



Article

Evolution of blockchain technology

Insights from the GitHub platform

[Jesus Leal Trujillo](https://dupress.deloitte.com/dup-us-en/authors/t/jesus-leal-trujillo.html) <<https://dupress.deloitte.com/dup-us-en/authors/t/jesus-leal-trujillo.html>> , [Steve Fromhart](https://dupress.deloitte.com/dup-us-en/authors/f/steve-fromhart.html) <<https://dupress.deloitte.com/dup-us-en/authors/f/steve-fromhart.html>> , [Val Srinivas](https://dupress.deloitte.com/dup-us-en/authors/s/val-srinivas.html) <<https://dupress.deloitte.com/dup-us-en/authors/s/val-srinivas.html>>

November 06, 2017

A new study of the global open-source platform, GitHub, offers key lessons on blockchain development—how projects have grown, what's likely to come next, and the implications for financial services firms.

Making sense of the noise

WE cannot predict the exact trajectory and impact of blockchain technology. But we also should not ignore its early stages of development and successes along with failures. Tracking this young technology's development could potentially maximize its potential to best serve us.

Figuring out how foundational technologies, such as the Internet or mobile, morph and grow is not easy. New technologies often attract a wide variety of developers, including many freelancers from around the world. The sheer number of developers, the types of problems they are trying to solve, and the geographic spread can make it difficult to anticipate where any new technology is headed.

But perhaps the fundamental difference with blockchain development is that it has largely been orchestrated in the open-source environment. Bitcoin, the original blockchain system, was birthed in open source.

Accordingly, in an effort to better understand the development of blockchain and its ecosystem, we have conducted an extensive data analysis of blockchain projects in an open-source environment. Our study appears to be the first empirical attempt to understand the evolution of blockchain using metadata available on GitHub, a global software collaboration platform.

We chose GitHub because it is the largest known software collaboration platform in the world, with more than 68 million projects and 24 million participants (figure 1). ¹ GitHub also

appears to host the most important projects for the blockchain community. ² The activity on GitHub provides a unique opportunity to identify who is behind blockchain's development, what type of programming is powering it, where the talent resides, how networks and communities of projects and developers are organized, and what risk factors exist for investing resources into repositories.

Figure 1. GitHub in numbers

- 24 million GitHub users
 - 68+ million repositories
 - 337 different languages
-
- **"Repositories"** are software projects that host code
 - **Watchers vs. committers:** A *watcher* follows the development of a project and a *committer* contributes to a project with code
 - **"Commits"** are contributions to code
 - **"Forking"** is copying a project into the work environment



Source: Deloitte analysis of GH Torrent data and GitHub API data, as of October 12, 2017.

Deloitte Insights | deloitte.com/insights

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Fig1.png>

Financial services firms seem to be leading the way in blockchain applicability; they currently have the most commercial use cases of blockchain in the marketplace. Our findings could help firms improve their ability to identify successful projects and opportunities based on how the blockchain ecosystem is evolving.

Unless otherwise cited, all data and statistics that we report on blockchain activity on GitHub in this paper is a result of our analysis

of the GH Torrent project and the GitHub API (see the sidebar, “Study methodology”).

STUDY METHODOLOGY

To conduct our study on GitHub, we utilized data collected by the GH Torrent project, a research initiative led by Georgios Gousios of Delft University of Technology, which monitors the GitHub public event timeline where all of the projects’ activities and modifications are recorded. ³ After this initial process was completed, the information was stored in a relational database. The database compiled by GH Torrent comprises more than 4.7 billion rows of information. To identify

relevant projects, we queried the GitHub API about keywords associated

Show more

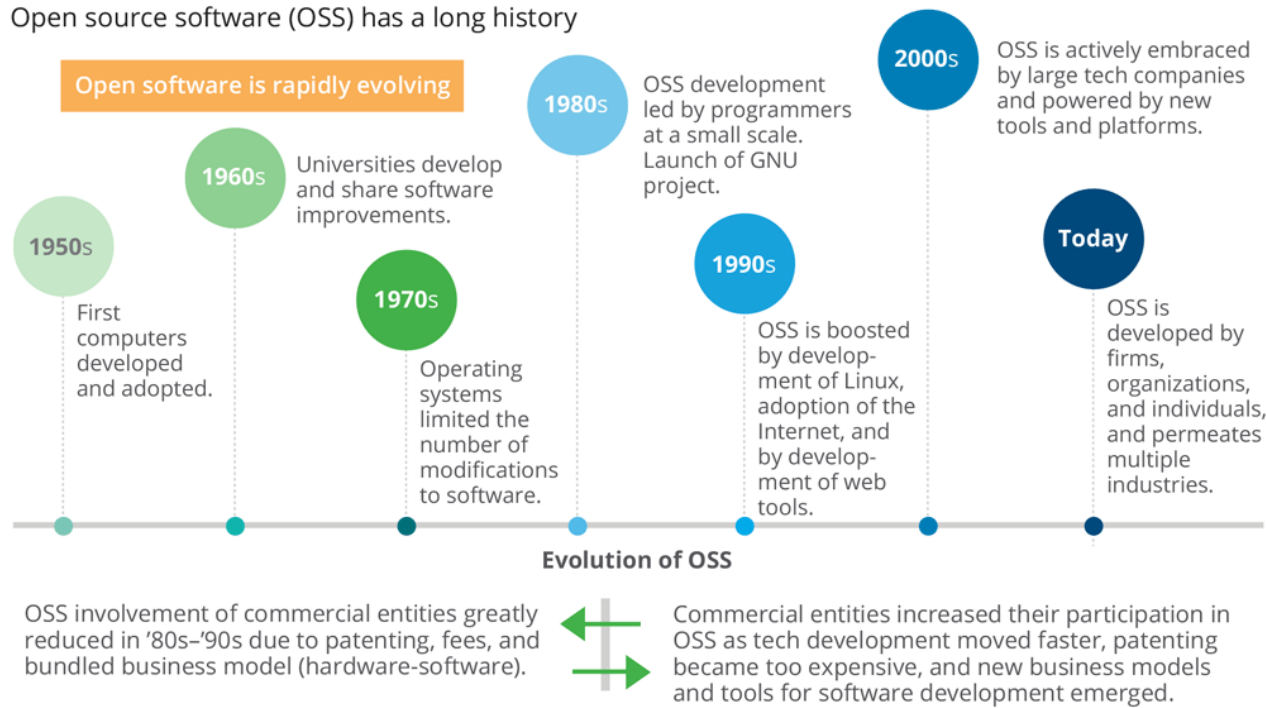


Blockchain thrives in an open world

While sharing software code in a public forum can be traced to the 1950s, open-source platforms have only become hubs for software development within the last 30 years (figure 2). ⁴ The Internet was a great enabler for scaling: Earlier, open-source activity had been mainly the realm of academia, but the Internet made it accessible to aficionados and experts of all stripes, amateur and professional, individual and commercial. ⁵ That said, there was a dip in relevance of software development on open source for a period when commercial entities that secured licenses and patents placed high fences around software code. ⁶ However, disruptive innovation has fostered an ever-increasing sharing economy, which has shifted important software development back to open-source platforms. ⁷

Figure 2. History of open source

Open source software (OSS) has a long history



Source: Longsight, available at <https://www.longsight.com/learning-center/history-open-source>, accessed September 12, 2017.

Deloitte Insights | deloitte.com/insights

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Fig2.png>

Open source could be the ideal petri dish for attracting a critical mass of blockchain coding efforts, talent, and overlapping objectives that accelerate an ecosystem with common standards. ⁸ It may also mitigate the cost that firms would pay to dedicate resources to a still largely experimental technology. Developing proofs of concept in an “intranet” blockchain learning platform does not seem as efficient as learning how to develop business solutions on an “Internet” blockchain. ⁹ At the current evolutionary stage of blockchain technology, it is likely to be in a developer’s best interest to develop, or watch the development of, blockchain solutions on open source. Blockchain appears to have a better chance to more quickly achieve rigorous protocols and standardization through open-source collaboration, which could make developing permissioned

blockchains easier and better.

Our primary unit of analysis on GitHub is the repository. A repository contains the relevant code and files behind projects, where the actual protocol and implementation of programs reside. Throughout this report we use the term “repository” and “project” interchangeably. We will also be looking at the two types of project authors: *users* — individuals with no known affiliation to an institution (*users*) ; and *organizations*— accounts associated with financial services firms, start-ups, research centers, or software foundations. ¹⁰

In the next three sections, we look at repositories—their authors, their chances of survival, and how they fit into communities and networks of communities; which programming languages are prevalent and why; and where talent resides. (See our interactive dashboard, where you can explore the GitHub ecosystem’s repositories, coding, and geographies in detail.)

GitHub in blockchain

Use the interactive visualizations below to explore the GitHub blockchain ecosystem’s repositories, coding, and geographies in greater detail.

Repositories by year

Most popular languages

Repositories by

Repositories by geography

Network visualization

Communities of

repositories



































































































































































































































































































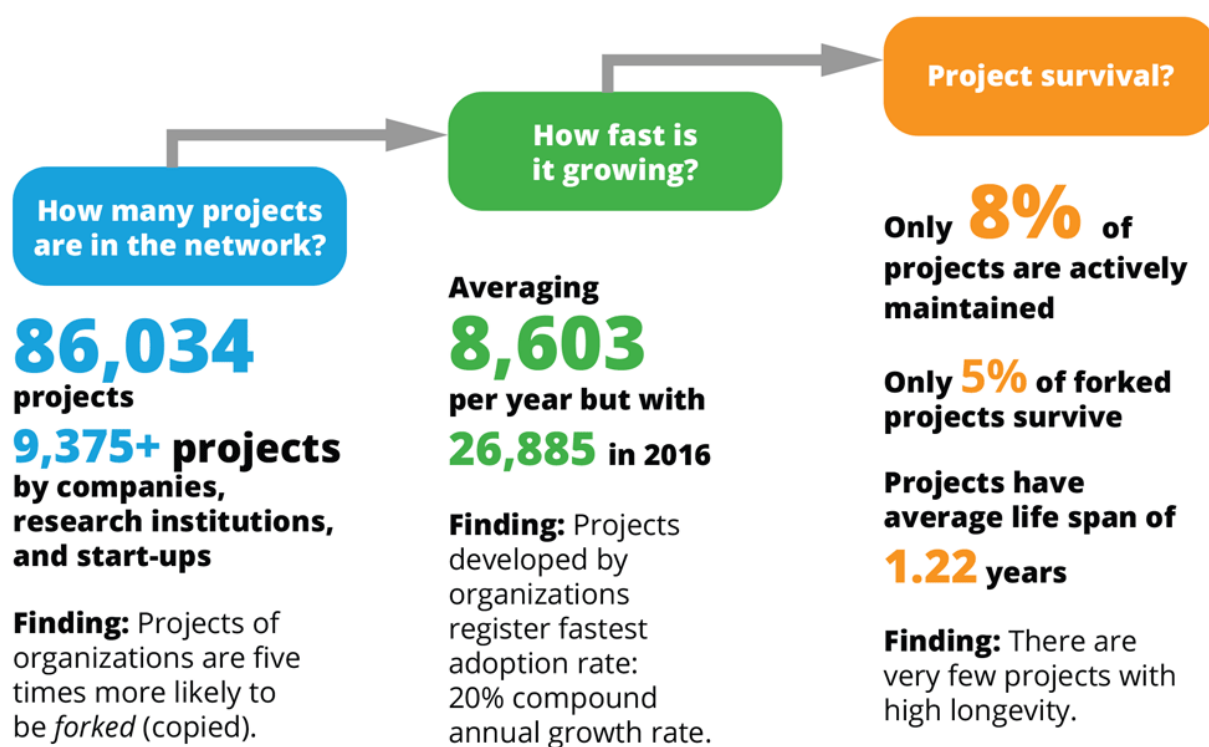


[[View interactive graphic fullscreen](#)]

Repositories reveal interesting trends about organizations

The core code supporting Bitcoin was published in April 2009. Since then, the number of projects on GitHub related to blockchain has grown significantly, averaging more than 8,600 new projects a year. In 2016 alone, there were almost 27,000 new projects (figure 3).

Figure 3. Blockchain on GitHub



Source: Deloitte analysis of GH Torrent data and GitHub API data, as of October 12, 2017.

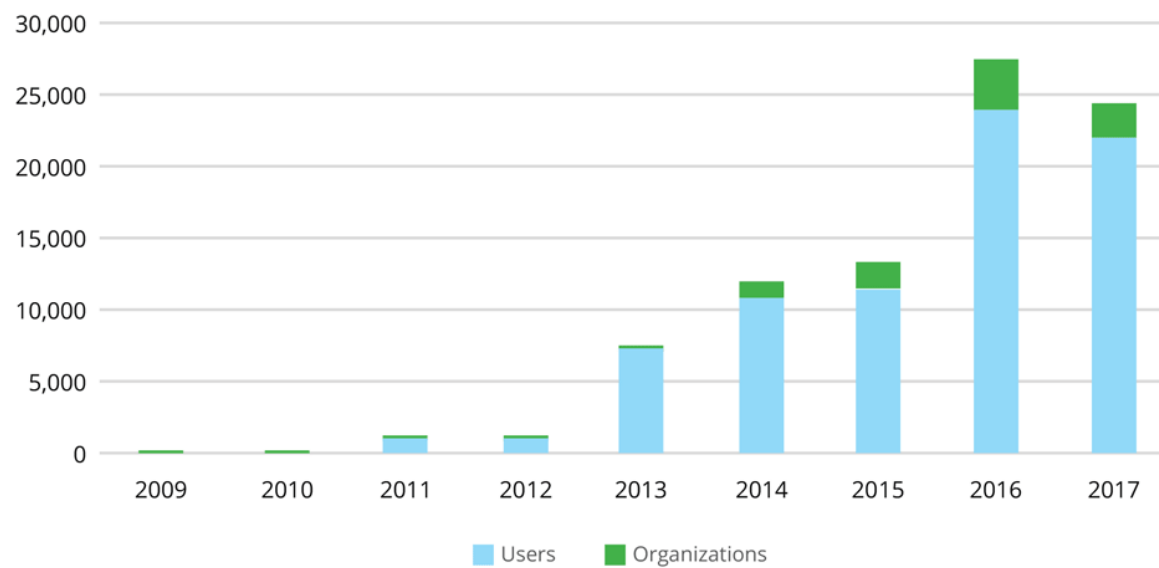
Deloitte Insights | deloitte.com/insights

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Fig3.png>

The growth in the number of projects has been matched by the rapid growth of content produced to develop these blockchain technologies. Please see figure 4, and [Repositories by year](#) in

our interactive dashboard.

Figure 4. New GitHub projects by author type, 2009–2017*



*Data for 2017 includes only the first six months of the year.
Source: Deloitte analysis of GH Torrent data and GitHub API data.

Deloitte Insights | deloitte.com/insights

https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Fig4.png

In analyzing blockchain repositories and their content, we noticed that increasingly more organizations appear to be getting involved. In 2010, organizations developed less than 1 percent of all projects. By 2017, their blockchain projects accounted for 11 percent (organizations currently account for 7 percent of total—not just blockchain—software development on GitHub). And recent data about the rate at which commercial organizations can find success with blockchain initiatives through open source seems promising; some high-profile, large commercial entities are already doing so. (Please refer to Repositories by organization in our interactive dashboard.)

Of particular significance, some projects that organizations have developed have resulted in new platforms (such as Ethereum, Ripple, Corda, and Quorum) which some developers now use to build applications. Organization-owned projects also tend to be updated more frequently than those developed by users, and are reportedly five times more likely to be copied, implying that the blockchain community has deemed them most relevant.

When a project is copied, all of the content becomes available to the account that copied the project, thus working as a de facto knowledge-transfer mechanism. This process is commonly referred to as a citation network (see Appendix for network definitions), ¹¹ where projects that are most often copied occupy a more central role in the network of projects, which we refer to as project centrality.

Under this rubric, some of the most central projects have been developed and maintained by organizations: Bitcoin core, the C++ and Go implementation of Ethereum, Python clients for Ethereum, and the Bitcoin Improvements Proposal. To interactively explore a depiction of the various networks in GitHub, please see [Network visualization](#) in our interactive dashboard.

When exploring the aforementioned interactive graph, keep in mind that the initial projects of Ethereum and Bitcoin are maintained by organizations (foundations), and that a vast amount of blockchain projects and applications in GitHub are actually built on top of these two projects. In short, organization-led projects are the backbone

code for thousands of other projects. Out of the 20 most central projects in the blockchain space measured by popularity, citation, and collaboration (see Appendix for network definitions), 18 were created and maintained by organizations (see table 1).

Table 1. Top 20 central repositories in the blockchain ecosystem

Project name	Author	User type	Total copies	Followers	Total contributors	Description
bitcoin	Bitcoin	Organization	7,588	11,729	627	Source code behind Bitcoin
go-ethereum	Ethereum	Organization	1,717	5,603	149	Official Go implementation of the Ethereum protocol
bitcoinjs-lib	bitcoinjs	Organization	500	1,478	62	Bitcoin-related functions implemented in pure JavaScript
Electrum	spesmilo	Organization	534	1,028	187	Bitcoin thin client on Electrum (wallet)
cpp-ethereum	Ethereum	Organization	905	1,332	119	Ethereum C++ client
bips	Bitcoin	Organization	527	762	158	Bitcoin improvement proposals
bitcoinj	bitcoinj	Organization	928	1,243	106	Java implementation of Bitcoin
Rippled	Ripple	Organization	397	1,267	54	Decentralized cryptocurrency blockchain daemon implementing the Ripple Consensus Ledger in C++
mist	Ethereum	Organization	612	2,752	47	Browse to explore Decentralized Apps (DAPPs) built on Ethereum
truffle	ConsenSys	Organization	181	768	31	Development environment, testing framework, and asset pipeline for Ethereum
pyethereum	Ethereum	Organization	348	1,135	57	Python core library of the Ethereum project
cgminer	Ckolivas	User	737	1,760	88	ASIC and FPGA miner in C for Bitcoin
ethereumj	Ethereum	Organization	350	598	58	Java implementation of the Ethereum yellowpaper
btcd	Btcsuite	Organization	380	1,272	59	An alternative full node Bitcoin implementation written in Go (golang)
testrpc	ethereumjs	Organization	154	559	33	Fast Ethereum RPC client for testing and development
bitcoinbook	bitcoinbook	Organization	735	2,204	40	Bitcoin book
ripple-client	Ripple	Organization	500	1,239	54	A UI for the Ripple payment network built using web technologies
EIPs	Ethereum	Organization	148	650	38	The Ethereum improvement proposal
embark-framework	iurimatias	User	98	515	37	Framework for serverless Decentralized Applications using Ethereum, IPFS, and other platforms
bcoin	bcoin-org	Organization	121	649	20	JavaScript Bitcoin library for node.js and browsers

Source: Deloitte analysis of GH Torrent data and GitHub API data.

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Table1.png>

Organizational commitment in open source appears to be dominating the core development of blockchain because it is most likely more demanding and purposeful than individual participation in development. Once resources are put into place by an organization, there is generally more incentive to ensure that the project is successful. Given that organization participants are tied to one another beyond any particular project, there is often greater accountability to one another, which also drives ongoing development.

BLOCKCHAIN EVOLUTION IS LARGELY ABOUT THE “COMMUNITY”

A *community* on open source is a group of developers with shared interests that develops and improves existing content. We identified 772 different blockchain communities on GitHub. Each community is typically defined by patterns of collaboration between these projects that can give rise to new applications. For example, the Ethereum platform was initially developed by two central figures in the Bitcoin project; their project has since evolved into the largest blockchain community, measured by active projects, on GitHub (see the sidebar, “Understanding the Ethereum ecosystem”).

In the blockchain space, communities of projects comprise at least 25 projects, with some large clusters including hundreds of projects (see [Communities of repositories](#) in our interactive dashboard).

By studying communities, we can explore how projects that have developed a specialization enable the creation of new applications. For example, we found that tools for enabling crowdsales and initial coin offerings (ICOs) are often connected to projects in large blockchain subcommunities: projects developing content for smart contracts, escrow accounts, and the core code behind Ethereum in the Go language. Not surprisingly, this seems to align with the predilection of many ICOs being offered on top of the Ethereum blockchain (for more information on ICOs, please read **“Initial coin offering: A new paradigm,” Deloitte**

<https://www2.deloitte.com/us/en/pages/consulting/articles/initial-coin-offering-a-new-paradigm.html> ¹²). Ethereum allows developers and start-ups to issue their own currency, including in the form of an ICO, on the Ethereum blockchain through smart contracts, which can seriously reduce the token and cryptocurrency barrier to entry. ¹³

An interesting example of how seemingly disparate communities connect is the Monero cryptocurrency, created in 2014. Monero has concerted different attributes than Bitcoin regarding its level of privacy (no reuse of addresses allowed), scalability (no blocksize limit), and security (more forced decentralization). ¹⁴ Still, the community that contains Monero and related projects has a strong connection to the community that contains the main Bitcoin repository.

It could potentially be especially important for blockchain

developers to pay close attention to communities. Our analysis reveals that many projects that specialize in particular industries or types of applications in the blockchain space that are enriching the ecosystem have strong community affiliations.

INTERPRETING THE MORTALITY RATE OF BLOCKCHAIN PROJECTS

The stark reality of open-source projects is that most are abandoned or do not achieve meaningful scale. Unfortunately, blockchain is not immune to this reality. Our analysis found that only 8 percent of projects are active, which we define as being updated at least once in the last six months. Here, organizations are a positive differentiator; while 7 percent of projects developed by users are active, 15 percent of projects developed by organizations are active.

The mortality rate of projects is often an essential factor in understanding project centrality and the emergence of protocols and best practices. For commercial purposes, since few projects will likely survive, understanding the factors that contribute to a project's mortality may be an essential skill for firms wishing to piggyback on a successful code, emulate successful projects, or build in-house capabilities. ¹⁵ Note that about 90 percent of projects developed on GitHub become idle, and the average life span of a project is about one year, with the highest mortality rate occurring within the first six months. Our analysis revealed 11 variables associated to a project becoming inactive. Of these variables, organizations should consider the following three in particular:

- First, perhaps not surprisingly, *projects developed by users* tend to have shorter life spans and to be stand-alone blockchain technology applications rather than foundational libraries that enable the creation of multiple applications. We found that users were more prone to tinker, developing and prototyping ideas that often do not gain traction.
- Second, is *the concentration of contributions to a project* . Our analysis indicates that most projects with high mortality rates are those in which one committer generated the vast majority of content, and, in fact, the median project (measured by number of committers) has only one committer (see Glossary in Appendix for definitions of types of repository participants). Often this sole, or lead, developer shifts attention to other projects or simply becomes disengaged. On the other hand, most projects that survive tend to have multiple committers with less concentration of activity attributed to one particular committer. Predictably, most organizations are structured in this latter manner.
- Third, *projects with higher numbers of copies (forks) of a project* are more prone to survive while those with few or no copies are more prone to stall as will their copies. In fact, regarding the latter, a copy of a project also tends to have less committers and hence higher mortality rates. This is not to say that there are not highly successful projects that are the products of forks, but generally projects that are forks of projects tend to have high

mortality rates due to a paucity of new committers.

Programming languages lean toward financial services

For potential developers, the question that often surfaces first is, “How should we start?” To help answer that question, it can be important to find out what’s under the hood of existing projects. Although not the most popular language when measured by number of blockchain repositories, we found that C++ was used most in the ecosystems’ central repositories. This was not surprising, given that C++ has been used for some time in the financial services industry to develop applications that demand efficient memory management, speed, and reliability. For the heavy lifting behind cryptocurrency projects (Bitcoin included), C++ is still the favored language. And for the most central repositories on GitHub, C++ accounts for almost one-half of all the content (see [Most popular languages](#) in our interactive dashboard.).

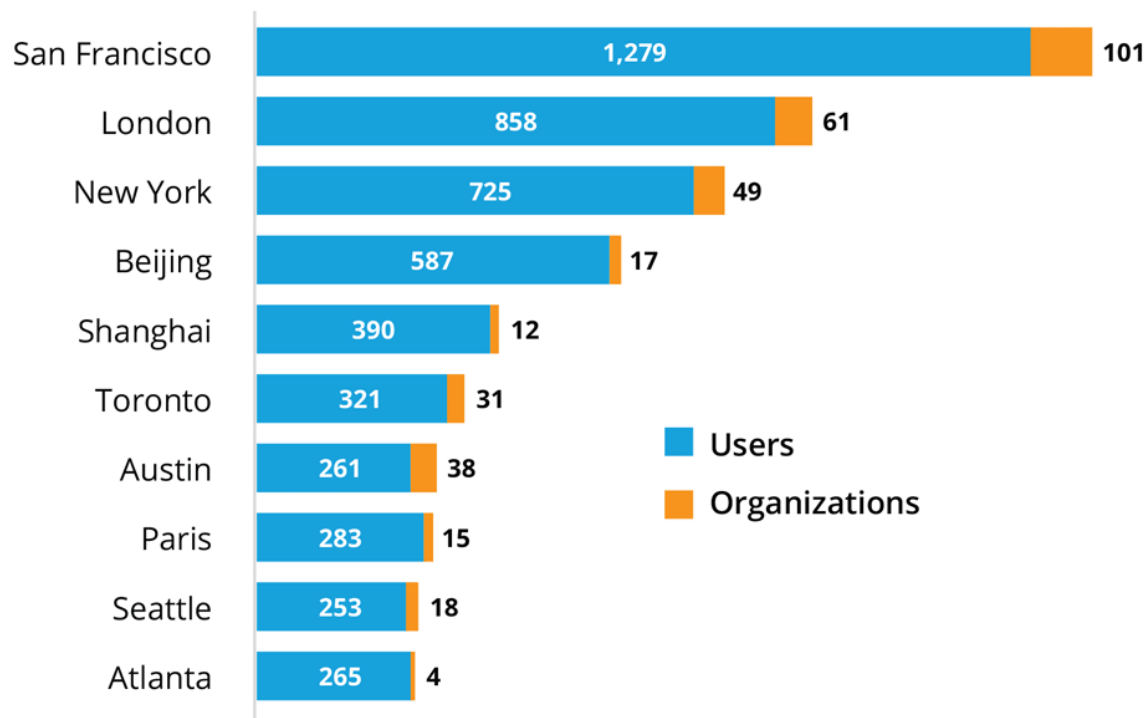
However, we also discovered that Go, the programming language developed by Google in 2009, appears to be gaining traction. It is now the second largest language used for blockchain-related projects. Go seems to have rapidly evolved from a fringe language to one of the centerpieces of the GitHub blockchain ecosystem. Just two years ago, in 2015, less than 2 percent of all of the content of projects in the blockchain space were developed in Go. Programmers attribute the meteoric rise of Go to its simplicity and

ability to scale. ¹⁶ And while financial services firms do reportedly rely on the memory management, speed, and reliability of C++, scalability appears to also be an exceptionally high priority for financial services firms that interact and transact with multiple and diffuse stakeholders. It also seems telling that Ethereum and Hyperledger projects, which both involve integrating other technologies into blockchain to expand its use beyond cryptocurrencies, reportedly favor Go.

Identifying blockchain talent by geography

Given that an important issue that financial institutions face is hiring the necessary talent to develop, implement, or maintain new technologies, we thought it would be helpful to know where top blockchain talent contributing to GitHub resides. Most GitHub project owners—developers who start repositories—live in North America or Europe, with San Francisco having the largest concentration. Interestingly, the next two most popular cities to find project owners are two traditional financial services hubs: London and New York. (See figure 5 and [Repositories by geography](#) in our interactive dashboard.)

Figure 5. Top 10 cities for blockchain development (number of projects)¹⁷



Source: Deloitte analysis of GH Torrent data and GitHub API data, as of October 12, 2017.

Deloitte Insights | deloitte.com/insights

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Fig5.png>

We found that projects coming from San Francisco are diverse; they include solutions for exchanges, wallets for cryptocurrencies, interfaces for different blockchains (for example, Ripple, Hyperledger, and Ethereum), and payment tools for cryptocurrencies, to name a few. The ecosystem in London is also varied, but features more projects connected to the Ethereum community, which would also imply more projects around accompanying technologies, such as digital identities, smart contracts, and open APIs. Participants in New York appear to be specializing in projects that are geared toward traditional financial services. It is also worth noting the high level of activity in China, specifically, Shanghai and Beijing. In both of these cities, most of the

projects pertain to cryptocurrencies and cryptocurrency exchanges, with an emphasis on scalability.

UNDERSTANDING THE ETHEREUM ECOSYSTEM

The Ethereum project is a decentralized platform for blockchain applications based on smart contracts. In 2013, Vitalik Buterin, an active Bitcoin developer, proposed the idea that became Ethereum; his goal was to help create applications that use blockchain technology beyond the cryptocurrency sphere. From its inception, Ethereum was designed to be a blockchain protocol that could enable any application to be written on top of it. ¹⁸ The Ethereum platform is composed of a

virtual machine that executes smart contracts (for an explanation of a

Show more



How could financial services use the GitHub analysis?

The data scientists of Deloitte developed and honed a methodology to analyze and organize GitHub data in order to better understand the evolution of a young, possibly transformative technology and its ecosystem. Our overall objective is to provide insights that help financial institutions make better, more informed decisions and avoid pitfalls.

From this effort, we have learned that financial services firms are involved in blockchain development on GitHub. There are essentially two types of participators on GitHub: the *committer* and the *watcher*. The committer makes commits, or contributions to code, while the watcher follows the development of a project

without making code contributions. So far, few financial services firms' employees are committers to projects on the firms' behalf. There are, however, a few high-profile financial services firms that not only commit, but actually have their own projects running under their brand with significant commits.

Nonetheless, financial services firms seem predominantly engaged as watchers of projects in GitHub. It is difficult to get an actual number on these watchers as they can be watching under handles or private email addresses. Regardless, our analysis can equip both financial services committers and watchers with perhaps a unique opportunity to gain access to a large and nuanced view of the blockchain ecosystem. Leveraging our analytical methodology, firms can now target multiple projects for possible involvement or learning, identify talent using a variety of metrics, see how changes in protocol and trends can point toward standardization and interoperability, and, finally, all of this and more can increase their understanding of blockchain's evolution.

Specifically, our analysis may enable financial institutions, and other firms, to:

- Identify pockets of opportunities for future innovation—where to invest, how much, and when
- Determine where competitors are playing already and identify gaps
- Understand and predict which languages are gaining/losing

ground using which types of applications, and invest accordingly

- Determine where talent and expertise exist and how best to leverage the geographic distribution of talent
- Assess partnerships and collaboration opportunities

It is our hope that these findings can arm the financial services industry with the data it may need to not only better identify successful projects and opportunities based on how the blockchain ecosystem is evolving, but to become influential participants, themselves, in how blockchain evolves.

Appendix

NETWORK ANALYSIS

We use several metrics commonly used in the field of network analysis, such as number of connections (*degree*), centrality (PageRank score), and clustering (*community detection*). ²¹ We defined three types of network connections in our analysis:

Collaboration measures the contribution of projects to each other. To build this network, we identified the repositories that received collaborations from each other in our universe of blockchain projects rather than the entire GitHub set.

Citation measures the use of a project's content by another project. Projects that are highly cited tend to have a high centrality score (see the next section). To build the network, we identified users who copied a repository, joining his or her projects with the

project that he or she copied.

Followers measures the popularity of a project within other projects. To build the network, we identified users who followed a given repository in our universe and joined these users' projects.

CENTRALITY ANALYSIS

To identify the most central repositories in our network, we used the PageRank (PR) score. Developed by Google, PageRank is a common metric to identify centrality in a network and has been widely used in several fields. We calculated the PR score for each of our three networks. ²² Once we obtained the PR score, we ranked the projects based on the value of that metric. We repeated the process for the three networks and created a composite score defined accordingly:

$$\text{Centrality score} = \frac{1}{(\text{Collaboration ranking} + \text{Citation ranking} + \text{Follower ranking})}$$

<https://dupress.deloitte.com/content/dam/dup-us-en/articles/4107_GitHub-blockchain/figures/4107_Centrality.png>

Centrality Score values that are closer to 1 indicate a more central role in the network.

IDENTIFYING COMMUNITIES

To identify communities in our network, we implemented a commonly used community detection algorithm for large graphs known as *fast greedy community detection* algorithm. ²³ The algorithm iterates through the different network connections, adding projects to a community until a local optimal value is reached. The algorithm repeats this process until there are no further improvements. We implemented the algorithm in the collaboration network.

PROJECT MORTALITY

To identify factors associated to a project becoming inactive, we implemented two classification models: a logistic regression and a random forest. ²⁴ The logistic regression was used to identify meaningful variables while the random forest is used to identify which projects become idle.

Authors

Jesus Leal Trujillo < <https://dupress.deloitte.com/dup-us-en/authors/t/jesus-leal-trujillo.html> > is a senior consultant and data scientist working for the research and insights group at Deloitte Services LP.

Stephen Fromhart < <https://dupress.deloitte.com/dup-us-en/authors/f/steve-fromhart.html> > is a manager at the Deloitte Center for Financial Services, Deloitte Services LP, covering the banking and capital markets sectors.

Val Srinivas < <https://dupress.deloitte.com/dup-us-en/authors/s/val-srinivas.html> > is the banking and securities research leader at the Deloitte Center for Financial Services, Deloitte Services LP.

Acknowledgements

The authors wish to thank the following Deloitte client service professionals for their insights and contributions to the report: **Dan Byler** , manager, Deloitte Services LP, and **Jason Lewris** , data scientist, Deloitte Services LP.

The authors also thank the following Deloitte professionals for their support and contributions to the report:

Cover image by: **Lucy Rose** <<http://lucyrozeillustration.com/>>

Topics in this article

Blockchain <<https://dupress.deloitte.com/dup-us-en/tags/blockchain.html>> , **Financial Services** <<https://dupress.deloitte.com/dup-us-en/tags/financial-services.html>> , **Disruptive innovation** <<https://dupress.deloitte.com/dup-us-en/tags/disruptive-innovation.html>> , **Emerging technologies** <<https://dupress.deloitte.com/dup-us-en/tags/emerging-technologies.html>>

Endnotes

- 1 GitHub, “ **GitHub is how people build software** <<https://github.com/about>> ,” accessed September 7, 2017.
- 2 While there are other popular software development platforms such as BitBucket, GitLab, and SourceForge, among others, none gained the widespread adoption of GitHub. See Cade Metz, “ **How GitHub conquered Google, Microsoft, and everyone else** <<https://www.wired.com/2015/03/github-conquered-google-microsoft-everyone-else/>> ,” *Wired* , March 12, 2015.
- 3 Gousios, Georgios, *The GHTorrent dataset and tool suite*, Proceeding , 2013, MSR '13 proceedings of the 10th Working Conference on Mining Software Repositories, pp. 233–236.
- 4 Michael Schwarz and Yuri Takhteyev, “Half a century of public software institutions: Open source as a solution to hold-up problem,” *Journal of Public Economic Theory*, Oxford 12.4, pp. 609–639, August 2010.
- 5 Aaron Weiss, “Open source moves to the mainstream,” *InformationWeek* , April 10, 2000.
- 6 Schwarz and Takhteyev, “Half a century of public software institutions: Open source as a solution to hold-up problem,” pp. 609–639 .
- 7 “ **History of open source** <<https://www.longsight.com/learning-center/history-open-source>> ” Longsight, accessed August 22, 2017.
- 8 Michael Widenius and Linus Nyman, “ **The business of open source software: A primer** <<https://timreview.ca/article/756>> ,” *Technology Innovation Management Review* , January 2014.
- 9 Joichi Ito, Neha Narula, and Robleh Ali, “ **The blockchain will do to the financial system what the Internet did to media** <<https://hbr.org/2017/03/the-blockchain-will-do-to-banks-and-law-firms-what-the-internet-did-to-media>> ,” *Harvard Business Review* , March 8, 2017.
- 10 For a comprehensive glossary of GitHub terms, refer to the organization’s glossary: <https://help.github.com/articles/github-glossary/> <<https://help.github.com/articles/github-glossary/>>
- 11 For more information on network analysis on GitHub, please see Dan Byler and Tiffany Schleeter, *The arc of innovation in the oil and gas industry* , Deloitte University Press, May 2017.

- 12 Rob Massey, Darshini Dalal, and Asha Dakshinamoorthy, “ **Initial coin offering: A new paradigm** <<https://www2.deloitte.com/us/en/pages/consulting/articles/initial-coin-offering-a-new-paradigm.html>> ,” Deloitte, August 2017.
- 13 Steven McKie, “ **Understanding the Ethereum ICO token hype** <<https://medium.com/blockchannel/understanding-the-ethereum-ico-token-hype-429481278f45>> ”, *Medium* , June 2014, accessed on October 3, 2017.
- 14 Olga Kharif, “ **Investors who missed Bitcoin rally go for Dash, Ether, Monero** <<https://www.bloomberg.com/news/articles/2017-03-08/investors-who-missed-bitcoin-rally-go-for-dash-ether-monero>> ,” *Bloomberg* , March 8, 2017.
- 15 We define stalled projects as repositories that have not had a contribution either by the owner of the project or by the community in more than 6 months.
- 16 Tyler Treat, “ **Go is unapologetically flawed, here’s why we use it** <<http://bravenewgeek.com/go-is-unapologetically-flawed-heres-why-we-use-it/>> ,” *Brave New Geek* , May 20, 2015.
- 17 We look at the projects for which their owners reported their locations.
- 18 Deep Patel, “ **Business in the age of Ethereum** <<https://techcrunch.com/2017/06/04/business-in-the-age-of-ethereum/>> ,” *TechCrunch* , June 4, 2017.
- 19 J.P. Buntinx, “ **What is the Ethereum Virtual Machine?** <<https://themerple.com/what-is-the-ethereum-virtual-machine/>> ,” *The Merkle* , May 1, 2017.
- 20 Robert Hackett, “ **Big business giants from Microsoft to J.P. Morgan are getting behind Ethereum** <<http://fortune.com/2017/02/28/ethereum-jpmorgan-microsoft-alliance/>> ,” *Fortune* , February 28, 2017.
- 21 For an overview of the methods described see Mark Newman, *Networks: An Introduction* , Oxford: Oxford University, 2010.
- 22 Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, “ **The PageRank citation ranking: Bringing order to the web** <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.775&rep=rep1&type=pdf>> , Stanford University InfoLab, January 29, 1998.
- 23 Newman, M.E.J., Girvan, M., “Finding and evaluating community structure in networks,” *Phys Rev E* 69, 026113, 2004.
- 24 For an overview of the methods described see Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning: Data mining, inference, and prediction* , Standford University, New York: Springer, 2016.

© 2017. See **Terms of Use** <<https://www2.deloitte.com/us/en/legal/terms-of-use.html>> for more information.

© 2017. See **Terms of Use** <<https://www2.deloitte.com/us/en/legal/terms-of-use.html>> for more information.

