

## **Intellibot Data Cleaner**

**Jerry Odero and Vincent Omwenga**

**Faculty of Information Technology, Strathmore University, Nairobi, Kenya.**

Data cleaning is an activity involving detecting and correcting errors and inconsistencies in a database, data warehouse or any data record of an organization. Kenya Revenue Authority (KRA) in its quest to be a fully data driven organization, is actively undertaking the data cleaning process. However, this process is currently manual and slow as it involves physical transfer of documents to be processed from the various stations, via different levels of management for approval, to the centralized return processing unit. A process, which might take at least a fortnight for the processing of one taxpayers ledger account. Furthermore, this whole process needs lots of man-hours, since there is a vast amount of data to be cleaned due to the many ledger accounts affected during the manual filing system that ended in 2014. There exists many data cleaning processes and approaches which are used to purge out dirty data, before it's loaded into the data warehouse. These processes vary depending on the data source, they are time consuming and expensive for organizations, in terms of skilled staff and the tools involved, hence this research proposed the application of RPA (Robotic Process Automation) to develop an intelligent bot (Intellibot) to be used in the transactional data cleaning exercise in Kenya Revenue Authority (KM). With the transition from legacy system to I-Tax and I-CMS systems for domestic and customs revenue management respectively, the researcher sought to find out the current data cleaning process in the legacy system. This research led to the development of an RPA system for the current manual data cleaning process implemented and tested using the Blue Prism platform. The system detected the errors using a knowledge-based model-, clustering them as errors due to uncaptured returns, uncaptured losses or credit re-adjustments. The intellibot system was able to load the ledgers, detect the errors and clean them with utmost precision. Experiments conducted on performance of the bots varied by seconds, in the first experiment. Also in the second performance test, there was a variance of seconds in cleaning the different errors detected, hence improving the data integrity significantly: free of errors, to be migrated to the I-Tax platform, thus support better decision making process in the organization, and a higher return on investments.

**Keywords:** RPA; Data Cleaning; Knowledge-based System; KRA; Blue Prism; Legacy Systems; Intellibot.