



Calidad del dato y depuración de bases de datos en salud pública

Jorge Mario Angulo Mosquera

Bacteriólogo, Msc en Epidemiología

Secretaría Departamental de Salud del Valle del Cauca

GOBERNACIÓN
Departamento del
Valle del Cauca



www.valledelcauca.gov.co/





Calidad del dato y depuración de bases de datos en salud pública

Jorge Mario Angulo Mosquera

Bacteriólogo, Msc en Epidemiología

Secretaría Departamental de Salud del Valle del Cauca

Definiciones



Dato:

Calidad del dato:

Eventos:

Eventos de interés en salud pública:

Definiciones



Dato:

- Un dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa.

Calidad del dato:

- Es la cualidad de un conjunto de información recogida en una base de datos o un sistema de información que reúne entre sus atributos la exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad y confiabilidad, para la toma de decisiones.

Eventos:

- Sucesos o circunstancias que pueden modificar o incidir en la situación de salud de un individuo o una comunidad y que, para efectos del presente decreto, se clasifican en condiciones fisiológicas, enfermedades, discapacidades y muertes; factores protectores y factores de riesgo relacionados con condiciones del medio ambiente, consumo y comportamiento; acciones de protección específica, detección temprana y atención de enfermedades y demás factores determinantes asociados.

Eventos de interés en salud pública:

- Aquellos eventos considerados como importantes o trascendentes para la salud colectiva por parte del Ministerio de la Protección Social, teniendo en cuenta criterios de frecuencia, gravedad, comportamiento epidemiológico, posibilidades de prevención, costo– efectividad de las intervenciones, e interés público; que, además, requieren ser enfrentados con medidas de salud pública.

Introducción y contexto



Importancia de la calidad de datos en salud pública:

- La calidad de los datos es esencial para la consistencia del reporte, la confianza de los usuarios y para la eficacia de los procesos operativos.

Estándares internacionales (FAIR, GATHER)

- FAIR (Findability, Accessibility, Interoperability, and Reuse)
- GATHER (Guidelines for Accurate and Transparent Health Estimates Reporting)

<https://data.who.int/es/about/data/who-data-principles>

<https://www.go-fair.org/fair-principles/>

https://cdn.who.int/media/docs/default-source/world-health-data-platform/gather/gather_statement_spanish.pdf?sfvrsn=a79ea9ad_10&download=true

GOBERNACIÓN
Departamento del
Valle del Cauca



Normativa Colombiana

DECRETO 1389 DE 2022

(Julio 28)

“Por el cual se adiciona el Título 24 a la Parte 2 del Libro 2 del Decreto Único 1078 de 2015, Reglamentario del Sector de Tecnologías de la Información y las Comunicaciones, con el fin de establecer los lineamientos generales para la gobernanza en la infraestructura de datos y se crea el Modelo de gobernanza de la infraestructura de datos”

EL PRESIDENTE DE LA REPÚBLICA DE COLOMBIA

GOBERNACIÓN
Departamento del
Valle del Cauca



Gobernanza de datos

Definición:

La gobernanza de datos es un conjunto de prácticas para tomar decisiones acerca de los datos y gestionarlos durante todo su ciclo de vida, con el objeto de optimizar la capacidad de la organización para utilizar datos que generen información de calidad a fin de fundamentar políticas, estrategias y gestión operativa.

Objetivos:

La gobernanza de datos permite a las organizaciones gestionar, proteger, mantener y utilizar datos de manera eficaz para generar información que mejore la calidad de la atención de salud, los resultados en materia de salud y el desempeño del sistema de salud.

Responsables:

cualquier organización que recopile, gestione o utilice datos de salud deberá practicar la gobernanza de datos.

Dimensiones:

integridad, consistencia, seguridad, privacidad

Gobernanza de datos



Calidad del dato en salud pública

Calidad de datos es la cualidad de un conjunto de información recogida en una base de datos, un sistema de información que reúne entre sus atributos los siguientes:

Exactitud:

El dato refleja la realidad.
Ej: Edad correcta del paciente.

Completitud:

Todos los campos necesarios están diligenciados.
Ej: Sexo y fecha de nacimiento registrados.

Coherencia:

Información lógica y sin contradicciones.
Ej: No mujer con cáncer de próstata.

Integridad:

Los datos mantienen sus relaciones lógicas.
Ej: Código de hospital existe en la tabla de instituciones.

Actualización:

Datos recientes y oportunos.
Ej: Notificación de brote sin retraso.

Accesibilidad:

El dato puede ser consultado con seguridad.
Ej: Investigador autorizado accede con confidencialidad.

Relevancia:

El dato es útil para el análisis.
Ej: Recolectar presión arterial, no color favorito.

Confiabilidad:

Fuente y proceso de captura seguros.
Ej: Registro de laboratorio estandarizado.

Aseguramiento de la calidad



Es el proceso de verificación de la fiabilidad y efectividad de los datos, que debe realizarse periódicamente.



Recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores vacíos

- Descubrimiento de datos.
- Perfilado de datos.
- Reglas de calidad de datos
- Monitorización de la calidad de los datos
- Reporte de calidad de datos
- Limpieza de datos

<https://bookdown.org/paranadagarcia/database/calidad-de-datos.html#calidad-de-datos>

Limpieza de datos

Limpieza de datos siempre esta asociado a los objetivos determinados en el proceso de captura de datos. Como se ha mencionado en forma reiterada esto esta definido por las características de la organización: ¿para qué se estan capturando datos? ¿qué tipo de datos se requieren? según las respuestas se requiere definir diversos procesos de limpieza o ninguno.

Un buen diseño inicial de captura de datos que esta alineado con los requerimientos establecidos, hará practicamente innecesario un proceso que es bastante costoso en recursos humanos.

La limpieza de los datos incluye 5 actividades principales:

- Depurar
- Corregir
- Estandarizar
- Relacionar
- Consolidar

Técnicas de limpieza y depuración

- El proceso de limpieza de datos permite identificar datos incompletos, incorrectos, inexactos, no pertinentes, duplicados, etc. y luego substituir, modificar o eliminar estos datos sucios ("data duty").
- Ejemplo:
 - Corrección

| nombres | apellidos | genero | calle | numero | sector | comuna |
|-------------|-----------|--------|---------------------|--------|-----------------|------------|
| CARLOS IVAN | ERNADEZ | F | COSTA SAN FRANCISCO | 890 | VILLA SAN PEDRO | CONCEPCION |

| nombres | apellidos | genero | calle | numero | sector | comuna |
|-------------|-----------|--------|---------------------|--------|-----------------|---------------------|
| CARLOS IVAN | HERNANDEZ | M | COSTA SAN FRANCISCO | 890 | VILLA SAN PEDRO | SAN PEDRO DE LA PAZ |

<https://bookdown.org/paranedagarcia/database/calidad-de-datos.html#calidad-de-datos>

Técnicas de limpieza y depuración

- Ejemplo:
 - Estandarización

| nombres | apellidos | genero | calle | numero | sector | comuna |
|-------------|-------------|--------|----------------------------|---------|--------|----------|
| CARLOS IVAN | ERNADEZ | H | O'HIGGINS | No. 890 | CENTRO | SANTIAGO |
| JUAN | VERA | M | BERNARDO O'HIGGINS | 700 | MAIPU | STGO |
| MARCELA | VARAS SOLIS | F | ALAMEDA BERNARDO O'HIGGINS | #660 | CENTRO | SANTIAGO |

| nombres | apellidos | genero | calle | numero | sector | comuna |
|---------|-----------|--------|----------------------------|--------|--------|----------|
| | | | ALAMEDA BERNARDO O'HIGGINS | 890 | | SANTIAGO |

<https://bookdown.org/paranedagarcia/database/calidad-de-datos.html#calidad-de-datos>

Análisis de datos

| Correo | Estado |
|--------------------|------------|
| usuario1@gmail.com | ✓ Válido |
| usuario@correo.com | ✗ Inválido |
| prueba@hotmail.com | ✓ Válido |

Transformación de datos

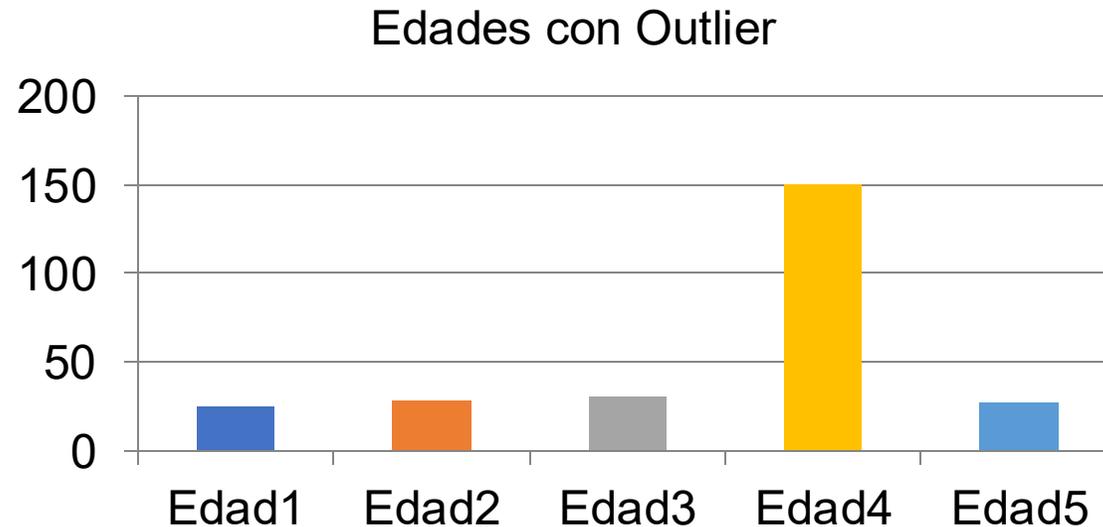
| fec_con_ | ini_sin_ | Tiempo de oportunidad Dx | Oportunidad diagnóstico |
|------------|------------|-----------------------------|----------------------------|
| 16/04/2024 | 26/03/2024 | 21 | No |
| 25/08/2024 | 21/08/2024 | 4 | No |
| 25/08/2024 | 24/08/2024 | 1 | Si |
| 27/08/2024 | 20/08/2024 | 4 | No |
| 28/08/2024 | 24/08/2024 | 4 | No |
| 26/08/2024 | 25/08/2024 | 3 | No |
| 30/08/2024 | 28/08/2024 | 2 | Si |
| 24/09/2024 | 22/09/2024 | 2 | Si |

<https://bookdown.org/paranedagarcia/database/calidad-de-datos.html#calidad-de-datos>

Eliminación de duplicados

| Nombre | Documento |
|---------------|-----------|
| Juan Pérez | 12345 |
| Juan A. Pérez | 12345 |
| Ana Gómez | 54321 |

Método estadístico



<https://bookdown.org/paranedagarcia/database/calidad-de-datos.html#calidad-de-datos>

Herramientas técnicas

- Software: R, Python, SPSS, Excel, EpiInfo, Stata, etc
- Validaciones automatizadas

Proyecto final

- Realizar descarga de base de datos.
- Revisar diccionario de datos.
- Realizar revisión de base de datos aplicando las técnicas de limpieza y depuración.
- Documentación de los hallazgos
- Corrección de la base de datos.

GRACIAS

GOBERNACIÓN
Departamento del
Valle del Cauca





Elaborado por:

Jorge Mario Angulo Mosquera

Secretaría Departamental de Salud del Valle del Cauca

The Global Health Network LAC – Universidad del Valle

Este material de aprendizaje electrónico es propiedad de la Red Global de Salud.

Usted es libre de compartir o adaptar este material, pero debe atribuirlo a The Global Health Network utilizando el enlace <https://lac.tghn.org/>

Mayor información en: proyecto.globalhealthnetworklac@correounivalle.edu.co