

1 **Interviewer:** So, I will start by telling you a bit about the project and about how the
2 interview will be conducted. The aim of this project is to map how researchers work
3 with research data today and what their needs are regarding data. The goal is to
4 develop the best possible support for researchers so that the researchers, you, can
5 focus on your research. The discussion today will be structured around a simplified
6 model of working with research data. I will show you what it looks like, let me share
7 my screen (...) do you see an image?

8 **Participant:** Yes, I can see it, yeah.

9 **Interviewer:** So, this model, it consists of different phases: planning, data
10 collection, analysis, sharing, preservation and reporting. And we will look
11 specifically at something called capabilities and it has nothing to do with a person's
12 capabilities, but rather it's a broad concept that describes the processes, the
13 information resources, competences, regulations, IT support services and so on
14 that relate to the e-infrastructure used by research. So that is all the support
15 services that are needed to be able to manage research data through the research
16 activity. We will not look at all processes connected to research today but to focus
17 on what e-infrastructure is needed to manage research data well.

18 And when we talk about e-infrastructure this is what we mean [note: shows
19 interview pictures] you don't need to memorize this this is just an illustration of
20 some examples to show things that could be considered e-infrastructure. At the
21 lowest level it can be the systems and networks that you need to store and work
22 with data, it can also be the infrastructures and tools that you use to collect or
23 analyze data, it can be the platforms that you use to collaborate around data or to
24 publish data, and at the very top are some examples of support services that you
25 might use or have access to. And all of this is e-infrastructure. It might not be
26 something that you think about in your daily work, but you think about it when it
27 doesn't work or when it's not available because it's necessary for your work as a
28 researcher. So please keep this image in mind during the interview because it
29 represents the environment that you as a researcher need to relate to and move in
30 during your everyday work.

31 **Participant:** Yes, thank you.

32 **Interviewer:** Today's interview will be done like this. We will start from a research
33 activity that is relevant to you, perhaps a current or recently completed project, it's
34 usually easier to talk about the specific case, but if you come up with comments
35 that are more general, they are also very welcome. I'm going to stop sharing and
36 [notetaker] will show you a different image (...) So this is the same model that I
37 showed you in the previous image but without any text around it. We will talk based
38 on this wheel with the different phases of a research project. And you can tell us
39 how you work with data at the different stages but you don't have to stick to them
40 strictly you can talk freely and [notetaker] will briefly note what you tell us on the
41 picture with the wheel and then we will check with you if you think that we have
42 caught what you have said and you can correct it if it's necessary.

43 And at the end we will ask you to indicate which figure from the ones up in the left
44 corner that you would choose to represent your overall satisfaction with each of the
45 phases happy for good, neutral for well, and frowning for dissatisfied and it can feel
46 uncomfortable to mark a lot with dissatisfied but be as honest as you can and it's
47 not about how you feel that your work has been done in the different phases but
48 rather like how does the environment with regard to digital tools and platforms how
49 does that work for you in the different phases. And as I mentioned earlier, we
50 record the audio during this interview, and it will be transcribed after which the
51 recording will be erased and what will be analyzed is the transcription of the
52 interview and the notes that we take during the interview. So, do you have any
53 questions about this before we continue?

54 **Participant:** Yeah, yeah, I don't have much question, but I think, is there any
55 questionnaire that you ask to each section for example on plan data collection and
56 others or I will say by myself?

57 **Interviewer:** You can speak freely, but I will ask you questions about what's if
58 [hesitates] if there are specific things I want to know. The topic is how you work
59 with data during the different parts so first, if you would like to tell us briefly about

60 your research in general and with some focus on the data that you collect or
61 generate in your work?

62 **Participant:** Yeah, thank you. So, yeah, as I previously mentioned in email, I am a
63 first-year PhD student and I just started working on data, some which includes like
64 phenotypic data as well as genotypic my research is mainly focused on genomic
65 tools and breeding strategy of *Lepidium campestre* as oil seed crop for Nordic
66 conditions. To fulfill that, we focus on genomic data, which can be generated in the
67 sequencing company and sent back to us, as well as the one we generate from
68 greenhouse as well as field experiment. We do have two separate data, which I
69 would say phenotypic data as well as genotypic data. I can say that on that like
70 from starting like from planning to final data analysis, so I wanted to know which
71 one you are looking for or I can proceed, for example, from the phenotypic data
72 how we can collect and what we are doing in that. Should I continue with that?

73 **Interviewer:** You can tell me about both if you think it's relevant. Is it the same
74 project?

75 **Participant:** Yes, it's the same project.

76 **Interviewer:** If we start with the planning phase for example, how did you(--) were
77 you part of planning this project or was it planned when you started already?

78 **Participant:** Actually, this is like, when we start our experiment, it has to be pre-
79 planned because we needed to know which variable we have to collect from that
80 experiment. So, we put those parameters in list and we will wait for that until the
81 plant grows and reach to that stage. Then it is pre-planned one. We do planning for
82 the experiment as well as data collection part.

83 **Interviewer:** This project, is it(--) Do you work with it alone or do you have
84 collaborators?

85 **Participant:** I work with my supervisors. We have research groups. So, in that
86 research group, I do my own project because I'm a PhD student. I have my own
87 single line. I focus on that and generate data based on that.

88 **Interviewer:** Okay, so do you have any(--) and part of this project, is it people at
89 [university] only, or are there also other organizations or universities involved?

90 **Participant:** Other organizations as well. The data is mainly generated by us, but
91 for other work, we do collaboration.

92 **Interviewer:** So, when you were planning the setup of the experiment and what
93 parameters you were going to look at, what tools did you use? Did you use Excel
94 sheets or?

95 **Participant:** Yes, Excel sheets. We do that like variables we needed to collect in
96 that. We use the Excel sheets.

97 **Interviewer:** Did you consult any support services during the planning of the
98 project, or did you do this?

99 **Participant:** Not really. I didn't do that. Maybe in the future, because I also took
100 [course] recently and I know what resources are there. And what, like if I need
101 support, they are already there. I knew that I will do in the future.

102 **Interviewer:** So, did you encounter any problems or difficulties during the planning
103 of the project?

104 **Participant:** Not really. I didn't face any challenge for planning.

105 **Interviewer:** Okay. So, if we move on to the data collection phase, can you
106 describe a little bit about how you collect the data, both the phenotypic and the
107 genotypic data?

108 **Participant:** Yeah, that is interesting. For the phenotypic data, what we do, I
109 usually prepare Google Drive with data collection sheets. Then I use my tablet. I do
110 have tablet. I just record the data directly to the Google Drive so that it would be
111 easily accessible to share and to do analysis and also to reduce the workload. If I
112 collect on some, like, paper sheet, I have to, again, copy to the computer as well. I
113 finish, like, at the same time, I do on Google Drive.

114 **Interviewer:** And this is the phenotypic data then?

115 **Participant:** Yes, yes.

116 **Interviewer:** And the genotypic, or the sequence data, how do you collect that data?

117 **Participant:** Yeah, for that usually we send the genotype, like the samples, to the
118 company. They do sequencing and they send us the genotypic data in, I think, in
119 CSV file form or something like that, if I remember correctly. We get directly from
120 the company, sequencing company platform.

121 **Interviewer:** You send the samples to them and they send you back files?

122 **Participant:** Yeah, yeah files.

123 **Interviewer:** Do they send them on email or do you have some other way of
124 exchanging files?

125 **Participant:** Yeah, that is a(--) I think like it comes to my supervisors, so I didn't
126 ask them like through what they get that. I think through email I would say that.

127 **Interviewer:** Do you know the name of the company that you use?

128 **Participant:** For example, the previous one was [global organization]. [Global
129 organization] yes that one, sequencing company.

130 **Interviewer:** Okay, so have you encountered any problems or difficulties in the
131 data collection?

132 **Participant:** Not much. The problem may be when I was in field that is another
133 environmental issue sometimes it is rainy. Other than no technical issue with data
134 collection.

135 **Interviewer:** Okay. I was also going to ask you if you had consulted any support
136 services? Then you had contacted someone to do the sequencing work for you. Is
137 there any other part of this that you had other people involved in or had support
138 with?

139 **Participant:** Yeah, for this, like other people was also involved. My supervisor is
140 also involved in this. And they do contact with the company and other necessary
141 things. I don't like directly contact the company by myself.

142 **Interviewer:** So, when you have received these sequence results, do you store
143 them on your own computer locally, or?

144 **Participant:** It's like a hard drive. We make also on hard drive and yeah, we do
145 that. And other [university] repository storage, I don't remember exactly, because
146 as I said, like, this is the beginning. I'm not in, in detail. And now everything and
147 talk more from that is a little bit like shallow in that, I would say.

148 **Interviewer:** So, do you know if, in this project, if you have had any costs or
149 applied for any money related to the data collection or data storage?

150 **Participant:** Not really. I did it by myself. I used my personal drive and others, so I
151 didn't pay anything for it.

152 **Interviewer:** Tell me a little bit about how you analyze the data.

153 **Participant:** That is interesting and once we collected the data we use different
154 software. As you showed, I use the R software mainly and I use R software so.
155 Others also like other software, the genomics and other genetic data a genetic
156 software, which is like Structure, Arlequin, so there are several software's which I
157 mainly use for data analysis. So, for this interview, I would say R software.

158 **Interviewer:** I have to [laughs] I think I need to make my window a little bit bigger
159 so I can(--) You can write R, [notetaker].

160 **Notetaker:** Okay.

161 **Participant:** R software yeah.

162 **Interviewer:** Did you use R Studio?

163 **Participant:** Yes.

164 **Interviewer:** When you use R, or?

165 **Participant:** Yes.

166 **Interviewer:** Do you collaborate during the analysis with someone or is it on your
167 own computer?

168 **Participant:** I do by myself but I(--) last time I asking some for support of
169 bioinformatics like to use [university] server. So, they directed me, they guided me,
170 in that so my plan is like, in the coming days or weeks I will use [university] server.

171 **Interviewer:** Okay. Have you had contact with [university data support function]?

172 **Participant:** I contacted, like I forgot their name, [university data support function].

173 **Interviewer:** Okay. Have you encountered any difficulties in this work?

174 **Participant:** I got email. I got like a credential on that. Then I will continue,
175 because I haven't started yet. Just I ask to use(--) I'm facilitating to use that
176 [university] server because it reduces my time when I analyze on the [university]
177 server. Like it substantially reduces the time that I spend on other software like R.

178 **Interviewer:** Okay. And for the phenotypic data, how do you analyze that?

179 **Participant:** Yes, that's almost the same with others. What we do, we see on the
180 Excel, I see on Excel how the data looks like. And after that, I just directly go to R
181 software. Based on that, I use different package from that and do the analysis in
182 the R Studio.

183 **Interviewer:** What kind of data are the phenotypic data? Is it measurements of
184 different kinds?

185 **Participant:** Yes, yes.

186 **Interviewer:** Are there also chemical analysis and things like that?

187 **Participant:** Not chemical analysis, it's like quantitative. For example, I can say
188 plant age and like the other one is inflorescence length and days to flowering, how
189 the plant matures, whether they mature early or late, we record the dates. So such

190 data is grouped into quantitative and qualitative data. We collect that and do the
191 analysis on that.

192 **Interviewer:** Is it all data that you type into your pad or do you have any
193 instruments that log automatically data digitally?

194 **Participant:** That is, for example, for plant data, I use tape, which is in centimeters.
195 Then I measure that and record that data directly from the greenhouse or from the
196 field.

197 **Interviewer:** And have you encountered any problems in this part of the work, any
198 difficulties?

199 **Participant:** Not a problem. I didn't encounter any problem with that.

200 **Interviewer:** No. Have you started this data analysis or is it... ?

201 **Participant:** Yes, yes.

202 **Interviewer:** ...also in the future?

203 **Participant:** Yes, I started the analysis of that one.

204 **Interviewer:** And this is also stored on your own computer or on a hard drive or?

205 **Participant:** Yes, yes, I also store that in [university], like there is a single drive or
206 OneDrive, as we know. I put on that and I also keep on Google Drive, my own
207 Google Drive and cloud platform, Google Cloud, yeah.

208 **Interviewer:** Do you collaborate with anyone around the analysis are you
209 exchanging the data with someone?

210 **Participant:** My research group, my supervisors.

211 **Interviewer:** And how do you share the data then?

212 **Participant:** Like with Google Drive and yes sometimes also if the data is small
213 size we(--) I send on email. With Excel sheets, I put on Excel sheet and just I send
214 some through email.

215 **Interviewer:** Okay, so have you, I guess you are not in the phase where you have
216 published anything yet?

217 **Participant:** Yes, I'm not on that baseline.

218 **Interviewer:** Have you talked about how you(--) what you plan to do further in the
219 project if you plan to publish articles or?

220 **Participant:** Yes, that is that is our future plan and also, as I said, I started analysis,
221 data analysis, some of that, so that will go to publication in the future. So based on
222 that once we analyze then they see our results, we will then go to the like selection
223 of the journal. We are also planned for some journals and based on that we will
224 develop the manuscript and then try to publish.

225 **Interviewer:** Have you thought about whether to publish the data, the raw data
226 from the studies. Do you know if the journals are asking for this?

227 **Participant:** I think that is what we expect because if we are going to publish, that
228 should be(--) The data should be publicly available, so it is (.) they're like, the data
229 will be asked for that.

230 **Interviewer:** Have you discussed where(--) Are you thinking about publishing the
231 data, do you know?

232 **Participant:** Like next Monday we will have a meeting that was our thought. Then
233 this meeting is today so [smiles]. I cannot say anything now myself I will meet my
234 supervisor by Monday and talk about which journal we are going for.

235 **Interviewer:** Yeah, do you have any idea in general where the research(--) The
236 kind of research that you do, where you would usually publish it, if you publish the
237 data?

238 **Participant:** Yes, like from my experience, because I have limited experience, this
239 time I would rather look to my supervisors to say "this journal fit with this" and also
240 they'll see my, like manuscript and suggest me in that.

241 **Interviewer:** Yeah.

242 **Participant:** So, it's good to refrain, to say "this journal at this stage" because I
243 have to talk with them.

244 **Interviewer:** Of course. I would also look at this last phase called reporting, just to
245 ask you do you know if you(--) This project, will be reported in any other way than
246 in articles like will you have(--) How is it funded, will you need to report back to your
247 funder for example, in some way?

248 **Participant:** Yes, like a report will be there because this one is, like, the funding
249 agent in its report at the end of the project, so reporting will be there. And also,
250 when we are going to publish, they will be partner of that. So, everything will be
251 clear and like everybody knows what is going on for the project. And the frequent
252 meeting is also there like when we discuss the project how the project is going on.
253 Through that I think the final report will be there, yeah.

254 **Interviewer:** Do you know if you will need to report anything about the data
255 management?

256 **Participant:** Yes, yes, we have to do that one also because like (.) from the
257 beginning when they do funding they really care about the data, so when I'm
258 saying we are sharing the articles together, they are, we are also sharing the data
259 and everything together yeah.

260 **Interviewer:** Which funder is it for this project?

261 **Participant:** Like [funder] I'm sorry, yeah.

262 **Interviewer:** Okay, so do you know if there was a data management plan in the
263 beginning of the project as well?

264 **Participant:** I'm not sure that, I'm sorry, no I I'm not sure that. Yes, from our side, I
265 do since I started this project, the planning was there, for what I'm doing.

266 **Interviewer:** Have you talked about, in your project, what will happen to the data in
267 the long term after the project is finished?

268 **Participant:** I wouldn't say it is finished. It has to be preserved for replication in the
269 secondary data analysis because(--) So for that we have to preserve in the
270 [university] database in other platform I think, yeah.

271 **Interviewer:** So that would be saved on the [university] server then? Is that the
272 plan?

273 **Participant:** Yes, yes.

274 **Interviewer:** Is there anything else you have discussed about what will happen to
275 the data in the future?

276 **Participant:** Yeah, it will be like stored in(--) what I'm sure is it will be stored, and
277 where and how would it be like clear at the end. And it will store for the long term,
278 what I would know.

279 **Interviewer:** Okay, so if you look at the things that we have written now is there
280 something you would like to correct or change or do you or add or do you feel that
281 this represents what you have described for us?

282 **Participant:** For the analysis you can add another software as well because we
283 are using from [university] software center, the analysis one. Different software
284 from that we use also Minitab, Arlequin and others, other software yeah (...) That is
285 the only thing I would say.

286 **Interviewer:** Okay [laughs]. If you would choose one of the figures up in the left
287 corner to put on the planning phase do you feel, not about how you performed your
288 work, but like how did you feel the work went with regards to, like the digital tools
289 that you used and so on?

290 **Participant:** Like for the planning, you mean?

291 **Interviewer:** Yeah, the planning part of the data management and the practicalities
292 around that?

293 **Participant:** Yeah for that, as I mentioned, we directly collect data so we haven't
294 faced any major problems in that, as I said others maybe, like environmental issue,

295 I mean, when we go for field, we face different, like, rainy and other seasons, so
296 that is the only case. From the perspective of planning and others, I haven't faced
297 much. Because I just, I did, like, single experiment in this. From greenhouse as
298 well as from field, I collected that. So, in these two, I didn't face major problems, I
299 would say. And also, for example, disease problem is not about technical issue,
300 and I would raise over here. I faced, like, on my plant, disease case or something,
301 that may affect my, like, what I planned to collect.

302 **Interviewer:** Yeah.

303 **Participant:** If it helps, that is the other one.

304 **Interviewer:** Okay. So, which of the smileys would you choose for the collection
305 data collection? Are you happy with it?

306 **Participant:** Yes, yes, I'm happy.

307 **Interviewer:** But as I'm thinking about the planning phase, you were not part of the
308 planning so much? Do you feel that worked out for you? Or did you feel like it was
309 difficult that you were not part of the planning?

310 **Participant:** You mean part of the(--) Can you say it again, please?

311 **Interviewer:** The planning phase, how do you feel like that has worked for you?
312 You were not so much part of the planning as I understand it, but do you feel like it
313 was planned(--) the planning was done well, so that it has worked, the plan has
314 worked for you?

315 **Participant:** Yeah, for example from the setting experiment to collecting data, I'm
316 in the plan. Because we develop design, how that experiment is going to be done.
317 So, I'm part of the planning. And maybe like the one (.) the problem may be with
318 like when we collect that and how sensitive that data is. There is no clear
319 guidelines, "this data is sensitive or it's not sensitive or something like that", that is
320 not clear, so we do that by ourself. Maybe that is the case, if the data is very
321 sensitive we have to put some restriction and we have to take care more, so that

322 that may be another problem. There is no clear guidelines “this is sensitive data or
323 not sensitive data”.

324 **Interviewer:** Actually, I have I have a last question and that was what your dream
325 scenario would be, if things, support system and the e-infrastructure, was perfect
326 so maybe that is could fit there, that one of the things that you would wish for is
327 clear guidelines around sensitive data, or?

328 **Participant:** Yeah!

329 **Interviewer:** So, you’re welcome to(--) We can we can jump to that if you want to
330 talk about what would you(--) Is there anything else that you would wish for if things
331 were?

332 **Participant:** That’s the one and also, I know that [university] library is preparing
333 different webinars in the platform. I was taking a look at that. We have to attend in
334 that. And still like (.) practical workshop, like particularly for new beginners. I think a
335 practical workshop is necessary to do proper planning, to final preservation of data.
336 (..) Yeah, that is the only two I would wish for.

337 **Interviewer:** Yeah, okay. So, if we go back to the different phases, If you look at
338 the analysis phase of the work with data, are you happy with that or frustrated or
339 neutral?

340 **Participant:** Yes, yes, I’m happy. And I also took one course, advanced statistics.
341 So, it’s still, like, they are willing to help us. If we need help, we can go and ask
342 them about software. From the bioinformatics aspect also I got help for what I
343 asked, so then I’m happy with that.

344 **Interviewer:** And the thought you have about publishing the articles and so on.
345 You haven’t started with that yet but are you feeling confident or are you worried do
346 you feel positive or negative about that?

347 **Participant:** I would say both if it is possible, [laughs] if it is possible, if not, I think
348 the positive side weigh more.

349 **Interviewer:** Okay.

350 **Participant:** Yeah.

351 **Interviewer:** What makes you feel positive?

352 **Participant:** Because we generated very good data. When I analyze that, I feel
353 confident with that. Then I'm waiting for the reviewers and others, for their
354 comments. I don't know how they do reviewing, with that I have some uncertainty
355 or feeling uncertainty maybe there. Still, I'm confident, so confidence way more
356 than the other.

357 **Interviewer:** Yeah, how do you feel about the long-term preservation like we talked
358 about do you have any thoughts on that general feeling?

359 **Participant:** Yes, for that, like once we publish that I'm sure that we will keep that
360 in the [university] data repository because we can also get, I think digital object
361 identifier, DOI for that and others. So, we will keep that for the long-term purpose.

362 **Interviewer:** Do you know which repository you are planning to use to get a DOI?

363 **Participant:** I have to Google for that, I'm not sure for that. But I may ask help
364 [notetaker] and others for that.

365 **Interviewer:** Absolutely you can. You can definitely turn to us for advice if you
366 have any questions regarding these things.

367 **Participant:** Yes.

368 **Interviewer:** But would that be a happy—

369 **Participant:** Yes.

370 **Interviewer:** ...smiley then for the preservation?

371 **Participant:** Yeah.

372 **Interviewer:** And the reporting? How do you feel about that?

373 **Participant:** The reporting also. Yes, I'm happy with that one as well, because that
374 is the original plan. When they do funding, they are looking for reports. That is also
375 what has to be done.

376 **Interviewer:** Is there anything else you would like to add? We are coming to the
377 end of the interview now. Is there something else?

378 **Participant:** This is all what I have, as a beginner! At the third year, I may say
379 more than this.

380 **Notetaker:** It's true. When we talk to PhD's, they are still learning, right?

381 **Participant:** Yes.

382 **Notetaker:** A senior researcher, for instance, your supervisor, would be able to
383 give other answers. But it's also interesting to see, when you're at the beginning,
384 what do you know about the possible help at [university], etcetera. So, it's fine,
385 really!

386 **Interviewer:** Yeah, this is very valuable for us.

387 **Notetaker:** Yes, absolutely! And it's quite natural that you learn as you go.

388 **Participant:** Yes, yes thank you.

389 **Interviewer:** So [notetaker], was there something you wanted to ask as well?

390 **Notetaker:** No, yes, you have already talked a little bit about practical workshops
391 as a dream scenario, and you have taken courses. So, you obviously think it's a
392 good idea with training and education on like on the side of the actual PhD project.
393 Do you think that would be for everyone?

394 **Participant:** Yeah.

395 **Notetaker:** Would you, this is just between us, but would you send your supervisor
396 to take a course, for instance? [laughs]

397 **Participant:** Yes, because when, like, after I took that course also, it changed me
398 a lot.

399 **Notetaker:** Yeah.

400 **Participant:** Because it tells us something is there. And the way we plan, the way
401 we think, and the way we implement. So, those thing is very important, when we
402 get also practical like what I wanted to say practical workshop. For example,
403 accessing every infrastructure of [university] where, we don't know exactly which
404 infrastructures are there. Sometimes we try to reach out, but we need help still. So,
405 some practical workshop can bridge the gap for that and make more confidence,
406 so that's what I wanted to say. Still the course is very good, besides that kind of
407 brainstorming, kind of showing and also contacting in person, how data was basis
408 operating, something like that, was my dream.

409 **Notetaker:** I can really relate. We have introductory, wel,I presentations workshops
410 for new PhD students in one faculty. In [campus] they are invited to kind of, maybe
411 not a practical workshop, but at least this is available you can contact these people
412 if you need help with this and that. And I would love it if all the faculties could do
413 this. Because you come to a new university and you don't(--) some may have been
414 students here but still, you don't know what's available.

415 **Participant:** Yes, that's for sure.

416 **Notetaker:** Yeah.

417 **Interviewer:** Okay, so if you didn't have anything else, [notetaker], then I think
418 we're done. And like I said, it's very valuable that you have contributed your
419 thoughts and experiences to this. Thank you so much.

420 **Participant:** You are welcome, thank you so much. Have a good day.

421 **Interviewer:** Yes, the same to you.

422 **Notetaker:** Bye.

423 **Participant:** Bye bye.