WEBVTT

1
00:00:00.329 --> 00:00:06.600
Manish Parashar: Software in science. I he has been supported by a number of NSF grants, including the

2
00:00:08.160 --> 00:00:15.030
Manish Parashar: NSF Career Award and the 2019 PKs award, as well as by the Sloan Foundation.

3
00:00:16.260 --> 00:00:21.120
Manish Parashar: We're very proud of James as a long term Boise.

4
00:00:22.230 --> 00:00:22.770
Manish Parashar: p-i

5
00:00:24.000 --> 00:00:38.670
Manish Parashar: James has published in the field of information science, computer supported corporate work and Information Sciences, as well as given keynotes and industry and fun and his and his on funding agency advisory panels.

6
00:00:39.360 --> 00:00:53.820
Manish Parashar: He recently contributed to the development of site as a system to improve incentives for high quality software work in science, my mapping from software to its requested citations.

7
00:00:55.440 --> 00:00:59.400
Manish Parashar: without delaying any further welcome James and over to you.

8
00:01:00.810 --> 00:01:04.890
James Howison: Well, thank you so much munition so a great pleasure to be here and thanks to everyone else.

9
00:01:05.910 --> 00:01:17.220
James Howison: So the big question of my research frames, what we'll talk about today is how does working on things made out of digital information changed the way we collectively work so

10
00:01:18.870 --> 00:01:31.500
James Howison: When we build things out of information when we build digital things. There are two really key performances and afforded says are things we can do with the stuff right so digital information can be copied.

11
00:01:32.250 --> 00:01:41.790
James Howison: Although it has it has high design costs. That's how costly to put together, but then ultra low instantiate costs cheap network distribution write once, run anywhere.

12

00:01:42.330 --> 00:01:50.310
James Howison: And that encourages to think of software as an artifact, something that once we created. We can easily share it. Everyone gets a car.

13
00:01:52.110 --> 00:01:58.950
James Howison: The second performance of digital information is that we can break it down into pieces and we can easily recombine it

14
00:01:59.970 --> 00:02:09.060
James Howison: And so that means the digital information is very flexible. Right. It can be patched can be rapid wrapped it can be extended, it can be recombined

15
00:02:09.660 --> 00:02:18.690
James Howison: And so that's really exciting, particularly in fast moving spaces like like science and other research means we can have lots of new ways to do things.

16
00:02:19.140 --> 00:02:28.470
James Howison: But there's a sting in the tail of recombination as an avoidance and as we'll see in this proposal dealing. Sorry. In this presentation, dealing with that.

17
00:02:29.580 --> 00:02:31.950
James Howison: needs us to think of software as an activity.

18
00:02:34.680 --> 00:02:37.680
James Howison: So today I'm going to talk about software sustainability.

19
00:02:38.520 --> 00:02:48.000
James Howison: And before we get started on give a definition there. And thanks to many years of work in the workshops on the on software sustainability, run by Dan cats and others.

20
00:02:48.630 --> 00:02:57.270
James Howison: And I'll define sustainability as the condition that results when the work needed to keep software scientifically useful is in fact undertaken.

21
00:03:01.320 --> 00:03:07.530
James Howison: Before we go any further, I was asked to do a slide about me. So, you know, it's very exciting to be here. I came from Australia.

22
00:03:08.460 --> 00:03:17.940
James Howison: To the US for a PhD as many international students do it's a well trodden path I did my PhD at Syracuse University with Kevin Cranston.

23
00:03:18.510 --> 00:03:29.970

James Howison: And I arrived, having worked a little bit in management consulting and with an undergraduate degree in economics, where I wrote a somewhat embarrassingly titled making the cyber world safe for capitalism.

24
00:03:30.720 --> 00:03:45.600
James Howison: On his thesis and did some master study in software engineering, just a little bit of programming. After that I spent some time in in Thailand doing old school distance learning paper and pen smuggled in on the back of trucks to refugee camps in Burma.

25
00:03:47.070 --> 00:03:56.070
James Howison: And I came out to my PhD thinking, you know, open source is all altruistic and open and nonprofit organizations in the developing world need help. So why don't we put these two together.

26
00:03:56.880 --> 00:04:05.700
James Howison: And of course, as most people do. I took that topic chopped it in half, chopped in half again kept going until I had a dissertation. And in fact, I'll get to talk a little bit about that today.

27
00:04:06.540 --> 00:04:12.930
James Howison: Then I had a wonderful formative postdoc at Carnegie Mellon with Jim herb slip. What we started studying scientific software.

28
00:04:14.790 --> 00:04:16.860
James Howison: deeply inspired by people studying

29
00:04:17.880 --> 00:04:28.560
James Howison: At the intersection of science and technology studies and scholarly communication and organization science so scholars like PAUL EDWARDS they repairs Janet for Tessie Christine Boardman

30
00:04:30.030 --> 00:04:38.520
James Howison: All crucial one of my most formative experiences was also NSF funded at the Consortium for the science of socio technical systems CSS t

31
00:04:39.810 --> 00:04:49.350
James Howison: People have been involved there. Steve soil and ladders and Tom Finn hold and many others. And so that was a summer camp kind of between my PhD and my postdoc built up those networks.

32
00:04:49.950 --> 00:04:58.170
James Howison: And last year, I was meant to, you know, to help give that back by helping organize CSS to of course that was put on hold true to the pandemic and we hope to get back to that very quickly.

33
00:04:59.490 --> 00:05:11.310
James Howison: Now I'm an associate professor at the School of Information. I've been very grateful for NSF support that that Minish mentioned, and I'm very proud to work with doctoral students, such as Hannah Cahoon and confined. Do you can

34
00:05:11.910 --> 00:05:19.020
James Howison: Face. As you can see that and hoping to pay it forward now through organizing CSS T hosting a CI fellow, which is very exciting.

35
00:05:19.560 --> 00:05:28.380
James Howison: Also developing site as.org which Minish mentioned and providing the soft side data set, which I hope to have a few minutes to talk about at the end. So that's me.

36
00:05:29.700 --> 00:05:39.780
James Howison: Okay, so four parts to today's presentation. First I want to give some images of open source peer production as well as software work in science based on

37
00:05:40.500 --> 00:05:42.030
James Howison: Ethnographic method fieldwork.

38
00:05:42.900 --> 00:05:51.300
James Howison: And then I want to talk at a conceptual level drawing on conversations with many people in this community over the years about why sustainability is difficult for software.

39
00:05:51.540 --> 00:06:00.570
James Howison: And particularly software in science during on some work of mine with Jim have slept. Then I want to talk about work coming out of my career grant, which is a

40
00:06:01.170 --> 00:06:08.640
James Howison: Routes to sustainability via peer production. And finally, I want to do a little bit of forward looking, and talk about

41
00:06:09.240 --> 00:06:24.180
James Howison: incentivizing sustainable ecosystems. What happens when this recombination possibility of digital information, you know, starts to dominate and how we might start to get a handle on that. I look forward to your questions. Okay.

42
00:06:25.500 --> 00:06:32.490
James Howison: So way back in grad school. I like many, many grad graduate students was responsible for organizing References

43
00:06:33.000 --> 00:06:40.320
James Howison: And the screenshot, you see here is the big desk project. And that's what I was using to organize big tech references and

44
00:06:40.950 --> 00:06:45.570
James Howison: I was working on a very small laptop. A LITTLE MAC 11 inch or something and

45
00:06:46.260 --> 00:06:55.320
James Howison: At the time, the software had two columns one for conference venue and one for journal. And I thought, well, look, those are never both fields. So we could merge them into the container calm.

46
00:06:56.160 --> 00:07:02.100
James Howison: And so being inspired by open source and wanting to explore it more I I dived into that and

47
00:07:03.000 --> 00:07:15.240
James Howison: And you can see here, and this comes comes from my dissertation, actually, you know, this is how it was built. So I had use of the software, leading to annoyance downloading the code copy and pasting the nearest feature.

48
00:07:15.750 --> 00:07:22.680
James Howison: The cycle of build test and debug 14 hours over two days, middle of the night stuff because you know I'm in grad school. There's no child to wake me up.

49
00:07:23.460 --> 00:07:33.990
James Howison: Early in the morning and then I submitted an almost working patch to the mailing list and immediately got a response from somebody who said, oh, this is great. Oh, I think you can advance it with this small change there.

50
00:07:35.130 --> 00:07:44.700
James Howison: And then I was able to pull it back and inspired by that contact, I was able to build test and debug it a little bit more. And then finally, check a working version into the repository.

51
00:07:46.500 --> 00:08:02.670
James Howison: So that was, that was me working on code built that was a small incremental change that happened quickly that was built on the code of others. It was built in a community in the sense that I didn't do it all on my own, but I did all the programming on it.

52
00:08:05.100 --> 00:08:05.490
James Howison: So when

53
00:08:06.540 --> 00:08:16.230
James Howison: As I did my observations of big desk. I also, you know, found. Found these emails and you can see this is way back in 2003 So way back, you know, towards the start of open source.

54
00:08:17.220 --> 00:08:24.480
James Howison: And so this was talking about web groups where you'd subscribe to somebody list of publications on their webpage and here's the main developer right saying

55
00:08:24.900 --> 00:08:32.190
James Howison: I really want to use this that the conditions have never quite been right either. I was waiting for this or that. But I ran out of free time.

56
00:08:34.770 --> 00:08:42.690
James Howison: So what happened was there was the project, saying, hey, we really want this feature, this would be great. But I don't have the time to do it and what didn't.

57
00:08:44.160 --> 00:08:52.350
James Howison: What didn't happen then was the project, setting down and building out something like this. And this is just an example Gantt chart I pulled off the internet and

58
00:08:53.400 --> 00:08:55.860
James Howison: It, it shows you know work breakdown

59
00:08:56.910 --> 00:09:05.370
James Howison: With people do jobs that feed into each other and eventually the outcomes you want. That's what didn't happen. What actually happened was for four years, four or more years later.

60
00:09:05.790 --> 00:09:10.620
James Howison: They actually may actually pat check that code in well actually it was an email with a patch back in those days.

61
00:09:11.010 --> 00:09:19.800
James Howison: It was much easier than I expected because the existing groups code was so easy to extend kudos, I wouldn't even have tried it so much hadn't already been solved. Well,

62
00:09:20.790 --> 00:09:31.770
James Howison: So that's an example of a group deferring work that was too complex to do in the way that was available to them, which was primarily one person working on their own over time.

63
00:09:32.970 --> 00:09:38.790
James Howison: Over a period of, you know, no more than two weeks, right, one person no dependencies. No risks on other people.

64
00:09:39.690 --> 00:09:46.620
James Howison: You know, in that work. And so I took those those insights and I suppose that the open source way and I went out and I looked at other projects.

65
00:09:46.860 --> 00:09:54.480
James Howison: And what I found in studying to instant messenger clients firing game was. Yeah, the vast majority of tasks or individual tasks.

66
00:09:54.870 --> 00:10:00.450
James Howison: Which was solo work right. So people working on their own in a short period of time and checking it in

67

00:10:00.780 --> 00:10:07.800
James Howison: And you can see here taking the digital traces of their activity and the shaded area is the inter release period.

68
00:10:08.070 --> 00:10:20.430
James Howison: And you can see the triangles are the actual checking in of code. So the squares early are the community saying, hey, we'd like to do these things. But then the work itself happens when the software is ready to support it and it happens quite quickly.

69
00:10:22.590 --> 00:10:32.070
James Howison: So my claim is not that all open source works like this. Clearly there's corporate involvement and a lot of things that we think of as primary open source projects now have

70
00:10:32.310 --> 00:10:42.930
James Howison: Work breakdowns internally and teams that run open source project offices and things like that. But when you think about what's special in different about the way open source organized

71
00:10:43.620 --> 00:10:53.010
James Howison: Is it uses this Forbidden City digital artifact such that work can be done in very small layers on top of each other their individual short and layered

72
00:10:53.460 --> 00:11:02.700
James Howison: And really complex work is deferred until it's easier and it doesn't always get easier. But this allows a process of production.

73
00:11:03.180 --> 00:11:11.610
James Howison: Which I think is really interesting. And this comes from a paper with Kevin Cranston that we published it in mice quarterly. And here what we can see if we start

74
00:11:12.690 --> 00:11:24.330
James Howison: You can see my mouse is maybe not. Well, we can see on the left is a growing user base using a changing code base. So the code going out into the community inspiring ideas which fly across the top.

75
00:11:24.780 --> 00:11:34.140
James Howison: And producing new developers that are attracted in and those stock of improvement ideas are referred to and people ask, are they motivating yet if they are

76
00:11:34.740 --> 00:11:49.470
James Howison: Then we think about can we do it at relatively low completion risk and if we do it jumps into the software and heads out of the tracks new users and contributors. If we can't, then it dives back and is stored until the conditions are right.

77
00:11:50.730 --> 00:12:00.480
James Howison: So that's an image of of scientists of open source. Now, and a bunch of scientific software work. So he

said, Well, how does a cubic kilometer of ice become a scientific paper.

78
00:12:01.170 --> 00:12:08.730
James Howison: And this is where can I get a debrief Jim hope sled. So we did interviews with the ice cube project. So you first find some ice. You build a big drill

79
00:12:09.360 --> 00:12:19.350
James Howison: You build some digital optical modules and sign in with your name and then you drop them way down in holes in Antarctica you gather up the data. So NSF sponsored work.

80
00:12:20.490 --> 00:12:30.510
James Howison: You store and analyze it after squeezing it back and eventually shipping it back you simulate the light in the ice, you have to simulate the atmosphere of these neutrinos you model it

81
00:12:31.050 --> 00:12:42.300
James Howison: And then of course you analyze it and this is where we really eventually we get plots and, of course, then you get together with a few of your closest friends and publish a paper. So of course software is everywhere. Through this workflow.

82
00:12:43.350 --> 00:12:52.680
James Howison: And and you can really get a feel for that. If you read. PAUL EDWARDS work on the fast machine about how all the things that go into making up a climate science fact.

83
00:12:53.640 --> 00:13:04.170
James Howison: In other work we studied scientific software development with three cases where we looked at focal papers were those papers. We then looked at all the software that they used in their

84
00:13:04.500 --> 00:13:14.760
James Howison: Workflow and we went out and interviewed all the developers of that software identifying the incentives and we worked down the stack. So here you can see an example of that there

85
00:13:15.600 --> 00:13:22.530
James Howison: And at the bottom you can see, you know, the top is the workflow that we extracted through interviews, because the papers don't have all this information in it.

86
00:13:23.070 --> 00:13:39.510
James Howison: And at the bottom you can see the pieces of software marked with the green stars. So what we did in that work was asked what motivating people to contribute to scientific software work and we looked at as we're talking to be able to talk about the pace of their work right.

87
00:13:41.850 --> 00:13:48.420
James Howison: So one of the things we found with the pace of the work was that people write papers every six, nine months, mostly

88
00:13:48.780 --> 00:13:55.230
James Howison: And so they're touching that software on a much longer time cycle then somebody working with bit tech, for example.

89
00:13:55.830 --> 00:14:07.650
James Howison: Now that's less true of infrastructure software which runs along in the background churning out simulations. But when the scientific end users are working with it, they're touching that software on much longer time cycles.

90
00:14:08.190 --> 00:14:24.030
James Howison: And that will be really important when we come back, talk about ecosystem complexity of the end. But what we found about motivations in science was we found software was motivated in a couple of different ways. So we found software framed as support. So here we've got a

91
00:14:26.340 --> 00:14:39.810
James Howison: Cycle of academic reputation where resources turn into research has resorts turns into publications publications and citations produce reputation, which feeds back to resources simplistic, but hopefully illustrative

92
00:14:40.740 --> 00:14:45.480
James Howison: Because of software, there can be framed as a service separately paid for feeding into research.

93
00:14:47.400 --> 00:15:02.100
James Howison: We saw it framed as collaboration service work, which is interesting in these large physics collaborations that just thrown out this distinction between research and other things and said, hey, that's all research and so software work is done in there and people don't break it out.

94
00:15:03.120 --> 00:15:17.010
James Howison: We also found incidental software where people were creating software as part of their regular research and passing it, you know, possibly, keeping it themselves, but maybe also sharing it with the community.

95
00:15:18.690 --> 00:15:21.240
James Howison: But just as an incidental to their other work.

96
00:15:22.620 --> 00:15:30.210
James Howison: And we also saw software as a parallel software practice. Right. And this is where people are working on research, but also working on software.

97
00:15:32.340 --> 00:15:39.180
James Howison: But you can see what this one is a little more complicated. So the software gets produced in the course of the research.

98
00:15:39.810 --> 00:15:53.430
James Howison: And then they might produce a software publication, which might get you a little bit of reputation but

then in order to get the citations from that you have to go through software releases maintenance improvement software support.

99
00:15:55.530 --> 00:16:04.110
James Howison: Liberty hopefully use which will eventually maybe produce some software citations, but other work has shown that that's relatively rare or at least disorganized.

100
00:16:06.810 --> 00:16:16.680
James Howison: So the route to academic reputation virus software is long and complicated and involves tons of work that other types of academic reputation doesn't involve

101
00:16:18.510 --> 00:16:29.610
James Howison: In other work gym hopes up. And I looked at improvements and when they were shared. So we looked at the work of ncb I'd lost and a bunch of improvements that we could find in the community.

102
00:16:30.180 --> 00:16:37.560
James Howison: And we looked at improvements that branch from that code. But we're eventually returned and given back to in CBI to take forward.

103
00:16:38.640 --> 00:16:40.500
James Howison: We also found one ones that weren't

104
00:16:41.520 --> 00:16:51.690
James Howison: And that distinction here was all of the improvements that were primarily motivated by academic reputation branched but never returned to the main line of development.

105
00:16:52.560 --> 00:17:00.660
James Howison: All other ones motivated by anything else, which might have been money fun said sense of duty to the open source community those branch them returned

106
00:17:01.680 --> 00:17:05.370
James Howison: Which suggests that we have some some issues in

107
00:17:06.720 --> 00:17:09.720
James Howison: Reputation in scientific software as a motivator.

108
00:17:11.460 --> 00:17:22.440
James Howison: So with that, hopefully illustrative background, just some images of these different types of work right dive into the second question here, which is why is sustainability so challenging

109
00:17:24.000 --> 00:17:33.180
James Howison: And so this is based on these interviews from CSC W, together with lots of work and conversations with people in this in the science community over time. And thank you to everybody for those

110
00:17:34.950 --> 00:17:50.490
James Howison: So here you can see scientists using software within a workflow and of course workflow systems or something that size has contribute a lot of support to in the science community has produced over the years. So here we can see our components that move over time.

111
00:17:51.780 --> 00:18:02.010
James Howison: Well, look, I'm at. And you can see these are workflow components pushing data to each other that eventually produce plots and, of course, each of these build up on dependencies.

112
00:18:02.430 --> 00:18:13.740
James Howison: That build on other dependencies. And so we have entire stacks of software that are built up. And of course, this, this is also based on other work by by Judith Seagal wrote some

113
00:18:14.790 --> 00:18:18.870
James Howison: Excellent stuff on scientific software development over time, as did many others.

114
00:18:19.950 --> 00:18:28.230
James Howison: So I'm going to use a fancy word and call this a software assemblage. Sometimes you hear a call stack, you might think of it as code that's in a directory on someone's computer somewhere.

115
00:18:28.620 --> 00:18:41.040
James Howison: This might end up as a workflow in a formalized workflow environment like galaxy or on a supercomputer, but more often. It runs, you know, ENCODE on someone's computer themselves after importing all these components.

116
00:18:42.240 --> 00:18:51.120
James Howison: So what we saw with the interviews and scientists working at scientists and users, working with the code was that they're working to reanimate these assemblages

117
00:18:51.900 --> 00:19:03.930
James Howison: So this is code. Sitting on someone's someone's computer us to produce plots republication that is then pick it up. Maybe by students, maybe by postdoc maybe by the researcher herself.

118
00:19:04.620 --> 00:19:18.450
James Howison: reanimated in order to do work. Now this very specifically, and our influence laughed when we said, well, we're just running it for reproducibility. And I said, well, no know you've got to want to do new science. So we have to extend this software.

119
00:19:19.740 --> 00:19:25.710
James Howison: And so what we're seeing is after they quote, get the plots, they leave that code from months, maybe years

120

00:19:26.790 --> 00:19:40.770
James Howison: And so what that means is the cycles that are driving potential participation in science scientific software work are quite long much I think longer than working industry. Now, I don't have any firm empirical data on that.

121
00:19:42.570 --> 00:19:49.140
James Howison: And here's where I'll pause and say no. My job as an organization, scientists here is to is to give you images that help you think about

122
00:19:49.440 --> 00:19:59.640
James Howison: Recapitulate the great deal of experience that each of you has in this space. So my job will be done today, if I give you some new ways of thinking about the work that you do and you see other people doing

123
00:20:00.960 --> 00:20:10.860
James Howison: So this piece of work is potentially much slower than an industry where software is used, day in and day out. And even when infrastructure is running all the time. The users visited every now and again

124
00:20:11.910 --> 00:20:18.840
James Howison: And when they come back, they come back to extend these workflows to use the software assemblage for new purposes for new science.

125
00:20:19.800 --> 00:20:33.780
James Howison: So within that context, let's ask what is needed for sustainability and as fast. And because we know from working software engineering that maintenance can be up to nine nine tenths, that the amount of work that goes in to using a piece of software over time.

126
00:20:35.160 --> 00:20:48.000
James Howison: And researchers and CCW have really noted this so beats Matt beats and Charlotte least studied this, and essentially found that the work of the work in long term in a piece of cyber infrastructure look very much like the work in its development.

127
00:20:49.110 --> 00:20:53.820
James Howison: Well, what sorts what drives the need for work to keep software sustainable.

128
00:20:54.810 --> 00:21:06.330
James Howison: So I'm going to argue that there's five real sources here. One is the difficulty of software production. It's hard to write software, right, it's hard to write cutting edge scientific software. It's a difficult challenge.

129
00:21:07.500 --> 00:21:10.650
James Howison: It's also hard for people to pick up and use software.

130
00:21:11.520 --> 00:21:19.290
James Howison: And so that can produce work because users come back and say it's not working for my purposes, and it can take a lot of time and effort.

131
00:21:19.560 --> 00:21:28.020
James Howison: To explain how to use it, or even to adjust to it, and often the developers haven't actually perceived the exact circumstances, the users are going to be using it in

132
00:21:29.700 --> 00:21:32.220
James Howison: But those things are true of all software production. Right.

133
00:21:33.420 --> 00:21:44.400
James Howison: Now the third one is the rise of new possibilities. And so what we found in our interviews wasn't people using code to do new science, which means there's a changing scientific frontier.

134
00:21:45.570 --> 00:21:53.250
James Howison: I believe that the founder, the founder of the NSF has had something to say about that. But the you know the ever changing the endless frontier.

135
00:21:54.570 --> 00:22:10.050
James Howison: And so if you think about it, I would say that the scientific frontier changes it more continually then, for example, the accounting frontier. Right. So everything about accounting software things do change, but they don't change at nearly the pace that work does in science.

136
00:22:11.100 --> 00:22:22.290
James Howison: It's also true that work comes. So this is like new scientific ideas, new models. Somebody invented that everybody needs to use right now to be doing cutting edge science.

137
00:22:23.040 --> 00:22:28.830
James Howison: And also the work also comes from changing, changing technological capabilities, including hardware and software.

138
00:22:29.760 --> 00:22:39.600
James Howison: new hardware new soccer that other people have written, new opportunities and we can see that now with quantum we're still seeing it with the work you know in parallel and GPUs.

139
00:22:40.320 --> 00:22:52.080
James Howison: Also argument returns for the end of this talk, that it comes from ecosystem complexity and this is the dependencies that all the software is able to be built upon which offers new capabilities but also new challenges.

140
00:22:52.980 --> 00:23:04.890
James Howison: So all this work coming in that needs to be done to keep this sun software scientifically over time. What are we going to do so, essentially, we've got two strategies, we need to suppress try and suppress these drivers of new work.

141
00:23:06.060 --> 00:23:14.010

James Howison: Which is quite difficult. Imagine as a you know people concerned about software and science of it we suppress the ability to pursue the scientific in frontier.

142
00:23:14.910 --> 00:23:17.130
James Howison: We can try to increase the efficiency of work.

143
00:23:17.580 --> 00:23:26.160
James Howison: Right, so that it is even more easily done with automated test suites and things like that and and you can see a lot of that in software engineering inspired work. It's very important.

144
00:23:26.550 --> 00:23:32.400
James Howison: But ultimately work is always going to be needed. And so we have to attract the resources to do that work.

145
00:23:33.390 --> 00:23:45.300
James Howison: And so here you can see the characteristically simple model of how resources are attracted so resources turn into software which eventually to hopefully turns into use.

146
00:23:46.140 --> 00:23:54.060
James Howison: That use has some impact in the world and that impact then produces new resources to the people who need to do the work.

147
00:23:55.980 --> 00:23:58.380
James Howison: Let's think about. Sorry, I think I said

148
00:23:59.550 --> 00:24:07.980
James Howison: Yeah. So here I'm going to talk about different modes of production for software or how the projects attract resources in different areas.

149
00:24:08.550 --> 00:24:17.460
James Howison: So I'll talk about three ideal types. The first is a commercial project. The second is open source. Peer production and the third is scientific grant making

150
00:24:19.500 --> 00:24:22.500
James Howison: So in a commercial project we see

151
00:24:24.060 --> 00:24:32.010
James Howison: That the resources money turns into software which turns into use. And really, what that means is purchasing software.

152
00:24:33.030 --> 00:24:43.650
James Howison: And that use is happening in some value producing activity. Right. So it's being used by business either to have a competitive advantage or to reduce costs.

153
00:24:44.730 --> 00:24:52.530
James Howison: And that's the impact. Either way, that impact produces resources either its new profit or its cost the money left over after cost reduction.

154
00:24:53.100 --> 00:24:57.810
James Howison: Which produces dollars which drives the software work that needs to be done.

155
00:24:58.410 --> 00:25:05.730
James Howison: So there's a pretty obvious circle right so as the maintenance. The new development exploring new opportunities exploring new use cases.

156
00:25:06.180 --> 00:25:17.580
James Howison: Those arise as the developers know of the work through sales and interaction with people who've paid them simultaneously producing the read the work and the resources.

157
00:25:19.590 --> 00:25:23.250
James Howison: Now, I would argue, an open source peer production, the situation is not that different.

158
00:25:24.060 --> 00:25:31.050
James Howison: And if you think about this, an industry, you know, you think about when Apple contributes to open source or to pick a better example IBM

159
00:25:31.590 --> 00:25:48.150
James Howison: Or you know any of these large companies or startups, what's happening is they've got software they want to influence for some value producing activity, some way that they attract resources and that impact on their world produces develop a time which they push back.

160
00:25:50.430 --> 00:25:51.660
James Howison: Now in grant making

161
00:25:52.980 --> 00:25:58.530
James Howison: We have resources given to projects which produce software which hopefully gets used in science.

162
00:25:58.920 --> 00:26:13.890
James Howison: That impact is observable in some form. Maybe it's citations. Maybe it's letters of support and collaboration. Maybe it's accolades like prizes right we've been seeing software mentioned in Nobel Prizes that's exciting because those then feed through grant peer review.

163
00:26:14.940 --> 00:26:16.710
James Howison: In order to produce resources.

164

00:26:17.790 --> 00:26:22.620
James Howison: So quite a different system. And notice also the people using it are funded by the same

165
00:26:23.670 --> 00:26:32.400
James Howison: pot of money. Typically, as would need to go to the software projects. Okay, so let's summarize. You know why sustainability is difficult.

166
00:26:32.790 --> 00:26:46.740
James Howison: First is the indirect resource attraction reputation is a great motivator, but also problematic. They're incompatible incentives for needed work right so citations literature impacts are not directly linked

167
00:26:47.760 --> 00:26:53.910
James Howison: And so what that means is that researchers have a lot of trouble, knowing exactly how their code is being used, whereas

168
00:26:55.830 --> 00:27:02.640
James Howison: People selling code in business know their customers and scientists often don't often are surprised by their use.

169
00:27:03.930 --> 00:27:19.590
James Howison: And grants also produce what one of our informants called a service center idea where I have been given the money by the community and I must do the work will come back to that. But I don't want to have too much on reputation because there's also the availability

170
00:27:20.970 --> 00:27:26.490
James Howison: Of skilled labor, as I said, this software working science is hard, it's cutting edge often

171
00:27:27.210 --> 00:27:38.340
James Howison: And so the pipelines for that labor, the programming language. The mental models. Those are harder to find than perhaps you know people doing work in some parts of industry.

172
00:27:38.940 --> 00:27:57.480
James Howison: And it's also true that the open source projects in industry tend to be once that for software that has already been demonstrated as needed and possible, right. So I think it's interesting thing about the software that's really successful in industry is typically lagging innovation.

173
00:27:59.280 --> 00:28:04.080
James Howison: Where's the software we really want to fund and make sustainable and sciences driving innovation.

174
00:28:04.980 --> 00:28:14.130
James Howison: The other thing is if we think back to what I call productive deferral in the open source community of waiting until a project is easy enough to do science will not wait

175

00:28:14.790 --> 00:28:23.910
James Howison: Right, if you, if you, as a project, don't do it one of your end users will found a new project and drive that scientific possibility that they perceive and drive it forward.

176
00:28:25.020 --> 00:28:36.180
James Howison: We also have these different time scales of use. And so when people are inspired to work on something, it might be a year before their code is available in which time the underlying code may have changed and made things difficult

177
00:28:37.980 --> 00:28:43.890
James Howison: Okay, so if that's what makes sustainable little difficult. What are we going to do about it.

178
00:28:45.330 --> 00:28:51.180
James Howison: And one possibility has been to be inspired by by open source peer production and there's a lot of connections here.

179
00:28:51.870 --> 00:28:59.550
James Howison: In some ways, and some authors have even written about open source as academic production right open source was inspired by academia. Well, now

180
00:29:00.030 --> 00:29:09.480
James Howison: Scientific software has been inspired by by peer production. And the idea here is, is a peer production of open sources, a route to sustainability in scientific software.

181
00:29:12.630 --> 00:29:17.460
James Howison: And the model, most often discussed as well. I'll eventually called reorganization.

182
00:29:18.180 --> 00:29:36.180
James Howison: Which is we fund some software and over time that project turns itself into an open source community by building software, people want to use bringing those people into the community and turning them into contributors harnessing their insight on the scientific frontier.

183
00:29:37.980 --> 00:29:39.690
James Howison: But there are other possibilities. That's we'll see

184
00:29:40.860 --> 00:29:41.790
James Howison: I'll pause here for a second.

185
00:29:42.960 --> 00:29:50.580
James Howison: Here we're actually talking about an organization science and science and technology studies is a literature on a transpose to open communities.

186
00:29:50.970 --> 00:30:03.270

James Howison: And we have much literature and plans and even funding programs on transfer to industry right so it such that if you go to a technology transfer office and you say, I want to open source my code. They say, I don't know who deals with that. It's not us.

187
00:30:04.830 --> 00:30:11.820
James Howison: Small, small caveat to that I believe at Johns Hopkins, they have tech transfer office that will help people create open communities.

188
00:30:13.320 --> 00:30:20.310
James Howison: So, you know, the big plan. Here's to develop a literal. What does it mean to go open in science. And so what we've been doing in work with Hannah Cahoon and

189
00:30:20.640 --> 00:30:30.360
James Howison: Stefan do in my NSF career funded project is looking at a number of case studies of successful transitions to peer production and then a panel study of projects.

190
00:30:31.710 --> 00:30:34.050
James Howison: Trying to move to peer production perhaps

191
00:30:35.340 --> 00:30:44.040
James Howison: So here I'll just talk about one of the case studies briefly so Enzo is software involved in simulations of galaxies. And so we've done interviews with

192
00:30:44.490 --> 00:30:53.160
James Howison: The producers of enzymes and enzymes, one of these projects that has been has been around for a long time and a model that we're starting to call the long center. Okay.

193
00:30:53.640 --> 00:31:04.620
James Howison: So Enzo starts out in the early 90s. It's a, you know, scientists wanting to learn and see starting to mess around create some code not sharing it using it locally.

194
00:31:05.250 --> 00:31:15.900
James Howison: And over time, scrounging together resources moves forward. So what we've done is a bunch of interviews observations that related workshops examination of websites publications

195
00:31:16.230 --> 00:31:29.490
James Howison: And then done qualitative analysis to create a timeline identifying episodes that lead to phase changes in the end of project over time. And what is eventually a successful peer production communities for scientific software code.

196
00:31:31.470 --> 00:31:45.840
James Howison: So, as described to us performance. The, the, quote, early days of the end zone project where a project in the lab. A few people in a lab together writing code, no central base no central code base to which changes are coming and going.

197
00:31:46.710 --> 00:31:51.810
James Howison: And copying and pasting features across personal branches, although I'm not even sure the

198
00:31:52.350 --> 00:32:01.110
James Howison: Performance use the word branch, they would say directories right so directories on a shared code that's where the code was living but all the work in occurring inside a single lab.

199
00:32:01.680 --> 00:32:13.890
James Howison: Through to a brand around the year 2000 where a grant and some additional you know scrounged up money and I are influenced that the robber fictional amount of 20 hours a week for graduate students.

200
00:32:14.880 --> 00:32:21.840
James Howison: Applying the graduate students work much more leading to a period co NZ 1.0 which are informative described as service center.

201
00:32:22.440 --> 00:32:26.130
James Howison: And here's a, here's a graph that helped enjoy push this forward. They hired

202
00:32:26.610 --> 00:32:36.330
James Howison: A software manager who set up version control and now you can see this central branch, what changes are coming to and from these diagrams are illustrative that not based on tested.

203
00:32:37.230 --> 00:32:49.200
James Howison: So a mainline branch internally and then release that branch through what we're calling a one way membrane. So you know that code was available, but there was no expectation of changes coming back.

204
00:32:50.640 --> 00:33:01.800
James Howison: But they did get insights into use with reports coming back from what they called the friendly users, which were other labs, some of which were postdocs from the original lab, but others were not

205
00:33:04.200 --> 00:33:04.560
James Howison: And

206
00:33:06.150 --> 00:33:20.700
James Howison: At this time, the knowledge of who was using the code. And particularly, who was developing the code was not super well spread. And so what happens. Eventually, is one of the descendants of the original lab ends up with a new academic appointment.

207
00:33:22.470 --> 00:33:29.850
James Howison: And it's a it's at Stanford and it comes with startup funds, right. So, a new source of resources beyond grants different types of usage.

208
00:33:30.390 --> 00:33:37.980
James Howison: And this wasn't had learned to various new versions floating out there, folks, if you like, through conversations at conferences and even reviewing papers.

209
00:33:38.670 --> 00:33:52.260
James Howison: Right, so that knowledge of what's happening out there. It's not coming through a sales, marketing, sales follow up team. It's not coming through bug reports and issued issue tracker is the path through this one way membrane. It's coming through scholarly communication. Right.

210
00:33:54.600 --> 00:34:09.900
James Howison: And so what they do is actually create what they call the week of code. And this was to fly use startup funds to fly in these people who had forks and to try to get them working together. And one thing that really struck me about this was they began with the code of the people not present.

211
00:34:12.360 --> 00:34:26.460
James Howison: It would have been so tempting to start with the with the old code, code. What they said was this code represents work. Let's start with the people who aren't here to talk and figure out if we can merge our code in around the edges so that we end up with

212
00:34:27.600 --> 00:34:36.720
James Howison: Something, it looks more like this where work is still happening in the labs, things are coming into a project core now so dissolving that one way membrane.

213
00:34:37.950 --> 00:34:44.850
James Howison: But this was a period they called the Wild West because he had the central branch to which people contribute. Now, this starts to look like peer production. Right.

214
00:34:46.290 --> 00:34:56.940
James Howison: With contributions coming you notes and stuff, but the developments still happening in it in separately in the external labs, which is not uncommon. Right. It's still taking this relatively long time frame.

215
00:34:57.480 --> 00:35:01.980
James Howison: And they talk about, you know, big code dumps would come in and we'd have to, like, you know, work with those

216
00:35:03.180 --> 00:35:12.990
James Howison: And one of the real challenges they brought up was autonomy concerns right where people were working on code, but I didn't know whether I could rely on them at my block my work.

217
00:35:13.680 --> 00:35:19.500
James Howison: And I actually think that people talk about ego in science, but this came up more in our analysis, which was

218

00:35:20.070 --> 00:35:28.830
James Howison: Autonomy concerns. Right. I want to be able to drive forward the scientific frontiers there. I want to reach for it. I'm reaching for it, but I can't. If I have to rely on somebody else.

219
00:35:30.000 --> 00:35:34.560
James Howison: And so there's a little bit of concern that that was happening here. As you know, elements were added.

220
00:35:35.010 --> 00:35:47.670
James Howison: And so eventually even though this already starts to look like quite a mature open source project they eventually develop a system called the curated branch with review gates were contributions which you'll notice now are becoming smaller

221
00:35:49.050 --> 00:35:49.290
James Howison: Than

222
00:35:50.700 --> 00:36:00.420
James Howison: Know, they've got it all right that you can see those contributions becoming smaller branching off and coming back branching off and coming back more often, but are also being checked through code review.

223
00:36:01.380 --> 00:36:09.960
James Howison: Okay so much what changed him. Okay, what didn't change is the motivations, the code is still a scientist and side effect of scientific inquiry, there is still channel.

224
00:36:10.710 --> 00:36:19.140
James Howison: But there are challenges coordination, after long periods of working apart. The, the leadership had to change. And as we, you know, move to the open period.

225
00:36:19.620 --> 00:36:27.960
James Howison: You know, we heard stories of the feelings of responsibility and emotional connection to the work as well as a fear of missing out of the discovery and the realization

226
00:36:28.320 --> 00:36:40.800
James Howison: Of possibilities and the real difficulty of passing on that leadership, with the emphasis there on responsibility and stewardship. Right. That was really hard, but they managed it and then deep concerns about giving up autonomy.

227
00:36:41.880 --> 00:36:45.600
James Howison: But they also didn't decide any club any governance.

228
00:36:46.770 --> 00:36:55.290
James Howison: At the week of coned they intentionally just focused on getting the collaboration tech together, which is somewhat contrary to the org science work on on good collaborations.

229

00:36:56.730 --> 00:37:09.420
James Howison: Okay. They were inspired, but they were working alongside each other and that leads us to a panel study of s have funded projects. The SI to program funded by the NSF NSF size.

230
00:37:10.230 --> 00:37:19.590
James Howison: And here we're examining other projects that were specifically asked to be sustainable up you know as part of their grant. So there were asteroid sustainability plans.

231
00:37:19.920 --> 00:37:27.330
James Howison: I don't know what all of them say, not all planned peer production or open sourcing as it was more cold, but that was a frequent one. Okay.

232
00:37:27.720 --> 00:37:34.290
James Howison: So what we're going to report on today is a study of 92 of these grants that had started by 2014

233
00:37:35.100 --> 00:37:43.470
James Howison: That ended up creating software so that reduced it to 92 grants. I'm sorry. So these are all the grants funded by her side to that started before 2014

234
00:37:44.070 --> 00:37:47.220
James Howison: So we so we could watch the whole period of the grant on a little after

235
00:37:48.000 --> 00:37:56.700
James Howison: An 84 were judged to intend to create software, right. So as I to contributed to workshops and other things like that. But the bulk of them intended to create software.

236
00:37:57.690 --> 00:38:07.080
James Howison: And these ad for grants funded work on 114 code bases. Right. And so we analyze each code base separately or at least the activity around them, which we call projects.

237
00:38:07.710 --> 00:38:15.900
James Howison: And again, great, thanks to our Hannah Kahuna did the vast bulk of this work, and more recently siphon do has joined us as we move to the interview stage.

238
00:38:16.710 --> 00:38:21.960
James Howison: So what we're doing is we've done content analysis over time, over over these five year period, looking at these projects.

239
00:38:22.320 --> 00:38:28.530
James Howison: We've started interviews with some and now we're working on more interviews, but I can show you the emerging results.

240
00:38:29.220 --> 00:38:34.800

James Howison: Because these are the sorts of things we were looking for. Here we're asking these projects, when they present to the world.

241
00:38:35.760 --> 00:38:45.930
James Howison: What can we learn right and we adopted a persona of a potential contributor right somebody who might want to have an idea of how to work with these with this code and more to contribute back

242
00:38:46.620 --> 00:38:56.280
James Howison: And we couldn't for things like does the project presents separately from the grant. What does it present, we have a grant and writing code, or is it we are a project and we received a grant

243
00:38:57.210 --> 00:39:02.040
James Howison: We've asked about things like in invitations to contribute, whether they highlight publications

244
00:39:02.790 --> 00:39:18.570
James Howison: Whether they have repositories. What's the collaborative setup. Do they have offline meetings online meetings and these codes emerged over time. And what I'm going to present today is a taxonomy of project types that were funded by the SI to grant proposal. So the first is peer production.

245
00:39:19.590 --> 00:39:28.110
James Howison: Second is what we call a tool group. The third is what we call an author group and then a lab and the fifth is a business.

246
00:39:30.120 --> 00:39:35.820
James Howison: And so here you can see characteristics of different project types. So vertically from the top to bottom of the different

247
00:39:37.200 --> 00:39:44.640
James Howison: Organizational configurations production lab to group. So on moving across are the characteristics that we looked at. So you can see

248
00:39:45.300 --> 00:39:54.180
James Howison: What we're displaying here is whether the group was named and surprisingly only four of these types of names off the groups don't talk about that in a second. A web presence.

249
00:39:54.810 --> 00:40:04.410
James Howison: Most of them have web presences with the group start. You can see all the groups being great. There are some of the, you know, at least organized least organization like entities that were funded

250
00:40:05.490 --> 00:40:07.980
James Howison: You see whether they have similar interests or not.

251
00:40:09.030 --> 00:40:16.260

James Howison: You can see whether they accept it outside contributions, but we could find evidence of that and whether they displayed an invitation to contribute on their website.

252
00:40:17.370 --> 00:40:21.510
James Howison: Can also see whether they featured publications and whether they discuss revenue.

253
00:40:23.010 --> 00:40:38.460
James Howison: So what we ended up doing here. Over time was doing what we call a genre analysis of online presences right so you think genres like horror movies mystery movies. And what I'm going to show now is is illustrative mock ups of the sorts of web presences that we found.

254
00:40:39.480 --> 00:40:49.920
James Howison: So this is a website of mock up of a peer production community. So you can see it's named after itself, it identifies as an open source community has a website as website. Kind of looks like this. Right.

255
00:40:50.820 --> 00:41:00.900
James Howison: There, you don't see individual features, you don't see very rarely see publications and almost never on the front page and you see a really welcoming and open invitation to contribute.

256
00:41:01.620 --> 00:41:08.910
James Howison: Incidentally, we expected to find differences in collaborative infrastructures, but we didn't ever on us as getting get up now and repositories.

257
00:41:11.340 --> 00:41:19.560
James Howison: And hopefully you're going to recognize some of these. So a tool group looks a little different. And one of the features of tool group is this NSF logo at the bottom right where they

258
00:41:20.220 --> 00:41:30.180
James Howison: They really identify as the grant is funding this software. And this is a communication about it with named we have a website with singular interest with just pursuing

259
00:41:30.630 --> 00:41:39.420
James Howison: You know this project on this website. Very rarely outside contributions or invitations to contributes mostly pop more often publications

260
00:41:41.760 --> 00:41:47.550
James Howison: All the groups, the least organized in here you can see what I'm displaying here is not a website, but actually a poster.

261
00:41:48.000 --> 00:41:52.170
James Howison: And so we most often found the presence of what we ended up calling author groups as

262
00:41:53.100 --> 00:42:05.730

James Howison: Si to PDFs, which were posters from the SI P eyes meeting. Right. That was, you know, the only web presence that we could find for some projects. And so these groups are rarely named that lists of authors write

263
00:42:07.080 --> 00:42:10.800
James Howison: All those postdocs grad students they very rarely have a website.

264
00:42:11.850 --> 00:42:17.490
James Howison: That group of authors that was presenting this way has a singular interest, they're going to you know they're doing this grant

265
00:42:19.110 --> 00:42:26.940
James Howison: They don't have outside contributors that are making notations to contribute. They always list publications, right, this is very much an academic genre.

266
00:42:28.530 --> 00:42:33.600
James Howison: Oh, these will often talk about an intent to open source right like towards the end.

267
00:42:35.700 --> 00:42:47.460
James Howison: You know, the, the fourth type here is a lab. And this is a really well established organizational form and scientific in Science Technology Studies and Steve shaping has written of lab life and lots of people have written about life in a lab.

268
00:42:48.000 --> 00:42:55.350
James Howison: And here we found projects where you know this is a lab typically named after the P eyes and be, you know, the house and lab or

269
00:42:56.430 --> 00:42:57.510
James Howison: You know the parish our lab.

270
00:42:58.830 --> 00:43:08.910
James Howison: And here we see, you know, the PR featured but also their team like people are there and they have a website, interesting enough labs have diverse interests they pursue multiple things over long periods of time.

271
00:43:09.390 --> 00:43:16.350
James Howison: And they're not just about a piece of software and you might actually see software as a sub feature on their website right

272
00:43:17.250 --> 00:43:31.500
James Howison: They might actually invite contributions, but in a very specific way they'll often say join my group, right, which really could to a potential outside contributor means like become an academic and write a grant with me, or at least become a postdoc and write, write code.

273
00:43:32.730 --> 00:43:40.020

James Howison: So it's almost as though a contribution would come by joining the organization through employment, of course, they always feature publications

274
00:43:40.500 --> 00:43:43.200
James Howison: And not ribbon and they but they don't discuss revenue.

275
00:43:43.860 --> 00:43:49.050
James Howison: Will come back to a little bit more about that. And of course, the, the fifth type here was businesses, we didn't find many of these. This will say,

276
00:43:49.230 --> 00:43:55.830
James Howison: But here, of course it's named for the website business talks about singular interests presents this piece of software is its product.

277
00:43:56.430 --> 00:44:06.690
James Howison: There's almost at least in these ones, they weren't invites to contribute, although that's obviously a possibility. I mean, that's not very rarely publications sometimes have white paper, which is a genre of

278
00:44:07.770 --> 00:44:16.350
James Howison: People you know seeking academic credibility for their new business and at least some discussion of revenue or at least plans of things you could purchase.

279
00:44:19.560 --> 00:44:23.580
James Howison: Okay, so now we're going to ask, will give him those like project taxonomy.

280
00:44:24.000 --> 00:44:30.900
James Howison: And we'll come back to a second because each of those has different challenges in transitioning to pay production. Let's examine how they changed over time.

281
00:44:31.440 --> 00:44:40.080
James Howison: And so what our research team has done is look at an allocate these project types to the projects at the start of their grant. And at the end of the grant.

282
00:44:40.800 --> 00:44:46.950
James Howison: And so here you can see the sort of relative size of the different taxonomy. So you can see lab and tool group with the two largest

283
00:44:47.220 --> 00:44:54.510
James Howison: Kind of makes sense in some ways an NSF funding program is designed to reach out to those communities and sponsor the workforce development through them.

284
00:44:55.080 --> 00:45:03.030
James Howison: But also some, some of these grants did go to existing peer production communities someone to author

groups. And as you said just a few into businesses.

285
00:45:03.870 --> 00:45:12.630
James Howison: Now, by the end of the grant. There was a little bit of change, although I want to highlight that it was mostly continuity and you can see those lines moving across the

286
00:45:13.410 --> 00:45:23.880
James Howison: Diagram, you know, very clearly, but the changes are there, right. So we can see, you know, a couple of P one P production project turned into a business, we can see that some

287
00:45:24.480 --> 00:45:27.090
James Howison: Two groups became peer production and so on so forth.

288
00:45:27.960 --> 00:45:38.310
James Howison: What we did next within wait even longer and go back and say, Are these projects active right. Does it seem that they are still producing and maintaining software that is scientifically useful.

289
00:45:39.300 --> 00:45:44.310
James Howison: And so that adds a third time period here where we can see we add a new category of became inactive.

290
00:45:47.700 --> 00:45:52.860
James Howison: So, I believe, the first thing that strikes you there. Is it it's different ratios in different organizational times

291
00:45:54.090 --> 00:46:06.540
James Howison: At the author groups largely became an active the labs and labs were kind of 5050 interestingly enough tool groups continued, but look at the top production largely continued

292
00:46:07.920 --> 00:46:20.940
James Howison: There is one that becomes inactive. So sort of examine that. Sorry. As a zoom in and a close up. And of course, the X x Y axis scale is changing here. But now we're just looking at projects became an active can see those different ratios.

293
00:46:24.300 --> 00:46:35.520
James Howison: These are projects that you know ended as labs. Right. So some tool groups actually sort of lost their additional contributors over time and became labs, then of course some labs end up becoming inactive.

294
00:46:36.420 --> 00:46:43.860
James Howison: And what we're doing as a project now is doing interviews on examples of, you know, that we think the represent these environments more

295
00:46:44.340 --> 00:46:51.480
James Howison: But I want to focus in on this one today for peer production. Now, of course, not all of these projects

with trying to accomplish. Peer production, but many work.

296
00:46:52.860 --> 00:46:59.340
James Howison: And these are the ones that end in peer production and as we can see, most of them continue to be active is one that doesn't

297
00:47:00.780 --> 00:47:12.090
James Howison: But the other thing that is key to notice here, and this is a key finding of the study so fun. It's not that surprising when you think about it, but the best way to accomplish peer production was to start as a production.

298
00:47:13.800 --> 00:47:23.790
James Howison: So those ones at the top all starters peer production receive a grant in some form, right, that grand contributes to their ongoing work and they continue to produce

299
00:47:24.450 --> 00:47:36.120
James Howison: On the other hand, we do see some changes. Right. Not that many. But we do see some changes. So the bottom. So if I zoom into those changes. We can see the projects that changed to become peer production.

300
00:47:37.020 --> 00:47:43.140
James Howison: And honestly, from an old science perspective, this is remarkable right changing your mode of production is isn't really huge right so

301
00:47:43.350 --> 00:47:49.200
James Howison: You've got to change your governance structure your collaboration structure. You've got to become open all these things really hot.

302
00:47:49.950 --> 00:47:58.920
James Howison: And what we see is that there's two ways these projects became peer production, most of them became peer production through what we're calling a handoff transition

303
00:47:59.790 --> 00:48:12.510
James Howison: And this is when the grant funded project hands their work off in some form to appear production community that already existed and and takes that code on and drives it forward super interesting.

304
00:48:13.740 --> 00:48:21.090
James Howison: What we're calling reorganization, which is perhaps what most people think of when they say open sourcing doesn't happen that often.

305
00:48:21.840 --> 00:48:34.110
James Howison: And so here we can see you know that that a few thing a few tool groups, right, which were sort of specifically funded for a piece of software did in fact reorganize themselves.

306
00:48:34.410 --> 00:48:43.830

James Howison: In relatively short period of time. Because remember, for example, the long center idea of Enzo with decades of work and then a grant that facilitated that reorganization.

307
00:48:45.720 --> 00:48:54.480
James Howison: And we do see that in some of these as well. So what we're doing at the moment is doing interviews diving more into these I can give you a little bit of a taste here so

308
00:48:56.160 --> 00:49:06.750
James Howison: So, you know, one of the handoffs, for example, we're all members of a professional organization and that professional organization had an open source project so their grant funded work that

309
00:49:07.800 --> 00:49:17.340
James Howison: Created code that eventually ended up in that open source project, but they didn't do the groundwork in the open source project, but they had connections with it in advance.

310
00:49:17.700 --> 00:49:25.080
James Howison: It took a long time it took, you know, a year and a half of grant time to get a pull request accepted at the destination project.

311
00:49:26.790 --> 00:49:30.540
James Howison: We have another example where a handoff occurred to a company that was running

312
00:49:31.080 --> 00:49:36.390
James Howison: Software and what we think happened and we're pursuing this further with interviews was a postdoc went to work for that company.

313
00:49:36.720 --> 00:49:48.750
James Howison: But then the company had experience in running peer production communities took that code wrapped it into their production community and you know now that community has dozens of participants in hundreds of contributions.

314
00:49:50.310 --> 00:50:03.900
James Howison: With the reorganization one. It was a project sponsored by national lab that actually had a long history and use the grant to fund the really difficult and distinct work of identifying a potential community of

315
00:50:04.560 --> 00:50:10.170
James Howison: Contributors reaching out to them and building, you know, building sustainability that way.

316
00:50:11.310 --> 00:50:20.010
James Howison: Okay, so what do we see here we've got the old changes a project in and of itself, it's relatively rare. We saw a little difference in collaboration tech use

317
00:50:20.790 --> 00:50:26.340

James Howison: We saw that hierarchy is something that has to change to move to peer production and that can be difficult.

318
00:50:26.820 --> 00:50:38.160
James Howison: But we also saw limited ability to diversify resources and a lot of that's hard to tell him awfully right which a fancy word to say. Birds of a feather flock together. And one of the things we've seen is that

319
00:50:39.300 --> 00:50:45.750
James Howison: Even in projects moving to peer production. A lot of it comes through descendants of the lab postdocs their students.

320
00:50:46.530 --> 00:50:55.860
James Howison: You know, going forward, which is great for trust building, but because those groups are still writing grants to the same agencies, they really bring in new sources of money. Right.

321
00:50:56.970 --> 00:51:05.460
James Howison: In some cases, transitioning to projects co sponsored by industry seems to help there. And in part because it's bringing in new sources of resources.

322
00:51:07.860 --> 00:51:17.790
James Howison: We also want to talk, you know, in the merging interviews people talking about how hard it is to hand off code to an existing peer production Community Code is not a gift.

323
00:51:18.660 --> 00:51:25.230
James Howison: A research paper in some ways as a gift to the world, you put it out there that sits there and you know it's archived somebody else looks after it.

324
00:51:25.830 --> 00:51:35.670
James Howison: Code is in fact a set of liabilities and responsibilities. So when a project accepts code. They're saying, we will maintain this for the future.

325
00:51:36.150 --> 00:51:51.600
James Howison: And that's a big request of somebody. So, unsurprisingly, what we're seeing is the projects that were able to transition to these communities already had connections to them. They were either directly funding peer production communities which managed to write code and continue

326
00:51:53.520 --> 00:52:04.620
James Howison: Or they were projects that people already had credibility in. And interestingly enough, and this will come in my recommendations at the end. That is something we can look for in grant proposals.

327
00:52:06.210 --> 00:52:16.980
James Howison: If we really care about sustainability in an area right if we want to fund things that are more than proofs of concept where proof of concepts are important. Don't get me wrong, and

328
00:52:17.490 --> 00:52:23.370
James Howison: Agencies want to fund discovery and that makes sense. But if sustainability is the answer. And we think production matters.

329
00:52:24.420 --> 00:52:27.000
James Howison: Then we can do things slightly differently.

330
00:52:28.050 --> 00:52:30.360
James Howison: I'll come back to that as well. So I want to fly here.

331
00:52:30.870 --> 00:52:41.250
James Howison: If you're really interested in what it takes to move something from a personal results to a community resource also take a look at this work from Jim hope slopes group with our trainer and others at CMU

332
00:52:41.880 --> 00:52:48.570
James Howison: Really detailed look at task and balance of responsibilities between those sharing a new bit of results and those benefiting from it.

333
00:52:49.740 --> 00:52:50.100
James Howison: Okay.

334
00:52:52.800 --> 00:53:01.020
James Howison: So now I want to talk. Finally, you know, towards the end of this talk here about ecosystem complexity and how this makes it even harder.

335
00:53:03.660 --> 00:53:08.190
James Howison: So what we talked about moment as individual projects, taking grant money.

336
00:53:09.390 --> 00:53:19.710
James Howison: Doing software work and seeing if they can be sustainable over time and also seeing if they can create peer production communities which largely means being open to outside contributions.

337
00:53:20.490 --> 00:53:33.030
James Howison: And that's important, and it's a good stop, but I wanted to return to another key cause of work which is ecosystem complexity because peer production doesn't all its own. You know, fundamentally, deal with this.

338
00:53:34.500 --> 00:53:43.320
James Howison: So in the time people are getting their plots doing their scientific work the world changes. And so the reanimation of code.

339
00:53:44.010 --> 00:53:54.660
James Howison: Encounters change updated packages new packages new possibilities. Now happens both in the

workflow. The data sources scientific frontier and also below in each of the dependencies.

340
00:53:55.620 --> 00:54:10.470
James Howison: And so each of those dependencies can become a source of new work for the code that is assembled around it. And so we people probably know the phrase breaking change right if we make a breaking change were acknowledging that our users will have to do some work to deal with it.

341
00:54:12.210 --> 00:54:21.870
James Howison: And impart this cascading complexity potentially cascading complexity. So one of the reasons that scholars have found that the work of maintenance looks a lot like the work of development.

342
00:54:22.140 --> 00:54:30.990
James Howison: Because you're constantly responding to new work. In fact, I think we should dump the word maintenance because it implies small amounts of work, but it's not. It's

343
00:54:32.040 --> 00:54:43.080
James Howison: Software as a continued activity. So the image I want to present here is the idea of these of rocks and drops falling into a pool with intersecting circles, creating incredibly complex patterns.

344
00:54:43.800 --> 00:54:51.510
James Howison: And those patterns are producing work. And if they're producing work, then they need resources available to do that work.

345
00:54:52.860 --> 00:54:59.490
James Howison: That's part of the definition of sustainability code is not so much an artifact, as it is an activity.

346
00:55:01.410 --> 00:55:08.130
James Howison: And so in emerging work from interviews that I started with Jim hopes live I'm following up with my group at at UT Austin.

347
00:55:10.110 --> 00:55:19.200
James Howison: And also with Matt German pray and Sean gardens, we're starting to look at what work holds a software ecosystem together. If anything, because it's incredibly hot

348
00:55:20.250 --> 00:55:28.830
James Howison: We're starting to categorize break up this maintenance category and talk about sensing work which is knowing what's happening, what's changing what's producing work out there.

349
00:55:29.760 --> 00:55:37.170
James Howison: The adjustment, which is actually doing that work and synchronization, which is making sure that work gets to the right spot.

350
00:55:38.370 --> 00:55:42.390

James Howison: So those so that the adjustment can scale out to a community.

351
00:55:44.130 --> 00:55:54.930
James Howison: So I want to set this up a little bit more because ecosystem position depends on two aspects. And this is, you know, just conceptual work that that I'm working on.

352
00:55:55.920 --> 00:56:10.740
James Howison: And so on the vertical axis there you can see the number of users, right. So this is the reuse of ordinance code can spread out of the Community take new opportunities and people are using it in the same way. So as you push up the y axis people using the code in the same way.

353
00:56:12.300 --> 00:56:14.670
James Howison: As you push across the x axis, though.

354
00:56:16.560 --> 00:56:28.500
James Howison: That's a change in diversity of use and what I meant. Don't mean there is different demographic groups using code, although that would be important to know about. Unfortunately, that's not something that I am researching. I hope people do.

355
00:56:29.040 --> 00:56:41.250
James Howison: But here I'm talking about recombination. Right. And so this is code being used in different ways. So perhaps the code has been used to do visualizations in one field and gets adopted in another field to do those visualizations.

356
00:56:42.390 --> 00:56:44.280
James Howison: And so that diversity of US changes.

357
00:56:46.590 --> 00:56:51.030
James Howison: And different types of needed work scales differently with these ecosystem positions.

358
00:56:51.630 --> 00:57:02.760
James Howison: So new feature development scales linearly linearly with more us and more diverse years and linear is still hot right as the use becomes more diverse, it becomes more difficult to develop those new features.

359
00:57:03.990 --> 00:57:11.340
James Howison: User support increases linearly and maybe it is even sub linearly. If we can get users starting to help each other.

360
00:57:11.940 --> 00:57:21.000
James Howison: And if someone wants to know what motivates us is to help each other. Read Laconia von Hippel 2003 it's about the early days of the Apache mailing list. Fantastic. By the way, the answer is learning

361
00:57:21.450 --> 00:57:35.700

James Howison: People learn by helping each other I'm sensing work knowing what people are doing out there scales linearly with diversity of use. So as diversity of use goes up scales linearly, but it's actually flat with use. Right. You know how people are using it.

362
00:57:36.570 --> 00:57:46.530
James Howison: But adjustments and synchronization. Now that scales exponentially with the diversity of us because that's more and more different components, each of which is a source of change over time.

363
00:57:48.000 --> 00:57:53.940
James Howison: That's really hard. So different production systems generate you know this work differently.

364
00:57:55.680 --> 00:58:02.310
James Howison: Sorry handle this work differently so markets provide insight through sales and sensing. At the same time, their general resources.

365
00:58:03.540 --> 00:58:16.740
James Howison: Now markets handle the explosive ecosystem complexity work by enforcing rules on platforms. So I always laugh when people say, hey, we want to build a science platform, just like iOS, and I think. Yeah. Are you going to

366
00:58:17.340 --> 00:58:26.070
James Howison: curate the App Store and in force use power to enforce rules on your users because that's anathema to the scientific

367
00:58:26.490 --> 00:58:39.000
James Howison: Scientific values, not saying it's wrong. And if a community decides to go that direction, more power to them, but the way markets control this explosive complexity is through an exercise of power to reduce recombination.

368
00:58:40.500 --> 00:58:47.070
James Howison: Open Source peer production handles this through being open to ideas from the outside, which do sensing an adjustment.

369
00:58:47.820 --> 00:59:01.320
James Howison: And the emergence of distributions. And one of the things that distributions do is push work upstream and that means taking an adjustment at the edge and finding the best place in the hierarchy for that work to benefit everybody around them.

370
00:59:03.360 --> 00:59:10.950
James Howison: Now grant funding has little here, there is this amazing period of insight in the proposal where people make their argument.

371
00:59:11.370 --> 00:59:21.690
James Howison: But there's very little systematic process, which gives insight, one would think publications, would that they lag and software use in publications is not as visible as it could be.

372
00:59:23.010 --> 00:59:36.540
James Howison: And so we end up with this situation where this really big challenge of going open is already there. But this really big challenge of managing ecosystem complexity is very hard as well. And so it's hard to track use in scientific workflows.

373
00:59:37.740 --> 00:59:39.390
James Howison: And that makes it particularly difficult

374
00:59:41.550 --> 00:59:47.130
James Howison: Okay. Two more minutes. And then I'm done, I'm dying to hear reactions and questions.

375
00:59:48.900 --> 01:00:02.430
James Howison: So also want to emphasize that different organizational forms that you might recognize spread out this ecosystem management work differently. So we'll talk about a few very quickly. So a grant startup, you know, you give a ground to a new piece of software.

376
01:00:03.540 --> 01:00:12.000
James Howison: They don't. They're like a stealth startup. They don't want to release to the world until it works because that will undermine their users. And so it's very hard for them to get that ongoing inside

377
01:00:12.960 --> 01:00:22.080
James Howison: The service center which is a grant funded to manage code for a community says we're funded to do the work for our users. We've taken a slice of science funding and we must serve them.

378
01:00:22.860 --> 01:00:30.420
James Howison: But then you end up with this one way membrane with were inside like requires whole user workshops not continual interactions.

379
01:00:31.770 --> 01:00:36.270
James Howison: And then we can have an open project which, in a way, is a massive achievement. Right.

380
01:00:36.990 --> 01:00:45.840
James Howison: Now we can have Emilio from project which has all the trappings of openness, but no outside contributions which means they're not really getting that sensing an adjustment work.

381
01:00:46.650 --> 01:00:50.040
James Howison: But at least others can sense you and your changes in that matters.

382
01:00:50.760 --> 01:00:59.190
James Howison: We have a passively open project. And this one is sort of what I talked about in the transitions. This is where projects have accomplished these transitions to peer production.

383
01:00:59.400 --> 01:01:07.620
James Howison: huge success production is really hard. It's hard in industry and we're actually seeing a reasonable amount of it in science, but it's not enough.

384
01:01:08.280 --> 01:01:17.310
James Howison: Because at least these projects are getting adjustment work fire contributions, but they're not yet doing the work necessary to get the synchronization.

385
01:01:18.030 --> 01:01:23.280
James Howison: And you can see here, you know, key actions for each of people in these areas. And that leads us to the sort of fifth

386
01:01:23.700 --> 01:01:33.900
James Howison: level of maturity here, which is an ecosystem player and these are people who be funded to build community. They know their ecosystem neighbors. They know how their software is combined with other software.

387
01:01:34.680 --> 01:01:46.020
James Howison: And they actually have into project connections that already pushing work upstream and downstream and in some ways I think this is what the Si tu integrations. The mid level grants were really hoping to propagate.

388
01:01:46.920 --> 01:01:57.120
James Howison: But it's also something that we could ask people to write about in funding programs show you show us your research competence. Show us your ecosystem management competence.

389
01:01:58.890 --> 01:02:02.460
James Howison: So here's some suggestions on improving sensing and

390
01:02:04.230 --> 01:02:10.200
James Howison: You know, one of the things we can do is improve the visibility of use in software.

391
01:02:10.890 --> 01:02:23.550
James Howison: And that particularly means, you know, a lot of time users are like, oh, this my private science can't, you can't, you know, investigate this but users of infrastructure have responsibility to demonstrate their usage and feed that back to people.

392
01:02:24.600 --> 01:02:32.250
James Howison: And we can also improve the discover ability of mentions in publications and that's work on pursuing now through the soft side data set, which is

393
01:02:32.970 --> 01:02:41.640
James Howison: Mentioned is hand coded and over 5000 publications designed for machine learning recognition. So go grab it. Do it. I really look forward to what people do with that work.

394
01:02:42.270 --> 01:02:50.550
James Howison: We can improve adjustment work. We could overcome the service center framing. Could we find a program that only funds contributions to quote other people's projects.

395
01:02:51.000 --> 01:03:01.140
James Howison: Could we write a grant solicitation that looks like that. Not what's your big idea. But how are you going to help somebody else's big idea that would drive ecosystem interaction.

396
01:03:03.810 --> 01:03:05.160
James Howison: But the big takeaway here.

397
01:03:06.660 --> 01:03:14.760
James Howison: And there's work being done on this. Now the big takeaway here is that recombination is a key importance of software but it leads to complexity.

398
01:03:15.780 --> 01:03:24.720
James Howison: visibility into that complexity and how it develops over time. That is the challenge of our day across industry across academia.

399
01:03:25.500 --> 01:03:42.780
James Howison: I would say at least markets and peer production in industry have methods and mechanisms but open source science work is struggling just to create open sustainable open projects, but we must drive visibility into these complex hierarchies as well. Thank you.

400
01:03:51.510 --> 01:03:58.380
Manish Parashar: Thank you so much for us absolutely brilliant talk. I mean, it's just amazing. So

401
01:04:00.600 --> 01:04:04.230
Manish Parashar: First, let's give James a virtual round of applause.

402
01:04:05.610 --> 01:04:08.820
Manish Parashar: And thank you again for this say excellent talk.

403
01:04:10.110 --> 01:04:18.630
Manish Parashar: And then we have time for questions. So I see. There's already one question in the Q AMP. A but there is a q&a button at the bottom.

404
01:04:19.950 --> 01:04:29.340
Manish Parashar: And you can go there and enter your questions and I will read it out and have James answer to the first question.

405
01:04:30.330 --> 01:04:43.890

Manish Parashar: States, although researchers don't use the software they write every day. They use an electronic lab notebooks, almost every day, especially in life sciences. Did you look at El an electric

406
01:04:44.370 --> 01:04:59.430
Manish Parashar: Electronic lab notebooks businesses in your research, what do the aliens fit, where do we fit in in the scientific software ecosystem. It does NSF have resources, including these tools, maybe the first part is here.

407
01:05:02.070 --> 01:05:07.110
James Howison: Yeah, thank you. I'm getting a little cross talk to someone, I wonder if people could. There we go.

408
01:05:08.250 --> 01:05:21.540
James Howison: Thank you dealer I. These are really interesting question. Right. So electronic lab notebooks. So things like Jupiter notebook and, you know, many others are really interesting source a potential source of insight into

409
01:05:22.680 --> 01:05:30.240
James Howison: Into software complexity. So we haven't specifically looked at those because we were focusing on the Si tu funding program.

410
01:05:31.560 --> 01:05:38.760
James Howison: But people have done actually by gathering up electronic lab notebooks that are, you know, shared on places like GitHub.

411
01:05:39.180 --> 01:05:50.070
James Howison: And then being able to look at the components that they use. And in a way, those give us insight into this code that's off the analysis level code that's often buried on on people's computers.

412
01:05:50.550 --> 01:06:03.810
James Howison: So I think it's a lot of potential there and I'm happy to follow up with some links of research that actually started to look at those. One thing you could check out is something called world of code, which is coming out of orders Marcus and Jim hope slips group.

413
01:06:05.580 --> 01:06:21.870
James Howison: Looking at that looking at that sort of work and now ln businesses. I'm not quite sure, to the extent that they are actually able to generate revenue that also gives insight into what their users are doing than they can handle ecosystem complexity.

414
01:06:23.550 --> 01:06:28.530
James Howison: I know that one promising development area is the

415
01:06:29.640 --> 01:06:37.740
James Howison: skipped over this one, but I'll leave it up as we talk here is a funding program by the Chan Zuckerberg institute that is specifically sponsoring that Jupiter.

416

01:06:38.940 --> 01:06:53.940
James Howison: It's called the central opens open source scientific components and they've specifically funded paper existing peer production. Right. And so they are grant funding programs would be really interesting place to look at that and then funding Jupiter notebooks. Okay, thank you.

417
01:07:01.710 --> 01:07:05.970
Manish Parashar: Thank you very much. There was another question.

418
01:07:07.140 --> 01:07:07.800
Manish Parashar: Here.

419
01:07:09.150 --> 01:07:20.910
Manish Parashar: Regarding visibility. Do you think there is an opportunity for analytics and analytic tools in software privacy focus alternatives to Google Analytics are gaining usage.

420
01:07:21.420 --> 01:07:31.710
Manish Parashar: On the web. Is there something beyond traces and package managers or similar that could provide anonymous visibility into day to day software usage.

421
01:07:34.410 --> 01:07:34.860
James Howison: Thank you.

422
01:07:36.570 --> 01:07:39.030
James Howison: Very, very hard right so

423
01:07:40.050 --> 01:07:52.500
James Howison: What Elliot is asking about here is what's often called telematics, right. So this is click stream data from software that users are using or an or click stream data and analytics on you know cloud services.

424
01:07:54.150 --> 01:08:03.810
James Howison: I've actually worked on this myself trying to build what we call scientific spy when you know with Jim's Jim hopes lunch group. We didn't call it that. But our potential users called, it's called it that

425
01:08:04.290 --> 01:08:11.070
James Howison: And there are there are real concerns around privacy that aren't that scientists expresses privacy right they express us

426
01:08:12.060 --> 01:08:23.010
James Howison: You know, concerned about competitive pressure and in some communities. You can see this right so we interviewed people in the structural biology community and they're even knowing that somebody is working on X molecule.

427
01:08:23.370 --> 01:08:31.680

James Howison: gives away a ton because apparently it's such a spouse search space that that's too much. So the idea of kind of having

428
01:08:32.220 --> 01:08:40.230
James Howison: Click stream data is difficult. What I will talk about one thing which is Bob McClay attack has a project called exalt which actually looks at

429
01:08:40.980 --> 01:08:46.560
James Howison: The code running on NSF supercomputers NSF funded supercomputers and it looks at the at

430
01:08:47.220 --> 01:09:02.220
James Howison: The libraries that are compiled for those jobs right so they have a sort of a formal intake system which tries to make sure that people are using those libraries and so Bob and tack actually make those data available. I think electronic notebooks are our real possibility but

431
01:09:03.750 --> 01:09:09.870
James Howison: But I think that science is even more scared of that and I think it's a bit of a problem, right, because yes privacy matters.

432
01:09:11.010 --> 01:09:15.240
James Howison: But these are public resources, we're talking about. And there are responsibilities.

433
01:09:15.930 --> 01:09:21.810
James Howison: And I always laugh because I have trouble getting scientific software people to actually genuinely request a citation.

434
01:09:22.140 --> 01:09:28.350
James Howison: Let alone. Remind people give insight and what they're doing to the code. Those are anathema to the openness. So we've got it. We've got to

435
01:09:28.680 --> 01:09:36.240
James Howison: You know, build around this, you know, can we build a consensus that part of your responsibility of being on public infrastructure of benefiting from public money.

436
01:09:36.660 --> 01:09:49.530
James Howison: Is showing you're backstage working in more detail. Can we get over this, you know, genuine and reasonable concerns about competitiveness, what are the forms. So that's a really, you know, interesting area for development. Thank you.

437
01:09:53.430 --> 01:09:55.410
Manish Parashar: Thank you. We have

438
01:09:57.510 --> 01:10:05.400

Manish Parashar: More questions coming in. So the next one is by rally Martin, can the lessons of software.

439
01:10:05.760 --> 01:10:17.280
Manish Parashar: Scientific software sustainability be extended to the development and maintaining the software management infrastructure of our data management infrastructure or is this a fundamentally different issue.

440
01:10:21.120 --> 01:10:22.410
James Howison: Thanks. Thanks rally.

441
01:10:23.820 --> 01:10:27.630
James Howison: So it's a question. Any question how similar our data and software.

442
01:10:28.920 --> 01:10:40.950
James Howison: And I think that one can do some extensions and there are other people doing really important research on this. So I would point you to Christine boardman's group at UCLA also

443
01:10:43.560 --> 01:10:50.490
James Howison: Well, start that, right, because the interesting thing about data is I don't know and I don't think anyone knows yet about

444
01:10:51.090 --> 01:10:56.580
James Howison: How live data sets are right, are we getting to the extent that you're getting lots of new

445
01:10:57.030 --> 01:11:04.800
James Howison: Contributions or transformations of data sets. Then they develop more like software as an activity that exists with all these dependencies.

446
01:11:05.160 --> 01:11:09.840
James Howison: If they're not, then I can see the more like publications, where the maintenance work is a little bit less.

447
01:11:10.290 --> 01:11:14.940
James Howison: So I suspect if you look at work like this publication called science fiction talking about how the

448
01:11:15.510 --> 01:11:22.560
James Howison: How difficult metadata is, you know, to understand and work with and metadata changes more than the underlying data. So I think that's important.

449
01:11:23.010 --> 01:11:31.410
James Howison: I want to dive back for a second to an earlier question has been asked for references summarizing best practices. Yes.

450
01:11:32.370 --> 01:11:39.600

James Howison: And I'll bring those backup. So on this slide, Hong Kong away, you'll see some important links and I really want to highlight the work

451
01:11:40.140 --> 01:11:48.000
James Howison: Of the UK software sustainability Institute and Neil true home as a director that so check that out. So, such a UK software sustainability Institute.

452
01:11:48.240 --> 01:11:54.000
James Howison: They have tons of this sort of important materials. Right. So there are a bunch of papers called 10 simple rules.

453
01:11:54.660 --> 01:12:02.490
James Howison: Some developed, but you know people who are at the NSF others by researchers and they are you know best practice guides that are really important.

454
01:12:03.030 --> 01:12:10.080
James Howison: Also want to highlight the work of the US research software sustainability institute received, you know, Dan cats and cardiogram

455
01:12:10.950 --> 01:12:22.830
James Howison: And they've funded workshops on Community incubation, but also best practice guides and you can find that that work linked from the WS SSP research computing.org.uk it start UK procession US based

456
01:12:23.370 --> 01:12:29.640
James Howison: So that's a group that was funded by the SI to programming. It's lot towards its Lodge, which is the largest still not happened.

457
01:12:30.060 --> 01:12:36.510
James Howison: But it could be really important because the workforce development that you're talking about. We need to fund workforce development in

458
01:12:37.110 --> 01:12:42.240
James Howison: The capabilities of going open, but also the capabilities of ecosystem competence.

459
01:12:42.750 --> 01:12:55.290
James Howison: And I really think that that you'll find that you know use useful papers there and I finally want to point to the Linux Foundation has these open source guides, which are industry based but really also do focus on ecosystem competence.

460
01:12:55.860 --> 01:13:04.470
James Howison: So again, this the work on this side is is crucially important and very, very interesting and I would do want to highlight that there is real ecosystem awareness in the new NSF planning.

461

01:13:05.670 --> 01:13:08.670
James Howison: 2019 blueprint documents. Thank

462
01:13:10.830 --> 01:13:11.490
Manish Parashar: Thank you.

463
01:13:12.600 --> 01:13:30.150
Manish Parashar: We have a few more questions coming in. The next one is from Martin Albert says, Does your research suggest to you that funding agencies should think about fostering sustainability inside the interest structure projects and new ways, and if so, what approaches might be appropriate.

464
01:13:30.480 --> 01:13:44.790
James Howison: Yeah. Fantastic question. Martin. Thanks. So I think that my data speaks to two things. One, the data speaks to actual a reasonable amount of success in generating sustainability via peer production.

465
01:13:45.300 --> 01:13:50.430
James Howison: Okay, there are like nine. I think it's about 19 peer production projects that are still active and going forward. Well,

466
01:13:51.150 --> 01:13:56.220
James Howison: It also speaks to the possibility of transitions to peer production as a mode of sustainability.

467
01:13:57.030 --> 01:14:09.180
James Howison: The key thing there though is the transitions are separate, just like commercialization is so I would really love to see things like I core and other efforts at NSF, which has a long tradition of funding.

468
01:14:09.540 --> 01:14:20.550
James Howison: Commercialization as a separate activity. And one of the things that's really key. There is one of the things that's funded, there is identification of potential markets. Now for businesses that means sales.

469
01:14:21.300 --> 01:14:34.350
James Howison: But open communities that means contributors just actually Weird, right. It's not us. It's definitely not users. It's contributors. So identifying an addressable market of contributors, we can fund that

470
01:14:34.800 --> 01:14:52.530
James Howison: So I think, I think funding existing peer production saying to people, hey, and an example of that. Look at the call for the CCI OSS program. They really nailed that. You know, full disclosure, I and many others that think along my my lines helped write those grant proposals.

471
01:14:53.730 --> 01:15:04.830
James Howison: I also think it's really key to fund visibility into infrastructure complexity and evolution. So that's a separate line. So, last point that I think that we could have additional merit.

472

01:15:06.390 --> 01:15:18.300
James Howison: Requirements in calls that prompt reviewers to discuss their project competencies. Right. So for people to really feature their ability to mentor new contributions to their project.

473
01:15:18.600 --> 01:15:26.790
James Howison: To really feature their ecosystem awareness. Right. And to really understand that those as I say them are aspects of intellectual merit.

474
01:15:27.420 --> 01:15:42.870
James Howison: Right. Knowing what you need to do to continue your scientifically scientific usefulness of the software is not about a good algorithm. It's about what it's about team community and ecosystem competencies and I do think the grand panels can judge those

475
01:15:48.150 --> 01:15:51.450
Manish Parashar: Thank you. Let's see. There's one more question.

476
01:15:52.860 --> 01:16:01.950
Manish Parashar: It'll be useful to hear some of your thoughts about what the scientific high level objective goals question would look like.

477
01:16:03.030 --> 01:16:09.390
Manish Parashar: Included in a solicitation. That's aimed that these kinds of sustainable ability related goals. You mentioned

478
01:16:10.050 --> 01:16:24.000
Manish Parashar: And how projects might migrate tensions or mitigate tensions between doing transition and studying transition that have some parallels to the soft fabulous article publication tension you called out earlier.

479
01:16:25.620 --> 01:16:27.150
James Howison: Yeah, fascinating is a lot there.

480
01:16:28.500 --> 01:16:34.380
James Howison: To me, the high level intellectual our high level objectives, goals questions.

481
01:16:35.430 --> 01:16:37.830
James Howison: It's about framing.

482
01:16:39.120 --> 01:16:51.240
James Howison: Community competencies and ecosystem competencies as intellectually difficult, challenging problems that we can derive generalizable, or at least transferable knowledge about

483
01:16:52.080 --> 01:17:00.510
James Howison: Okay, so it's not where it's not only about how do we build software. But how do we do community

transitions project transitions community building.

484
01:17:00.930 --> 01:17:17.100
James Howison: Ecosystem knowledge. How do we do it in a way that builds the builds knowledge and transferable knowledge. Right. So, you know, to me, that's a lot to ask people who are already working really hard to build software. So I think

485
01:17:18.480 --> 01:17:29.880
James Howison: I think the collaboration is a really key here. So, for example, adding into a solicitation evidence of it of understanding your addressable market of potential contributors

486
01:17:30.450 --> 01:17:37.980
James Howison: And discussing the project activities that will lead to those contributors coming on board. Right. And so that might be mentorship.

487
01:17:39.540 --> 01:17:44.820
James Howison: But also discussing one of the intellectual challenges there wasn't actually intellectually difficult

488
01:17:45.210 --> 01:17:54.300
James Howison: And that's going to naturally lead. Hopefully, I think that you'll find there's some work in science and technology studies or collaborations with organizational researchers, so somebody wants to say

489
01:17:55.200 --> 01:18:07.800
James Howison: You know the intellectual challenges. How do I as the leader of this project manage a leadership transition. Right. And how do I do it in a way that helps other people know how to do it. How do I write about that.

490
01:18:08.940 --> 01:18:19.350
James Howison: So I think, I think that's key on the other side I think observe ability and analysis of ecosystem complexity is an area that size could really get into right

491
01:18:19.710 --> 01:18:27.120
James Howison: So there are a number of firms black duck software is one of them that do large scale source code analysis to identify dependencies within companies.

492
01:18:27.330 --> 01:18:39.870
James Howison: Right. So essentially, they're looking at the company's procurement process and saying, Have you created dependencies on on, you know, a bit of code. And it's it by one person and not not a big balls work in industry and roads and bridges really crucial to that.

493
01:18:41.730 --> 01:18:53.100
James Howison: And so I think that size could fund a project or a grant proposal that's really looking at how we get at this. So the sorts of things. La. It was asking about how do we observe these dependencies, perhaps looking at lab notebooks.

494
01:18:53.490 --> 01:19:06.630
James Howison: And then how do we analyze the dynamics within them. And it's kind of network science, plus, plus, right. What does it look like to have a component that's a lot bottleneck as the work moves around. I think there's real intellectual challenges there. Hopefully that was useful.

495
01:19:13.410 --> 01:19:19.710
Manish Parashar: Thank you. Any other questions. I don't see any more in the Q AMP a box.

496
01:19:21.210 --> 01:19:42.540
Manish Parashar: I would like to remind everybody that after that there is a sec. A post lecture office are at 3pm Eastern scheduled for today and and we can continue the conversation there. So I'm going to pause for a little while to see if any other questions come in.

497
01:19:51.480 --> 01:20:13.230
Manish Parashar: If not, I'm going to thank James again for a wonderful talk. And some very, very timely and important issues that we're all struggling with. So thank you for your insights and your presentation and I look forward to join you again during the

498
01:20:14.790 --> 01:20:24.960
Manish Parashar: Post lecture office hours and I invite everybody to join. Join us there. So thank you again. And with that, I'm going to close this session.

499
01:20:25.380 --> 01:20:29.610
James Howison: Fantastic. Thank you. Thank you very much money. No, and everybody