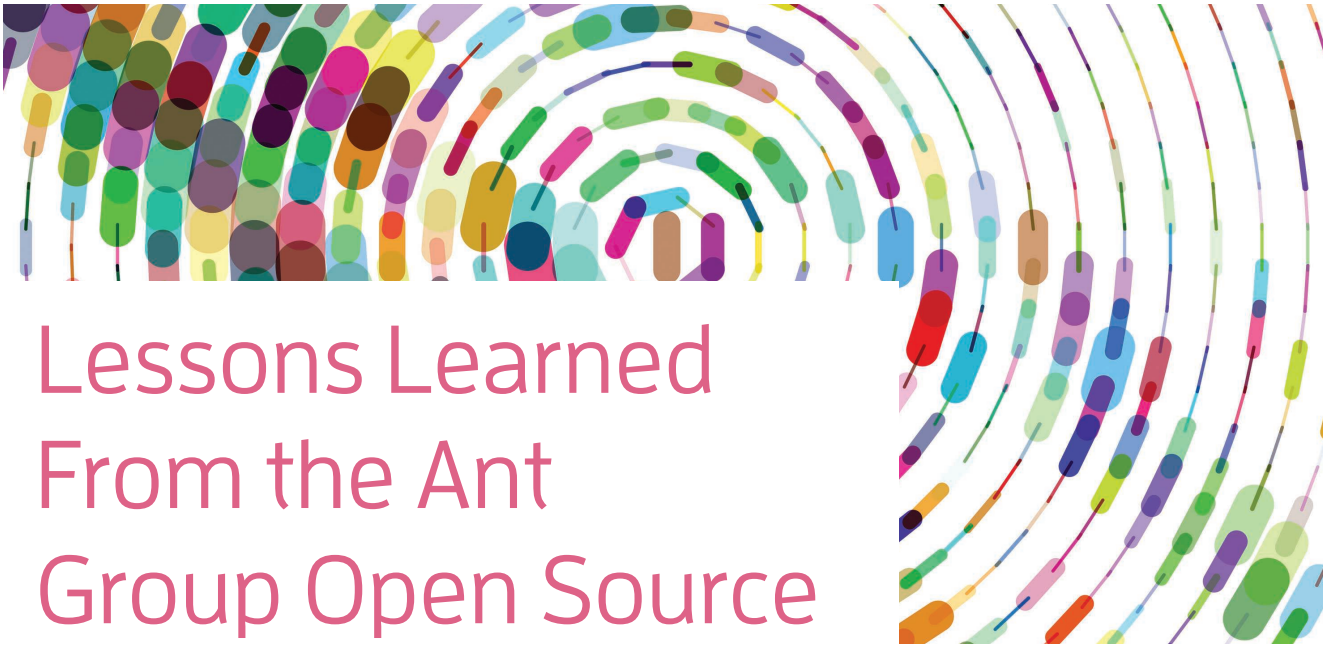

Lessons Learned From the Ant Group Open Source Program Office

Xiaoya Xia and **Wei Wang,** East China Normal University
Shengyu Zhao, Tongji University
Sikang Bian, Ant Group
Rong Wang, Alibaba Group

This article introduces the open source program office best practices of a Chinese tech giant, Ant Group. We summarize roles and challenges and discuss the tool sets adapted by Ant Group.

Open source software (OSS) has created some of our most widely used technologies, including operating systems, web browsers, and databases. OSS is becoming the de facto infrastructure of the digital world and is widely used today by companies in a broad range of industries. Embracing OSS is not merely an option for companies to consider but a critical path to achieving technical innovation.

By observing the GitHub event log, we found that the overall activity of global open source has been rising significantly over the last five years with an accelerating trend, especially postpandemic. The number of active repositories is maintaining a growth rate of about 24%.[1] In such a context, China is also experiencing a rapid and even greater OSS growth. The establishment of the OpenAtom Foundation has spurred the introduction of various national policies to encourage OSS development. China-based tech giants, such as Alibaba, Ant Group, Baidu, Tencent, Huawei, etc. account for the preponderance of significant open source contributions. Figure 1 illustrates the overall activity rank[2] of Chinese companies during 2017–2021, calculated by OpenDigger[3] using GitHub event logs. As companies are becoming important players, teams often referred to as open source program offices (OSPOs) are set up to deal with OSS-related affairs.

THE INTERNAL AND EXTERNAL DIPLOMAT

The OSPO is not only a governance entity to ensure the compliance and safety of OSS but also a strategic initiative

Digital Object Identifier 10.1109/MC.2022.3219638
Date of current version: 5 April 2023

0018-9162/23©2023IEEE

EDITOR **DIRK RIEHLE**
Friedrich Alexander-University of Erlangen Nürnberg;
dirk.riehle@fau.de

that could lead to competitive advantages. Companies getting involved in open source can be simplified into three stages: using, participating, and creating. During each phase, issues may emerge that involve different departments:

- *R&D Department*: select open source components and develop OSS
- *Marketing Department*: extend open source's influence and brand building
- *Legal and Security Department*: perform OSS compliance and risk management.

The OSPO's main focus is to coordinate among different departments internally and respond to other open source stakeholders externally. Although the TODO Group systematically illustrates OSPO roles, behaviors, size, and responsibilities using a mind map,[6] it has also pointed out that the functionality, team structure, and focus can vary based on different situations and business goals.

Currently, the most common form of OSPO in China is to serve as virtual departments, as they are in a trial stage. For example, Alibaba has set up the Alibaba Open Source Technology Oversight Committee, which is a virtual team that has made open source-related decisions for years. It was not until 2019 that it established an OSPO and onboarded full-time-equivalent employees. Now the compact, hybrid team is working closely with the Alibaba Open Source Technology Oversight Committee on open source strategy, tool automation, project facilitation, developer and academia outreach, and so on.

Similarly, the Ant Group OSPO was formed at the beginning of 2021. Ant Group was formerly known as *Ant Financial*, an affiliate company of Alibaba Group, and owns the world's largest mobile payment platform, Alipay. Ant Group had open source projects long before it had an OSPO. The projects span multiple technology stacks, including the front end and back end, such as Ant Design, EggJS, OceanBase, SOFAStack, etc. As a virtual department, the Ant Group OSPO team is lean and agile, comprising core employees who are formerly from different business departments and mainly led by the technology department.

The Technology Oversight Committee usually focuses solely on the technology domain, while the OSPO involves legal, security, software development, marketing, and management. An apt metaphor exists, describing open source program managers as "diplomats," both external and internal. Their most prominent function is to build communication bridges and collaborate across multiple teams, thus achieving horizontal empowerment.

From a pragmatic perspective, the OSPO needs to solve the problems that inevitably occur while using, participating in, and creating open source projects. From another point of view, the OSPO can create and sustain a strong company culture that promotes values, empowers the open source ecosystem, and attracts talent at the same time.

FROM THE EDITOR

Open source program offices (OSPOs) are virtual or dedicated organizational units inside companies with the mandate to set open source policies, ensure license compliance, manage project engagement, and do various other open source tasks. In previous articles, we reviewed OSPOs in general. In this article, Xia et al. take us into the domain of Chinese OSPOs and how they both are similar to and differ from other OSPOs. Look forward to learning about the principles and practices of OSPOs and, in particular, Chinese OSPOs. As always, be happy, be healthy, and keep on hacking! —*Dirk Riehle*

ANT GROUP'S OSPO BEST PRACTICES

Challenges faced

When Ant Group cold-started the OSPO, it realized that the need for an OSPO comes from the fact that promoting

(index)	rank	company	active repos	activity
0	1	'Alibaba'	885	78336.02
1	2	'Baidu'	244	60428.08
2	3	'Ant group'	348	53370.07
3	4	'PingCAP'	121	46647.69
4	5	'Tencent'	294	25703.63
5	6	'Huawei'	118	10095.76
6	7	'QingCloud'	48	9396.54
7	8	'Fit2Cloud'	63	9002.51
8	9	'Vesoft'	43	7988.11
9	10	'JD'	51	6488.17

FIGURE 1. The activity rank of Chinese companies in the past five years.

open source as a corporation is intrinsically harder because there are higher overall costs and risk exposure, and it takes longer to build momentum.

Participating in open source is much easier as an individual: you just do it. There are obvious costs, such as time and energy, and relatively clear goals, such as social recognition. It is easy to evaluate. On the corporate side, things are different.

Reason 1: Overall costs and risk exposure are higher. A company would always take care of compliance risks, legal risks, and security risks. Mistakes in this regard can create chain consequences; for instance, you never want to provide products with security vulnerabilities or license issues. The cost is not only in maintenance and repair but also in the damage to the company's reputation. Although proper use of licenses provides certain levels of protection, nontrivial compliance or business risks lurk in many places.

Similarly, the overall costs of operating companies' own open source projects are also considerably higher. A company needs to thoroughly think through a project's governance strategy depending on the nature of the project and the vision of the company.

There is also the cold-start problem. It is not easy to get someone who has never been exposed to open source to fully embrace the OSS and open source community best practices. For example, engineers are less motivated to write documentation, and it also takes a significant amount of effort to build up habits for asynchronous (async) communications, which brings up reason 2.

Reason 2: It takes time to build momentum. Another challenge is that it takes time to gather momentum to identify and implement the best open source practices. Investing in open source is typically a long-term plan that requires dedication. Why? Some immediate problems are relatively quickly addressed, like providing open source license consulting for teams. However, many other things, including but not limited to "the culture of being open," async communication practices, etc. require a much longer time until their potential benefits are fully realized.

Corporate investment in open source is similar to investing in research institutions. We need to set the vision on long-term benefits instead of evaluating short-term gains. Committing to open source ways requires top-down dedication and bottom-up understanding.

With these challenges, why would a corporation still want to do it? This is because, compared to individuals who participate in open source, the potential "reward" for being open is pretty high. For instance, a To Business tooling software as a service, if done right, can directly benefit from being open source.

Still, it requires more than one project or a team's effort to make a company's open source ventures successful. That is why we need to measure risks, costs, and gains.

Tool sets for metrics, tracking, and monitoring

As for-profit organizations, companies will set certain business goals; however, it may take a longer payback period to assess the return on investment, with some of the efforts being hard to quantify. Before collecting information and tracking the metrics, it is important to clarify the goal in the first place. That said, there are some common aspects that open source program managers would always want to know: the general health of the project they open sourced; how influential it is, which adds up to the corporate reputation; the ability to attract talent; and how to identify devoted external contributors.

An earlier column introduced the Community Health Analytics Open Source Software (CHAOSS) project,[4] which defines and tracks key metrics of OSS communities. Here, we illustrate another tool set adapted by Alibaba and Ant Group to track key contributors, projects, and recognize success.

OpenDigger. OpenDigger[7] conducts systematic studies using data science methods, where it measures, models, and analyzes projects, developers, and processes. One of the valuable data assets of OpenDigger is the continuously collected global GitHub behavior logs and the software dependencies that currently support three programming language ecosystems: JavaScript, PHP, and Python. With the global data, measurement aspects, computational models, and labels, the data service project can quickly generate analysis reports for any projects, user groups, and institutions that are active on GitHub as well as perform horizontal comparisons within and across organizations.

OpenDigger implements quantitative CHAOSS metrics, integrating them in the Jupyter notebook environment, where community managers can query the result interactively. Figure 2 shows the number of bus factor contributors and their trends of different GitHub organizations in 2015.

Companies should be wary of falling into the trap of chasing numbers when quantifying effort. That is why OpenDigger is also committed to performing measurements from the graph's perspective. The statistical method can be easily cheated by spamming the events if a project wants to get a higher rank. While in the graph, the node value will not be influenced too much if an account opens a lot of pull requests but never collaborates with others.

The real world of software development is a social coding scenario where developers and projects are interconnected by activities and relationships. The OSPO cares about the ecosystem-level data in which connections and relations truly matter.

Figure 3 describes the primary idea of constructing the graph. There are three developers (blue) and three projects (pink). Developers are

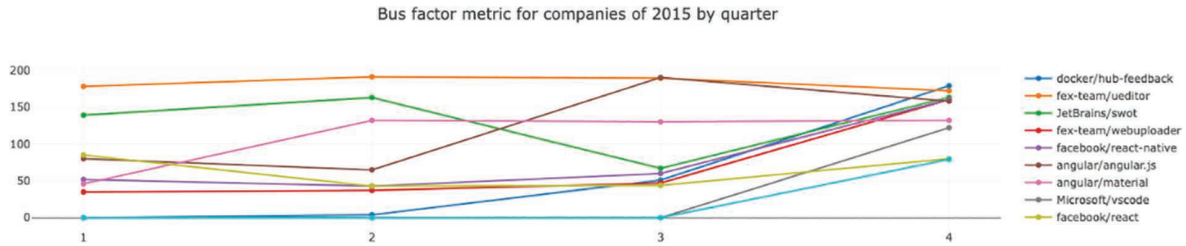

FIGURE 2. The bus factor metric for companies by quarter in 2015.

"active" in projects, projects are dependent on other projects, and developers are following other developers. The "active" relationship contains any actions that can be performed by a developer, like activities in issues and pull requests, making a release, attending a meeting, and so on. The node value is calculated by a PageRank-like centrality algorithm and is defined as the influence of a developer or a project.

OpenDigger empowers upper-layer productization tools, reports, and dashboards as a project that enables data and measurement.

Hypercrx. Hypercrx[8] is a chrome extension that further expands the playability of the graph visualization. Once installed, it renders the interactive networks on any GitHub projects' homesites and GitHub users' profile pages. As can be seen in Figure 4, the networks demonstrate project connections (left) and developer connections (right). Projects are connected by developers, and developers are connected by collaborations on issues and pull requests. Clicking the node allows redirection to the corresponding project or user site. The play tool provides an exploratory open source observation setting of the ecological relationship, serving as a convenient portal for community managers to track contributor inflows and outflows.

The dashboard. With the organization-level data collected, the dashboard presented in Figure 5 can effectively provide a bird's-eye view of the open source projects maintained by corporations like Alibaba and Ant Group. The most common symptoms to keep an eye out for include (but are not limited to) active projects, community growth, license distribution, and project decay (projects with no commit in the past, say, six months).

The dashboard helps with decision making from a strategy perspective. On the other hand, it is also an incentive. As mentioned earlier, using a computational model (such as a graph) might be better than focusing on a single quantitative dimension. The metrics we observe really depend on how we want to assess the work but will also impact the value orientation of internal work at the same time.

The principle and practices

Ant Group summarizes the overall OSPO role as one principle: the OSPO should be the go-to entity for both internal and external open source-related matters.

In engineering terms, we identify the OSPO as a "façade" or a "public application programming interface (API)" that abstracts internal capabilities and interacts with the external world. Why use this public API metaphor? Because a public API is designed to last. It represents a contract with the external communities and entities.

TODO Group helped normalize and standardize the OSPO. It became

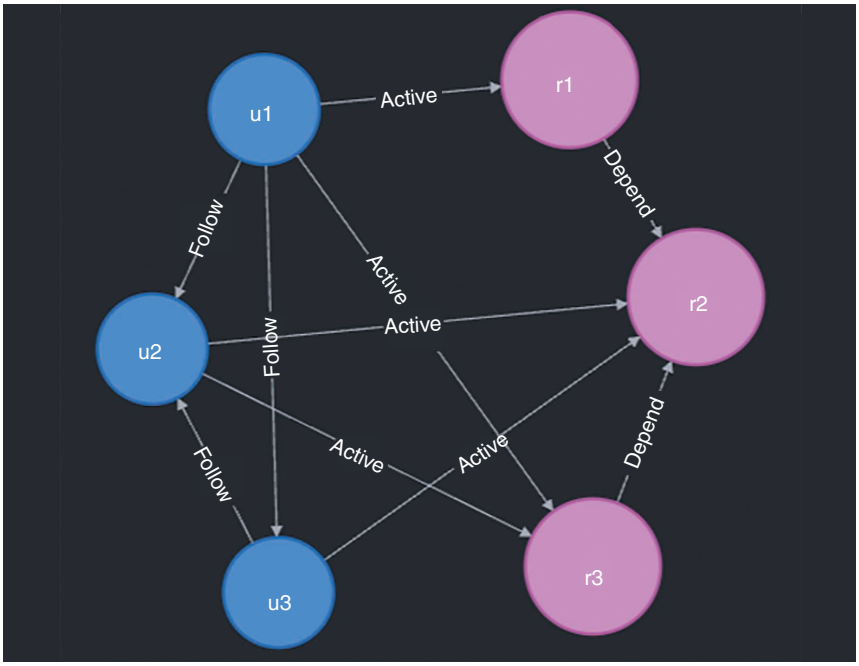

FIGURE 3. The developer and project graph demonstrates a collaborative interconnectedness. u: user; r: repo.

part of common terminology and the potential common belief that we can all fall back on. We don't need to start from scratch to explain that "we need a governing body that can help the company manage open source compliance matters as well as blah blah blah"; instead, we can say, "Let's build an OSPO."

Practice 1: The OSPO is a necessary infrastructure to enable developers and their teams to open source compliantly and confidently. This practice aims to help internal engineering teams. The primary focus of this work is spearheading processes and best practices for projects and community members to learn from and follow. If we drill down one level and can be more specific, the scope of work would cover "using open source internally"; "contributing and setting up new open source projects"; and all of the software composition analysis, static application security testing tools, communication tools, and utility tools that would help to achieve that. Metrics are also key aspects, as observability is typically a critical factor in infrastructure.

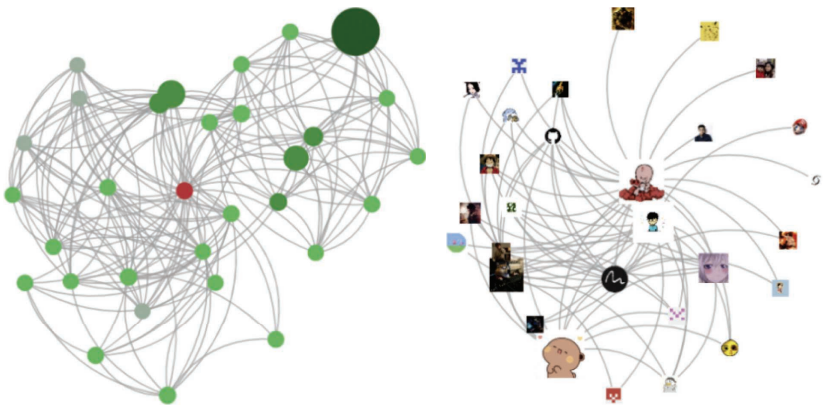

FIGURE 4. (a) Hypercrx project networks and (b) developer networks.

Practice 2: For OSS project teams, the OSPO is a technology enabler. It is a catalyst for the reaction. A catalyst itself does not achieve much, but it either makes an impossible reaction happen or enhances the rate or results of the reaction.

The Ant OSPO helps projects as they grow by providing strategic planning for goal setting, governance consultation, and operation plan design as well as potential go-to-market strategies to empower projects to be professional and successful. The metrics provide North Star guidance, so the team will have some impartial and professional references to which they can directly relate.

Practice 3: On the external side, OSPO could be the "deal maker."

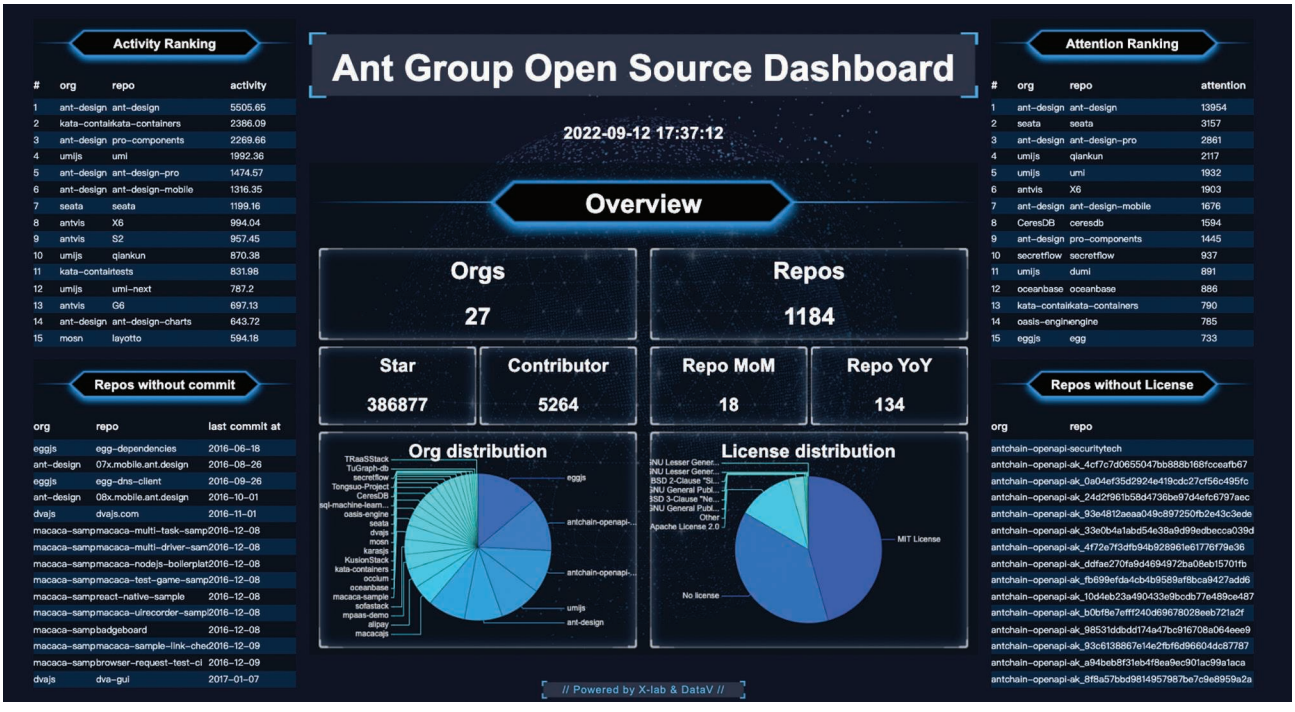

FIGURE 5. A bird's-eye view of the Ant Group dashboard.

The book *The Rainforest: The Secrets to Building Next Silicon Valley*[4] states that even though there are many other places with a similar concentration of talent and investors, what made Silicon Valley so successful were the network hubs. They are well-connected individuals with a "giver" style. They connect start-ups with investors as well as high-potential individuals with mentors and opportunities. The book refers to the existence of "deal makers" with a positive connotation. As the "façade" that connects with the external world and interfaces with communities, foundations, and other collaborators, the OSPO has the potential to be this "deal maker." It connects demand with potential solutions, which many open source communities and entities can directly benefit from.

Practice 4: Find all of the alliances to build open source methodologies, tools, and so much more. Security and compliance teams are typically the first step in building alliances. Furthermore, the technology branding team would also need to be willing to cooperate. After finding alliance members, the next step would be to refine and standardize the methodology. Ant Group formed a working group with bimonthly open source application reviews. Once we have methodologies, we will have tooling needs. Some of the tools are self-developed to meet customized needs, while others can seek external partners. For example, the tool sets illustrated in this article are based on academic outreach and are well adapted.

Practice 5: Explore innovative paths based on predecessors. The TODO Group offered standardized approaches and resources. However, not all OSPOs are created equal. It is important to identify your own style by focusing on solving the problems at hand.

If we return to the analogy of a catalyst, the OSPO catalyzes the incubation of to-be-open-sourced candidate projects and empowers them for success.

First, Ant Group adopts the lifecycles Cloud Native Computing Foundation used internally and provides end-to-end support for its OSS projects, depending on the stage of development. A project starts from the sandbox stage, in which the OSPO helps the team to refine and focus on its primary goals. Once the project passes joint review, it enters the incubation stage. During this phase, the OSPO team focuses on designing the operation strategies as well as potential commercialization strategies. It is a "rinse and repeat" process that the OSPO works closely with the project teams on.

Second, Ant Group has InnerSource as a staging environment in which people can put their projects to learn and grow. Offering engineers a playground internally significantly reduces barriers to contributing and sharing. Engineers who are reticent to contribute to open source directly felt more comfortable in an inner source environment. InnerSource is also the perfect playground for setting up good practices like documentation rules, async communications rules, and so on.

REFERENCES

1. "GitHub 2020 digital insight report," X-lab, Shanghai, China, 2021. [Online]. Available: http://oss.x-lab.info/github-insight-report-2020-en.pdf
2. X. Xia, Z. Weng, S. Zhao, and W. Wang, "Exploring activity and contributors on GitHub: Who, what, when, and where," in *Proc. 29th Asia-Pacific Softw. Eng. Conf. (APSEC)*, 2022.
3. "OpenDigger." GitHub. Accessed: Aug. 10, 2022. [Online]. Available: https://github.com/X-lab2017/open-digger
4. S. Goggins, M. Germonprez, and K. Lumbard, "Making open source project health transparent," *Computer*, vol. 54, no. 8, pp. 104–111, 2021, doi: 10.1109/MC.2021.3084015.
5. V. Hwang and G. Horowitt, *The Rainforest: The Secret to Building the Next Silicon Valley*. Scotts Valley, CA, USA: Createspace Independent Publishing Platform, 2012.
6. "OSPO Mindmap." Accessed: Aug. 10, 2022. [Online]. Available: https://ospomindmap.todogroup.org/
7. "OpenDigger." GitHub. Accessed: Aug. 10, 2022. [Online]. Available: https://github.com/X-lab2017/open-digger
8. "Hypercrx." GitHub. Accessed: Aug. 10, 2022. [Online]. Available: https://github.com/hypertrons/hypertrons-crx

XIAOYA XIA is a Ph.D. candidate at the School of Data Science and Engineering, East China Normal University, Shanghai 200062, China. Contact her at xiaoya@stu.ecnu.edu.cn.

WEI WANG is a professor at the School of Data Science and Engineering, East China Normal University, Shanghai 200062, China. Contact him at wwang@dase.ecnu.edu.cn.

SHENGYU ZHAO is a Ph.D. candidate at the School of Electronic and Information Engineering, Tongji University, Shanghai 201804, China. Contact him at frank_zsy@tongji.edu.cn.

SIKANG BIAN is a technical strategist at Ant Group Open Source Program Office, Hangzhou 310063, China. Contact him at biansikang.bsk@antgroup.com.

RONG WANG is an operations manager at Alibaba Group Open Source Program Office, Hangzhou 311121, China. Contact her at tunan.wr@alibaba-inc.com.