

## LF AI & Data Foundation Technical Advisory Council (TAC) Meeting Minutes February 8 2024 (6:00am - 7:00am PST) via Zoom

TAC Meetings are bi-weekly and open to everyone in the LF AI & Data community.

### Voting Notes

Under the original charter rules we need 50% attendance from the voting members (represented by premier members + graduate projects) to meet quorum. To ensure quorum is met more consistently, the TAC implemented the following attendance and voting eligibility rule effective Dec 1, 2022: Voting members of the TAC who miss 2 consecutive meetings will not be allowed to vote and must then attend 2 consecutive meetings before their voting privileges are reinstated.

With the new attendance and voting rule adopted by the TAC, the number of eligible voting members varies per week and is tracked [here](#). Please see meeting minutes on the [TAC Wiki](#) for details on attendance and voting eligibility on a per meeting basis.

Please see the [TAC Wiki](#) for more information about voting.

Please see current voting members, including alternates, on the [TAC Wiki](#).

Voting details for today's meeting: 7 voting members were eligible to vote and 6 of them were present.. Quorum was therefore met for this meeting.

TAC Voting Member	Voting Representative	2/8/2024
<b>Quorum</b>	<b>Attendance/No. Eligible Votes</b>	<i>6 out of 7 eligible were present</i>
4paradigm	Zhongyi Jerry Tan	
Amazon Web Services	Brian Granger	
	Mark Atwood	
Baidu	Jun Zhang	
	Alternate: Daxiang Dong	
	Alternate: Yanjun Ma	
Ericsson	Rani Yadav-Ranjan	
Huawei	Howard (Huang Zhipeng)	
	Alternate: Charlotte (Xiaoman Hu)	
	Alternate: Leon (Hui Wang)	
IBM	Susan Malaika	X
	Alternate: Beat Buesser	
	Alternate: Alexandre Eichenberger	
Nokia	Michael Rooke	X
	Alternate: Jonne Soininen	
OPPO	Jimmy (Hongmin Xu)	
SAS	*Nancy Rausch	

	Alternate: JP Trawinski	
ZTE	Wei Meng	
	Alternate: Liya Yuan	
Adversarial Robustness Tool Project	Beat Buesser	X
	Alternate: Kevin Eykholt	
Angel Project	Jun Yao	
Egeria Project	Mandy Chessell	X
	Alternate: Nigel Jones	
	Alternate: David Radley	
	Alternate: Maryna Strelchuk	
	Alternate: Ljupcho Palashevski	
	Alternate: Chris Grote	
Flyte Project	Ketan Umare	
Horovod Project	Travis Addair	
Milvus Project	Xiaofan Luan	
	Alternate: Jun Gu	
ONNX Project	Alexandre Eichenberger	X
	Alternate: Jim Spohrer	
	Alternate: Prasanth Pulavarthi	
	Alternate: Andreas Fehlner	
Pyro Project	Fritz Obermeyer	
OpenLineage	Julien Le Dem	
	Alternate: Mandy Chessell	
	Alternate: Michael Robinson	X
Marquez	Willy Lulciuc	
	Alternate: TBD	
	Alternate: TBD	

### Note for Voting Members

Please ensure you attend the bi-weekly TAC meetings to maintain voting eligibility. If you have not already provided an alternate representative, please email Nancy Rausch (Nancy.Rausch@sas.com), TAC Chair AND operations@lfaidata.foundation to designate an alternate representative. It is critical to meet quorum during the meetings especially when there are voting items on the agenda.

---

### **LF Staff Attendees**

- Lucy Hyde, Senior Program Manager, Linux Foundation
- Matt White, GenAI Commons

### **Non-LF Staff Attendees**

- Christina Certo
- Oita Coleman
- Alexandre Eichenberger
- James Harroun
- Ofer Hermoni
- Vini Jaiswal
- Nero Okwa
- Juan Luis Cano Rodríguez
- Joel Schwarzmann
- Hakan Sonmez
- Merel Theisen

### **Call to Order**

Lucy Hyde (LH) called the meeting to order at 6:05am Pacific and Nathan Southern, though not present, took minutes after the fact from the recording.

LH reviewed the LF Antitrust Policy notice, reminded everyone that TAC calls are recorded, and presented all social media links.

### **Agenda**

LH reviewed the agenda for the meeting and noted that it would chiefly focus on the Kedro Project review. There were no further changes or additional topics added.

### **Approval of Minutes**

With quorum achieved, the TAC sought minutes approval for the January 25, 2024 meeting. Alexandre Eichenberger (AH) issued a first motion for approval; Mandy Chessell (MC) put forward a second motion for approval. All voting members who were present expressed approval, with no dissents, and LH marked the minutes approved.

### **Kedro Project Annual Review**

Juan Luis Cano Rodriguez (JR) presented the review of Kedro. JR noted that Kedro is an Open Source Python Library used to create reproducible, maintainable and modular data science code, and that uses software engineering best practices to help the user build production-ready data science pipelines. Kedro was donated to LF AI & Data by McKinsey and Quantum Black in early 2021. Yetunde Data, the former product manager for the project until 2023, served as chair until JR assumed this role. The source code is hosted on GitHub. Recently the project exceeded 9000 Github Stars. Community engagement is high.

JR then presented a demo of the Kedro pipeline. JR noted that Kedro-Viz is a web application developed by the project that enables the user to see into pipelines and interact with them. In said demo, JR

illustrated how the process begins with numerous artifacts/datasets, that are connected with nodes, that in turn produce a series of results (plots, datasets, etc.) The datasets/pipelines can be nested, and in turn broken down further.

Kedro-Viz offers a visual representation of the platform; everything is powered under the hood by Kedro, as well as its plug-ins, which are Python-based frameworks for data science.

JR then listed some of the key organizational contributors to his project, including McKinsey & Company, Getindata, Société Générale, and Voltron Data.

JR noted that over 200 individuals are contributing code and documentation, and that, per statistics from LFX Insights, the project is healthy. In 2023, Kedro also surpassed 1,500 users in its public Slack following a migration from Discord. JR also stated that the project has 81 packages or plugins - some of which are showcases on how to use Kedro for specific things that can be found in the open, and some of which can be remixed or reused in existing projects and maintained by the TSC, or individuals outside of the TSC.

JR stated that Kedro also published a three-hour video course on Youtube that it used to reach a much broader audience and establish a new workstream for producing more video content for social media engagement. Kedro now has a broad presence on Facebook, LinkedIn and the Fediverse (Mastodon).

JR added that as a product of the project's growth, it has added two additional organizations to the TSC. The Kedro documentation was announced as 2023's winner of the UK Technical Documentation Awards. The project was recently listed as one of Datacamp's top MLOP tools and ranked 7th on OSSInsights MLOP Tools.

JR then listed several recent technical developments on Kedro, including: the release of kedro 0.19, kedro-datasets 2.0 and kedro-viz 6.x and 7.x; decoupled kedro and kedro-datasets; the replacement of legacy configuration loaders with an OmegaConf-based one; and the shipment of shareable URLs in kedro-viz.

JR then reviewed areas in which the project needs help, which include:

- *Organic Growth:* Though the project has many GitHub stars, it has received relatively few mentions in Python Conferences, on Reddit, on Medium, and the like and would benefit from this.
- *Champions Program:* Kedro needs to more regularly engage power users to build a stronger sense of community.
- *Greater TSC Involvement:* At present, the majority of non-McKinsey coding activity happens in satellite repositories (plugins).
- *Data-driven Decision Making:* Kedro needs telemetry collection that respects user privacy and yet supports evolution - and is working in tandem with the Linux Foundation and McKinsey to effectuate this.

JR concluded his presentation by discussing the project's feelings about working with LF AI & Data. He praised the responsiveness of foundation leadership and LF IT, but expressed some uncertainty regarding the status of LF Insights Portals and Security dashboards, as well as a desire for Kedro to set up its own Snyk. He also conveyed an interest on behalf of the project to collaborate with other LF AI & Data projects, such as OpenLineage, Amundsen and Feast.

## **TAC Open Discussion**

LH suggested that JR connect with Linux Foundation marketing coordinators Jen Shelby (JS) and Carly Akerly (CA) to strategically publicize Kedro's progress through the LF blog and social media channels.

LH also suggested that Kedro become involved in the Ambassador Program at the Linux Foundation - which provides a way to promote community members, give them a podium to discuss the work that is ongoing, and interact more intensively with members of the community, with added benefits including training discounts, swag and more.

LH also asked JR about the datasets Kedro uses. She noted that recently, the general manager of DroneCode reached out. They have a large quantity of publicly-available data (much of it originating from drones) but no one to use it or process it. LH expressed a willingness to put DroneCode in touch with the Kedro team, if said team is seeking unique datasets with which to work.

JR clarified that this may not be relevant, as their term 'datasets' applies to connectors to different data sources, not to 'datasets' in the conventional sense. However, the Kedro team is interested in defining and working on a use case regarding how to use data in unstructured forms.

Michael Robinson (MR) of OpenLineage (OL) stated that his project possesses a strategic vision similar to Kedro's, regarding the tactical direction of the project's integrations. MR asked JR to confirm that Kedro's plugins are externally managed and maintained, and also asked him to comment on the approach Kedro took within the community to make this happen.

JR responded that Kedro has been observing and following OL for some time, and took note of the fact that static lineage took place in OL release 1.0 but hasn't yet defined a use case for OL on Kedro. JR also foresees potential collaboration between Kedro and Marquez. Regarding plugins: several levels exist within Kedro. The project uses a tool called Orbits that models community engagement as a series of orbits around the core. In the innermost orbit, Kedro maintains its own plugins, including: Kedro Airflow, Kedro Docker, Kedro Telemetry and Kedro Datasets. These are managed in the Kedro Github organization. In addition to this, JR outlined various plugins created by the Kedro TSC members, housed on their own individual Github organizations and used very frequently.

Finally, JR stated that some Kedro users upload random plugins to their own Github repositories, but that the project has little to no control over these.

MR then asked JR if Kedro has been able to identify the most popular tools. In response, JR stated that this is gauged by a combination of factors; for instance, the Kedro team sees, in its slack, increased activity (and requests) around certain plugins, and comparatively little activity around others. In general, the Kedro team takes a heterogeneous approach to measuring tool popularity. It has observed that certain plugins outside of those maintained by the original team are extremely popular among community members.

MR thanked JR and expressed an interest in connecting off-meeting and exploring synergies and possible collaborations between OpenLineage and Kedro.

JR concluded by thanking the team for the opportunity to present a project update.

## **TAC Announcements**

LH concluded the meeting with the following announcements:

- The next TAC meeting on Feb. 22nd will feature the graduation of the DocArray project
- OpenKedro will present at the March 7th TAC

LH also asked anyone with a future TAC agenda item to reach out to the group via the mailing list: [tac-general@lists.lfaidata.foundation](mailto:tac-general@lists.lfaidata.foundation).

### **Adjournment**

With no additional business, the meeting adjourned at 6:32am Pacific.

### **Action Items**

JR: connect with Linux Foundation marketing coordinators Jen Shelby (JS) and Carly Akerly (CA) to strategically publicize Kedro's progress through the LF blog and social media channels.

JR: connect with LH to discuss and move forward with Kedro's involvement in the Linux Foundation's Ambassador Program.

JR and MR: connect off meeting to explore synergies and overlaps between OpenLineage and Kedro.

Anyone on the TAC interested in learning more about Kedro should reach out to JR off-meeting.

### **Takeaways**

LH: Touch base with Nancy Rausch and/or Ibrahim Haddad after the meeting to verify OpenKedro's presentation at the March 7th TAC, and if applicable, coordinate the details with JR.

### **Chat Transcription**

1

00:01:11.540 --> 00:01:12.490

Juan Luis Cano Rodríguez: Hello!

2

00:01:14.860 --> 00:01:15.990

Nero Okwa (Kedro): Hello!

3

00:01:19.210 --> 00:01:29.119

Lucy Hyde: I am helping Nancy out today, and we'll she's not able to make it yet, so we'll give the rest of the folks here a couple of minutes before they come in.

4

00:01:29.810 --> 00:01:30.750

Juan Luis Cano Rodríguez: Fantastic.

5

00:02:34.340 --> 00:02:40.960

Lucy Hyde: Good morning, everyone. We're just. I'm helping Nancy out this morning. She's unable to make it. I'll give everybody a couple more minutes to

6

00:02:41.320 --> 00:02:47.459

Lucy Hyde: come on to the meeting. I know that it's very early for Pacific right now. So

7

00:02:53.500 --> 00:02:55.589

Lucy Hyde: good morning, Michael. Good morning, Oita.

8

00:03:00.980 --> 00:03:02.570

Oita Coleman: Good morning, everyone.

9

00:03:07.010 --> 00:03:13.319

Oita Coleman: Afternoon for those sewer and other towns.

10

00:03:15.970 --> 00:03:17.290

Michael Robinson | OpenLineage: Good morning, everyone.

11

00:03:22.700 --> 00:03:31.190

Lucy Hyde: I'll probably repeat this a few more times. I'm hoping out Nancy this morning. She's unable to make it so. We'll give everybody a couple of more minutes here and try and get started at 5 after.

12

00:03:34.600 --> 00:03:35.920

Oita Coleman: Thanks, Lucy.

13

00:06:17.740 --> 00:06:25.980

Lucy Hyde: Good morning, everybody, and we'll give it a couple more minutes here. Nancy isn't able to make it this morning, so I am covering for her. We'll get started at about 5 after

14

00:07:14.770 --> 00:07:18.450

Lucy Hyde: alrighty. Let me just share my screen.

15

00:07:20.190 --> 00:07:22.010

Lucy Hyde: This may

16

00:07:26.570 --> 00:07:29.829

Lucy Hyde: awesome welcome. Oh, no! I'm

17

00:07:30.150 --> 00:07:32.759

Lucy Hyde: in the wrong tech meeting that would.

18

00:07:32.830 --> 00:07:37.569

Lucy Hyde: Oh, they were me for 1 s. I'm confusing myself

19

00:07:41.940 --> 00:07:43.590

Lucy Hyde: all right trying again.

20

00:07:43.660 --> 00:07:54.429

Lucy Hyde: Welcome to the February eighth hack meeting. We have Ketro's Annual Review today. I am covering for Nancy today she was unable to make it this morning.

21

00:07:55.780 --> 00:08:21.149

Lucy Hyde: Alright. This is our anti trust policy that everybody is probably extraordinarily familiar with at this point. We abide by applicable antitrust and open competition laws. They adhere to all sorts of things, meeting agendas. You know, meetings, groups, and if you need a copy of our anti trust policy, it's available on our website. If you have any questions about the anti-trust policy.

22

00:08:21.150 --> 00:08:28.130

Lucy Hyde: you are welcome to reach out, and we can put you in contact with Andy up to growth. Or you are welcome to get back to your respective legal counsel.

23

00:08:31.070 --> 00:08:42.860

Lucy Hyde: Just a reminder that all Tac calls are recorded and they will be available to view on the wiki and our long list of sites. If you type in Lfa, I data it should bring up

24

00:08:43.159 --> 00:08:46.449

Lucy Hyde: a our website which points to all of these as well.

25

00:08:48.640 --> 00:08:55.490

Lucy Hyde: All right. I do actually believe we have quorum this morning. And then we have the kitchen project review

26

00:08:58.790 --> 00:09:06.020

Lucy Hyde: alright, and just a general reminder to please change your zoom, display name so that we know who you are and where you go

27

00:09:08.130 --> 00:09:09.570

Lucy Hyde: remembers.

28

00:09:10.290 --> 00:09:18.210

Lucy Hyde: All right. This is a motion to approve. The January 25 min from the January 20 fifth tac call. Can I have a first motion?



29

00:09:20.580 --> 00:09:22.509

Alexandre Eichenberger: Yeah, Alex Ekenberger onyx.

30

00:09:22.600 --> 00:09:24.470

Lucy Hyde: Thank you. And a second

31

00:09:24.770 --> 00:09:33.919

Mandy Chessell (egeria): Armandy chest aligaria. Thank you very much. All voting members in favor of approving the minutes. Can you come on mute, and say, aye, aye.

32

00:09:34.450 --> 00:09:35.750

Lucy Hyde: thank you

33

00:09:35.840 --> 00:09:41.220

Lucy Hyde: that the minutes for the January 20 fifth tech call are approved, and they will be available on the wiki.

34

00:09:42.840 --> 00:09:54.360

Lucy Hyde: All right. With that being said, I will hand it over to the Kedra team to present their annual review. Would you like to screen, share, or I can continue to screen, share? It's whatever you prefer.

35

00:09:54.600 --> 00:09:59.359

Juan Luis Cano Rodríguez: if it's possible. May I screen shirt and made some slight modifications to the slides.

36

00:09:59.420 --> 00:10:00.919

Lucy Hyde: Absolutely. Hold on one sec.

37

00:10:10.610 --> 00:10:24.510

Juan Luis Cano Rodríguez: Alright, so well, Hello, everyone. And thanks for joining. So my name is Juan Luis. I'm currently working as product manager for Cedro at quantum black, which is part of Mckinsey. and we want to give you an annual review for Kevlar today.

38

00:10:25.530 --> 00:10:39.330

Juan Luis Cano Rodríguez: So first of all, a little bit of context about the projects. So Cedro is an open source. Python Library for production, 20 data science. It was donated to the Linux Foundation, AI, and data by Mckinsey and Quantum black

39

00:10:39.590 --> 00:10:44.970

Juan Luis Cano Rodríguez: in early 2021 and

40

00:10:45.680 --> 00:10:56.440

Juan Luis Cano Rodríguez: It has been chaired initially by Yetundada, who was the former product manager for the project until last year I took over.

41

00:10:56.560 --> 00:11:07.069

Juan Luis Cano Rodríguez: You can see here some links. The source code is hosted on Github. Here's the website. and these are the new foundation mailing lists.

42

00:11:07.130 --> 00:11:24.780

Juan Luis Cano Rodríguez: and we included some numbers about community engagement. We're going to talk a little bit about this in a moment. But wanted to mention that we surpassed 9,000 Gifa stars, and very recently, which is one of the achievements we want to highlight

43

00:11:25.910 --> 00:11:28.659

Juan Luis Cano Rodríguez: to give you a bit of an impression of how

44

00:11:28.770 --> 00:11:35.250

Juan Luis Cano Rodríguez: Kedro and Kelvis looks like. I have a very quick demo, and I want to show you

45

00:11:36.030 --> 00:11:37.180

Juan Luis Cano Rodríguez: so.

46

00:11:37.370 --> 00:11:47.379

Juan Luis Cano Rodríguez: This is how a Kedro pipeline looks like we are looking at Kedrovis, which is a web application that we develop

47

00:11:47.530 --> 00:11:56.509

Juan Luis Cano Rodríguez: and that allows users to see how the pipelines look like interact with them, and and so on. So you see, there's a number of

48

00:11:56.720 --> 00:12:11.410

Juan Luis Cano Rodríguez: artifacts or datasets that are on top, and then a number of nodes that connect those data sets that then produce some other results, such as plots or other datasets, and so on.

49

00:12:11.810 --> 00:12:17.339

Juan Luis Cano Rodríguez: These datasets and pipelines can be nested. So, for example.

50

00:12:17.370 --> 00:12:20.440

Juan Luis Cano Rodríguez: and you can break down some of the pipelines and so on.

51

00:12:20.870 --> 00:12:36.979

Juan Luis Cano Rodríguez: Kdrovis is a tool that does a visual representation of the pipeline, and then everything is powered under the hood by K. Pro. As well as its plugins, and which are python based frameworks for data science.

52

00:12:37.990 --> 00:12:40.100

Juan Luis Cano Rodríguez: So, going back to the presentation.

53

00:12:41.960 --> 00:12:48.599

Juan Luis Cano Rodríguez: The organizations that are currently contributing to Cedro are listed here, Mckinsey and Company, as the

54

00:12:48.700 --> 00:12:53.870

Juan Luis Cano Rodríguez: originating organization that initially donated the project.

55

00:12:54.270 --> 00:13:03.729

Juan Luis Cano Rodríguez: and we had a colleague, from working for a startup called Claypot AI, that has been acquired by voltron data

56

00:13:03.750 --> 00:13:05.099

Juan Luis Cano Rodríguez: very recently.

57

00:13:05.280 --> 00:13:25.499

Juan Luis Cano Rodríguez: and this year we were able to include 2 other organizations to the Keto Tsc. One is. One of them is getting data, abolish data consultancy, and another one is Societa General, a French insurance company. They're both using Kedro in their projects and also actively contributing

58

00:13:25.620 --> 00:13:30.939

Juan Luis Cano Rodríguez: to the discussions and in the form of social plugins and extensions, and so on.

59

00:13:32.750 --> 00:13:46.839

Juan Luis Cano Rodríguez: Here's some numbers about the project, so you can see that as I was mentioning before, we have now more than 9 0 stars. There's a large number of people contributing

60

00:13:46.880 --> 00:13:50.630

Juan Luis Cano Rodríguez: and code and documentation over 200

61

00:13:50.700 --> 00:14:06.570

Juan Luis Cano Rodríguez: at this point, and you can see a healthy evolution of the project. There are some statistics as well in the insights portal, all the foundation, and these were a little bit more up to date.

62

00:14:08.730 --> 00:14:16.079

Juan Luis Cano Rodríguez: So a few words on some key achievements we made in 2023. So in terms of community growth.

63

00:14:16.180 --> 00:14:19.770

Juan Luis Cano Rodríguez: developer advocacy growing the Tsc, and so on.

64

00:14:19.870 --> 00:14:33.390

Juan Luis Cano Rodríguez: we surpassed 1,500 members in our public slack, and which is and shared among the Tsc members, internal and external community participants.

65

00:14:33.400 --> 00:14:39.120

Juan Luis Cano Rodríguez: and and so on. And this was a big achievement for us because we were migrating from

66

00:14:39.190 --> 00:14:46.309

Juan Luis Cano Rodríguez: discord a few years back. And now we have a frightened community that's participating.

67

00:14:46.590 --> 00:14:49.470

Juan Luis Cano Rodríguez: We also surpassed 9,000 Github stars.

68

00:14:49.860 --> 00:15:02.420

Juan Luis Cano Rodríguez: There's at the moment 81 packages or plugins for cater. Some of them are showcases on how to use cable for specific things that can be found in the open.

69

00:15:02.500 --> 00:15:13.540

Juan Luis Cano Rodríguez: Some of those are plugins that can be remixed and reused in existing projects and maintained by the Tsc. Or people that do not yet belong to Tsc.

70

00:15:14.610 --> 00:15:20.680

Juan Luis Cano Rodríguez: We published a 3 h video course on Youtube. and which was

71

00:15:20.750 --> 00:15:31.649

Juan Luis Cano Rodríguez: big undertaking as well, and it has served us to reach a much broader audience and kickstart new workstream

72

00:15:31.670 --> 00:15:36.149

Juan Luis Cano Rodríguez: for producing more video content for social media engagement.

73

00:15:36.940 --> 00:15:40.599

Juan Luis Cano Rodríguez: As I mentioned, we have added 2 new organizations to the Tsc.

74

00:15:41.830 --> 00:15:50.559

Juan Luis Cano Rodríguez: Also, the kid documentation was announced as the oil winner of the UK. Technical communication awards last year, which we're very proud of.

75

00:15:50.920 --> 00:16:01.370

Juan Luis Cano Rodríguez: We have established stable social media presence on LinkedIn Youtube and the failures, thanks to the master server that the linguist foundation provides.

76

00:16:02.080 --> 00:16:03.290

Juan Luis Cano Rodríguez: and

77

00:16:03.310 --> 00:16:18.939

Juan Luis Cano Rodríguez: one of the data camps, tops, envelops tools. Sorry one. It was listed as one of the tops. envelopes, tools, and in data times blog and ranked as seventh

78

00:16:19.160 --> 00:16:22.170

Juan Luis Cano Rodríguez: top mouse tool in Oss insights as well

79

00:16:23.290 --> 00:16:27.170

Juan Luis Cano Rodríguez: in terms of new features and functionalities. We released

80

00:16:27.240 --> 00:16:36.190

Juan Luis Cano Rodríguez: new major versions across the board in different parts of the project, so cetera 19, and

81

00:16:36.600 --> 00:16:47.379

Juan Luis Cano Rodríguez: has been on our a year and a half in the making. And it's a new major release that helps us undertake several deprecation

82

00:16:48.000 --> 00:16:52.869

Juan Luis Cano Rodríguez: cycles remove some outdated functionalities and so on.

83

00:16:53.090 --> 00:17:04.680

Juan Luis Cano Rodríguez: We also need to schedule datasets 2.0, with lots of contributions from the community. And Kendall Vis has released 2 series of major releases 6

84

00:17:04.710 --> 00:17:06.159

Juan Luis Cano Rodríguez: and 7.

85

00:17:06.700 --> 00:17:17.869

Juan Luis Cano Rodríguez: We managed to decouple Kedro and cedar datasets which have been a long time in the making. So now the Kedro core is much more lightweight and individual datasets can be installed separately.

86

00:17:18.300 --> 00:17:29.450

Juan Luis Cano Rodríguez: We have replaced our legacy configuration loaders with one based on Omega Comp, which is a very established library for dynamic configuration loading, and we're now leveraging that

87

00:17:29.870 --> 00:17:40.650

Juan Luis Cano Rodríguez: and Kobe's shipped shareable Urls, which allow users to quickly share their dashboards with people that do not have Caleb's installed

88

00:17:42.530 --> 00:17:47.980

Juan Luis Cano Rodríguez: in terms of areas the brain could use some help on. There's a few of them. So on one hand.

89

00:17:48.140 --> 00:17:54.759

Juan Luis Cano Rodríguez: We see some stagnation of organic growth. So there's lots of giga stars, and it keeps growing.

90

00:17:54.790 --> 00:18:01.600

Juan Luis Cano Rodríguez: But when we go to where practitioners in the community shared knowledge such as

91

00:18:01.640 --> 00:18:08.259

Juan Luis Cano Rodríguez: python conferences, credit threads, medium blog posts, and so on. And we don't see lots of

92

00:18:08.270 --> 00:18:20.479

Juan Luis Cano Rodríguez: content generated by the users. So it's an area that we're actively working on. We also would like to kick start a champions program. So there's already

93

00:18:20.500 --> 00:18:34.330

Juan Luis Cano Rodríguez: a fair number of key contributors, power users that are not directly involved in the Tsc. Just yet, but they are showing lots of potential, and they participate consistently in Github, slack, and so on.

94

00:18:34.650 --> 00:18:37.569

Juan Luis Cano Rodríguez: And we would like to hear from other members of

95

00:18:37.940 --> 00:18:40.490

and the Leo's foundation Tac

96

00:18:40.580 --> 00:18:51.270

Juan Luis Cano Rodríguez: of what has worked for them to build a stronger sense of community. Also, the way the category is set up as a constellation of repositories.

97

00:18:51.330 --> 00:18:55.860

Juan Luis Cano Rodríguez: Makes it so that the majority of

98

00:18:56.020 --> 00:19:03.249

Juan Luis Cano Rodríguez: coding activity that is not by Mckinsey, is happening in these plugins, and so on.

99

00:19:03.260 --> 00:19:12.160

Juan Luis Cano Rodríguez: But in Keto Core and Keros most of the contribution comes from Mackenzie, so we are wondering if at some point we will

100

00:19:12.220 --> 00:19:23.710

Juan Luis Cano Rodríguez: want to diversify this contribution and make other Tsc members contribute more actively to the roadmap and the core of the project as well.

101

00:19:24.790 --> 00:19:34.170

Juan Luis Cano Rodríguez: And finally, we are using user telemetry to support our data-driven decision-making process.

102

00:19:34.190 --> 00:19:39.689

Juan Luis Cano Rodríguez: And this has been in place for some years. But we were taking a very conservative approach

103

00:19:39.770 --> 00:19:54.130

Juan Luis Cano Rodríguez: by having an opt-in and collection mechanism and a separate package for the telemetry collection, and so on. And now we're working with the legal teams of Mckinsey and Linux Foundation, AI and data as well.

104

00:19:54.190 --> 00:20:09.389

Juan Luis Cano Rodríguez: so that they can assist us in making this process collect less data and make sure that we're not collecting any personal data with the goal of enabling an opt out mechanism and be able to

105

00:20:09.560 --> 00:20:12.429

Juan Luis Cano Rodríguez: collect higher volume

106

00:20:12.470 --> 00:20:14.369

Juan Luis Cano Rodríguez: of data that can improve

107

00:20:14.620 --> 00:20:16.080

our decision process.

108

00:20:18.300 --> 00:20:23.029

Juan Luis Cano Rodríguez: and finally, in terms of feedback on working with Lf. A. And data. And

109

00:20:23.340 --> 00:20:34.099

Juan Luis Cano Rodríguez: of course, Jacqueline, when she was still working at Lf. And now, Lucy, have been always really helpful and responsive, and we cannot say otherwise.

110

00:20:34.130 --> 00:20:40.329

Juan Luis Cano Rodríguez: and we frequently ask the It team to change some Dns

111

00:20:41.120 --> 00:20:48.899

Juan Luis Cano Rodríguez: records for our slack invites and so on, and they're also always super responsive. So we are. We are very grateful for that.

112

00:20:49.320 --> 00:21:06.289

Juan Luis Cano Rodríguez: Having the one password as a shared resource as well as the mailing list, is really useful for us. To avoid single points of failure, and also make sure that we can easily share resources among different members of the Tsc.

113

00:21:07.270 --> 00:21:08.610

Juan Luis Cano Rodríguez: We had to say, there's

114

00:21:08.730 --> 00:21:28.739



Juan Luis Cano Rodríguez: some unclear status on different insights, portals, security dashboards, and so on. This is already in progress when conversations with Iraham. To understand what is the status of the security dashboards? Considering setting up our own sneak, dashboard, and so on.

115

00:21:29.190 --> 00:21:42.359

Juan Luis Cano Rodríguez: And potentially, we think that could be closer across collaboration between adjacent projects. And, for example, Kedro and the open lineage or keto and Amazon as a data catalog, keto and fist as a

116

00:21:42.500 --> 00:21:45.180

Juan Luis Cano Rodríguez: feature store, and so on.

117

00:21:46.260 --> 00:21:53.130

Juan Luis Cano Rodríguez: So though that is it for my site, and I'll be happy to take any questions from the ITC. Thanks a lot.

118

00:21:56.640 --> 00:22:21.759

Lucy Hyde: It looks like you've shown a lot of progress over the past year. It's really lovely to see. I had. I have actually 3 things now that I think of it. The first is if you'd like to connect with Jen Shelby, and we also brought on a new marketing contractor, Carly Ackerley to promote some of the really good milestones that you've made in the last year, and she can publish it to the Lf blog and also put it on the social media channels.

119

00:22:21.840 --> 00:22:24.689

Lucy Hyde: She's really Jen is amazing.

120

00:22:24.900 --> 00:22:29.749

Lucy Hyde: The second is so we do have

121

00:22:29.930 --> 00:22:36.569

Lucy Hyde: the like documentation and infrastructure for what we call an ambassador program at the Lf, and basically it's like a

122

00:22:36.580 --> 00:22:54.119

Lucy Hyde: a way to promote community members kind of give them a you know. A podium to speak about the project and all of the work that's going on, and, you know, really interact with folks out in the community. And then on the kind of the flip side we can offer them discounted training some Lf, swag that type of thing.

123

00:22:54.170 --> 00:23:13.229

Lucy Hyde: So if you're interested in that, more than happy to meet offline and I can, I can get you started with that as well. That's of interest to Ketro. And the last thing is, I'm curious about the data sets that you use and this is really for everyone. So the general manager of Drone code reached out. They have

124

00:23:13.340 --> 00:23:33.419

Lucy Hyde: tons of publicly available data, but they don't really have anybody to use it or process it, and it it because it does come from drones and other Uav. It is a little bit outside of the norm of what you'd see, but I'm also more than happy to send you the resources there. If you're looking for a unique data set to play around with

125

00:23:35.270 --> 00:23:52.999

Juan Luis Cano Rodríguez: fantastic thanks. A Lot Lucy, for just for clarification. What we call data sets are actually connectors to different data sources and not data sets in the sense of Cds of observations. But we would be super interested in maybe working on a use case on how to use schedule to

126

00:23:53.070 --> 00:23:54.520

Juan Luis Cano Rodríguez: process

127

00:23:54.570 --> 00:24:07.050

Juan Luis Cano Rodríguez: non structured data of different forms, because we really have connectors for different formats. And we're working on promotional material, technical blog posts and so on. So yeah, super happy to collaborate there.

128

00:24:07.230 --> 00:24:09.350

Lucy Hyde: Perfect. That sounds good. Thanks.

129

00:24:12.600 --> 00:24:15.249

Michael Robinson | OpenLineage: I have a question.

130

00:24:16.540 --> 00:24:24.079

Michael Robinson | OpenLineage: This is Michael with open lineage. I really enjoyed your presentation, and it's a great looking project.

131

00:24:24.730 --> 00:24:26.750

Michael Robinson | OpenLineage: we're a

132

00:24:27.380 --> 00:24:38.669

Michael Robinson | OpenLineage: I don't know if we're in a similar stage or not, but we have. It sounds like a similar strategic vision in terms of where we ultimately want our integrations to be.

133

00:24:38.820 --> 00:24:53.060

Michael Robinson | OpenLineage: It am. Am I correct in understanding that your your plugins are externally managed and and maintained? And if so.

134

00:24:53.190 --> 00:24:59.339

Michael Robinson | OpenLineage: could you talk about the the approach you took within the community to make that happen?

135

00:25:00.430 --> 00:25:17.699

Juan Luis Cano Rodríguez: Yes, so indeed! We've been looking at open lineage, and for some time already. There's no proof of concept just yet, but I was pretty much looking forward for the static lineage to take place, which already happens in all one.

136

00:25:17.910 --> 00:25:23.730

Juan Luis Cano Rodríguez: And I've been training around with Marquez as well. And so I think there's lots of potential collaboration there

137

00:25:24.050 --> 00:25:35.970

Juan Luis Cano Rodríguez: for the plugins. There are several levels we follow, or we're using a tool called orbit that models community engagement as a series of orbits around the core.

138

00:25:36.200 --> 00:25:47.970

Juan Luis Cano Rodríguez: And in the innermost orbit E, we have our own plugins that we maintain. So ketro airflow, cetera docker Keto telemetry and keto data sets. Those are managed in the

139

00:25:48.000 --> 00:25:52.180

Juan Luis Cano Rodríguez: Cedar Github organization and is pretty much the Tsc.

140

00:25:52.380 --> 00:25:56.679

Juan Luis Cano Rodríguez: And mostly quantum black team committing to them.

141

00:25:56.920 --> 00:26:07.650

Juan Luis Cano Rodríguez: Then there are other plugins that the Sc. Members create created, but they are on their own individual Github organizations.

142

00:26:07.950 --> 00:26:10.110

Juan Luis Cano Rodríguez: But they're pretty much

143

00:26:10.300 --> 00:26:13.770

Juan Luis Cano Rodríguez: among the most widely used

144

00:26:13.940 --> 00:26:19.219

Juan Luis Cano Rodríguez: cadre plugins. So they are really important for the project as a whole, as an ecosystem.

145

00:26:19.990 --> 00:26:22.870

Juan Luis Cano Rodríguez: And then there are other plugins that

146

00:26:23.170 --> 00:26:45.200

Juan Luis Cano Rodríguez: random users upload to their github or Github repositories, and we don't have control over those, but we try to discover them when they exist, and we aggregate them in a repository that follows the awesome kind of listings. So we have an awesome ketro. Repository for these things.

147

00:26:46.620 --> 00:27:01.949

Michael Robinson | OpenLineage: Thank you for that. That's really interesting. I don't wanna I don't wanna hog the discussion session here. But, in terms of your plugins, have you been able to identify the most popular tools?

148

00:27:03.580 --> 00:27:11.510

Juan Luis Cano Rodríguez: Yeah, it's usually, you know, a combination of factors. I'm not usually a fan of just looking at

149

00:27:11.630 --> 00:27:16.920

Juan Luis Cano Rodríguez: one-dimensional matrix, like the house stars, and so on, although that usually gives a good indication.

150

00:27:16.980 --> 00:27:34.830

Juan Luis Cano Rodríguez: But we see, for example, in our slack Workspace, lots of requests and questions around certain plugins and very little activity, and around different others. So we do have a sense of which ones are the most

151

00:27:35.090 --> 00:27:36.659

Juan Luis Cano Rodríguez: popular ones.

152

00:27:36.920 --> 00:27:48.049

Juan Luis Cano Rodríguez: That's also, you know, issue activity in general pull requests coming in. So we try to take a heterogeneous approach to measuring.

153

00:27:48.300 --> 00:27:49.670

Juan Luis Cano Rodríguez: Popularity.

154

00:27:49.730 --> 00:27:53.879

Juan Luis Cano Rodríguez: And we have identified that some of the plugins there.

155

00:27:54.120 --> 00:28:01.990

Juan Luis Cano Rodríguez: let's say, outside of what the original Qv team maintains are really popular among the community.

156

00:28:05.210 --> 00:28:11.540

Michael Robinson | OpenLineage: Okay, thank you. I'd love to continue talking with you guys. Maybe

157

00:28:12.260 --> 00:28:16.969

Michael Robinson | OpenLineage: maybe in slack we could exchange information that'd be great

158

00:28:17.380 --> 00:28:18.400

Juan Luis Cano Rodríguez: absolutely.

159

00:28:29.180 --> 00:28:31.610

Lucy Hyde: Are there any other questions for the catch routine?

160

00:28:37.750 --> 00:28:38.990

Lucy Hyde: Alright.

161

00:28:40.660 --> 00:28:51.810

Lucy Hyde: that was a really cool presentation, and I'll connect with you, offline and and send you some stuff about for I'll link you up with Jen and some stuff about the ambassador program. So it looks like there's a really good opportunity there.

162

00:28:52.760 --> 00:28:59.880

Juan Luis Cano Rodríguez: Fantastic thanks, a lot, everyone. And this was a quick overview of the achievements. And but of course.

163

00:28:59.900 --> 00:29:11.499

Juan Luis Cano Rodríguez: Ketro is very much enjoyed when looking at it from a technical angle. So if anybody wants a deep dive on how it works and so on, feel free to reach out.

164

00:29:14.980 --> 00:29:16.699

Lucy Hyde: Excellent! Thank you.

165

00:29:17.470 --> 00:29:22.699

Lucy Hyde: Alright. So for our upcoming tech meetings.

166

00:29:22.800 --> 00:29:38.520

Lucy Hyde: We have graduation of Dockeray coming up next and then open schedule on March March seventh. Excuse me, we're always looking for special topics as well. So by all means, if you have a topic idea or agenda, please send it to the tag general list, and we will put you on the schedule

167

00:29:40.220 --> 00:29:51.559

Lucy Hyde: and then our attack meeting details, we have migrated to Lfx. If you experience any Lfx issues with accessing the meeting. Please reach out and we will file a book report.

168

00:29:52.880 --> 00:29:59.220

Mandy Chessell (egeria): And this is sorry, Lucy, this is this is Mandy. My meeting invites coming up an hour later.

169

00:29:59.480 --> 00:30:18.990

Lucy Hyde: Oh, joy! We usually see that sometimes with delay savings when it affects. But if you wouldn't mind either, just forwarding me the invite that shows it an hour later, or a screenshot. That would be wonderful.

170

00:30:19.130 --> 00:30:21.550

Mandy Chessell (egeria): Okay, alright, thanks

171

00:30:21.560 --> 00:30:29.540

Juan Luis Cano Rodríguez: also, Lucy. Sorry to interrupt. I saw open cable in the agenda for the march. Events. Is that

172

00:30:29.750 --> 00:30:46.360

Juan Luis Cano Rodríguez: Is that correct? Because I was not usually they manage the schedule and inserting inserting.

173

00:30:46.830 --> 00:30:58.789

Lucy Hyde: And that wraps up our meeting today. as anything. If you have any questions or need anything, please feel free to reach out to me, and more than happy to me, or to chat via email or slack

174

00:30:59.200 --> 00:31:02.599

Lucy Hyde: And I hope that everyone has a wonderful Thursday.

175

00:31:06.200 --> 00:31:10.789

Juan Luis Cano Rodríguez: Thanks everyone. Bye, bye.

