Meeting of the LF AI & Data Technical Advisory Council (TAC)

December 15, 2022

LF AI & DATA

Antitrust Policy

› 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

LF AI & DATA

Reminder: LF AI & Data Useful Links

› Web site: lfaidata.foundation

› Wiki: wiki.lfaidata.foundation

› GitHub: github.com/lfaidata

› Landscape: https://landscape.lfaidata.foundation or https://l.lfaidata.foundation

› Mail Lists: https://lists.lfaidata.foundation

› Slack: https://slack.lfaidata.foundation

› Youtube: https://www.youtube.com/channel/UCfasaeqXJBCAJMNO9HcHfbA

› LF AI Logos: https://github.com/lfaidata/artwork/tree/master/lfaidata

› LF AI Presentation Template: https://drive.google.com/file/d/1eiDNJvXCqSZHT4Zk_-czASlz2GTBRZk2/view?usp=sharing

› Events Page on LF AI Website: https://lfaidata.foundation/events/

› Events Calendar on LF AI Wiki (subscribe available): https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=12091544

› Event Wiki Pages: https://wiki.lfaidata.foundation/display/DL/LF+AI+Data+Foundation+Events

Agenda

› Roll Call (2 mins)

› Approval of Minutes from previous meeting (2 mins)

› OpenLineage moving from Sandbox to Incubation (40 min)

› LF AI General Updates (2 min)

› Open Discussion (2 min)

TAC Voting Members - Please note

Please ensure that you do the following to facilitate smooth procedural quorum and voting processes:

- Change your Zoom display name to include your First/Last Name, Company/Project Represented
    - example: Nancy Rausch, SAS
- State your First/Last Name and Company/Project when submitting a motion
    - example: First motion, Nancy Rausch/SAS

LF AI & DATA

TAC Voting Members - Please note

› TAC members must attend consistently to maintain their voting status

› After 2 absences voting members will lose voting privileges

› Voting privileges will only be reinstated after attending 2 meetings in a row

TAC Voting Members

Note: we still need a few designated backups specified on [wiki](#)

Member Representatives (8 out of 16 required for quorum)

Member Company or Graduated Project	Membership Level or Project Level	Voting Eligibility	Country	TAC Representative	Designated TAC Representative Alternates
4paradigm	Premier	Voting Member	China	Zhongyi Tan	
Baidu	Premier	Voting Member	China	Ti Zhou	Daxiang Dong, Yanjun Ma
Ericsson	Premier	Voting Member	Sweden	Rani Yadav-Ranjan	
Huawei	Premier	Voting Member	China	Howard (Huang Zhipeng)	Charlotte (Xiaoman Hu) , Leon (Hui Wang)
Nokia	Premier	Voting Member	Finland	@ Michael Rooke	@ Jonne Soininen
OPPO	Premier	Voting Member	China	Jimmy (Hongmin Xu)	
SAS	Premier	Voting Member	USA	*Nancy Rausch	JP Trawinski
ZTE	Premier	Voting Member	China	Wei Meng	Liya Yuan
Adversarial Robustness Toolbox Project	Graduated Technical Project	Voting Member	USA	Beat Buesser	
Angel Project	Graduated Technical Project	Voting Member	China	Bruce Tao	Huaming Rao
Egeria Project	Graduated Technical Project	Voting Member	UK	Mandy Chessell	Nigel Jones, David Radley, Maryna Strelchuk, Ljupcho Palashevski, Chris Grote
Flyte Project	Graduated Technical Project	Voting Member	USA	Ketan Umare	
Horovod Project	Graduated Technical Project	Voting Member	USA	Travis Addair	
Milvus Project	Graduated Technical Project	Voting Member	China	Xiaofan Luan	Jun Gu
ONNX Project	Graduated Technical Project	Voting Member	USA	Alexandre Eichenberger	Prasanth Pulavarthi, Jim Spohrer
Pyro Project	Graduated Technical Project	Voting Member	USA	Fritz Obermeyer	

*Current TAC Chairperson

☐LF AI & DATA

Minutes approval

Approval of November 17, 2022 Minutes

Draft minutes from the November 17 TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

› That the minutes of the November 17 meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

LF AI & DATA

Approval of December 1, 2022 Minutes

Draft minutes from the December 1 TAC call were previously distributed to the TAC members via the mailing list

Proposed Resolution:

› That the minutes of the December 1 meeting of the Technical Advisory Council of the LF AI & Data Foundation are hereby approved.

Background

› **December 2020:** open sourced
› **May 2021:** joined the LFAI
  › 300 GitHub stars
  › Collaboration with Marquez
  › Sandbox
› **Today:**
  › 1.1K GitHub stars
  › Collaborations with Microsoft, Snowflake, Airflow, Egeria, Northwestern Mutual, Amundsen
  › Ready for Incubation

Adoption
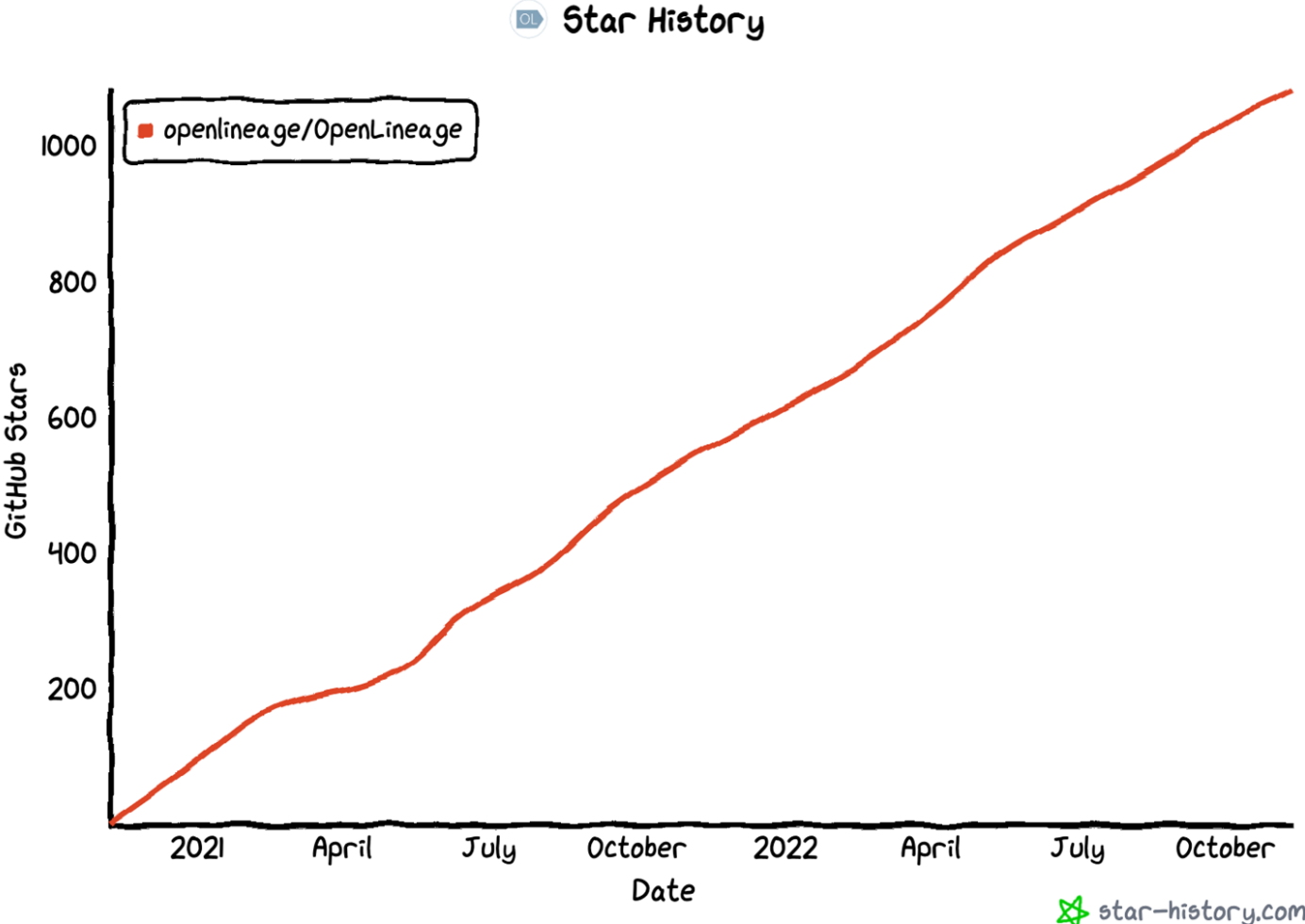
Adoption since 2021

	November 2022	December 2021	Change
Stars	1.1K	600	**+83%**
Forks	128	44	**+190%**

Progress since 2021

- **Contributors**
  - › 184 total contributors (Microsoft, Astronomer, HSBC, UBS, Northwestern Mutual, NTT Data Deutschland, others)
- **Best Practices**
  - › Core Infrastructure Best Practices Silver Badge (Nov. 2022)
- **Commits**
  - › 2.09K, averaging ~44 per week in the past 11 months
- **Governance and Procedures**
  - › TSC, Contributor Guide, Charter, CoC, Code Quality Assurance
- **LFAI Collaborations**
  - › Egeria, Marquez, Amundsen

LF AI & DATA

Top Contributors

Organization	Commits	LOC
Datakin	629	90,650
GetInData sp. z o.o. sp. k.	456	101,205
Unknown	364	10,696
Datakin Inc	238	19,773
Individual - No Account	235	65,537
UBS AG	44	28,127
Astronomer	38	3,036
Microsoft Corporation	9	724
HSBC	8	1,361
NTT DATA Deutschland GmbH	8	709
Fairtide Pte. Limited	5	29
Northwestern Mutual Life Insurance Company	4	3,600
Societe Generale Corporate and Investment Banking	2	0
Neal Analytics, LLC	1	143
Palantir Technologies Inc.	1	0
Six Group AG	1	65

Top Contributors

Commits

- Astronomer 1.9%
- UBS AG 2.2%
- Individual - No Account 11.5%
- Datakin Inc 11.6%
- Unknown 17.8%
- Datakin 30.8%
- GetInData sp. z o.o.... 22.3%
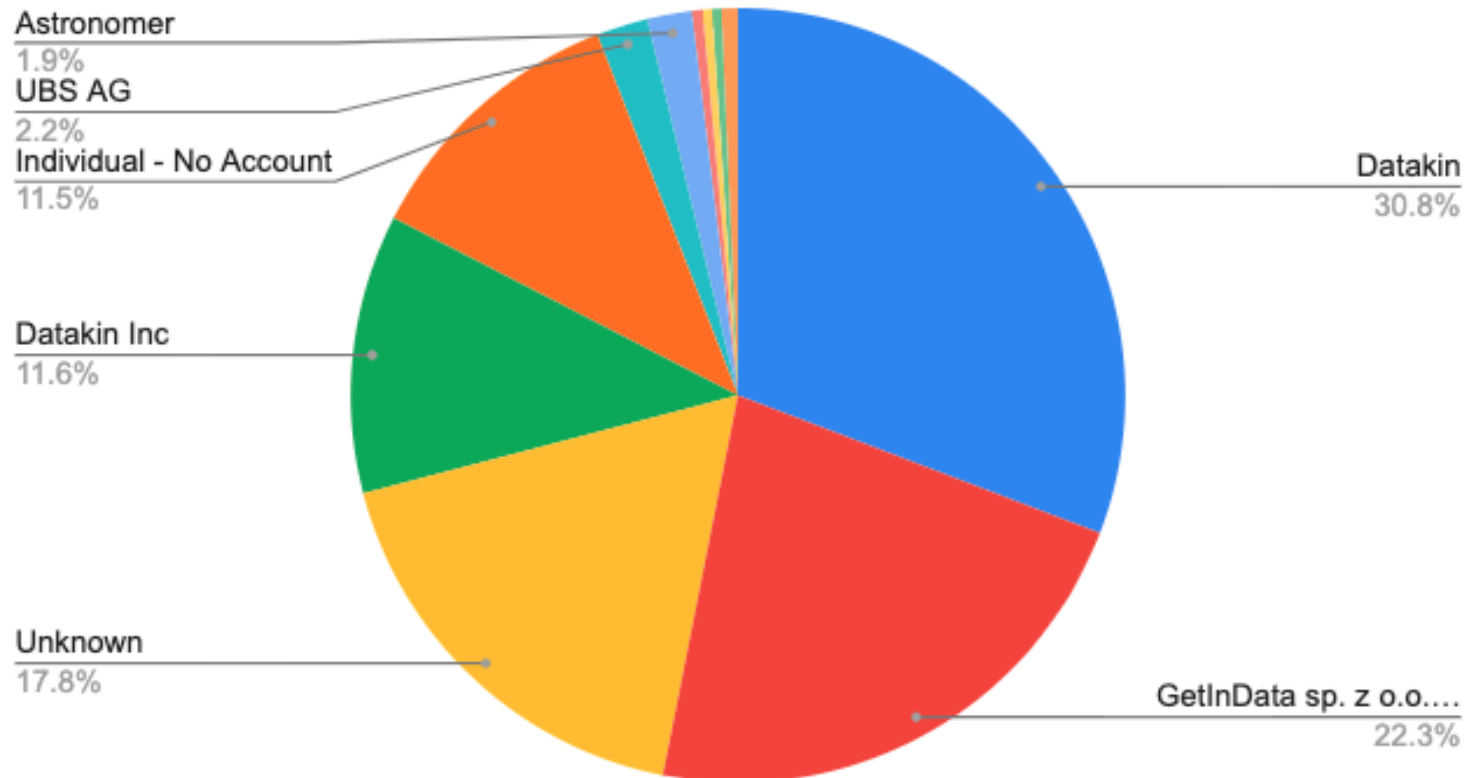
Top Contributors

LOC

- Astronomer 0.8%
- Datakin Inc 4.7%
- UBS AG 4.9%
- Individual - No Account 17.4%
- Unknown 19.7%
- GetInData sp. z o.o.... 30.0%
- Datakin 21.1%

15DEC2022
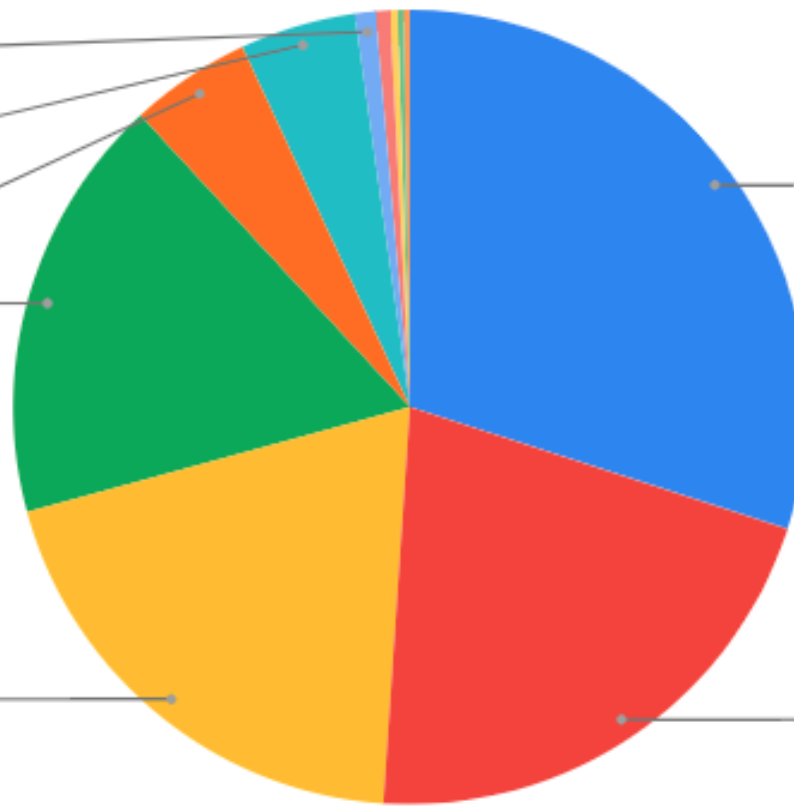
Top Contributors

	November 2022	December 2021	Change
% Commits outside Astronomer/Datakin/GiD	35	14.5	**+141%**
% LOC outside Astronomer/Datakin/GiD	35	13	**+169%**
Commits outside Astronomer/Datakin/GiD	720	15	**+4700%**
LOC outside Astronomer/Datakin/GiD	114,027	17,153	**+565%**

Notable Collaborations

› Spark integration (Microsoft)

› Snowflake adapter (Snowflake)

› Egeria integration (Egeria)

› Amundsen table lineage extractor (Amundsen)

› Dagster integration (Northwestern Mutual)

› OpenLineage scanner (in use by MANTA)

› Rust parser interface for SQL integration (supported by Microsoft)

› Interactive developer env, default extractor, extractors for SQL check operators (Airflow)

Best Practices Badge

Open SSF Best Practices
Core Infrastructure Silver Badge
Earned November 2022

Contributions over Time

Contributions over Time

Additions and Deletions per week

15DEC2022
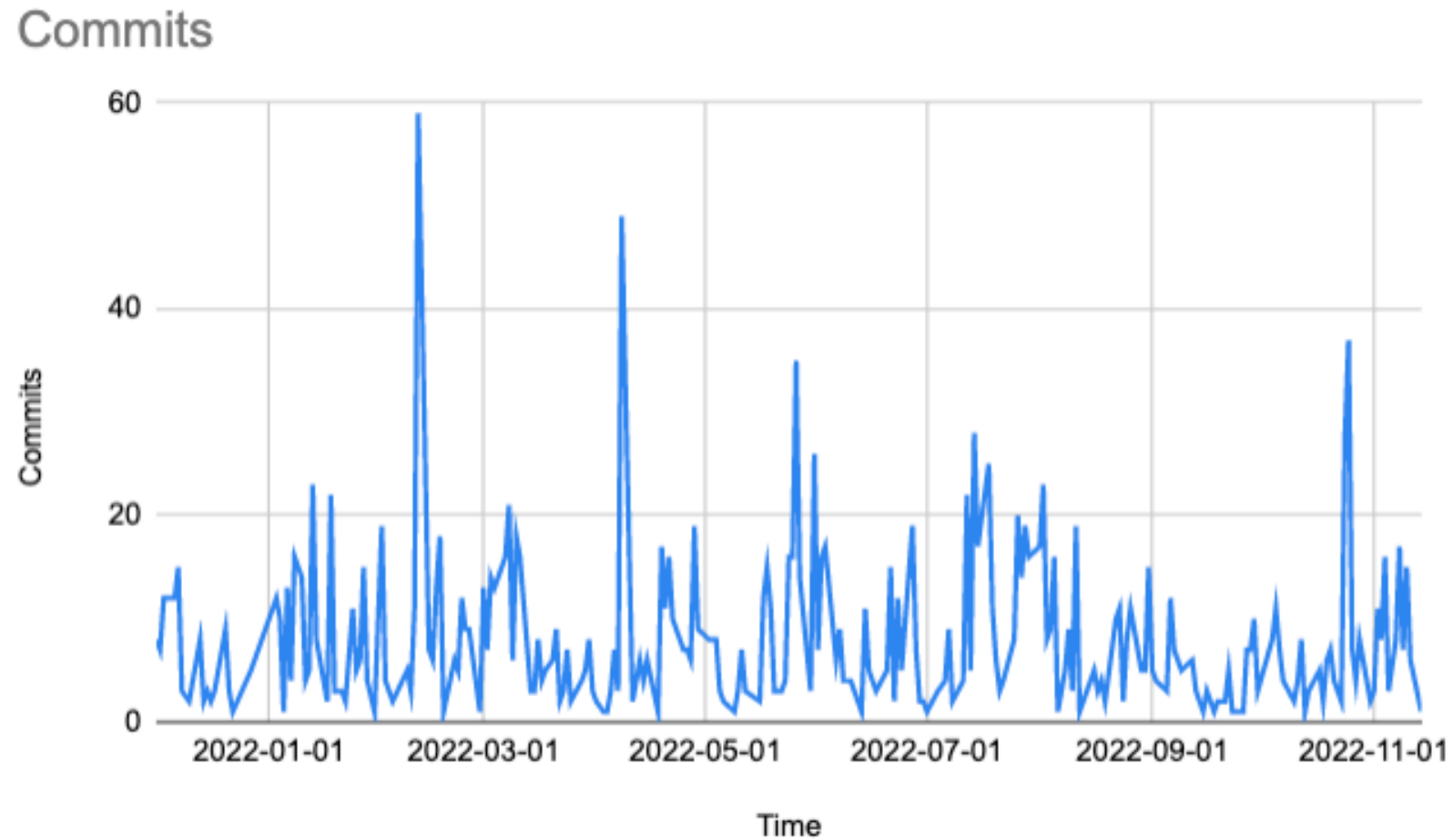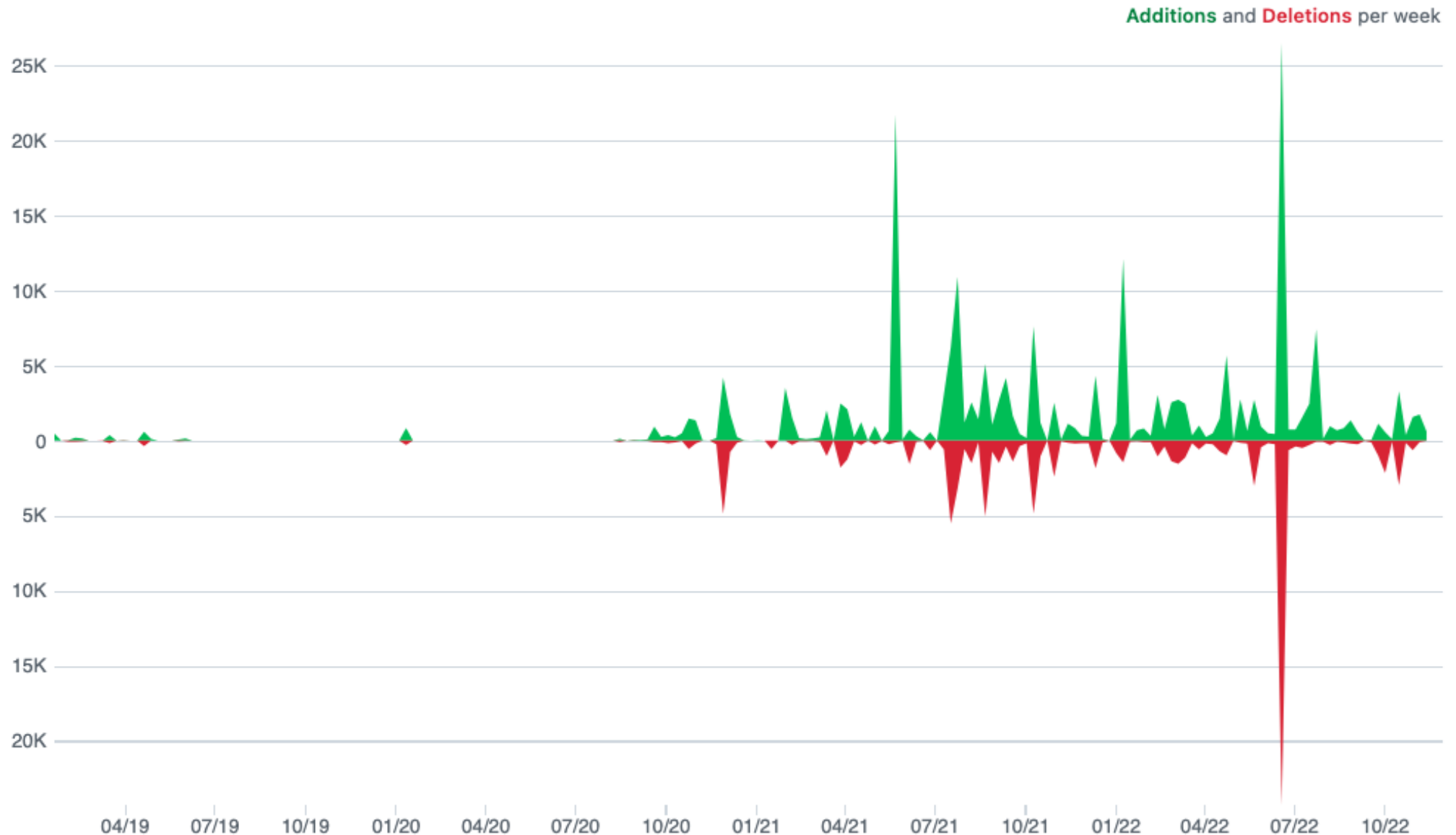
Governance and Procedures

› **Governance document**
https://github.com/OpenLineage/OpenLineage/blob/main/GOVERNANCE.md:
  › Release authorization process
  › Committer election process
  › Leadership structure
› **Technical Steering Committee document**
https://github.com/OpenLineage/OpenLineage/blob/main/TECHNICAL_STEERING_COMMITTEE.md
› **Code of Conduct document**
https://github.com/OpenLineage/OpenLineage/blob/main/CODE_OF_CONDUCT.md
› **Code Quality Assurance document**
https://github.com/OpenLineage/OpenLineage/blob/main/CODE_QUALITY_AND_SECURITY.md
› **Releasing document containing release process**
https://github.com/OpenLineage/OpenLineage/blob/main/RELEASING.md
› **New contributor guide**
https://github.com/OpenLineage/OpenLineage/blob/main/CONTRIBUTING.md

Technical Steering Committee

Chairperson: **Julien Le Dem** (Astronomer, Marquez, Apache Arrow, Apache Parquet)

Voting members:

Mandy Chessell (Egeria, Pragmatic Data Research, Ltd.)	**Maciej Obuchowski** (GetInData, Marquez)
Daniel Henneberger (prev Datakin)	**Paweł Leszczyński** (GetInData, Marquez)
Drew Banin (dbt Labs)	**Will Johnson** (Microsoft)
James Campbell (Superconductive)	**Michael Robinson** (Astronomer, Marquez)
Ryan Blue (Apache Iceberg)	**Tomasz Nazarewicz** (GetInData, Marquez)
Willy Lulciuc (Astronomer, Marquez)	**Jakub Dardziński** (GetInData, Marquez)
Zhamak Dehghani (Stealth)	**Howard Yoo** (Astronomer, Marquez)
Ross Turk (Astronomer, Marquez)	

LFAI Collaborations

› **OpenLineage + Egeria**
  › Completed direct integration
  › Egeria's integration daemon supports the OpenLineage API for local processing engines.
  › Egeria contributed the OpenLineage proxy backend in November 2021.
› **OpenLineage + Marquez**
  › Ongoing direct integration
  › Marquez is the reference implementation of the OpenLineage standard, acting as the backend to receive OpenLineage events from all the processing engines OpenLineage supports. Marquez also visualizes the lineage graph in map form and offers analytical data in its UI.
› **OpenLineage + Amundsen**
  › Completed integration
  › Amundsen's `OpenLineageTableLineageExtractor` extracts table lineage information from OpenLineage events.

Incubation Summary

› Joined as Sandbox project, May 2021

› Many unique contributors (184)

› Numerous involved partners (Microsoft, Northwestern Mutual, Snowflake, Airflow, etc.)

› High GitHub star count (1.1K)

› Substantial flow of commits (44 per week avg)

› CII Silver badge

› Collaborations within the LFAI (Egeria, Marquez, Amundsen)

› TSC members from Astronomer, dbt, Superconductive, etc.

Please vote to approve OpenLineage for Incubation status in the LFAI.

□LF AI & DATA

2023 Roadmap

› Improve the SQL parser
› Add job and run facets to the Flink integration
› Add job/run dependencies facet to Airflow integration
› Add DAG run listener to Airflow integration
› Add ability to annotate datasets outside job runs
› Add facets validator to Python client
› Extend column lineage beyond Spark integration
› Automate updating of docs
› Add dbt Cloud support to Airflow integration
› Improve OpenAPI code generation
› Add support for Apache Atlas
› Add facet to report DAG metadata

Formal request

The authors hereby formally request the graduation of OpenLineage from Sandbox to Incubation in LF AI & Data

TAC Open Discussion

LF AI & DATA

LF AI General Updates

Upcoming TAC Meetings

Upcoming TAC Meetings

› December 29, 2022 – Holiday break, no meeting

› January 12, 2023 – TonY annual review, Angel annual review (tentative)

Please note we are always open to special topics as well.

If you have a topic idea or agenda item, please send agenda topic requests to
tac-general@lists.lfaidata.foundation

LF AI & DATA

Open Discussion

LF AI & DATA

TAC Meeting Details

› To subscribe to the TAC Group Calendar, visit the wiki:
  https://wiki.lfaidata.foundation/x/cQB2 _____

› 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

□LF AI & DATA

Legal Notice

□LF AI & DATA

15DEC2022