



EXPERIENCIA EN CALIDAD DE DATOS

Diego Rosselli

Objetivos

- Introducción Calidad Datos

1

- Experiencia en Trazabilidad trámites

2

- Lecciones aprendidas

3

Objetivos

- **Introducción Calidad Datos**

1

- Experiencia en Trazabilidad trámites

2

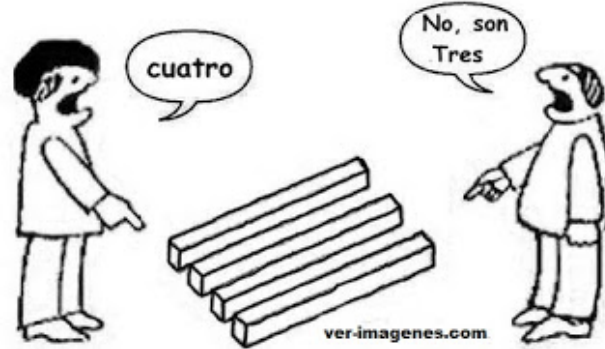
- Lecciones aprendidas

3

Introducción

1

¿ Son 3 o son 4 ?



Calidad:

- La percibimos
- La definimos
- La medimos



Calidad - Valor - Expectativas

1

Puede ser correcto, preciso, actualizado y **no colaborar** con las **necesidades** del usuario.



Puede ser correcto, preciso, actualizado y **no alcanzar** o **por el contrario superar** las **expectativas** del usuario.

Adecuación al uso

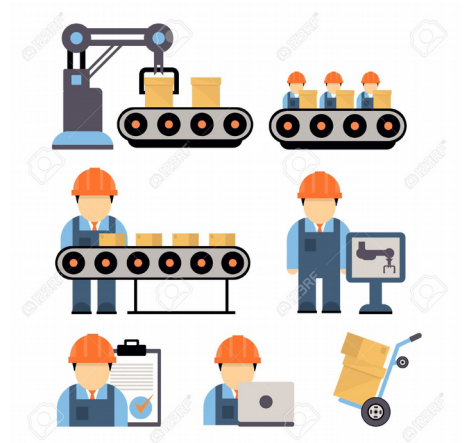
- 3 mt x 1,50 mt aproximadamente
- 2,98 mt x 1,37 mt
- desarmable, pieza más grande rectangular 1,10 mt x 1,00 mt



Causas de problemas de calidad

1

- Producción: ingreso manual, sensores, unificar diferentes fuentes



Causas de problemas de calidad

1

- Almacenamiento: ausencia de formato común, diseño inadecuado de base de datos



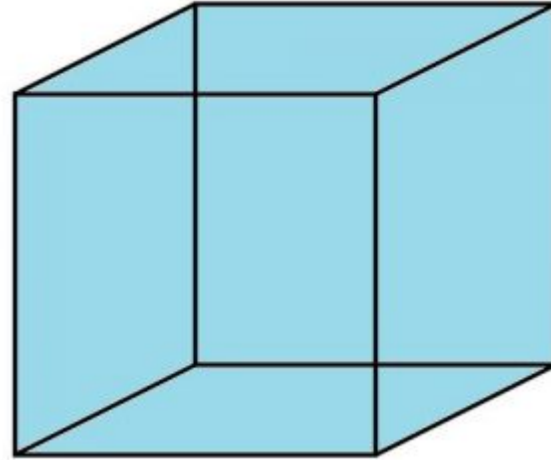
Causas de problemas de calidad

- Utilización: actualización, cambios de requerimientos, errores de interpretación



Multi-dimensión de la calidad

- Completitud
- Unicidad
- Exactitud
- Consistencia
- Actualidad
- Otras



Dimensión Unicidad

1

| CI | Nombre | Calle | Nro. | F.Nto. | Edad |
|----|--------|---------|------|--------|------|
| 1 | José | Calle A | 1 | A | 30 |
| 11 | José | Calle A | 1 | A | 30 |
| 1 | José | Calle A | 1 | A | 30 |
| 3 | María | Calle B | 2 | B | 28 |
| 4 | Juana | Calle C | 3 | C | 29 |

Dimensión Completitud

1

| CI | Nombre | Calle | Nro. | F.Nto. | Edad |
|----|--------|---------|------|--------|------|
| 1 | José | Calle A | 1 | A | 30 |
| 2 | Pepe | | | C | |
| 4 | Juana | Calle B | 2 | | 20 |
| 3 | María | Calle C | | D | 40 |

Falta

Inferir

NC

Estimar

Unicidad: Medición y Registro

1

| CI | Nombre | Calle | Calidad de cada línea | | | |
|----|--------|---------|-----------------------|---|----|---|
| 1 | José | Calle A | 1 | A | 30 | 0 |
| 11 | José | Calle A | 1 | A | 30 | 0 |
| 1 | José | Calle A | 1 | A | 30 | 0 |
| 3 | María | Calle B | 2 | B | 28 | 1 |
| 4 | Juana | Calle C | 3 | C | 29 | 1 |

Unicidad: Medición y Registro

1

| CI | Nombre | Calle | Calidad de cada línea | | | |
|----|--------|---------|-----------------------|---|----|---|
| 1 | José | Calle A | 1 | A | 30 | 3 |
| 11 | José | Calle A | 1 | A | 30 | 3 |
| 1 | José | Calle A | 1 | A | 30 | 3 |
| 3 | María | Calle B | 2 | B | 28 | 1 |
| 4 | Juana | Calle C | 3 | C | 29 | 1 |

Complejidad: Medición y Registro

| CI | Nombre | Calle | Nro. | F.Nto. | Edad |
|----|--------|---------|------|--------|------|
| 1 | José | Calle A | 1 | A | 30 |
| 2 | Pepe | | | C | |
| 4 | Juana | Calle B | 2 | | 20 |
| 3 | María | Calle C | | D | 40 |

0

1

Calidad de cada celda

Dimensión Completitud

Falta

Inferir

1

| CI | Nombre | Calle | Nro. | F.Nto. | Edad |
|----|--------|---------|------|--------|------|
| 1 | José | Calle A | 1 | A | 30 |
| 2 | Pepe | | | C | |
| 4 | Juana | Calle B | 2 | | 20 |
| 3 | María | Calle C | | D | 40 |

NC

Estimar

Objetivos

- Introducción Calidad Datos

1



- **Experiencia en Trazabilidad trámites**

2

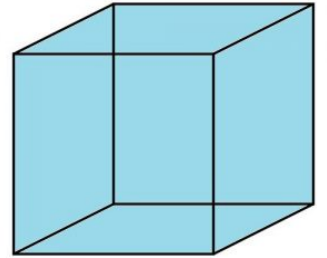
- Lecciones aprendidas

3

Estrategia

- BD para resultados 
- Mediciones con Pentaho PDI 
 - Simplifica implementación
 - Permite automatizar ejecución
 - Libre uso

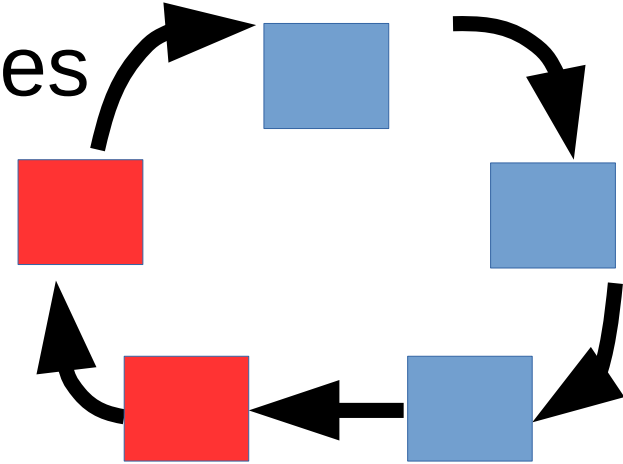
- Selección de Dimensiones relevantes:
 - Completitud
 - Unicidad
 - Exactitud: Correctitud sintáctica y semántica
 - Consistencia: integridad entre entidades



Estrategia

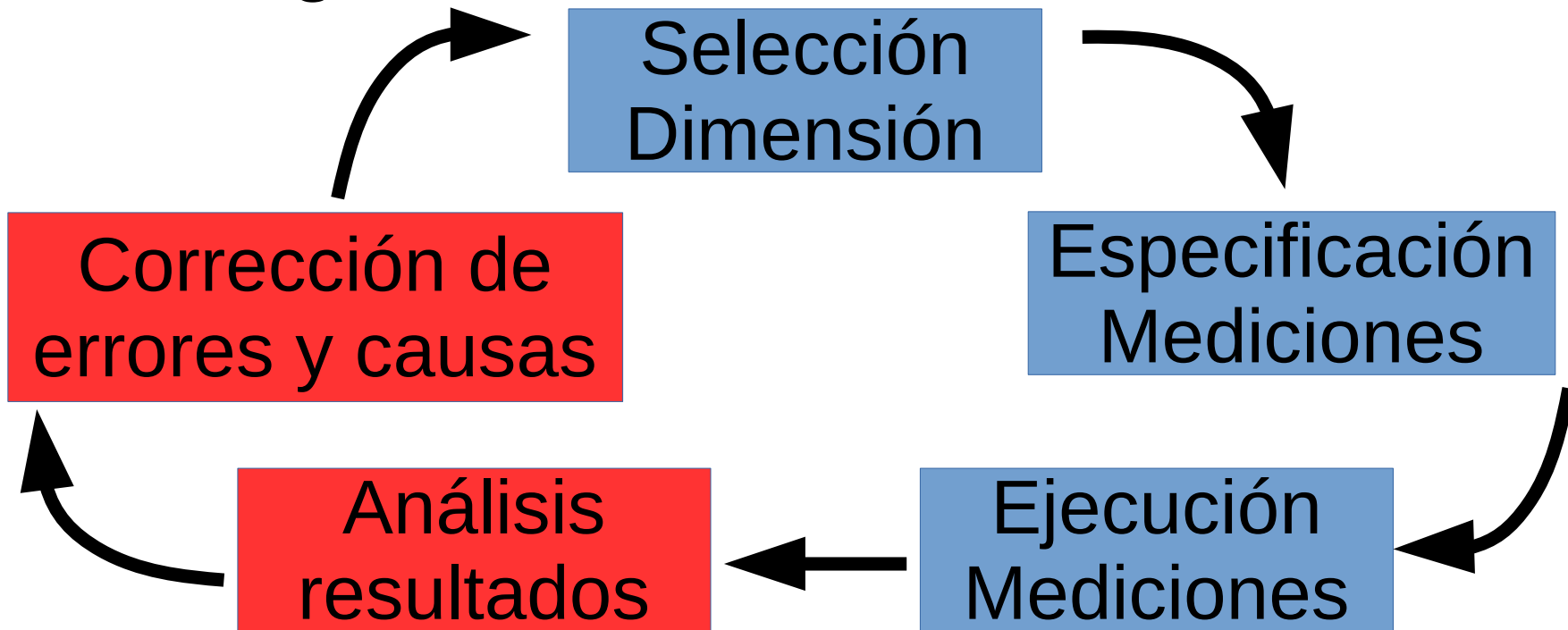
2

- Proyecto dividido en iteraciones



- Una dimensión por iteración

Estrategia



Objetivos

- Introducción Calidad Datos

1

- Experiencia en Trazabilidad trámites

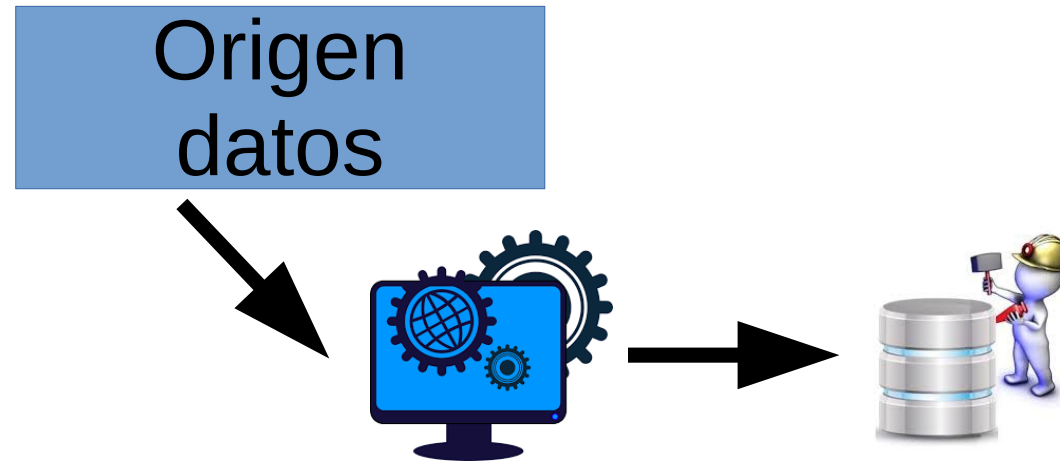
2

- **Lecciones aprendidas**

3

Correcciones

- Esfuerzo importante de corrección de datos y errores



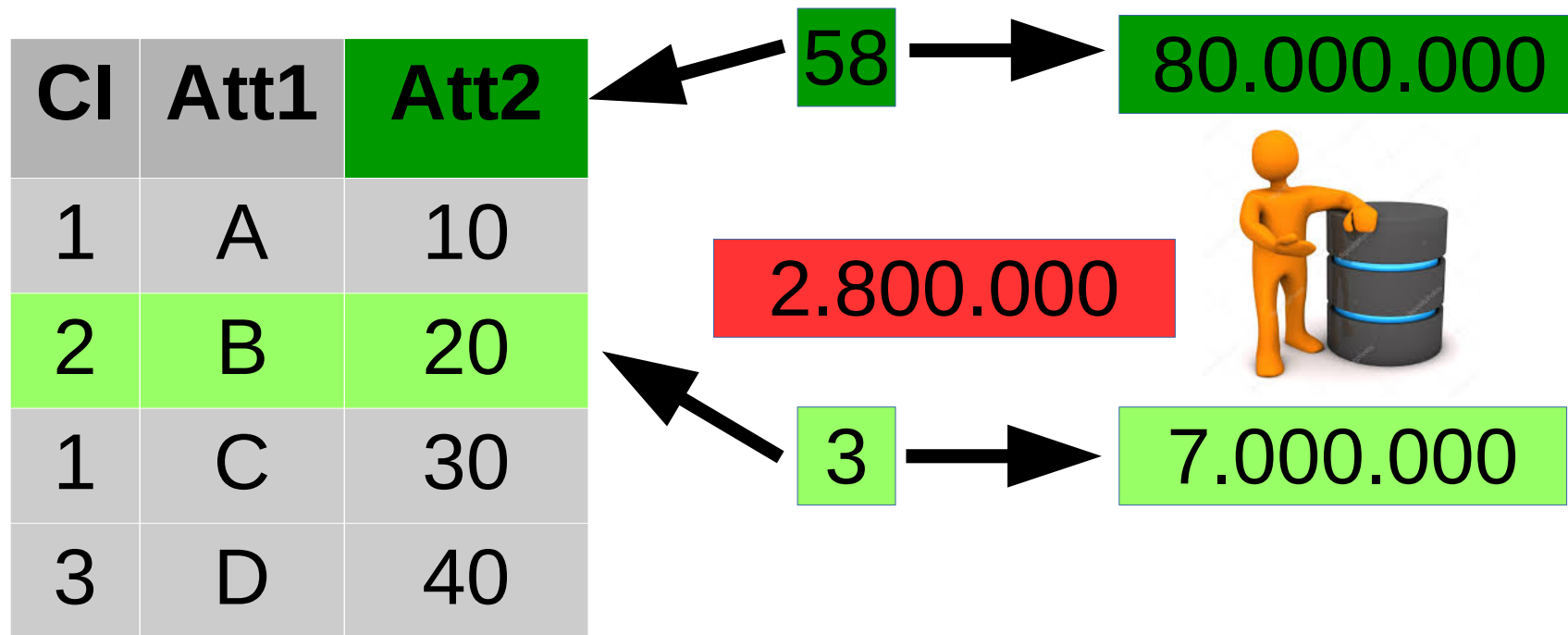
Datos generados en mediciones

| CI | Att1 | Att2 | | |
|----|------|------|---|---|
| 1 | A | 10 | ← | 4 |
| 2 | B | 20 | ← | 1 |
| 1 | C | 30 | | |
| 3 | D | 40 | | |

Diagram illustrating data generated in measurements. The table shows columns CI, Att1, and Att2. The Att2 column is highlighted in green. To the right of the table, two rows are highlighted in green, each with a value in a green box and arrows pointing to the corresponding Att2 value:

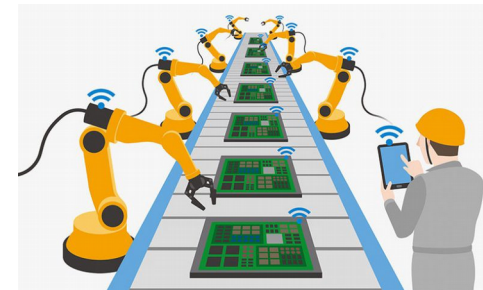
- Row 1: Value 4, arrows pointing to Att2 values 10 and 20.
- Row 2: Value 1, arrows pointing to Att2 values 20 and 30.

Datos generados en mediciones



Automatización y Paneles

- Automatización de mediciones, para sucesivas ejecuciones
- Diseñar panel de resultados para mostrar resultados a usuarios



Construir comunidad

- Compartir experiencias
- Dudas/Ayuda



diego.rosselli@agesic.gub.uy





¡MUCHAS GRACIAS!

diego.rosselli@agesic.gub.uy