Quality-Informed Semi-Automated Event Log Generation for Process Mining

R. Andrews[a,*], C.G.J. van Dun[c,*], M.T. Wynn[a], W. Kratsch[b], M.K.E. Röglinger[c], A.H.M. ter Hofstede[a]

[a] *Queensland University of Technology, Brisbane, Australia*
[b] *FIM Research Center, University of Augsburg, Augsburg, Germany*
[c] *FIM Research Center, University of Bayreuth, Bayreuth, Germany*

Abstract

Process mining, as any form of data analysis, relies heavily on the quality of input data to generate accurate and reliable results. A fit-for-purpose event log nearly always requires time-consuming, manual pre-processing to extract events from source data, with data quality dependent on the analyst's domain knowledge and skills. Despite much being written about data quality in general, a generalisable framework for analysing event data quality issues when extracting logs for process mining remains unrealised. Following the DSR paradigm, we present RDB2Log, a quality-aware, semi-automated approach for extracting event logs from relational data. We validated RDB2Log's design against design objectives extracted from literature and competing artifacts, evaluated its design and performance with process mining experts, implemented a prototype with a defined set of quality metrics, and applied it in laboratory settings and in a real-world case study. The evaluation shows that RDB2Log is understandable, of relevance in current research, and supports process mining in practice.

Keywords: process mining, data quality, event log, log extraction

* Corresponding author *Email addresses:* r.andrews@qut.edu.au (R. Andrews), christopher.vandun@fim-rc.de (C.G.J. van Dun), m.wynn@qut.edu.au (M.T. Wynn), wolfgang.kratsch@fim-rc.de (W. Kratsch), maximilian.roeglinger@fim-rc.de (M.K.E. Röglinger), a.terhofstede@qut.edu.au (A.H.M. ter Hofstede)