WEBINAR SUMMARY

NOVEMBER 21, 2024

DARKREADING

Safeguarding GitHub Data to Fuel Web Innovation

Presenters: **James Ciesielski,** CTO and Co-Founder, Rewind
Vivien Lacourba, Head of Systems Team, W3C
Denis Ah-Kang, Web Developer & Systems Engineer, W3C

Moderator: **Terry Sweeney,** Contributing Editor, Dark Reading

KEY TAKEAWAYS

- In the Shared Responsibility Model, safeguarding account-level data falls squarely on the user.
- Protecting the code base from disruption depends on a well-executed backup and recovery strategy.
- Build or buy?
- Rewind's industry-leading solution helps customers protect against data loss.

in partnership with

rewind

OVERVIEW

Data breaches and accidental repository deletion are just two threats that put GitHub data—and businesses—at risk. Most users do not understand the Shared Responsibility Model, which means that SaaS providers such as GitHub only protect their platform with system-wide disaster recovery. Responsibility for safeguarding account-level data falls squarely on the user.

For World Wide Web Consortium (W3C), an international multi-stakeholder community that develops open web standards, data security is always top of mind. When its in-house backup and recovery solution was unable to adequately address the organization's complex data security needs, W3C turned to Rewind. Rewind offers a robust data backup and recovery solution for SaaS applications such as GitHub, helping to prevent data loss. Rewind provides data recovery within minutes, minimizing service downtime, wasted developer cycles, and lost revenue due to security incidents, human error, or other threats.

KEY TAKEAWAYS

In the Shared Responsibility Model, safeguarding account-level data falls squarely on the user.

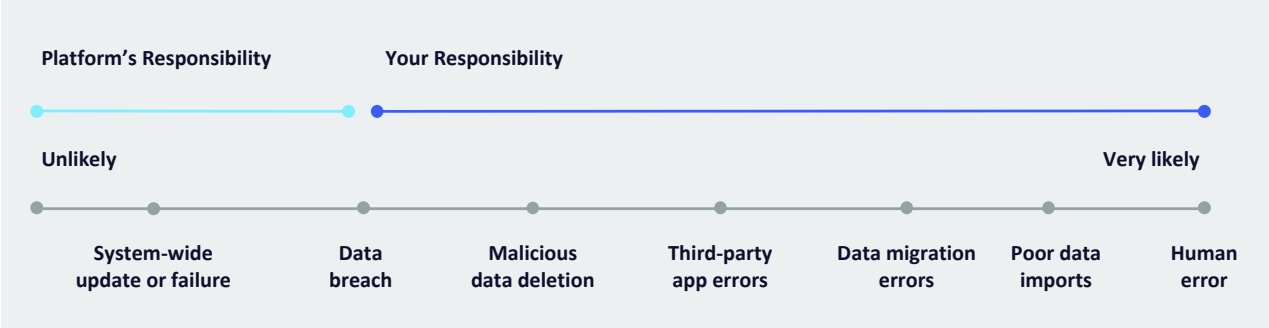
For organizations using cloud-based services such as GitHub, understanding their security responsibility is crucial to preventing data loss. Cloud-based vendors use language in their terms and conditions that represents the Shared Responsibility Model, which describes where the division of security and operational responsibility exists between a cloud service provider (including SaaS) and its customers.

Of critical importance is understanding that the responsibility for safeguarding account-level data falls on the user.

> "The cloud service provider [is] responsible for their infrastructure, and you, the user, are responsible for everything you do within that infrastructure."
>
> *James Ciesielski, Rewind*

In the results of a Forrester study, 95% of companies using SaaS incorrectly assumed the SaaS provider could recover data in the event of data loss. Misunderstanding the Shared Responsibility Model can lead to significant problems, as evidenced by the 58% of companies that reported SaaS data loss in the last 12 months.

Figure 1: Visualizing the Shared Responsibility Model



Protecting the code base from disruption depends on a well-executed backup and recovery strategy.

Whether to protect the code base or any other critical system, the most effective security is a multi-layered effort. This includes enforcing multi-factor authentication, practicing least-privilege access controls, conducting tabletop activities and regular security audits, and maintaining a strong incident response plan anchored in a backup and recovery strategy.

When it comes to backups, Gartner reported that in 2024, only 15% of enterprises prioritize backups of their SaaS applications. However, it projects that by 2028, that number will rise to 75% of enterprises. Unfortunately, it often takes an incident before a company reconsiders its approach to backup and recovery.

Getting ahead of data loss should be a priority in every organization's disaster recovery plan. Developing and implementing a backup strategy will significantly reduce the likelihood of a major disruption and ensure teams can stay productive—even in the event of a security breach.

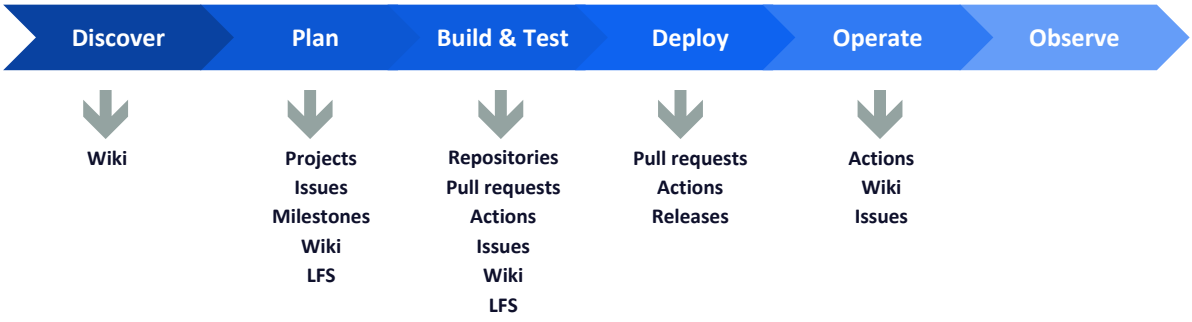
> "Backups are not really just about recovery. They're about resilience—and resilience is what keeps your team moving forward, no matter what comes your way."
>
> *James Ciesielski, Rewind*

Preventing data loss is especially critical for data stored in GitHub. Organizations of all sizes and across all industries use GitHub not only as a source code repository, but also in other myriad ways throughout the software development lifecycle, such as for capturing critical documentation, recording information about product releases, managing infrastructure, and more. The loss of data associated with projects, pull requests, wikis, and other functions would often have a severe, if not catastrophic, impact on business operations.

GitHub backups are key to safeguarding source code and enabling rapid recovery of the code base, which helps to minimize loss of productivity and reputational damage.

> "A lot of different critical components outside of source code live within GitHub today. . . What would happen if any one of those things disappeared? Usually, the impact is orders of magnitude beyond what you would initially think."
>
> *James Ciesielski, Rewind*

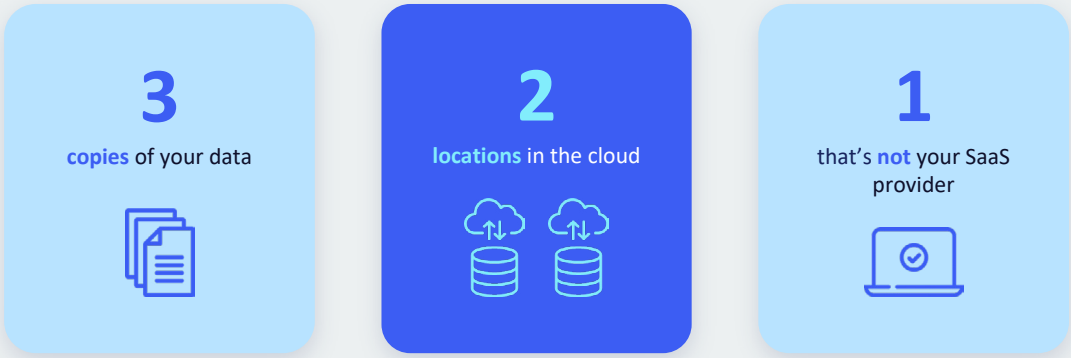
Figure 2: GitHub plays a fundamental role across the software development lifecycle



Backup Best Practices

Best practices recommend the 3-2-1 rule for SaaS backups: Have 3 copies of the data in 2 different locations in the cloud, of which 1 location cannot be the SaaS provider.

Figure 3: The 3-2-1- backup rule



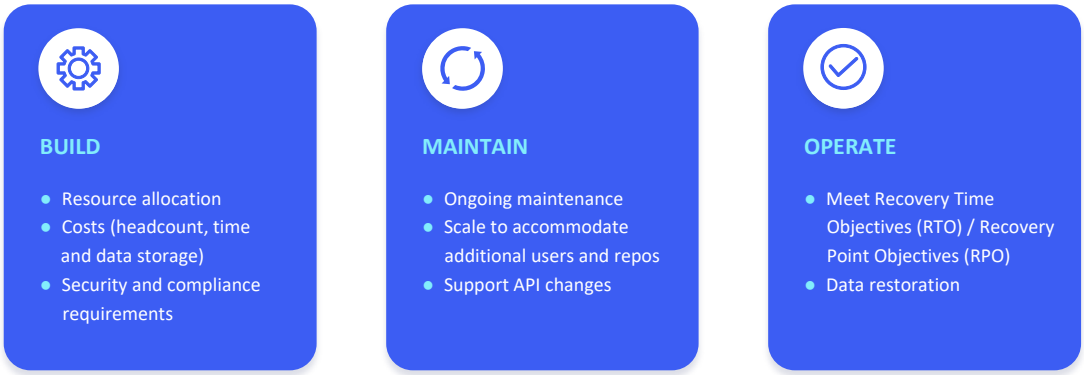
Build or buy?

When considering the best backup strategy, enterprises are often faced with the question: Build or buy?

Both offer different benefits.

- **Build:** Building in-house allows specific controls to be implemented in ways that are meaningful to the business. Knowledgeable in-house engineers can usually build a custom solution quickly and at relatively low cost. However, costs can rise significantly over time. Organizations increasingly depend on SaaS products for business operations. These products are constantly changing, requiring corresponding consistent and often frequent updates to custom solutions. Companies with large GitHub accounts require a solution that can scale, requiring additional costly in-house storage. And ensuring security of the solution against continuously evolving cyberthreats adds even greater cost to an in-house service.
- **Buy:** Buying a solution from a third-party vendor enables businesses to access a solution already refined by industry expertise. A third-party solution allows companies to delegate responsibility to the provider to meet new security needs and challenges, rather than having to invest internal resources into maintaining safe, secure backups. Shifting focus away from maintaining an in-house backup service offers a low-maintenance solution for data loss protection and allows organizations to invest their valuable resources in improving the core business, instead.

Figure 4: Building a backup solution requires significant initial as well as ongoing investment



Rewind's industry-leading solution helps customers protect against data loss.

Rewind prioritizes security at every level, offering a scalable, always up-to-date backup and recovery solution for customers worldwide. And Rewind's configurability ensures that every business can implement a data security approach that works best for that organization. Out of the box, Rewind provides customers the ability to decide where their GitHub data gets stored.

Rewind's Cloud Sync automatically mirrors companies' data to a big three cloud provider of their choice: Azure, AWS, or Google Cloud. Cloud Sync makes it easy for organizations to follow the 3-2-1 rule for SaaS backups, storing three copies of data in two different locations in the cloud, one of which is not in the Rewind infrastructure.

> "Some SaaS products offer a rudimentary backup offering, but we liken that to backing up your hard drive to your hard drive. . . They're not offering you the best possible solution."
>
> *James Ciesielski, Rewind*

Case Study: W3C

W3C, the World Wide Web Consortium, is an international nonprofit created in 1994 by Tim Berners-Lee. W3C has over 350 member organizations, including large companies, such as Apple, Amazon, Mozilla, Google, Alibaba, Sony, and the BBC, as well as smaller companies, research centers, and universities.

Approximately 15,000 web experts comprise the W3C community, working together to develop the standards of the web and make them available under the W3C royalty-free license. Some of W3C's best-known technical standards include HTML, CSS, SVG, XML, and WebRTC.

As the web evolved, growing into a stable, robust platform for users and businewsses, W3C began expanding its use of third-party providers, eventually shifting some operations to SaaS solutions. Several years ago, W3C moved its specification development to GitHub. Today, W3C's GitHub usage is extensive, with over 1,800 repositories shared among more than 20 GitHub organizations and accessible by 3,000+ GitHub members with different privileges.

"Technical standards—that's a collaborative effort, and the GitHub platform is a perfect fit. . . It's also best-known to a lot of developers."

Vivien Lacourba, W3C

Mistakes can and do happen. But while some mistakes create minor inconveniences, others can have significant consequences. Early in its use of GitHub, W3C experienced a significant loss due to human error—a major repository was accidentally deleted by a user with admin permissions. At the time, W3C had not implemented proper backups. Fortunately, the problem was identified shortly after it occurred and GitHub support was able to restore the repository.

The incident spurred W3C to implement a backup solution to protect the organization against data loss or if a future move away from GitHub became necessary.

Initially, W3C built custom scripts in-house to query the GitHub API on a regular basis; however, the number of repositories quickly maxed out the number of allowed requests, rendering the approach impractical. The solution also had issues with the restoration process, as some data was missing from the pull request. To keep up with the API update—as GitHub regularly adds new features—required more resources than W3C could spare.

Given the importance of GitHub for the organization, W3C decided to pivot, identifying Rewind as the best solution to ensure proper backups. Rewind's extremely easy-to-use user interface, multiple backup versions, and support for API changes addressed the key challenges W3C was experiencing with custom scripts. Additionally, Rewind's user management system made it easy for W3C to manage access to backups for specific users. Today, W3C has migrated all of its GitHub organizations to Rewind.

"Rewind provides us with the scalability, security, storage and support we need. With our growing list of repositories, we don't need to worry about API rate limiting or storage space anymore. We can focus on other projects that make the web a better place."

Denis Ah-Kang, W3C

W3C

Over 350 member organizations

Approximately 15,000 web experts

Over 1,800 repositories shared

3,000+ GitHub members can access

Figure 5: Key benefits of Rewind for W3C's backup and recovery of GitHub

Easy backup/restoration process

Keep up with the GitHub API updates

Users management

Security

ADDITIONAL INFORMATION

To learn more, visit W3C and Rewind

BIOGRAPHIES

James Ciesielski

CTO and Co-Founder
Rewind

After completing a Bachelor of Math, Computer Science/Software Engineering at the University of Waterloo, James has over 20 years of experience building highly scalable software and services in the fields of telecommunications, media, and financial technology in both enterprise and start-up environments. An experienced technical leader, James has successfully overseen the development and launch of a variety of software products.

Vivien Lacourba

Head of Systems Team
W3C

Vivien joined W3C in May 2003 as the W3C Webmaster at the MIT/CSAIL host site in Cambridge, MA, USA. Since September 2004 Vivien is working as a Systems & Network Engineer for W3C Europe at the ERCIM host site in Sophia-Antipolis, France. Vivien graduated in September 2003 from the Polytech Nice Sophia engineering school (formerly known as ESSI) in Sophia-Antipolis, France.

He holds an engineering degree in Computer Science, specializing in Networks. In June 2000, he received a two year degree in Computer Programming at the University of Lyon, France.

Denis Ah-Kang

Web Developer & Systems Engineer
W3C

Denis joined W3C in August 2011, as part of the Systems Team, to become the W3C Webmaster at the MIT host site in Cambridge, MA, USA. Between 2013 and 2014, he joined the Interaction Domain to work on the HTML5 test suite. He is now working on maintaining the W3C infrastructure and is involved in the development of the publications tools. Prior to join W3C, Denis worked for various consulting companies as a software developer.

Denis is currently based in Reunion Island.

Terry Sweeney (Moderator)

Contributing Editor
Dark Reading

Terry Sweeney is a Los Angeles-based writer and editor who's covered business technology for three decades. He's written about cyber security for more than 15 years and was one of the founding editors of Dark Reading. Sweeney has covered enterprise networking extensively, as well as its supporting technologies like storage, wireless, cloud-based apps and the emerging Internet of Things. He's been a contributing editor to The Washington Post, Crain's New York Business, Red Herring, Information Week, Network World, SearchAWS.com, and Stadium Tech Report.