

1 **Interviewer:** Let's start with the short question just to set the ground. Can you briefly tell
2 us about your research with focus on the data you collect or generate?

3 **Participant:** Okay, so a couple of different types of data. You asked me to think about
4 recent research activities and a particular project that I'm doing together with a PhD
5 student, and some other colleagues come to mind. That is a project where we have
6 been collecting survey data from teachers based in Sweden, [country] and [country].
7 And so, respondents or participants are asked to take a survey that we have created,
8 where they self-report on things like their language, their English language proficiency,
9 their approaches to teaching and then some general demographic details also including
10 their teaching experience and that sort of thing. In that project, we're also testing with
11 means of objective test measures, teachers' English proficiency. And that's the next step
12 of the project. So, survey data, test data, interview data, at the very last phase of that
13 project we will also be interviewing teachers. And looking back at my other research
14 activities in the last decade or so, I would say it's data, one of those three, test data,
15 interview data, survey data, primarily.

16 **Interviewer:** Okay, sounds interesting. So, with this, we can go through this data
17 lifecycle. We put a few recurring questions here that apply to maybe not all, but most of
18 the different phases. Like what do you do with the data during each phase? How does
19 that work for you? What tools do you use? What data do you store and where do you
20 store it? Is this associated to any costs, and do you get support or do you need support?
21 How did you experience data management in this phase? And did you encounter any
22 problems? So, let's start with the Planning phase. What do you do with the data and how
23 do you plan with regards to data?

24 **Participant:** Okay. So, it obviously is very much dependent on the project. If I'm, again,
25 thinking about our most immediate needs in one of the current projects. If we think about
26 the survey phase, which precedes the test phase and the interview phase, what we do
27 is(-- This is an online survey. So, we send this out to targeted participants, who are part
28 of a purposeful sampling of teachers, and they respond to the questionnaire, which is
29 anonymous. And that's deliberate, because some of the questions that are in there can
30 be potentially sensitive, not sensitive in the legal sense of the word, but potentially

sensitive for them as individuals and for their career progression, and so on. They will respond to that, and we use for that purpose Microsoft survey tool or Microsoft Forms, I think it's called, and the data then sits within the Microsoft cloud. We download the file in the form of an Excel sheet, and we store that locally here. Initially, on a shared drive and afterwards on a secure drive, on a hard drive, not attached to the internet. We do that for all the data despite the fact that this is anonymous, so it's really(--) It can't be traced back but just to be on the safe side. The test data is a bit more complex in the sense that we are partnering with [international test institute] which is as you know, one of the world-leading testing institutes responsible for the IELTS test, and so on. And we're partnering with them, and they've offered us free licenses for all the teachers that we want to test. That data will be processed only by [international test institute] so an external party, and they will then provide us with a raw dataset with completely anonymous or pseudo-anonymous data, in the sense that we need to connect the survey results with the test results. But we're thinking of a way of doing that without involving names. So, some kind of pseudo-anonymization is probably going to happen there. The interview data(--) We haven't really discussed the details of this, but if I go back to our project plan, what we say is that we will interview people either in person or online and record the interviews and use some kind of AI transcription service. And then store that data, after it's been cleaned and anonymized. Again, store that in the first instance on a shared drive, and then locally

Interviewer: Okay, I think partially you talked already about the Create and Collect as well and you told us a little bit about the tools. Is any of those what you said, is this associated to costs?

Participant: Yes, but costs that are being covered in-kind by our collaborator in [city of mentioned international test institute] so there(--) They have a vested interest in trying their new test on a new, potentially new cohort of testees. But (.) from my perspective, no costs other than personnel costs. There's a PhD student working in the project. I use some of my time for the data collection, but no license fees because they've been, as I said, the test provider is giving us access to the test for free. No, I can't think of any(--) Possibly some cost for incentives! Though those are very small incentives in the

61 neighborhood of 100 kronor so, so roughly 10 euros, because that's what the tax
62 authorities in Sweden allow.

63 **Interviewer:** Okay. Did you encounter any problems in this planning or data collection
64 phase?

65 **Participant:** Oh, yes!! [laughs]. Data collection is never a smooth sailing. You can plan it
66 and think that you have thought of every possible aspect of data collection, but my
67 experience is that, when that plan is put to the test in real life it is something else entirely.
68 So, we are relying on people giving up their time to be involved in this research. It's not a
69 lot of time; the online survey takes approximately 25 minutes up to 45 minutes for some
70 people to complete. But it is(--). We have to ask people to give up their time and not
71 everybody's willing to do that. So, it's(--). You need to send out lots of requests for
72 interviews for survey participation for test participation, and you can only plan so far. At
73 the end of the day, we are in the hands of our research participants in that regard and
74 they may or may not turn up on the day! It's certainly happened. We've collected data
75 here locally with students as well. I came bringing 50 baguettes with drinks and the
76 students had not been informed by their teacher on the day that we were coming. And
77 they were going to go off and do something else. So, it's(--). Because I deal in people, it
78 is very much(--). I'm very conscious of the fact that I'm asking for people's valuable time
79 to spend with me or with one of the other people in the project, to help us in the data
80 collection. They are the data. And for that reason, I think it's very rare for something to
81 play out exactly as planned. Sorry about the long answer.

82 **Interviewer:** No, I understand. That's interesting. And thinking about how you handle
83 the data, did you get support from IT or was everything just clear for you right from the
84 start when you joined the university?

85 **Participant:** Oh, when I joined the university, that was a long time ago now! I would say
86 I'm absolutely reliant on support. I've had a lot of support from local IT. In here at
87 [department], I've had(--). In another project, I'm working with colleagues in [country] and
88 [country], we're starting data collection now with collecting classroom data in October.
89 I'm meeting with a central IT person in the room where we're going to collect data
90 tomorrow. (.) Has there been any other level of support? Well, I speak to colleagues, of

91 course. I have not reached out to [local University data support function] or you or your
92 colleagues. I know that you come highly recommended. [Colleague] and [colleague]
93 spoke very highly of you. I know that there is such a resource available to us, but I just
94 haven't felt the need to do that yet.

95 **Interviewer:** Thanks for the warm words! And that's completely fine [laughs]. Okay, let's
96 continue with the data analysis phase. What do you do with your data there?

97 **Participant:** Again, it depends on the nature of the data, of course. But if we(--). Let's
98 take the three steps. So, with survey data, or survey-like data, thinking back about what
99 I've done in the past, what you do is, we analyze closed questions, perhaps with the
100 help of descriptive and inferential statistics. We analyze open questions, open survey
101 questions with the help of, for example, thematic analysis. Test data is very much just
102 numbers, right or wrong in the case of most of the test data that we work with. The
103 interview data is, again, very much question of different types of content or a thematic
104 analysis.

105 **Interviewer:** And what tools do you use for those? You said it's like statistical analysis
106 or thematic analysis, but--

107 **Participant:** Yeah, so (.) I'm not a statistician. I always seek the help of others in that. I
108 know that when I worked with colleagues or with statisticians, it's usually SPSS-based
109 because it's easy enough to work with data across platforms and across universities,
110 etcetera., exchange data if it's in SPSS. We use Excel a lot, and I'm not great at Excel
111 either, but I'm a lot better at that than I am at SPSS. I've done some work with NVivo in
112 the past (.) for interviews and survey data both. Yes, I think that's, that's probably, those
113 are the main tools I would say. So again, not very sophisticated, but, but our (.) In many
114 ways, our data is not sophisticated, but I have thought about, professional development
115 in this regard, it's just it's difficult to find the time for it! [laughs slightly embarrassed]
116 Things just keep ticking on and I do want to take some statistics courses, I'd like to learn
117 some programming, I'd like to learn R for example, it seems to be a very useful tool for
118 the kinds of investigations that we want to do sometimes. I know that some of my
119 colleagues use it. I have (.) only started to scrape the surface of sort of the AI horizon
120 when it comes to qualitative data analysis, but again it's (.) I just haven't had time to

121 dedicate myself to, or immerse myself in that very very vast area of professional
122 development yet.

123 **Interviewer:** Yes, but exciting plans! So, with the status quo how it is now, did you or do
124 you usually encounter problems in this phase or is it more that you have plans to do
125 more in the future? As it is now, do you encounter any problems, you mentioned sharing
126 data between different universities or anything like that, do you encounter problems?

127 **Participant:** Some(--) I mean we've had issues with(--) Because one of my closest
128 collaborators for many, many years, used to be based in [region in Asia]. There was
129 always a difficulty with third-party infringement and data security. The people in the data
130 protection office here were very cautious about means of sharing data with the people in
131 [region in Asia], saying things like "anything you share, beware that it is, despite the fact
132 that this is in [region in Asia], it is just as if you were sharing it with somebody in [capital
133 in Asia]". And so on. So yes, we had some issues with that, but made it work anyway.
134 And (.) other than that no real sharing issues. Collaborations have worked well. Another
135 project that I'm currently involved in [Nordic country] they use a super secure platform
136 that I struggled to get into, as do most of the other collaborators [laughs] on that project
137 so it's probably too secure. That's where all the data in that project sits and will sit when
138 we collect data here in classrooms in October, so that will always be transferred directly
139 to that secure platform.

140 **Interviewer:** Okay. Making data available, how does that work for you?

141 **Participant:** Not very well because (.) We are as I said before in the business of dealing
142 with people. People are either wary of, or afraid of, you know, overtly afraid that their
143 data will somehow find its way to the wrong people. If it's students, they are concerned
144 about teachers seeing their results on tests and their responses on surveys, etcetera.
145 When it's teachers they can, there can be concerns about data being shared with
146 management, etcetera. And this is despite the fact that the vast, vast majority of data
147 that we handle is anonymous or pseudo-anonymous. That means that one of the
148 promises we usually make to our research participants is that we will not share the data
149 in any shape or form outside the project. And we realize(--) I certainly realized that that
150 is a loss to other researchers in the field and in neighboring fields. The exception being

151 completely anonymous, numerical data. I'm less concerned about that. But as soon as
152 we're dealing with qualitative responses that can (.) With some research participants we
153 feel that we are honor bound by our pledge to participants not to disclose this data or to
154 share this data outside the project. To not do that.

155 **Interviewer:** So that means, if I understand you correctly also, if you publish something
156 if you write a publication, you don't publish data or parts of the data with that? You(--)
157 But you have the conclusions in the paper?

158 **Participant:** Yes, typically no. And I would say that it is still standard practice within my
159 field. There is (..) it would be(--) I see, and that's why I've sort of(--) This is on my radar
160 because I do think it's (.) theoretically very valuable to have access to datasets from
161 other people. But I would say still very much standard practice within research in my
162 field and closely neighboring fields not to publish datasets, in any way, yes.

163 **Interviewer:** Okay, then let's talk about data preservation. While the data will probably
164 not used be anymore after the project, you still preserve it in any way?

165 **Participant:** We do and (.) thinking back, what we have done is (.) it's either been (.) So
166 I have(--) One form of preservation is (.) this [shows a bunch of papers]! So we
167 sometimes collect, still collect data in paper format and these are the original files
168 submitted or, sorry, documents submitted by test takers for example, with consent forms
169 and everything associated with the testing, including their responses on the test.
170 Because it has their names on them, we have we've never logged these anywhere but
171 on paper. So, I keep them in a in a locked drawer in my room.

172 **Interviewer:** So, they don't exist in a digital format is that correct?

173 **Participant:** Those don't exist in digital format but other times we have scanned files like
174 that and put it on a on a seemingly secure server. It used to be Box, OneDrive these
175 days. I am very unsure how secure they actually are, if there was ever somebody
176 wanting to access that information or there was a data leak, just like last week. I'm
177 always wary of where I store information like that, especially if it has names on them. I'm
178 probably less, I wouldn't say less careful, but less troubled by less secure storage of
179 completely anonymous data. Survey responses that have been submitted anonymously

180 and it is impossible to trace back to somebody else. I will often store them just on
181 OneDrive, for example.

182 **Interviewer:** Yes, I understand. Thinking about long term, did you get any support with
183 archiving data, like from the library, similar?

184 **Participant:** Yes, I've had information. I know that that there are endless possibilities for
185 archiving. I mean we have a legal obligation to store research data for a number of years,
186 possibly five years, I don't remember. And then, when it comes to sharing data. The day
187 I decide to do that, I know that there is a lot of very knowledgeable people just down the
188 corridor from me here who can support me in that so(--) Yes, there's much to be learned.

189 **Interviewer:** Okay, so that means currently you're having your personal ways of
190 preserving the data but no, like [university]-wide solution that you are using?

191 **Participant:** No.

192 **Interviewer:** Okay

193 **Participant:** If you don't count in OneDrive.

194 **Interviewer:** No. I understand. Yes. Okay, let's come to the last point: Reporting. And I
195 mean we touched a bit upon this publication thing. But in general, how do you report
196 results from your studies, what are your channels of communication?

197 **Participant:** 95 percent I would say research articles. The remaining five percent will be
198 the occasional book chapter but I'm not particularly fond of book chapters because they
199 don't have the same kind of impact that a research article does. And I've also written
200 some reports, some government reports and some reports in our report series here at
201 the department. Occasionally (.) so yes, that's also, those are all the book chapters
202 thinking about it. (.) So it's probably the traditional routes, research articles primarily,
203 occasionally book chapters, occasionally reports.

204 **Interviewer:** Do you have any, or do you usually try to reach the general public with
205 your research? Is there like, I don't know, debate articles?

206 **Participant:** Oh yes, we've been quite successful in that regard. My research, which
207 deals with the use of English in higher education for learning and language development
208 perspectives, there's a lot of interest from the public in that. A lot of the work that I have
209 done, the results have been written up as press releases. We've had a good success
210 with, together with the media department here at [university]. Our local communicator is
211 very keen on that. I share it on social media like LinkedIn, ResearchGate, and that sort
212 of thing. And obviously, the government reports that I've written are read by a very,
213 potentially very wide audience, of not just, maybe not general general public, like the
214 man and woman in the street, but a broad audience outside of academia in policymaking
215 roles, political roles, etcetera.

216 **Interviewer:** Yes, right! Great! Anything else that you want to add to this last reporting
217 phase when it comes to data? I understand before that you usually don't report data
218 itself but the results?

219 **Participant:** Yes, we don't report data. I think primarily because we are usually not able
220 to share the data itself and that's why we can't do that. Well, it makes no sense to do
221 that.

222 **Interviewer:** Yes, okay. So, I would like you to add one of the three smiley faces that
223 you see on the top of the slide to each of the phases. And they should represent how
224 you feel during that phase, so your very general experience, "does this work well does
225 this not work well?"

226 **Participant:** Well, okay, so that's my angle here: how do I feel in whether it(--). Does it
227 work well, does it not work well or is it just so and so, is that it?

228 **Interviewer:** (.) Well I, I would assume it's the same thing, right? You feel good if it
229 works well or (.) I'm not--

230 **Participant:** You can be satisfied and think that "this is working well but I'm not, I'm not
231 feeling, it doesn't give me joy" but... (.) So it can be a smooth ride but not(--). Maybe it's
232 not the most fun thing to do and vice versa

233 **Interviewer:** Okay, it's more about "do you think this works well?" Okay. And sort of if it
234 works well, even if you didn't have any support then you can still put a smiley face,
235 because you didn't miss it, okay. So...

236 **Participant:** Okay, so Planning, I like planning! I think it's one of the things that also
237 gives me joy. It usually works well! It's realizing the plan, putting the plan into action,
238 that's sometimes derails [laughs], so a happy phase for plan!

239 **Interviewer:** Okay, collection, Create and Collect the data?

240 **Participant:** As I said, far from always unproblematic. I still like it! I'm very much an
241 empirical applied researcher, I like working with data, but, your question was not that.
242 Does it(--) Is it always a smooth ride? No! So maybe the middle face.

243 **Interviewer:** Analysis phase, how does that go for you?

244 **Participant:** Again, far from always a smooth sail, so I would say the middle face.

245 **Interviewer:** And making available?

246 **Participant:** Because we don't, it sometimes troubles me. I would put the sad face there.
247 I'd like to do it, but it's just not possible. It's a condition of the research that we do, that
248 we can't always share it even if we wanted to.

249 **Interviewer:** Yes, I understand. Okay, data preservation?

250 **Participant:** Yes, I would go with the middle face. It's(--) Or the neutral face. It's usually
251 not a problem. Can be issues but, but nothing major.

252 **Interviewer:** Yes, and reporting your results?

253 **Participant:** Reporting the results, not the data?!

254 **Interviewer:** Well, we talked about the data before, the raw data at the make available.
255 So here I would probably focus on (.) well, whichever you want to include. So, it can be
256 results and raw data.

257 **Participant:** Okay, I'll put the happy face there. It usually works really well. Yes!

258 **Interviewer:** Okay, great! Now my last question to you, please describe your dream
259 scenario when it comes to research with data, and it can be small or big things,
260 whatever you would like to see improved.

261 **Participant:** Okay, dream scenario is the inclusion of a data person, maybe a data
262 engineer like yourselves in the planning process. A statistician in the planning process. (.)
263 A way around the condition that I mentioned before that we cannot make the data
264 available to other researchers, and potentially other stakeholders. There are lots of
265 different stakeholders for my type of research. So, I'd like to, to find ways to, to make it
266 available, without sacrificing the availability of research participants. I would love for
267 there to be access to research assistants. As an investigator, you always have to be part
268 of the research data collection phase, but perhaps I would like to be less hands-on than I
269 sometimes have to be. If we had research assistants that could help with data collection,
270 data preparation, data management, that would be very helpful and would probably
271 mean that I could spend my time on things that I'm more qualified for. (.) Preservation,
272 yes, well, I guess with the help of a data engineer at the beginning of the planning, the
273 question of data preservation would be a non-issue, because then that would also
274 almost solve itself. Yes, I think that's it!

275 **Interviewer:** Okay, great thank you! So, let's take a moment to read through the notes
276 and then you can tell us if you agree or if you think we should change or add something.

277 **Participant:** Okay, do you want me to read through it now by myself?

278 **Interviewer:** Yes, that would be great.

279 **Participant:** Okay, I'll start with the Planning; I'll go quiet for a little bit. (..) Well, what I
280 would say is that the cost that we have are personnel fees, in the sense that it is the cost
281 of staff in data collection. We don't(--) There are no other direct costs related to it, other
282 than small pots of money that, I mean I have applied for funding, successfully, many
283 times in the past. For example, we conducted a survey this year with a representative
284 random sample of students in Sweden that required us to use an external provider of
285 that service, and that was a cost of maybe 50 000 kronor, but usually it's small pots of
286 money like that.

287 **Interviewer:** Okay, so that means, in the planning phase, that's where you were thinking,
288 right? The last part is no personnel fees and that should be only personnel fees, usually?

289 **Participant:** Primarily staffing costs, I would say. Occasionally small funding to make
290 data, to access research participants. Okay, moving on to Create and Collect, local IT.
291 So this is(--) You asked about support, different people to collect. The Create, Collect is
292 fine (..) Analyze. (..) Yep, fine! (..) Yep. Yep, looks good!

293 **Interviewer:** Great! Thank you. Do you have any questions before I end the recording?

294 **Participant:** No, no.