

1 [Interviewer introduces the interview by describing what is meant by e-infrastructures
2 and the research data life cycle, as well how the interview will be structured]

3 **Interviewer:** So, then starting with the planning, which is probably what you're involved
4 a lot in right now. And here are the questions about data, sort of. How is the data aspect
5 of the planning phase working for you? Are you using any tools? Have you identified
6 any data that you will be using yet? And identified costs? Received any support about it?
7 Any issues that you've met here?

8 **Participant:** Yeah, I, for now I think everything is very well, clear, I would say. My
9 supervisor, from the beginning in the research plan, he has already stated which are the,
10 which data we can use. For example, there is the(--). Since I'm working on [field of
11 research], if I did, there's the [university] [dataset]. Yeah, so it's already there we can
12 already use it and work on it. There are also some previous works, so from the PI and
13 also from the literature, which can which we can start(--). Like, the artifacts that they
14 have produced before, so for now I started working on that, using that, using what's
15 already available. Until now, for the first task, I would say, yes, the data is already(--). I
16 don't have an issue with data since I'm mostly starting, not from scratch, but from
17 continuing previous works.

18 **Interviewer:** Okay.

19 **Participant:** Then for the tools I think for now I'm not [inaudible] too much. It's more
20 paper writing and also, I'm not using too much tools except there was the a recent mail
21 about ChatGPT Edu education accounts. This is something that I actually need, I'm
22 trying to set it up but it's still(--). I have, I'm actually having an issue with that since, you
23 recall you need the access to [neighboring university's] account so.

24 **Interviewer:** Right so you need to get help from [university], maybe?

25 **Participant:** Yeah, I mean, you need to order, I have like(--). I was doing that before the
26 meeting actually, but there is some information that may be missing, so I would still try
27 to look more into it. Because I think this is something that is really useful, it's always
28 good to have an AI assistant. But they're making the free plan very, very restrictive so if

29 you have just you try to understand one paper you put it on you upload it to, for example,
30 ChatGPT to query about it and then after a few prompts you have already reached the
31 daily limit, and then you will have to wait for tomorrow to keep digging a bit.

32 [both laugh]

33 **Participant:** It's good to have an account that can help you have access to more
34 prompts, I would say, yes.

35 **Interviewer:** That's good. So ChatGPT and sort of LLM tools are really useful for you?

36 **Participant:** Yeah, sure. I mean, it's only, it's really useful for research and for
37 automating. For example, if you write you need to, you want to reformulate some
38 paragraphs, if you feel that you have written things correctly. Or also, like when you
39 don't understand the paper, instead of just contacting the main authors, which may
40 find(--) to take too much time to answer, and also, it's very useful to just query in LLM
41 about it.

42 **Interviewer:** Okay, I see. Yeah, that sounds really useful, actually. I haven't tried it yet.
43 For papers, at least. For code, I've used it a bit, and so on.

44 **Participant:** Yeah, for vibe coding, yes. But you have to always be in the middle,
45 because you don't know if the code is correct, not correct, vulnerable, not vulnerable.

46 **Interviewer:** Yeah, yeah, exactly. You sort of need to be very specific and very critical,
47 but give some ideas. Alright, alright. So, you have already in a way started or been
48 involved or received data, so you're essentially also a little bit in the collection phase,
49 perhaps, because you have identified datasets and you know what you're going to use,
50 a little bit at least?

51 **Participant:** Yeah, I guess. I mean, you know this phase better than me, so if you feel
52 that we're already in that phase... [chuckles]

53 **Interviewer:** Oh, no, well, let's see. So, you've sort of received information about the
54 datasets a little bit, right?

55 **Participant:** Right, exactly, yes.

56 **Interviewer:** From your supervisor, yeah. And have you had a look at them? Have you
57 used any tools to take a look at them? Or have you had any issues with formats and
58 storage and sharing, and things like that, perhaps?

59 **Participant:** Not really, actually. I had a task for you to work digging, like trying to
60 understand and plot the data from the dataset for my PhD interview. So I worked with it
61 before. It's on Zenodo. So my supervisor just gave me the password and then you just
62 download the dataset and you're free to use it. But I don't know if (--) for now it's still, if
63 it's enough for my PhD, if there are things that we need to update. But for the beginning,
64 I think it's a good baseline to start with.

65 **Interviewer:** Yeah, yeah. That's good. And you use open data as a node, so that's
66 interesting.

67 **Participant:** Yeah, exactly.

68 **Interviewer:** I guess then, from analysis to reporting, you don't really maybe have so
69 much to say at this stage, right? Or do you?

70 **Participant:** Yeah, I mean, I'm working on a first survey paper, so I don't know if it's...
71 (..) But no artifacts, no source code, so not too much.

72 **Interviewer:** Sorry, what was the paper about?

73 **Participant:** It's a survey on [subject]. It's more of a theoretical feasibility analysis so,
74 yeah, I mean you, are we(--) I did it as a poster I presented it to [inaudible] so maybe it
75 can fit in the Make Available.

76 **Interviewer:** Yeah definitely, Make Available, and even(--) Yeah, was this as part of
77 your employment at [university] as well?

78 **Participant:** Yeah, it was first(--) Yeah, work that I worked, that they started doing the
79 first task you mentioned, it's kind of a pre-PhD paper, but still fits in the main context of
80 the work.

81 **Interviewer:** Okay, that makes sense. So, you didn't have any data there, perhaps? Or
82 you had maybe a smaller amount of data that you use or maybe not a classical
83 definition of data as raw data but something else?

84 **Participant:** I would say papers, if we can count them as data.

85 **Interviewer:** Yeah, possibly I mean.

86 **Participant:** Yeah, so it was a paper that this server like was already started by my PI a
87 few years ago, so there was already a list of papers that he was reading and
88 summarizing and working on, so I started based on that. It's not starting from scratch,
89 as I said in the beginning, therefore I think, for find(--) like finding the frame of your work
90 was quite simpler, because you already have a baseline and then you just need to find
91 similar works and update based on more recent works, from the gap of the few years
92 that the work was stored somewhere in on overlay so.

93 **Interviewer:** Right, yeah, in a sense this was also like a literature review of sorts?

94 **Participant:** Exactly yes, yes.

95 **Interviewer:** I see, okay. This is all very interesting for us, because just to give you
96 some other background as well. We're looking at the needs of different types of
97 researchers, so this includes professors, normal researchers but even PhD students, so
98 even if not all the sections are filled for us it's still very useful to see that okay there is a
99 process which we need to support early on as well, so that's very good.

100 **Participant:** Okay, yeah, great.

101 **Interviewer:** Let's see, is there anything else you would like to add to this before we go
102 on to the next bit? I'm thinking of the whole life cycle now.

103 **Participant:** I don't know. I mean I think I've said nearly everything, if there is a specific
104 question like...?

105 **Interviewer:** One specific question that we have then, is the dream scenario which is: if
106 you had to scale away all problems and just have a perfect world, with perfect support,

107 is there anything which you would have wished to have been different until now, in your
108 experience at [university] with relation to data?

109 **Participant:** Ah, with relation to data? To be honest, no idea. I mean of course you
110 would prefer to have much more (..) how to say, guidelines, but in the meantime you're
111 in the PhD but you have to be autonomous, so it has it's a thread off that. But it's(--) It
112 might be difficult in the beginning to, like, find your way, find your research. But it's the
113 main point of a PhD actually, so... [chuckles]

114 **Interviewer:** Yes, yeah, that's true [laughs]. And when you say guidelines, did you have
115 anything specific in mind there? Guidelines for something specific?

116 **Participant:** In the work(--) Like when you're working, it's good to have(--) I mean, what
117 happened to me is that I arrived maybe in the end of the year, so there was the summer
118 break, there was a lot of delays, some people were not here, so it was difficult to keep a
119 constant(--) working in a constant manner. There was a few days in meetings, there
120 was a few stuff, I guess. That was the main issue. Otherwise, I think some things were
121 really fine for now.

122 **Interviewer:** So, if I understand correctly then you sort of missed the support from your
123 local unit in getting started, because it happened to be.

124 **Participant:** No, not really. Because I mean, not from the administrative perspective,
125 because I was really well. Everything was well explained and detailed and was really
126 well supported by all sides actually. But more let's say the research part, so you're
127 reading the papers, you're trying to summarize, trying to find your ways and all, and
128 then you try to make something of it before meeting with the supervisor or co-supervisor.

129 **Interviewer:** Right, yeah, okay. That makes sense. Anything else that you want to add
130 here to the dream scenario?

131 **Participant:** Not really, I think, I think that was maybe it, yes.

132 **Interviewer:** Yep, that was good. For the last bit, this is a(--) Just spicing things up a
133 little bit, you know, so we have three smileys on top there, happy, neutral and sad. If
134 you had to put a smiley on the sections that you have been involved in at least, or the

135 phases that you would have been involved in, what phase would you put on planning
136 and creating and maybe making available?

137 **Participant:** I think I'm happy with the planning, because it's good(--). It's actually a very
138 good thing like, when you start you like, from(--). I remember from the very first day the
139 PI presented to me the first task that I can start working on so that you don't get lost. I
140 think this is something that might happen a lot with other PhD students where, I mean
141 I've talked to so many of them where I met in conferences and all, that they said during
142 the first six months of PhD you just don't know your track, you're trying to find your work,
143 trying to understand your work. So, I think I was in a bit of a much better condition
144 because I already have something to start working on from the very first day. I think it's
145 a good smiley for the planning phase. Now for the create and collect, I think it's also,
146 maybe neutral because there were not many things in this phase probably.

147 **Interviewer:** Okay, so neutral then. Alright, well if you've managed(--), is there any other
148 stage which you want to add a face to?

149 **Participant:** I don't know what's(--). What do you usually have in analyze, preserve, and
150 report, what are the usual answers that I know, maybe?

151 **Interviewer:** Oh, well, Analysis usually that's when they're actually using the(--). That's
152 the rough part of the work or the hard part of the(--). Digging in part of the work, so to
153 speak. When they already have the data, now they're doing analysis of the data to test
154 some hypothesis, right. There they're usually talking about specific tools they used or
155 specific software about storage a lot, "I stored my data here, but I had a problem
156 sharing it with someone else", that sort of thing. And Making Available, then that is the
157 sharing in Zenodo or publishing or somehow making the data available, the data that
158 has been either produced or which is the result of an analysis, so primary or secondary
159 data depending on the scenario. Then Preservation, that's usually about more, where
160 the data is stored during, but especially after, the project. What happens to the data
161 afterwards. And Reporting is a bit of a funny one because reporting maybe has more to
162 do about the reporting of the work done during the project, and to what was done with
163 the data, so that is usually quite empty and linked to the publication of an article,

164 perhaps a project report. There is something like that, things like that. And perhaps this
165 you haven't encountered yet, because, of course, you're still very early in your PhD, so
166 that's completely fine.

167 **Participant:** Yeah, exactly, yeah. I think maybe make available(--) I remember, like,
168 when I was, before I presented the process, since it's(--) The paper is still not published,
169 we put it on ResearchGate, like made it available so that we can have feedback from
170 people. I think this is something also that can be mentioned.

171 **Interviewer:** Yes, yeah, so ResearchGate.

172 **Participant:** You can put, yeah, papers in open access, so that you have feedback
173 before you submit.

174 **Interviewer:** Like a pre-print sort of thing?

175 **Participant:** Yeah exactly. Because I was still(--) I had still not finished the paper and
176 then there was [conference name], a conference, and I wanted to present it as a poster,
177 so I had to put the link to the paper on the poster. I saw the PI and we decided to put it
178 on the ResearchGate instead of archive, because ResearchGate is easy to upload and
179 re-upload the update instantly, you don't need to wait for a few days. So, yeah, at least
180 you can keep a first draft of the paper somewhere and then you get feedback, and then,
181 once you're updated, once you finish the work, you can re-update or re-upload, or just
182 delete the draft that you had put on your profile.

183 **Interviewer:** Okay. I see. Alright, I think well from my end, I'm very happy and satisfied
184 with this. Is there anything else that you would like to add?

185 **Participant:** No, not really. Maybe a smiley to make a very...

186 **Interviewer:** Ah, yes, yes, okay, that's true. Good. Perfect. Alright, well, [name], thanks
187 a lot for your time. I really appreciate it.

188 **Participant** Oh, you're welcome. I hope that it has been useful for you.

189 [Interviewer wraps up the interview]