

An Empirical Study on the Survival Rate of GitHub Projects*

Adem Ait¹, Javier Luis Cánovas Izquierdo¹, and Jordi Cabot^{1,2}

¹ IN3 - UOC, Barcelona, Spain
{aaait.mimoune,jcanovasi}@uoc.edu
² ICREA, Barcelona, Spain
jordi.cabot@icrea.cat

Keywords: Open Source Analysis, Survival Analysis, Mining Software Repositories, Empirical Study.

Published in: MSR 2022 - Proceedings of 19th IEEE/ACM International Conference on Mining Software Repositories, pp. 365–375, 2022.

Impact Factor: GII-GRIN-SCIE Class 2

DOI: <https://doi.org/10.1145/3524842.3527941>

Abstract. The number of Open Source projects hosted in social coding platforms such as GitHub is constantly growing. However, many of these projects are not regularly maintained and some are even abandoned shortly after they were created. In this paper we analyze early project development dynamics in software projects hosted on GitHub, including their survival rate. To this aim, we collected all 1,127 GitHub repositories from four different ecosystems (i.e., NPM packages, R packages, WordPress plugins and Laravel packages) created in 2016. We stored their activity in a time series database and analyzed their activity evolution along their lifespan, from 2016 to now. Our results reveal that the prototypical development process consists of intensive coding-driven active periods followed by long periods of inactivity. More importantly, we have found that a significant number of projects die in the first year of existence with the survival rate decreasing year after year. In fact, the probability of surviving longer than five years is less than 50% though some types of projects have better chances of survival.

* This work has been partially funded by the Spanish government (LOCOS project - PID2020-114615RB-I00)

