Title: The Impact of Design and UML Modeling on Codebase Quality and Sustainability

Author[1]: Omar Badreddin, University of Texas at El Paso, Email: obbadreddin@utep.edu
Author[2]: Khandoker Rahad, University of Texas at El Paso, Email: karahad@miners.utep.edu
Presenter: Khandoker Rahad

Abstract: The general consensus of researchers and practitioners is that up- front and continuous software design using modeling languages such as UML improve code quality and reliability particularly as the software evolves over time. Software designs and models help in managing the underlying code complexities which are crucial for sustainability. Recently, there has been increasing evidence suggesting broader adoption of modeling languages such as UML. However, our understanding of the impact of using such modeling and design languages remains limited. This paper reports on a study that aims to characterize this impact on code quality and sustainability. We identify a sample of open source software repositories with extensive use of designs and modeling and compare their code qualities with similar code-centric repositories. Our evaluation focuses on various code quality attributes such as code smells and technical debt. We also conduct code evolution analysis over five-year period and collect additional data from questionnaires and interviews with active repository contributors. This study finds that repositories with significant use of models and design activities are associated with reduced critical code smells but are also associated with increase in non-critical code smells. The study also finds that modeling and design activities are associated with significant reduction in measures of technical debt. Analyzing code evolution over five-year period reveals that UML repositories start with significantly lower technical debt density measures but tend to decline over time.