THE LINUX FOUNDATION

WHITEPAPER

Open Source: The Missing Data and Management Layer

Executive Summary

For many years, open source software played the Rebel Alliance against the Empire of proprietary software. Today, open source is the dominant force in technology. Linux is the most popular operating system outside of personal computing. MySQL is the most popular database. The Internet is powered by open source software like Apache and NGINX. The most popular programming languages are all implemented in open source: Java, Go, Rust, C, and JavaScript. The communities that build open source ecosystems are living, breathing creatures, filled with energy and bound together by mutual interest in collaboration and sharing.

While the Force that is open source is obviously powerful — the growth and dominance of open source makes this self-evident — the open source revolution remains incomplete. While the ethos of open source is transparency and collaboration, the tool chain to automate, visualize, analyze and manage open source software production remains scattered, siloed, and of varying quality. Developers today are more and more focused on active community engagement to recruit others to help, mentor them as they contribute more, and enable them to be more active in the community.

> "It's an energy field created by all living things. It surrounds us and penetrates us; it binds the galaxy together."
>
> **— OBI WAN KENOBI, STAR WARS**

Organizations dependent on open source communities are looking to shift towards active community engagement themselves.

At an organization level, engineers, architects, documentation writers, Open Source Program Office professionals, lawyers, and more are trying to stay coordinated as an organization and ensure their engagement in the communities they participate in is productive and rewarding, for the employees and the organization. They face demands to better manage and foster participation, support their project leads and maintainers, and synchronize activities across their various teams involved. Many could benefit from a centralized source of information about their activities, tooling to simplify and streamline the multiple components of consuming and managing open source projects, and a solution to visualize and analyze the open source universe quickly and accurately on key parameters and indicators.

- For organizations that wish to better understand how to coordinate internal participation in open source and measure outcomes, gathering all the different pieces of data — code contributions, event talks and attendance, project governance and technical advisory committee roles — is painful and time consuming.
- For CTOs and engineering leads looking to build a cohesive open source strategy, obtaining objective information on the health, governance, and security of open source projects is two parts manual labor plus one-part tribal knowledge.
- For project maintainers, the legal and operational side of launching a project can be overwhelming and growing an open source project often leads to burnout.
- For individuals, keeping track of all their open source impacts requires manual aggregation of data and information from multiple sources on a frequent basis.

The Linux Foundation's LFX Platform is designed to address these issues and more. LFX probes metrics and workflow APIs from numerous developer and operations tools, and then frames them into context for open source projects with innovative modeling. LFX also adds intelligence to create outcome driven KPIs and provides a best practices driven, vendor agnostic control plane for operating and scaling open source projects.

As such, LFX functions as a single pane of glass for active community engagement and open source activity — a collaboration and tooling engine, and a platform upon which the already powerful open source movement can grow even more quickly and successfully.

Contents

The Universe of Open Source

A useful construct for thinking about open source is a mirror image of the physical universe. Projects are akin to planets, each with its own size, gravity, and trajectory. The gravity of planets acts on each other. The rise of Linux served to drive subsequent growth of MySQL, Apache, and the php programming language, to name one example. Clouds of projects tend to form galaxies of closely related entities that are often used together. Kubernetes is composed primarily of Golang but as a large planetary system, Kubernetes also supports other projects written in other languages. Programming languages are like the elements, the base level of composition of projects and galaxies. Languages also float on their own, like clouds of cosmic dust.

People and organizations contribute to projects — some closer in, some further out. The closer in, the more involved a person or organization is in the project. The most significant contributor(s) serves as the maintainer of a project, something that both individuals and organizations can do. Primary contributors and active ecosystem participants engage a few rings out. Active participants might be people developers who submit PRs or serve as project ambassadors, marketing people who work to promote the project via blogs and videos for their companies, senior technologists and engineers who serve on technical oversight committees, attorneys, financial experts and public relations professionals who advise a project, and more.

Participating further out are periodic contributors and one-way consumers of a project, who care about what's happening with the project but not enough to maintain an active presence or to attempt to influence and improve the project. It is important to note that at any contribution ring, an individual or organization can provide a critical input or capability. For example, while most bug reports are of lower severity, a developer in a distant ring who reports a Zero-Day bug delivers a mission critical benefit to a project and all the organizations that depend on that project. For this reason, it is important to maintain both a line of sight and open communications with all participants in projects, at every level.

Much of the data and information that makes up the open source universe is, not surprisingly, open to see.

Much of the data and information that makes up the open source universe is, not surprisingly, open to see. The major code repositories, GitHub and GitLab, both offer APIs that allow third-parties to track all activity on open projects. Activity in social media channels and public project chat channels are likewise relatively easy to capture. Blog posts, conference talks, and documentation are all public, and all can be captured. Some pieces of data should remain private, such as event participation, sponsorship activity, educational activity and project email. In cases where a project is hosted in a foundation — such as the Linux Foundation — there is an opportunity to aggregate the public and semi-private data into a privacy respecting, opt-in unified data layer. This data layer could serve as the foundation for creating an operational and analysis layer that would enable all

stakeholders to more easily interact with and manage open source projects. By aggregating the data and building a collaboration engine, all participants in the enterprise of open source will benefit from sharing knowledge and boosting network effects. Ultimately, such a data layer could dramatically accelerate the pace of innovation and adoption in open source, further accelerating the ascendancy of open technology development with all its benefits to society and humanity.

The Crushing Burden of Operating An Open Source Project

Developers looking to release and grow an open source project must do a lot more than ship code. Today open source project maintainers must staple together multiple tools to manage, observe, build, distribute and secure projects. Not surprisingly, key administrative, legal, security and operational tasks are neglected because maintainers focus on the code — the main reason they started the project in the first place. That said, neglecting these areas can harm the viability and adoption of projects and introduce unacceptable risk. Prior to release, maintainers need to perform the following necessary tasks if they wish their community to grow and succeed over the long haul.

Legal and Governance Requirements

Each open source project requires a basic legal foundation to support subsequent growth. Project founders must first choose a license type, a decision which has long-ranging consequences for the health, use and community growth of a project. Projects maintainers must also choose a legal entity type and set up governance of the entity. Project maintainers must also set up bylaws and rules that govern how a project runs; for example, establishing both a project board of directors as well as a project technical advisory committee.

Legal tasks include: researching and selecting a license; researching and putting in place bylaws and a governance structure; putting in place a Code of Conduct, securing brand marks and copyright for logos and name (to prevent freeriders and deceptive use of the mark). A project needs to set up security policies and security infrastructure to scan code, check for dangerous dependencies, and access controls to limit access to the main branch of code. Additionally, in order to redistribute contributions, it is necessary to ensure that the project has the necessary rights. A Contributor License Agreement is an agreement that grants the rights needed for the contribution to be redistributed as part of the project.

Marketing Requirements

Marketing is essential for the success of open source projects. Marketing activities are responsible for getting the word out and driving adoption beyond organic demand. Rarely do project maintainers have marketing experience. Some larger companies provide marketing support of open source projects but this requires creating additional tooling, as open source does not fit neatly into enterprise marketing paradigms.

Initial marketing tasks include: setting up a website and a blog; setting up social media accounts; setting up a chat capability (Slack or Discourse); creating a project launch including initial content, PR; securing event time and creating a tech talk about the project. Ideally, the project maintainers should also build out a strategy for growth and identify tools they will need to execute on their strategy. To drive adoption, maintainers will need to create good tutorials and write product documentation (which is arguably a technical task but increasingly viewed as baseline marketing). Lastly, projects need to maintain a CRM database of

everyone that interacts with the project in order to deliver the right communications to the right people and to power the email lists. An integrated view of a project participant, sponsor, or sustaining organization is important to ensure that the right people have access to the right documents and information.

Post-Launch It Just Gets Busier

After launch, ongoing tasks only magnify. Governance requires constant meetings with advisory committees and boards. Ongoing marketing requires: monitoring chat and social media; keeping fresh content on the website and in social channels; as well as writing blogs and documents; applying for speaking slots and answering calls for proposals; create project meetups and events; identify and nurture potential project ambassadors; and create a newsletter and email list and maintain that list.

On the technical infrastructure side, every project needs to: create and manage repositories (and who has access to the code), create a build and release stack to incorporate pull requests and new versions of the project code; write additional documentation to continue drive adoption; compile and write release notes; monitor and supply answers to technical questions surfaced in chat or social media; and create a security policy, email and response strategy. For security, the project maintainers need to pursue security audits and fix bugs that represent security risks in a timely fashion. For the project to be used in regulated settings, such as those requiring FIPS compliance, patches must be applied within a limited time frame that can enforce crisis mode and derail software shipping schedules.

Even for companies that seek to host open source projects and encourage teams to generate them, the scope of requirements forces serious consideration because of the resources required to check all the boxes.

This is only a partial list. Note that none of this relates to actually writing new code. For maintainers, the list is overwhelming. Even for companies that seek to host open source projects and encourage teams to generate them, the scope of requirements forces serious consideration because of the resources required to check all the boxes.

Risk to the Open Source Software Supply Chain

In the past three years, the interconnectedness of the open source universe has become also a critical source of security vulnerabilities. Organizations and governments building applications for software and hardware atop open source code are taking on all the vulnerabilities that might be present in the open source projects. Because open source is not only pervasive but also extremely varied, this increasingly complex and extended software supply chain has now become an attack surface of applications. Organizations that use open source, technology companies build open source projects, and maintainers of open source projects all are vulnerable to upstream and downstream dependencies in software supply chains.

For example, changes in ownership of a simple JavaScript library that is widely deployed in software compilers might allow malicious actors to insert

malware or secret-sniffing scripts surreptitiously, giving them access to privileged data or allowing them to remotely execute code. Another layer of dependency is the infrastructure of open source — the build servers and continuous integration pipelines used by many projects. Should any of these trusted systems or components be compromised — as we witnessed with the hack of SolarWinds' build server or the log4j "Log4Shell" vulnerability in 2021 — the entire supply chain is at risk. What makes all of this even more risky is now the cyberattackers are as likely to be state actors as organized crime. The level of sophistication of attackers is higher than ever and they are fluent in open source.

For organizations building with open source, gaining better insights into the state of their open source supply chain security is of paramount importance in maintaining a strong security posture. For project maintainers, building secure software and following secure software build and distribution practices is critical to the long-term success and adoption of their project. As decision-making authority in building applications has "shifted left" to smaller and smaller application teams and to individual developers, the need for clear determinations of whether open source code is secure has become obvious.

Should any of these trusted systems or components be compromised — as we witnessed with the hack of SolarWinds' build server or the log4j "Log4Shell" vulnerability in 2021 — the entire supply chain is at risk.

The Challenge to Enterprises of Managing Open Source Participation at Scale

The world's leading technology companies such as Apple, Facebook, Amazon and Google all have detailed open source strategies. They release open source projects regularly, sponsor open source conferences, and submit pull requests and bugs to open source maintainers. As we see in our work with the TODO Group, open source program offices (OSPOs) are rapidly proliferating into other industries beyond technology. Comcast, Bloomberg, BMW, Home Depot and dozens of other businesses and organizations that rely on open source software have OPSOs and open source strategies. Some of these organizations, like Bloomberg, have an explicit "open source first" policy. Some, like Comcast and Facebook, have built major pieces of open source software in house which were then released and became large, independent community-based projects. All of these organizations care deeply how they show up in the world of open source. In most of them, the OSPO reports directly to the CTO and open source has become a critical part of the business and technology strategy.

As open source has become more mission critical, so has the need to manage an enterprise's open source presence including code contributions, project participation in governance roles, sponsorships, and content

generation. Sophisticated organizations wish to automate manual processes around open source legal processes for their employees, such as Contributor License Agreements and approved projects for code contributions. In addition, the security of the open source supply chain and of all open source projects incorporated in company infrastructure and products becomes crucial in deciding which projects to adopt and support, and which projects to avoid. Related to security, enterprise requires detailed and accurate metrics on community health to drive better long-term decisions on which open source projects to consume.

The open source-centric enterprises and their leadership manually aggregate all the different sources of this information, or they spend internal resources to build tooling to automate the capture and analysis of open source metrics for their organization. Creating and maintaining this tooling is expensive, time-consuming, and challenging given the fast-changing nature of open source. For individuals working in these enterprises (or on their own), they lack a dashboard or single destination to see and manage their open source presence or share their open source accomplishments.

The open source-centric enterprises and their leadership manually aggregate all the different sources of this information, or they spend internal resources to build tooling to automate the capture and analysis of open source metrics for their organization.

For industry groups, open source foundations are becoming the vehicle for broad collaboration across multiple software projects that benefit all participants in that sector. Aggregating and analyzing code contributions, tracking employee skill development through completed training and certifications, monitoring event participation, and recording leadership roles in these industry-driven foundations becomes an important way to determine ROI and a mechanism for keeping track of many complex interactions.

LFX: Turning The Force of Open Source Into An Actionable, Extensible Data Layer

At the Linux Foundation, our mandate is to further the growth, development, and impact of open source around the world. To truly achieve it, we need to address the aforementioned challenges and unlock the true potential of open source. As the largest home of open source so\ftware projects in the world today and the largest producer of open source technology events, the Linux Foundation sits in a unique and privileged position to serve hundreds of projects and standards bodies, thousands of member organizations, dozens of industries and millions of individual members. Over the last five years we designed and tested systems and solutions that addressed the key challenges facing open source maintainers, technology leaders, individuals, and projects and industries, collectively.

Based on learnings from this process, we created LFX — a digital toolkit platform for growing, managing, consuming, securing, promoting and, most importantly, building open source technology, for any purpose and anyone.

The architecture and design of LFX mirror the strengths of open source software design. LFX is modular, extensible and API-driven. It is designed to be pluggable and easily integrate the data sources and tools that are already in use by organizations rather than force them to change their work processes.

> Based on learnings from this process, we created LFX — a digital toolkit platform for growing, managing, consuming, securing, promoting and, most importantly, building open source technology, for any purpose and anyone.

With LFX, organizations in the open source ecosystem and projects can simply integrate tool stacks they use and love for any relevant function including CI/CD, version control, code analysis and security. Out-of-the-box LFX includes dozens of integrations for relevant tools. LFX does not interfere with or manage the tools, rather it adds automation to provision and scale tools with an agnostic control plane for open source projects. More importantly, it gathers valuable metrics and workflows based on APIs to drive outcome drive actions for operating and growing an open source project.

By creating a holistic and configurable view of projects, organizations, foundations and more, LFX makes it much easier to understand what is happening in open source, from the most granular to the universal. With a few clicks, LFX users can go from high level dashboards and global visualizations engines down to single actions or activities. With easy to set up filters on hundreds of parameters, the LFX can telescope and microscope all the way from views and analysis of the total universe of open source down to health, membership, participation and development velocity of specific companies, projects or foundations. LFX can even provide more granular views down to individuals, code repositories and lines of code.

The LFX will always allow organizations to opt out of sharing any data that is not publicly accessible in open source forums (GitHub, conferences, public email listservs) on the platform; Giving everyone agency and control over their data is crucial to maintaining an ethos of privacy and trust.

The Linux Foundation is confident — given the history of the foundation as a non-profit trusted neutral third-party and facilitator of open source innovation — that the community will recognize the value in sharing data and information for the greater benefit of building a stronger body of knowledge and best practices for open source ecosystems.

LFX pulls in data from a variety of sources that help paint a more holistic picture of open source community engagement and activity, including:

- source control software (e.g., Git, GitHub or, GitLab)
- CI/CD platforms (e.g., Jenkins, CircleCI, Travis CI, and GitHub Actions)
- project management (e.g., Jira, GitHub Issues)
- registries (e.g., Docker Hub)
- documentation (e.g., Confluence Wiki)
- marketing automation (e.g., social media and blogging platforms)
- event management platforms (e.g., physical event attendance, speaking engagements, sponsorships, webinar attendance, webinar presentations)

LFX Supply Chain Management Tools

The LFX data engine then allows users to query and visualize this data with dashboards, charts, directories and other well-known data constructs. The LFX Platform uses APIs and custom connectors to unify a suite of useful tools and functionality, creating a single place where the key persona of open source can perform the bulk of their work. LFX also enables user-defined automated workflows and processes to eliminate the manual work common to managing open source.

LFX for Technology Leadership and OSPOs

For enterprises looking to more effectively manage and visualize their open source activities, the Organization Dashboard in LFX provides a variety of tools and data analysis capabilities to accommodate the needs of each persona in the open source ecosystem. The Organization Dashboard allows for quick queries, filters and centralized views to show which employees of an organization are working on which open source project and what their specific contributions have been.

Organizations can also understand which leadership groups in LF projects their employees play central roles. With easy-to-use and customized visualization engines, dashboards, query engines and maps, the Organization Dashboard provides a comprehensive view to determine ROI, participation, and where gaps remain in OSS strategy and execution.

The LFX platform also provides visibility to all Linux Foundation sponsor benefits for member

organizations — showing which benefits are being actively utilized and which are not (e.g., is the member company taking advantage of conference and training discounts, reading exclusive LF research, or participating in governing committees).

With its extensible and pluggable architecture LFX will support an organization's specific data sources allowing them to integrate their own tools, and seamlessly connect to their workflows and leverage the advanced data aggregation, visualization, and reporting capabilities of LFX.

More specifically, using the Organization Dashboard:

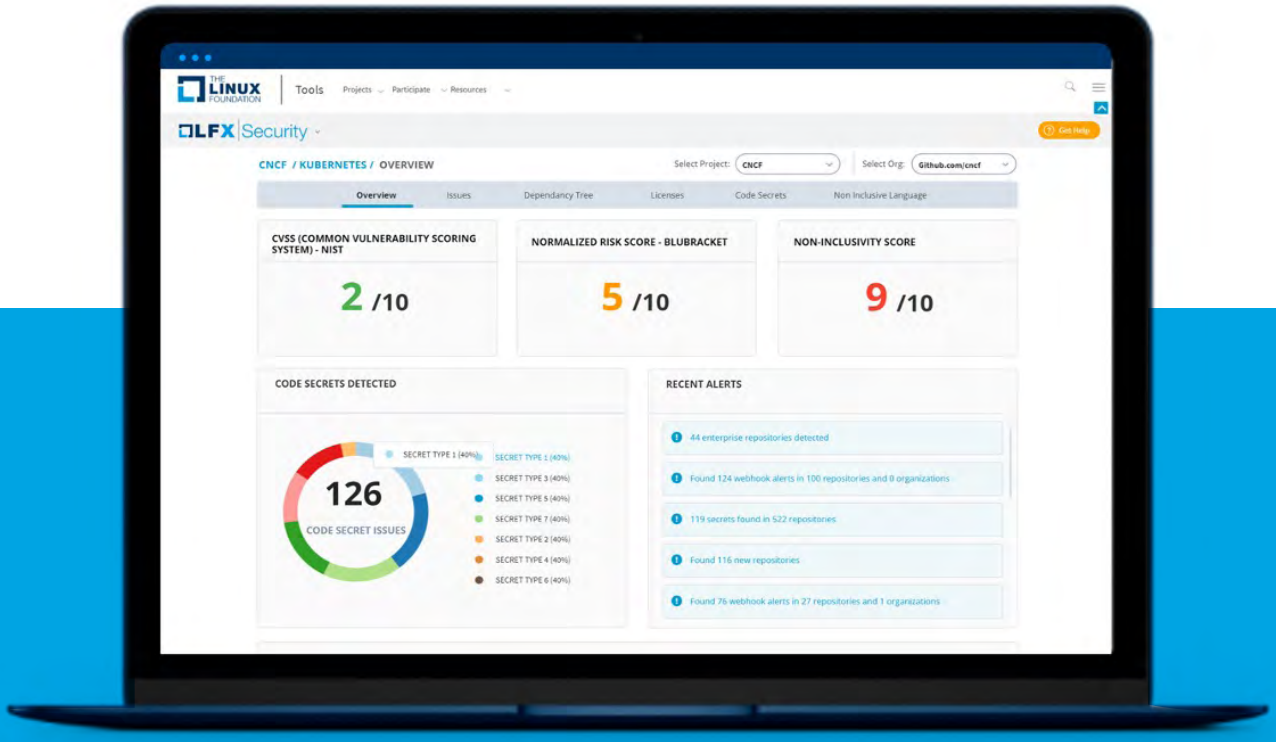
- **Employee Participation** — OSPO leadership can examine how active their organization's employees are in any project or foundation via the leaderboards for top projects their developers are contributing to as well as the top developer talent.
- **Code Contributions** — OSPO and tech leadership can compare how many commits, pull requests, and lines of code their organization contributes to different projects and foundations and compare contributions from other organizations.
- **Event Participation** — Organizations have easy access to data about which Linux Foundation events their employees attend and what their participation levels were. OPSO leadership can determine the interest level of employees in specific events and identify which employees speaking at the event drew the largest audiences. The LFX events data also allows for aggregated views of sponsorship spend, gender of attendees, number of talks given, and booth visitors.
- **Compliance** — For open source projects they contribute to, companies can safeguard against compliance risks by signing corporate level license agreements that cover their developers' compliance requirements and provide legal coverage for thousands of employees in a few clicks.
- **Content** — Content teams can report on what content has been created by whom from blogs to talks to webinars to email newsletter mentions.
- **Communications** — The in-line CRM, allows for OSPO and technology leadership to manage with bulk actions all their employees and filter and communicate with them based on community governance groups like Technical Committees, Boards, Marketing roles, or other specific attributes.
- **Analysis** — Tools for analyzing project health and good governance can inform decisions on where to invest contribution resources or sponsorship dollars, or, more importantly, what projects pass muster for inclusion in an organization's technology infrastructure roadmap.
- **Visualizations** — Time-series visualizations and granular activity breakdowns of each project or foundation illuminate velocity, rate of growth, and longer term trajectory of a project or foundation.
- **Security Ratings** — Developed from the Security tool, clear security ratings for each project make it easy to understand and compare security levels per project based on code coverage and analysis, patching behavior and project criticality. LFX packages best-of-breed open source security tooling with industry standards for Software Bill of Materials to allow project maintainers to provide more proactive and comprehensive security coverage and issue management.
- **Talent Management** — Talent management modules that allow OSPOs to track certifications, mentorships, and other educational achievements as well as organizational progress towards educational goals.
- **Diversity and Inclusion** — LFX breaks down diversity and inclusion measurements for projects, foundations and organizations that allow leadership to track progress or identify gaps in D&I efforts to assign for enhanced remediation or resources.

LFX for Open Source for Maintainers and Project Support Teams

The LFX platform takes all the best practices for managing open source projects collected from decades of experience at the Linux Foundation and puts them at the fingertips of open source project maintainers and project participants. A project can use the LFX platform to manage membership, automate operational tasks, streamline finance and legal processes, and stand up a technology infrastructure, all within a single platform. With the platform, project maintainers and their fellow contributors can:

- Organize governance resources including legal entities and official committees, project documentation, export controls, and trademarks.
- Manage a project's financial operations including membership tiers, pricing, billing, and accounting.
- Set up code and version release pipelines with a selection of CI/CD tools. This pipeline-in-a-box is entirely inline and hosted by LFX to facilitate commits, code merges, builds, security checks and code analysis, and, finally packaging and source code distribution.
- Centralize management of all project IT services including source control, domains, mailing lists, membership lists and levels, cloud instances, and collaboration tool.
- Deploy role-based access to all your tools including source control, financials, marketing automation, legal and event organizations solutions. Set granular permissions per project for administrators, project managers, and community stakeholders.

A project can use the LFX platform to manage membership, automate operational tasks, streamline finance and legal processes, and stand up a technology infrastructure, all within a single platform.



LFX for Open Source Security

For enterprises, organizations, project maintainers, and all consumers of open source software, delivering rock-solid security is essential. The LFX platform builds on the work of the Core Infrastructure Initiative and the Open Source Security Foundation and best practices garnered from years of work on open source security by the Linux Foundation to create a comprehensive suite of capabilities and tools to help projects improve their security. LFX integrates all the existing security tooling. Capabilities include:

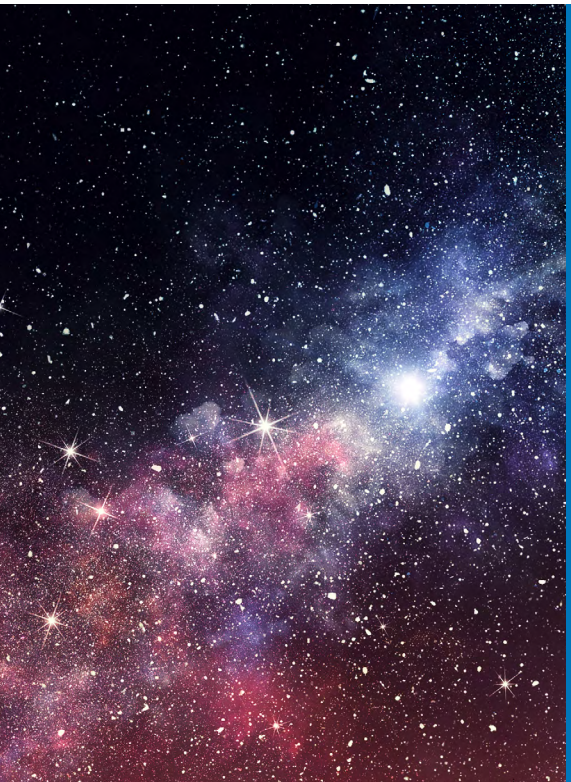
- Dependency risk analysis by open source software dependency management platform Snyk
- Security test coverage with software composition analysis (SCA)
- Patch and vulnerability management visualizations and tracking
- Secret leak and non-inclusive language detection by BluBracket
- Secure version control and code pipeline management
- Best practices guides for setting up security policies and response processes
- CVE response ratings, CII badge status and other certification completion notifications

Conclusion: Magnifying the Force of Open Source With Better Data, Better Tools

The LFX platform will subsequently expand into other offerings including mentorship listings and management (think of it as Google Summer of Code in a box), integration of additional mailing list tools and other communications or marketing tools, integration of additional security testing and code analysis capabilities, and explanation of project coverage beyond the Linux Foundation to other open source projects published in major code repositories.

The primary obstacles facing the open source movement are surmountable with smart data and integration, intelligent packaging of critical functionalities, and accessible user experiences to enable anyone — from engineers to non-technical marketing and events personnel — to improve open source project management, participation and outcomes. Just as the Force in Star Wars is the animating power behind progress and light, open source is paving a way to a non-zero-sum future built for the benefit of all. Mapping and channeling the power of open source will be radically simplified when all the pieces needed to manage, visualize, nurture and grow open source communities can be found in a single platform.

By combining all the different streams of open source work into a useful platform, LFX will improve communication and collaboration, simplify management, surface the best projects and project leaders, and provide insightful guidance based on real data captured at scale, across the widest variety of projects ever collected into a single source of information. Rather than go to a Jedi master to learn about the force, any CTO, developer, marketer, project volunteer, user or student seeking an internship can turn to LFX to get whatever they need. With this free flow of information will come previously unattainable information economies of scale that will accelerate open source innovation for the benefit of all.



THE LINUX FOUNDATION

The Linux Foundation promotes, protects and standardizes Linux by providing unified resources and services needed for open source to successfully compete with closed platforms.

To learn more about The Linux Foundation or our other initiatives please visit us at www.linuxfoundation.org.