

1 **Interviewer:** And I'll start the recording. Right, thank you for participating in this
2 interview. You make a difference by contributing your story. The aim of the project is to
3 map how researchers work with research data today and what their needs are. The goal
4 is to develop the best possible support for you so that you can focus on your research.
5 To structure the discussion we used a simplified model of working with research data
6 based on the phases of planning, data collection, analysis, accessibility sharing,
7 preservation and reporting. We look specifically at capabilities, a broad concept that
8 describes processes, information resources, competences, regulations, IT support, etc.
9 in relation to the e-infrastructures used by the research. Quite simply, all the support
10 services that are needed to be able to manage research data well through all phases of
11 our research activity. We also place special emphasis on investigating what e-
12 infrastructure is needed to manage research data. E-infrastructure is usually not
13 something you think about in your daily work until it's not available or does not work but
14 you need it for your work as a researcher. And I'm going to share this image. (.) Please
15 keep these different levels of e-infrastructure in mind throughout the interview as it
16 represents the environment that you as a researcher must relate to and move in in your
17 everyday work.

18 We will start from a research activity that is relevant for you, perhaps a current or
19 recently completed project. It's usually easier to talk about a specific case. If you come
20 up with comments of a more general kind those are also very welcome. You should
21 preferably not speak about all abilities and repeat "we are going to be using this model
22 to be talking later". You should not necessarily speak about all the capabilities and keep
23 in mind that this is a model which means that this is a simplification of what reality looks
24 like but it's needed for us to speak the same language and develop the right support. In
25 the rest of the interview, we'll only talk using this very basic wheel with the different
26 phases of our research project. You can tell us how you work with the data at the
27 different stages, but you don't have to stick to them strictly. One of us will briefly note
28 what you tell us on the picture with the wheel. Then we check that what you think have
29 caught, and can correct it if necessary. At the end, we will ask you to indicate which
30 emoji you would chose to represent your overall satisfaction for each of the faces: happy

31 for good, neutral for well, frowning for dissatisfied. We know from experience that it can
32 feel uncomfortable to mark a lot of dissatisfied, but we want to ask you to be as honest
33 as you can. We will record the audio during the interview and it will be transcribed, after
34 which the recording will be erased. What is analyzed is the transcription of the interview
35 and the notes we take during the interview. Do you have any questions before we get
36 started?

37 **Participant:** No.

38 **Interviewer:** No, then I will stop sharing and we will share the note. Right, so let's start.
39 Can you briefly tell us about your research with focus on the data you collect or
40 generate?

41 **Participant:** Okay. Well, I guess I do two different or maybe even three different types of
42 research in terms of data. So, I come from a background doing behavioral research, so
43 mostly experimental studies, which would typically mean then, that I generate data on
44 my own based on a panel or something like that. But I also do research based on data
45 that is being made available to us from our partner companies. So, it could be behavioral
46 data from a customer relationship management program or something like that. It could
47 also be like financial data or other internal data from an organization. And then I guess
48 the third type of data that I work with is based more on interviews of various types with
49 practitioners or sometimes with consumers.

50 **Interviewer:** Right. Now let's move on to talking about a specific project that stands out
51 to you. We will show you the picture with the wheel and my colleague will take notes
52 which you will see the text form from in the meantime. The focus is on the data
53 management throughout the project you are thinking of - before, during and after the
54 project. You are welcome to talk completely freely based on what comes to mind. We
55 have a few things that we want to cover during the interview and may ask some specific
56 questions to capture everything. So, question number two. When you see this picture of
57 a project's phases, can you describe what you do with data during the different phases
58 one at a time? For the planning phase, how did you reason about how to handle the
59 data before collecting or generating them?

Participant: So, I guess I will take the case then of when you get access to data through a partner of some kind. So, I have this big project where we have a lot of data from an agency that works on influencer marketing. OK, that's not so specific, but it's like a vast amount of data of what's been posted, how consumers reacted to it, who has posted it, et cetera, et cetera. And I would say that from this perspective, when you plan, it's more like you talk to various companies and you see if they are willing to collaborate. And then you have to explore what kind of data do they have? What would they be willing to share and who would be a partner who could draw some data from their systems, so to say. And typically this will be very exploratory at the beginning, because you don't really know how they work with data and you don't really know what data they have or how they sort it. And in this case specifically, it was also like you typically collaborate with others, right? So my co-author was the one who met with the agency first and who started the conversation. And then I was brought up and brought into it a little bit later. So this is not behavioral data, which is denoted there, right? This is an example of where we work with data from a company, because I think when I do data that is primary data that I create myself, I think the data management process isn't that tricky, right? But for these types of projects, it's a lot more trickier and it has to do with both like, what can they share and how can they share? And typically this varies a lot also between different companies, how they view it, but many of these more digital companies are quite open to sharing and they typically have quite good data and they collect most of the things so you could really get a rich view of what's going on by collaborating with them.

Interviewer: So do you feel that the planning process works well or did you experience any issues?

Participant: I think like in the planning process, the issues are mostly related to like non-disclosure arrangement and things like that, like what can we share, how can we share it and how can we kind of merge our interests? But, as a researcher I feel that I have quite a good amount of experience in that. But of course there could be things that would be good to have as support when it comes to these like NDAs and how to kind of talk to companies, because typically they would give you something they would give to consultants, right? Which isn't really what we are when we collaborate like this.

90 **Interviewer:** So do you feel that you would need support regarding anything else except
91 for the NDAs during the planning phase?

92 **Participant:** Not really [hesitates] but related and the first step that comes after is the
93 kind of receiving data which is really tricky for me right now, I don't have a good
94 infrastructure for these like really big data sets and also not good support for kind of
95 transferring the data. So in some cases we have to kind of go to a company with a hard
96 drive and, you know, get it directly out of their system, in others we have to use a co-
97 author who can use other kind of structures that they have available at their universities
98 and sometimes we get to work on the platform of that the data provider, but in this
99 specific case then it was my co-author who solved this in [their] academic environment,
100 which is outside of Europe, so [they] has a little bit more freedom I would say [laughs].

101 **Interviewer:** But then when you talk about data and cooperation with co-authors outside
102 of Europe, how do you think about that during the planning phase?

103 **Participant:** [hesitates] I don't know. I mean typically in these types of collaborations
104 somebody has like the lead for the data and in this case then it was not me in the project
105 but my co-author. So based on the institutional setting that [they] was in, [they] managed
106 that and this specific company was also not very worried about these things, right? So it
107 was more us than them I would say.

108 **Interviewer:** Right, do you want to move on to the create collect phase or do you have
109 anything to add on the planning phase?

110 **Participant:** Well, I guess, create and collect. Again, this is not behavioral data. This is
111 data from companies and as I mentioned, right? I feel that I'm very constrained in terms
112 of receiving data. So, I mean, this is typically something that you start talking about quite
113 early and there are requirements from the company but there's also restrictions on us,
114 right? How to store and what to do with it. So, this is something that [hesitates] where
115 we're trying to build a better infrastructure in in the research group where I'm at as well,
116 because we have a good kind of relationship with many organizations that would be
117 willing to share under the right conditions, but they need us to tell them how the right
118 conditions are, and what would be a good thing for both sides in terms of this. And one

119 thing is, for instance, in this case that we use a sample, the data set is so big. I mean, I
120 don't even have resources to open it based on the infrastructure that I have here.

121 **Interviewer:** How big is big?

122 **Participant:** I don't know the details because I don't remember, but it's huge. It's over
123 20 000 influencers across 300 [hesitates] sorry [hesitates] I don't know, it's massive. It's
124 three years of data for [hesitates] I don't remember.

125 **Interviewer:** Okay.

126 **Participant:** It's gigantic.

127 **Interviewer:** If you lacked the infrastructure locally to handle that amount of data, what
128 tools, software and services did you use on your part to work with that data set?

129 **Participant:** So basically what we did is that, we work on it like over Zoom and it's done,
130 not by me, but I'm part of the process and watch the process being performed by others.
131 And then we also, in this case, had to bring in data scientists and also some people who
132 were used to working on this data and they did part of the processing without us.

133 **Interviewer:** So you only did proxy work on the data, you didn't do any hands-on work?

134 **Participant:** No, in this case I couldn't do any hands-on work.

135 **Interviewer:** Okay. Did you turn to any support function at the university to see if you
136 could handle the data in any other way or didn't you feel that was necessary?

137 **Participant:** In this case it was my co-author who was the lead. I didn't do it, but if I had,
138 I wouldn't know who to ask to be honest.

139 **Interviewer:** What did the cost look like during the project for you? Was it mainly work
140 hours or did you have to pony up money for any software or something like that?

141 **Participant:** No, in this case it was work hours for me and then the data scientists that
142 we brought on worked within [large tec company], so they provided a lot of resources.

143 **Interviewer:** What did you do with the data during the create and collect phase?

144 **Participant:** So first, of course, we looked at it to see what we had and then we did a lot
145 of extra kind of analysis with different machine vision algorithms and things like that to
146 code the content of various posts and the comments on those posts. We also did some
147 validation studies with people in panels, who looked at some of these posts to make
148 sure that the machine vision algorithm was not crazy. So, there was a lot of different
149 things like that taking place to be able to work with the data for research purposes.

150 **Interviewer:** And you didn't store any data yourself during the create and collect phase,
151 everything was co-op?

152 **Participant:** Right.

153 **Interviewer:** Do you want to add anything else to the create and collect phase or should
154 we move on to the next one?

155 **Participant:** We can move on to the next one.

156 **Interviewer:** So, the next one is analyze. Do you want to talk in general about the
157 analyze phase and how you experienced that?

158 **Participant:** Yeah, so again, this is a restriction often in the project. So, in this case, it
159 wasn't, because others were doing the analytical work with me as kind of a partner, but
160 not hands-on. What we've done when we have had the analytical work here, is that we
161 have had some kind of Amazon account to help us or other kind of cloud services, but I
162 haven't been in charge of that, so I don't know the details of it. We've used help from the
163 [research center] to set up, and then we pay on a usage basis. And then of course, if
164 you work with some companies, such as several of the retail companies, they don't allow
165 you to use Amazon for their data, then you have to find a different solution.

166 **Interviewer:** What solution is that as a rule?

167 **Participant:** I'm not sure. It could be other kinds of software providers. [hesitates] I'm
168 not sure what we've used to be honest. I've only been part of the Amazon Web Service
169 projects, but I know that they've used others.

170 **Interviewer:** And the partners are concerned that Amazon will access the data in some
171 way if it's stored on their servers?

172 **Participant:** It's a competitor for them and they have a policy internally that they don't
173 use it, and then they don't want us to use it either.

174 **Interviewer:** Right. So, basically the same question for the analyze phase (.) Did you
175 need support for anything during this phase? Did you know if you had access to this
176 support at [university]?

177 **Participant:** No, so the solution here was, of course, to go through the co-authors and I
178 wouldn't know if we had it here.

179 **Interviewer:** But what support if any were you lacking? What would you have liked to
180 ask, if anything?

181 **Participant:** It's difficult for this specific project but I guess the support would have been
182 [hesitates] It would have been nice to know if I could do some analytical work myself,
183 what infrastructure we had for that and things like that. But my view was that we didn't
184 have it, so I just didn't really explore it.

185 **Interviewer:** Do you have anything you want to add to the analyze phase?

186 **Participant:** No, I'm good.

187 **Interviewer:** So, then we'll move on to the make available phase. Do you want to talk,
188 as with the other phases, in general what you did with the data during the analyze
189 phase?

190 **Participant:** So, here of course it's tricky because when you have the data from a
191 company you cannot like use [hesitates] share all the data freely. We use [online
192 repository service] and we share parts of the like [hesitates] all the code and also, like,
193 descriptives and on a level where it's comfortable for the partner and still make sense
194 from a research perspective to get a better sense of what we've done and what the data
195 really is. The co-authors were in charge of this, but we have like an [online repository

196 service] page for the project where we've uploaded a lot of aggregated information or
197 overview information, but not the actual data.

198 **Interviewer:** How do you feel this phase worked?

199 **Participant:** I mean, given that I had co-authors who were quite used to the process, it
200 was rather smooth. So, and I mean, the [online repository service], it's kind of easy to
201 use.

202 **Interviewer:** So, you didn't experience any issues during the make available phase?

203 **Participant:** I mean, of course, I mean, it relates to this first phrase with what you
204 cannot and can do based on your agreement, right? But even that [hesitates] that was
205 already solved, I didn't see this as a big problem. Then, of course, some of the journals
206 are requesting a lot of open data type things, but they typically understand that if you
207 work on company data, you cannot just share it freely. So sometimes you also share
208 more with the reviewers than you would do on your [online repository service]. But in this
209 specific case, we didn't have to do that. They were happy with what we provided there.

210 **Interviewer:** Right. You talked about code and so on. Which software and tools did you
211 use to produce the code? And what did the code do?

212 **Participant:** So this was not me, right? I think [they] used mostly like R and various
213 packages.

214 **Interviewer:** And the R code was what you posted when you made it available
215 afterwards?

216 **Participant:** Right. In the end, we ended up doing something really weird that nobody
217 has done before. So, I don't know the details of it. But yes, all the code is available on
218 the [online repository service].

219 **Interviewer:** Right. Did you store anything else apart from the code?

220 **Participant:** Yeah, all the descriptives and a lot of visuals of the data structure and
221 things like that, but not the actual data.

222 **Interviewer:** Which formats and software did you use for those?

223 **Participants:** I didn't do them, but I mean, they're all kind of uploaded on an HDM. It's
224 the format of the [online repository service], right? So, [they] just uploaded the output
225 there. I wasn't part of the process.

226 **Interviewer:** Did you need any support from the university during this phase? Or did you
227 feel that everything worked?

228 **Participant:** So, this was the co-author who was in charge of it. So, I didn't have to do
229 the work for this HTML.

230 **Interviewer:** How did you discuss making available data during the project? Did you
231 discuss pros and cons and so on and the reasons for making the data available?

232 **Participant:** Yeah, I mean, so both we discussed, of course, what we were allowed
233 based on the agreement with the partner. But then also in terms of data, since it was so
234 big, the data set, so the processing demands for making too much available and the
235 time needed to do certain things were restrictions, right? So, we were within the author
236 team, we discussed what is needed. And then, of course, we got some reviewer
237 feedback on what they wanted and we tried to kind of balance the openness with the
238 resources available and the restrictions we had on sharing, I guess.

239 **Interviewer:** And when you were contemplating processing a big data set, did you only
240 consider local resources at your university? Or did you consider if there were any
241 national infrastructure that you could have used as opposed to something locally at the
242 university?

243 **Participant:** Again, this wasn't my resources that we drew on, so it was based on the
244 university resources that they had. But then also one of the partners was within [large
245 tech company], [they] was employed there, right? So, [they] had more resources than
246 most of us for working on the data.

247 **Interviewer:** Ok. Do you want to add anything to this phase or do you want to move on
248 to the next one?

249 **Participant:** Move on to the next one.

250 **Interviewer:** Right. So, the next one is preserve.

251 **Participant:** Yes.

252 **Interviewer:** So, how have you thought about preserving the data? And how do you feel
253 that has worked?

254 **Participant:** This is a bit tricky. I think we haven't really ended the project. We still have
255 some ideas on how to work with the data and so on. But I guess, like, preserving the
256 [online repository service] and what we put there is the main source of, like, preserving
257 the project. And we can add on when we do more, or if we do more papers on it. So, to
258 link what we've done with the data set. Whether or not we are allowed to keep the data
259 for a longer time, I don't remember the details, to be honest, in the agreement and it's
260 my co-author who leads this as well, right? So, [they] knows more about it than I do. But
261 I guess like the way the [online repository service] looks now, there's a [hesitates]
262 preserving at least the current contributions that we're making, so that it can be like, you
263 have access to the code and you have access to the descriptors, you can look at it and
264 do other types of studies based on the aggregate so to say rather than the specific data
265 set.

266 **Interviewer:** But you haven't been allowed to store any of the data that you worked on
267 long term?

268 **Participant:** I don't remember how that looks, so I'm not sure if we [hesitates] how long
269 we can keep it to be honest, but I'm thinking now more on preserving it for like
270 [hesitates] based on the research that we published, rather than the data that we have in
271 general for the project.

272 **Interviewer:** But, if someone asks for the data in five or ten years, have you stored any
273 documentation or something like that, so someone could ask for the same specific data
274 in the same specific way from the data company?

275 **Participant:** Yeah, I mean, that's the idea, that the [online repository service] and the
276 way we store it there, because the data we will never be able to give to anyone based

277 on how the agreement looks. But then whether we will be able to keep it forever, I'm not
278 sure. I don't remember, and I don't have the data myself, so I have been fortunate
279 enough not to need to think about that.

280 **Interviewer:** And were there any costs associated with the preserve phase that you
281 know about for you locally or at where they preserve the data?

282 **Participant:** [hesitates] The [online repository service] you don't pay, right? As far as I
283 know. Then, of course, I mean you need to have storage to have the data on, but that's
284 part of the infrastructure that they have. I don't know if they pay for it in any extra way,
285 I'm not sure, but I don't think so.

286 **Interviewer:** At your co-authors university you mean?

287 **Participant:** Yes.

288 **Interviewer:** Did you need any support during the preserve phase or would you have
289 wanted any support during the support phase?

290 **Participant:** I have a co-author who's very comfortable with this, so [they] did most of
291 the guidance and most of the actual work. So, I guess, that's typically how you would in
292 a research project you will get people together who can do various parts of the process
293 and help each other out. So, some are stronger on the data management, some are
294 stronger on the topic and so on. So, I guess that's how I typically work with this.

295 **Interviewer:** And then we come to the last phase, which is the report phase. How did
296 you work with data during this phase? Did you experience any issues during this phase?

297 **Participant:** Not really. I mean, I guess, reporting, we do it continuously throughout the
298 project, right? So, initially with the partner there's a lot of more descriptives that we
299 share with them. Things that they haven't had the time or resources to look at in terms of
300 what you can learn from the data. And then, of course, like conference papers and
301 presentations and then articles. So, I don't know, there's [hesitates] yeah [hesitates] no
302 really, issues there I think.

303 **Interviewer:** Were you missing any tools, software or support during this phase?

304 **Participant:** No.

305 **Interviewer:** Right. So, then we come to the last square, which is the dream scenario.

306 So, if you could go through the whole process again, what would the dream scenario
307 have looked like?

308 **Participant:** So, the dream scenario would be that I could do this internally here, rather

309 than through co-authors. And then, of course, I would have wanted to have like a clear

310 infrastructure, like this is how I can transfer the data, this is where I can store it.

311 Potentially, also, some support in kind of doing some of the initial analysis and guidance

312 on where to kind of get the processing capabilities needed for it. Then, I'm not [hesitates]

313 I mean [hesitates] the type of modeling that we ended up doing, and that's not part of my

314 skill set, so I wouldn't do that myself. So, then I would need to also know [hesitates] how

315 could I collaborate with others based on the data that is now in our infrastructure and

316 what are the tools made available for such co-authors to come in and work on it as well?

317 (...) I guess it would be good to have the research assistance that they tend to have at

318 [university department] or [university department], where they have people who can

319 really work on the data and do things initially for you and help you set it up in a good

320 way.

321 **Interviewer:** Do you have any dream scenario for the rest of the process, besides

322 working with the data locally?

323 **Participant:** I guess that was the question, right? I mean, like (.) writing a paper, I have

324 to do that myself, right? And putting it on [online repository framework], that's not so

325 tricky either, right? It's based on the initial work. So, I think, it's the early work and up to

326 the analysis that is the tricky part in this process.

327 **Interviewer:** Right. Did you have any dream scenario support, except for research

328 assistance and the infrastructure?

329 **Participant:** No, not really.

330 **Interviewer:** Right. So, now we are approaching the end of the interview. Can you look
331 at the picture of what you have said and tell us if there is anything you want to change or
332 something that should be added?

333 **Participant:** (...) I guess in the plan [phase], it doesn't really talk about the like non-
334 disclosure things. Maybe it's implicit, but I don't see it.

335 **Interviewer:** So, we can add NDAs to the plan phase.

336 **Participant:** I think that would make sense.

337 **Interviewer:** Is there anything specific regarding NDAs that you want to add to the
338 planning phase?

339 **Participant:** It would be good to have support for that.

340 **Interviewer:** Right.

341 **Participant:** (...) I guess also potentially for like ethics approval, things like that, if that's
342 needed. It wasn't in this case, but (--) Because, they had in this case the agency had like
343 an agreement with all the influencers, that their data could be used for this purpose. So,
344 it was already dealt with.

345

346 **Interviewer:** But basically more legal support during the planning phase.

347 **Participant:** Yes.

348 **Interviewer:** Right. We now want you to add one of the three smileys to describe where
349 you were satisfied, where you have no opinion, or where you were dissatisfied or feeling
350 annoyed, one for each face. Would you like to give a brief justification of your choices so
351 that we can more clearly see what infrastructure needs to be developed?

352 **Participant:** Yeah, I guess with the plan I'm neutral, because, I mean, also that comes
353 with experience I have set up for doing it, but I guess it could be easier. And I see that,
354 especially when you're more junior, you really, like, struggle with this because there are
355 many things you don't know that you need to know, and you don't know who to ask and
356 you don't know what to do. The create collect, I think, like the infrastructure we have

357 doesn't really work. So, maybe a negative smiley for the create and collect. (...) At least,
358 maybe there is available at the school, but we don't know about it. And then analyze,
359 maybe neutral. Here there are some people we can ask, but there's no clear direction for
360 it, but it's solvable. Make available, I think that's fine. I think the [online repository
361 service] works, so I could be happy with that. (...) Preserve, I don't know enough about
362 it, so maybe neutral or even a bit dissatisfied. (...) And like report, yeah, sure. I mean,
363 that's not a problem, so I'm happy with that. That's what we do, that's what being a
364 researcher is to me, so that works well.

365 **Interviewer:** Is there anything you would like to add before we end the interview?

366 **Participant:** No, I guess, just that now we use this specific case. Of course, the
367 challenges are quite different if you do experimental work or if you do interviews, but I
368 think this is good.

369 **Interviewer:** Right. So many thanks for your participation. I'm just going to stop the
370 recording.