

1 **Interviewer:** Thank you. My name is [Name], and I have [Name] with me for taking
2 notes. So, let me tell you about the purpose of the project, and then I'll tell you about
3 how the interview will be done, and then we can talk about (.) your research and stuff.

4 So, the point of the project is to figure out how researchers work with research data, and
5 not just how they should work, you know? The goal is to develop better support for
6 scientists in handling research data. So, for the structure of discussion we have a model
7 with, of working with research data, based on the phases of planning and data
8 collection, analysis, accessibility, and sharing, preservation, and reporting.

9 We're going to look at the capabilities; a broad concept that describes processes
10 information resources, competences, regulations, IT support services, etc. in relation to
11 e-infrastructures used by research. So, that's all the support services that are needed to
12 (.) manage research data through all the phases of a research activity. (.) So, I'm going
13 to show you a little picture.

14 (.)

15 **Interviewer:** When I push the record button, it moves the share button, so... [mumbling]
16 There you go.

17 This picture [shares image on screen] is supposed to illustrate the various levels of
18 services that can be relevant at the basic level. There's the basic infrastructure, actual
19 storage hardware and networks for transmitting data and things like this. Then there's
20 the research infrastructures for generating data or processing data, things like NAISS or
21 SNIC or something. And then we have tools and research data services at a higher
22 level where whatever that might be. Things like Zenodo maybe or... things like this.

23 So... (.) Here's the actual map of the model of (.) research that we want to use in this
24 discussion. As you can see, there's different phases. Here is the planning phase where
25 you're planning a research project, generating research data, and then analyzing the
26 research data, making it available to others potentially, preserving it for later use and for
27 the... To guarantee replicability and things, and then also the reporting phase, what
28 you're doing when you're actually writing the paper. So, there's all kinds of activities

29 associated with each of these phases, and they're in the gray boxes around. There's
30 other things that you do as well, but that's the capabilities that was mentioned. So, you
31 can keep them in mind. Things like planning, research, data management, and
32 generating the data. Trying to see here, haven I mentioned everything... [reflecting].
33 Yes! So, during today's interview, we'll be looking at this page instead.

34 (..) [note: interviewer and notetaker discussing technical collaboration]

35 So initially, we'll talk briefly about your research. And then... I'll ask you to decide on a
36 particular research project. And then we'll go through the project, each of these phases
37 and look at what you did and how it worked and that's how it will go. Then at the end
38 you'll have an opportunity to use these smiley faces to basically rank each phase
39 according to how it makes you feel. [titters] And then that's that. Does that sound
40 reasonable?

41 **Participant:** Yeah, but I have a question. When I select a project, is it one that has
42 finished, one that is starting, or?

43 **Interviewer:** Ideally one that that is finished or is nearly finished. One that you have
44 experiences to talk about.

45 **Participant:** Yes.

46 **Interviewer:** But before we do that you can just talk about your research in general. So
47 we get a feel for what you do.

48 **Participant:** Yes, so, my research field is in batteries energy storage. It's developing
49 new materials and understand how they behave inside a battery device. So, that's in
50 general terms. So, it's connected to chemistry and materials science.

51 **Interviewer:** Do you do simulations or are you in the lab making the batteries?

52 **Participant:** No, experimental. I don't(--) And if I do simulations is with collaborators.

53 **Interviewers:** Okay, excellent.

54 **Participant:** Not myself.

55 **Interviewer:** That's great. Do you have a research project in mind?

56 **Participant:** Yes, I have one that is submitted so I think that's sort of close enough to
57 an end. The topic of that project is to develop a fluorine free electrolytes, so one of the
58 components of the batteries to make it less toxic and more sustainable. And it's on the
59 behavior inside the battery.

60 **Interviewer:** Okay, fantastic. Then please, the planning phase... How... Can you
61 describe what you did with data during the planning phase?

62 **Participant:** Well, if I understand correctly here the plan is, like to plan the project, how
63 I will you know what data will I gather right? Is that what it meant?

64 **Interviewer:** Yeah.

65 **Participant:** So, in that sense, yeah, I plan the experiments that I think would answer
66 the research questions that I have and then the type of data that I normally work with is
67 just, I mean, text or data files, Excel-files kind of so I sort of know in advance what type
68 of variables and what type of data I gain, then I go and generate it. And if there is
69 something that I was missing, then I go back to the plan and say, okay, there's one
70 experiment that is missing, so I need extra data to confirm this hypothesis. So, I think I
71 reiterate this.

72 **Interviewer:** So, in the plan, you have a sort of space for completing more experiments.

73 **Participant:** Yes. Yes. I think that's always open or I not always everything I plan
74 becomes what is useful. I have to plan new things. Or new experiments.

75 **Interviewer:** And do you, did you plan for what happens after the analysis phase as
76 well?

77 **Participant:** Not really, I mean, I store all the data that I get, even if it's failure analysis,
78 but yeah, I don't know. I don't have a plan for it. I just save it in case I need it at some
79 point during the project. And of course, the aim is to publish and to store the valuable
80 data at least for a few years, as a backup. But I don't have any other plan beyond that
81 when I collect the data.

82 **Interviewer:** So, the data collection step I guess we can call that the experiment itself.

83 **Participant:** Yes.

84 **Interviewer:** How does that go?

85 **Participant:** That goes, I mean, some experiments go well, some experiments don't go
86 well. And then I repeat it until I get either reproducible data or answers to the questions.
87 So, if something is not clear, then I repeat the experiment or I plan new experiments.

88 **Interviewer:** Is the data from the experiment, does it come from like a measurement
89 machine directly into a computer, or?

90 **Participant:** Yes. Yeah, I think at least in this project, and I think in most of my projects,
91 yes, it is gathered by the computer. But then, I mean, this now we go to the analyze
92 phase. So, the data that is generated is not how I plot it. So, I need to change units or
93 do some calculations to plot it in the way that shows the information that I'm looking for.

94 **Interviewer:** So, it sounds like you have a pretty tight, like, logical loop between the
95 planning, the data creation and analysis where you're going back and forth.

96 **Participant:** Yes.

97 **Interviewer:** Do you have any problems with this the data collection stuff does it work
98 the connection between instrument, and does it go directly to your computer, or does it
99 go to?

100 **Participant:** No, it goes to the computer linked to the machine and then some of them
101 are on the network so I can access the data from my computer in some other cases, no,
102 I have to go to the lab and transfer the data in a USB or whatever and then analyze it in
103 my computer.

104 **Interviewer:** Yes, [notetaker].

105 **Notetaker:** I didn't quite get it. When you say that most data are gathered by a
106 computer, is that the computer synthesizes the data or is it reading from an instrument?

107 **Participant:** It reads from an instrument. So, it's the computer connected to the
108 instrument. Generating the data and collecting it.

109 **Interviewer:** And some of those computers are connected to the network, others, you
110 have to walk over with USB.

111 **Participant:** Now we're trying to improve that in the lab. Yeah. Yeah. I don't know if it
112 works. I mean, I don't go to the lab anymore but yeah that was the case when I was
113 doing the research myself.

114 **Interviewer:** Okay. Is that seen as a problem for like PhD students and postdocs doing
115 the work?

116 **Participant:** For me when I was a postdoc it was, because then I have to be present in
117 the lab and then if I am taking the data I am bothering people that want to start the
118 measurements, so it's easier if it's automatically uploaded and then anyone can take it
119 at any time. You don't have to be present, and you don't have to bother someone else
120 using the computer.

121 **Interviewer:** But don't you have to be present in the lab anyway since you're working
122 with physical materials, they need to be put into a machine?

123 **Participant:** Yes, but a lot of the battery data especially, you run the experiment and
124 the battery is cycling for months, so during those months you don't need to be physically
125 present if you collect the data and you check if the battery is running well then I let it
126 run. I mean I will be in the lab anyway, but sometimes it's nice to do it from the office or
127 from home.

128 **Interviewer:** Yeah, it makes sense for it to upload, great thanks. Tell us more about the
129 analysis phase, how does that work with the data handling?

130 **Participant:** So that differs more from the experiment. So, if I now think of the data that
131 I have for this paper a lot of the battery performance is very straightforward because it's
132 just changing the units or, you know, making or calculating a ratio. So, the type of data
133 analysis is just plotting what I get from the computer, from the equipment. But some
134 other parts require more analysis. Like data fitting. Yeah. What was it called? When I

135 separate one peak into several peaks. Yeah. I forgot the word. But anyway, it requires
136 more analysis and more data handling in that case.

137 **Interviewer:** But this is all done on your laptop, basically.

138 **Participant:** Yeah, that's all done in my laptop. And for that, I don't do it myself because
139 this is, yeah, I had a postdoc or collaborator doing it. But to access the software to
140 analyze the data, you have to be with the VPN. So that's a little bit more tricky. Or you
141 have to be physically in [city].

142 And then other type of data that is microscopy, then that's images. Then I have to go to
143 the equipment and what I gather is just images, different magnifications or different
144 spots in the sample. So that is not a spreadsheet, that's an image. And that often I don't
145 analyze it in my case. Or if I analyze it, it's just measuring lengths or sizes. And that I
146 need to do it in the, I do it in the equipment, in the laboratory.

147 **Interviewer:** Yeah.

148 **Participant:** Because the software for that one is only available at the computer from
149 the instrument. So that I cannot install in my own computer. I don't have the license.

150 **Interviewer:** Yeah. So, it sounds to me like some of the most annoying parts is license
151 stuff. It requires you connecting via VPN, for example. That's why you need a license,
152 the VPN, right?

153 **Participant:** Yeah, yeah. But because, I mean, analyzing the data, even if it's tricky to
154 do it, it's part of my job. I have to do it.

155 **Interviewer:** Well, of course.

156 **Participant:** I can, yeah. And then the VPN is something that is maybe someone else. If
157 we get an extra license or things like that, that can be fixed for me. But it's not
158 something I can do.

159 **Interviewer:** If you have problems with these things, to whom do you turn for support?

160 **Participant:** So, in the case of microscopy, I mean, I don't do anything about it. I
161 understand that it's, yeah, if we only have two licenses for the whole [city], it's clearly
162 someone has decided that no more will be bought, that's my guess, I guess it's because
163 it's too expensive or it requires a lot of, you know, computer power or whatever and so
164 it's not something I have to pursue because I don't think it takes that much time from
165 myself even if I have to go to the equipment, it's not something I have to do every day,
166 so it's not a fight that I need. And then, in the one that I have to fit the data and so on, I
167 know some of my colleagues are developing code in Python to ease or facilitate the
168 analysis and the plotting, so that is simplifying the the work or the analysis part.

169 (..)

170 **Interviewer:** Thanks, I think we can move on to the making available phase. So that's(-
171 -) What we mean by that is sharing data with (..) potentially other collaborators but also
172 more widely. If there's data availability (..) aspect to publication or something.

173 **Participant:** Yeah. (..) This project, for example, it's part of a European project. We did
174 not necessarily share the data... I mean. So, we share the results, we share the
175 conclusions with the partners, but we didn't share the spreadsheet (..) with the partners.
176 I guess no one requested it or it wasn't necessary. We provided the plots or the results
177 as part of the reporting or upon request, but not the raw data.

178 Then we have written the paper, and we submitted it for this Chem Archive previous to
179 submit it to a journal and now it's under review in the journal open access. But the raw
180 data is not published. We just have a statement saying that we will provide it upon
181 request.

182 **Interviewer:** Got it.

183 **Participant:** I would like personally to be better at this and make it openly available.

184 **Notetaker:** When you say that you're sharing the data between partners throughout this
185 European project, how do you share? What kind of service do you use?

186 **Participant:** We share the plots or the conclusions, so it's graphs or images.

187 **Notetaker:** Yes, but how? Do you send email, Dropbox?

188 **Participant:** Ah, sorry. Either emails or(--) In the project meetings, we show the data or
189 in the reports. We write the conclusions and put the graphs and that report goes to the
190 partners and the European Commission for the Project Officer. As part of the
191 deliverables and reporting, yeah.

192 (..)

193 **Interviewer:** When data is provided upon request, how do you do that?

194 **Participant:** You mean sharing the data? No one has asked. [inaudible]

195 **Interviewer:** Okay, so you haven't had to solve that problem.

196 **Participant:** No [laughs]. No. I have asked other authors. And I've never gotten an
197 answer.

198 **Interviewer:** Okay.

199 **Participant:** But me, I haven't ever received, as corresponding author, I've never
200 received a request.

201 **Interviewer:** That's... Doesn't sound great. [laughs]

202 **Participant:** Yeah, I mean, that was maybe a while ago. Maybe our openness wasn't as
203 open then.

204 **Interviewer:** Yeah, I can see that. Do you know how, I mean, if someone would ask for
205 data, do you know how you would do it?

206 **Participant:** I think I have at least the figures that are going to the manuscripts backed
207 up. Then I would go to those files and then export it as a spreadsheet or just, yeah, and
208 share that. (..) I guess. Depends on what they need, if they need a specific point then I
209 will provide that point, if they need the whole graph then I will provide that. But now that
210 I'm saying this, in the ones that are not open access, so all papers of mine with the
211 copyright that I haven't thought about it, if I'm allowed to share or not or if it belongs to
212 the publisher. I don't know. Maybe I would check with the journal first. [titters]

213 **Interviewer:** Yeah, that's interesting. (...) Yeah, okay, thanks, that was interesting.
214 Then Report. (.) I guess we already talked about that during the Making Available(--)
215 Oh, yeah, wait! We forgot this preserve step. That's the idea that, I mean long-term
216 storage of the data, you said that you store the data long term, where do you store it
217 and why and how?

218 **Participant:** Yeah. The data from my PhD(--). So before I moved to Sweden, I have it on
219 a USB, like external hard drive. Then the data from [university], the one that I
220 generated, I have it in... Also, in the external hard drive and also I have it on the
221 network, how is it called, the x... I call it the x, I don't know how it's called. (.) Like the
222 Argos?

223 **Interviewer:** Yeah. You're [mumbling]

224 **Participant:** But then I run out of space there, so then the... That's the one I use every
225 day, so the really old one, I now have it in the... the data portal. So, there I have old
226 data. Then part of the data I have it in Box. So, it's a little bit(--). For different purposes(-
227 -) Because, for example, the problem with the Argos is that I need to be connected to
228 the VPN to access it. So, if I'm analyzing data on a plane or somewhere without
229 internet, then I cannot do that. So that's when I started to work on Box, because Box
230 sort of worked easy. I can do the Argos, I can download it and synchronize, but it
231 became very messy. So, I moved to Box for analyzing the data. And then once it's
232 analyzed and so on, then it's moved to the Argos. You know to gather it in the university
233 storage system, but I'm running out of storage everywhere. [laughs] Yeah.

234 **Interviewer:** Do you have it recorded anywhere where all this data is, the different
235 parts, so someone else could find it?

236 **Participant:** I do not have that.

237 **Interviewer:** Do you have it written down so you can find it yourself if you forget?
238 [laughs]

239 **Participant:** Not really. I mean, I sort of think I know where it is, but maybe I should be
240 a little bit more careful on where specific projects are.

241 **Notetaker:** You say you're running out of storage which is kind of a common theme for
242 the researchers, but can you give us some idea about the amount of storage in
243 gigabytes terabytes, just so we get a rough idea about the order of magnitude?

244 **Participant:** Yeah, maybe I have to say that I'm the kind of person that prefers to store
245 everything and maybe I should clean things a bit, but let me check Box how much I
246 have because I don't know (...) How do I find that information? I have 100 gigabytes in
247 Box and then Argos, I'm not sure how much I have to be honest. Should I check?

248 **Interviewer:** No, you have I think the normal amounts, I would guess. You're not paying
249 for any additional storage?

250 **Participant:** I don't think so.

251 **Interviewer:** Have you considered what to do now that you're running out of storage?

252 **Participant:** So, now that I have the data portal, then I'm putting things there. Now I
253 think I'm fine. Well, Box was at some point really filling up and then I started moving
254 things from Box to the data portal. So now, yeah, my data portal, I use it as a backup.
255 That's where I store things that I don't use anymore.

256 **Interviewer:** Great. So yeah, and then we have the reporting phase, which I think,
257 again, we discussed already. Or do you have anything more to add to that?

258 **Participant:** No, just as part of the European projects or, or national projects. And then
259 they(--) Not in every case, I posted in Chem Archive. I normally do it when I am in a
260 hurry or I want to cite or show that it's published. And then I publish it in journals. And
261 then if we go back to Make Available, I just realized that in some projects, the project
262 provides like a cloud or storage unit where people upload their presentations. In those
263 projects where that is provided, then I upload it there.

264 **Interviewer:** Excellent. That brings us to the end. Is there anything you want to add to
265 anything? Was there any, as far as I gathered, nowhere do you have any special sort of
266 costs associated with data stuff. You don't have an account or anything or any special
267 funding for data storage or any special service associated with data?

268 **Participant:** Not with cost. And now that we can publish open access through the
269 library, then no need to worry about it.

270 **Interviewer:** Yeah. And any support of, I mean, and you don't typically turn to anyone
271 else for help either. You don't ask the library if they have any good ideas about what to
272 do with your data or anything, I guess.

273 **Participant:** No, sometimes I discuss with [colleague]. And then he's the one who told
274 me all these different places where we can do things. I guess he has helped me a bit.

275 **Interviewer:** Your local support network, so to speak. Colleagues.

276 **Participant:** Yes.

277 **Interviewer:** Excellent. Then, yeah, we now want you to use these smileys. You can
278 start with the(--) In the same order we went through or not, it's up to you. But is there
279 any part that is particularly easy where you have a smiley face?

280 **Participant:** Yes, I guess the Planning is easy. I'm used to it. And that Create and
281 Analyze. I think those are where I'm comfortable. (...) Yeah. Then I guess making it
282 available, I can put the second face (note: neutral smiley face) because I think I could
283 be better at making it more available than what I'm doing now. Preserve, I'll put the
284 same second face because, yeah, as I mentioned, maybe I should be more organized
285 or, you know, have clear guidelines where people can find things among my data. And
286 then report, I think I can put the smiley face as well.

287 **Notetaker:** A quick question. You put a neutral face on preserve while you're fairly self-
288 critical, but I guess we're primarily interested in how you are experiencing the various
289 tools that the university and other parties are providing, not necessarily(--) I think this
290 phase should reflect the services, not your own.

291 **Participant:** Ah, right. Then put the sad phase, no? Because the fact that I have to split
292 the data into different things, it's not an idea. And then in that case, you can put the
293 second phase in analyze. Because maybe that's something we can improve.

294 **Interviewer:** That makes sense. So that's done. Now you get to describe your dream
295 scenario. If everything was perfect, what would be different?

296 **Participant:** Wow. So what, in these specific parts you mean, right? So Analyze(--) I
297 think the Planning and Collection would be the same or I think that I don't know how I
298 would change that. But then in Analyze, I think optimizing more the coding or data
299 analysis, so it's more, it's faster. So, if I have a code where I put the data from the
300 instrument and gives me the plus that I want, that would be... [hesitates]

301 **Interviewer:** You want a more integrated pipeline from the instrument to sort of a ready
302 analysis?

303 **Participant:** Yes. Yes. And so that would be ideal. And, you know, and then I guess
304 more license that gives me freedom to work whenever, whatever. And then making it
305 available, maybe having more guidelines on how to do this in a better way, maybe
306 protocols that are easier to follow, maybe that would help organize the data and publish
307 it.

308 And then in the Preserve... Having more space or storage space and maybe separate
309 what is just backup from all projects that is published compared to the data that I'm still
310 dealing with it currently. Separating those two might be a good way. And then having
311 more data so I can have everything in one space. I guess that I can access to it even
312 offline. Yeah.

313 And then report, I think it's fine. Here what we didn't discuss now with, at least in the
314 battery field, the journals ask to report certain technicalities or details and they are trying
315 to have that standardized but it's not yet(--) So then each journal ask in a different way
316 or their own template so maybe that could be also standardized so that the reporting of
317 the data becomes easier. And with that more people would do it, I hope.

318 If I can add on the Planning, then planning to make data available and preserve it in a
319 nice way. I think that would also help. Maybe you would store or analyze and collect
320 data. Or you will deal with data differently if you know you will have to make it available.

321 (..)

322 So yeah, it's tricky... in a dream scenario. That would be the case.

323 (..)

324 **Interviewer:** Well, that was a lot of concrete ideas. I like that. Thank you very much, do
325 you have anything else you want to add or think about?

326 **Participant:** No. I'm wondering if you're going to give me any advice? Or not. [laughs]

327 **Interviewer:** We can get to some advice after we conclude the interview. We are done.

328 Let's see if I forgot something... No! Thank you very much. I'm going to take this
329 material, transcribe it, and then the analysis people are going to do their analysis [titters]
330 on it. Then hopefully, it'll become input on how we can improve things. So, thank you
331 very much.

332 **Participant:** Thank you.