

1 **Interviewer:** Okie doke. And then now we can go to the Milky Way, which you as an
2 [researcher] probably know pretty well. So, this is the entire data management lifecycle.
3 You start in the planning phase where you plan a bit like: how are you going to manage
4 your data, what kind of data are you going to use, how long are you going to keep it for.
5 Then you have the data generation process, so you acquire your data, you find what is
6 relevant and then you start processing it, you go into the analysis phase. Afterwards you
7 might want to make it available and you also might have to preserve it for future use and
8 in the end, you have a report on like what you found in the data and we will discuss this
9 entire Milky Way and just bit by bit, guided by a few questions.

10 I think then we can go to slide number three. So we have some recurring questions that
11 we'll go through, like what do you do with the data, how does it work, how does it work
12 for you, any tools that you use, what do you store, where do you store it, any costs that
13 are involved that are maybe a bit of a hindrance, get any support or do you not want
14 support, do you need it, any problems that you encounter, etcetera. So, first in the
15 planning phase how do you plan what kind of data you get?

16 **Participant:** Okay, yeah. I think this is a comprehensive framework to think about what
17 data or what research looks like. The planning phase is essentially consists of two
18 steps, one is to think up a science goal or a question. That's really the night science the,
19 yeah, I don't know how to describe how you do that. I think a lot of communication with
20 people is important but... [hesitates] and having a lot of collaborations with people that
21 work slightly outside of your own research. I think the next step is more important for
22 me, is I need to write these plans as proposals to [facilities] and that is prior to the
23 creation or collection, we posit several science ideas and say give us x amount of
24 observing time or analyzing time on a certain [facility] or [institute] to work with. I think
25 especially the second step is part of the research I work a lot on. This is technical
26 writing that you need to learn and get better at during your academic career.

27 **Interviewer:** Do you get any support in developing these skills or in this process?

28 **Participant:** No, only within colleagues you get feedback. And so that's both
29 internationally, but also within [university]. There is, yeah, the support is very incidental

30 and not pre-planned. Typically, we organize for large proposal deadlines, specific
31 meetings, but that's all up to us ourselves to decide to sit down and think about these.

32 **Interviewer:** So, this is really an emergent process, but it's not really a top-down
33 framework that you're using.

34 **Participant:** Exactly. There's no top-down framework, but I have always appreciated
35 when we organize meetings around this, so it doesn't come as a surprise, and you're
36 prepared when the deadline approaches.

37 **Interviewer:** Any specific tools that you use to determine which [facilities] you write to?

38 **Participant:** No, I think, especially within [research field], there are very linked options
39 between(--) Things that could help, for example, is workshops or conferences or specific
40 research papers that highlight use cases that you hadn't thought of before. But I would
41 call that more research than forward planning. Yeah, I could go into concrete examples
42 perhaps, but yeah, I think conferences and workshops are the best way to broaden your
43 own horizon on what other [facilities] exist. But otherwise, typically within your field of
44 research, you use a subset of [equipment] that are interesting for your research, and
45 you know those as a person working within the field quite narrowly or quite well.

46 **Interviewer:** Yeah. So, this is really just knowledge that comes with the field.

47 **Participant:** Yeah. Yeah. I think so.

48 **Interviewer:** Okay. Yeah. Yeah, I think, yeah, do you encounter any problems during
49 this planning process?

50 **Participant:** No, it's just, it's very labor intensive and the outcomes are very unforgiving
51 and hard to predict. So currently in my career stage I have enough data but five years
52 ago or so I didn't yet and that's a set of challenges that you need to work with as an
53 [researcher]. You have to generate this, so perhaps it is a little bit in between the
54 creation collection and the planning stage. It could also be that data that I need for my
55 science as I identified in question one or in point one, sorry, already exists on an archive
56 somewhere. And then it's actually possible to go and scavenge this or devise a science
57 plan that can use the data that is, yeah, that's already there. Those are called archives.

58 **Interviewer:** Yeah. Just for our understanding, what is the success rate for the
59 [equipment] that you typically subscribe to?

60 **Participant:** 10% or so.

61 **Interviewer:** 10%?

62 **Participant:** Yeah. One in seven is typically the ratio people use.

63 **Interviewer:** Okay that is (.) that is good to know. I think then we can move to the
64 create and collect phase, so this is about the collection phase, how does your data
65 collection work?

66 **Participant:** It varies from [facility] to [facility]. Some [facilities] have(--) Let's first briefly
67 discuss the archive, about half of the [facilities] have good archives, while the other half
68 do not. If a [facility] has a bad archive that means you just have to email individual
69 people and hope that they read your email, care enough about you to look up in the
70 internal systems. Surely there must be a way of figuring this out, but I haven't worked it
71 out yet. An alternative, there are places with good archives where there's simply a web
72 address where you type in your [reference] or the science projects that you've heard of
73 and you can collect the data and it comes in mostly uncalibrated form to you and then
74 requires a lot of work to become publication ready. So that's on the side of archive.
75 Then in terms of the collection, after your proposal has been approved, they will set up
76 specific, or they mostly sometimes you have to do it yourself, set up specific [collecting]
77 plans which then get executed on the site of the [facility], hopefully. Sometimes bad
78 weather gets in between or it doesn't get scheduled because the priority isn't high
79 enough or it's not feasible because the [equipment] weren't available. So, there's
80 practical constraints as well on their side. But yeah, the data collection is relatively
81 straightforward.

82 **Interviewer:** Do you take this data yourself and do you travel to the [facility]? Do you
83 operate the [equipment] yourself?

84 **Participant:** Smaller [equipment] work like that. And [type of equipment] can still have
85 this way of operating. But the larger the institutes, typically speaking, it's operated by

86 the [facility] themselves. So [facility1] is operated by the(--) so one of the big [facilities] I
87 use is operated by a set of scientists on their end and operators. The same with the
88 [facility2] as well. But other [facilities] like [facility3], for example, require hands-on time
89 to [operate] the [facility] and take all the necessary [measurements], including
90 calibrations.

91 **Interviewer:** Do you encounter any issues or any problems when you do that?

92 **Participant:** No. There is sometimes that you find small mistakes in your setups when
93 you did the planning phase. Then you have to correct them along the way. But yeah.

94 **Interviewer:** And any costs? Are there any costs involved with this?

95 **Participant:** Potentially, it's possible to go [measure] with some of these [facilities] on
96 site. But traveling to [location] for a few hours of [measuring] is costly. So typically, that's
97 all done remote. So yeah, that's often cost limited.

98 **Interviewer:** Yeah, but the data gathering itself... I mean, like gathering the data, does
99 that come with any cost or is that free?

100 **Participant:** No, typically, that's free.

101 **Interviewer:** Have you ever encountered that you had to pay for [facility] time?

102 **Participant:** No I know some [facilities], especially when they're no longer as wanted,
103 can arrange some sort of payment scheme that you pay money to take time with a
104 [facility] but most of the time these [facilities] are funded from the government or
105 international collaborations of governments, so effectively free or already paid.

106 **Interviewer:** And the data and(--) So okay, the [facility] has taken the data, it's taken its
107 [data] or whatever form the data has, do you then take this on a hard disk with you or?

108 **Participant:** Most of the time this is stored online and you collect this data through a
109 web server to your own server to analyze the data.

110 **Participant:** And you're the only one who has access to it?

111 **Interviewer:** Yeah, typically there's a proprietary time, roughly one year for most
112 [facilities], sometimes two or three and sometimes none at all. But there's a certain level
113 of protection on your data, but yeah, I think that's it.

114 **Participant:** Okay so you have like a head start of a year or two years or three years?

115 **Interviewer:** Exactly.

116 **Interviewer:** Yeah, okay, perfect. I think now we can go into the analysis phase. How
117 do you do your data analysis?

118 **Participant:** This is a very haphazard process that depends very strongly from [facility],
119 from instruments to instruments, so a [facility] might have multiple instruments and each
120 instrument typically builds their own data reduction pipeline to achieve publication ready
121 data that takes care of noise that got in the way or interference or artifacts because
122 there's an issue in the detector. What typically happens is that you specialize in a
123 handful of these [equipment] and [measuring] methods. I have about five that I can do
124 myself, but sometimes the [facility] is good enough that they will put you in touch with
125 someone who can calibrate it for you, and then they can help you along the way to
126 make sure that your data becomes publication ready. But because we are often
127 operating on this(--) on the limits of what can be done, we want, and the [researchers]
128 themselves want, to have final say on the different decisions that are being taken to
129 identify what is really noise, what is really signal, what is really there, so I think it will be
130 very difficult to get around this [hesitates] unfortunately, but some of these codes are
131 based on, yeah (.) codes from the 1990s - 1980s even. So, it's good to keep in mind
132 that this can be very old software with a lot of things that just require a lot of time to
133 learn.

134 **Interviewer:** Yeah, and do you ever write your own software for this data analysis or?

135 **Participant:** Yeah okay, so I think the analysis that I was talking to or about is going
136 from the [facilities] to science ready data, but that data then needs to be analyzed or
137 yeah, adapted to actually address the questions that you want to answer. That code is
138 typically written by the researchers themselves fortunately now(--) so in [research field]
139 there's two main codes that people use one of them is python and the other is [field-

140 specific software]. Both systems kind of have a lot of large packages or collections of
141 code that will address key questions or key tasks that you need to do with your data to
142 move from one step to the next. So [research field] in particular has I want to say maybe
143 a hundred different packages or codes that people have developed. Most of them now
144 fortunately within the python framework to achieve specific scientific questions. I say a
145 hundred it could be a thousand as well.

146 **Interviewer:** Yeah, yeah. Are there any costs associated with this analysis process?

147 **Participant:** So fortunately, the community is moving towards Python pretty rapidly.
148 Previously [field-specific software] and MATLAB were used which are paid services, but
149 currently everything is moving towards a free model or a free methodology. In part of
150 analyzing, I have to say that a lot of the data collection can be very big, very large
151 amounts of data. Some projects amounting into the terabytes, multiple terabytes, so this
152 requires servers to host the data and then provide access. Within [university] we have, I
153 have access to two different servers both of which are adequate for the tasks at hand.

154 **Interviewer:** Yeah, is that adequate(--) Does that mean like you're happy with them or
155 is that like they could be better?

156 **Participant:** I am happy with them. I think it could be like any, like many of these
157 processes' costs about half a day to run per [measurement] that I study. Meaning that if
158 a double as fast computer was invested I(--) there are moments in the analysis project
159 that I could work twice as fast, but it's for me not a very limiting point. I know that some
160 people have much faster computers, and they can iterate faster. But, yeah, at a certain
161 moment, I think we're approaching diminishing returns.

162 **Interviewer:** Yeah, okay. And, yeah, I was going to say, you've already mentioned the
163 data that you store. Any other problems that you might encounter in the analysis phase?

164 **Participant:** (...) No if I think of something I'll let you know.

165 **Interviewer:** Okay, perfect, thank you. Then we go to the making available part of the
166 life cycle, so how do you make your data available and if so, how?

167 **Participant:** This is a change that's recently happening within the [research field]
168 community that the data availability statements is now mandatory on most journals. I
169 think this simple change has already made me a lot more aware of how to present data
170 for future use. Typically I don't make my data fully public. I will make myself available if
171 people have questions about my research. I can send them exactly the files that they
172 need. My email is always included in the research work that I do. [hesitates] But I am
173 moving towards uploading most code that I used for the entire paper onto a GitHub
174 page and also, where possible, including direct links to where the data is stored that I
175 used for the research or provide the reduced versions of the data available for people to
176 download straight from the GitHub page. So, GitHub typically is quite limited to the total
177 amount of, well, size of files that you can store there, so in raw megabytes. There are
178 also additional places I can store larger amounts of data. But typically, I make that
179 available by direct contact with me instead of putting that on GitHub.

180 **Interviewer:** Yeah. Do you get any support with making your data available?

181 **Participant:** No, I don't. It would perhaps be nice, in part because writing code for
182 yourself is very different than writing code for someone else to read. First of all, it would
183 make code a lot more efficient and likely remove issues or bugs if people were more
184 incentivized to place their code online. I don't think that's a bad habit to get into, but it
185 would be quite a large amount of work or change my workflow quite significantly if I now
186 wanted to place all my analysis online.

187 **Interviewer:** Yeah.

188 **Participant:** It wouldn't be bad, however.

189 **Interviewer:** Yeah. I noticed that you use the availability of data and the availability of
190 code, you use that more or less interchangeably.

191 **Participant:** Okay, yeah.

192 **Interviewer:** Yeah. Just out of curiosity, is this because you have this data pipeline
193 where you use your own scripts to go from raw data to science products or?

194 **Participant:** [hesitates] I think that, yeah, there's a fundamental difference between
195 data and code. Code is indeed the step that transforms the already good data into the
196 conclusions that you want to draw in your paper or at least in the results part of your
197 paper. I think both would be good to make available to the public and would be a good
198 initiative. I think, or I'm involved in several projects that now put their data online as well
199 in large repositories, often associated with [facilities] themselves. But it is hard to
200 prescribe to other [researchers] how they should analyze their data. And if somebody
201 else gave me their calibrated data, if I was a good scientist and I really wanted to write
202 something meaningful about this data, I would reanalyze everything myself to make
203 sure that I understood every step. Because otherwise I think, yeah, so making the data
204 available is not necessarily the final stop on the trajectory of making my data public.
205 Because many people would want to reanalyze their data. Instead, I see it more as a
206 continuum of how perfect the data needs to be. So typically, what we publish online is
207 about 90% or can address 90% of the science cases. But there might be 10% of the
208 science cases where people need to go again, recalibrate the data, clean it again to
209 make sure there aren't any very small wiggles that would affect a specific science goal.
210 Sorry, maybe this is a little complicated.

211 **Interviewer:** But this is really good information for us to have because we interview
212 people from all different specializations. And it's really helpful to understand the different
213 challenges that people have. And people have different ideas of what they consider
214 data. And the fact that for you, sort of like the script or the analysis procedure is very
215 tightly coupled to the data and you said this cleaning process, that is good to know.

216 **Participant:** Yeah, that's definitely correct.

217 **Interviewer:** Any costs that you encounter when making the data available?

218 **Participant:** I mean, since time is money, it does take a lot of time to make everything
219 available. And as an opportunity cost, I could also be doing other things with my time
220 instead of pushing everything to a GitHub page in a way that anyone can just grab my
221 data and run it. I think there are enough servers around in the [research field]
222 community that if I wanted to host more raw data somewhere, I would always have

223 contacts at either [facilities] that have a large place to host this data or I also think
224 [university] has a service for this, but I've forgotten off the top of my head. But I think it's
225 possible to also host it through [university].

226 **Interviewer:** Yeah. Any other problems that you encountered?

227 **Participant:** [hesitates] No, I think it's a good step that the community is slowly taking.
228 It's also good to keep in mind that this is starting to overlap very closely with computer
229 science. And that within the community, there are people who still code in Fortran and
230 who insist on using [field-specific language] for everything or so there is quite a
231 mismatch in specialties. Like if we solve this issue of putting everything online and
232 readable, if people put Fortran code online that still wouldn't make me be able to
233 interact with that code directly.

234 **Interviewer:** Yeah, I understand. So, there is still(--) There is no unified tooling to use
235 even though you're--

236 **Participant:** [interrupts] It is slowly changing and I think this is a good thing but it's also
237 good to keep in mind that there's a reason people chose this specialty and not computer
238 science as their(--) some people's priorities might be different than other people's
239 priorities. I think having an open collaboration and showing how you analyze the data is
240 good. That doesn't mean that people are wanting to spend out of every 10 hours they
241 work, one hour documenting everything very accurately online and putting it on GitHub,
242 because there is a strong cost associated to putting everything online.

243 **Interviewer:** Yeah. Okay. That is very useful information. I think then we can go to the
244 data preservation part. So, like, after you're done with the data, like, do you preserve
245 your data? How do you preserve it? What do you use for this?

246 **Participant:** Yeah. So, typically, each university I've worked at has their own servers.
247 One time I moved my data from the [country 1] server where I stored my data to the
248 [country 2] server, which was a bit of a pain. But now I also have two hard drives that I
249 use that store about 10 terabytes worth of data from my projects and associated to my
250 projects locally, as well as storing it online on the server from our division.

251 **Interviewer:** So, you have a university resource, you have your own personal resource,
252 and you have a division resource. Do you understand that correctly?

253 **Participant:** No, there's only two resources. The division resource and the university
254 resource are the same.

255 **Interviewer:** Oh, okay.

256 **Participant:** But typically, there's one university resource and there's one personal
257 resource. Most of the time, [inaudible] it's worth of text that is personally written locally
258 on my laptop as well and on the cloud. But any large data can often be generated within
259 a day or two and wouldn't be that much work to recalibrate and get back to publication
260 ready quality data. Typically, I put that just on a couple of hard drives, but if they fail,
261 nothing is lost.

262 **Interviewer:** No. You have the raw initial data, but once you've set up your calibration
263 pipeline to get to the clean data, then it's very easy to rerun that?

264 **Participant:** Yeah, that's relatively easy. And that's also specifically true for the few
265 [facilities] I've specialized in. If there are some other data of [facilities] that I do not know
266 how to recalibrate and those I also store in the cloud, because if I lose those, then I
267 don't know how to regenerate them.

268 **Interviewer:** Yeah, I see. And they don't have their own, the [facilities] don't have their
269 own database of the cleaned data, for example?

270 **Participant:** No that is correct. Maybe a good example to highlight here is a a really
271 nice example of what the [institute] in [city] has recently done is essentially it's a one-
272 person team that decided to make all sorts of analysis scripts for [facility] data and I
273 think that's now one of the most trustworthy databases of data that comes straight out of
274 [facility]. They host it themselves, they put in all the effort themselves, and it's resulting
275 in many extra publications and high-quality data products, but at the cost of an entire
276 FTE (full time equivalent) plus probably a large amount of storage. That's(--) but yeah, I
277 think that that's a very cost-effective way of processing the [facility] data, it's just

278 unfortunate that it wasn't undertaken from within [facility] themselves because yeah for
279 some reason they don't see that as important.

280 Similar ideas can be developed also for many different [facilities], most people working
281 with [facility] always, or the institute I work with, say that the online calibrated data is
282 really nice but transitioning from that to publication level is still one extra step that needs
283 to be done and that because that step is in the way it's still one step away from being
284 publication ready and but making that transition would actually be very competitive. That
285 would be easily worth a couple of FTE, I think. But it's also, yeah—

286 **Interviewer:** So, there are challenges in sort of standardizing this process that might be
287 overcome by a joint community effort, if I might summarize.

288 **Participant:** It could be done like that. Many people think it's a good idea, getting the
289 money or the time or overcoming the opportunity cost of not doing anything else aside
290 from that is a hard step to bridge.

291 **Interviewer:** Yeah. [hesitates] Yeah, I think then we've done the preservation stage. I
292 think then we would like to talk a little bit about reporting like okay, how do you report
293 about your data?

294 **Participant:** One part of it is to do [hesitates] to have an online web page where people
295 can access this data. Part of any good data set is also to transform it into a paper and
296 link subsequently to where the data is available and where the data can be preserved or
297 acquired, I think those are all very useful steps to report online data. You can mention it
298 during conferences, but in the end it's quite hard to do much more than that.

299 **Interviewer:** Yeah, this online website that is your own personal website or?

300 **Participant:** Typically not. Typically, it's linked through a bespoke website linked to the
301 server of that data. If your project is just called project x then you would have
302 projectx.com and then the data would be linked there.

303 **Interviewer:** Okay so this, I'm trying to understand this, this is a project for example and
304 that's, as part of this project you've done several [measurements]?

305 **Participant:** Yeah.

306 **Interviewer:** And then with different [facilities] and you pull the data together and then
307 for this project you make a project website and that's where you store some of the
308 output data after?

309 **Participant:** Exactly, exactly and then typically the project, you can have several
310 projects to work on, but then that project has our own website and that's where all that(-
311 -) so it's a very project-based analysis or project or reporting structure that typically
312 [research field] works with. Counter to that the entire [facility] archive that I talked to you
313 about that is hosted from [city] that is not project-based because that's associated to an
314 individual [facility]. And yet that feels like more of the job of an [facility] themselves to
315 figure out than it feels of the job of an individual researcher.

316 **Interviewer:** Yeah. And organizing one of these websites that's typical or common to do
317 within [research field]?

318 **Participant:** Yeah, I think it's becoming more common. It is a step that is taken without
319 any(--). There's no one forcing you to take that step. But it is a good or a wise choice to
320 make if you want to have a large dissemination of your data and make your projects
321 popular among the community.

322 **Interviewer:** Yeah. Are there any costs associated with this? Do you get support for
323 making such a website?

324 **Participant:** No support, yes, there's costs. Although most of the time, they're
325 organized through universities and then the cost is quite low. You would have to have
326 your own(--). If I wanted to set one up right now and I wouldn't have support from other
327 universities; I would have to ask for funding from within the project I'm currently being
328 funded from.

329 **Interviewer:** Okay. Yeah. Are there any problems that you encounter when doing this?

330 **Participant:** Well, it's not trivial. And yet, probably at the university, there's enough
331 people who know exactly how to do it, or set up a web page that can just host all this
332 data. But in the giant scope of things, I think it's not as difficult as, for example, making

333 all your data public, or your code public. I think that that's more difficult because yeah a
334 website can be made in half a day while transforming all your codes to be legible and
335 robust so that anyone can use it and that it can run on any system is a lot more work.

336 **Interviewer:** Yeah, yeah, I understand. Any particular tooling you use in this case for
337 the website for example?

338 **Participant:** Typically, I host my website through GitHub, make a GitHub page and
339 have that link through a domain I buy.

340 **Interviewer:** Yeah, okay. I think then we're also done with reporting unless I'm
341 forgetting something. Then we come to the smiley phase, so can you, like the planning
342 phase, are you happy with it, are you neutral about it, are you very sad about it?

343 **Participant:** I think it's the funnest part of [research field], that is the most fun part. I'll
344 go happy phase.

345 **Interviewer:** And the creation of the data collection phase how do you feel about that?

346 **Participant:** Data collection phase is [hesitates] yeah, I feel neither happy nor sad
347 about it is just a step that is necessary to take but it depends very strongly on the
348 [facility] and it's going to be hard to make easier without having specialists on each
349 [facility] available.

350 **Interviewer:** Yeah. The analysis part?

351 **Participant:** I like it. I think it's good, but it requires a lot of work. So, I'll go smiley face.

352 **Interviewer:** Yeah. And availability?

353 **Participant:** I'll go neutral in part because I think this is a still changing environment and
354 as the community changes, as more journals require everything to become more public,
355 maybe that'll turn into a smiley face, because then I can rely on it also steadily "oh
356 there's a paper I can just grab their code and do exactly their work" instead of recreating
357 someone's work takes months.

358 **Interviewer:** Yeah and there it's like you're sort of waiting for getting a settled way of
359 doing things.

360 **Participant:** Yeah, I think so. I think the community is moving slowly.

361 **Interviewer:** Yeah. Data preservation? How do you feel about data preservation?

362 **Participant:** I think neutral because I think it's good enough, but I surely can imagine it
363 can get better.

364 **Interviewer:** Yeah.

365 **Participant:** And reporting, I will go exactly the same. I also feel neutral. I haven't felt a
366 lot of help, but also the need for help is pretty low.

367 **Interviewer:** Yes. Okay. Now, in your dream scenario, what would this *Vintergatan*, this
368 galaxy of data management, what would it look like in the different phases? How does
369 your dream scenario work?

370 **Participant:** Okay, I think that's a good question actually. So what parts are actually,
371 what parts are hard to be stuck on? I think the collaborative process is a spectacular
372 part to participate in. Of course, it would be nice if there was more [measuring] time
373 available, but that's not something we can change. Yeah, I wish universities would or
374 [facilities] would change slightly how the creation slash collection of data would happen
375 and it would be nice to see for example more of these archival projects maybe even get
376 funding from within a university. They could say something like, we're now so involved
377 in this instrument or in this institute and there exists already so much data let's invest a
378 year to set everything up, instead of what often happens is that people collect a lot of
379 data, they pick the one or two raisins or diamonds out of it and then they leave
380 everything else while there's a lot of good science that's yet to be done on this front. I
381 think that that's a that's a type of science maybe linked to a type of analysis that would
382 be good to fund from within the university but would be really hard to sell because
383 you're not [hesitates] you're doing steady good science instead of frontier high risk high
384 reward stuff, you just know that you will collect something really valuable and in process
385 of doing that, I think, instead.

386 Data availability, it wouldn't be bad if there was some sort of standardized way that
387 [university] would encourage people to place their data online. For example, my typical
388 way is I place all the data that I use to make a paper online so that every graph has its
389 own Python script that you can just run and it produces exactly this data for example. I
390 don't know what would be a better alternative to this, but I think [university] can think
391 about how to manage putting data online and facilitates this collaboration of availability,
392 mostly on the side of the code, but maybe also a repository of, we see that you publish
393 the paper, put all the data that you use to make this paper on this repository, and then
394 we'll keep it for five years, 10 years. And here's the link that you can go and download it
395 from. For example, something like that wouldn't be bad. That's not provided from
396 journals. That's not provided from [facilities]. But then again, to standardize this within
397 the entire [research field] community might be difficult.

398 Yeah, I think, yeah, that and maybe moving on to the preserve and reporting sides, it
399 could be interesting to have a bunch of people that are really fast with making websites
400 or interfaces for people to collect data from, maybe also using the online archival tools
401 that exist. There is the [institute] within [research field], for example, where you can just
402 upload terabytes worth of data you need to that is responsible for [research field]
403 science so they exist but maybe spreading it in a visually appealing way through a nice
404 website maybe similar to how [infrastructure] also helps with outreach or with online
405 data dissemination.

406 **Interviewer:** Yeah, I think that is it, so I think I will stop the recording here.