

1 **Interviewer:** Good! So, hello again! [excited] I'm [Name]. Well, you know [Name] (.),
2 she will take notes.

3 The aim of the project, KAFFE it is called, is to map how researchers work with
4 research data today and what their needs are. And the goal is to develop (.) the best
5 possible support for you so that you can focus on your research. To structure the
6 discussion, we will use a simplified model of working with research data based on the
7 phases of planning, data collection, data analysis, accessibility and sharing, data
8 preservation and reporting in the end. We look specifically at capabilities, and this is a
9 very broad concept that describes all the processes or information resources that you
10 might need to get along this process, and you'll see a little bit more what this is in a
11 moment.

12 Of course, our focus or our emphasis is on what e-infrastructures are needed to
13 manage research data well, and e-infrastructure is something that you as a researcher,
14 I guess, are not usually thinking about unless it doesn't work. So (.) [notetaker] maybe
15 you could show the image (.) the four levels of infrastructure.

16 **Notetaker:** Yeah, sure. (.) I'm sharing now. (...) Are you able to see my screen now?

17 **Participant:** Yep, maybe you can zoom a bit. Just make it full-screen. [inaudible] That's
18 good.

19 **Notetaker:** Yeah, sure. [speaks slowly] [inaudible]

20 **Participant:** If you click on slide show (...) Yep, and then start from first slide (.) Yeah.

21 **Interviewer:** So, the main points are here that there are different (.) infrastructures that
22 (.) accompany you on your way through research projects and on your way through
23 working with data. So, on the level one, we see the IT infrastructure, which was needed
24 for data storage, backup and all of this. And level two is a bit more specific to data. So,
25 it's about research, data creation and processing. And there you see some
26 infrastructures listed on the right, like for example, SNIC, providing compute servers. On
27 the level three, we have research data management. So, this is (.) a very broad
28 category of tools that are required for managing active data. So, while in the process,

29 not so much for [inaudible] data sharing and analyzing. And finally, we have the
30 research data services on level four, which is more like guidelines and information that
31 can be helpful to have for research data management. Okay, so this is what we mean
32 when we talk about infrastructure. Please keep this image in mind throughout the
33 interview, as it represents sort of the environment that you relate to as a researcher.

34 The interview now more specifically goes like this: we will start from a research activity
35 that is relevant for you. So, it could be a current or recently completed project. And we
36 do that because it is usually easier to talk about specific projects. But if you come up
37 with more general comments, that's absolutely [inaudible]. Now (.) I (...) we can go to
38 the next slide. And there, you can see the model that we're talking about.

39 So, you see (.) the data (.)

40 **Participant:** Sorry, may I interrupt? (.) [hesitant] I see (.) I hear some background noise;
41 I don't know if it's on your side or [notetaker] side? (.) So, maybe (.) one of you can
42 mute yourself (.) Whoever it is [inaudible] [chuckles]

43 **Notetaker:** Yes, sure.

44 **Interviewer:** We can try with [Name].

45 **Participant:** Yeah, let's (.)

46 **Interviewer:** [inaudible] my headphones.

47 **Notetaker:** Let me (.) just (.) mute myself. [speaks quietly]

48 (...)

49 **Interviewer:** Okay, let me know if it persists. (.) Good.

50 **Interviewer:** Okay, so what you see here is like this wheel or the cycle. That's like the
51 data life cycle. And in gray boxes, you see these capabilities that are relevant for you to,
52 I mean, you do or you possibly would (.) like to have support with that. So, it's like this
53 define the project's data, create a data management plan, process data, analyze data,
54 all the way to store data, long term archive data and reporting. And (.) So (.) you do not
55 need in this interview to talk about all the capabilities. This is just to give you an idea of
56 what specific aspects we are interested in learning about how you handle these and
57 where you think you could have more support.

58 **Participant:** Okay. Good. (.) So, I guess I can start with what the current practice is and
59 maybe in (.) towards the end I can say a few words on how I would like it to be.

60 **Interviewer:** That sounds perfect and I am (.) So, in the rest of the interview we will
61 essentially see this wheel without (.) without the capabilities but [Notetaker] will, while
62 you are talking, take notes on what you're saying and you will see what [s/he] is writing
63 down in this figure.

64 **Participant:** Okay!

65 **Interviewer:** And then (...) Yes. Do you have any questions before we get started?

66 **Participant:** No, no, it's fine.

67 **Interviewer:** Okay, so I'll start with a very general question just to (.) sort of set the
68 stage before we get into the specifics. Can you briefly tell us about your research with
69 focus on the data you collect or generate?

70 **Participant:** Yeah, so I am interested in robotics and AI research. And in particular, I'm
71 interested in sensory motor control of robots. That means how we move, what kind of
72 reflexes we have. And how can we make optimal actions in dynamically changing
73 environments, pretty much like animals or humans that we are able to (--) we are very
74 versatile, right, we can adapt to situations and how can we enable this on robotic
75 systems. In particular, I am more focused on the algorithm side of it (.) So... How to
76 develop algorithms that can, for example, take into account sensor measurements from
77 the robot and maybe the collected past experience that the robot has and transform this
78 into actions that preserve the safety of the robot and also the safety of its environment
79 so that it does not hurt the humans or other infrastructure surrounding it. So, this is the,
80 let's say, general scope of my research.

81 **Interviewer:** Okay, great. (.) So, let's move on to talking about maybe a recently
82 completed project or one that you are currently working on (.) So, when you see this
83 picture with the faces [speaks slowly] and now we'll move to the one without these
84 capabilities. When you see this picture, can you describe what you do with the data
85 during the different phases? One at a time and we put a few questions here that are

86 well interesting for us to know for all the phases essentially. So, what do you do with the
87 data? How does it work? What tools do you use? (.) Do you get support and does it cost
88 something and, well, how did you experience data management in this phase and what
89 were the problems, if any? So, just talk freely! [laughs]

90 **Participant:** Yeah, so, for the planning phase, I would say that really depends on the
91 context or the topic of the project because we have so much variety in robotic systems
92 like you can have these six degree of freedom robotic manipulators (.) quadruped
93 robots, humanoid systems, right, and all of them have very different structure. So (.) the
94 type of data that you collect and what it means for us is also very different. So, yeah (.)
95 the (.) let's say, whenever a project starts you will see, okay what kind of application
96 scenario the robot must operate into, and what kind of robot architecture are we using.
97 So, based on that, we will think about how to manage the data. And (...) what I can also
98 say at this stage is that maybe unlike some other fields, right, we are dealing with high-
99 frequency time series data that is collected from sensors or motors in the robot. So, the
100 size of the dataset, right, can be very large. So, we also have to often look into
101 compression opportunities and how can we (.) yeah, how can we manage a large
102 amount of data (.) efficiently. So yeah, I would say at the planning stage, the key
103 questions to ask is, okay, what robot are we using? What is the intended application
104 scenario? Is there already some reference point in the literature where people, other
105 researchers have, let's say, recorded or released some datasets? So, then we also look
106 into the state of the art or basically the published datasets, right? And can we learn from
107 some of the best practices that are out there? Yeah, otherwise we have to, quite often
108 we have to devise our own strategy because there is just so much diversity in the
109 system that we are dealing with.

110 **Interviewer:** Okay, maybe we can go to a specific project. I don't know if you recently, I
111 don't know submitted a paper or something [titters], where you just sort of decided on
112 one specific robot. So then, in the planning phase... You usually look for freely [speaks
113 slowly] openly available data as well? And if that does not exist you create it yourself?

114 **Participant:** Yeah. That's great. So, I mean, I can talk about the [previous] project that
115 we have been working on. I mean, it started when I was in [country] but still, we keep

116 developing it and there the idea is to, like, to create an open repository of robot skills
117 [speaks slowly] and the data collected from robots also how the... those algorithms
118 perform. So, we have leaderboards to track the performance of algorithms or learning
119 algorithms, or control algorithms. So, yeah maybe I will use that as an application
120 scenario throughout the cycle because it will cover like most of the [inaudible]

121 **Interviewer:** Yeah, that sounds good. (...) Okay, so then, we had the planning phase
122 and then you went on to create collect data. [speaks slowly] And in this case, it's
123 probably more about creating?

124 **Participant:** Yeah, I guess I touched upon this a bit already, that it's mostly the time
125 series data, right, and collected from motors from sensors on the robot, sometimes
126 cameras, so image data and I think what probably sets us apart from other AI fields is
127 that our data is quite high frequency. So, you can imagine that you are collecting sensor
128 data at one kilohertz, right, and if you have, let's say, 30 joints in the robot then, yeah,
129 you have 30 doubles that you (--) So, for one second you will have thousand rows in a
130 CSV file and 30 columns just for the position part, and then velocities, and acceleration
131 torque... So, you can imagine the bulk of data a robot can produce within a matter of
132 seconds. So, dealing with that is not easy.

133 **Interviewer:** Yeah, so, thinking about data handling there, are there any specific tools
134 that you use or any (.) I assume it requires a certain IT infrastructure as well to handle
135 this?

136 **Participant:** Yeah. So, I mean, in the robotics community, we have some robot
137 operating systems. So, it's called ROS, for example, or ROC. There are many such
138 open source projects. And each of these operating systems would provide some kind of
139 logging functionality and those (.) logging functionality would, for example, without
140 interfering with what the robot is doing, yeah, it will reserve maybe certain threads on
141 the CPU to log the data and write this in some kind of bag files. In ROS, it's called ROS
142 bag. In other software it could be POCO log. Yeah, there are several of these different
143 formats. And they... Because we are dealing with large amount of data, we typically
144 have some compression applied on it. So, you cannot open the file and simply read

145 what's there. You would need another kind of parsing solution that would parse those
146 log files. And then, yeah, you can convert them into CSV or any format, maybe some
147 pandas DataFrame, whatever you need for further analysis. [speaks slowly]

148 **Interviewer:** Yeah, so maybe we add, like this compression issue, to the create collect
149 (.) gray box. [titters]

150 **Participant:** Yeah.

151 **Interviewer:** Yeah, okay. I think this... now you put it under Make Available [Notetaker],
152 but I think it could already go into the Create/Collect before it needs to be in the right
153 format to (.) such that [inaudible] they are able to analyze it.

154 **Participant:** Yeah.

155 **Interviewer:** Yeah, exactly. (.) And then I (...) do you set this up all by yourself? Or did
156 you get any support from somewhere? [inaudible]

157 **Participant:** Not from university, but we have our own, like, open source community,
158 where [inaudible] people are developing open source software. So, we quite often rely
159 on those tools.

160 **Interviewer:** So, you're saying not from infrastructures essentially?

161 **Participant:** No. [speaks quietly]

162 **Interviewer:** Okay. (.) Would you have liked support from infrastructures there? Or is it
163 like, it's just part of your research so [inaudible]

164 **Participant:** I mean, of course [laughs], we would like that because sometimes it's very
165 time consuming to deal with these kind of issues. But also, I feel that maybe different
166 communities have very different needs with respect to data. So, probably you would
167 need [chuckles] a bunch of engineers in a university just dealing with that domain
168 specific problem. And hence... We have... [speaking slowly] Yeah, some externally
169 funded open source community, for example, there is Open Source Robotics
170 Foundation that receives money from... from investors and so on, right.

171 **Interviewer:** [inaudible] open source?

172 **Participant:** Open Source Robotics Foundation. They are the ones who are developing
173 and maintaining ROS... Robot Operating System.

174 **Interviewer:** Okay.

175 **Participant:** So, yeah. I mean, the work is happening somewhere centrally. I'm not sure
176 if universities have the (.) let's say... incentives to invest [titters] into this directly,
177 because it might consume a bit of resources. But what the universities can do is to act
178 as a bridge between those foundations and the end researchers, right. Yeah, so I think
179 that's where universities could potentially contribute.

180 **Interviewer:** Do more maybe? [inaudible] what you are saying.

181 **Participant:** Yeah, yeah.

182 **Interviewer:** Okay! (.) Let's talk about the... Analyze... The yellow bar here. [speaks
183 slowly]

184 **Participant:** Yeah, here as I was saying (.) like, compression is when you collect, right.
185 And then [giggles], if you want do something with the data you have to...

186 **Interviewer:** Decompress?

187 **Participant:** Decompress. Decompress is the right word. So, yeah. Basically, parsing
188 the data, decompressing. Decompressing(!) it, parsing it. Maybe... Sometimes in
189 robotics we also have to do a deal with up sampling and down sampling quite a bit,
190 right. Because if you have very high frequency data, maybe the things that you want to
191 study can be captured even if the data was low frequency, right, depending on what you
192 want to study. So, then you would want to reduce the frequency of what we want so
193 there are some mathematical tools that you would need, like filtering, up sampling and
194 down sampling methods to maybe create another intermediate dataset that is more
195 easy to handle. And... So... [speaks slowly] Maybe at this stage you would convert a
196 binary, like, highly compressed file like HDF5, for example, to something which is, yeah,

197 more like human readable, so it's like a CSV, YAML, JSON or any of these formats,
198 right [speaking quietly].

199 **Interviewer:** Yep. And to do these things both the, let's say, pre-processing and the
200 later analysis of the dataset you use Python or are there any specific frameworks or
201 tools for this?

202 **Participant:** Yeah, I think when it comes to offline analysis, Python is the de facto
203 language there. But the... Let's say the logging part, right, and
204 compression/decompression, these are usually written in C++.

205 **Interviewer:** Okay.

206 **Participant:** But... There might be a Python wrapper around it, but yeah, natively they
207 are C++.

208 **Interviewer:** To make it faster.

209 **Participant:** Faster... Yep. [speaks quietly]

210 **Interviewer:** Yeah, understood. (...) Make available in the sense of... publishing,
211 sharing with others. [speaks slowly] How does that work for you?

212 **Participant:** So, in my particular case... I mean, we don't publish all the data. Let's say,
213 only the meaningful one, right, that has some meaning. So, for this, we quite often use
214 GitHub, so the Git Large File System [note: means Storage] LFS. Or we sometimes also
215 use Zenodo, like to give a numbering and release or some kind of version control. So,
216 those are the platforms that I personally have been using: But again, it depends on the
217 application, and some people, they host the dataset on their own website on a file
218 server and then you can download it from there. But I am just more of a fan of like Git
219 [note: means GitHub] [giggles]. So yeah, having version control...

220 **Interviewer:** And then I assume this comes without a cost?

221 **Participant:** Yes, yes, it's free. Yeah, at least at our scale and that we have been
222 operating it, it has been free.

223 **Interviewer:** Okay. So, does it mean that you, what you're publishing finally is a much
224 smaller dataset than the one logging in the beginning?

225 **Participant:** Yeah, those we almost never publish, yeah.

226 **Interviewer:** Yeah, okay. Well, that's... That's good to know that sort of the size
227 challenge is mitigated on the way.

228 **Participant:** Yeah. (.) But I must also say that just because it's like this doesn't mean it
229 should stay like this [laughs]. So, especially now, like, you know with the success of
230 generative AI algorithms, which are like, kind of very data-hungry. There are also, like,
231 big initiatives from Google, for example, or like big tech, who want to create very large
232 data-sharing possibilities from robots.

233 **Interviewer:** To store the raw data then.

234 **Participant:** Sorry?

235 **Interviewer:** To store the raw data then.

236 **Participant:** Yeah, yeah, to store the raw data. So yeah, things are changing a bit.
237 Again, because as I was saying, it's so much data... [chuckles] Yeah, we need very
238 high... Yeah, good software engineering, possibly well-funded teams, you know, who
239 can dedicatedly... Yeah, work on this problem, yeah.

240 **Interviewer:** Yep. Yeah, I understand. And then, I mean there's also the question if you
241 have published something that (.) and I don't know if this applies to your [project], do
242 you want to make it visible such that other researchers use it or is it more for internal
243 use?

244 **Participant:** No, for sure it's visible and we do several activities to, like, increase its
245 popularity. So, I mean, the most basic thing is to publish papers, present at
246 conferences, right, or journals, but then we are also using that in our teaching. So, we
247 encourage the use of those dataset in our, like, assignments so that students maybe (.)
248 solve some task using, you know, some small chunk of data from this ecosystem. Then
249 we are also organizing some international competitions where the teams from, like, all

250 over the world can compete. And we do that, like at the top robotics and AI conferences.
251 So, every (--) We have been doing this now since three years, so the third one is going
252 to be in the US. I'm going there at the end of this month. And again, we have like many
253 teams from different parts of the world competing using the data and the system and the
254 infrastructure that we have created. So, we do share it on our mailing lists, for example,
255 our community-building lists. That's a big... That's a source of attraction for (.) for the
256 work we have been doing.

257 **Interviewer:** Yeah. Yeah, okay. Anything more you want to say at this point? (.)
258 Otherwise, we move on to preserving the data. And there... It's a question... [hesitant].
259 Yeah, how do you do that, which parts are preserved, how are they preserved long-
260 term, how's that going?

261 **Participant:** Yeah, I think(--) I mean for now I think it's good, but what we... What I said
262 didn't make available [sic!] and then it's also a good bridge to the preserving part. So,
263 we are using, as I said like Git [note: means GitHub], Zenodo, right. And those are not
264 just... So, the server space we are not (.) owning them.

265 **Interviewer:** Yeah [speaks quietly]

266 **Participant:** Just trust on this... On this infrastructure that GitHub does not close down
267 [chuckles] and yeah... It will be [speaks quietly]

268 **Interviewer:** You're not owning them and the university is not owning it either.

269 **Participant:** Yeah, yeah.

270 **Interviewer:** Yeah, I understand.

271 **Participant:** But I don't know if owning it is also [chuckles] a good idea. I mean unless
272 it's a really open... open source, like very permissive licensing. So... (.) Yeah, I mean
273 it's good if, if there is something that we can trust, right [chuckles] and... And... It's...But
274 yeah, usually we just rely on what is available for free [titters].

275 **Interviewer:** Yeah, no, I... Exactly. I was just wondering if there are any (.) university
276 services that you are aware of where you could put your data to archive it, to make sure
277 that it's [inaudible].

278 **Participant:** No, I mean, no, we haven't. I'm at least... I'm also new in Sweden so I
279 don't know [chuckles].

280 **Interviewer:** Yeah [chuckles]. Good. So now when we, when you say it's the same as
281 previously so GitHub, Zenodo, mainly those open (.) infrastructures (.) that is for the
282 post-process data. What about the raw data? Do you keep it? Is it... (.)

283 **Participant:** I mean it's probably on some hard disk. Hard disk [laughs]. But yeah, we
284 don't come back to it often. Once it's published then... Yeah, then we don't go back to it.

285 **Interviewer:** Okay. It's... So, that the raw data is not published and (--) I mean, it might
286 be somewhere, possibly findable. Not visible for anyone externally.

287 **Participant:** Exactly.

288 **Interviewer:** Yep. Okay, I understand. So, then we can come to the last part which is (.)
289 reporting. Do you do any final report of (.) of the project. I mean, I understand there's a
290 publication and all of this...

291 **Participant:** There's publication, there is the project report itself that goes to the funding
292 agencies and... And sometimes, yeah, we also make websites where, like, it's a web
293 page that describes how to read the data or so. So, those pages are, again, hosted on
294 GitHub. And, yeah, that's the engagement point, let's say, with the outside world, right?
295 Otherwise, we might also have internal GitLab documentation that we use. Yeah, but for
296 the outside world it's mostly documentation on webpages, publications (.) or project
297 reports (.) through the funding agency, for example.

298 **Interviewer:** Yes. Okay! (...) I don't think there's necessarily much more to cover for
299 this wheel, as of now. So, we look at the picture of the notes of what you have said. Is
300 there anything you want to change or add from what has been written down? [speaks
301 slowly]

302 **Participant:** I will read it quickly.

303 **Interviewer:** Yeah. I was feeling that maybe we can add somewhere that you mainly
304 use Python and C++ for the both create, collect, and analyze?

305 **Participant:** Yeah, I think that's good to mention. So, in the analysis part, I would put
306 Python. Maybe create, collect, could be C++.

307 **Interviewer:** Okay, yeah!

308 (...)

309 **Interviewer:** Oh yeah, I wanted to ask you one more thing. Are there any... This is
310 robotics data, so I assume it's a no, but are there any legal considerations or restrictions
311 regarding to any of your data?

312 **Participant:** I would say that really depends on the, like, I mean, we are mostly doing
313 funding from either(--) So, the project we do are either funded from the government or
314 from industry. In case of industry grants, quite often they don't want to release the data,
315 right, so then we don't. And if it's funded through public money, then quite often there is
316 an obligation to make things open source. Then we do make it open source so(--)
317 Depends on who is funding.

318 **Interviewer:** Okay, good. (.) Yes. And maybe we can also add the point here that the...
319 At the preserve for the raw data, it's external drives. Somehow, we can know that then
320 it's not publicly available. Yeah, we cannot use it. (...) Okay... [speaks slowly]. Good.
321 Anything else you want to mention or change, [Participant], about this note?

322 **Participant:** No, I think (.) that's fine.

323 **Interviewer:** Good. And we now want you to add one of the three smileys to each of the
324 phases describing how well you were satisfied with (.) Well, how this is going for you,
325 including if you would have liked more support at some point. So, if you're very happy
326 you put a smiley face, if you're neutral, the neutral one, and if you really felt like there
327 could have been more support or something didn't work well, you put the unhappy face.

328 **Participant:** I mean as I said in the beginning, I'm not really using any university
329 infrastructure here so... Yeah, I mean (.) we are mostly using open source tools that are
330 developed outside. [inaudible]

331 **Interviewer:** And it is completely fine. So, it should reflect your overall perspective
332 [smiles] and so to say, even if you're not using them, but you feel unhappy at this stage
333 then it probably means you don't know if they exist or they don't exist. So, I mean this...
334 Yeah (.)

335 **Participant:** Yeah... (.) Yeah, I guess neutral is okay. [chuckles]

336 **Interviewer:** For the planning phase?

337 **Participant:** Oh, okay, I have to do this for each one?

338 **Interviewer:** For every phase, please, yes.

339 **Participant:** Okay. And is this smiley reflecting what I feel about that step or is it
340 specifically related to the support I expect from the university?

341 **Interviewer:** It is what you feel about this step, which in a way reflects if... I mean, I
342 know you said you don't use support. So, if you put a happy face, that sort of means
343 you're happy even without support. If you put an unhappy face, then it means you're
344 unhappy about the process, meaning that somehow you lack support, no matter if it's
345 because it's not there or because you're not aware of it.

346 **Participant:** Oh, okay, I see, I see.

347 **Interviewer:** So, it is really just how you feel about–

348 **Participant:** Feel about each of these steps and where we could have help. Right,
349 okay. The first step, you can put a smiley. (.) So, I think this is entirely the researcher's
350 job to know what's your application scenario, what is your source of data and so on. So,
351 I don't think we need support on this particular front. The second one is "create and
352 collect". I think here I would put a neutral face. I think (.) we can be a bit more ambitious
353 [on the university sides in Europe to also take an active part, let's say, in supporting
354 these kinds of needs that are very relevant for the future. So right now, most of this

355 infrastructure is being developed in the US with the support of big tech, basically. But
356 sometimes they are usually spinoffs from universities. So, for example, ROS was a
357 spinoff from, I think, Stanford. So, I think, if universities in Europe also take a bit more
358 bolder steps, (.) that would be really good. And as I said, like the problem in robotics is
359 high frequency data, right? And lots of data, right. So, how do you effectively collect it? I
360 mean, right now (--) We are also entering this world where you want to collect data from
361 multiple robots in batches. So, imagine like 10 robots of the same type doing some
362 activity, right? And you want to collect data from 10 different, real-time data from 10
363 different hardware, right, doing something. So, there are no good, like, technologies to
364 enable this kind of work. So, at least not that I'm aware of but I do know, like this kind of
365 batch learning is happening right now at the bit in the US, like in big companies. And
366 yeah, I think, universities, they don't have access to that at the moment. So, a bit more,
367 like independence, on this front would be really good.

368 **Interviewer:** Okay, can we add to the notes that are there then, that most currently
369 available infrastructures are developed in the US. And maybe just as a bullet at the very
370 end. And then it would be good for European universities to...

371 **Participant:** To step in, yeah.

372 **Interviewer:** To step in, exactly. (...) It would be good if European universities (.) step
373 in. (...) E U R O P E... Yes. [giggle] It is very small on a screen. Probably it wants to
374 capitalize it or so... (...) Okay. Good. So that's a neutral one. Analyze?

375 **Participant:** I think here I'm happy. And mostly the tools already exist. And we can be
376 independent on this front.

377 **Interviewer:** Yeah. And it's also part of the research, right? I guess.

378 **Participant:** Yeah.

379 **Interviewer:** Making available.

380 **Participant:** I guess... [hesitant] Again, maybe neutral, because, yeah, we do have
381 some things that we can use, but maybe it's also good. If there were, like university-
382 based solutions (.) to host data and...

383 **Interviewer:** Yeah.

384 **Participant:** [inaudible] right.

385 **Interviewer:** Yeah.

386 **Participant:** Yeah, so neutral.

387 **Interviewer:** And I do believe the university is working on this. [giggle]

388 **Participant:** Okay. [chuckles]

389 **Interviewer:** Preserving data, like archiving.

390 **Participant:** I think I'm fine with this one also. Again, we can't invent GitHub here.
391 [laughs] So, what's there (.) is probably all pretty fine. Report publication(--). So here, I
392 mean, I would like if, like, [university] has more open access agreements, for example,
393 that would help because... So, for example, one of the things I was quite surprised
394 about when I started here was that we did not have any open access agreement with
395 IEEE. Which is such a big publisher and so many people from electrical engineering,
396 computer science, mechanical... They publish there. And then [titters] I kind of pushed
397 for it. And now we did get an open access agreement with IEEE. But yeah, many other
398 big publishers probably (--) I mean we should look into this, like, pushing for more open
399 access right, in the, yeah [inaudible]

400 **Interviewer:** Yeah. (.) Yeah, maybe we can add this as well as a bullet point.

401 **Participant:** Yeah. So, again a neutral face.

402 **Interviewer:** Yeah, and it would be good to have more open access agreements. (...)
403 Great, so... Last question from our side: what would be your dream scenario? And here
404 I'm referring to like maybe any part of this data processing life cycle where you feel that
405 things are not getting ideal... What would be your dream scenario there? (...)

406 **Participant:** I think (.) it would be good if we support systems, let's say in universities
407 they are more ambitious, so this is again coming to that point, right (.), that quite often
408 they are not dealing with like high frequency data. And also, sometimes there is like a

409 clear hierarchy, so if you solve the hard problems the easy problems are solved. But
410 quite often like the efforts at universities are about dealing with simpler systems and
411 how you can help other, like domains, which are (.) which are a bit easier to deal with.
412 So, I think... The problem of handling data in robotics is a hard one, but if we crack that
413 one, then it will affect all other fields. So instead of thinking short-term solutions, we
414 should think of longer-term solutions and more ambitious [giggle] goals. And I... And I
415 think if this mindset changes across universities (.) then we can create counterparts to
416 the US ecosystem [giggles], which we are heavily relying on at the moment.

417 **Interviewer:** Yeah, yeah, I understand. Deal with more complex data domains by the
418 support systems... [speaks slowly] Is it across universities or is it more like—

419 **Participant:** I mean it has to start somewhere right, and if it's [chuckle] a good initiative
420 then people will join, right. So...

421 **Interviewer:** Maybe one can... We could note there, so the (.) less complex fields will
422 benefit as well.

423 **Participant:** Yeah, yeah.

424 **Interviewer:** So, that it's clear that this is... It's not a one thing focus [titters], but more
425 like taking into account the general ecosystem. (...)

426 I think what [participant] said is that they will naturally be included in the solution
427 [speaks to notetaker].

428 So, to shift from [inaudible] tasks to more difficult ones, but then you have solved the
429 simple ones along the way.

430 **Participant:** Yeah.

431 (...)

432 **Interviewer:** Can you write that? [speaks to notetaker] [inaudible]

433 Yeah, I mean like similar to what you wrote, less complex domains will benefit
434 automatically or something like that. (...) Okay. Anything else you want to mention or let
435 us know?

436 **Participant:** I feel like turning one of the smileys to a sad one [laughs]. So, let's change
437 the second one to a sad one.

438 **Interviewer:** Okay! [giggles]

439 [inaudible, talking over each other]

440 **Interviewer:** I've used them evenly.

441 **Participant:** "No, not be happy, it should be sad". [sarcastic]

442 [both laughing]

443 **Interviewer:** The opposite direction! [laughs]

444 **Participant:** Yeah. [laughs]

445 (...)

446 **Participant:** Alright.

447 **Interviewer:** Good.

448 **Participant:** Otherwise... Yeah, thanks for this initiative. I think it's good that you are
449 contacting researchers across domains asking about their needs.

450 **Interviewer:** Yeah, thanks a lot for taking the time to talk to us. If you have any
451 questions afterwards, feel free to reach out anytime.

452 **Participant:** Yes, sure.

453 **Interviewer:** Good!

454 **Participant:** Okay, thank you.

455 **Interviewer:** Thank you, have a nice weekend.

456 **Participant:** Have a nice day.