Process mining is widely used to discover, analyze, and improve business processes based on event data extracted from IT systems, stored in so-called event logs. A key task in this regard is process discovery, which aims to reconstruct how a process was truly executed. To do so, process discovery strives to establish an accurate process model on the basis of the recorded behavior captured in an event log. Using such event logs as basis for discovery has an important limitation, however: It limits the scope of analysis to back-end events, i.e., secondary, indirect events that were triggered by the actual user activity. User activities that do not result in such back-end events or take place in productivity applications such as Excel and Outlook, are thus not recorded in event logs and, therefore, invisible to traditional process mining and discovery techniques.

To avoid this problem and be able to elicit a comprehensive view on business processes, the goal of this proposal is to enable process discovery based on user interaction (UI) logs, rather than on traditional event logs. In essence, a UI log is a collection of recorded interactions performed on GUI components, such as clicks on buttons or keyboard entries in text areas. The benefit of using UI logs is that they can be obtained for any business process of which the activities are performed on a computer, regardless of the specific applications required for it. Available logging software is then able to extract and store relevant data such as the interaction type (e.g. click or keyboard stroke), the time, and context (e.g., the GUI element and URL) in a UI log.

However, eliciting process information from such UI logs is a complex task, for which various problems need to be overcome. Specifically, we need to address two problem areas, each with its own specific challenges: 1) data transformation and 2) process representation. The problem area of data transformation is concerned with turning the input UI log into an event log that carries proper event labels, is free of noise, and has a proper case identifier. The problem area of process representation is concerned with deriving a useful representation from the low-level events from the logs. This entails grouping events into higher-level events, generating proper labels for these higher-level events, and providing the user with an effective process representation that balances between high and low-level information.

The proposed project will address these challenges by combining behavioral process analysis with a novel semantic angle. The proposed project will result in the development of approaches that address the aforementioned challenges in an automated manner, ultimately covering the entire pipeline from UI log to an informative process representation. In this way, the successful project will considerably advance state-of-the-art research in process mining, particularly for situations involving raw, low-level event data.