

A Data Quality in Use model for Big Data

Merino Jorge, Caballero Ismael, Rivas Bibiano, Serrano Manuel, Piattini Mario

*Alarcos Research Group
Escuela Superior de Informática. Universidad de Castilla-La Mancha
Paseo de la Universidad 4, 13071, Ciudad Real*

Abstract

Beyond the hype of Big Data, something within business intelligence projects is indeed changing. This is mainly because Big Data is not only about data, but also about a complete conceptual and technological stack including raw and processed data, storage, ways of managing data, processing and analytics. A challenge that becomes even trickier is the management of the quality of the data in Big Data environments. More than ever before the need for assessing the Quality-in-Use gains importance since the real contribution — business value — of data can be only estimated in its context of use. Although there exists different Data Quality models for assessing the quality of regular data, none of them has been adapted to Big Data. To fill this gap, we propose the “3As Data Quality-in-Use model”, which is composed of three Data Quality characteristics for assessing the levels of Data Quality-in-Use in Big Data projects: Contextual Adequacy, Operational Adequacy and Temporal Adequacy. The model can be integrated into any sort of Big Data project, as it is independent of any pre-conditions or technologies. The paper shows the way to use the model with a working example. The model accomplishes every challenge related to Data Quality program aimed for Big Data. The main conclusion is that the model can be used as an appropriate way to obtain the Quality-in-Use levels of the input

Email address: {jorge.merino, ismael.caballero, bibiano.rivas, manuel.serrano, mario.piattini}@uclm.es (Merino Jorge, Caballero Ismael, Rivas Bibiano, Serrano Manuel, Piattini Mario)

URL: <http://alarcos.esi.uclm.es/> (Merino Jorge, Caballero Ismael, Rivas Bibiano, Serrano Manuel, Piattini Mario)