A Wayback Machine for Source Code

...The “Software Heritage” project (https://www.softwareheritage.org/) is a sort of Wayback Machine for software. The project plans to create an archive of computer code source files as they appear on the web — an undertaking that has implications not just for history, but for science and research, too.

Since 2015, archivists at the Software Heritage project, which is hosted by the French Institute for Research in Computer Science and Automation, have been collecting open source code available at various online repositories and websites. To date, the archive contains more than 4 billion source files from more than 80 million projects, says Roberto Di Cosmo, a computer scientist who is directing the project in Paris. In cases where open source code disappears, or the server it is stored on is hacked, destroyed or lost, the platform aims to become the go-to place for a backup version.

In the coming weeks, Di Cosmo and colleagues plan to release the archive for anyone to access for the first time. Adding code to the platform, however, will continue in the same fashion, Di Cosmo says. He speculates that the archive currently contains only around a quarter of the world’s open source software, noting that code is often published in hard-to-access places on the internet.

Read more at Undark (https://undark.org/article/software-heritage-source-code-archive/)