

Marco de trabajo 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].



Temas

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



Principales Componentes del Marco de trabajo



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

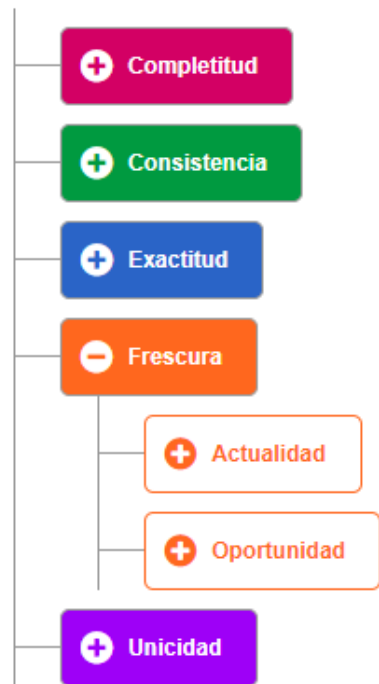
- ❑ Un elemento clave para la gestión de calidad de datos es contar con un **modelo de calidad de datos** que considere las características específicas del escenario de trabajo:
 - ❑ p. ej. requerimientos, problemas, prioridades
- ❑ El **Modelo de Calidad de Datos de Referencia (MCDR)** tiene como fin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



Descripción General del Modelo

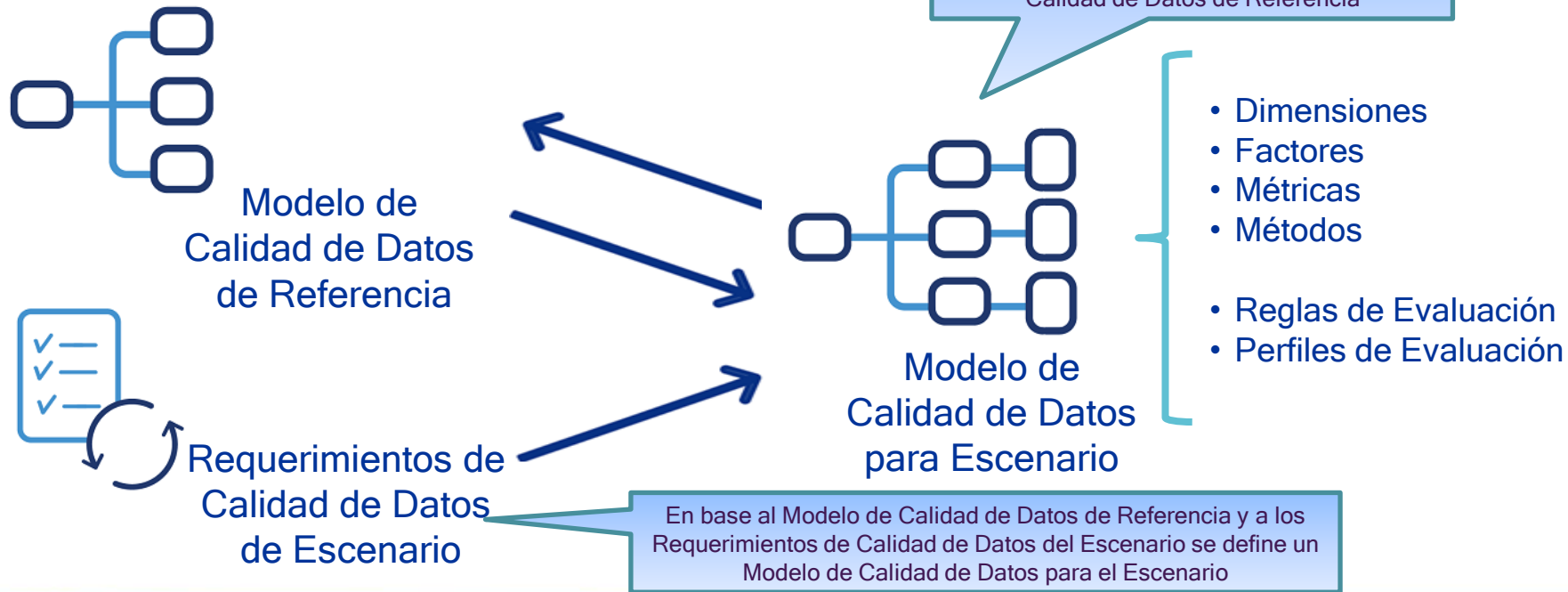
Introducción

- ❑ 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

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





Marco Conceptual Asociado

Introducción

- ❑ El **Modelo Conceptual** del *marco de trabajo* 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

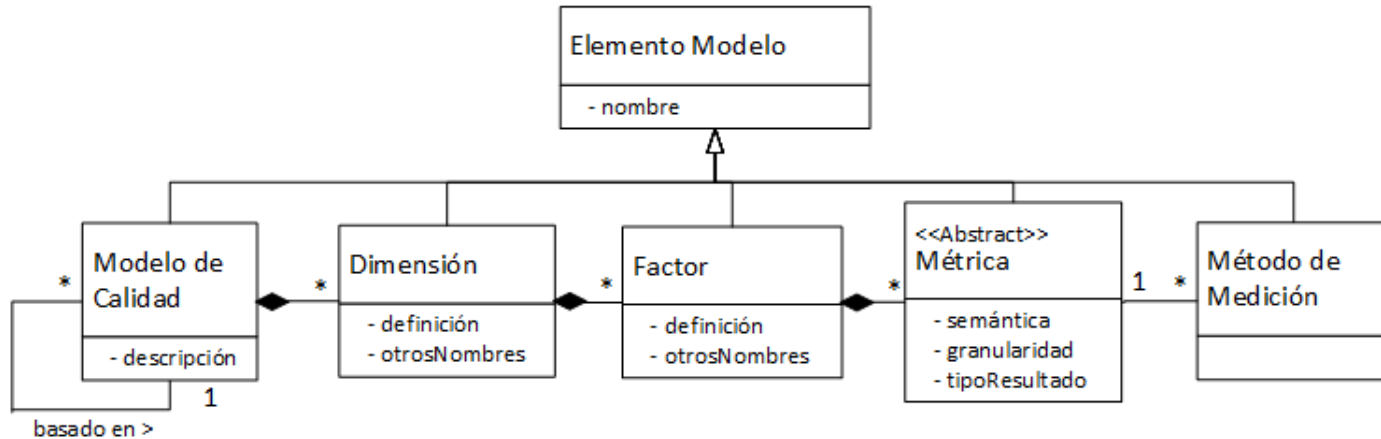




Marco Conceptual Asociado

Elementos Modelo

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





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.
- ❑ 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
- ❑ un **Método de Medición** es un proceso que implementa una métrica y se encarga de tomar medidas para la misma

exactitud

correctitud
sintáctica

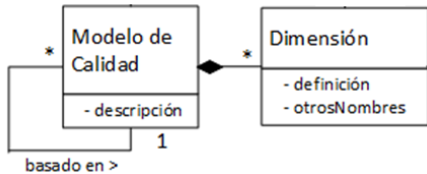
porcentaje de datos sintácticamente correctos





Marco Conceptual Asociado

Elementos Modelo - Ejemplos de Instancias

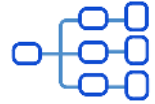


BASADO EN



Modelo de Calidad de Datos para Energía

BASADO EN

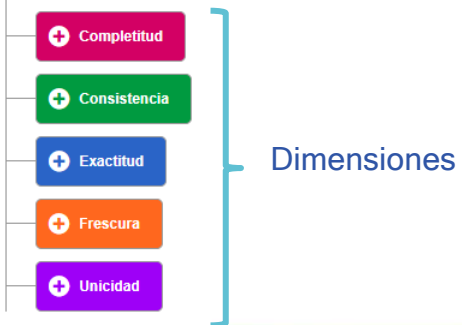


Modelo de Calidad de Datos para Escenario 2

BASADO EN



Modelo de Calidad de Datos para Escenario 4



Modelo de Calidad de Datos para Salud

BASADO EN



Modelo de Calidad de Datos para Escenario 1

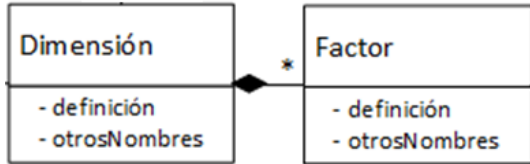




Marco Conceptual Asociado

Elementos Modelo - Ejemplo Dimensión

Especificación de la dimensión Unicidad para el modelo de referencia



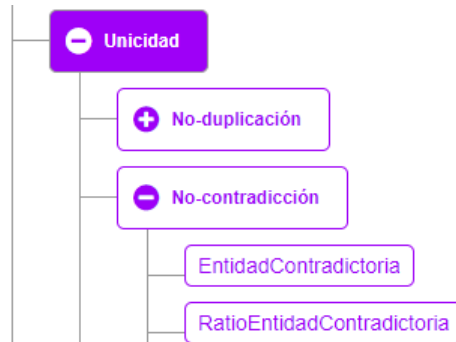
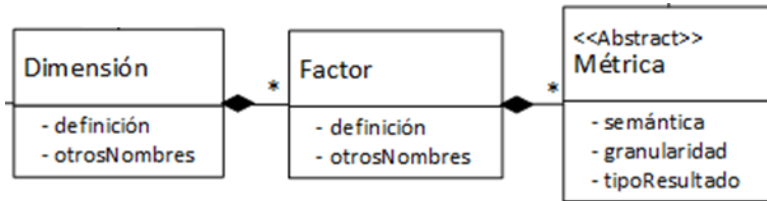
Unicidad	
Definición	Captura el grado en el que un dato del mundo real es representado en forma única. (Adaptado de [BS16])
Otros Nombres	Uniqueness (ingl.). En [BS16] se habla también de redundancia para el caso de linked data. La norma [ISO13] la considera como parte del factor Compleción y lo establece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



Marco Conceptual Asociado

Elementos Modelo - Ejemplo Factor

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



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

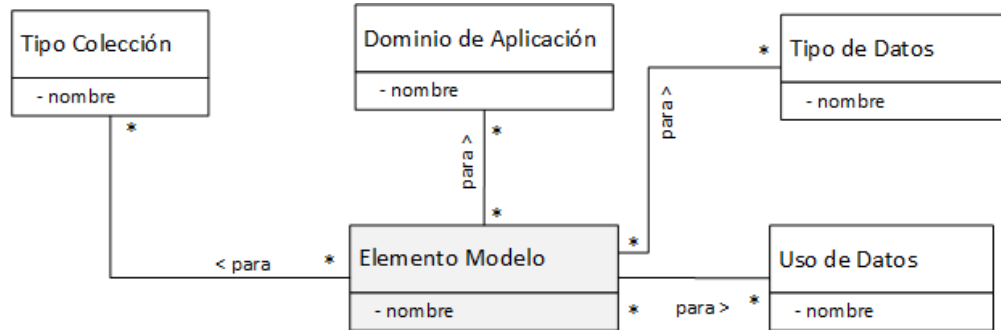
base de datos relacional, documental

dominios de aplicación

usos de datos

operativa, toma de decisiones

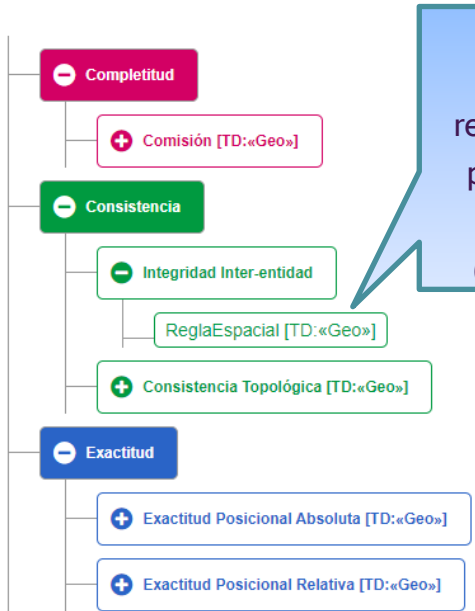
salud, transporte, energía



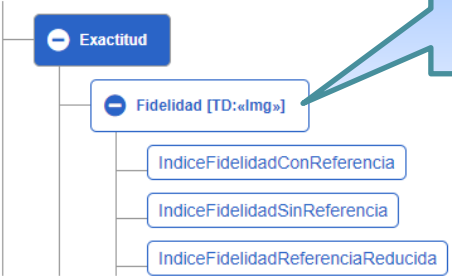


Marco Conceptual Asociado

Aplicabilidad de Elementos Modelo - Ejemplos



elementos del modelo de referencia filtrados por tipo de datos geográfico ([TD:«Geo»])



elementos del modelo de referencia filtrados por tipo de datos imagen ([TD:«Img»])

métrica definida para el dominio de aplicación Salud ([DA:«Salud»])

Integridad Intra-entidad	
Definición	Captura la satisfacción de reglas entre atributos de una misma entidad. (Adaptado de [BS16])
Otros nombres	Relation Integrity (ingl.)
Métricas	ReglaIntegridadIntraEntidad(Sexo,Enfermedad) [DA:«Salud»]. Indica si una enfermedad de una persona es compatible con el sexo de esa persona.



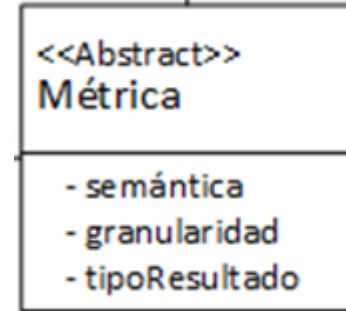


Marco Conceptual Asociado

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 especifica sobre qué se realiza la medición

Granularidad considerada en marco de trabajo	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

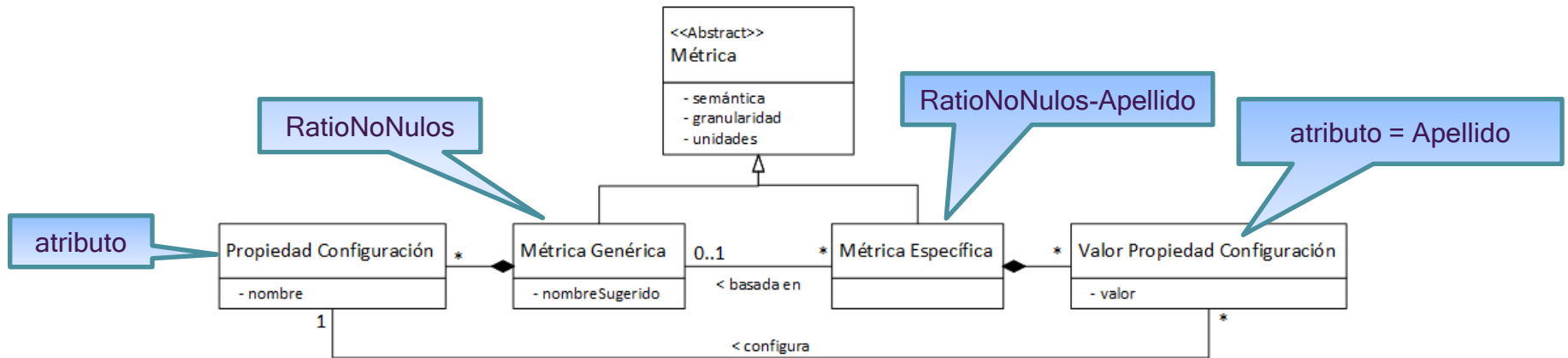




Marco Conceptual Asociado

Métricas Genéricas y Específicas

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

❑ Métricas Específicas

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

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

Formato(cédula, DNIC)	
Semántica	Indica si el valor del atributo numDoc cumple con el formato de cédula de identidad uruguaya establecido por DNIC.
Valor Propiedades	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	
Semántica	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.
Granularidad	instanciaAtributo
Tipo de Resultado	Boolean

❑ Métrica Agregada

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

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



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

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

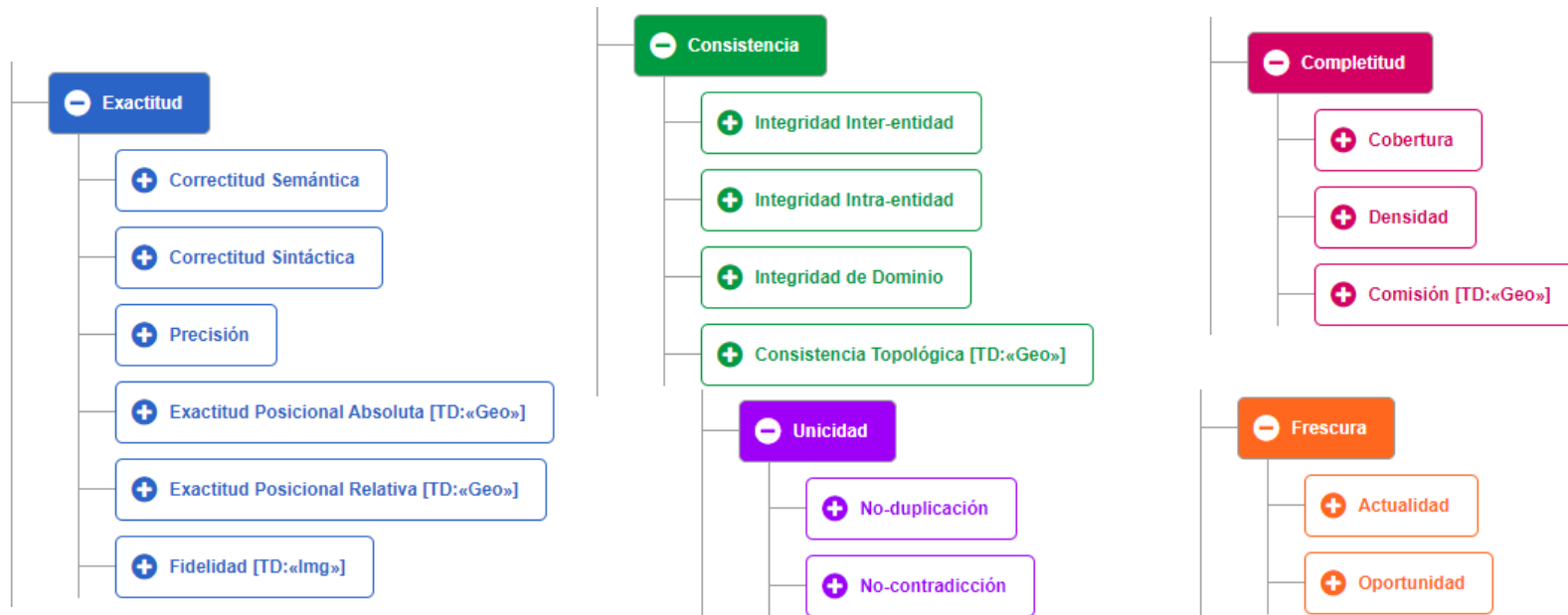




Dimensiones y Factores de Calidad

Introducción

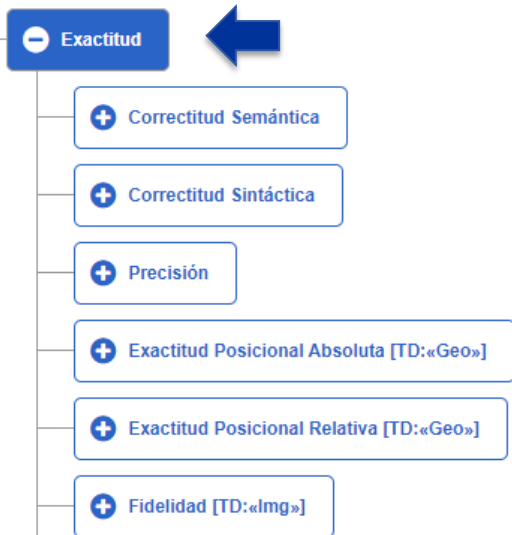
- El modelo define 5 dimensiones y 17 factores de calidad





Dimensiones y Factores de Calidad

Dimensión: Exactitud



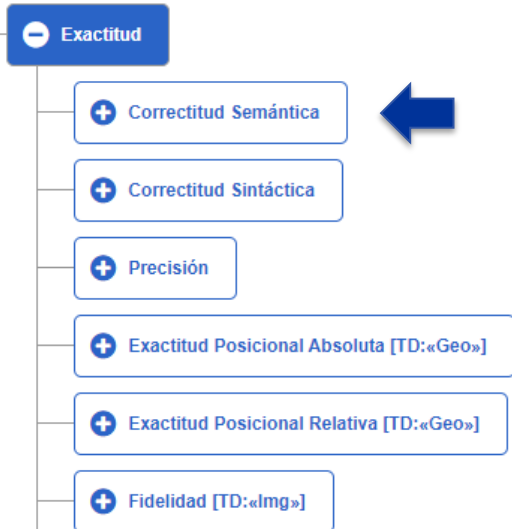
Exactitud	
Definición	Proximidad entre un valor de datos v y un valor de datos v' , considerado como la representación correcta del fenómeno del mundo real que v intenta representar. (Adaptado de [BS16])
Otros Nombres	Accuracy (ingl.), correctitud
Factores	Correctitud Semántica Correctitud Sintáctica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





Dimensiones y Factores de Calidad

Factor: Correctitud Semántica



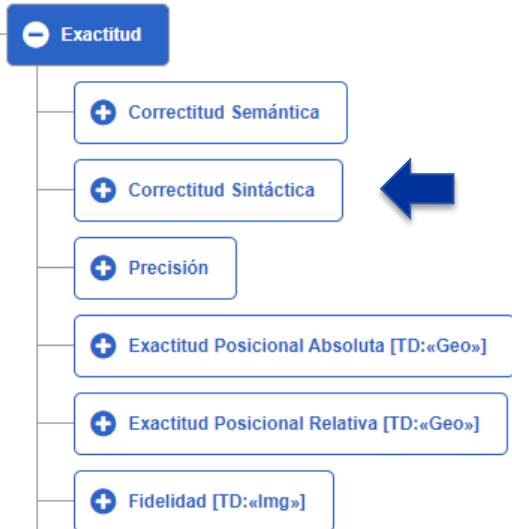
Correctitud Semántica	
Definición	Proximidad entre el valor v de un atributo y su verdadero valor v' . (Adaptado de [BS16])
Otros Nombres	Semantic Correctness (ingl.)
Métricas Genéricas	CorrectitudSemDebil CorrectitudSemFuerte RatioCorrectitudSemFuerte RatioCorrectitudSemDébil





Dimensiones y Factores de Calidad

Factor: Correctitud Sintáctica



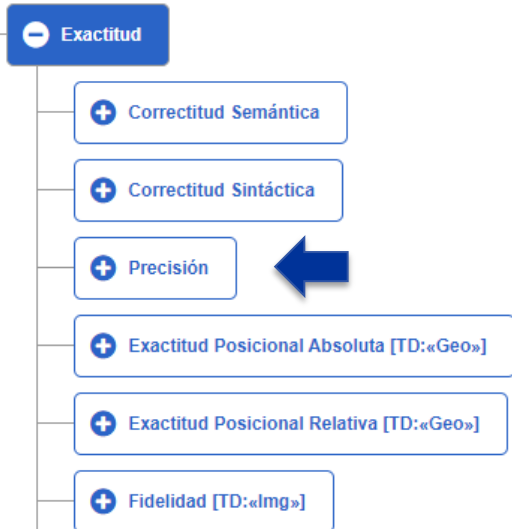
Correctitud Sintáctica	
Definición	Proximidad entre el valor v de un atributo y los elementos del dominio de definición de dicho atributo. (Adaptado de [BS16])
Otros Nombres	Syntactic Correctness (ingl.)
Métricas Genéricas	Formato
Métricas Específicas	Formato(Pais, ISOAlpha3) Formato(Enfermedad, CIE10)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





Dimensiones y Factores de Calidad

Factor: Precisión



Precisión	
Definición	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])
Otros Nombres	Precision (ingl.)
Métricas Genéricas	Escala ErrorEstandar



Dimensiones y Factores de Calidad

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

- Exactitud

+ Correctitud Semántica

+ Correctitud Sintáctica

+ Precisión

+ Exactitud Posicional Absoluta [TD:«Geo»]

+ Exactitud Posicional Relativa [TD:«Geo»]

+ Fidelidad [TD:«lmg»]



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





Dimensiones y Factores de Calidad

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

- Exactitud

+ Correctitud Semántica

+ Correctitud Sintáctica

+ Precisión

+ Exactitud Posicional Absoluta [TD:«Geo»]

+ Exactitud Posicional Relativa [TD:«Geo»]

+ Fidelidad [TD:«lmg»]



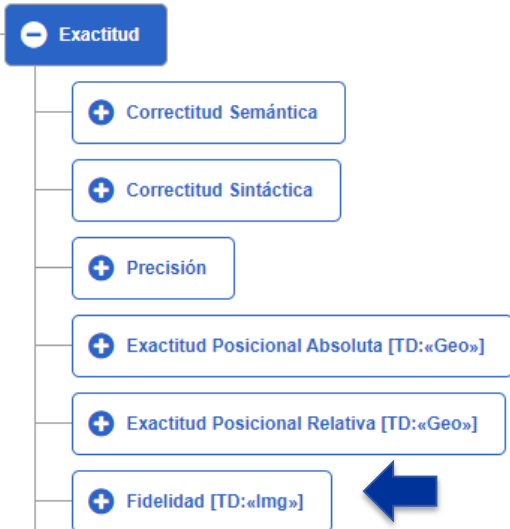
Exactitud Posicional Relativa (TD:«Geo»)	
Definición	Proximidad de las posiciones relativas de los objetos geográficos de un conjunto de datos a sus respectivas posiciones relativas verdaderas o aceptadas como tales [ISO13]
Otros Nombres	Relative positional accuracy (ingl.), Exactitud posicional interna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





Dimensiones y Factores de Calidad

Factor: Fidelidad [TD:«lmg»]

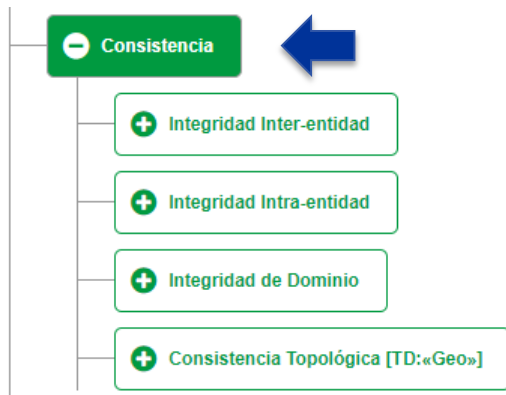


Fidelidad (TD:«lmg»)	
Definición	Captura el grado de concordancia aparente de una imagen con el original. [BS16]
Otros Nombres	Fidelity (ingl.), faithfulness (ingl.), genuineness (ingl.)
Métricas Genéricas	IndiceFidelidadConReferencia IndiceFidelidadSinReferencia IndiceFidelidadReferenciaReducida



Dimensiones y Factores de Calidad

Dimensión: Consistencia



Consistencia	
Definición	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])
Otros Nombres	Consistency (ingl.), cohesión, coherencia
Factores	Integridad Inter-entidad Integridad Intra-entidad Integridad de Dominio Consistencia Topológica (TD:«Geo»)





Dimensiones y Factores de Calidad

Factor: Integridad Inter-entidad



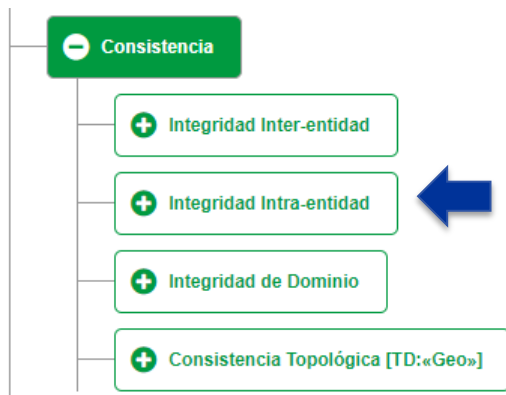
Integridad Inter-entidad	
Definición	Captura la satisfacción de reglas entre atributos de diferentes entidades de negocio. (Adaptado de [BS16])
Otros Nombres	Referential Integrity (ingl.), Integridad Inter-relacion
Métricas Genéricas	ReglaIntegridadInterEntidad ReglaEspacial [TD:«Geo»]
Métricas Específicas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





Dimensiones y Factores de Calidad

Factor: Integridad Intra-entidad



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





Dimensiones y Factores de Calidad

Factor: Integridad de Dominio



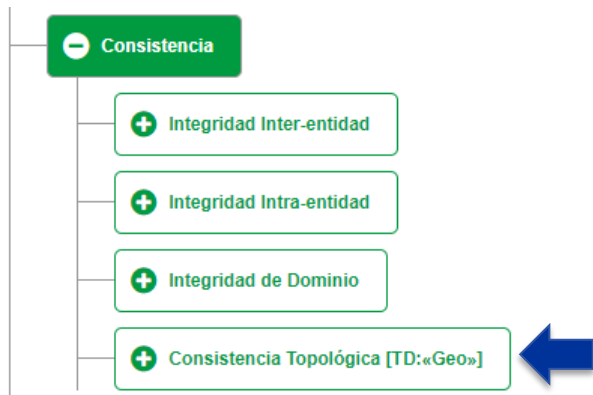
Integridad de Dominio	
Definición	Captura la satisfacción de reglas sobre los valores posibles que puede tomar un atributo. (Adaptado de [BS16])
Otros Nombres	Domain Integrity (ingl.)
Métricas Genéricas	ValoresPosiblesPorExtensión ValoresPosiblesPorComprensión
Métricas Específicas	ValoresPosiblesPE(Sexo,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





Dimensiones y Factores de Calidad

Factor: Consistencia Topológica [TD:«Geo»]



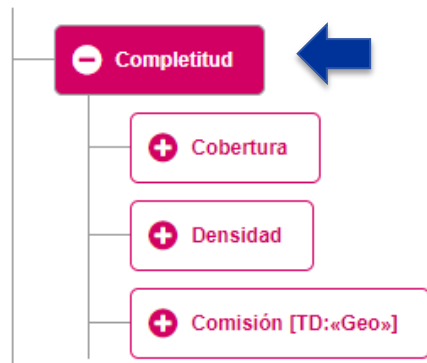
Consistencia Topológica (TD:«Geo»)	
Definición	Corrección de las características topológicas codificadas explícitamente. Las características topológicas de un conjunto de datos describen las relaciones geométricas entre los ítems del conjunto de datos que no son alteradas por transformaciones elásticas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





Dimensiones y Factores de Calidad

Dimensión: Completitud



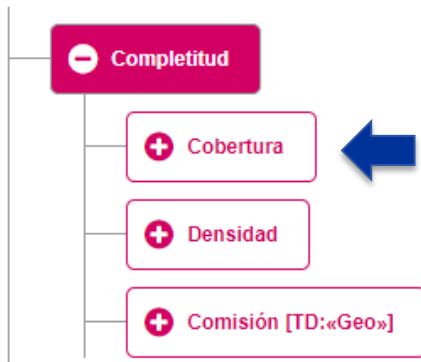
Compleitud	
Definición	Captura la medida en que los datos son de la amplitud, profundidad y alcance suficientes para una determinada tarea. (Adaptado de [BS16])
Otros Nombres	Completeness (ingl.)
Factores	Cobertura Densidad Comisión (TD:«Geo»)





Dimensiones y Factores de Calidad

Factor: Cobertura



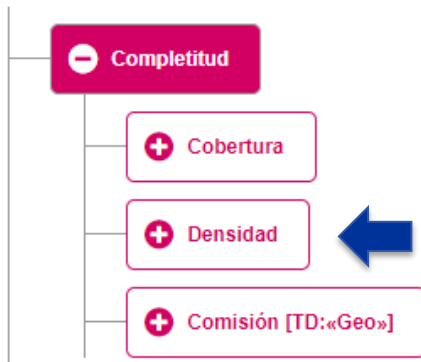
Cobertura	
Definición	Captura la proporción entre la cantidad de entidades existentes en una determinada colección de datos, y el total de entidades que deberían existir en dicha colección. La cobertura varía si se utiliza la Asunción de Mundo Cerrado, según la cual una colección de datos debería contener 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])
Otros Nombres	Coverage (ingl.)
Métricas Genéricas	RatioCobertura





Dimensiones y Factores de Calidad

Factor: Densidad

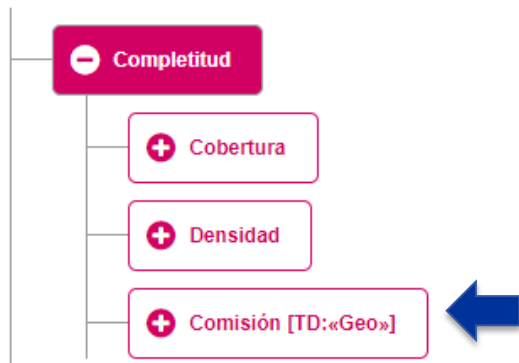


Densidad	
Definición	Captura la proporción entre la cantidad de instancias de atributo con valores no nulos y el total de instancias de dicho atributo (Adaptado de [BS16]). Un valor nulo de una instancia de atributo A de una entidad E puede interpretarse de varias maneras: <ol style="list-style-type: none">1. E no posee A2. se desconoce si E posee A o no3. E posee A pero se desconoce su valor
Otros Nombres	Density (ingl.)
Métricas Genéricas	NoNulo DensidadPonderada RatioNoNulos RatioDensidadPonderada



Dimensiones y Factores de Calidad

Factor: Comisión [TD:<<Geo>>]

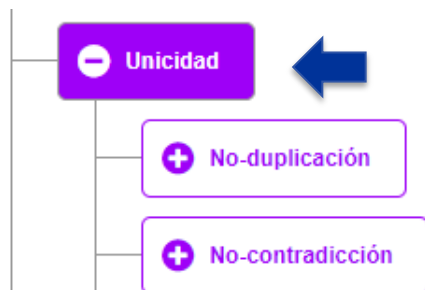


Comisión (TD:<<Geo>>)	
Definición	Datos excedentes presentes en un conjunto de datos [ISO13]
Otros Nombres	Commission (ingl.)
Métricas Genéricas	ÍtemExcedente ÍndiceltemsExcedentes



Dimensiones y Factores de Calidad

Dimensión: Unicidad



Unicidad	
Definición	Captura el grado en el que un dato del mundo real es representado en forma única. (Adaptado de [BS16])
Otros Nombres	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.
Factores	No-duplicación No-contradicción



Dimensiones y Factores de Calidad

Factor: No-duplicación



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





Dimensiones y Factores de Calidad

Factor: No-contradicción



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



Dimensiones y Factores de Calidad

Dimensión: Frescura



Frescura	
Definición	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])
Otros Nombres	Freshness (ingl.), exactitud temporal
Factores	Actualidad Oportunidad



Dimensiones y Factores de Calidad

Factor: Actualidad



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





Dimensiones y Factores de Calidad

Factor: Oportunidad

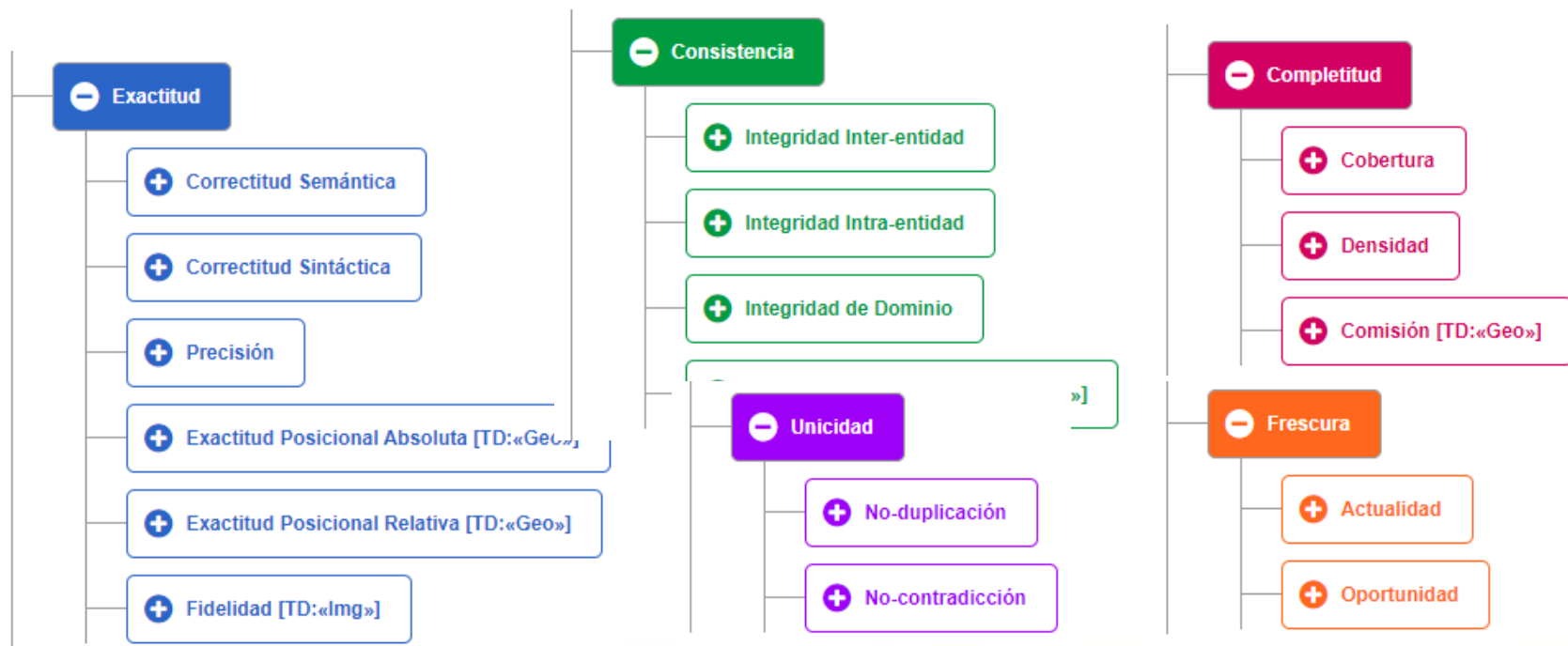


Oportunidad	
Definición	Captura la demora que existe entre la actualización de un dato y el momento en el que éste se encuentra disponible para ser utilizado.
Otros Nombres	Timeliness (ingl.)
Métricas Genéricas	OportunidadAtributoPorFecha OportunidadAtributoPorIntervalo OportunidadEntPorFecha OportunidadEntPorIntervalo:



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 (Complejidad - Cobertura)
 - ❑ EntidadDuplicada (Unicidad - No-duplicación)
 - ❑ DesactualizaciónPorFormato (Frescura - Actualidad)





Detalle de Métricas de Calidad

Métrica: Formato

Exactitud

+ Correctitud Semántica

- Correctitud Sintáctica

Formato

+ Precisión

+ Exactitud Posicional Absoluta [TD:«Geo»]

+ Exactitud Posicional Relativa [TD:«Geo»]

+ Fidelidad [TD:«lmg»]

Métrica Genérica: Formato

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

Métrica Específica: Formato(Enfermedad, CIE10) [DA:«Salud»]

Semántica	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).
Basada En	Formato
Propiedades Configuración	1. Atr=Enfermedad 2. Estandar(Enfermedad)=«CIE-10»





Detalle de Métricas de Calidad

Métrica: ValoresPosiblesPorExtensión

Métrica Genérica: ValoresPosiblesPorExtensión	
Semántica	Indica si el valor de un atributo se encuentra dentro de un dominio definido por extensión.
Granularidad	instanciaAtributo
Tipo de Resultado	Boolean
Nombre Sugerido	ValoresPosiblesPE(Atr)
Propiedades Configuración	1. Atributo=Atr 2. DominioPorExtension(Atributo)

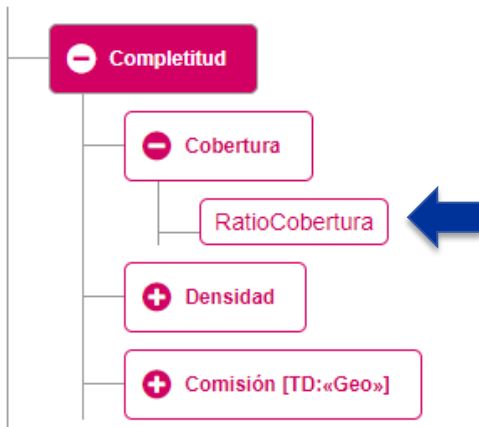
Métrica Específica: ValoresPosiblesPE(Sexo, AGESIC)	
Semántica	Indica si el valor de un atributo Sexo se encuentra dentro del conjunto de valores definidos en el Vocabulario de Persona de AGESIC ([AGE18])
Basada En	ValoresPosiblesPorExtensión
Propiedades Configuración	1. Atributo=Sexo 2. DominioPorExtension(Atributo)={{1-Masculino, 2-Femenino, 3-Desconocido, 4-Indeterminado, 9-NoAplica}}





Detalle de Métricas de Calidad

Métrica: RatioCobertura



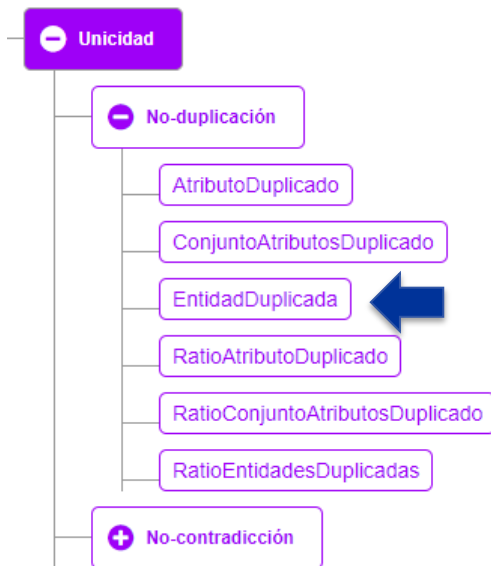
Métrica Genérica: RatioCobertura	
Semántica	Proporción entre la cantidad de instancias de una entidad y el número total de instancias de un referencial de esa entidad.
Granularidad	entidad
Tipo de Resultado	Intervalo real [0.0, 1.0].
Nombre Sugerido	RatioCobertura(Entidad, Referencial)
Propiedades Configuración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





Detalle de Métricas de Calidad

Métrica: EntidadDuplicada



Métrica Genérica: EntidadDuplicada	
Semántica	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.
Granularidad	instanciaEntidad
Tipo de Resultado	Boolean
Nombre Sugerido	EntidadDuplicada(Ent, ConjuntoAtributos)
Propiedades Configuración	1. Entidad=Ent 2. ConjuntoClave=(Atr1,...,AtrN)





Detalle de Métricas de Calidad

Métrica: DesactualizaciónPorFormato

− Frescura

− Actualidad

DesactualizaciónPorFecha

DesactualizaciónPorCambios

DesactualizaciónPorFormato

+ Oportunidad



Métrica Genérica: DesactualizaciónPorFormato	
Semántica	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.
Granularidad	instanciaAtributo
Tipo de Resultado	Boolean
Nombre Sugerido	DesactualizaciónPorFormato(Atr,Formato)
Propiedades Configuración	1. Atributo=Atr 2. Formato= Expresión regular o reglas que cumple formato vigente.

Métrica Específica: DesactualizaciónPorFormato(TelefonoFijo, FormatoPNN)	
Semántica	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.
Basada En	DesactualizaciónPorFormato
Propiedades Configuración	1. Atr=TelefonoFijo 2. Formato=(Length(TelefonoFijo)==8)



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>.



Uruguay
Presidencia

<>agesic

www.gub.uy/agesic

