

# Framework para la Gestión de la Calidad de Datos en Gobierno Digital

Modelo de Calidad de Datos de Referencia

## Temas

---

- Descripción General del Modelo
- Marco Conceptual Asociado
- Dimensiones y Factores de Calidad
- Detalle de Métricas de Calidad

Los temas se presentan en forma de “guía rápida”. Por más detalles consultar el documento: “Framework para la Gestión de la Calidad de Datos en Gobierno Digital” [AGE2020].

2

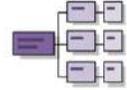
## Temas

---

- Descripción General del Modelo
- Marco Conceptual Asociado
- Dimensiones y Factores de Calidad
- Detalle de Métricas de Calidad

# Descripción General del Modelo

## Introducción



Marco Teórico



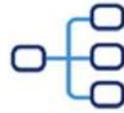
Proceso de Gestión de  
Calidad de Datos



Marco Conceptual



Caso de estudio



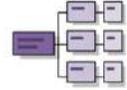
Modelo de Calidad de  
Datos de Referencia



Recursos de soporte

# Descripción General del Modelo

## Introducción



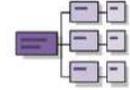
- ❑ Un elemento clave para la gestión de calidad de datos es contar con un **modelo de calidad de datos** que considere las características específicas del escenario de trabajo:
  - p. ej. requerimientos, problemas, prioridades
- ❑ El **Modelo de Calidad de Datos de Referencia (MCDR)** tiene como fin guiar y facilitar la definición de modelos de calidad de datos para escenarios de trabajo específicos en un contexto de gobierno digital

5

El MCDR es aplicable en contextos de **gobierno digital**, donde distintas **organizaciones** llevan a cabo **procesos de negocio** y colaboran entre sí con el fin de ofrecer **servicios públicos** a los **ciudadanos**.

# Descripción General del Modelo

## Introducción



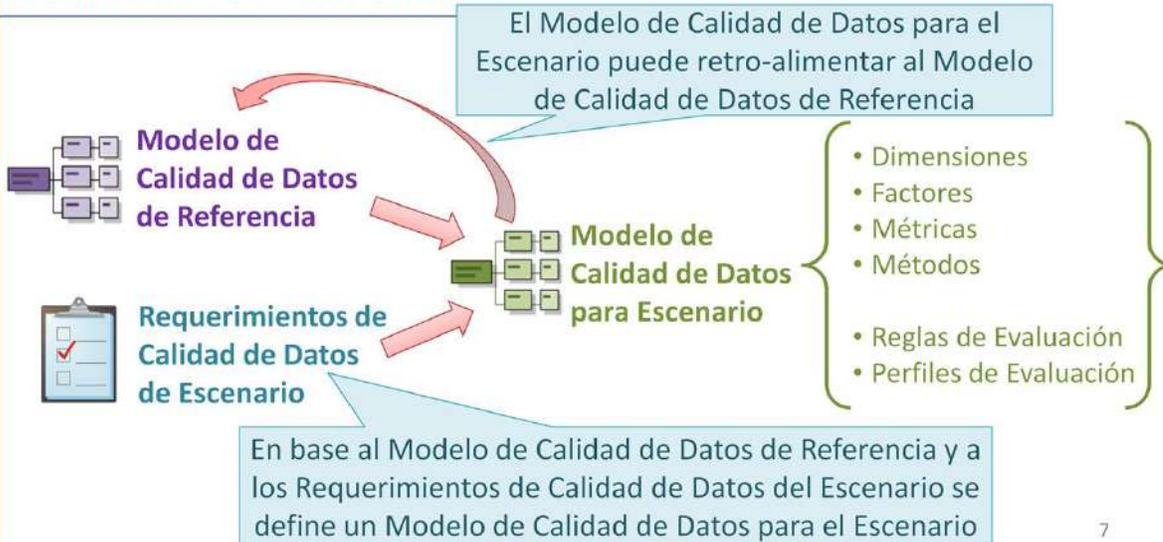
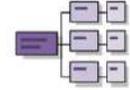
□ El MCDR provee un conjunto, extensible e instanciable al contexto de aplicación, de elementos de calidad de datos:

- 5 dimensiones de calidad
  - ❖ p. ej. Frescura
- 17 factores de calidad
  - ❖ p. ej. Actualidad, Oportunidad
- Más de 40 métricas de calidad
  - ❖ p. ej. DesactualizaciónPorFecha



# Descripción General del Modelo

## Utilización del Modelo



## Temas

---

- Introducción
- Marco Conceptual Asociado
- Dimensiones y Factores de Calidad
- Detalle de Métricas de Calidad

# Marco Conceptual Asociado

## Introducción

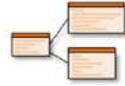
---



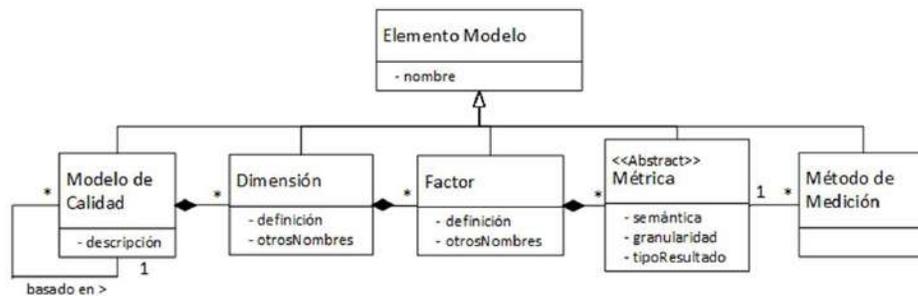
- El **Modelo Conceptual** del *framework* incluye conceptos que permiten la especificación del **Modelo de Calidad de Datos de Referencia**, así como de modelos de calidad de datos para escenarios específicos:
  - Elementos Modelo
  - Aplicabilidad de Elementos Modelo
  - Métricas

# Marco Conceptual Asociado

## Elementos Modelo



- Los conceptos **Elemento Modelo** permiten especificar modelos de calidad de datos y sus elementos (i. e. dimensiones, factores, métricas y métodos).



10

## Marco Conceptual Asociado

### Elementos Modelo

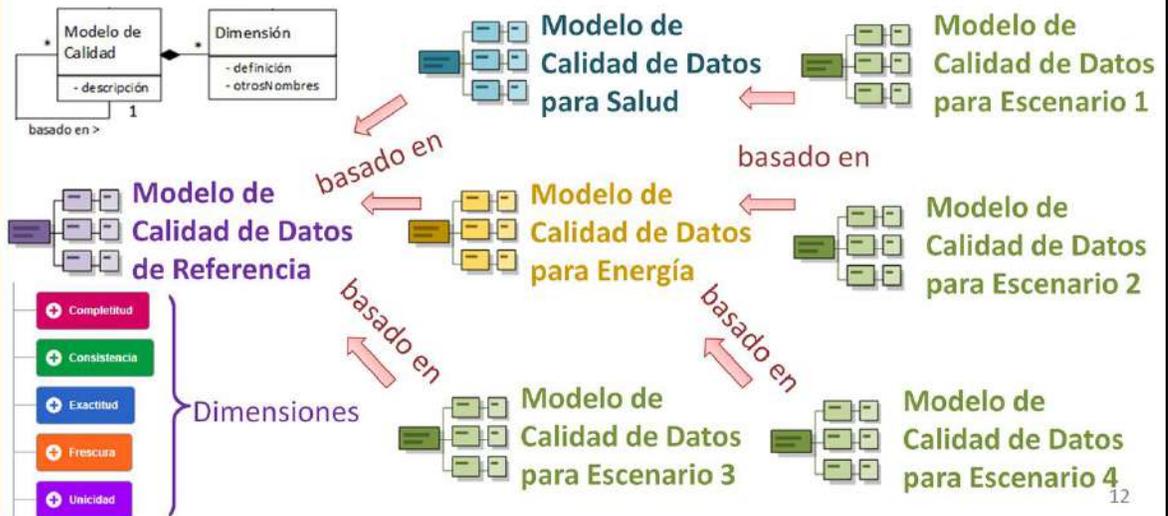
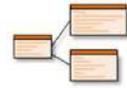


- ❑ Un **Modelo de Calidad** está compuesto por dimensiones que representan las cuestiones de calidad de datos relevantes para un determinado escenario o conjunto de escenarios. exactitud
- ❑ una **Dimensión** captura una faceta a alto nivel de la calidad
- ❑ un **Factor** representa un aspecto particular de una dimensión
- ❑ una **Métrica** es un instrumento que define la forma de medir un factor de calidad correctitud sintáctica  
porcentaje de datos sintácticamente correctos
- ❑ un **Método de Medición** es un proceso que implementa una métrica y se encarga de tomar medidas para la misma

11

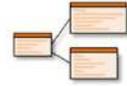
# Marco Conceptual Asociado

## Elementos Modelo – Ejemplos de Instancias

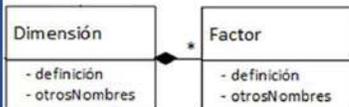


# Marco Conceptual Asociado

## Elementos Modelo – Ejemplo Dimensión



- ❑ Especificación de la dimensión Unicidad para el modelo de referencia

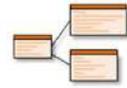


Unicidad	
<b>Definición</b>	Captura el grado en el que un dato del mundo real es representado en forma única. (Adaptado de [BS16])
<b>Otros Nombres</b>	Uniqueness (ingl.). En [BS16] se habla también de redundancia para el caso de linked data. La norma [ISO13] la considera como parte del factor Compleción y lo establece como una medida para este factor.
<b>Factores</b>	No-duplicación No-contradicción

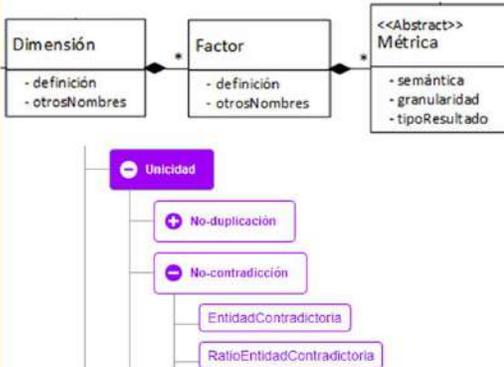
13

# Marco Conceptual Asociado

## Elementos Modelo – Ejemplo Factor



- Especificación del factor No-contradicción de la dimensión Unicidad del modelo de referencia



No-contradicción	
<b>Definición</b>	Captura el grado de duplicación (o repetición) de una misma instancia de entidad del mundo real que es representada con datos contradictorios.
<b>Otros Nombres</b>	Contradiction-free (ingl.)
<b>Métricas</b>	EntidadContradictoria RatioEntidadContradictoria

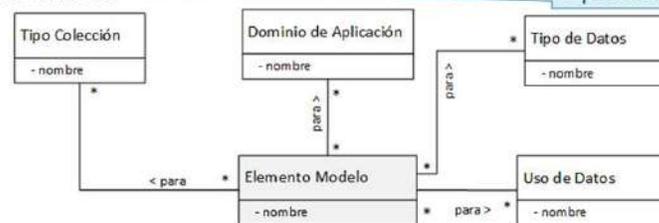
# Marco Conceptual Asociado

## Aplicabilidad de Elementos Modelo



□ Los Elementos Modelo pueden aplicar a uno o más:

- tipos de datos — alfanumérico, imagen, geográfico
- tipos de colecciones de datos — base de datos relacional, documental
- dominios de aplicación — salud, transporte, energía
- usos de datos — operativa, toma de decisiones

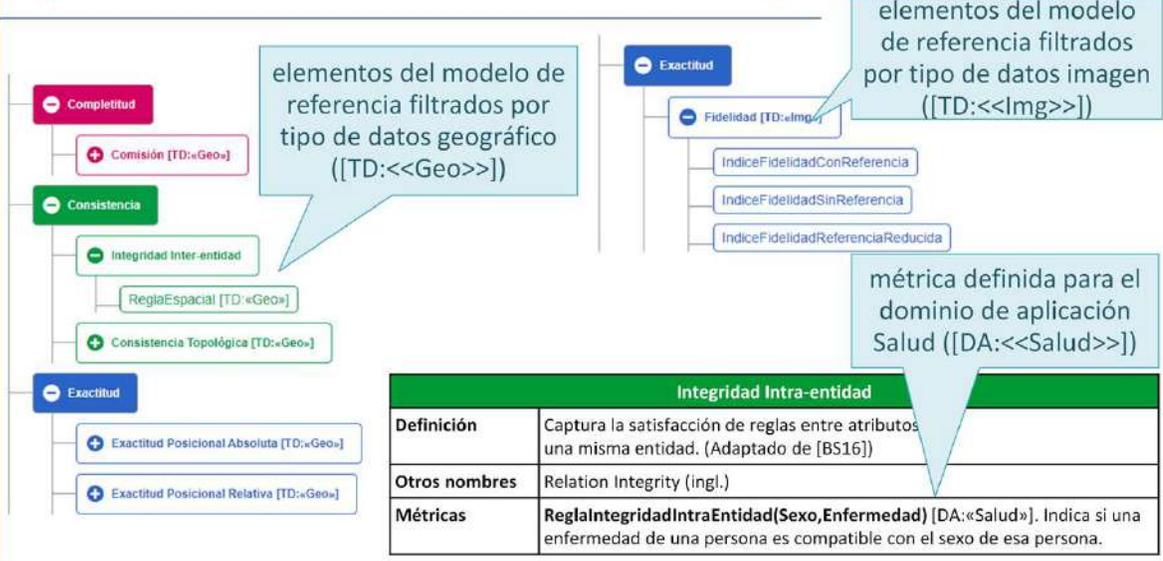
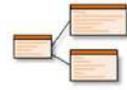


15

Los elementos del modelo (p. ej. dimensiones, factores, métricas) pueden aplicar a uno o más tipos de datos (p. ej. alfanumérico, imagen, geográfico), tipos de colecciones de datos (p. ej. base de datos relacional, base de datos documental), dominios de aplicación (p. ej. salud, transporte, energía) y usos de datos (p. ej. operativa, toma de decisiones).

# Marco Conceptual Asociado

## Aplicabilidad de Elementos Modelo - Ejemplos

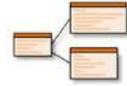


Por ejemplo, un modelo puede tener un factor Consistencia-Topológica que aplica únicamente al tipo de datos Geográfico.

Asimismo, un modelo puede tener una métrica Consistencia-Enfermedad-Sexo que aplica únicamente al dominio de la Salud.

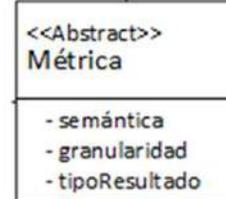
# Marco Conceptual Asociado

## Métricas: Definición



❑ Para definir una métrica es necesario especificar:

- semántica (qué se mide)
- granularidad (sobre qué se mide)
- tipo de resultado (p. ej. porcentaje)

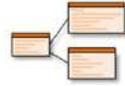


❑ Ejemplo:

RatioNoNulos-DireccionCiente	
Semántica	Proporción de clientes que tienen un valor no nulo para la dirección.
Granularidad	atributo
Tipo de Resultado	Intervalo real [0.0, 1.0]

# Marco Conceptual Asociado

## Métricas: Granularidad

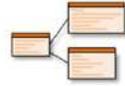


- ❑ La granularidad de una métrica específica sobre qué se realiza la medición

Granularidad considerada en Framework	Correspondencia con Modelo Relacional
instanciaAtributo	celda
atributo	columna
conjuntoAtributos	conjunto de columnas
instanciaEntidad	tupla
entidad	tabla
conjuntoEntidades	conjunto de tablas
colección	base de datos
conjuntoColecciones	conjunto de bases de datos
organización	organización
conjuntoOrganizaciones	conjunto de organizaciones

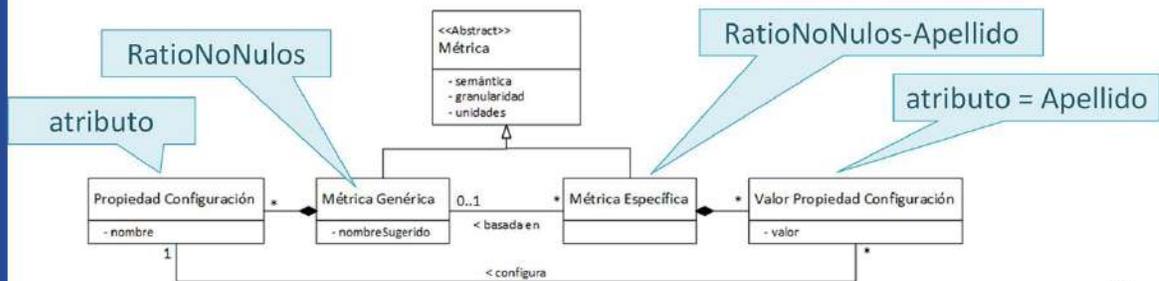
# Marco Conceptual Asociado

## Métricas Genéricas y Específicas



□ El *framework* distingue entre:

- Métricas Genéricas
- Métricas Específicas



## Marco Conceptual Asociado

### Métricas Genéricas y Específicas

---



#### ❑ Métricas Genéricas

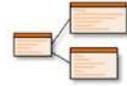
- representan una familia de métricas con características similares
  - ❖ p. ej. RatioNoNulos, Formato
- tienen como objetivo guiar, agilizar y uniformizar la creación de nuevas métricas

#### ❑ Métricas Específicas

- pueden refinar métricas genéricas para abordar requerimientos más específicos
  - ❖ p. ej. RatioNoNulos-Apellido, RatioNoNulos-Pais, Formato-Pais-ISOAlpha3
- el refinamiento se realiza:
  - ❖ estableciendo valores para las Propiedades de Configuración (p. ej. atributo: Apellido, País) que define la métrica genérica
  - ❖ especificando una semántica más concreta

# Marco Conceptual Asociado

## Métricas Genéricas y Específicas: Ejemplo



### ☐ Métrica Genérica

Formato	
<b>Semántica</b>	Indica si el valor de un atributo cumple con el formato definido para ese atributo según algún estándar o diccionario.
<b>Granularidad</b>	instanciaAtributo
<b>Tipo de Resultado</b>	Boolean
<b>Nombre Sugerido</b>	Formato (atributo, estándar o diccionario)
<b>Propiedades Configuración</b>	atributo estándar o diccionario

### ☐ Métricas Específicas

Formato(país, ISOAlpha2)	
<b>Semántica</b>	Indica si el código de un país está en ISO 3166-1 alpha-2.
<b>Valor Propiedades</b>	atributo = país, estándar = ISO 3166-1 alpha-2

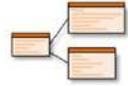
Formato(país, ISOAlpha3)	
<b>Semántica</b>	Indica si el código de un país está en ISO 3166-1 alpha-3.
<b>Valor Propiedades</b>	atributo = país, estándar = ISO 3166-1 alpha-3

Formato(cédula, DNIC)	
<b>Semántica</b>	Indica si el valor del atributo numDoc cumple con el formato de cédula de identidad uruguaya establecido por DNIC.
<b>Valor Propiedades</b>	atributo = cédula, estándar = DNIC

## Marco Conceptual Asociado

### Métricas por Agregación

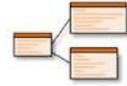
---



- ❑ Las métricas definidas por agregación, son métricas que pueden calcularse a partir de métricas ya existentes
- ❑ M2 es una métrica agregada de M1 si la semántica de M2 puede expresarse como un cálculo que agrega un conjunto de resultados de M1
- ❑ Además, si M1 está definida para la mínima granularidad posible, entonces M1 es una métrica atómica.

# Marco Conceptual Asociado

## Métricas por Agregación: Ejemplo



### ☐ Métrica Atómica

NoNulo	
<b>Semántica</b>	Indica si una instancia de atributo tiene un valor no nulo. Puede ser necesario especificar un diccionario con todos los valores del atributo que se consideran nulos o vacíos, cuando existe más de una posibilidad.
<b>Granularidad</b>	instanciaAtributo
<b>Tipo de Resultado</b>	Boolean

### ☐ Métrica Agregada

RatioNoNulos	
<b>Semántica</b>	Proporción de valores no nulos según la métrica NoNulo, sobre el total de instancias de atributo sobre las que se realizó la medición.
<b>Granularidad</b>	atributo
<b>Tipo de Resultado</b>	Intervalo real [0, 1]

Persona	Persona	Persona	
- Nombre: Juan	- Nombre: José	- Nombre: María	RatioNoNulos = 0,3333
- Apellido: NULL	- Apellido: NULL	- Apellido: Perez	
NoNulo= false	NoNulo= false	NoNulo= true	

23

# Marco Conceptual Asociado

## Métricas por Agregación: Tipos



□ Algunos tipos muy utilizados de métricas agregadas son:

Métrica Agregada	Semántica
Ratio	Proporción de valores en «true» sobre el total de valores medidos.
RatioUmbral	Proporción de valores iguales o mayores a un umbral, sobre el total de valores medidos.
Promedio	Valor promedio de todos los valores medidos.
PromedioPonderado	Promedio ponderado de los valores medidos, en donde cada valor se multiplica por un coeficiente o peso.

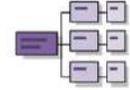
## Temas

---

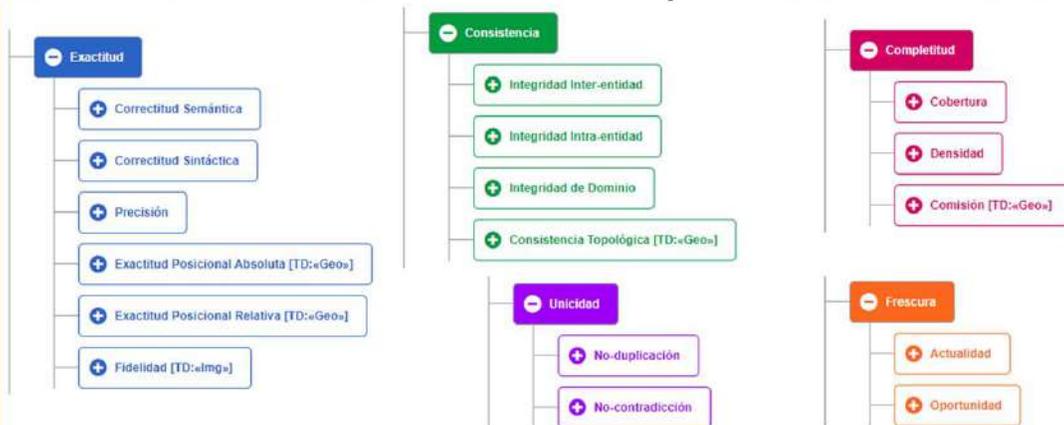
- Introducción
- Marco Conceptual Asociado
- Dimensiones y Factores de Calidad
- Detalle de Métricas de Calidad

# Dimensiones y Factores de Calidad

## Introducción

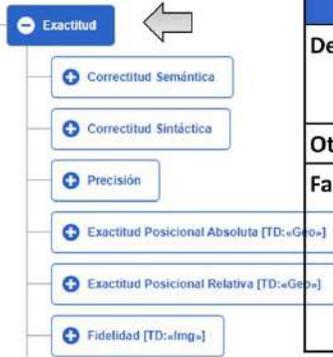
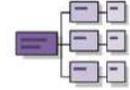


El modelo define 5 dimensiones y 17 factores de calidad



# Dimensiones y Factores de Calidad

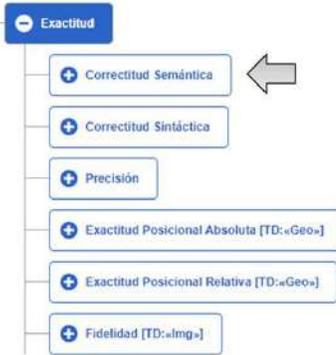
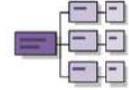
## Dimensión: Exactitud



Exactitud	
<b>Definición</b>	Proximidad entre un valor de datos $v$ y un valor de datos $v'$ , considerado como la representación correcta del fenómeno del mundo real que $v$ intenta representar. (Adaptado de [BS16])
<b>Otros Nombres</b>	Accuracy (ingl.), correctitud
<b>Factores</b>	Correctitud Semántica Correctitud Sintáctica Precisión Exactitud Posicional Absoluta Exactitud Posicional Relativa Fidelidad

# Dimensiones y Factores de Calidad

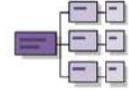
## Factor: Correctitud Semántica



Correctitud Semántica	
<b>Definición</b>	Proximidad entre el valor $v$ de un atributo y su verdadero valor $v'$ . (Adaptado de [BS16])
<b>Otros Nombres</b>	Semantic Correctness (ingl.)
<b>Métricas Genéricas</b>	CorrectitudSemDebil CorrectitudSemFuerte RatioCorrectitudSemFuerte RatioCorrectitudSemDébil

# Dimensiones y Factores de Calidad

## Factor: Correctitud Sintáctica



### Exactitud

+ Correctitud Semántica

+ Correctitud Sintáctica

+ Precisión

+ Exactitud Posicional Absoluta [TD:«Geo»]

+ Exactitud Posicional Relativa [TD:«Geo»]

+ Fidelidad [TD:«Img»]

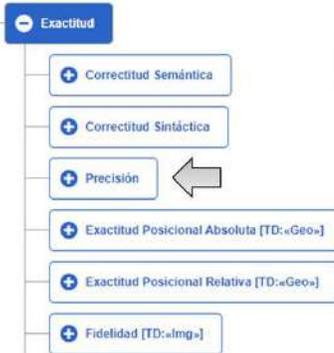
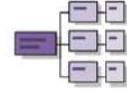


### Correctitud Sintáctica

<b>Definición</b>	Proximidad entre el valor v de un atributo y los elementos del dominio de definición de dicho atributo. (Adaptado de [BS16])
<b>Otros Nombres</b>	Syntactic Correctness (ingl.)
<b>Métricas Genéricas</b>	Formato
<b>Métricas Específicas</b>	Formato(Pais, ISOAlpha3) Formato(Enfermedad, CIE10) [DA:«Salud»] Formato(NumeroDocumento, DNIC)

# Dimensiones y Factores de Calidad

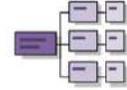
## Factor: Precisión



Precisión	
<b>Definición</b>	Captura el grado de detalle que posee un dato que lo hace útil para un determinado uso o que permite discriminarlo de otros datos que no son exactamente iguales. (Adaptado de [BS16]) y [ISO08])
<b>Otros Nombres</b>	Precision (ingl.)
<b>Métricas Genéricas</b>	Escala ErrorEstandar

# Dimensiones y Factores de Calidad

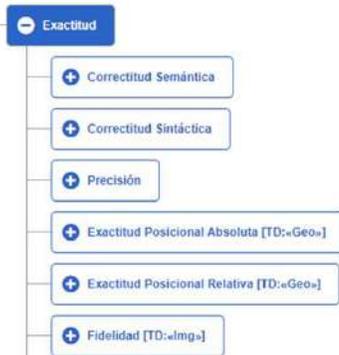
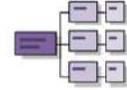
## Factor: Exactitud Posicional Absoluta [TD:«Geo»]



Exactitud Posicional Absoluta (TD:«Geo»)	
<b>Definición</b>	Proximidad de los valores reportados de las coordenadas a los valores verdaderos o aceptados como tales [ISO13]
<b>Otros Nombres</b>	Absolute positional accuracy (ingl.), Exactitud posicional externa
<b>Métricas Genéricas</b>	ÍndiceErroresPosicionalesPorUmbral ValorMedioIncertidumbrePosicional

# Dimensiones y Factores de Calidad

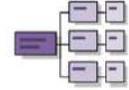
## Factor: Exactitud Posicional Relativa [TD:«Geo»]



Exactitud Posicional Relativa (TD:«Geo»)	
<b>Definición</b>	Proximidad de las posiciones relativas de los objetos geográficos de un conjunto de datos a sus respectivas posiciones relativas verdaderas o aceptadas como tales [ISO13]
<b>Otros Nombres</b>	Relative positional accuracy (ingl.), Exactitud posicional interna
<b>Métricas Genéricas</b>	ErrorHorizontalRelativo

# Dimensiones y Factores de Calidad

## Factor: Fidelidad [TD:«Img»]



### Exactitud

+ Correctitud Semántica

+ Correctitud Sintáctica

+ Precisión

+ Exactitud Posicional Absoluta [TD:«Geo»]

+ Exactitud Posicional Relativa [TD:«Geo»]

+ Fidelidad [TD:«Img»]

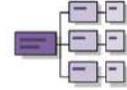


### Fidelidad (TD:«Img»)

<b>Definición</b>	Captura el grado de concordancia aparente de una imagen con el original. [BS16]
<b>Otros Nombres</b>	Fidelity (ingl.), faithfulness (ingl.), genuineness (ingl.)
<b>Métricas Genéricas</b>	IndiceFidelidadConReferencia IndiceFidelidadSinReferencia IndiceFidelidadReferenciaReducida

# Dimensiones y Factores de Calidad

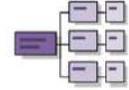
## Dimensión: Consistencia



Consistencia	
<b>Definición</b>	Captura la violación de las reglas semánticas definidas sobre un conjunto de entidades de negocio o de sus atributos. En un modelo relacional, las restricciones de integridad son un ejemplo de tales reglas semánticas. (Adaptado de [BS16])
<b>Otros Nombres</b>	Consistency (ingl.), cohesión, coherencia
<b>Factores</b>	Integridad Inter-entidad Integridad Intra-entidad Integridad de Dominio Consistencia Topológica (TD:«Geo»)

# Dimensiones y Factores de Calidad

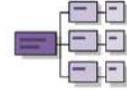
## Factor: Integridad Inter-entidad



Integridad Inter-entidad	
<b>Definición</b>	Captura la satisfacción de reglas entre atributos de diferentes entidades de negocio. (Adaptado de [BS16])
<b>Otros Nombres</b>	Referential Integrity (ingl.), Integridad Inter-relacion
<b>Métricas Genéricas</b>	ReglaIntegridadInterEntidad ReglaEspacial [TD:«Geo»]
<b>Métricas Específicas</b>	ReglaIntegridadInterEntidad(Sexo, Enfermedad) [DA:«Salud»]

# Dimensiones y Factores de Calidad

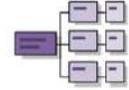
## Factor: Integridad Intra-entidad



Integridad Intra-entidad	
<b>Definición</b>	Definición Captura la satisfacción de reglas entre atributos de una misma entidad. (Adaptado de [BS16])
<b>Otros Nombres</b>	Relation Integrity (ingl.)
<b>Métricas Genéricas</b>	ReglaIntegridadIntraEntidad RatIoIntegridadIntraEntidad
<b>Métricas Específicas</b>	ReglaIntegridadIntraEntidad(Sexo,Enfermedad) [DA:«Salud»]

# Dimensiones y Factores de Calidad

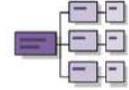
## Factor: Integridad de Dominio



Integridad de Dominio	
<b>Definición</b>	Captura la satisfacción de reglas sobre los valores posibles que puede tomar un atributo. (Adaptado de [BS16])
<b>Otros Nombres</b>	Domain Integrity (ingl.)
<b>Métricas Genéricas</b>	ValoresPosiblesPorExtensión ValoresPosiblesPorComprensión
<b>Métricas Específicas</b>	ValoresPosiblesPE(Sexo, AGESIC)

# Dimensiones y Factores de Calidad

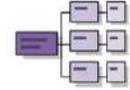
## Factor: Consistencia Topológica [TD:«Geo»]



Consistencia Topológica (TD:«Geo»)	
<b>Definición</b>	Corrección de las características topológicas codificadas explícitamente. Las características topológicas de un conjunto de datos describen las relaciones geométricas entre los ítems del conjunto de datos que no son alteradas por transformaciones elásticas (rubber-sheet transformations) [ISO13]
<b>Otros Nombres</b>	Topological Consistency (ingl.)
<b>Métricas Genéricas</b>	ÍndiceFallosConexiónNodosEnlace

# Dimensiones y Factores de Calidad

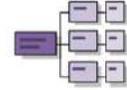
## Dimensión: Completitud



Completitud	
<b>Definición</b>	Captura la medida en que los datos son de la amplitud, profundidad y alcance suficientes para una determinada tarea. (Adaptado de [BS16])
<b>Otros Nombres</b>	Completeness (ingl.)
<b>Factores</b>	Cobertura Densidad Comisión (TD:«Geo»)

# Dimensiones y Factores de Calidad

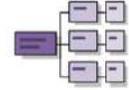
## Factor: Cobertura



Cobertura	
<b>Definición</b>	Captura la proporción entre la cantidad de entidades existentes en una determinada colección de datos, y el total de entidades que deberían existir en dicha colección. La cobertura varía si se utiliza la Asunción de Mundo Cerrado, según la cual una colección de datos debería contener todas las entidades de un tipo, o si se utiliza la Asunción de Mundo Abierto, según la cual una colección de datos puede ser una representación parcial de las entidades del mundo real. (Adaptado de [BS16])
<b>Otros Nombres</b>	Coverage (ingl.)
<b>Métricas Genéricas</b>	RatioCobertura

# Dimensiones y Factores de Calidad

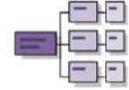
## Factor: Densidad



Densidad	
<b>Definición</b>	Captura la proporción entre la cantidad de instancias de atributo con valores no nulos y el total de instancias de dicho atributo (Adaptado de [BS16]). Un valor nulo de una instancia de atributo A de una entidad E puede interpretarse de varias maneras: <ol style="list-style-type: none"><li>1. E no posee A</li><li>2. se desconoce si E posee A o no</li><li>3. E posee A pero se desconoce su valor</li></ol>
<b>Otros Nombres</b>	Density (ingl.)
<b>Métricas Genéricas</b>	NoNulo DensidadPonderada RatioNoNulos RatioDensidadPonderada

# Dimensiones y Factores de Calidad

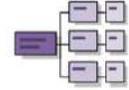
## Factor: Comisión [TD:<<Geo>>]



Comisión (TD:<<Geo>>)	
<b>Definición</b>	Datos excedentes presentes en un conjunto de datos [ISO13]
<b>Otros Nombres</b>	Commission (ingl.)
<b>Métricas Genéricas</b>	ÍtemExcedente ÍndiceItemsExcedentes

# Dimensiones y Factores de Calidad

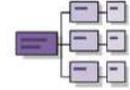
## Dimensión: Unicidad



Compleitud	
<b>Definición</b>	Captura el grado en el que un dato del mundo real es representado en forma única. (Adaptado de [BS16])
<b>Otros Nombres</b>	Uniqueness (ingl.). En [BS16] se habla también de redundancia para el caso de linked data. La norma [ISO13] la considera como parte del factor Compleción y lo establece como una medida para este factor.
<b>Factores</b>	No-duplicación No-contradicción

# Dimensiones y Factores de Calidad

## Factor: No-duplicación



Unicidad

No-duplicación

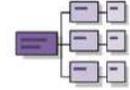
No-contradicción



No-duplicación	
<b>Definición</b>	Captura el grado de duplicación (o repetición) de un mismo dato.
<b>Otros Nombres</b>	Duplication-free (ingl.)
<b>Métricas Genéricas</b>	AtributoDuplicado ConjuntoAtributosDuplicado EntidadDuplicada RatioAtributoDuplicado RatioConjuntoAtributosDuplicado RatioEntidadesDuplicadas

# Dimensiones y Factores de Calidad

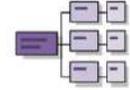
## Factor: No-contradicción



No-contradicción	
<b>Definición</b>	Captura el grado de duplicación (o repetición) de una misma instancia de entidad del mundo real que es representada con datos contradictorios.
<b>Otros Nombres</b>	Contradiction-free (ingl.)
<b>Métricas Genéricas</b>	EntidadContradictoria RatioEntidadContradictoria

# Dimensiones y Factores de Calidad

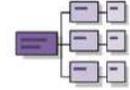
## Dimensión: Frescura



Frescura	
<b>Definición</b>	Captura la rapidez con la que los cambios en el mundo real son reflejados en la actualización de los datos. La frescura es un tipo de exactitud no-estructural dependiente de la variable tiempo, lo que implica que un dato que un determinado momento es correcto, puede no serlo en otro momento. (Adaptado de [BS16])
<b>Otros Nombres</b>	Freshness (ingl.), exactitud temporal
<b>Factores</b>	Actualidad Oportunidad

# Dimensiones y Factores de Calidad

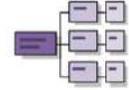
## Factor: Actualidad



Actualidad	
<b>Definición</b>	Captura el tiempo de demora entre un cambio en el mundo real y la correspondiente actualización de los datos.
<b>Otros Nombres</b>	Currency (ingl.)
<b>Métricas Genéricas</b>	DesactualizaciónPorFecha DesactualizaciónPorCambios DesactualizaciónPorFormato

# Dimensiones y Factores de Calidad

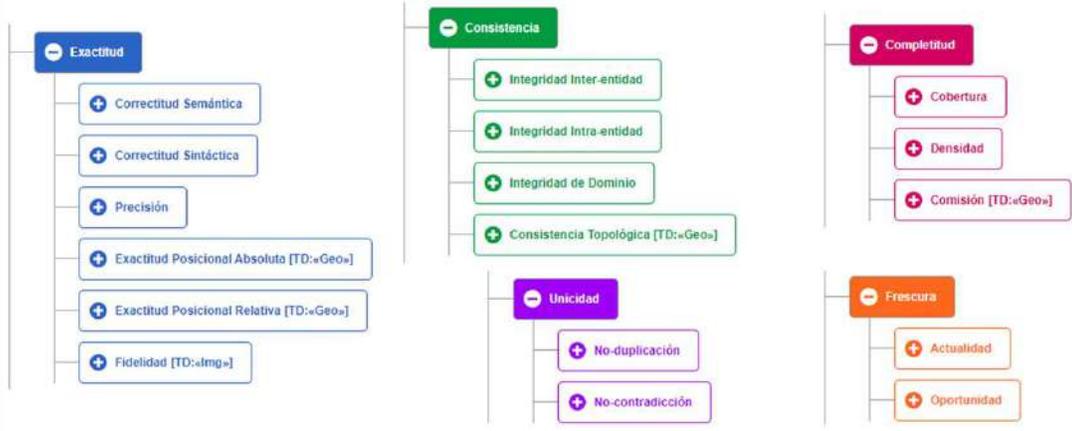
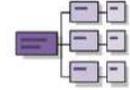
## Factor: Oportunidad



Oportunidad	
<b>Definición</b>	Captura la demora que existe entre la actualización de un dato y el momento en el que éste se encuentra disponible para ser utilizado.
<b>Otros Nombres</b>	Timeliness (ingl.)
<b>Métricas Genéricas</b>	OportunidadAtributoPorFecha OportunidadAtributoPorIntervalo OportunidadEntPorFecha OportunidadEntPorIntervalo:

# Dimensiones y Factores de Calidad

## Resumen



## Temas

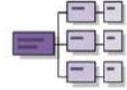
---

- Descripción General del Modelo
- Marco Conceptual Asociado
- Dimensiones y Factores de Calidad
- Detalle de Métricas de Calidad**

# Detalle de Métricas de Calidad

## Introducción

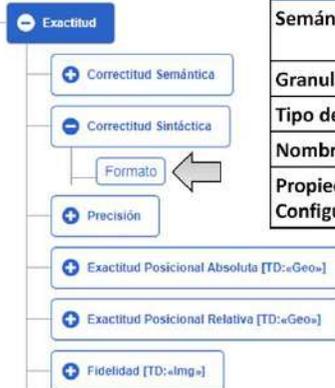
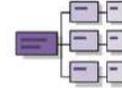
---



- ❑ El modelo brinda detalles de más de 40 métricas, entre las que se encuentran:
  - Formato (Exactitud – Correctitud Sintáctica)
  - ValoresPosiblesPorExtensión (Consistencia – Integridad de Dominio)
  - RatioCobertura (Compleitud – Cobertura)
  - EntidadDuplicada (Unicidad – No-duplicación)
  - DesactualizaciónPorFormato (Frescura – Actualidad)

# Detalle de Métricas de Calidad

## Métrica: Formato

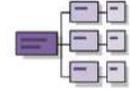


Métrica Genérica: Formato	
<b>Semántica</b>	Indica si el valor de un atributo cumple con el formato definido para ese atributo según algún estándar o diccionario.
<b>Granularidad</b>	instanciaAtributo
<b>Tipo de Resultado</b>	Boolean
<b>Nombre Sugerido</b>	Formato (Atr, Estándar)
<b>Propiedades</b>	1. Atributo=Atr
<b>Configuración</b>	2. Estandar(Atributo) o Diccionario(Atributo)

Métrica Específica: Formato(Enfermedad, CIE10) [DA:«Salud»]	
<b>Semántica</b>	Indica si el código de una enfermedad está en el formato CIE-10 (comienza con una letra válida y tiene una cantidad menor o igual a 6 dígitos).
<b>Basada En</b>	Formato
<b>Propiedades</b>	1. Atr=Enfermedad
<b>Configuración</b>	2. Estandar(Enfermedad)=«CIE-10»

# Detalle de Métricas de Calidad

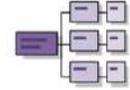
## Métrica: ValoresPosiblesPorExtensión



<ul style="list-style-type: none"> <li>Consistencia             <ul style="list-style-type: none"> <li>+ Integridad Inter-entidad</li> <li>+ Integridad Intra-entidad</li> <li>- Integridad de Dominio                 <ul style="list-style-type: none"> <li>ValoresPosiblesPorExtensión</li> <li>ValoresPosiblesPorComprensión</li> </ul> </li> <li>+ Consistencia Topológica [TD:«Geo»]</li> </ul> </li> </ul>	<b>Métrica Genérica: ValoresPosiblesPorExtensión</b>	
	<b>Semántica</b>	Indica si el valor de un atributo se encuentra dentro de un dominio definido por extensión.
	<b>Granularidad</b>	instanciaAtributo
	<b>Tipo de Resultado</b>	Boolean
	<b>Nombre Sugerido</b>	ValoresPosiblesPE(Atr)
	<b>Propiedades Configuración</b>	1. Atributo=Atr 2. DominioPorExtension(Atributo)
<b>Métrica Específica: ValoresPosiblesPE(Sexo, AGESIC)</b>		
<b>Semántica</b>	Indica si el valor de un atributo Sexo se encuentra dentro del conjunto de valores definidos en el Vocabulario de Persona de AGESIC ([AGE18])	
<b>Basada En</b>	ValoresPosiblesPorExtensión	
<b>Propiedades Configuración</b>	1. Atributo=Sexo 2. DominioPorExtension(Atributo)={{1-Masculino, 2-Femenino, 3-Desconocido, 4-Indeterminado, 9-NoAplica}}	
	53	

# Detalle de Métricas de Calidad

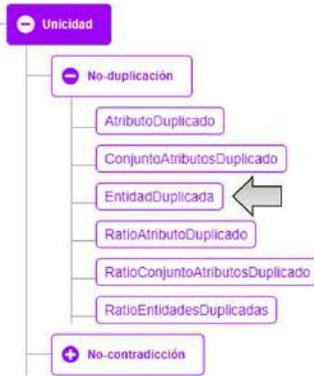
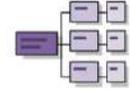
## Métrica: RatioCobertura



Métrica Genérica: RatioCobertura	
<b>Semántica</b>	Proporción entre la cantidad de instancias de una entidad y el número total de instancias de un referencial de esa entidad.
<b>Granularidad</b>	entidad
<b>Tipo de Resultado</b>	Intervalo real [0.0, 1.0].
<b>Nombre Sugerido</b>	RatioCobertura(Entidad, Referencial)
<b>Propiedades Configuración</b>	1. Entidad=Ent 2. Referencial(Entidad)

# Detalle de Métricas de Calidad

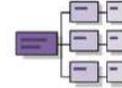
## Métrica: EntidadDuplicada



Métrica Genérica: EntidadDuplicada	
<b>Semántica</b>	Indica si existe, para una instancia de entidad, al menos otra instancia más que representa el mismo objeto del mundo real, con los mismos datos o algún dato faltante. Se debe especificar un conjunto de atributos que permita identificar unívocamente una instancia de entidad. Los demás atributos de la entidad que no pertenezcan a dicho conjunto deben tener los mismos valores en las dos instancias, o ser nulos en alguna de ellas, para que se consideren duplicados exactos.
<b>Granularidad</b>	instanciaEntidad
<b>Tipo de Resultado</b>	Boolean
<b>Nombre Sugerido</b>	EntidadDuplicada(Ent, ConjuntoAtributos)
<b>Propiedades Configuración</b>	1. Entidad=Ent 2. ConjuntoClave=(Atr1,...,AtrN)

# Detalle de Métricas de Calidad

## Métrica: DesactualizaciónPorFormato



Métrica Genérica: DesactualizaciónPorFormato	
<b>Semántica</b>	Indica si un dato se encuentra desactualizado en base a reglas sintácticas que establecen cuando un dato es actual y cuando no, en base al formato vigente de su tipo de dato.
<b>Granularidad</b>	instanciaAtributo
<b>Tipo de Resultado</b>	Boolean
<b>Nombre Sugerido</b>	DesactualizaciónPorFormato(Atr,Formato)
<b>Propiedades Configuración</b>	1. Atributo=Atr 2. Formato= Expresión regular o reglas que cumple formato vigente.

Métrica Específica: DesactualizaciónPorFormato(TelefonoFijo, FormatoPNN)	
<b>Semántica</b>	Indica si un número de telefonía fija se encuentra desactualizado en base al formato vigente del Plan Nacional de Numeración (PNN), que establece que a partir de 2010, todos los teléfono pasaron a tener 8 dígitos.
<b>Basada En</b>	DesactualizaciónPorFormato
<b>Propiedades Configuración</b>	1. Atr=TelefonoFijo 2. Formato=(Length(TelefonoFijo)==8)

56

## Referencias

---

**[AGE2020]** AGESIC. Marco de referencia para la gestión de calidad de datos. <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/marco-referencia-para-gestion-calidad-datos>

**[BS16]** Carlo Batini y Monica Scannapieco. Data and Information Quality. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-24106-7.

**[AGE18]** AGESIC. Vocabulario de Persona. Inf. téc. 1.0. Dic. de 2018. URL: <https://catalogodatos.gub.uy/dataset/agesic-vocabulario-de-persona> (visitado 20-12-2019).

**[ISO08]** ISO/IEC. ISO/IEC 25012:2008 - Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model. Estándar. International Organization for Standardization (ISO), dic. de 2008. URL: <https://www.iso.org/standard/35736.html>.

**[ISO13]** ISO. ISO 19157:2013 - Geographic information – Data quality. Estándar. International Organization for Standardization (ISO), dic. de 2013. URL: <https://www.iso.org/standard/32575.html>. 57