WEBVTT

1 00:00:01.883 --> 00:00:02.883 <v ->All right.</v>

2 00:00:04.054 --> 00:00:08.160 In the interest of time, let's go ahead and get started.

3 00:00:08.160 --> 00:00:08.993 Hey everybody,

4 00:00:08.993 --> 00:00:13.710 thank you so much for coming today and this week seminar.

5 00:00:13.710 --> 00:00:16.350 It's my pleasure to introduce Stephen Larsson

6 00:00:16.350 --> 00:00:19.023 and Adria Haimann from Metacell.

7 00:00:20.160 --> 00:00:23.730 This is a few words of context here.

8 00:00:23.730 --> 00:00:25.740 We've talked about, we've had people,

9 00:00:25.740 --> 00:00:28.140 we started this semester with somebody from the hospital.

10 00:00:28.140 --> 00:00:30.120 We've had people from academia,

11 00:00:30.120 --> 00:00:32.940 we've had people from pharmaceutical companies.

12 00:00:32.940 --> 00:00:36.750 And so very excited to present something different.

13 00:00:36.750 --> 00:00:40.320 So Metacell is a company that works

14 00:00:40.320 --> 00:00:42.657 in sort of the research space.

15 00:00:42.657 --> 00:00:44.160 Near and dear to my heart.

16 00:00:44.160 --> 00:00:46.410 They've been, from their beginning, I think,

17 00:00:46.410 --> 00:00:49.923 very active in the computational neuroscience community.

18 00:00:52.044 --> 00:00:56.010 We both contributed to a project called NetPyNE

19 00:00:56.010 --> 00:00:59.763 for building models of computational neurons.

20 00:01:00.810 --> 00:01:04.200 But more broadly, they work in the greater

21 00:01:04.200 --> 00:01:06.381 health informatics space.

22 00:01:06.381 --> 00:01:09.570 And they're going to tell us a little bit

23 00:01:09.570 --> 00:01:11.670 about how we can enhance biostatistics

24 00:01:11.670 --> 00:01:12.750 and health informatics research

25 00:01:12.750 --> 00:01:15.960 through collaborative cloud-based data science tools.

26 00:01:15.960 --> 00:01:17.403 So let's welcome them.

27 00:01:19.500 --> 00:01:21.870 <v ->Thank you very much. Good afternoon everyone.</v>

28 00:01:21.870 --> 00:01:23.700 I can see some of the back of your heads,

29 00:01:23.700 --> 00:01:25.650 so I can imagine that I'm also, you know,

30 00:01:25.650 --> 00:01:27.250 virtually looking at your faces.

31 00:01:28.200 --> 00:01:29.580 Thanks so much for having us.

32 00:01:29.580 --> 00:01:32.940 I'm Adria Haimann and I work alongside Stephen at MetaCell.

33 00:01:32.940 --> 00:01:35.370 And as already mentioned, today we're gonna share with you

34 00:01:35.370 --> 00:01:39.060 some insights into how academics are using cloud-based

35 00:01:39.060 --> 00:01:41.610 collaboration tools to enhance their research.

36 00:01:41.610 --> 00:01:43.230 But before I kind of begin with this,

37 00:01:43.230 --> 00:01:45.240 I wanna provide you with some context.

38 00:01:45.240 --> 00:01:48.390 So, 10 years ago I was in your position,

39 00:01:48.390 --> 00:01:50.310 I was studying health economics

40 00:01:50.310 --> 00:01:51.690 at the London School of Economics,

41 00:01:51.690 --> 00:01:53.730 and I had joined a research team

42 00:01:53.730 --> 00:01:55.560 at the European Observatory for Health.

43 00:01:55.560 --> 00:01:57.210 And I was relatively new to this field

44 00:01:57.210 --> 00:01:59.730 and kind of found myself in a Catch 22

45 00:01:59.730 --> 00:02:01.590 that maybe you can relate to.

46 00:02:01.590 --> 00:02:04.920 So I wanted to know how can someone or a student or postdoc

47 00:02:04.920 --> 00:02:07.710 or researcher discover the best way to collaborate

48 00:02:07.710 --> 00:02:09.630 on their research and use new tools

49 00:02:09.630 --> 00:02:11.790 if you have fairly minimal experience,

50 00:02:11.790 --> 00:02:14.340 neither academia or in industry.

51 00:02:14.340 --> 00:02:17.100 So that's essentially what we want to show you today

52 00:02:17.100 --> 00:02:19.320 and what we'd love to share with you,

53 00:02:19.320 --> 00:02:20.880 if you could go to the next slide,

54 00:02:20.880 --> 00:02:23.940 which is kind of a collection of key topics

55 00:02:23.940 --> 00:02:27.000 of how researchers are doing just that,

56 00:02:27.000 --> 00:02:29.220 while also getting the most out of their data.

57 00:02:29.220 --> 00:02:30.690 So during this seminar,

58 00:02:30.690 --> 00:02:32.880 we're gonna cover different methods that you can share

59 00:02:32.880 --> 00:02:36.480 data analysis and introduce you to a specific cloud-based

60 00:02:36.480 --> 00:02:37.750 collaboration platform

61 00:02:37.750 --> 00:02:40.710 that we've created called Cloud Workspaces.

62 00:02:40.710 --> 00:02:42.750 And then we'll run you through some examples

63 00:02:42.750 --> 00:02:45.030 of how researchers are using this platform,

64 00:02:45.030 --> 00:02:48.060 as well as how we've formed an industry partnership.

65 00:02:48.060 --> 00:02:50.190 And then lastly, we wanna show you kind of other ways

66 00:02:50.190 --> 00:02:52.800 that this tool can be used in academic settings.

67 00:02:52.800 --> 00:02:55.470 And then of course, we'll open it up to you guys

68 00:02:55.470 --> 00:02:57.180 and encourage you to ask us questions

69 00:02:57.180 --> 00:02:59.100 on any of these topics.

70 00:02:59.100 --> 00:03:02.010 So I'll hand over to Stephen now.

71 00:03:02.010 --> 00:03:03.900 <v ->Thanks Adria for that great introduction.</v>

72 00:03:03.900 --> 00:03:05.523 And hello to all of you.

73 00:03:06.530 --> 00:03:11.530 I currently see you as tiny, tiny pixels on my screen

74 00:03:12.150 --> 00:03:13.410 because of the way this is viewed.

75 00:03:13.410 --> 00:03:16.140 So as much as I'd love to be there in person

76 00:03:16.140 --> 00:03:17.460 and looking into the whites of your eyes,

77 00:03:17.460 --> 00:03:18.390 I'm not gonna get that chance.

78 00:03:18.390 --> 00:03:22.677 But, I think we have a really good robust discussion

79 00:03:22.677 --> 00:03:26.283 for you guys that I hope you'll find very interesting.

80 00:03:27.420 --> 00:03:30.390 And thank you very much again to Robert for the invitation.

81 00:03:30.390 --> 00:03:33.603 So similar backstory on myself,

82 00:03:34.830 --> 00:03:39.830 I went through undergraduate training at MIT

83 00:03:39.930 --> 00:03:43.050 in computer science, did a master's in AI

84 00:03:43.050 --> 00:03:44.850 before it was cool again,

85 00:03:44.850 --> 00:03:49.850 and then shipped off to UCSD for a PhD

86 00:03:50.760 --> 00:03:54.090 in neuroscience with a computational specialization.

87 00:03:54.090 --> 00:03:59.090 So very much familiar with the academic experience

88 00:03:59.420 --> 00:04:04.420 and I'm really excited to share with you

89 00:04:05.640 --> 00:04:08.746 some of the things that I've learned since leaving academia.

90 00:04:08.746 --> 00:04:10.020 And one of those things

91 00:04:10.020 --> 00:04:13.589 has been to start this company, MetaCell,

92 00:04:13.589 --> 00:04:16.110 which I basically started as I was wrapping up my PhD

93 00:04:16.110 --> 00:04:21.110 and I kind of realized that I wanted to serve science

94 00:04:21.510 --> 00:04:25.210 in a different way than was gonna be possible

95 00:04:26.710 --> 00:04:29.130 just within the confines of academia

96 00:04:29.130 --> 00:04:31.493 because I realized that I was a builder

97 00:04:31.493 --> 00:04:36.240 and to build software that could,

98 00:04:36.240 --> 00:04:40.170 software tools that could be useful to, you know,

99 00:04:40.170 --> 00:04:43.260 tools that I would wanted to have had as myself,

100 00:04:43.260 --> 00:04:44.190 a graduate student.

101 00:04:44.190 --> 00:04:47.520 I would need to kind of put a professional team of folks

102 00:04:47.520 --> 00:04:50.842 together that, you know, really came outta industry

103 00:04:50.842 --> 00:04:54.210 and that are kind of high hard to higher end academia.

104 00:04:54.210 --> 00:04:57.660 So the story of this slide is, since then,

105 00:04:57.660 --> 00:04:59.160 all the different great groups

106 00:04:59.160 --> 00:05:00.900 that we've had a chance to work with,

107 00:05:00.900 --> 00:05:04.934 and you'll see a really kind of motley crew of logos

108 00:05:04.934 --> 00:05:08.040 that are present here from, you know,

109 00:05:08.040 --> 00:05:09.880 really, really big pharma companies

110 00:05:12.180 --> 00:05:13.920 like Yale, you guys are on here,

111 00:05:13.920 --> 00:05:17.820 other universities that we've had the chance to work with,

112 00:05:17.820 --> 00:05:21.030 and then biotech companies,

113 00:05:21.030 --> 00:05:24.810 med device companies that we work with some,

114 00:05:24.810 --> 00:05:27.543 some of the US lots internationally.

115 00:05:29.310 --> 00:05:31.140 And realizing that, you know,

116 00:05:31.140 --> 00:05:33.690 the core thing that unifies all the work

117 00:05:33.690 --> 00:05:36.308 that we've been doing over time is the way

118 00:05:36.308 --> 00:05:39.840 that sort of math and computation can help us

119 00:05:39.840 --> 00:05:41.340 understand the life sciences.

120 00:05:41.340 --> 00:05:45.600 So hence I come to you today in a biostatistics seminar

121 00:05:45.600 --> 00:05:46.860 to talk about, you know,

122 00:05:46.860 --> 00:05:50.081 some of the other pieces of the puzzle

123 00:05:50.081 --> 00:05:55.081 that go into advancing the life sciences in that way.

124 00:05:55.993 --> 00:06:00.993 So, let's start with a really simple, simple example, right?

125 00:06:03.780 --> 00:06:07.860 So let's say you're doing some kind of analysis

126 00:06:07.860 --> 00:06:11.433 on some kind of bio data, okay?

127 00:06:13.020 --> 00:06:15.670 Perhaps in the statistics context, you're using SaaS.

128 00:06:16.980 --> 00:06:19.800 In a computational neuroscience context,

129 00:06:19.800 --> 00:06:24.213 you may be using Python and the Python suite of tools.

130 00:06:25.560 --> 00:06:28.860 Some in the statistics field are using R open source,

131 00:06:28.860 --> 00:06:30.330 you know, statistics packages.

132 00:06:30.330 --> 00:06:33.210 Whatever it is, you've got some data, you know,

133 00:06:33.210 --> 00:06:35.070 maybe you're analyzing it on behalf of yourself,

134 00:06:35.070 --> 00:06:36.816 maybe you're analyzing on behalf of your lab,

135 00:06:36.816 --> 00:06:38.340 the group that you're working with.

136 00:06:38.340 --> 00:06:40.710 Maybe you're analyzing it in terms of a company.

137 00:06:40.710 --> 00:06:41.760 Whatever it is,

138 00:06:41.760 --> 00:06:44.384 you wanna share that data analysis with somebody else.

139 00:06:44.384 --> 00:06:46.530 You're probably gonna have to gather

140 00:06:46.530 --> 00:06:49.710 some history of those commands together.

141 00:06:49.710 --> 00:06:52.650 Maybe it's packaged up as a script, maybe not.

142 00:06:52.650 --> 00:06:54.420 You're gonna send that file

143 00:06:54.420 --> 00:06:56.640 to somebody else very often.

144 00:06:56.640 --> 00:06:58.860 And then you're also gonna wanna somehow

145 00:06:58.860 --> 00:07:00.540 collect the outputs of that, right?

146 00:07:00.540 --> 00:07:04.984 The figures, the diagrams, the summary statistics,

147 00:07:04.984 --> 00:07:07.950 the result of T-tests, you know,

148 00:07:07.950 --> 00:07:09.210 things like this, right?

149 00:07:09.210 --> 00:07:12.240 And send that output somewhere, right?

150 00:07:12.240 --> 00:07:16.050 So, you know, that is a problem time immemorial.

151 00:07:16.050 --> 00:07:20.145 And you know, as long as I've been, you know,

152 00:07:20.145 --> 00:07:23.400 working in this space still, you know,

153 00:07:23.400 --> 00:07:25.140 it's very common to just do this

154 00:07:25.140 --> 00:07:28.920 and it's maybe send this over email, right?

155 00:07:28.920 --> 00:07:31.530 It's still a practice that I'm sure you know, happens.

156 00:07:31.530 --> 00:07:34.528 And so, and that's probably just fine, you know,

157 00:07:34.528 --> 00:07:37.380 in many small circumstances.

158 00:07:37.380 --> 00:07:41.403 But as that scales up, there's problems of reproducibility,

159 00:07:42.330 --> 00:07:44.400 there's problems of, you know,

160 00:07:44.400 --> 00:07:46.110 keeping track of who sent what.

161 00:07:46.110 --> 00:07:48.360 Email is not a great file management system.

162 00:07:48.360 --> 00:07:53.360 So we've been thinking a lot over the course of our company,

163 00:07:54.750 --> 00:07:56.137 which is, we've been around now,

164 00:07:56.137 --> 00:07:59.790 this is our 13th year about how, you know,

165 00:07:59.790 --> 00:08:02.130 the cloud and the internet basically can come into that

166 00:08:02.130 --> 00:08:04.980 in any better way than sending email along.

167 00:08:04.980 --> 00:08:08.010 And so we've thought a lot about, you know,

168 00:08:08.010 --> 00:08:10.980 what starts to happen when there's a computer that lives

169 00:08:10.980 --> 00:08:14.940 in the cloud that multiple people can jump into and join.

170 00:08:14.940 --> 00:08:17.640 And what is, you know, how does that work in general?

171 00:08:17.640 --> 00:08:22.470 It's something that we're not only just us doing, right?

172 00:08:22.470 --> 00:08:24.420 This is an idea that's been there for a while.

173 00:08:24.420 --> 00:08:27.210 Anybody familiar with like, say Python Notebooks, right,

174 00:08:27.210 --> 00:08:28.770 are aware of this idea.

175 00:08:28.770 --> 00:08:30.503 There's tools like Google Colab,

176 00:08:31.356 --> 00:08:34.200 and then we've even been talking to major universities,

177 00:08:34.200 --> 00:08:35.370 like we've been having a conversation

178 00:08:35.370 --> 00:08:36.930 with Harvard Medical School,

179 00:08:36.930 --> 00:08:39.211 where they've been working collaboration with Amazon

180 00:08:39.211 --> 00:08:42.845 to kind of work together with them to set up computers

181 00:08:42.845 --> 00:08:44.310 that are in the cloud.

182 00:08:44.310 --> 00:08:48.990 Similarly, of course, there's gonna be what happens with,

183 00:08:48.990 --> 00:08:50.310 at like, at your local university

184 00:08:50.310 --> 00:08:52.230 with your local computing infrastructure.

185 00:08:52.230 --> 00:08:55.740 Typically that's based around supercomputers that are there

186 00:08:55.740 --> 00:08:58.980 for doing like really powerful computations or calculations.

187 00:08:58.980 --> 00:09:00.600 Things that are very data intensive.

188 00:09:00.600 --> 00:09:02.370 A workspace in the cloud is sort of in between.

189 00:09:02.370 --> 00:09:05.194 So it's kind of like, you know,

190 00:09:05.194 --> 00:09:08.580 just a laptop that isn't your physical laptop,

191 00:09:08.580 --> 00:09:11.010 but it's like a laptop that's somewhere else in the cloud

192 00:09:11.010 --> 00:09:13.770 that you can log into and do some analysis with.

193 00:09:13.770 --> 00:09:16.380 And it basically lives as long as you wanna do that analysis

194 00:09:16.380 --> 00:09:17.700 and then it goes away

195 00:09:17.700 --> 00:09:19.770 if you don't need that analysis anymore

196 00:09:19.770 --> 00:09:22.200 or it can stay there as long as your lab is around, right?

197 00:09:22.200 --> 00:09:24.840 And then go away if you don't need it anymore.

198 00:09:24.840 --> 00:09:27.300 So the idea is then in this story,

199 00:09:27.300 --> 00:09:29.100 instead of just gathering the history of commands,

200 00:09:29.100 --> 00:09:31.228 sending the file and sending the output of the file,

201 00:09:31.228 --> 00:09:34.470 what if, right you could do all that in the context

202 00:09:34.470 --> 00:09:36.753 of a computer that multiple people

203 00:09:36.753 --> 00:09:38.940 can join and look at, right?

204 00:09:38.940 --> 00:09:40.380 Work in that same environment.

205 00:09:40.380 --> 00:09:41.213 When you log out,

206 00:09:41.213 --> 00:09:43.203 it's exactly where you left it, right?

207 00:09:43.203 --> 00:09:46.680 Like if you know your computer gets misplaced

208 00:09:46.680 --> 00:09:49.905 or you drop it, you know, off a bridge into a river,

209 00:09:49.905 --> 00:09:51.450 like, doesn't matter 'cause

210 00:09:51.450 --> 00:09:53.520 all this stuff is preserved, right?

211 00:09:53.520 --> 00:09:57.312 So, how does that idea start to change the basic practice

212 00:09:57.312 --> 00:10:01.560 of interacting with data and doing analysis like this

213 00:10:01.560 --> 00:10:05.400 if you were to change that one variable okay?

214 00:10:05.400 --> 00:10:08.820 So that's sort of the starting premise for our chat today.

215 00:10:08.820 --> 00:10:13.260 So, you know, what that might look like is, you know,

216 00:10:13.260 --> 00:10:15.996 a session one-on-one or two-on-one with multiple people

217 00:10:15.996 --> 00:10:20.996 where you get, you know, perhaps one of you in the future.

218 00:10:21.964 --> 00:10:24.176 In the case that we've been doing in our company,

219 00:10:24.176 --> 00:10:28.230 one of our staff members, who has experience

220 00:10:28.230 --> 00:10:31.810 in doing a different kind of data analysis.

221 00:10:31.810 --> 00:10:35.970 In our case, we work on a variety of problems,

222 00:10:35.970 --> 00:10:37.260 but one of the major ones we worked on

223 00:10:37.260 --> 00:10:39.437 is like the imaging of calcium signals

224 00:10:42.284 --> 00:10:44.573 in neural tissue okay?

225 00:10:44.573 --> 00:10:48.780 But you know, you might be on a call like this one and just

226 00:10:48.780 --> 00:10:50.430 the same way that you might meet with your lab members on a

227 00:10:50.430 --> 00:10:53.580 Zoom call, you might meet with someone

228 00:10:53.580 --> 00:10:55.590 with experience in data analysis or biostatistics

229 00:10:55.590 --> 00:11:00.590 that is not in your lab or not in your even organization.

230 00:11:00.960 --> 00:11:02.040 It might be somewhere remote,

231 00:11:02.040 --> 00:11:06.499 maybe at another university or in a company like ours.

232 00:11:06.499 --> 00:11:11.499 But what they might get as the experience of that is

233 00:11:13.026 --> 00:11:16.770 jointly logging into this workspace that lives in the cloud.

234 00:11:16.770 --> 00:11:19.890 And if SaaS is the thing you wanna use,

235 00:11:19.890 --> 00:11:22.320 you might find a whole SaaS instance there

236 00:11:22.320 --> 00:11:24.570 in a desktop that you can log into.

237 00:11:24.570 --> 00:11:27.330 But the point being that multiple people now can type on it

238 00:11:27.330 --> 00:11:30.060 as opposed to like physically handing your laptop around

239 00:11:30.060 --> 00:11:33.240 in the lab or even just screen sharing it

240 00:11:33.240 --> 00:11:35.310 in some kind of a lab meeting, right?

241 00:11:35.310 --> 00:11:37.590 It's actually allowing for people to jump into the same

242 00:11:37.590 --> 00:11:40.096 application and literally like trade off

243 00:11:40.096 --> 00:11:42.570 on like typing commands into it.

244 00:11:42.570 --> 00:11:45.720 Kind of like what you get with a Google Document

245 00:11:45.720 --> 00:11:47.580 or a Google Spreadsheet, right?

246 00:11:47.580 --> 00:11:48.990 That real-time collaboration,

247 00:11:48.990 --> 00:11:50.790 but now for any kind of application.

248 00:11:51.630 --> 00:11:54.390 So that's one experience you might have.

249 00:11:54.390 --> 00:11:55.530 Not just SaaS, right?

250 00:11:55.530 --> 00:11:57.600 So a Jupyter Notebook, as I mentioned before,

251 00:11:57.600 --> 00:11:58.890 is another thing that you can use.

252 00:11:58.890 --> 00:12:00.660 And those of you who might be using,

253 00:12:00.660 --> 00:12:02.979 again, the more open source technologies,

254 00:12:02.979 --> 00:12:05.490 if you might be using R Statistics or using Python

255 00:12:05.490 --> 00:12:08.070 or whatnot, you'd be familiar with, you know,

256 00:12:08.070 --> 00:12:10.729 a Jupyter Notebook.

257 00:12:10.729 --> 00:12:13.050 So it's based around, you know,

258 00:12:13.050 --> 00:12:16.350 this idea of putting a computer in the cloud,

259 00:12:16.350 --> 00:12:18.300 multiple folks logging into it,

260 00:12:18.300 --> 00:12:21.390 and then being able to sort of transport

261 00:12:21.390 --> 00:12:24.600 your expertise around the world.

262 00:12:24.600 --> 00:12:29.600 Because in addition to the knowledge of doing analysis

263 00:12:30.540 --> 00:12:31.593 being shipped around,

264 00:12:32.460 --> 00:12:34.470 data can also come into this workspace

265 00:12:34.470 --> 00:12:37.923 as an intermediate space that's private to a given lab,

266 00:12:39.210 --> 00:12:43.041 but allows for a different kind of model on sharing data

267 00:12:43.041 --> 00:12:46.740 where it sort of stays under the control of the lab,

268 00:12:46.740 --> 00:12:48.745 you know, whoever puts it there can take it back,

269 00:12:48.745 --> 00:12:50.520 that kind of thing.

270 00:12:50.520 --> 00:12:53.730 Okay so we've been exploring this model

271 00:12:53.730 --> 00:12:57.300 and we've also been talking to other organizations

272 00:12:57.300 --> 00:13:00.030 and universities about this model and how to use it,

273 00:13:00.030 --> 00:13:01.680 how to implement it, right?

274 00:13:01.680 --> 00:13:05.081 As I mentioned, we've been talking to folks like

275 00:13:05.081 --> 00:13:07.740 at Harvard Medical School that partner with Amazon

276 00:13:07.740 --> 00:13:10.890 to bring these sorts of instances into their

277 00:13:10.890 --> 00:13:12.570 labs and what can be done with it.

278 00:13:12.570 --> 00:13:14.010 So I'm gonna wanna talk a little bit

279 00:13:14.010 --> 00:13:16.410 about like some of those details,

280 00:13:16.410 --> 00:13:19.410 and I'm saying it here in the context of our product,

281 00:13:19.410 --> 00:13:20.400 but I'm not trying to sell you anything.

282 00:13:20.400 --> 00:13:21.300 I'm really trying to talk about it

283 00:13:21.300 --> 00:13:23.880 more in the context of what can be done.

284 00:13:23.880 --> 00:13:26.763 So thinking about it, like,

285 00:13:27.900 --> 00:13:29.340 so I mentioned SaaS as an example.

286 00:13:29.340 --> 00:13:31.170 I mentioned Jupyter Notebooks as an example,

287 00:13:31.170 --> 00:13:33.525 but there might be other kinds of software

288 00:13:33.525 --> 00:13:35.583 that are more particular to a use case,

289 00:13:35.583 --> 00:13:38.010 like MATLAB's another one that could be installed.

290 00:13:38.010 --> 00:13:39.898 But there might be even more specific software

291 00:13:39.898 --> 00:13:43.680 that might need to be set up or run.

292 00:13:43.680 --> 00:13:46.599 Sometimes, for example, survey software

293 00:13:46.599 --> 00:13:50.650 where you might collect data from a very particular kind of

294 00:13:51.540 --> 00:13:53.820 survey system and you need something to work with it.

295 00:13:53.820 --> 00:13:55.260 So imagine that,

296 00:13:55.260 --> 00:13:57.561 like for the use case that you might have, right,

297 00:13:57.561 --> 00:14:01.500 you could have a workspace that is set up

298 00:14:01.500 --> 00:14:03.330 so that all that software comes pre-built

299 00:14:03.330 --> 00:14:04.280 once you set it up.

300 00:14:05.208 --> 00:14:07.410 Much like, you know, having laptops

301 00:14:07.410 --> 00:14:10.200 that have come pre-configured with a certain set of tools,

302 00:14:10.200 --> 00:14:11.790 but instead of handing out physical laptops,

303 00:14:11.790 --> 00:14:13.980 it's on the cloud.

304 00:14:13.980 --> 00:14:14.850 The virtual collaboration,

305 00:14:14.850 --> 00:14:18.180 I think I've gone through a lot, the multiple workspace,

306 00:14:18.180 --> 00:14:20.340 I think I mentioned also.

307 00:14:20.340 --> 00:14:23.220 Data security I kinda mentioned, you know,

308 00:14:23.220 --> 00:14:25.530 anybody who's doing data analysis

309 00:14:25.530 --> 00:14:28.710 with anybody who has, you know,

310 00:14:28.710 --> 00:14:30.120 talking to somebody that they weren't the ones

311 00:14:30.120 --> 00:14:32.370 to collect it, I'm sure has run into challenges

312 00:14:32.370 --> 00:14:36.690 where folks are reticent to, you know, share data.

313 00:14:36.690 --> 00:14:38.430 So that's why in this context,

314 00:14:38.430 --> 00:14:40.830 it's really important to note that like, you know,

315 00:14:40.830 --> 00:14:42.360 we can lock that environment down

316 00:14:42.360 --> 00:14:44.310 and make sure that only the people that can log into it

317 00:14:44.310 --> 00:14:47.400 have access to it, that's a really important point.

318 00:14:47.400 --> 00:14:49.140 So it's not really like the data

319 00:14:49.140 --> 00:14:50.520 are going out of somebody's control.

320 00:14:50.520 --> 00:14:51.540 Again, they're kept in a place

321 00:14:51.540 --> 00:14:53.490 where anybody who wants to can remove

322 00:14:53.490 --> 00:14:55.563 that data again and delete it.

323 00:14:56.580 --> 00:15:00.664 And then if there were to be very computationally aggressive

324 00:15:00.664 --> 00:15:04.353 things to do, it's very easy to scale it up.

325 00:15:05.360 --> 00:15:09.510 And that's something that folks also like.

326 00:15:09.510 --> 00:15:13.710 So how, you know, how are ways that this kind of workspace

327 00:15:13.710 --> 00:15:16.680 can support biostatistics research

328 00:15:16.680 --> 00:15:18.270 and data analysis in general.

329 00:15:18.270 --> 00:15:20.280 So I mentioned data science as a service

330 00:15:20.280 --> 00:15:21.990 a little bit in this example.

331 00:15:21.990 --> 00:15:25.547 So this would be the case where any organization

332 00:15:25.547 --> 00:15:28.880 who say doesn't have biostatistics

333 00:15:28.880 --> 00:15:32.082 or data science expertise local to them

334 00:15:32.082 --> 00:15:36.090 might be interested in sort of renting time

335 00:15:36.090 --> 00:15:40.020 or having some part-time person come in to help with that.

336 00:15:40.020 --> 00:15:42.401 And that's a model that we've seen work well

337 00:15:42.401 --> 00:15:44.250 both for labs and for companies.

338 00:15:44.250 --> 00:15:48.510 One way in which labs really like it is new PIs

339 00:15:48.510 --> 00:15:51.150 with a startup package that just, you know,

340 00:15:51.150 --> 00:15:53.970 first few weeks into their appointment

341 00:15:53.970 --> 00:15:56.760 with an R one, right, no staff yet.

342 00:15:56.760 --> 00:16:01.323 Nobody, but they're coming in with data from their previous,

343 00:16:03.182 --> 00:16:05.744 you know, from their postdoc basically.

344 00:16:05.744 --> 00:16:07.020 And what do they do, right?

345 00:16:07.020 --> 00:16:10.350 They need to write grants, they need to like hire staff,

346 00:16:10.350 --> 00:16:11.610 they need to do all these things.

347 00:16:11.610 --> 00:16:15.330 So we've actually found labs are very happy

348 00:16:15.330 --> 00:16:18.660 in that circumstance just to get going, you know,

349 00:16:18.660 --> 00:16:20.100 to be like, "Hey, I have this data,

350 00:16:20.100 --> 00:16:21.300 I haven't analyzed it yet.

351 00:16:21.300 --> 00:16:22.707 I really wanna put in my grant proposals.

352 00:16:22.707 --> 00:16:26.662 I just need somebody to kind of sit with me virtually

353 00:16:26.662 --> 00:16:30.449 and run through this data,

354 00:16:30.449 --> 00:16:32.850 so that I can get these figures

355 00:16:32.850 --> 00:16:34.260 made and get my grant out, right?"

356 00:16:34.260 --> 00:16:35.880 And I just don't have time

357 00:16:35.880 --> 00:16:37.200 to bring on a full person to do that.

358 00:16:37.200 --> 00:16:40.290 So data sciences service can be very useful for that.

359 00:16:40.290 --> 00:16:42.240 Data standardization and sharing as a service.

360 00:16:42.240 --> 00:16:45.750 So, you know, I'm not sure how much it's affecting folks

361 00:16:45.750 --> 00:16:48.287 in the room, but the NIIH over time

362 00:16:48.287 --> 00:16:53.287 has gotten increasingly serious about making data sharing

363 00:16:54.720 --> 00:16:56.175 happen for real for real,

364 00:16:56.175 --> 00:16:58.260 and not for fake for real, right?

365 00:16:58.260 --> 00:17:00.570 And so this year in particular,

366 00:17:00.570 --> 00:17:04.680 a new policy from NIIH has come out, DMS policy,

367 00:17:04.680 --> 00:17:08.610 where they're really, really asking for even, you know,

368 00:17:08.610 --> 00:17:10.860 grant proposals to have a whole data management

369 00:17:10.860 --> 00:17:14.880 strategy figured out upon submission.

370 00:17:14.880 --> 00:17:19.320 And even, you know, saying you need to set aside

371 00:17:19.320 --> 00:17:20.153 some budget for that

372 00:17:20.153 --> 00:17:22.440 'cause it turns out data sharing doesn't happen for free,

373 00:17:22.440 --> 00:17:24.060 doesn't happen for free, you know,

374 00:17:24.060 --> 00:17:25.817 for PIs for their time, right?

375 00:17:25.817 --> 00:17:29.190 So that's also something where, okay,

376 00:17:29.190 --> 00:17:30.420 I don't have the expertise to figure out

377 00:17:30.420 --> 00:17:34.110 which of the billion databases I might share my data in.

378 00:17:34.110 --> 00:17:35.880 Could somebody come in and help do that?

379 00:17:35.880 --> 00:17:36.990 Well how do you do that?

380 00:17:36.990 --> 00:17:41.220 You know, when I did work in the neuroinformatics

381 00:17:41.220 --> 00:17:43.170 space as a graduate student

382 00:17:43.170 --> 00:17:46.530 and I was trying to help figure out for neuroscientists

383 00:17:46.530 --> 00:17:50.100 how to get data that they had, you know, collected

384 00:17:50.100 --> 00:17:54.600 in a very laborious process of experimental collection,

385 00:17:54.600 --> 00:17:56.880 was trying to help them share their data

386 00:17:56.880 --> 00:17:59.489 'cause they wanted to comply with these policies

387 00:17:59.489 --> 00:18:04.290 even back then, you know, very frequently I would

388 00:18:04.290 --> 00:18:05.407 get the challenge of like,

389 00:18:05.407 --> 00:18:08.040 "Yeah, it's in a hard drive under my desk, right?

390 00:18:08.040 --> 00:18:10.080 Physical hard drive sitting under my desk, right?"

391 00:18:10.080 --> 00:18:14.310 Like, okay, so you can go pick it up and like take it away

392 00:18:14.310 --> 00:18:15.330 and do something with it.

393 00:18:15.330 --> 00:18:18.960 But you know, they don’t have the expertise, you know,

394 00:18:18.960 --> 00:18:22.200 locally to even know, okay, now we’re gonna plug it in

395 00:18:22.200 --> 00:18:23.250 and we gotta look through it

396 00:18:23.250 --> 00:18:26.520 and like, oh, the PhD student is left three years ago.

397 00:18:26.520 --> 00:18:27.353 And like, how do I do that?

398 00:18:27.353 --> 00:18:31.200 So the idea of, okay, if all we can do is like take that

399 00:18:31.200 --> 00:18:32.580 hard drive from under the desk

400 00:18:32.580 --> 00:18:37.320 and like plug it in the cloud, share it on Dropbox,

401 00:18:37.320 --> 00:18:38.643 okay, something like this or you know,

402 00:18:38.643 --> 00:18:40.813 have a conduit to get it to the cloud,

403 00:18:40.813 --> 00:18:43.385 share that folder in a workspace online

404 00:18:43.385 --> 00:18:47.220 and then have somebody else that does this all the time

405 00:18:47.220 --> 00:18:49.320 like go through all that and do their best to start,

406 00:18:49.320 --> 00:18:50.940 you know, documenting what they find,

407 00:18:50.940 --> 00:18:54.120 maybe raising questions that they might find, you know,

408 00:18:54.120 --> 00:18:55.357 to present to the PI,

409 00:18:55.357 --> 00:18:57.900 ”Hey, I know your PhD student left three years ago,

410 00:18:57.900 --> 00:18:59.160 but you know, can you tell me a little bit

411 00:18:59.160 --> 00:19:00.840 about this experimental methodology?”

412 00:19:00.840 --> 00:19:03.390 There’s now at least a hope that you can start,

413 00:19:03.390 --> 00:19:04.650 you know, standardizing that data,

414 00:19:04.650 --> 00:19:05.971 sharing it in a better way,

415 00:19:05.971 --> 00:19:09.097 making the NIIH not come kick down your door

416 00:19:09.097 --> 00:19:11.040 with the data sharing police force

417 00:19:11.040 --> 00:19:13.680 that I'm sure they're setting up now.

418 00:19:13.680 --> 00:19:14.580 Okay probably not.

419 00:19:15.519 --> 00:19:20.519 Okay a third way is through workshops.

420 00:19:20.700 --> 00:19:22.920 And I'll have some specific examples

421 00:19:22.920 --> 00:19:24.780 a little bit later about this one.

422 00:19:24.780 --> 00:19:26.520 But if you think about, you know,

423 00:19:26.520 --> 00:19:29.670 the experience of either physically traveling

424 00:19:29.670 --> 00:19:31.440 or doing what we're doing here

425 00:19:31.440 --> 00:19:35.760 and then being exposed to software, right?

426 00:19:35.760 --> 00:19:37.230 It's one thing to have slides show

427 00:19:37.230 --> 00:19:39.060 you pretty pictures of what software looks like.

428 00:19:39.060 --> 00:19:42.787 And it's another thing to say basically like,

429 00:19:42.787 --> 00:19:47.120 "Hey, log into, like go right now on your laptops

430 00:19:47.120 --> 00:19:49.740 and go hit this address"

431 00:19:49.740 --> 00:19:52.860 and like, here's your login and like while I'm explaining it

432 00:19:52.860 --> 00:19:56.569 to you, check it out, play with it, right?

433 00:19:56.569 --> 00:20:00.450 So we've actually found that also to be a really valuable

434 00:20:00.450 --> 00:20:05.450 way to do an extra level of education and demonstration,

435 00:20:05.460 --> 00:20:08.790 especially for tools built in academia,

436 00:20:08.790 --> 00:20:10.920 which generally have a pretty small audience, right?

437 00:20:10.920 --> 00:20:14.010 Not a lot of people use them maybe necessarily,

438 00:20:14.010 --> 00:20:15.780 or it's like a very niche community.

439 00:20:15.780 --> 00:20:17.700 So the total number of humans is not great.

440 00:20:17.700 --> 00:20:21.000 So to have the ability right now in a live session

441 00:20:21.000 --> 00:20:23.820 to be like, let me show you this software you log in right

442 00:20:23.820 --> 00:20:27.030 now, play with it can move the needle a lot on getting folks

443 00:20:27.030 --> 00:20:30.600 to use stuff that that there will really be tools

444 00:20:30.600 --> 00:20:32.910 that they will actually help them a lot.

445 00:20:32.910 --> 00:20:34.500 And then lastly, you know,

446 00:20:34.500 --> 00:20:37.710 collaborations between labs, right?

447 00:20:37.710 --> 00:20:39.690 Hey, we just set up a consortia,

448 00:20:39.690 --> 00:20:41.160 it's a five lab consortia

449 00:20:41.160 --> 00:20:43.770 and we're all studying this thing, right?

450 00:20:43.770 --> 00:20:46.080 It's a collaboration between the folks that are generating

451 00:20:46.080 --> 00:20:47.580 the data and the folks are gonna analyze the data.

452 00:20:47.580 --> 00:20:49.973 Okay, great, we got this really smart set of mathematicians

453 00:20:49.973 --> 00:20:53.004 who are gonna do all these great statistics, awesome.

454 00:20:53.004 --> 00:20:55.170 How do you get the data from point A to point B?

455 00:20:55.170 --> 00:20:57.826 Well email, right?

456 00:20:57.826 --> 00:21:00.720 So what if you can improve that, right?

457 00:21:00.720 --> 00:21:04.200 Or you know, the context of, you know,

458 00:21:04.200 --> 00:21:06.360 we also find companies wanna collaborate with each other's

459 00:21:06.360 --> 00:21:08.160 and then universities and companies wanna collaborate

460 00:21:08.160 --> 00:21:09.630 with each other also, right?

461 00:21:09.630 --> 00:21:13.089 So in ways that I haven't already listed,

462 00:21:13.089 --> 00:21:16.860 but just collaborations of whatever variety.

463 00:21:16.860 --> 00:21:19.467 So when it comes down to those things, right,

464 00:21:19.467 --> 00:21:22.350 it's one step better than just sharing on Dropbox

465 00:21:22.350 --> 00:21:24.240 and being like, here are the data, go check it out

466 00:21:24.240 --> 00:21:28.233 'cause you're keeping the analysis all together, right?

467 00:21:29.220 --> 00:21:31.170 It adds a layer of reproducibility

468 00:21:31.170 --> 00:21:32.430 to those kinds of collaborations,

469 00:21:32.430 --> 00:21:36.330 which are hard to match in addition to all the other things,

470 00:21:36.330 --> 00:21:39.303 all the great best practices for reproducibility.

471 00:21:40.140 --> 00:21:42.840 Okay so that's four ways to use cloud workspaces

472 00:21:42.840 --> 00:21:44.390 support biostatistics research.

473 00:21:46.890 --> 00:21:50.640 So let's, you know, I think I've kind of walked through this

474 00:21:50.640 --> 00:21:51.960 example already verbally,

475 00:21:51.960 --> 00:21:53.884 but I did have a slide specifically for it.

476 00:21:53.884 --> 00:21:57.090 So like this happens in research all the time.

477 00:21:57.090 --> 00:21:59.550 There's a lab that needs a particular analysis completed

478 00:21:59.550 --> 00:22:01.290 and they don't have the expertise in lab.

479 00:22:01.290 --> 00:22:02.130 What can be done?

480 00:22:02.130 --> 00:22:04.290 So typically the alternatives are, you know,

481 00:22:04.290 --> 00:22:06.780 bring in some student or a postdoc or collaborate

482 00:22:06.780 --> 00:22:09.480 with a lab that has some mathematical expertise

483 00:22:09.480 --> 00:22:10.800 to perform analysis.

484 00:22:10.800 --> 00:22:13.410 But that can be quite time consuming, you know,

485 00:22:13.410 --> 00:22:16.350 that might not deliver the results you're looking for.

486 00:22:16.350 --> 00:22:20.040 Secondly, right for folks who might, you know,

487 00:22:20.040 --> 00:22:21.450 be in a position, like I mentioned

488 00:22:21.450 --> 00:22:25.200 with early lab set up, right?

489 00:22:25.200 --> 00:22:27.265 Engaging some part-time data scientists from industry

490 00:22:27.265 --> 00:22:30.840 could help work on particular problems as needed.

491 00:22:30.840 --> 00:22:32.892 And that’s interesting both perhaps

492 00:22:32.892 --> 00:22:34.920 from the perspective of me as a company,

493 00:22:34.920 --> 00:22:38.370 but also maybe interesting for yourselves

494 00:22:38.370 --> 00:22:41.001 thinking about a path through industry

495 00:22:41.001 --> 00:22:44.580 where you might be able to do biostatistics

496 00:22:44.580 --> 00:22:48.813 for multiple organizations at once, not just one at a time.

497 00:22:49.890 --> 00:22:50.850 And then it’s also interesting,

498 00:22:50.850 --> 00:22:52.980 as I mentioned from the perspective folks

499 00:22:52.980 --> 00:22:55.613 that have the problem that need to get the analysis done.

500 00:22:57.300 --> 00:23:02.300 Okay so some case studies, does this happen?

501 00:23:02.520 --> 00:23:05.340 I sort of mentioned abstractly, it does,

502 00:23:05.340 --> 00:23:10.340 but these are five cases that we’ve worked on in our company

503 00:23:10.410 --> 00:23:14.070 and they are, many of them have a,

504 00:23:14.070 --> 00:23:15.150 well they all have the theme

505 00:23:15.150 --> 00:23:17.670 of being calcium imaging data, okay?

506 00:23:17.670 --> 00:23:20.370 So here, you know, swap out biostatistics

507 00:23:20.370 --> 00:23:23.102 for looking at data that comes from a microscope.

508 00:23:23.102 --> 00:23:25.050 But at the end of the day,

509 00:23:25.050 --> 00:23:29.283 that data from a microscope is basically a video stream,

510 00:23:31.470 --> 00:23:33.360 generally black and white images

511 00:23:33.360 --> 00:23:35.790 that then have to be post-processed.

512 00:23:35.790 --> 00:23:38.926 And from that video stream there’s a spatial component

513 00:23:38.926 --> 00:23:42.970 of looking at a field of neurons under a microscope

514 00:23:44.250 --> 00:23:45.393 and a time component.

515 00:23:46.303 --> 00:23:48.960 Like how did those, you know,

516 00:23:48.960 --> 00:23:51.360 neurons activity change over time.

517 00:23:51.360 --> 00:23:54.000 But there's a lot of like statistical challenges

518 00:23:54.000 --> 00:23:55.016 that have to go into that.

519 00:23:55.016 --> 00:23:58.020 You need to separate the neurons out from each other, okay?

520 00:23:58.020 --> 00:24:00.090 They kind of overlapped on each other.

521 00:24:00.090 --> 00:24:04.080 So looking at a video stream, you're not always sure, right?

522 00:24:04.080 --> 00:24:06.240 If I'm looking at one neuron or two neurons.

523 00:24:06.240 --> 00:24:07.710 So you have to do some spatial analysis

524 00:24:07.710 --> 00:24:09.450 to separate those out.

525 00:24:09.450 --> 00:24:12.930 And then you wanna do some sort of peak finding over time.

526 00:24:12.930 --> 00:24:15.240 What you kind of wanna extract out is a time series

527 00:24:15.240 --> 00:24:16.860 of however many neurons you've detected

528 00:24:16.860 --> 00:24:18.870 in your field of view

529 00:24:18.870 --> 00:24:21.450 and then start to do some additional analysis.

530 00:24:21.450 --> 00:24:24.240 And that additional analysis will be based on

531 00:24:24.240 --> 00:24:26.040 the specifics of the experimental setup

532 00:24:26.040 --> 00:24:29.253 and like, you know, what part of brain were you looking at?

533 00:24:30.390 --> 00:24:33.189 What was your protocol that you applied

534 00:24:33.189 --> 00:24:36.690 and what kind of expectations

535 00:24:36.690 --> 00:24:40.113 do you have about the time series that you extracted?

536 00:24:41.220 --> 00:24:45.150 So these organizations that we work with, I guess, you know,

537 00:24:45.150 --> 00:24:47.580 four out of five are universities.

538 00:24:47.580 --> 00:24:50.820 So DGIST is Institute of Science and Technology

539 00:24:50.820 --> 00:24:55.820 in South Korea, McGill University in Canada,

540 00:24:58.140 --> 00:25:03.140 University of Penn, UPenn and University of Alabama.

541 00:25:03.589 --> 00:25:08.589 And then Maze, which is a small pharma company

542 00:25:09.360 --> 00:25:13.590 in San Francisco and they're all doing calcium imaging work.

543 00:25:13.590 --> 00:25:17.850 And I think we served all of these organizations

544 00:25:17.850 --> 00:25:20.643 within the same span of about six months.

545 00:25:21.568 --> 00:25:26.501 Each one of them had brought different data to the table.

546 00:25:26.501 --> 00:25:29.160 They're all generally in this form of video data

547 00:25:29.160 --> 00:25:32.223 with the calcium imaging to extract.

548 00:25:33.060 --> 00:25:33.930 All five of them were served

549 00:25:33.930 --> 00:25:37.530 by the same data scientist on our side,

550 00:25:37.530 --> 00:25:39.480 gentleman whose picture you saw earlier

551 00:25:41.100 --> 00:25:43.620 but they had very different scientific protocols, right?

552 00:25:43.620 --> 00:25:47.010 So it wasn't necessary that one person full-time

553 00:25:47.010 --> 00:25:49.560 over six months worked on each of these projects, right?

554 00:25:49.560 --> 00:25:51.786 Instead we have one individual,

555 00:25:51.786 --> 00:25:54.300 who's able to jump from project to project

556 00:25:54.300 --> 00:25:59.300 and check back in with multiple PIs/business leaders,

557 00:26:01.410 --> 00:26:04.980 managers to check in on the results of that, right?

558 00:26:04.980 --> 00:26:07.593 And that person never left their home, right?

559 00:26:08.490 --> 00:26:13.170 So our company is also fully remote, which is nice.

560 00:26:13.170 --> 00:26:16.620 And so I think that's a really powerful demonstration

561 00:26:16.620 --> 00:26:19.363 of what's possible for this kind of analysis,

562 00:26:19.363 --> 00:26:24.363 whereby, you know, essentially organizations

563 00:26:25.260 --> 00:26:26.970 in multiple different countries

564 00:26:26.970 --> 00:26:29.370 and different continent in one case, right,

565 00:26:29.370 --> 00:26:32.539 can all be served by the same person doing roughly

566 00:26:32.539 --> 00:26:36.021 having roughly the same skillset of data analysis

567 00:26:36.021 --> 00:26:40.230 but working on data that addresses very different scientific

568 00:26:40.230 --> 00:26:43.140 questions all at the same time.

569 00:26:43.140 --> 00:26:46.590 Okay, so that's a thing.

570 00:26:46.590 --> 00:26:49.080 And, in each one of these, I should say

571 00:26:49.080 --> 00:26:51.450 been done in this collaboration model that I mentioned

572 00:26:51.450 --> 00:26:56.450 where there's one workspace per organization, right?

573 00:26:56.670 --> 00:26:59.340 So each organization has their own workspace,

574 00:26:59.340 --> 00:27:01.380 they log into it, they can see the results

575 00:27:01.380 --> 00:27:04.200 of the data science work that happens.

576 00:27:04.200 --> 00:27:06.240 They have all in one way or the other,

577 00:27:06.240 --> 00:27:09.325 put data into the workspace, right?

578 00:27:09.325 --> 00:27:12.840 And, they've all sort of been able to pull figures back out

579 00:27:12.840 --> 00:27:17.840 again and direct the flow of analysis in the direction

580 00:27:18.810 --> 00:27:21.630 that they wanted through Zoom calls,

581 00:27:21.630 --> 00:27:23.340 like the one that I mentioned

582 00:27:23.340 --> 00:27:24.810 generally on like a weekly basis

583 00:27:24.810 --> 00:27:26.823 or every couple weeks check in.

584 00:27:28.159 --> 00:27:33.159 So yeah, a little bit more about the team behind that

585 00:27:33.930 --> 00:27:35.490 in terms of thinking about like what it takes

586 00:27:35.490 --> 00:27:36.490 to make that happen.

587 00:27:37.410 --> 00:27:39.360 While there is a little bit of like finding those labs

588 00:27:39.360 --> 00:27:41.423 and figuring out that they have that problem,

589 00:27:42.390 --> 00:27:44.790 which are not taken care of

590 00:27:44.790 --> 00:27:46.200 by the individuals on this screen.

591 00:27:46.200 --> 00:27:50.040 But I mentioned, I mentioned Phil, the PhD;

592 00:27:50.040 --> 00:27:52.140 another PhD, who's worked with us

593 00:27:52.140 --> 00:27:54.570 as data scientist is Marcus.

594 00:27:54.570 --> 00:27:56.790 And then kind of orchestrating behind the scenes,

595 00:27:56.790 --> 00:27:59.306 the standing up of these workspaces

596 00:27:59.306 --> 00:28:02.133 is a software architect, Zoran.

597 00:28:04.380 --> 00:28:06.960 Phil in the New York area, New York City area.

598 00:28:06.960 --> 00:28:11.960 Marcus is in China and Zoran is in the Netherlands.

599 00:28:12.720 --> 00:28:16.230 So again, interesting to think about the different

600 00:28:16.230 --> 00:28:19.170 geographies where folks come from being able to serve people

601 00:28:19.170 --> 00:28:20.730 in different geographies,

602 00:28:20.730 --> 00:28:23.010 but all of them when it comes to a project,

603 00:28:23.010 --> 00:28:27.180 like the center organizing node is a workspace.

604 00:28:27.180 --> 00:28:28.410 That is the thing that helps

605 00:28:28.410 --> 00:28:30.851 coordinate a lot of this together.

606 00:28:30.851 --> 00:28:33.101 There are a few other technologies that help.

607 00:28:34.190 --> 00:28:36.570 Those of you familiar with like a Kanban board

608 00:28:36.570 --> 00:28:39.247 or just really any kind of task driven software,

609 00:28:39.247 --> 00:28:41.580 you know, you can bring that to bear as well.

610 00:28:41.580 --> 00:28:44.070 So one of the ways you can organize work a little bit better

611 00:28:44.070 --> 00:28:46.440 than just sending emails back and forth

612 00:28:46.440 --> 00:28:49.830 is to encapsulate each task,

613 00:28:49.830 --> 00:28:52.974 break each task down into a card on a Kanban board.

614 00:28:52.974 --> 00:28:55.693 We like the tool called Trello,

615 00:28:55.693 --> 00:28:57.690 but there's lots of them out there

616 00:28:57.690 --> 00:28:59.550 that can be used for such things.

617 00:28:59.550 --> 00:29:01.800 And then, you know, one card per task

618 00:29:01.800 --> 00:29:03.870 is a nice way to organize things.

619 00:29:03.870 --> 00:29:06.810 And then using a practice from software engineering,

620 00:29:06.810 --> 00:29:09.480 you can actually sort of estimate

621 00:29:09.480 --> 00:29:11.807 in roughly how many hours, you know,

622 00:29:11.807 --> 00:29:14.610 the data scientists might think it would take

623 00:29:14.610 --> 00:29:16.380 to do a given task

624 00:29:16.380 --> 00:29:18.240 and then use that as a way to figure out

625 00:29:18.240 --> 00:29:19.560 like how long it's gonna take

626 00:29:19.560 --> 00:29:21.480 to do a certain kind of analysis.

627 00:29:21.480 --> 00:29:23.130 This is a practice that we actually use

628 00:29:23.130 --> 00:29:24.900 across my company for all sorts of tasks,

629 00:29:24.900 --> 00:29:26.250 not just data science,

630 00:29:26.250 --> 00:29:28.410 really organizing kind of everything that we do

631 00:29:28.410 --> 00:29:30.900 on the basis of making cards like this

632 00:29:30.900 --> 00:29:31.770 and moving things across.

633 00:29:31.770 --> 00:29:32.670 And I'm still surprised

634 00:29:32.670 --> 00:29:35.043 how many organizations don't use this.

635 00:29:36.001 --> 00:29:37.590 I have lots of friends in academia

636 00:29:37.590 --> 00:29:38.640 that do this just for their labs.

637 00:29:38.640 --> 00:29:39.919 You guys might do this in your labs, I don't know.

638 00:29:39.919 --> 00:29:43.560 But for organizing oneself,

639 00:29:43.560 --> 00:29:45.690 even if you do meet in person,

640 00:29:45.690 --> 00:29:47.875 having this sort of set up in the cloud

641 00:29:47.875 --> 00:29:50.943 can be very helpful for organizing work.

642 00:29:51.840 --> 00:29:53.610 Not sure how new or not new this is

643 00:29:53.610 --> 00:29:57.300 to those of you in the room, but something we use.

644 00:29:57.300 --> 00:29:58.440 And then of course there’s Slack,

645 00:29:58.440 --> 00:30:01.743 which I think has pretty good adoption amongst academia.

646 00:30:03.360 --> 00:30:06.219 We do find almost every lab that we talk to

647 00:30:06.219 --> 00:30:08.883 pretty much is on Slack or some version of it.

648 00:30:09.780 --> 00:30:12.210 Companies are using Microsoft Teams,

649 00:30:12.210 --> 00:30:13.470 which I personally like less,

650 00:30:13.470 --> 00:30:16.620 but you know, but we use that too.

651 00:30:16.620 --> 00:30:18.123 But basically, you know,

652 00:30:20.430 --> 00:30:23.490 one thing that we do that maybe others don’t do

653 00:30:23.490 --> 00:30:25.800 is to connect a Kanban board like

654 00:30:25.800 --> 00:30:28.410 the one that you saw to spit out notifications

655 00:30:28.410 --> 00:30:31.020 in a Slack channel at the same time,

656 00:30:31.020 --> 00:30:33.850 which can be really nice if you are a Slack based person

657 00:30:34.740 --> 00:30:37.260 to just like be able to see how tasks are changing

658 00:30:37.260 --> 00:30:39.600 and evolving in the feed,

659 00:30:39.600 --> 00:30:41.880 which then doesn’t require an extra conversation, right?

660 00:30:41.880 --> 00:30:45.210 Like ”Hey, so we agreed on Monday that you were gonna,

661 00:30:45.210 --> 00:30:50.210 you know, do that t-test on this survey data,

662 00:30:50.430 --> 00:30:52.410 how’s that going right?”

663 00:30:52.410 --> 00:30:54.960 Well if they’ve moved that card,

664 00:30:54.960 --> 00:30:58.110 which was like T-test on survey data from the to-do column

665 00:30:58.110 --> 00:30:59.280 to the doing column,

666 00:30:59.280 --> 00:31:01.560 a little notification's gonna pop up in Slack.

667 00:31:01.560 --> 00:31:03.930 And then when they write a comment like, "Yep, you know,

668 00:31:03.930 --> 00:31:06.750 I ran the test and wasn't statistically significant,"

669 00:31:06.750 --> 00:31:09.210 then that's gonna pop up also.

670 00:31:09.210 --> 00:31:11.460 That comment will then be relayed into Slack.

671 00:31:11.460 --> 00:31:12.600 So then when you go back to check in,

672 00:31:12.600 --> 00:31:13.470 you don't have to ask that question.

673 00:31:13.470 --> 00:31:15.360 It's like, "Yep, I saw that it happened

674 00:31:15.360 --> 00:31:18.090 and by the way I saw that it happened on Tuesday,

675 00:31:18.090 --> 00:31:20.060 you know, now it's Wednesday, you know.

676 00:31:20.060 --> 00:31:22.530 I forgot to check back in with you about it."

677 00:31:22.530 --> 00:31:24.933 So like that idea of asynchronous work can happen

678 00:31:24.933 --> 00:31:28.728 in this cloud-based context also, which again,

679 00:31:28.728 --> 00:31:31.380 like we use also in all other parts

680 00:31:31.380 --> 00:31:33.327 of our company can be really helpful

681 00:31:33.327 --> 00:31:35.943 for moving projects along in lots of ways.

682 00:31:37.410 --> 00:31:41.850 So yeah I've told you a lot

683 00:31:41.850 --> 00:31:44.370 about a particular example then of doing work.

684 00:31:44.370 --> 00:31:46.667 I wanna call Adria back in here

685 00:31:46.667 --> 00:31:51.667 to extend a little bit more in a partnership example

686 00:31:51.870 --> 00:31:53.490 that we've had some experience with.

687 00:31:53.490 --> 00:31:55.163 So back to you Adria.

688 00:31:55.163 --> 00:31:58.080 <v ->Thanks, so one thing that Stephen mentioned was, you know,</v>

689 00:31:58.080 --> 00:31:59.670 another challenge we might face is,

690 00:31:59.670 --> 00:32:02.850 okay, where do we go find people who have data that

691 00:32:02.850 --> 00:32:03.930 they might need help with?

692 00:32:03.930 --> 00:32:08.280 And we were thinking about where does data come from, right?

693 00:32:08.280 --> 00:32:11.640 And so one area that data's generated

694 00:32:11.640 --> 00:32:14.735 from is through devices and manufacturers

695 00:32:14.735 --> 00:32:17.190 make devices that are sitting in labs.

696 00:32:17.190 --> 00:32:19.770 So we thought of the idea of let's have discussions

697 00:32:19.770 --> 00:32:21.090 with these manufacturers

698 00:32:21.090 --> 00:32:23.640 and see if we could form some sort of partnership.

699 00:32:23.640 --> 00:32:26.820 Now when you're forming a partnership in industry,

700 00:32:26.820 --> 00:32:29.370 you need to think about why that would benefit both sides

701 00:32:29.370 --> 00:32:32.550 in order to kind of engage your perspective partner

702 00:32:32.550 --> 00:32:34.140 as to why they should talk to you right?

703 00:32:34.140 --> 00:32:36.810 So one thing that we identified was that

704 00:32:36.810 --> 00:32:38.640 a key aim of manufacturers

705 00:32:38.640 --> 00:32:40.920 is to provide additional support

706 00:32:40.920 --> 00:32:42.861 to their customers or make sure,

707 00:32:42.861 --> 00:32:45.210 hey, I have a customer or a lab that has data

708 00:32:45.210 --> 00:32:48.450 and then what if there's an aspect of their data

709 00:32:48.450 --> 00:32:50.640 they don't know how to do something

710 00:32:50.640 --> 00:32:51.990 or they don't know what to do,

711 00:32:51.990 --> 00:32:53.976 maybe they'll stop using my device down the line

712 00:32:53.976 --> 00:32:57.390 because the data's just not useful to them at this point

713 00:32:57.390 --> 00:32:58.950 'cause they're lacking a skillset.

714 00:32:58.950 --> 00:33:00.900 So we thought of an idea whereby

715 00:33:00.900 --> 00:33:03.150 we could approach device manufacturers

716 00:33:03.150 --> 00:33:05.130 and kind of explain what Stephen explained

717 00:33:05.130 --> 00:33:08.827 about our data science as a service offering and say,

718 00:33:08.827 --> 00:33:11.490 "Hey look, we could form a partnership with you,

719 00:33:11.490 --> 00:33:15.180 whereby as an offering, in addition to extending a warranty

720 00:33:15.180 --> 00:33:19.020 on your device, you could offer custom analysis support

721 00:33:19.020 --> 00:33:22.140 or data science support to any interested customers,

722 00:33:22.140 --> 00:33:24.180 whereby they could use cloud workspaces

723 00:33:24.180 --> 00:33:25.860 to put their data that they're collecting

724 00:33:25.860 --> 00:33:27.628 and then they could work with someone like Phil

725 00:33:27.628 --> 00:33:30.870 to solve a challenge that they might have."

726 00:33:30.870 --> 00:33:33.000 And so we actually successfully

727 00:33:33.000 --> 00:33:36.150 did form such a partnership quite recently.

728 00:33:36.150 --> 00:33:38.190 And if you go to the next slide,

729 00:33:38.190 --> 00:33:40.153 you'll see, so we are now working

730 00:33:40.153 --> 00:33:42.780 with a company called Neurophotometrics.

731 00:33:42.780 --> 00:33:45.990 They produce a device that does the imaging

732 00:33:45.990 --> 00:33:48.064 that Stephen previously described.

733 00:33:48.064 --> 00:33:52.800 And what our partnership involves is we essentially offer

734 00:33:52.800 --> 00:33:56.280 cloud workspaces as a solution to their customers,

735 00:33:56.280 --> 00:33:58.680 whereby when they collect their data,

736 00:33:58.680 --> 00:34:01.590 they can then work on our cloud workspaces alongside Phil

737 00:34:01.590 --> 00:34:03.187 or ourselves and we can work with them

738 00:34:03.187 --> 00:34:05.850 to solve any challenges they might need.

739 00:34:05.850 --> 00:34:08.403 Now who are these customers of Neurophotometrics?

740 00:34:08.403 --> 00:34:10.530 They are a bunch of different labs kind of

741 00:34:10.530 --> 00:34:11.675 all over the world as well.

742 00:34:11.675 --> 00:34:14.070 Mostly academics, some in industry as well.

743 00:34:14.070 --> 00:34:17.105 And so it's that way for us as an organization

744 00:34:17.105 --> 00:34:19.740 to kind of find potential labs

745 00:34:19.740 --> 00:34:21.840 we didn't even know had the challenge.

746 00:34:21.840 --> 00:34:24.510 And then it's also solving the problem

747 00:34:24.510 --> 00:34:26.340 for NeuroPhotometrics of how do you keep your

748 00:34:26.340 --> 00:34:29.010 customers happy if you don't really offer a service

749 00:34:29.010 --> 00:34:30.750 they're already kind of asking of you

750 00:34:30.750 --> 00:34:32.790 as a follow-on for providing this device.

751 00:34:32.790 --> 00:34:36.909 So, so far the partnership is fairly new.

752 00:34:36.909 --> 00:34:39.630 It seems to be working quite well so far

753 00:34:39.630 --> 00:34:40.650 and we're meeting new people

754 00:34:40.650 --> 00:34:42.510 and already getting kind of more projects

755 00:34:42.510 --> 00:34:44.677 like Stephen described for Phil to work on.

756 00:34:44.677 --> 00:34:46.020 So we'll see how it goes.

757 00:34:46.020 --> 00:34:47.490 But this is just one way to show you

758 00:34:47.490 --> 00:34:49.440 that it's not just about kind

759 00:34:49.440 --> 00:34:50.970 of solving a problem for a customer,

760 00:34:50.970 --> 00:34:52.710 it's about where do you find your customers

761 00:34:52.710 --> 00:34:55.623 and that could be through an industry partnership.

762 00:34:57.341 --> 00:35:00.543 <v ->Awesome, thanks for that.</v>

763 00:35:01.500 --> 00:35:06.500 So I mentioned one other model earlier, which is workshops.

764 00:35:08.220 --> 00:35:10.920 I think I talked about that example for a bit.

765 00:35:10.920 --> 00:35:15.920 And we have done a few of them actually as well

766 00:35:17.040 --> 00:35:18.450 in the computational neuroscience space.

767 00:35:18.450 --> 00:35:20.610 So now the space near and dear

768 00:35:20.610 --> 00:35:23.823 to our work with Robert.

769 00:35:25.050 --> 00:35:28.137 So one of those projects was a collaboration

770 00:35:28.137 --> 00:35:30.510 actually Brown University on something

771 00:35:30.510 --> 00:35:32.733 called the Human Neocortical Neurosolver.

772 00:35:34.170 --> 00:35:37.620 We have kind of a neuroscience bias in the company.

773 00:35:37.620 --> 00:35:39.240 We like doing those sorts of things.

774 00:35:39.240 --> 00:35:42.753 So we did a workshop also.

775 00:35:44.070 --> 00:35:46.165 We helped facilitate a workshop

776 00:35:46.165 --> 00:35:49.200 that allowed a software tool

777 00:35:49.200 --> 00:35:54.200 that came out of this particular collaboration to be shown.

778 00:35:56.190 --> 00:36:00.240 And, let me show you a little bit more.

779 00:36:00.240 --> 00:36:03.510 So in this case, I'm actually gonna switch

780 00:36:03.510 --> 00:36:05.037 away from the Human Neocortical Neurosolver

781 00:36:05.037 --> 00:36:07.470 and also show you an example with NetPyNE,

782 00:36:07.470 --> 00:36:09.490 which is the thing that Robert mentioned earlier

783 00:36:09.490 --> 00:36:11.280 that we work with as well.

784 00:36:11.280 --> 00:36:12.750 It's similar to HNN.

785 00:36:12.750 --> 00:36:15.270 In both cases there's a computational model

786 00:36:15.270 --> 00:36:16.230 of a neuron, okay?

787 00:36:16.230 --> 00:36:18.120 Just think of like, you know,

788 00:36:18.120 --> 00:36:21.990 a spatial model of a neuron that has a cell body

789 00:36:21.990 --> 00:36:25.053 and has an axon and dendrite, that kind of thing.

790 00:36:25.053 --> 00:36:27.995 And you wanna simulate something about it.

791 00:36:27.995 --> 00:36:32.995 And so you have a specialized piece of software

792 00:36:34.410 --> 00:36:38.190 that knows how to look at the model of a neuron,

793 00:36:38.190 --> 00:36:39.570 the way that it's shaped

794 00:36:39.570 --> 00:36:44.340 and how to get signals out of it basically, right?

795 00:36:44.340 --> 00:36:48.747 So in collaboration with NetPyNE also a software platform

796 00:36:50.070 --> 00:36:51.663 called Open Source Brain at UCL

797 00:36:51.663 --> 00:36:54.150 that we've been partnering with for a while.

798 00:36:54.150 --> 00:36:57.847 You might have something that looks like this.

799 00:36:57.847 --> 00:37:01.930 So what you can do in a workshop context

800 00:37:02.850 --> 00:37:05.310 with something like a workspace that's really exciting,

801 00:37:05.310 --> 00:37:07.050 as I mentioned to you before is have people

802 00:37:07.050 --> 00:37:09.270 put hands on with the software itself.

803 00:37:09.270 --> 00:37:10.980 And this is one of those pictures

804 00:37:10.980 --> 00:37:13.530 from one of those workshop that we did,

805 00:37:13.530 --> 00:37:15.660 I think this one was specifically NetPyNE

806 00:37:15.660 --> 00:37:17.610 where you can kind of see what everybody's looking at.

807 00:37:17.610 --> 00:37:20.160 So everybody brought laptops in, right?

808 00:37:20.160 --> 00:37:22.920 And they're able to launch in this case

809 00:37:22.920 --> 00:37:24.663 they're literally, you can see several of 'em,

810 00:37:24.663 --> 00:37:27.150 like this one up in front and this one over here,

811 00:37:27.150 --> 00:37:28.952 they literally have exactly the same screen up

812 00:37:28.952 --> 00:37:32.910 that is being shown, you know, in the screen share,

813 00:37:32.910 --> 00:37:34.260 not because they're logged into a Zoom,

814 00:37:34.260 --> 00:37:36.960 but 'cause they're actually logged into essentially

815 00:37:36.960 --> 00:37:40.050 a workspace environment where they can also like, you know,

816 00:37:40.050 --> 00:37:41.070 change parameters around.

817 00:37:41.070 --> 00:37:43.466 So you can get this hands-on tutorial effect

818 00:37:43.466 --> 00:37:46.438 in a workshop, in this context.

819 00:37:46.438 --> 00:37:50.220 That is kind of hard to do any other way

820 00:37:50.220 --> 00:37:51.453 if you don't have that.

821 00:37:52.710 --> 00:37:54.900 If it's deployed as web-based software,

822 00:37:54.900 --> 00:37:56.250 that makes it a little bit easier.

823 00:37:56.250 --> 00:37:57.260 But if it's not, you know,

824 00:37:57.260 --> 00:37:58.620 if it's something that's traditionally supposed

825 00:37:58.620 --> 00:37:59.453 to be on a desktop,

826 00:37:59.453 --> 00:38:03.180 then this is kind of the only way to do something like that.

827 00:38:03.180 --> 00:38:06.300 And this was at a academic conference,

828 00:38:06.300 --> 00:38:08.373 I think CNS that gets held.

829 00:38:09.330 --> 00:38:14.330 So yeah, from all that today then

830 00:38:14.970 --> 00:38:17.430 kind of wrapping up the part where I just,

831 00:38:17.430 --> 00:38:19.860 we just talk at you and I hope those questions

832 00:38:19.860 --> 00:38:23.130 that you guys have, what do we sort of talk about today?

833 00:38:23.130 --> 00:38:26.280 Like how can some cloud-based data science tools

834 00:38:26.280 --> 00:38:29.220 help enhance the ability to do biostatistics

835 00:38:29.220 --> 00:38:30.780 health informatics research?

836 00:38:30.780 --> 00:38:32.460 I've been, you know, leaning on some examples

837 00:38:32.460 --> 00:38:33.630 that are heavily neuroscience based,

838 00:38:33.630 --> 00:38:35.820 but we kind of think that that's not the thing

839 00:38:35.820 --> 00:38:37.110 that's particular to this, right?

840 00:38:37.110 --> 00:38:39.900 It's still, you know, as I started at the beginning,

841 00:38:39.900 --> 00:38:42.330 you know, doing some analysis, you know,

842 00:38:42.330 --> 00:38:45.180 sharing the results of the commands

843 00:38:45.180 --> 00:38:47.040 that we're using in the analysis

844 00:38:47.040 --> 00:38:48.185 and then sharing the output of that analysis, right?

845 00:38:48.185 --> 00:38:49.500 Like that's where we began.

846 00:38:49.500 --> 00:38:51.390 I think that's common to every technique.

847 00:38:51.390 --> 00:38:52.770 We're bringing some kind of science and math

848 00:38:52.770 --> 00:38:54.780 to bear on some data, right?

849 00:38:54.780 --> 00:38:56.640 So what we're finding is that, you know,

850 00:38:56.640 --> 00:38:58.793 by using cloud-based platforms

851 00:38:58.793 --> 00:39:01.702 really can help us facilitate collaborative research,

852 00:39:01.702 --> 00:39:04.916 allowing colleagues to share data and work together.

853 00:39:04.916 --> 00:39:07.710 You can help labs efficiently gain access

854 00:39:07.710 --> 00:39:10.489 to additional data science support if that's desirable.

855 00:39:10.489 --> 00:39:13.530 That they, you know, otherwise might struggle to get

856 00:39:13.530 --> 00:39:15.150 or is just kind of unaffordable.

857 00:39:15.150 --> 00:39:18.558 Doesn't make sense 'cause there's too much of a person.

858 00:39:18.558 --> 00:39:21.330 And then finally in the last example, right,

859 00:39:21.330 --> 00:39:22.650 you can facilitate, you know,

860 00:39:22.650 --> 00:39:25.620 distance workshops that allow much more immediate

861 00:39:25.620 --> 00:39:28.473 hands-on experience with certain software.

862 00:39:29.340 --> 00:39:34.340 So with all that, I will thank you all for listening

863 00:39:35.610 --> 00:39:37.890 to us for a full 40 minutes

864 00:39:37.890 --> 00:39:41.070 and happy to take any questions that you have on this

865 00:39:41.070 --> 00:39:43.530 or any other thing I can help directly.

866 00:39:43.530 --> 00:39:44.530 Thank you very much.

867 00:39:46.243 --> 00:39:47.910 <v ->Thank you so much.</v>

868 00:39:49.620 --> 00:39:53.193 Does anybody have any questions for our presenters?

869 00:39:57.060 --> 00:39:59.853 I'll start if there's no questions.

870 00:40:01.080 --> 00:40:05.253 So data science is a service growth industry.

871 00:40:06.990 --> 00:40:08.193 People want jobs.

872 00:40:10.350 --> 00:40:12.400 What's your take on the industry on that?

873 00:40:13.320 --> 00:40:18.320 <v ->We are about 18 months into our exploration of the market.</v>

874 00:40:21.660 --> 00:40:24.003 We have seen growth so far.

875 00:40:25.140 --> 00:40:27.183 We think there's more to go.

876 00:40:28.380 --> 00:40:29.940 I showed you those five labs,

877 00:40:29.940 --> 00:40:34.500 I think in total maybe served certainly more than a dozen,

878 00:40:34.500 --> 00:40:38.490 I wanna say maybe like 15 and like labs plus companies or so

879 00:40:38.490 --> 00:40:41.253 15, 16, in those 18 months.

880 00:40:42.534 --> 00:40:45.330 We had to figure out lots of other stuff along the way.

881 00:40:45.330 --> 00:40:50.260 But we think there's a need, you know, like I mentioned

882 00:40:52.110 --> 00:40:56.160 and folks that have the skillset to, you know,

883 00:40:56.160 --> 00:40:57.510 provide that data science service

884 00:40:57.510 --> 00:40:59.223 that are continually in demand.

885 00:41:00.390 --> 00:41:03.063 So I'm gonna say yes, it's growing.

886 00:41:03.990 --> 00:41:07.606 We're always wondering in industry how fast, you know,

887 00:41:07.606 --> 00:41:09.510 that's always the question,

888 00:41:09.510 --> 00:41:11.223 but it's definitely not shrinking.

889 00:41:13.050 --> 00:41:15.350 <v Robert>Alright, that's an exciting option.</v>

890 00:41:17.956 --> 00:41:19.710 <v Participant>Yeah just really quick,</v>

891 00:41:19.710 --> 00:41:22.140 what happens with authorship?

892 00:41:22.140 --> 00:41:26.370 If you work with the lab very closely on a project,

893 00:41:26.370 --> 00:41:29.253 they come out with a really good publication.

894 00:41:31.350 --> 00:41:35.790 How do you deal with that in this industry?

895 00:41:35.790 --> 00:41:38.673 <v ->Yeah, great question. Thank you.</v>

896 00:41:40.320 --> 00:41:42.933 So as a company,

897 00:41:44.040 --> 00:41:49.040 we don't require to have our data scientists listed

898 00:41:51.390 --> 00:41:53.823 as co-authors on papers.

899 00:41:55.200 --> 00:42:00.100 I think from an ethical perspective

900 00:42:02.250 --> 00:42:04.620 in the case where the contribution that the data scientist

901 00:42:04.620 --> 00:42:06.850 has made are very significant

902 00:42:09.150 --> 00:42:13.290 you know, sometimes PIs have asked the question to us,

903 00:42:13.290 --> 00:42:15.150 you know, what sort of acknowledgement

904 00:42:15.150 --> 00:42:17.910 would you like of the data scientist?

905 00:42:17.910 --> 00:42:20.730 And if the PI feels that, say, you know,

906 00:42:20.730 --> 00:42:22.800 someone who has a PhD who works with us

907 00:42:22.800 --> 00:42:25.623 has done enough work that it merits authorship,

908 00:42:26.730 --> 00:42:28.320 they're free to add that person.

909 00:42:28.320 --> 00:42:29.700 We don't require that.

910 00:42:29.700 --> 00:42:33.390 Otherwise, you know, an acknowledgements nice always right?

911 00:42:33.390 --> 00:42:35.493 But also not required.

912 00:42:37.140 --> 00:42:39.750 I think, you know, sometimes the nature

913 00:42:39.750 --> 00:42:41.970 of the contribution really matters.

914 00:42:41.970 --> 00:42:46.860 So, you know, as a company it's a little bit

915 00:42:46.860 --> 00:42:49.050 like how much do you acknowledge

916 00:42:49.050 --> 00:42:52.730 the vendor of your microscope, right?

917 00:42:52.730 --> 00:42:55.830 You might say, okay, I did this on a Nikon microscope

918 00:42:55.830 --> 00:42:58.080 or you know, but you might write that more

919 00:42:58.080 --> 00:42:58.913 as a method section.

920 00:42:58.913 --> 00:43:00.027 And then if like a technician came out

921 00:43:00.027 --> 00:43:01.890 and like helped you calibrate it,

922 00:43:01.890 --> 00:43:02.723 you're probably not gonna give

923 00:43:02.723 --> 00:43:04.524 that person an authorship either.

924 00:43:04.524 --> 00:43:06.990 But you might acknowledge them if they did extensive help

925 00:43:06.990 --> 00:43:09.840 that like led to some novel process.

926 00:43:09.840 --> 00:43:13.300 So on the whole, it's a case by case conversation

927 00:43:14.640 --> 00:43:17.100 that scales based on the level of the contribution,

928 00:43:17.100 --> 00:43:18.780 but it's not the first thing that we think of.

929 00:43:18.780 --> 00:43:21.258 It's not like, "Hey, because we did anything for you,

930 00:43:21.258 --> 00:43:23.220 please put us on a paper."

931 00:43:23.220 --> 00:43:24.240 Definitely don't do it that way.

932 00:43:24.240 --> 00:43:26.970 It's more the opposite, which is like, you know,

933 00:43:26.970 --> 00:43:27.990 we're gonna do a thing for you.

934 00:43:27.990 --> 00:43:30.256 Probably, you don't need to cite us.

935 00:43:30.256 --> 00:43:32.640 But if it gets up to a certain point

936 00:43:32.640 --> 00:43:34.770 and we kind of mutually agree that that's appropriate,

937 00:43:34.770 --> 00:43:36.513 then we're happy to discuss that.

938 00:43:41.190 --> 00:43:42.390 <v ->Thank you for sharing Stephen.</v>

939 00:43:42.390 --> 00:43:44.010 So I have a quick question too.

940 00:43:44.010 --> 00:43:46.323 So if you're running on data sets,

941 00:43:47.400 --> 00:43:50.040 one cell may take really long time to run,

942 00:43:50.040 --> 00:43:52.620 then how do you solve the concurrency issue?

943 00:43:52.620 --> 00:43:56.130 Let's say there's multiple people collaborating online

944 00:43:56.130 --> 00:44:00.060 that when the cell is running,

945 00:44:00.060 --> 00:44:04.470 what if some other, another party just clicked stop

946 00:44:04.470 --> 00:44:05.610 or doing something random?

947 00:44:05.610 --> 00:44:08.463 How do you solve the issue that people are on the same page

948 00:44:08.463 --> 00:44:11.433 when something takes really long time to run?

949 00:44:12.510 --> 00:44:13.980 <v ->Yeah, great question.</v>

950 00:44:13.980 --> 00:44:17.550 So a few ways,

951 00:44:17.550 --> 00:44:21.870 one nice thing about a cloud workspace is that

952 00:44:21.870 --> 00:44:25.350 we can expand the number of processors

953 00:44:25.350 --> 00:44:27.960 and the amount of memory kind of

954 00:44:27.960 --> 00:44:30.620 behind the scenes transparently.

955 00:44:30.620 --> 00:44:34.860 So basically you can like log out of the workspace

956 00:44:34.860 --> 00:44:37.890 and in five minutes log back into the workspace

957 00:44:37.890 --> 00:44:39.600 and we've like doubled the processing speed

958 00:44:39.600 --> 00:44:41.730 and like doubled the memory.

959 00:44:41.730 --> 00:44:44.804 So we tend to keep our default instance

960 00:44:44.804 --> 00:44:47.460 at like a reasonable like laptop,

961 00:44:47.460 --> 00:44:48.930 like probably not a high end.

962 00:44:48.930 --> 00:44:52.350 And then when we discover cases like what you're talking

963 00:44:52.350 --> 00:44:55.500 about where like, yeah, no, that cell requires a lot

964 00:44:55.500 --> 00:44:56.850 and we kind of know a little bit in advance,

965 00:44:56.850 --> 00:44:59.490 like we're gonna wanna run that a lot, right?

966 00:44:59.490 --> 00:45:00.930 We might do this, which was we might

967 00:45:00.930 --> 00:45:02.520 like just beef it up, right?

968 00:45:02.520 --> 00:45:06.505 And that's cool that we can do that.

969 00:45:06.505 --> 00:45:09.720 And then the question becomes like,

970 00:45:09.720 --> 00:45:11.940 does that need to run, you know, 24/7,

971 00:45:11.940 --> 00:45:13.170 does it need to run every day,

972 00:45:13.170 --> 00:45:14.970 every week, every month right?

973 00:45:14.970 --> 00:45:16.050 We think a little bit about that

974 00:45:16.050 --> 00:45:18.390 because then there's some additional costs on our side.

975 00:45:18.390 --> 00:45:20.220 If you're gonna do it for like an afternoon,

976 00:45:20.220 --> 00:45:23.930 it's like really not, it's not worth making any additional,

977 00:45:23.930 --> 00:45:26.700 you know, requests of somebody.

978 00:45:26.700 --> 00:45:28.140 But there's another part of your question I wanna get at

979 00:45:28.140 --> 00:45:33.120 too, which is like maybe overriding each other, right?

980 00:45:33.120 --> 00:45:34.170 So that can happen.

981 00:45:34.170 --> 00:45:37.710 And that's a little bit like software specific.

982 00:45:37.710 --> 00:45:42.690 So like in a Jupyter Notebook, you could,

983 00:45:42.690 --> 00:45:45.180 if you don't coordinate a little bit with your lab member,

984 00:45:45.180 --> 00:45:48.810 like overwrite something in one cell at one time, right?

985 00:45:48.810 --> 00:45:50.010 The other person didn't notice.

986 00:45:50.010 --> 00:45:53.163 So for that, we have some best practices, you know.

987 00:45:54.480 --> 00:45:58.770 By far the most common, you know, example that we see is,

988 00:45:58.770 --> 00:46:01.140 is like two or fewer people collaborating,

989 00:46:01.140 --> 00:46:02.550 but if it were three or four,

990 00:46:02.550 --> 00:46:04.860 we'd probably recommend that they do a best practice

991 00:46:04.860 --> 00:46:08.070 of like, you know, while you're doing work that's separate

992 00:46:08.070 --> 00:46:10.350 and you're not like talking to each other,

993 00:46:10.350 --> 00:46:12.645 do work on separate copies of the thing, right?

994 00:46:12.645 --> 00:46:14.940 And then come together in a meeting

995 00:46:14.940 --> 00:46:17.220 and like put it back together, right?

996 00:46:17.220 --> 00:46:19.980 Usually is the better practice if you're say,

997 00:46:19.980 --> 00:46:22.211 working on a Jupyter Notebook,

998 00:46:22.211 --> 00:46:24.716 and you know, communicate, you know,

999 00:46:24.716 --> 00:46:28.140 using some other method like a meeting like this.

1000 00:46:28.140 --> 00:46:29.670 So yeah so those are the two aspects.

1001 00:46:29.670 --> 00:46:31.830 On the one side, if it's computation intensive,

1002 00:46:31.830 --> 00:46:33.210 we can make it bigger.

1003 00:46:33.210 --> 00:46:34.824 If it's actually about people writing each other,

1004 00:46:34.824 --> 00:46:36.720 we recommend some best practices

1005 00:46:36.720 --> 00:46:38.870 for communicating outside of the workspace.

1006 00:46:42.090 --> 00:46:42.993 <v ->Other questions?</v>

1007 00:46:47.310 --> 00:46:49.830 All right, I have one more question.

1008 00:46:49.830 --> 00:46:53.400 So like in the old days,

1009 00:46:53.400 --> 00:46:56.904 people would buy a nice computer for their lab or maybe a

1010 00:46:56.904 --> 00:47:00.150 couple of nice computers and like then everybody

1011 00:47:00.150 --> 00:47:04.798 would log in at that and it was a one-time cost, right?

1012 00:47:04.798 --> 00:47:08.940 And so how have you found, I don't know,

1013 00:47:08.940 --> 00:47:13.940 I mean, so it's a very different model for

1014 00:47:14.057 --> 00:47:17.970 both academia industry, wherever that's trying

1015 00:47:17.970 --> 00:47:21.480 to transition from this one time cost

1016 00:47:21.480 --> 00:47:24.480 where now, you know, you might still be using this computer

1017 00:47:24.480 --> 00:47:27.430 10 years later for good and ill

1018 00:47:29.070 --> 00:47:34.070 versus sort of this continuous cloud-based thing.

1019 00:47:34.080 --> 00:47:34.913 I don't know,

1020 00:47:34.913 --> 00:47:38.613 do you have any words of wisdom on this transition?

1021 00:47:39.480 --> 00:47:42.090 Because it seems like, you know, you pay

1022 00:47:42.090 --> 00:47:46.050 for a cloud computer and if it’s on constantly,

1023 00:47:46.050 --> 00:47:48.150 it eats up a lot of money.

1024 00:47:48.150 --> 00:47:49.110 <v ->Yeah, yeah.</v>

1025 00:47:49.110 --> 00:47:50.853 So really good question.

1026 00:47:53.370 --> 00:47:54.203 So I think and-

1027 00:47:54.203 --> 00:47:57.900 <v ->Lose control of your data also, which to some extent,</v>

1028 00:47:57.900 --> 00:48:00.480 like somebody else has your data.

1029 00:48:00.480 --> 00:48:01.740 <v ->In theory, yes.</v>

1030 00:48:01.740 --> 00:48:05.910 But you know, I think some of this is just like a journey

1031 00:48:05.910 --> 00:48:09.120 and a transition that, you know, scientists are making.

1032 00:48:09.120 --> 00:48:11.070 Those of us, like yourself,

1033 00:48:11.070 --> 00:48:13.050 we’re more software engineer minded,

1034 00:48:13.050 --> 00:48:15.930 have been comfortable with the idea of say, you know,

1035 00:48:15.930 --> 00:48:18.090 like all of our company’s data, for example,

1036 00:48:18.090 --> 00:48:20.550 is kind of in Google’s clouds,

1037 00:48:20.550 --> 00:48:22.320 Google’s workspace technically.

1038 00:48:22.320 --> 00:48:24.900 None of it is sitting under my desk, right?

1039 00:48:24.900 --> 00:48:28.170 But we’ve gotten a level of comfort about data ownership

1040 00:48:28.170 --> 00:48:31.830 based on essentially trust and agreements

1041 00:48:31.830 --> 00:48:34.440 and our understanding of how certain sections

1042 00:48:34.440 --> 00:48:37.860 of disk are like cordoned off, you know, for ourselves

1043 00:48:37.860 --> 00:48:40.110 and lying on some of those best practices.

1044 00:48:40.110 --> 00:48:42.933 But to get to the heart of your question,

1045 00:48:43.920 --> 00:48:45.300 I think the best metaphor is like

1046 00:48:45.300 --> 00:48:48.330 buying a house versus renting an apartment, right?

1047 00:48:48.330 --> 00:48:50.847 So, you know, going down to Apple

1048 00:48:50.847 --> 00:48:54.510 and picking up a laptop or Dell or whatever you wanna use,

1049 00:48:54.510 --> 00:48:56.130 right, is that's the buy model.

1050 00:48:56.130 --> 00:48:57.870 And we're super comfortable with that.

1051 00:48:57.870 --> 00:49:00.630 The cloud model is more the like renting the apartment.

1052 00:49:00.630 --> 00:49:03.270 And certainly people make the choice,

1053 00:49:03.270 --> 00:49:04.800 you know, not to rent sometimes

1054 00:49:04.800 --> 00:49:06.783 because it's like, doesn't work out economically, right?

1055 00:49:06.783 --> 00:49:09.120 It's like, "Hey, I'm throwing money away."

1056 00:49:09.120 --> 00:49:10.650 Sometimes people throw, right?

1057 00:49:10.650 --> 00:49:12.630 But what is the advantage of renting, right?

1058 00:49:12.630 --> 00:49:15.510 The advantage of renting is, you know,

1059 00:49:15.510 --> 00:49:17.190 if a thing breaks in your rented apartment,

1060 00:49:17.190 --> 00:49:19.740 it's not on you to go pay extra money to go fix it.

1061 00:49:19.740 --> 00:49:21.090 That's on the person who owns it.

1062 00:49:21.090 --> 00:49:23.857 Similarly, if something breaks with your cloud workspace,

1063 00:49:23.857 --> 00:49:25.987 you know, you call us and you're like,

1064 00:49:25.987 --> 00:49:27.450 "Hey, this thing didn't work,

1065 00:49:27.450 --> 00:49:29.010 please fix it, right?"

1066 00:49:29.010 --> 00:49:30.510 And then there's this scaling thing, right?

1067 00:49:30.510 --> 00:49:32.167 Which is like, if you go back to Apple and you're like,

1068 00:49:32.167 --> 00:49:37.167 "Actually can you add like double the CPU

1069 00:49:37.440 --> 00:49:38.880 and double the memory?"

1070 00:49:38.880 --> 00:49:41.040 They'll be like, yes, you can pay us for that,

1071 00:49:41.040 --> 00:49:42.840 but it's gonna take a while, right?

1072 00:49:42.840 --> 00:49:44.490 And it's not gonna happen flexibly and scalably.

1073 00:49:44.490 --> 00:49:48.241 So I think it fits into a different space, right?

1074 00:49:48.241 --> 00:49:49.710 Obviously these two come together,

1075 00:49:49.710 --> 00:49:52.470 I'm talking to you on a physical laptop that I own, right?

1076 00:49:52.470 --> 00:49:55.612 But I'm also using cloud instances to do things.

1077 00:49:55.612 --> 00:49:59.850 So I think it's like, it fits into this niche where like,

1078 00:49:59.850 --> 00:50:02.910 actually the most useful computer for this purpose,

1079 00:50:02.910 --> 00:50:05.130 this collaborative purpose

1080 00:50:05.130 --> 00:50:08.070 is a rented one, right rather than an owned one.

1081 00:50:08.070 --> 00:50:10.560 And you know, maybe that means when I'm not using it,

1082 00:50:10.560 --> 00:50:13.110 I'm not paying for it at all, basically, right?

1083 00:50:13.110 --> 00:50:14.910 Like, if I'm like paused on this collaboration,

1084 00:50:14.910 --> 00:50:16.680 then I'm like actually not paying for it at all,

1085 00:50:16.680 --> 00:50:18.420 but then I can bring 'em back and six months and start

1086 00:50:18.420 --> 00:50:19.530 paying for it again.

1087 00:50:19.530 --> 00:50:22.410 So this is what I hope that folks take away is like,

1088 00:50:22.410 --> 00:50:24.300 it opens up a lot of new possibilities.

1089 00:50:24.300 --> 00:50:25.800 And the ones that we've gotten

1090 00:50:25.800 --> 00:50:26.940 are certainly not the only ones.

1091 00:50:26.940 --> 00:50:28.410 There's just like lots more

1092 00:50:28.410 --> 00:50:30.063 that you can imagine or envision.

1093 00:50:31.620 --> 00:50:34.679 But, but yeah, it's a mindset change

1094 00:50:34.679 --> 00:50:36.840 and it's one that I think, you know,

1095 00:50:36.840 --> 00:50:40.293 requires some adapting, yeah.

1096 00:50:42.030 --> 00:50:43.860 <v ->All right. Thank you so much.</v>

1097 00:50:43.860 --> 00:50:45.390 <v ->I have a question for you guys</v>

1098 00:50:45.390 --> 00:50:48.427 if there's not another question for me.

1099 00:50:48.427 --> 00:50:51.420 <v ->There's a question on the screen.</v>

1100 00:50:51.420 --> 00:50:52.770 <v ->Sorry, I have a question.</v>

1101 00:50:53.760 --> 00:50:56.763 I think piggy-backing off of that question-

1102 00:50:57.930 --> 00:50:59.550 <v ->Hi hello. Hi Noelle.</v>

1103 00:50:59.550 --> 00:51:00.753 <v ->Actually Hi.</v>

1104 00:51:01.890 --> 00:51:06.890 I used to like physical like pieces of data

1105 00:51:08.280 --> 00:51:10.470 and like having physical hard drives.

1106 00:51:10.470 --> 00:51:15.470 So like what is the security for data that's on the cloud?

1107 00:51:16.410 --> 00:51:18.680 <v ->Yeah, so folks like,</v>

1108 00:51:24.270 --> 00:51:29.270 we ourselves build these cloud instances

1109 00:51:29.700 --> 00:51:32.460 on the back of three major providers,

1110 00:51:32.460 --> 00:51:33.420 whose names you'll recognize,

1111 00:51:33.420 --> 00:51:37.200 Amazon, Google, and Microsoft okay?

1112 00:51:37.200 --> 00:51:39.660 Those are the big three cloud providers

1113 00:51:39.660 --> 00:51:42.990 and they make a guarantee to us

1114 00:51:42.990 --> 00:51:45.570 and then we make a guarantee to our customers

1115 00:51:45.570 --> 00:51:46.920 about the data protection.

1116 00:51:46.920 --> 00:51:48.690 So it's kind of like a layer cake.

1117 00:51:48.690 --> 00:51:51.780 And the foundation of it begins with, do you trust Amazon?

1118 00:51:51.780 --> 00:51:53.220 Do you trust Google? Do you trust Microsoft?

1119 00:51:53.220 --> 00:51:55.564 Some people say yes, some people say no,

1120 00:51:55.564 --> 00:51:58.950 but fundamentally they are the ones that, you know,

1121 00:51:58.950 --> 00:52:03.600 build data centers, right where the physical aspect

1122 00:52:03.600 --> 00:52:05.250 of these computers actually live.

1123 00:52:05.250 --> 00:52:07.222 So, you know, this virtual computer,

1124 00:52:07.222 --> 00:52:09.337 maybe if you go and like, 1125 00:52:09.337 --> 00:52:12.180 ”Hey, show me the hard drive where this lives.”

1126 00:52:12.180 --> 00:52:13.590 You’re gonna go out to like, I don’t know,

1127 00:52:13.590 --> 00:52:17.850 Washington State near some power plant basically,

1128 00:52:17.850 --> 00:52:20.700 where it’s very economical to set this up, right?

1129 00:52:20.700 --> 00:52:24.690 So they then guarantee like,

1130 00:52:24.690 --> 00:52:26.850 how do you know that that’s safe, right?

1131 00:52:26.850 --> 00:52:30.375 Well they guarantee that they’re following industry

1132 00:52:30.375 --> 00:52:35.370 standards to secure those facilities, to lock them down,

1133 00:52:35.370 --> 00:52:40.370 to like continually maintain and manage the networks

1134 00:52:40.830 --> 00:52:43.560 that are there to patch the servers

1135 00:52:43.560 --> 00:52:46.950 that they’re using to keep ahead of any security faults.

1136 00:52:46.950 --> 00:52:48.840 So there’s one layer of this

1137 00:52:48.840 --> 00:52:52.170 where we rely on these big providers to do their jobs.

1138 00:52:52.170 --> 00:52:57.150 And despite the last 15, 20 years of like hacks

1139 00:52:57.150 --> 00:52:59.760 that you’ve heard about whatnot that happened in industry,

1140 00:52:59.760 --> 00:53:03.142 these three providers so far have managed to avoid

1141 00:53:03.142 --> 00:53:05.220 being hacked in any major way.

1142 00:53:05.220 --> 00:53:07.680 Like you’ve not heard of like Amazon getting hacked,

1143 00:53:07.680 --> 00:53:09.600 Google getting hacked, Microsoft getting hacked.

1144 00:53:09.600 --> 00:53:12.930 If tomorrow Amazon gets hacked, then yeah,

1145 00:53:12.930 --> 00:53:14.250 we’re all worried okay?

1146 00:53:14.250 --> 00:53:16.260 And then we probably would need to shift around.

1147 00:53:16.260 --> 00:53:18.630 But so there's a fundamental guarantee

1148 00:53:18.630 --> 00:53:20.940 that like all cloud kind of relies on

1149 00:53:20.940 --> 00:53:22.500 and it's like good to talk about it

1150 00:53:22.500 --> 00:53:26.850 because like we all have to kind of trust these,

1151 00:53:26.850 --> 00:53:28.860 you know, these large providers.

1152 00:53:28.860 --> 00:53:31.380 But they also invest,

1153 00:53:31.380 --> 00:53:34.230 I'd say millions or hundreds of millions of dollars

1154 00:53:34.230 --> 00:53:35.490 in computer security.

1155 00:53:35.490 --> 00:53:38.160 Like if you're in the field of computer security,

1156 00:53:38.160 --> 00:53:40.860 like, you know these guys because they are sort

1157 00:53:40.860 --> 00:53:43.602 of world leaders in this sort of thing.

1158 00:53:43.602 --> 00:53:47.610 Microsoft, you know, notably was involved in doing some

1159 00:53:47.610 --> 00:53:51.510 forensic analysis on like Russian hacking back in 2016.

1160 00:53:51.510 --> 00:53:54.600 Like they were some of the first people to notice

1161 00:53:54.600 --> 00:53:57.930 that a state actor like Russia was on the scene

1162 00:53:57.930 --> 00:54:00.480 doing the various things, taking over computers.

1163 00:54:00.480 --> 00:54:04.500 So generally the community of software engineers

1164 00:54:04.500 --> 00:54:07.145 that do cloud work know these things

1165 00:54:07.145 --> 00:54:11.070 and kind of rely on Google, Amazon, and Microsoft

1166 00:54:11.070 --> 00:54:14.430 to like make these investments in computer security.

1167 00:54:14.430 --> 00:54:18.446 And notably like, I don't go like set up my own data center

1168 00:54:18.446 --> 00:54:20.550 because I know that I would have to invest millions

1169 00:54:20.550 --> 00:54:24.930 of dollars in having an equivalently good computer security

1170 00:54:24.930 --> 00:54:27.343 team to like watch out for Russia,

1171 00:54:27.343 --> 00:54:30.030 who by the way also invests hundreds of millions of dollars

1172 00:54:30.030 --> 00:54:30.990 to try to hack these things.

1173 00:54:30.990 --> 00:54:34.530 So, the world of computer security is a problem.

1174 00:54:34.530 --> 00:54:36.600 So there's that level of trust, okay?

1175 00:54:36.600 --> 00:54:39.390 And then on top of that, you have to trust one more level,

1176 00:54:39.390 --> 00:54:41.120 which is the group that like sets up the workspace.

1177 00:54:41.120 --> 00:54:42.840 So you kinda have to trust, like if it's from us,

1178 00:54:42.840 --> 00:54:45.330 you have to kind of trust us that we're not screwing

1179 00:54:45.330 --> 00:54:48.300 something up on top of all of those protections

1180 00:54:48.300 --> 00:54:51.240 'cause it is possible to do that at the level of like,

1181 00:54:51.240 --> 00:54:55.110 you know, Jupyter Notebook that our logins are well used.

1182 00:54:55.110 --> 00:54:58.663 So we also invest in using industry standard

1183 00:54:58.663 --> 00:55:01.770 like login protocols, so that only the people that we say

1184 00:55:01.770 --> 00:55:03.955 can log in can log in, right?

1185 00:55:03.955 --> 00:55:07.440 There's a layer of software security there that, you know,

1186 00:55:07.440 --> 00:55:11.190 we have to be on top of patching at one level also.

1187 00:55:11.190 --> 00:55:13.170 So these are all the things that make that secure.

1188 00:55:13.170 --> 00:55:14.760 And the last thing would be like,

1189 00:55:14.760 --> 00:55:18.391 do you or don't you trust us to like not to, 1190 00:55:18.391 --> 00:55:21.330 to not go in and do something nefarious with your data

1191 00:55:21.330 --> 00:55:23.490 even though we’re the only ones that can control it.

1192 00:55:23.490 --> 00:55:25.080 So you trust that nobody else can get into it,

1193 00:55:25.080 --> 00:55:25.913 but do you trust us?

1194 00:55:25.913 --> 00:55:27.180 And then that becomes,

1195 00:55:27.180 --> 00:55:29.015 yeah a question of like, you know,

1196 00:55:29.015 --> 00:55:32.070 going back and checking your references, you know,

1197 00:55:32.070 --> 00:55:34.800 talking to other PIs, making sure that something nefarious

1198 00:55:34.800 --> 00:55:36.990 hasn’t happened, you know, there.

1199 00:55:36.990 --> 00:55:39.150 And you probably wanna gain some confidence on that.

1200 00:55:39.150 --> 00:55:41.580 But what we’ve found is that organizations

1201 00:55:41.580 --> 00:55:43.170 are getting more and more comfortable with that.

1202 00:55:43.170 --> 00:55:46.470 Dropbox is a publicly traded company,

1203 00:55:46.470 --> 00:55:47.940 lots of people put stuff on Dropbox.

1204 00:55:47.940 --> 00:55:48.930 When you put something on Dropbox,

1205 00:55:48.930 --> 00:55:51.000 you’re essentially trusting Dropbox.

1206 00:55:51.000 --> 00:55:52.920 Dropbox is also built on one of these

1207 00:55:52.920 --> 00:55:55.314 three providers same way, right?

1208 00:55:55.314 --> 00:55:57.349 So it’s that kind of idea

1209 00:55:57.349 --> 00:56:00.762 that takes some getting used to but you know,

1210 00:56:00.762 --> 00:56:04.590 becomes increasingly useful to do this kind of work on.

1211 00:56:04.590 --> 00:56:07.289 And we see large banks and large pharma companies

1212 00:56:07.289 --> 00:56:10.374 having taken their time to also adopt cloud

1213 00:56:10.374 --> 00:56:12.510 large financial institutions.

1214 00:56:12.510 --> 00:56:14.640 But over time there's been increasing comfort

1215 00:56:14.640 --> 00:56:17.370 as some of these security questions

1216 00:56:17.370 --> 00:56:19.503 have been, you know, asked and answered.

1217 00:56:20.460 --> 00:56:21.960 So bit of a long answer,

1218 00:56:21.960 --> 00:56:25.143 but thank you for the question 'cause it's important.

1219 00:56:26.610 --> 00:56:27.630 <v ->Alright, thanks so much.</v>

1220 00:56:27.630 --> 00:56:28.770 In the interest of time,

1221 00:56:28.770 --> 00:56:31.942 I think we're gonna have to stop it here, thanks again.

1222 00:56:31.942 --> 00:56:36.942 Really appreciate. (audio garbles)

1223 00:56:36.960 --> 00:56:39.690 <v ->Thank you guys. Thank you all for your time.</v>

1224 00:56:39.690 --> 00:56:40.640 <v ->Have a great day.</v>