

Medición, evaluación y certificación de calidad de datos en la transformación digital

Fernando Gualo
Data Quality Team SL (DQTeam)
Instituto de Tecnologías y Sistemas de Información, Camino de Moledores SN, Ciudad Real, 13005

En la transformación digital y en organizaciones data centric, el dato toma un valor diferencial para el desempeño de la organización independientemente del uso (BI, IA, etc.) o infraestructura (Big data, Data Lake, etc.) que se trate. Es por ello por lo que asegurar niveles de calidad adecuados es básico para obtener éxito en las operaciones basadas en datos.

En la actualidad es posible medir y establecer un nivel de calidad de datos de los repositorios organizacionales basado en el contexto de uso o fin que se le vaya a dar a los datos. Para facilitar esta labor, ISO/IEC 25012 define un modelo general de calidad de datos con un conjunto de características que representan distintas dimensiones del dato. Estas características de calidad de datos se pueden clasificar en tres grupos: inherentes, dependientes del sistema o mixtas. Del mismo modo ISO/IEC 25024 proporciona un conjunto de medidas de calidad de datos que se pueden utilizar para medir y evaluar la calidad de los datos en términos de las características definidas en ISO/IEC 25012. Cada una de las propiedades define una o varias medidas, y se mide aplicando un método de medición. Del mismo modo, se dispone de una función de medición que permite computar las medidas para la obtención de una medida de calidad. A partir de la combinación de las distintas medidas de calidad es posible establecer el valor de la característica de calidad de datos.

Utilizando este modelo de calidad es posible realizar evaluaciones de calidad de datos y valorar cuantitativamente la calidad de los datos de un repositorio. Para llevar a cabo de manera correcta y adecuada una evaluación de calidad de datos es necesario conocer el contexto de uso de los datos puesto que este definirá los requisitos de datos que se han de cumplir. Este conjunto de requisitos de datos representa las restricciones técnicas y de negocio que han de cumplir los datos para aportar valor al fin que se les vaya a dar.

En la actualidad, es posible evaluar 5 de las características de calidad de datos existentes en ISO/IEC 25012. Para demostrar la aplicabilidad del modelo, se han llevado a cabo evaluaciones de calidad de datos de multitud de repositorios de datos organizacionales de distinta naturaleza y con distinto contexto de uso. Además, para aportar un valor añadido a la evaluación de calidad de datos, desde una entidad de certificación autorizada como AENOR es posible conseguir una certificación ISO/IEC 25012 sobre el nivel de calidad de los datos.