

1 **Interviewer:** Great. Maybe before we take notes, we can have like (.) start with a simple
2 question. Maybe you can briefly tell us about your research with focus on the data you
3 are collecting or generate?

4 **Participant:** Yes [titters] good question! I usually don't talk about, from this angle so
5 please ask if it's not clear. So, I work mainly with buildings and urban neighborhoods,
6 trying to use digital methods to improve the sustainability. So one typical example is
7 energy simulation of buildings. So then we would need data about the existing buildings
8 for example for renovation so we need to, know the address of the building the heights
9 the width maybe we have a 3D model of the building so we have these coordinates and,
10 ideally we also have what we call sometimes metadata, so information about the age of
11 the building, what (--) who is using it how it's used, how many people live there these
12 kinds of data (.) Yes this is so, one concrete example of what we do I think with data
13 [smiles].

14 **Interviewer:** Okay right! And maybe now we can move to a more specific research
15 project that you are doing, currently, because it will be more easy to talk about the data
16 there. Maybe we can start with planning how you do actually plan to collect the data or
17 how you work with this, first phase, of the data management?

18 **Participant:** Yes, I think not, very strategic but I can(--) I have now a project(--) This is
19 maybe a good example, it's funded by [government agency] and it's about renovation
20 strategies for buildings. We have three partners involved, and those are building owners
21 and one, for example, is (.) [municipal housing foundation]. They have 100 buildings and
22 they have different types of data meaning they have to, what I said, the address of the
23 building, the size, they have an energy performance declaration. They also have indoor
24 temperatures measured for every building. If we get the data, I'm not sure how it will (--)
25 usually it's in huge chaotic excel files [titters] and we would need to map it ourselves, in
26 a way, or come up with a format to store and sort it (.) there is no (.) For data that is not
27 related to the geometry there is no clear way of how to do it. For data that is related to
28 the geometry, we work with the digital twin city centers at [university] and they have a
29 digital twin city platform and we get data from them, and we can then use it in different,
30 for example 3D tools we work a lot with Rhino [note: 3D computer graphics software]

31 and get it and then a native format in the specific software that we need and then we
32 convert it in this format that we need.

33 **Interviewer:** Okay, as I understand, you said, the issues to plan the data, later you get it
34 in like chaos and not structured? Or do you face any other issues while planning?
35 Maybe we can just speak about planning for now?

36 **Participant:** Yes, and this way, when collaborating with stakeholders(--). Usually the
37 problem is, in our sector, that there is often no harmonized format of how to store this
38 data and even if there is one person who don't use it. [laughs]

39 **Interviewer:** Okay, and do you have any type of support during the planning phase?

40 **Participant:** (..) I mean in this project I just have a postdoc who is skilled, who [laughs]
41 went through a course for this data management plan, who knows much more on this
42 than me personally (..) other than that, no, there's no (--)

43 **Interviewer:** Okay.

44 **Participant:** There's no support. What we started now at [university department] and
45 you probably know this from [Name] because [they] work with you, I think, or somehow
46 related to [the university data support unit] that they work on this data repository to
47 collect data. But I think this is more than for later phase Make Available so in the
48 Planning, no, I think it's not really.

49 **Interviewer:** Okay, just in case, like (..) I did not notice, you spoke about if there is a cost
50 for this phase or not? The other question, is there any legal issues during this phase?

51 **Participant:** (..) So costs, I think usually it's just our time and of course that's a cost but
52 that's not [laughs] that we have to time. Legal issues is (..) in this project no, because the
53 stakeholders were open to share but, in other projects or when we collaborate for
54 example, there's one project with [municipal-owned energy company]. Where you have
55 to promise that you either, I don't know, you don't share the data or you destroy it and
56 that's a lot of discussion, I was not involved. But then it's a pity that we cannot use the
57 data from a colleague in my group, because somehow, it's limited to them. This postdoc
58 who works now in the project [they] got data from (..) I think the energy agency and, I

59 mean there(--) Our head of department had to sign 10 forms. We actually had to submit
60 a data management plan so there is one but I [laughs], I don't know what is in it. Then it
61 was delivered on a USB drive that only [they] is allowed to use so(--) And very weird. I
62 think that's maybe more in the second phase already now, we'll get the data but yes.
63 This is a bit, the legal issues, that if we get data from other stakeholders, they have their
64 own strange way of handling this.

65 **Interviewer:** Okay, we are already in the second phase. Maybe you can add more about
66 (--) Is it the same, like the cost for example, from your time, hours or there is other
67 resources for cost?

68 **Participant:** For these energy performance declaration then we needed to pay a little
69 service but I think it was, it is not (--) It was maybe 3 000 kronor for the whole project so
70 it's not really something that [inaudible] so, yes.

71 **Interviewer:** The main issues in this phase, during this phase of collection or creation,
72 you said to transfer the data from a resource to another?

73 **Participant:** Yes and (..) usually it (--) [hesitates and sighs] One big issue is that the
74 stakeholders you get involved in the project, and they, they are in the project on paper,
75 they don't deliver the data. Because, often we talk to somebody higher up then they say
76 "yes, some person in the company should deliver the data" but then it never happens
77 either because they don't have the data or they don't take the time. So, I think that's the
78 biggest challenge [laughs]. If we get it, it's chaotic but we can find a solution for it, but if
79 we don't get it then we'll get a bit stuck, so this is a big risk in the project!

80 **Interviewer:** Okay, if you want we can move to the next phase if you don't have
81 something to add here to this phase?

82 **Participant:** I think we can go to the next.

83 **Interviewer:** Okay, great! Then, can you please tell us what kind of tools you are using
84 to analyze your data and how long that takes, for example?

85 **Participant:** (.) Have to think a bit [smiles] because(--) Usually we use it as this data that
86 we get as inputs for our simulation, so in a way we create simulation data ourselves.

87 And we don't spend lots of time analyzing the inputs, as such. It's more that we create
88 than the simulations and then, we use this to analyze (.) So in the example of the
89 building renovations for example, 100 buildings, we simulate one million solutions of how
90 to them and then we just go through them and we'll check if they're plausible, sort them,
91 "what would be the best solutions"? Yes, filter the data, but it's not, it's not a complex
92 analysis [titters].

93 **Interviewer:** Okay, and for the analysis again, do you have a special cost for that and
94 do you get like collaboration from outside of the department?

95 **Participant:** No, but I think this is the phase where we would, where we know best what
96 we do. So I think there's the least need for support.

97 **Interviewer:** Okay. And what is the issues that you can face during this phase?

98 **Participant:** (.) I mean, in this project, I don't see issues in the data as such, other than,
99 I mean, we can make errors in our simulation [laughs] but that's not (--) we're not
100 dependent on others we (--). I have another project I didn't mention, but there, it's
101 different, it's more on urban mining, we collect infrared data for example from materials
102 and this is something where it's outside our field and there we discussed maybe for
103 example electric engineering or they have some other measures to (--) or other sensors
104 to actually collect and then analyze the data, which is beyond our current knowledge. So
105 I think, in this kind of project [titters], it would be nice to have support on data collection
106 and analysis, in a way together.

107 **Interviewer:** Okay. If you think you don't have something to add here, maybe we can go
108 to Make Available?

109 **Participant:** Yes. I think in most cases, it's linked to a publication where we then
110 publish a dataset. The last publication was just like some file in Zenodo or what's the
111 platform?

112 **Interviewer:** I think Make Available and Report they are different, maybe the publication.

113 **Participant:** Aha, OK!

114 **Notetaker:** Zenodo, yes, that's the one.

115 **Interviewer:** So, it's like make it available for other researchers, for example, where you
116 can make it available for them or, and is it–

117 **Notetaker:** Yes, like Zenodo.

118 **Interviewer:** [hesitates] Okay.

119 **Participant:** Yes [laughs] I think yes, I don't know, the difference to Report then, but
120 [laughs]

121 **Interviewer:** And is it, like, you are making available the, like starting from the raw data
122 or after you, make it, like metadata or how you(--) What kind of level of the data you can
123 show and share it to the others?

124 **Participant:** Usually, I would say it's some (.) curated data or it's not exactly what we
125 get from the stakeholders, but it's maybe harmonized in the format. If we are allowed to
126 do it, we work a lot with land material where the [national research centre] gets the data
127 for the geometry. And there we are not allowed to show any raw data, they just(--) it's in
128 there. So, there we only showed the results, or we show some aggregated input data,
129 something.

130 **Interviewer:** Okay, and did you face any issues during this phase, again?

131 **Participant:** Yes, we had one publication where we used the land material data where it
132 was so difficult to find out whether we can use it or not, where we then decided to(--) We
133 had a nice online viewer, we made an "access only" with a password, because people
134 were afraid that it would violate their condition. Yes, I think this unclear legal situation is
135 an issue.

136 **Interviewer:** Okay, then Preserve. How you store your data and, again, are you storing
137 it in all the levels from raw, processing, analyzing, or only the results you are storing and
138 how you do this?

139 **Participant:** Yes, that's [hesitates and laughs] maybe the thing where we also, why we
140 ask this repository to make available but also get better at storing. So far, for example all

141 PhD students are responsible to share it, and then, if they leave, in a way, hand it over
142 to me others in the group in an understandable way, including the raw data. But
143 sometimes this doesn't happen, or not in a very good way, so the documentation around
144 the data is sometimes lacking. With this postdoc who has, is interested in that, we
145 started more, putting small projects results on GitHub (.) openly if it's possible, mostly
146 openly yes. Or use such a git structure also internally, but this is maybe only, so far 20%
147 of the projects that do it, max.

148 **Interviewer:** Okay, so what I understood, you use GitHub as a tool?

149 **Participant:** Yes.

150 **Interviewer:** And you are not saving the documentations for everything?

151 **Participant:** (.) [hesitates] No. So we try, but I would say also with the documentation
152 maybe 20% is documented in an understandable way [laughs] and lots of (.) yes.

153 **Interviewer:** Okay, so the only way that you are preserving your data GitHub, you are
154 not using clusters or?

155 **Participant:** No.

156 **Interviewer:** Okay.

157 **Participant:** And I mean I try to have this project data, but they're just stored on my, like
158 OneDrive [university] thing.

159 **Interviewer:** So, you are using hard drive?

160 **Participant:** No, not even. I mean, I used to just have a hard drive as a backup but I
161 think I haven't [laughs] done a backup in two years!

162 **Interviewer:** Okay, and are you planning for example to save it for longer, like use, for
163 example, the clusters?

164 **Participant:** (.) No. I sometimes thought I should think or should make a plan for it but
165 [laughs] there is currently no plan and I think this would be also something, where it
166 would be nice to say "okay, this is a easy way to do it and an easy and safe way" where

167 I think external support or guideline would help. Or maybe they exist and I just don't
168 know it [smiles] but yes.

169 **Interviewer:** I just want to understand one point. Like is it enough, like for the next
170 researchers if they want to use your data in their research later, in some years. Is it
171 enough, what are you documenting and storing or?

172 **Participant:** I mean (.) I don't know! I mean, I started at [university] six years ago
173 [laughs]! Most PhD students are not finished yet. But I think now we are in this phase
174 where people start leaving, and I had one person who left. And we could not really use
175 [their] data. There are many reasons for this, but part of this reason was the lack of
176 documentation.

177 **Interviewer:** Okay, so the issue here, the documentation, maybe we can write that in
178 this phase.

179 **Notetaker:** I just wanted to ask. You do have the in-house repository, right?

180 **Participant:** I mean yes, we use this local, but I think, at the end, we just use the
181 GitHub, so we don't have any additional.

182 **Notetaker:** Okay.

183 **Interviewer:** Okay. Yes. Maybe if you don't have something to add, you can tell us how
184 you publish your data.

185 **Participant:** Yes, I think this is probably what I mentioned then under Make Available.
186 So, if it's a journal paper and a database link to it, either it's(--) Yes, it's a link online. I
187 think twice, we actually had a paper called data in brief, where it was a separate
188 publication about this dataset. And then they have an additional format how to document
189 and structure it. Otherwise, it's (.) often it's just tables [titters] that are added like
190 appendix to a paper or some online thing.

191 **Interviewer:** Okay, great. Thank you. Could you please look at the notes that my
192 colleague took and maybe you agree or disagree with some of them? You can correct
193 for us, if we put something, not needed or something?

194 **Participant:** (..) Yes, I'm looking.

195 **Interviewer:** Great!

196 **Participant:** (..) No I think that's good, I don't see anything to change. Maybe a general
197 comment is that(--) My feeling is we, me or we as a research area, should do it better!
198 [laughs] But there's... (..)

199 **Interviewer:** Okay.

200 **Participant:** There's somehow little time planned for this, and(--) It, if you don't do it, it
201 doesn't hurt directly; it's only if something goes wrong! So that's, I think why [laughs]
202 we're not doing it actively now!

203 **Notetaker:** It's a systemic problem, I think, so it's(--) Don't feel that you have to carry all
204 the guilt yourself!

205 **Participant:** Yes! I mean that's why I said "great that you do this interview"! I don't know
206 what comes out of it [laughs]. But I feel that this(--) well, yes.

207 **Interviewer:** Yes, maybe, now we can do something else by, if you see these smileys,
208 three smileys. Maybe you can pick if you are satisfied like happy, unhappy, neutral about
209 each face of these six phases. For example, about Planning. What is the level of
210 satisfaction?

211 **Participant:** [sighs] (..) I would say, it's neutral.

212 **Interviewer:** Neutral, for the planning [speaks to notetaker].

213 Yes. And for the Create/Collect? What do you think?

214 **Participant:** If we do it, it's fine! If it's with external stakeholders, it's a pain! So, I would
215 do the negative! [laughs]

216 **Interviewer:** Okay.

217 **Participant:** With Analyze, I'm happy with analyze!

218 **Interviewer:** Make available?

219 **Participant:** I think that's, maybe neutral. [smiles]

220 **Interviewer:** Okay.

221 **Participant:** Preserve is probably (.) not unhappy yet, but could be very unhappy, if it
222 goes wrong [laughs], so yes. Maybe unhappy. And Report, I think that's neutral I don't (-
223 -). It was not a big thing, so far.

224 **Interviewer:** Okay, thank you. We have one last question from our side so we can come
225 to the end. Can you please describe for us what is your dream scenario here? I'm
226 referring to the phases where you are unhappy or neutral. What do you think it's the best
227 scenario to deal with your research data, in regard to these phases, of course?

228 **Participant:** Yes, good question! [laughs]

229 **Interviewer:** I think you are happy only with the analysis, so you must have, a dream
230 scenario for the other phases.

231 **Participant:** Yes so (.) I think I'm(--) [laughs] I am not a type of person who likes getting
232 into the details of stuff, which I think are needed here. I would somehow, would
233 somebody do it for me? And this could be either some, artificial intelligence, somewhat
234 that just reminds me or finds the data that need to be safe, informs me about risk, or this
235 is a person. So, now I outsource this to the postdoc, but [they] cannot do everything. It
236 could be also a person at the [university], as an assistant [titters] who does the job for
237 me. I think with the unhappiness with other external stakeholders, I don't know what to
238 do because I think it's on their side and this didn't change in the last 10 years. I don't
239 know how, this would change much. But I think, regarding our plan (.) some (.)
240 assistance will make it easy, a way to do it (.) would help. Because I assume there are
241 guidelines and best practice and lots of stuff how to do it. But it means extra effort to, for
242 me to learn it, and then to do it [titters], so I think (.) yes, external assistant would be
243 good! Make available (.) I think that's less effort, but in the case that there are legal
244 issues, then, I think we would need more legal support. Because there I would also not
245 trust some chatbot, and if I, for example, reach out to [legal support unit at university] oh
246 [sighs exhausted], it takes so long [inaudible] until I get an answer! And it's also difficult
247 exactly what to ask for, so yes.

248 **Interviewer:** Okay, thank you again. Is there anything you would like to add before we
249 end the interview?

250 **Participant:** I cannot think of anything right now.

251 **Interviewer:** Okay. Thank you very much again, for giving us from your time. And I will
252 stop the recording.