

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

Fernando Gualo, Ismael Caballero, Mario Piattini, Javier Verdugo, Moisés Rodríguez

Ávila, 28 Septiembre 2022



Contenidos

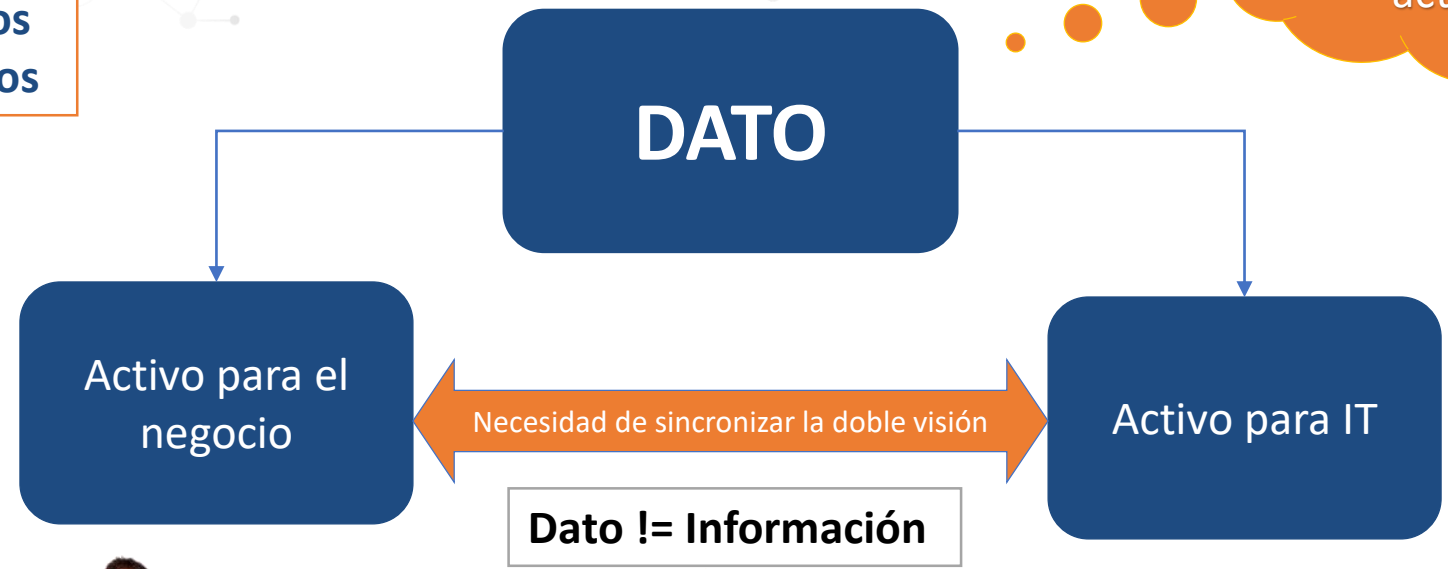
1. **Introducción**
2. Desarrollo/Descripción
3. Resultados y discusión
4. Conclusiones

1. Introducción

1.1. Doble visión del dato como activo organizacional (1/2)

Doble visión de los datos como activos

Idea de "valor" de los datos como activo



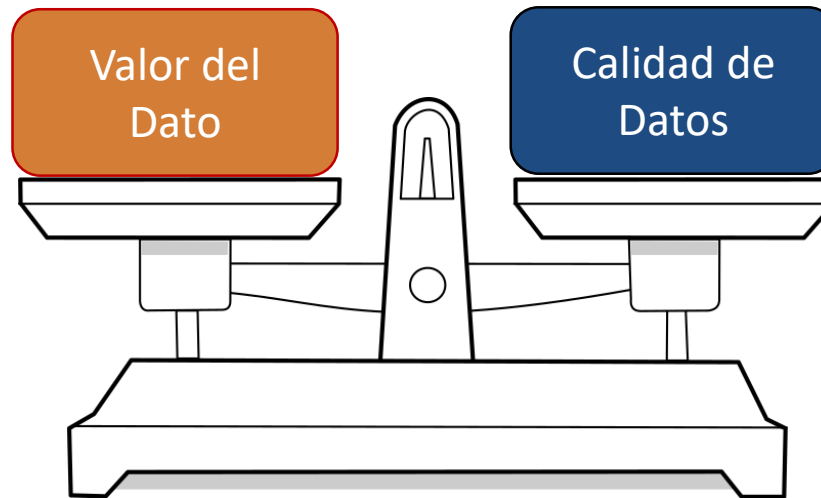
Dato != Información

Necesidad de datos de calidad para alinear ambas visiones



1. Introducción

1.1. Doble visión del dato como activo organizacional (1/2)



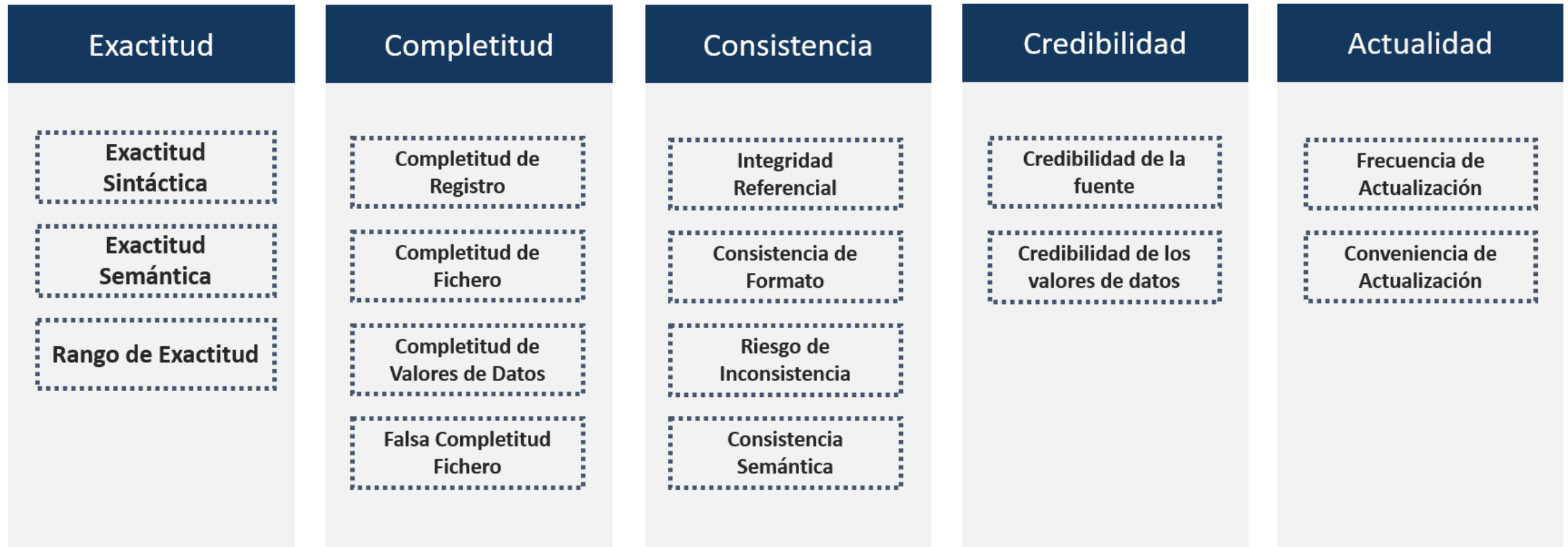
Necesidad de entornos de Gestión de Calidad de Datos

Contenidos

1. Introducción
- 2. Desarrollo/Descripción**
3. Resultados y discusión
4. Conclusiones

2. Desarrollo/ Descripción

2.1. Modelo de Calidad de Datos basado en ISO/IEC 25012 (1/4)



2. Desarrollo/ Descripción

2.1. Modelo de Calidad de Datos basado en ISO/IEC 25012 (2/4)

Característica de calidad de datos	Compleitud
Propiedad de la calidad de datos	Compleitud de Registro (COMP_REG)
Perspectiva	Inherente
Descripción de la medida	Compleitud de los registros previstos en un fichero de datos
Fórmula de cálculo	$X=A/B$ A = Número de atributos con valor no nulo dentro de un registro B = Número de atributos totales del registro
Escala	Ratio
Rango de valores	[0.0, 1.0]

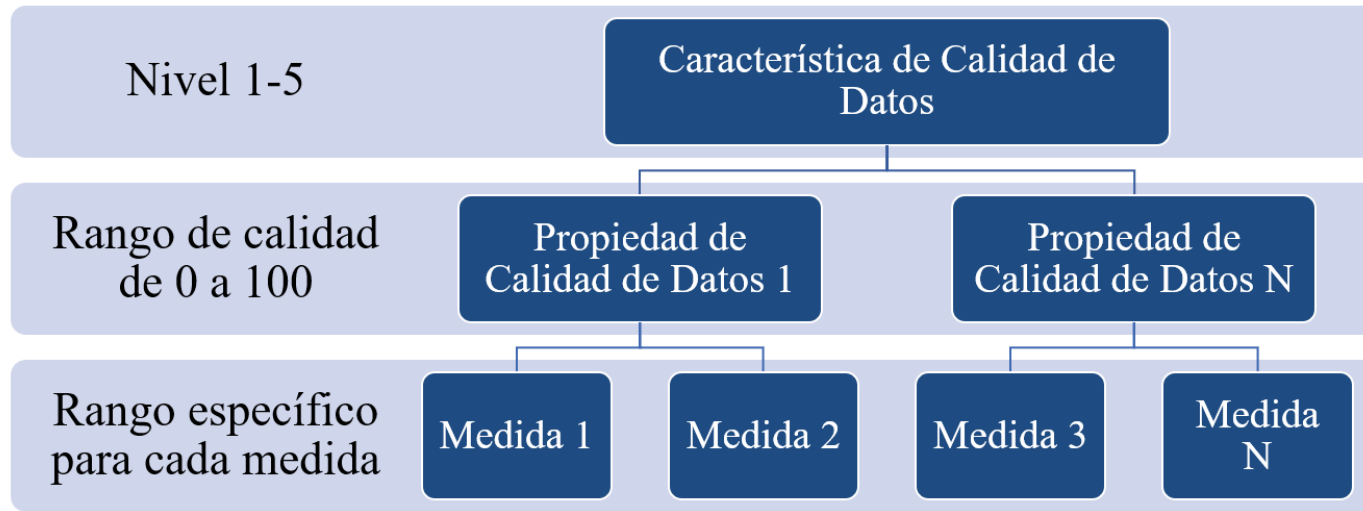
2. Desarrollo/ Descripción

2.1. Modelo de Calidad de Datos basado en ISO/IEC 25012 (3/4)

Valor de medición de la propiedad	Nivel	Descripción
0-25	1	Calidad deficiente
25-50	2	Calidad insuficiente
50-75	3	Buena Calidad
75-95	4	Muy buena calidad
95-100	5	Calidad excelente

2. Desarrollo/ Descripción

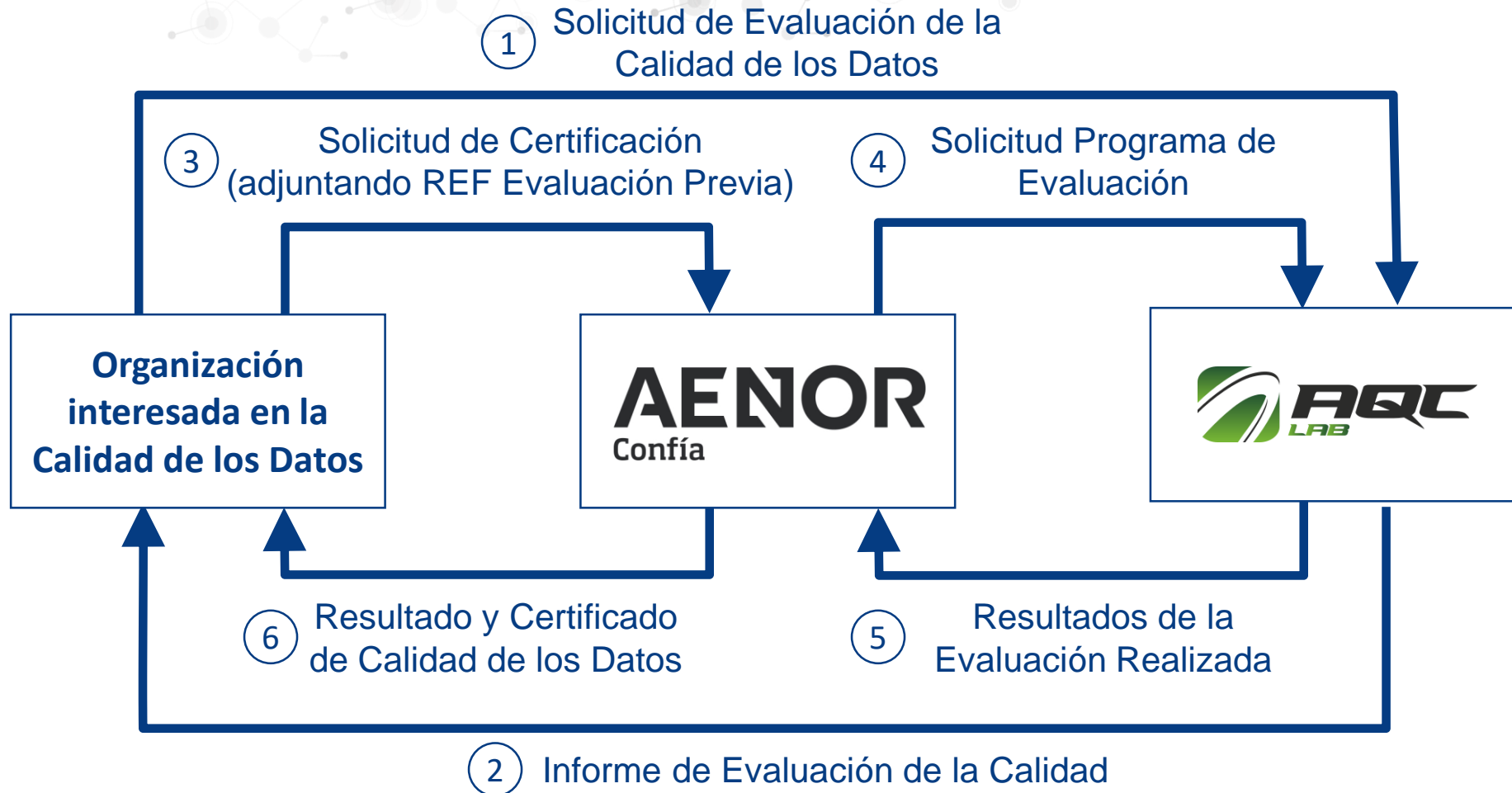
2.1. Modelo de Calidad de Datos basado en ISO/IEC 25012 (4/4)



		Niveles				Valor de Característica de calidad
		1	2	3	4	
Rango	0	-	-	-	-	0
	1	3	3	3	3	1
	2	2	3	3	3	2
	3	0	1	2	3	3
	4	0	0	0	3	4
	5	0	0	0	0	5

2. Desarrollo/ Descripción

2.2. Entorno de Certificación de Calidad de Datos



Contenidos

1. Introducción
2. Desarrollo/Descripción
3. **Resultados y discusión**
4. Conclusiones

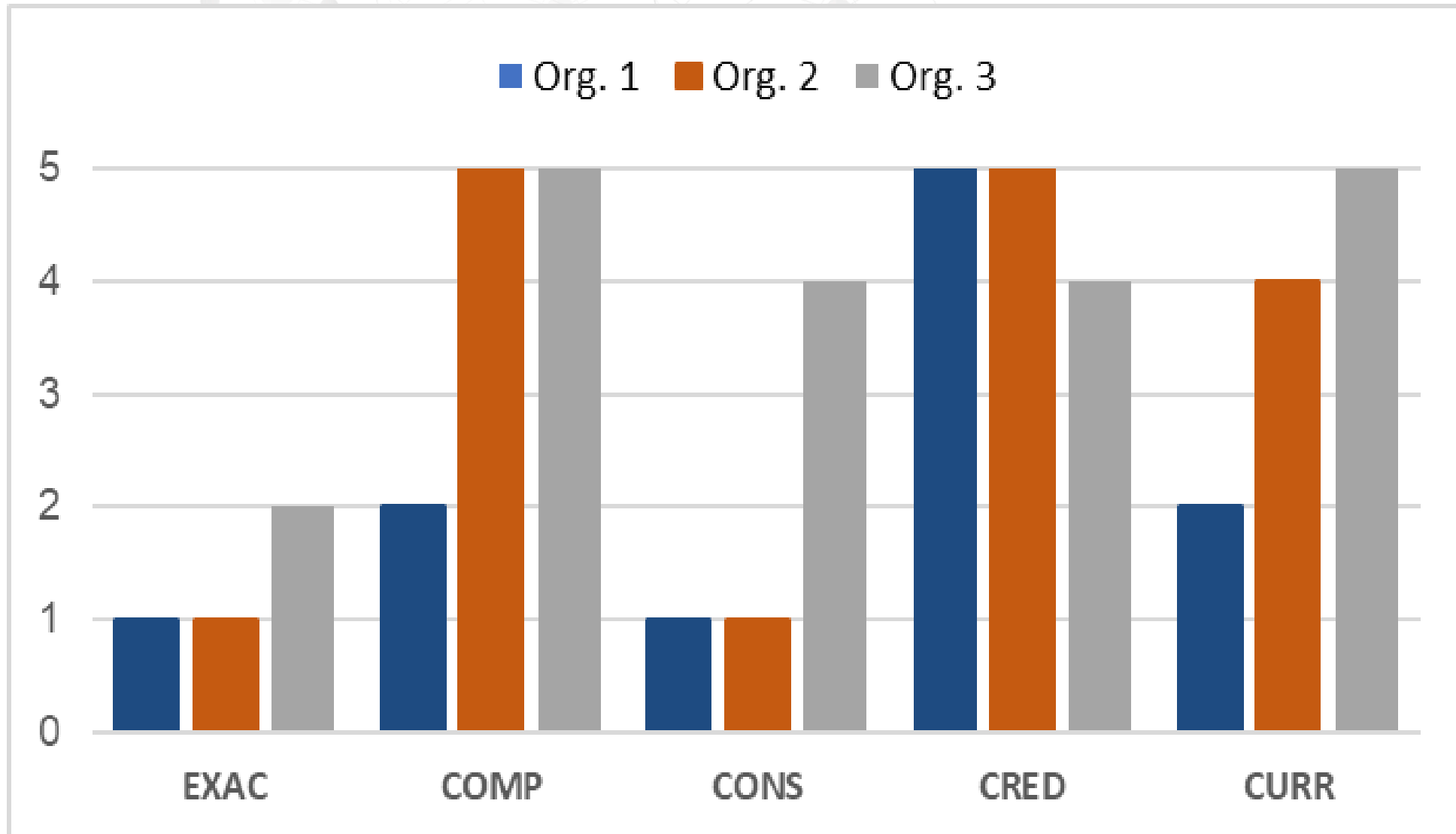
3. Resultados y discusión

3.1. Experiencias de aplicación

		Organización 1	Organización 2	Organización 3
Base de datos		Oracle 18c	Teradata 15.0	SQLServer 2017
Número de entidades de datos		46	30	24
Volumen de datos		750 millones	200 millones	100 millones
Características evaluadas		Exactitud, Completitud, Consistencia, Credibilidad, Actualidad		
Reglas por característica	Exactitud	189	94	89
	Completitud	131	100	78
	Consistencia	340	176	91
	Credibilidad	72	48	54
	Actualidad	81	70	63
Total de reglas de negocio		813	488	375

3. Resultados y discusión

3.2. Resultados primera evaluación



3. Resultados y discusión

3.3. Fortalezas y debilidades encontradas

Característica	Fortalezas	Debilidades
Exactitud	-	Los valores de los datos deben coincidir con ciertas expresiones regulares definidas y deben estar definidos dentro de los rangos de aceptabilidad del dominio de valores definido para los atributos.
Completitud	Los archivos de datos contienen un número adecuado de registros para representar las entidades requeridas, así como los valores necesarios para las diferentes entidades representadas por los datos.	Es necesario revisar las definiciones de datos completos, puesto que hay muchos casos en que no se dispone de valores para todos los atributos que son necesarios.
Consistencia	Las referencias a valores se hacen de manera correcta	Hay casos en los que no se cumple con determinados formatos para datos de similar naturaleza o tipo
Credibilidad	La credibilidad normalmente de los datos suele ser alta debido a que se llevan a cabo procesos de integración de datos	Es necesario revisar el nivel de calidad de las fuentes que nutren los repositorios de la organización, debido a que al estar fuera la calidad puede no ser la correcta
Actualidad	Los datos por norma general se actualizan en los tiempos debidos para disponer siempre de una versión actual	Es necesario revisar eventos de actualización de datos, así como los diferentes caminos.

3. Resultados y discusión

3.4. Discusión (1/2)

- Las organizaciones no están familiarizadas con la importancia de los requisitos de datos y hay que darles soporte en la identificación
- Las organizaciones a lo largo de los años y cambios en sus sistemas de información han ido documentando algunos cambios realizados
- Las organizaciones son más conscientes de la integridad de datos y actualidad de datos, por lo tanto, los resultados de la evaluación para estas características y sus propiedades suelen ser más altos que para las demás.
- Algunos sistemas de información de las organizaciones no son lo suficientemente potentes para soportar el estrés de consultas en periodos cortos de tiempo, y ha sido necesario virtualizar o sacar de sistemas productivos algunas partes de repositorios.

3. Resultados y discusión

3.4. Discusión (2/2)

- Existen problemas de calidad de datos similares tales como: valores de los datos que no cumplen con las reglas de sintaxis específicas, o existencia de registros cuyos valores para algunos atributos no pertenecen a los rangos esperados o permitidos.
- Las organizaciones tienden a definir valores por defecto no eficientes en sus repositorios de datos.
- Los jefes de TI e informática de las organizaciones evaluadas reportan que la identificación de problemas de calidad de datos les ha sido útil para definir planes de mitigación y extrapolar controles a todo el sistema de información, no solo al repositorio.
- A partir de los resultados de la evaluación ha sido posible definir mecanismos de control y aseguramiento de calidad que habilita a que la calidad de los datos nuevos integrados en sus sistemas de información se mantenga, y no se reduzca.

Contenidos

1. Introducción
2. Desarrollo/Descripción
3. Resultados y discusión
4. **Conclusiones**

5. Conclusiones

5.1. Conclusiones (1/2)

1

Los datos son un **elemento clave** en la economía de los datos.

2

Los **costes de la falta de calidad de los datos** pueden acabar siendo elevados, pero debe hacerse un estudio entre el **coste de la mejora** de los datos y los **beneficios reales** de su limpieza.

3

La **calidad de los datos** va más allá del perfilado/limpieza de datos y, por supuesto, de cualquier herramienta: la **potencia** (por ejemplo, las herramientas) **sin control** (por ejemplo, las metodologías y los modelos) **no produce resultados óptimos**.

5. Conclusiones

5.1. Conclusiones (2/2)

4

La evaluación y certificación de la calidad de los datos ayuda a **garantizar la sostenibilidad de la organización** a largo plazo.

5

La evaluación y certificación de la calidad de los datos ayuda a **conocer mejor los aspectos internos** (de gestión y tecnológicos) de la empresa y las formas de trabajo de la organización.

6

El **conocimiento adquirido** a través del proceso de evaluación y certificación de la calidad de los datos ayuda a **respaldar** mejor las **futuras iniciativas de gestión de la calidad de los datos** de la organización.

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

Fernando Gualo, Ismael Caballero, Mario Piattini, Javier Verdugo, Moisés Rodríguez

Ávila, 28 Septiembre 2022



Medición y evaluación de calidad de datos

Obtención del perfil de calidad para las propiedades de calidad de datos

Propiedad de Calidad de Datos	Resultado medición
Integridad referencial	100
Riesgo Inconsistencia Datos Maestros	100
Consistencia semántica de Datos Maestros	84
Consistencia de formato de Datos Maestros	95

Nivel	Valor de Calidad
1	[0, 20]
2	(20, 40]
3	(40, 70]
4	(70, 85]
5	(85, 100]

(Mapeo de los valores de nivel de calidad por propiedad a un vector de perfil de calidad)

3 Matriz de rangos para el valor de calidad

Rango	Niveles				Valor de Característica de calidad
	1	2	3	4	
0	-	-	-	-	0
1	4	4	4	4	1
2	3	4	4	4	2
3	0	2	3	4	3
4	0	0	0	4	4
5	0	0	0	0	5

5 Función por perfiles

[0,0,0,1,3]

(Función por perfiles para obtener el nivel de calidad)

Característica de Calidad de Datos	Nivel de Calidad
Consistencia	4

(Nivel de calidad de la característica)

4 Vector perfil de calidad

Obtención del nivel de calidad de la característica de calidad de datos a partir de función por perfil

(Resultados de medición para propiedad de calidad de datos)

$$CONS_FORM_DM = \frac{\text{NúmeroElementosFormatoConsistenteVariosFicheros}}{\text{NúmeroElementosDatosFormatoDefinido}}$$

2 Reglas de negocio relacionadas con consistencia de formato de datos maestros

1 Método de medición para cada una de las propiedades de calidad de datos

Beneficios de la evaluación de DQ

- ① Conocer el valor de los datos para tomar mejores decisiones
- ② Obtener reconocimiento frente a la competencia
- ③ Aumentar la satisfacción y confianza de los clientes
- ④ Utilizar datos de calidad en la toma de decisiones con el uso de BI e IA
- ⑤ Detectar y mitigar los riesgos derivados de los datos en toda la organización
- ⑥ Gestionar los riesgos relacionados con los datos y asegurar cumplimiento

Entorno de Evaluación y Certificación de Calidad de Datos

