Background
Data quality aspects, such as timeliness, ordering and completeness, are well-researched in the context of traditional database systems and data at rest. For streaming data, which is typically unbound, deriving these qualities is critical for applications building on streaming data.

Problem
Approaches to quantify data quality aspects depend on different characteristics of the data streams themselves (intermittent/event-based streams vs. continuous streams) and the applications built on top, which typically include stream processing, relying on different windowing techniques (sliding/tumbling windows). Defining and quantifying appropriate metrics to represent these aspects is challenging and implementations are lacking.

Goals
Your goal is to systematically research data quality and quantitative metrics of these in scientific literature, in particular in the streaming data context. Based on personal interest, a subset of data qualities is to be chosen and a mechanism to capture relevant metrics is to be prototypically implemented. The implementation can be done either for a messaging middleware, a stream processing framework or an RPC framework with an explicit notion of streams.

Requirements:
• Interest in data quality aspects
• No fear of extensive literature research and confusing terminology
• Programming skills in a modern programming language
• Desirable: first experience with any stream processing or RPC framework

Contact: Dominik Ernst
de@ise.tu-berlin.de