



# **UNIVERSIDAD TECNOLÓGICA DE SAN LUIS RIO COLORADO**

**Investigación calidad en los datos**

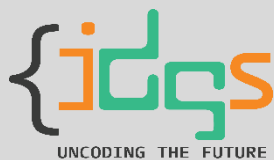
**MTRA. JULIA ELIZABETH GARCIA HERRERA**

**ALUMNO: VICTOR MANUEL GALVAN COVARRUBIAS**

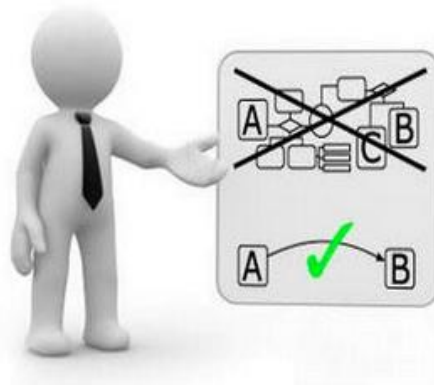
**ING. EN DESARROLLO Y GESTIÓN DE SOFTWARE**

San Luis Río Colorado, Sonora

Julio, 2022



## Calidad en los datos



### Definición

Calidad de datos es la cualidad de un conjunto de información en una base de datos, un sistema de información que reúne entre sus atributos la **exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad y confiabilidad** necesarias para resultar útiles al procesamiento, análisis y cualquier otro fin que un usuario quiera darles.

### Beneficios

1. Ahorro de tiempo y recursos.
2. Mejor capacidad de toma de decisiones y elaboración de estrategias más oportunas y eficientes basadas en el análisis de datos de calidad, confiables y validados.
3. Los riesgos en los proyectos se reducen.
4. Mejora la confianza de los clientes y la imagen de la empresa.
5. Ayuda a optimizar la captación de clientes y su fidelización.
6. Personalización de la oferta y el trato con el cliente.
7. Mayor rentabilidad de las campañas.
8. Conocer las expectativas y necesidades de clientes con más precisión.

### Problemas

1. **Gran variedad de formatos y fuentes** de las que proviene la información, lo que dificulta el establecimiento de una norma o métrica estandarizada que sirva para la normalización de la información.

2. **La velocidad a la que se genera la información** es muy grande, lo que ocasiona problemas para su medición y gestión de manera fiable.
3. **El volumen de información y datos** que se generan a diario es inmenso, lo que puede provocar falta de precisión en los análisis y complejidad en los procesos.

### Retos

La diversidad de fuentes de datos aporta abundantes tipos de datos y estructuras de datos complejas y aumenta la dificultad de la integración de datos.

El volumen de datos es tremendo, y es difícil juzgar la calidad de los datos dentro de un tiempo razonable.

Los datos cambian muy rápido y la "puntualidad" de los datos es muy corta, lo que requiere mayores requisitos para la tecnología de procesamiento.

No hay demasiados estándares de calidad de datos unificados y aprobados y la investigación sobre la calidad de datos de big data.

### Estándares jerárquicos

#### 1. Disponibilidad:

##### 1.1. Accesibilidad:

1.1.1. Si se proporciona una interfaz de acceso a datos

1.1.2. Los datos pueden hacerse fácilmente públicos o fáciles de adquirir

##### 1.2. Oportunidad:

1.2.1. Dentro de un tiempo dado, si los datos llegan a tiempo

1.2.2. Si los datos se actualizan regularmente

1.2.3. Si el intervalo de tiempo entre la recopilación y el procesamiento de los datos hasta la liberación cumple los requisitos

#### 2. Usabilidad:

##### 2.1. Credibilidad:

2.1.1. Los datos provienen de organizaciones especializadas de un país, campo o industria

2.1.2. Expertos o especialistas auditan regularmente y comprueban la exactitud del contenido de los datos

2.1.3. Los datos existen en el rango de valores conocidos o aceptables

### 3. Confiabilidad:

#### 3.1. Exactitud

- 3.1.1. Los datos proporcionados son precisos
- 3.1.2. La representación de datos (o valor) refleja bien el estado real de la información de origen
- 3.1.3. La representación de información (datos) no causará ambigüedad

#### 3.2. Consistencia:

- 3.2.1. Después de procesar los datos, sus conceptos, dominios de valor y formatos todavía coinciden como antes de procesar
- 3.2.2. Durante un cierto tiempo, los datos permanecen consistentes y verificables
- 3.2.3. Todos los datos son consistentes o verificables

#### 3.3. Integridad:

- 3.3.1. El formato de los datos es claro y cumple los criterios
- 3.3.2. Los datos son consistentes con la integridad estructural
- 3.3.3. Los datos son consistentes con la integridad del contenido

#### 3.4. Completitud:

- 3.4.1. Si una deficiencia de un componente afectará el uso de los datos para datos con componentes múltiples
- 3.4.2. Si una deficiencia de un componente afectará la precisión y la integridad de los datos

### 4. Pertinencia:

#### 4.1. Conveniencia:

- 4.1.1. Los datos recogidos no coinciden completamente con el tema, pero exponen un aspecto
- 4.1.2. La mayoría de los conjuntos de datos recuperados están dentro del tema de recuperación que los usuarios necesitan
- 4.1.3. El tema de la información proporciona coincidencias con el tema de recuperación de los usuarios

### 5. Calidad de presentación:

#### 5.1. Legibilidad:

- 5.1.1. Los datos (contenido, formato, etc.) son claros y comprensibles
- 5.1.2. Es fácil juzgar que los datos facilitados satisfacen las necesidades
- 5.1.3. La descripción de los datos, la clasificación y el contenido de codificación satisfacen la especificación y son fáciles de entender

**David Loshin**



Presidente de Knowledge Integrity, Inc, es un líder intelectual reconocido y consultor experto en las áreas de calidad de datos, gestión de datos maestros e inteligencia comercial. Su libro, "Business Intelligence: The Savvy Manager's Guide" (junio de 2003) ha sido aclamado como un recurso que permite a los lectores "comprender la inteligencia empresarial, las disciplinas de gestión empresarial, el almacenamiento de datos y cómo todas las piezas funcionan juntas". Su libro, "Gestión de datos maestros", ha sido respaldado por líderes de la industria de gestión de datos.

El establece que la gestión de la calidad de los datos empieza con la identificación y la medición del efecto de los resultados empresariales. Se definen las reglas, se fijan los objetivos de rendimiento y se implementan los métodos de mejora de la calidad, así como la limpieza específica de datos y se implementan procesos de mejora.