Fork-based (or branch-based) development is a lightweight mechanism that allows developers to collaborate with or without explicit coordination. Recent advances in distributed version control systems (e.g., ‘git clone’) and social coding platforms have made fork-based development relatively easy and popular by providing support for tracking changes across multiple forks with a common vocabulary and mechanism for integrating changes back. However, fork-based development has well-known downsides. When developers each create their own fork and develop independently, their contributions are usually not easily visible to others, unless they make an active attempt to merging their changes back into the original project. When the number of forks grows, it becomes very difficult to keep track of decentralized development activity in many forks. The key problem is that it is difficult to maintain an overview of what happens in individual forks and thus of the project’s scope and direction. Furthermore, the problem of lacking an overview of forks can lead to several additional problems and inefficient practices: lost contributions, redundant development, fragmented communities, and so on.

In this dissertation, I mixed a wide range of research methods to understand the problem space and the solution space. Specifically, I first design measures to quantify how serious are these inefficiencies, then I developed two complementary strategies to alleviate the problem: First, during the process of sampling 1311 GitHub projects and quantifying the inefficiencies, also by reaching out to developers who have used forks, I recognized that there are differences among projects. Therefore, I identified natural interventions and suggesting evidence-based guidance for projects that are inefficient. Moreover, I observed that the notion of forking has changed since the invention of fork-based development, so I conducted mixed-method experiment to understand the perception of forking by interviewing developers and identified future research directions. Second, I designed tooling interventions to improve the awareness in the fork-based development environment and help developers to detect redundant development to reduce developers’ unnecessary effort. To evaluate the effectiveness and usefulness of these awareness tools, I conducted both quantitative and qualitative studies.

Committee: Dr. Christian Kästner (Chair), Dr. James Herbsleb, Dr. Laura Dabbish, Dr. Andrzej Wąsowski (IT University of Copenhagen)