possible, up to the request, you can share the data with the external people?

Participant: As I said, yes, upon request. And that I need to ask for permission from the, again, *Trafikverket* for each single person. So, I think that I need to always get permission for it.

Interviewer: Okay. Thank you very much. Maybe you can just… take a short look at the notes that my colleague wrote and please try to tell us if there is something you need to correct or change in these boxes.

(...)

Participant: Yeah, I think, in general, that's good.

Interviewer: Okay, we have another request now for you. If you look in the bottom you can see three smileys there. Can you please use those smileys to describe how you feel about these different phases? Like, are you satisfied, for example, with the planning or you don't have any opinion, or you are dissatisfied. [titters]

Participant: So, for all six phases? I should–

Interviewer: Yeah, one by one. Like, for the Planning, are you happy about planning or do you think you face (.) issues and you are not happy with it, something like that.

Participant: I think I go with “no comments” for the Planning.

Interviewer: Okay.

Participant: Because we could do what we could do [laughs]. We did what we could do based on that, that’s why I think (.) no way to change it. The collected data, I think... There, I’m a bit unhappy because… Because I think we could do it smarter, so that’s one thing that I can say. Analyze the data, I think it was fun, so I’m happy.

Interviewer: Okay, good.

Participant: Yeah, and then Make it available... Mostly it’s extra work, so for me, which I’m lazy, I go with I’m unhappy with that because, yeah... (.) And also Preserve, yeah, probably no comment–

Interviewer: This is obvious, it will be nice if you be more honest about that. Because you said you have some issues with the Preserving. [smiles]

Participant: Yeah, that’s why I have no comments about that because I think–

Interviewer: But... You don’t have comment or you are happy or unhappy?

Participant: No comments... Because they were buying hard disks, so...

Interviewer: Okay, thank you.

Participant: About my “make availability” I’ve been unhappiest, because I needed to do some stuff there... (.) Work... Yeah, I’m happy about that because my thesis is published now.

Interviewer: Great! Then... I think we are now approaching the end of the interview, but we have one last request from you. Maybe by looking at the notes that you helped us to write and these smileys that you put for every phase. Maybe you can describe for us your... Like, dream scenario of how you deal with your data (.) in your (.) work.

Participant: So, I think this was an experience for me. If I go... If(--) With this knowledge that I have now or the experience that I have now... If I go back (.) I would do it in another way, of course. So, I could do it really much, much, much better because I know that there are good tools available outside. And also, I could select (.) different sensors, different companies, different providers, different acquisition system. Also, I would push it to do the measurements not in the north, maybe somewhere close by, that I could go and have better access. Or I could initiate a collaboration with the [university located up north] which is closer to the location. I could have somebody there, which they do some hard work which I didn't need to pay extra. I'm sure that they would have been happy about that. Because, of course, you give them something free and that you want something back from them, that would be good. I know that... As I said, also for storing data, of course, I have seen these (.) cheaper things which are coming out nowadays. Different class system and other things. I think there could be also useful because by recording them, you can always have access to them. But hard disks that I need to connect to my computer, look into and find. It's really tricky to find data in your hard disk because if you are working from your office, then you need to take your hard disk, which is heavy always. Or if you work from home, then it's a bit, you know, headache. All the things I would say, yeah, maybe there are better ways to do it. And I would say, yeah, this is my comment. I would have done it in a better way.

Interviewer: Yeah. Thank you again. And do you have anything else you want to add before we end this interview?

Participant: Yeah. One thing that I realized during the interview, I think you mentioned that [university research data support] do support. What I understood is that is just too great. I think it would have been nice to somehow let the PhD's, researchers know about this because I feel like we don't know about it. At least push (.) head of divisions to mention that in division meetings because I think there, people they listen. If you put it on a website probably less people read it. If you put it on somewhere that it's in the website of... your website, I'm not sure if people read it because they don't know that it exists.

Interviewer: You are right.

Participant: So, one way I would say because in a division meeting they normally announced what is what, what is available, what is not. People and each division has it every one month, so. It's good to know and "okay, you are a head of division, we are going to give this kind of support" by listing some "okay, we have a space, we have this, we have this, we do this kind of support with short description like in one or two paragraphs, a few bullet points. Then they just copy it in their presentation they said "okay, we have received this information from there" and then they do these kind of things for us. Yeah, it's good if you do measurements if you play with the data, it's good to call them and contact them, and get support. I would... I didn't know about many things which I realized today, of course. I will let others know about this, but I think it would have been nice if(--) I know that other people also have these issues in, at least, in our division, I know that people are struggling with the data, storing it, recording it, analyzing it. So, if such a good thing exists, I think it would have been nice that it would have been shared in a better way.

Interviewer: Okay, thank you very much for the feedback. We are putting this information in the website, as you said, maybe it's not readable by everyone. Also, we have also newsletter, I think, every two weeks or one month. But this is a good note, maybe in the meetings or the divisions. Yeah.

Participant: I think the best way is the head of divisions, because when we want to even buy a simple coffee machine, they mention it in the division meeting. So, we know that, "okay, there will be a coffee machine soon", or it's broken or whatever. So simple things or really advanced, and we hear it there. So, I would suggest there is a good place to share something really useful. Thank you so much.

Interviewer: We will transfer your notes [smiles]. Thank you very much for your time and talking to us about your data.

Participant: Thank you.

Interviewer: And I will stop recording.