

Technical Advisory Council Meeting

August 27, 2020

 THE **LINUX** FOUNDATION

 LF AI

Antitrust Policy Notice

- › Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.
- › Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Undergone LLP, which provides legal counsel to the Linux Foundation.

Recording of Calls

Reminder:

TAC calls are recorded and available for viewing on the [TAC Wiki](#)

Reminder: LF AI Useful Links

Web site: lfai.foundation
Wiki: wiki.lfai.foundation
GitHub: github.com/lfai
Landscape: landscape.lfai.foundation or l.lfai.foundation
Mail Lists: <https://lists.lfai.foundation>

LF AI Logos: <https://github.com/lfai/artwork/tree/master/lfai>
LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing

Events Page on LF AI Website: <https://lfai.foundation/events/>
Events Calendar on LF AI Wiki (subscribe available):
<https://wiki.lfai.foundation/pages/viewpage.action?pageId=12091544>
Event Wiki Pages: <https://wiki.lfai.foundation/display/DL/LF+AI+Foundation+Events>

Agenda

- › Roll Call
- › Approval of Minutes
- › ModelDB Presentation
- › ODPi Presentation
- › Upcoming TAC Meetings
- › Open Discussion

TAC Voting Members

Member	Contact	Email
AT&T	Reuben Klein	rk1518@att.com
Baidu	Daxiang Dong	dongdaxiang@baidu.com
Ericsson	Rani Yadav-Ranjan	rani.yadav-ranjan@ericsson.com
Huawei	Huang Zhipeng	huangzhipeng@huawei.com
Nokia	Pantelis Monogioudis	pantelis.monogioudis@nokia.com
Tech Mahindra	Nikunj Nirmal	nn006444@techmahindra.com
Tencent	Bruce Tao	brucetao@tencent.com
Zilliz	Jun Gu	jun.gu@zilliz.com
ZTE	Wei Meng	meng.wei2@zte.com.cn
Acumos AI Project	Nat Subramanian	natarajan.subramanian@techmahindra.com
Angel Project	Bruce Tao	brucetao@tencent.com
ONNX Project	Jim Spohrer*	spohrer@us.ibm.com

* TAC Chairperson

Approval of Minutes

Draft minutes from the August 13th meeting of the TAC were previously distributed to the TAC members

Proposed Resolution:

- › That the minutes of the August 13th meeting of the Technical Advisory Council of the LF AI Foundation are hereby approved

Invited Presentation

Model Versioning with ModelDB

Conrado Miranda
CTO, Verta

Model versioning with ModelDB



Conrado Miranda, CTO at Verta



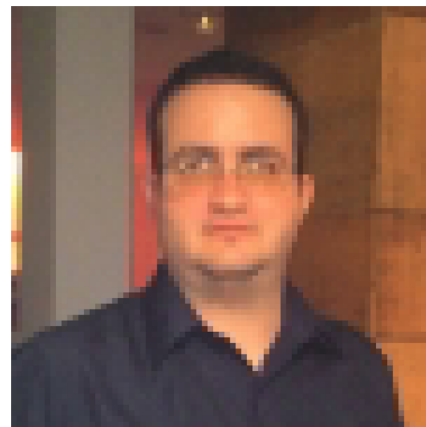
About us



Manasi Vartak
CEO, Verta

MIT CSAIL PhD
ModelDB Creator,
Leader, and Evangelist

ML @ Twitter, Google,
Facebook



Conrado Miranda
CTO, Verta

Engineering Manager
NVIDIA Drive Platform

Tech Lead, Cortex Team
Twitter's model training &
serving infra

Background

- Manasi created ModelDB, the first model management system
- ModelDB focused on tracking and experimentation
- Customers want reproducibility. Reproducibility requires versioning
- Model versioning *was* an open problem!
- ModelDB 2.0 re-built focused on versioning first

Why didn't the industry actually had versioning?

- Software analogy
 - You log that the code was built on Linux
 - You log that the compiler had an argument “-DENABLE_LOG”
 - You log that the test coverage is 95%
 - You save the compiled binary
- Do you have software versioning?

Why didn't the industry actually had versioning?

- Software analogy
 - You log that the ~~code was built on Linux~~ model was built with PyTorch
 - You log that the ~~compiler had an argument "-DENABLE_LOG"~~ you used SGD optimize
 - You log that the ~~test coverage is 95%~~ the precision is 95%
 - You save the ~~compiled binary~~ trained model
- Do you have ~~software~~ model versioning?

Ingredients for reproducibility

- Code: the actual code to perform the training (Git, Jupyter notebooks)
- Dataset: the data used to feed the model (S3, DB, CSV)
- Configuration: the configuration to your code (hyperparameters)
- Environment: the setup around the code (Docker, Conda, hardware)

- If you don't save something, you cannot reproduce the model.
- If you can't go back in time, you cannot reproduce the model.

How to do model versioning?

- Git is the gold standard for code versioning, but it has drawbacks
- ML needs to scale. Think 100k branches on every experiment
 - Even git-first systems like DVC had to add flexibility of using other storage [\[1\]](#)
- ML needs to be discoverable
 - Imagine not being able to navigate your source code
- ML types have strong meaning
 - Diff and merges between ingredients is more than just JSON files

ModelDB versioning

- API-first design with horizontally scalable service
 - CLI, Python and Scala clients, WebApp interface
- Git-like versioning for all ingredients and ML steps
 - Bring your own versioned repository when appropriate
- ML ingredients are first class
 - Specialized logic for diffs, merges and search for each type

Quick demo

- We'll use our [basic tutorial](#) as reference
- Get in touch to schedule a more complete demo later with more time!

Get in touch!

- We'll help you with open source setup or enterprise solution. Reach out in [Slack](#) or hello@verta.ai
- We love making data scientists more productive. Reach out with requests or to share a story with conrado@verta.ai
- We have a technical series on versioning coming up! [Stay tuned](#) and signup for our [newsletter](#)

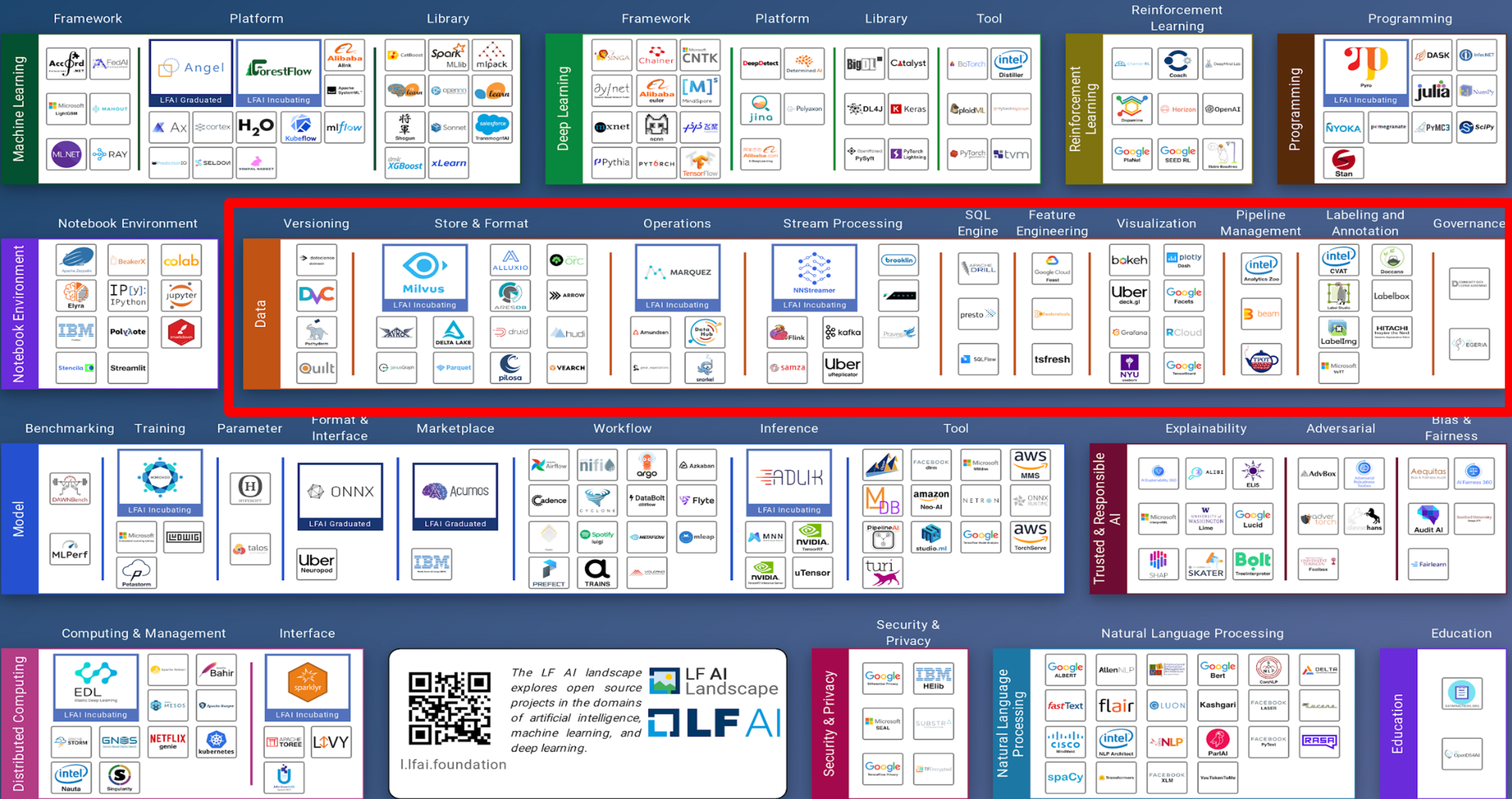
Invited Presentation


John Mertic


ODPi

ODPi Projects
Overview
John Mertic
Director, ODPi





 The LF AI landscape explores open source projects in the domains of artificial intelligence, machine learning, and deep learning.



l.fai.foundation

ODPi Project Portfolio



Incubation

Effort to bring together thought leaders across the industry for vendor neutral guidance on the intersection of BI and AI

<https://www.odpi.org/projects/bi-ai>



Graduated

Egeria is the world's first open source metadata standard. It provides open APIs, event formats, types and integration logic so organizations can share data management and governance across the entire enterprise without reformatting or restricting the data to a single format, platform, or vendor product.

<https://www.odpi.org/projects/egeria>



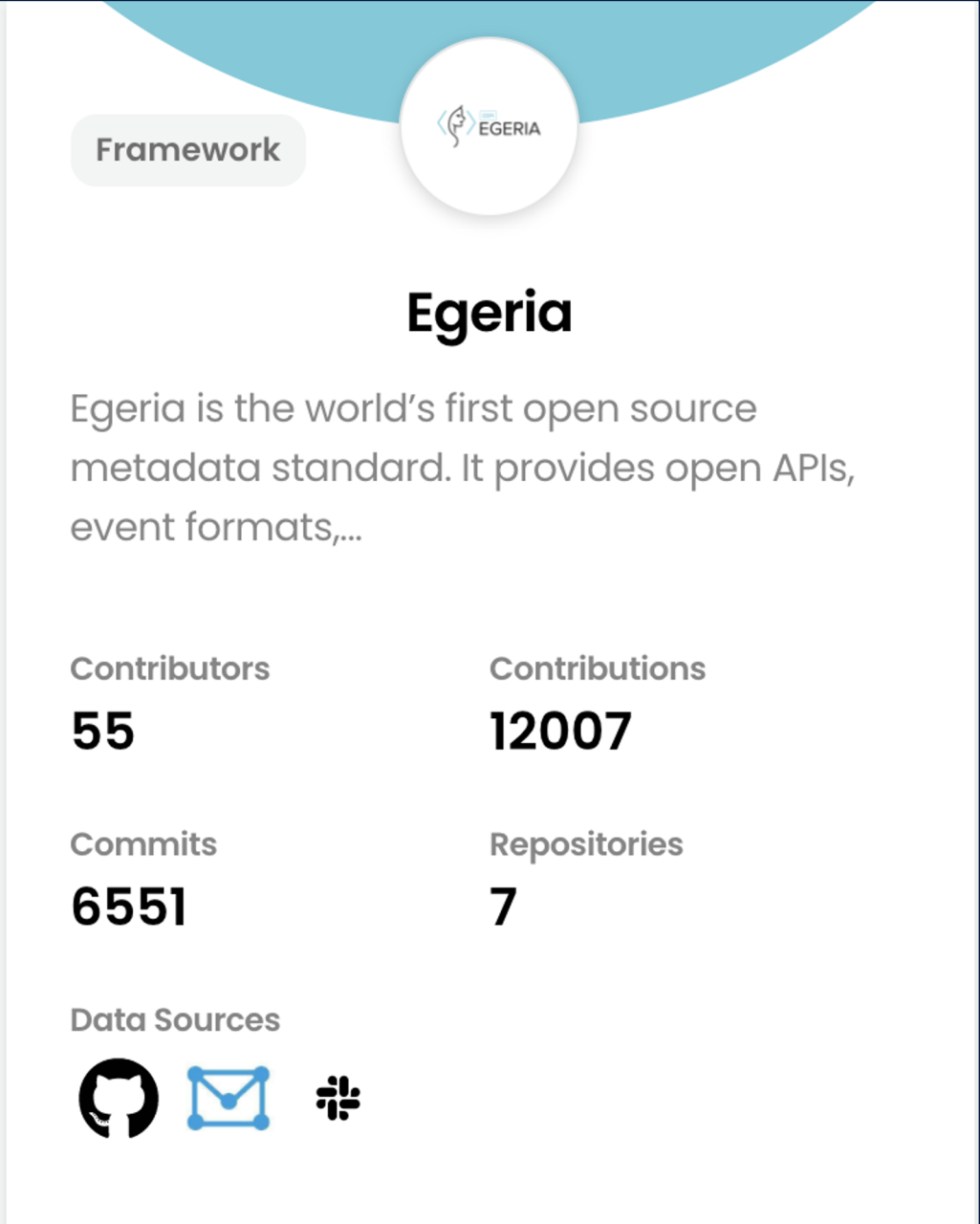
Incubation

OpenDS4All is an open source project built to accelerate the creation of data science curricula at academic institutions.

<https://www.odpi.org/projects/egeria>

Egeria

- ODPi Graduate level project
- Launched in August 2018
- Major releases:
 - 1.0 -- January 2019
 - 2.0 -- June 2020
- Current maintainers from IBM, ING Bank, and SAS
- Other contributors Expedia Group, GCHQ, and OCRV (Russia) having contributed over time.
- Deliverables: Apache 2 licensed integration service and APIs, as well as best practices for metadata management in an organization



The image shows a project overview card for Egeria. At the top, there is a teal curved header with a circular logo containing the Egeria icon and the word 'EGERIA'. Below the logo, the word 'Framework' is written in a grey rounded box. The main title 'Egeria' is in a large, bold, black font. A short description follows: 'Egeria is the world's first open source metadata standard. It provides open APIs, event formats,...'. Below this, there are two columns of statistics. The first column lists 'Contributors' with a value of '55'. The second column lists 'Contributions' with a value of '12007'. Below these, the first column lists 'Commits' with a value of '6551' and the second column lists 'Repositories' with a value of '7'. At the bottom, there is a section for 'Data Sources' with three icons: GitHub, a blue envelope icon, and a black multi-dot icon.




Framework

Egeria

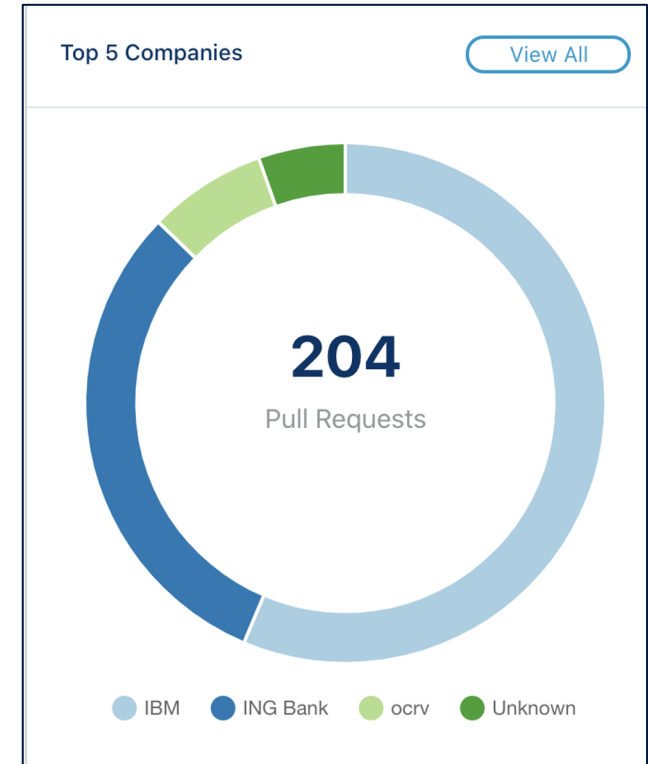
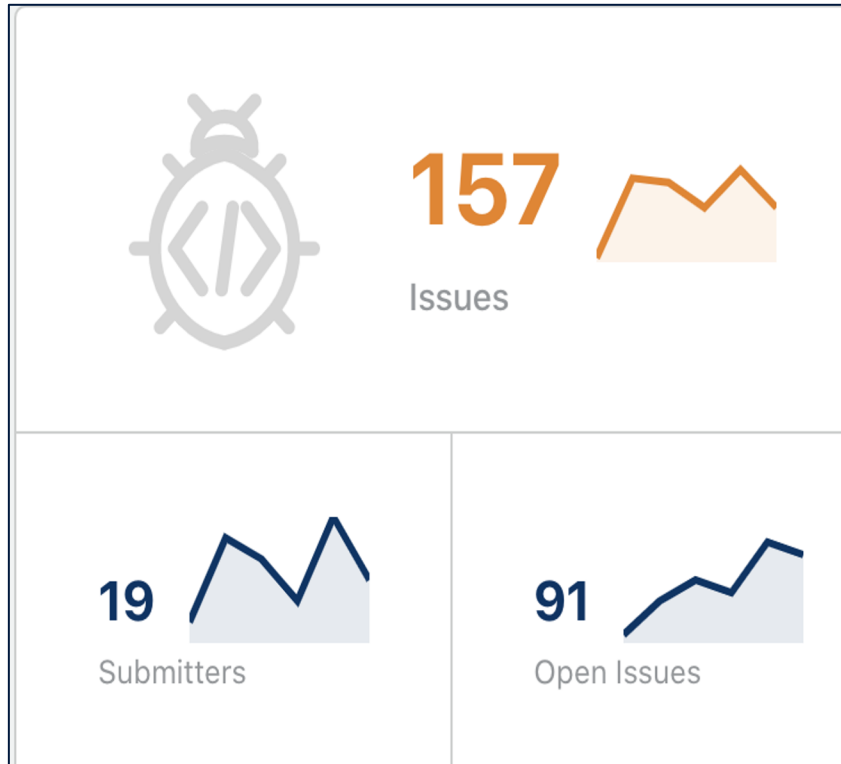
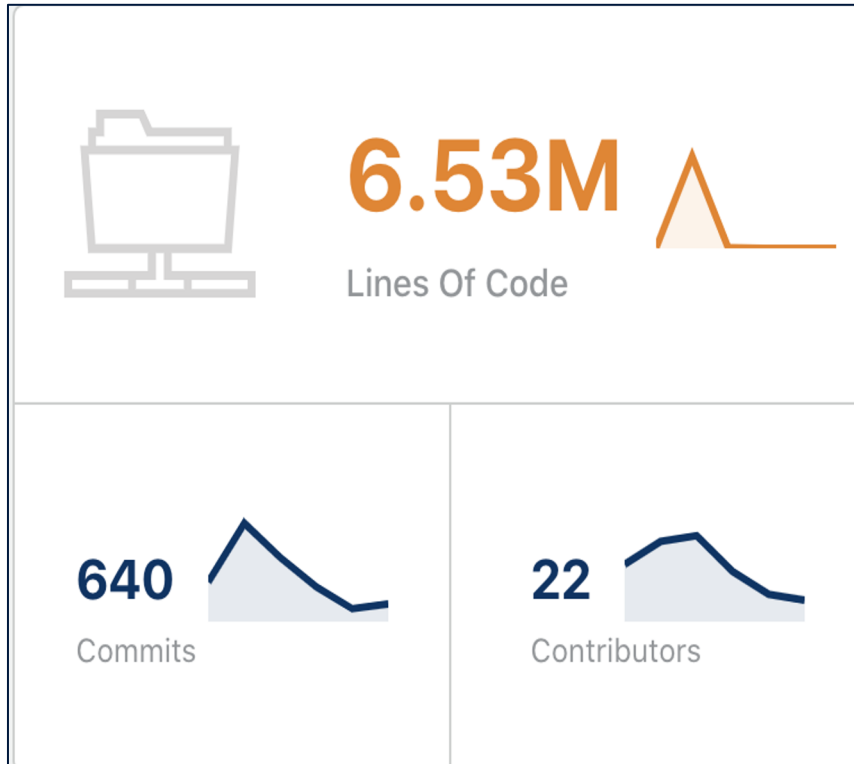
Egeria is the world's first open source metadata standard. It provides open APIs, event formats,...

Contributors	Contributions
55	12007
Commits	Repositories
6551	7

Data Sources



Egeria dev analytics



Egeria Conformance Program

Launched in January 2019, the program enables vendors to showcase support for their products able to integrate in with other Egeria Conformant offerings.

Integrations patterns include:

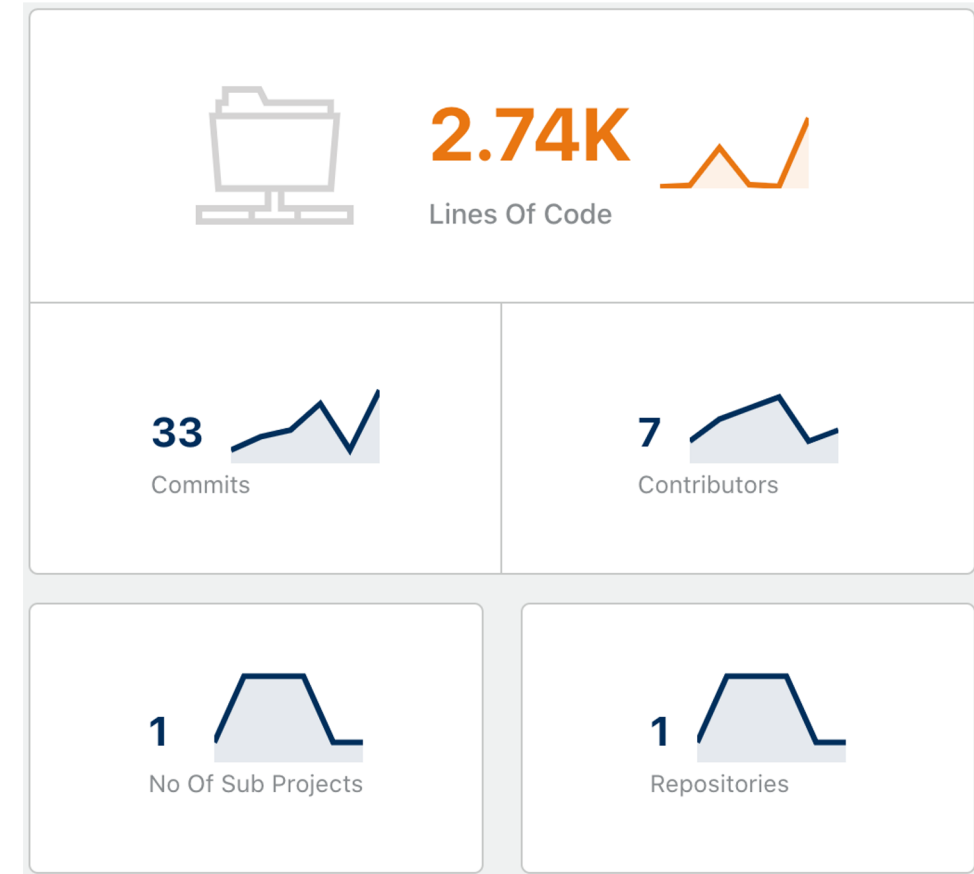
- Using Egeria as middleware between metadata repositories and other data tools
- Integrating Egeria as an API in a metadata repository
- Implementing the Egeria APIs in a vendor's product

No current solutions, but ones on the way from Cloudera, IBM, and SAS in their roadmaps.



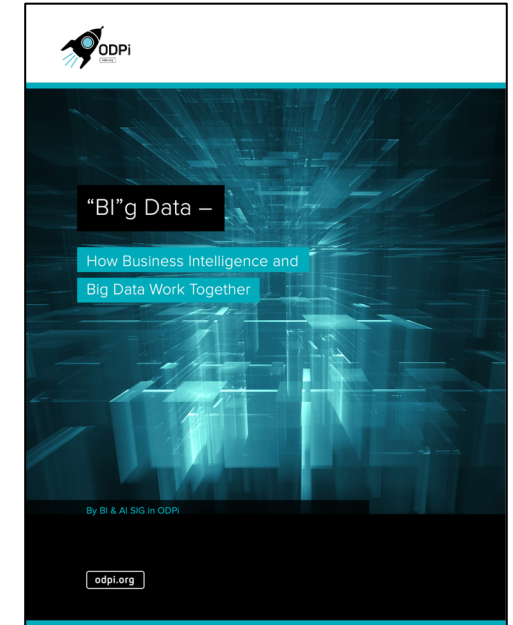
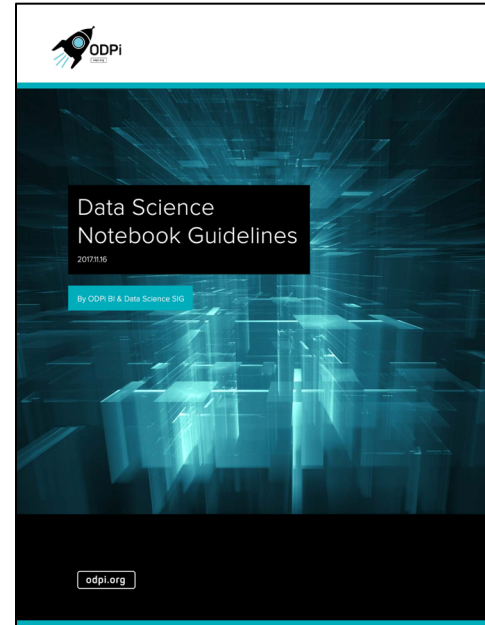
OpenDS4All

- › Joined as an Incubation Project in June 2019
- › Announced initial release February 2020
- › Initial contributions from IBM and University of Pennsylvania, with SAS and Oklahoma State University additionally engaging
- › Deliverables are a “starter set” of academic materials for that any higher education institution could use for a data science course.



BI and AI

- Incubation project accepted in early 2019
- Collaboration amongst thought leaders from Cloudera, IBM, Index Analytics, SAS, and more data vendors that delivers annual publications on topics related to BI, AI, and Big Data.
- 2020 focus is a "Bridge" initiative, allowing BI tools connecting to an AI training model to see the progress and results updating real time across multiple algorithms. Discussion at <https://lists.odpi.org/g/odpi-project-bi-ai/message/82>



Learn more about ODPi

- <https://odpi.org>
- ODPi Github - <https://github.com/odpi>
- Slack - <https://slack.odpi.org>

LF AI General Updates

OLFAI


















PREMIER
MEMBER

IBM®

Members

Premier	 AT&T	 Bai 百度	 ERICSSON	 HUAWEI				
	 NOKIA	 Tech Mahindra	 Tencent 腾讯	 ZILLIZ Reinvent Data Science				
General	 DiDi	 inwinSTACK	 orange	 redhat	 R Studio	 XENONSTACK		
Associate	 AI for people	 AMBIANT			 MAIEI	 NYU	 PennState Great Valley	

Projects

 Acumos ★ 10 LF Artificial Intelligence Foundation	 Angel ★ 5,927 LF Artificial Intelligence Foundation	 ONNX ★ 8,782 LF Artificial Intelligence Foundation		
 Adlik ★ 120 LF Artificial Intelligence Foundation	 Adversarial Robustness Toolkit ★ 1,641 LF Artificial Intelligence Foundation	 AI Explainability 360 Toolkit ★ 604 LF Artificial Intelligence Foundation	 AI Fairness 360 Toolkit ★ 1,022 LF Artificial Intelligence Foundation	 Amundsen ★ 1,106 LF Artificial Intelligence Foundation
 EDL Elastic Deep Learning ★ 91 LF Artificial Intelligence Foundation	 ForestFlow ★ 35 LF Artificial Intelligence Foundation	 Horovod ★ 9,687 LF Artificial Intelligence Foundation	 Ludwig ★ 6,888 LF Artificial Intelligence Foundation	 Marquez ★ 331 LF Artificial Intelligence Foundation
 Milvus ★ 3,791 LF Artificial Intelligence Foundation	 NNStreamer ★ 258 LF Artificial Intelligence Foundation	 Pyro ★ 6,352 LF Artificial Intelligence Foundation	 sparklyr ★ 744 LF Artificial Intelligence Foundation	

A Growing Community

- 17 projects, 77 repos
- Over 7.5 M LoC
- 1120+ active developer
- 50 known organizations contributing code
- 22 universities contributing code

* Data as of 08/21/2020

Companies hosting projects in LF AI

<https://landscape.lfai.foundation/format=hosting>



Looking to host a project with LF AI

Hosted project stages and life cycle: <https://lfai.foundation/project-stages-and-lifecycle/>

Offered services for hosted projects:
<https://lfai.foundation/services-for-projects/>

Contact:

Jim Spohrer (TAC Chair) and Ibrahim Haddad (ED, LF AI)

Promoting Upcoming Project Releases

We promote project releases via a blog post and on LF AI [Twitter](#) and/or [LinkedIn](#) social channels

For links to details on upcoming releases for LF AI hosted projects visit the [Technical Project Releases wiki](#)

If you are an LF AI hosted project and would like LF AI to promote your release, reach out to pr@lfai.foundation to coordinate in advance (min 2 wks) of your expected release date.

Note on quorum

As LF AI is growing, we now have 12 voting members on the TAC.

TAC representative - please ensure you attend the bi-weekly calls or email Jacqueline/Ibrahim to designate an alternate representative when you can not make it.

We need to ensure quorum on the calls especially when we have items to vote on.

Updates from the Outreach Committee

LF AI PR/Comms

- › Please follow LF AI on [Twitter](#) & [LinkedIn](#) and help amplify news via your social networks - Please retweet and share!
 - › Also watch for news updates via the tac-general mail list
 - › View recent announcement on the [LF AI Blog](#)
- › Open call to publish project/committee updates or other relevant content on the [LF AI Blog](#)
- › To discuss more details on participation or upcoming announcements, please email pr@lfai.foundation

Call to Participate in Ongoing Efforts

Trusted AI

- › **Leadership:**
Animesh Singh (IBM), Souad Ouali (Orange), and Jeff Cao (Tencent)
- › **Goal:** Create policies, guidelines, tooling and use cases by industry
- › **Github:**
<https://github.com/lfai/trusted-ai>
- › **Wiki:**
<https://wiki.lfai.foundation/display/DL/Trusted+AI+Committee>
- › **To participate:**
<https://lists.lfai.foundation/g/trustedai-committee/>
- › **Next call:** Bi-weekly on Thursdays at 7am PT, subscribe to group calendar on wiki
<https://wiki.lfai.foundation/pages/viewpage.action?pageld=12091895>

ML Workflow & Interop

- › **Leadership:**
Huang “Howard” Zhipeng (Huawei)
- › **Goal:**
Define an ML Workflow and promote cross project integration
- › **Wiki:**
<https://wiki.lfai.foundation/display/DL/ML+Workflow+Committee>
- › **To participate:**
<https://lists.lfai.foundation/g/mlworkflow-committee>
- › **Next call:** Every 4 weeks on Thursdays at 7:00 am PT, subscribe to group calendar on wiki
<https://wiki.lfai.foundation/pages/viewpage.action?pageld=18481242>

Launching an effort to create AI Ethics Training

Initial developed course by the LF: Ethics in AI and Big Data - published on edX platform:

<https://www.edx.org/course/ethics-in-ai-and-big-data>

The goal is to build 2 more modules and package all 3 as a professional certificate - a requirement for edX

- > The LF would cover the cost of the production and promotion
- > The course would be offered for free
- > The credit of the course will go to content creator and their organizations
- > Initial interested parties: IBM, AI for People, Montreal AI Ethics Institute, Ethical ML Institute
- > **To participate:**
<https://lists.lfai.foundation/g/aiethics-training>

Upcoming TAC Meetings

Upcoming TAC Meetings

- › **September 10:** Incubation projects voting
 - › [Delta project](#) – Yunbo Wang (Didi), Incubation proposal and vote
 - › [SOAJS project](#) – Antoine Hage, Incubation proposal and vote (tentative)

- › **September 24 :**
 - › **TBD**

Please send agenda topic requests to tac-general@lists.lfai.foundation

TAC Meeting Details

- › To subscribe to the TAC Group Calendar, visit the wiki: <https://wiki.lfai.foundation/x/XQB2>
- › Join from PC, Mac, Linux, iOS or Android: <https://zoom.us/j/430697670>
- › Or iPhone one-tap:
 - › US: +16465588656,,430697670# or +16699006833,,430697670#
- › Or Telephone:
 - › Dial(for higher quality, dial a number based on your current location):
 - › US: +1 646 558 8656 or +1 669 900 6833 or +1 855 880 1246 (Toll Free) or +1 877 369 0926 (Toll Free)
- › Meeting ID: 430 697 670
- › International numbers available: <https://zoom.us/u/achYtcw7uN>

Open Discussion

Legal Notices

- › The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.
- › Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.
- › Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.
- › TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.
- › Facebook and the "f" logo are trademarks of Facebook or its affiliates.
- › LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.
- › YouTube and the YouTube icon are trademarks of YouTube or its affiliates.
- › All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.
- › The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>. each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.
- › Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.