

Seminario: "Hacia una gestión institucional mediante estándares de información geográfica"

ISO 19157: Calidad de datos y su relación con las normas de la familia ISO 19100

Elena Chicaiza M.

elena.chicaiza@ipgh.gob.ec

 [@gachisbar](https://twitter.com/gachisbar)



UNIVERSIDAD
DE SANTIAGO
DE CHILE



Objetivos

- Conocer el alcance de la norma ISO 19157
- Conocer su vinculación con las normas de la familia ISO 19100
- Conocer cómo se trata el proceso de evaluación en ISO 19157
- Conocer las medidas de calidad y su importancia en la estandarización.
- Conocer las opciones de informar
- Tener una visión crítica de ISO 19157

La Norma ISO 9000:2015 define que: “ *...la calidad está determinada por la capacidad para satisfacer a los clientes, y por el impacto **previsto y el no previsto** sobre las partes interesadas pertinentes. La calidad de los productos y servicios incluye no solo su función y desempeño previstos, sino también su valor percibido y el beneficio para el cliente”*

**Partes interesadas pertinentes o
“stakeholders”**

Resultados de la calidad → equilibrados en todos sus grupos de interés

- Quién
- Requisitos
- Necesidades
- Expectativas
- Valor
- Beneficio para el cliente.



Aptitud para el uso

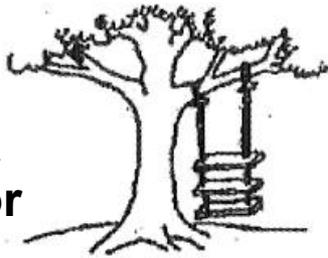
Definir y diseñar las métricas empleadas en la calidad para los diferentes usos; un mismo producto

{ **Útil**
Inadecuado

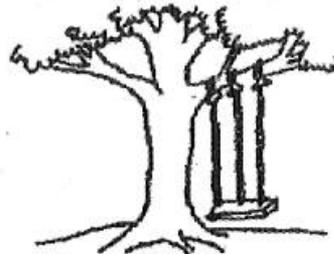


(Floridi e Illari, 2014)

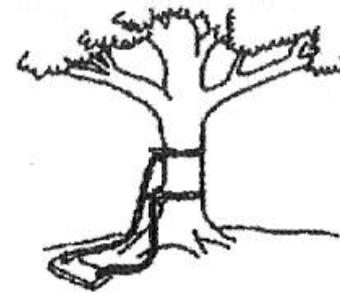
Perspectiva del productor



As requested by the manager



As described in the specifications

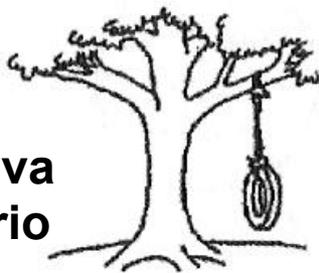


As designed by the analyst

Exactitud posicional, temática, temporal, etc



Perspectiva del Usuario



What the user wanted



As produced

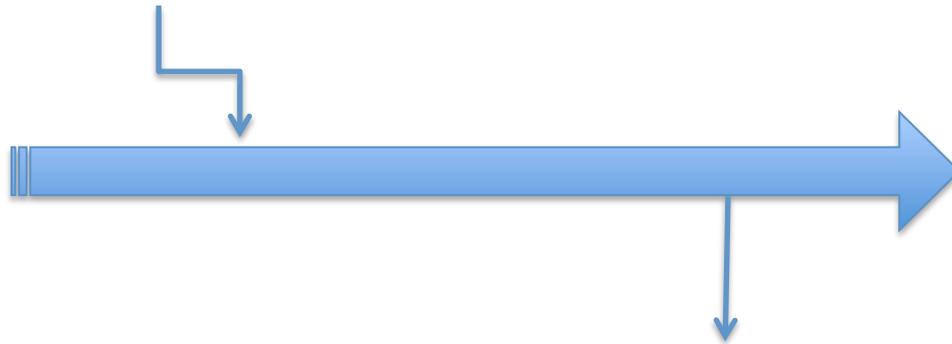
"Aptitud para el uso"

INTERNAL QUALITY

HAY un problema de comunicación:

- Descripciones detalladas
- Especificaciones cuantitativas y cualitativas

Surge en la academia durante los años 80 y principio de los 90 → **control posicional** únicamente orientado al producto



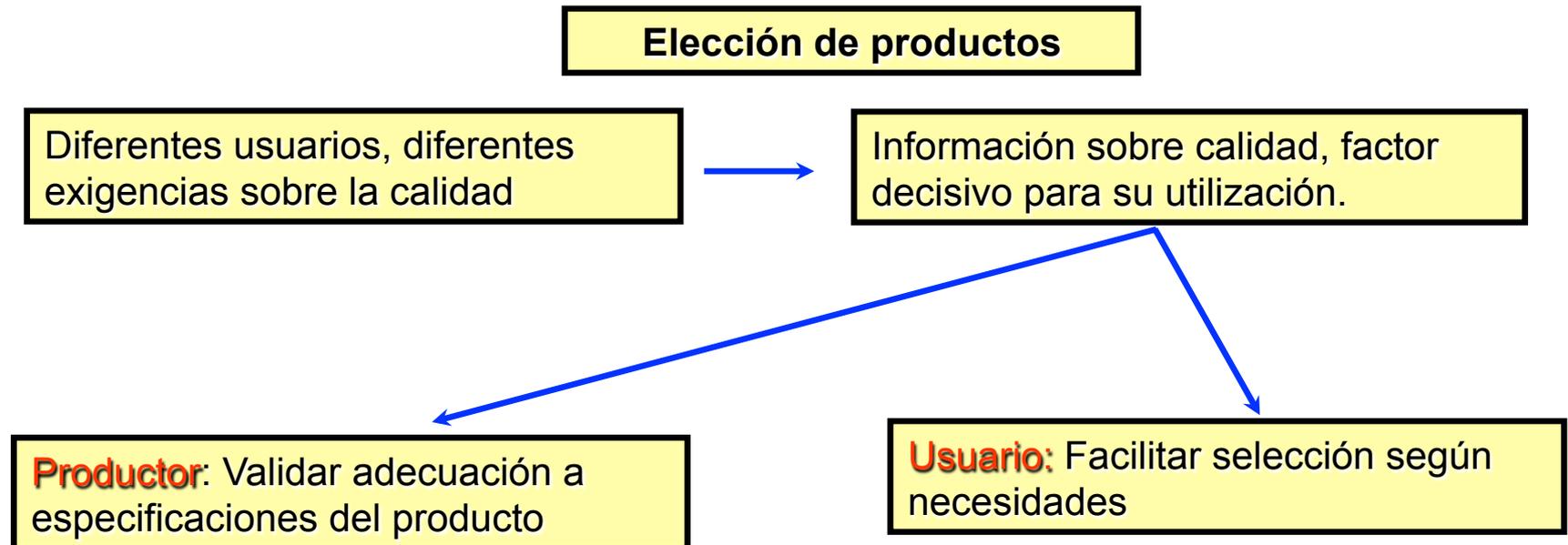
consideración de la **metacalidad** y la **usabilidad** como elementos de la calidad.

*La tendencia de crecimiento continuo de generación y uso de la IG requiere especial atención en cuanto a la normalización y calidad, como **claves de éxito** (Ariza-López et al., 2013) al momento de la toma de decisiones basadas en IG.*

La importancia de la Normalización

Organización	Número de casos	%
FGDC	40	20.3
ISO	55	27.9
MIL	83	42.1
OGC	19	9.6
Total	197	100

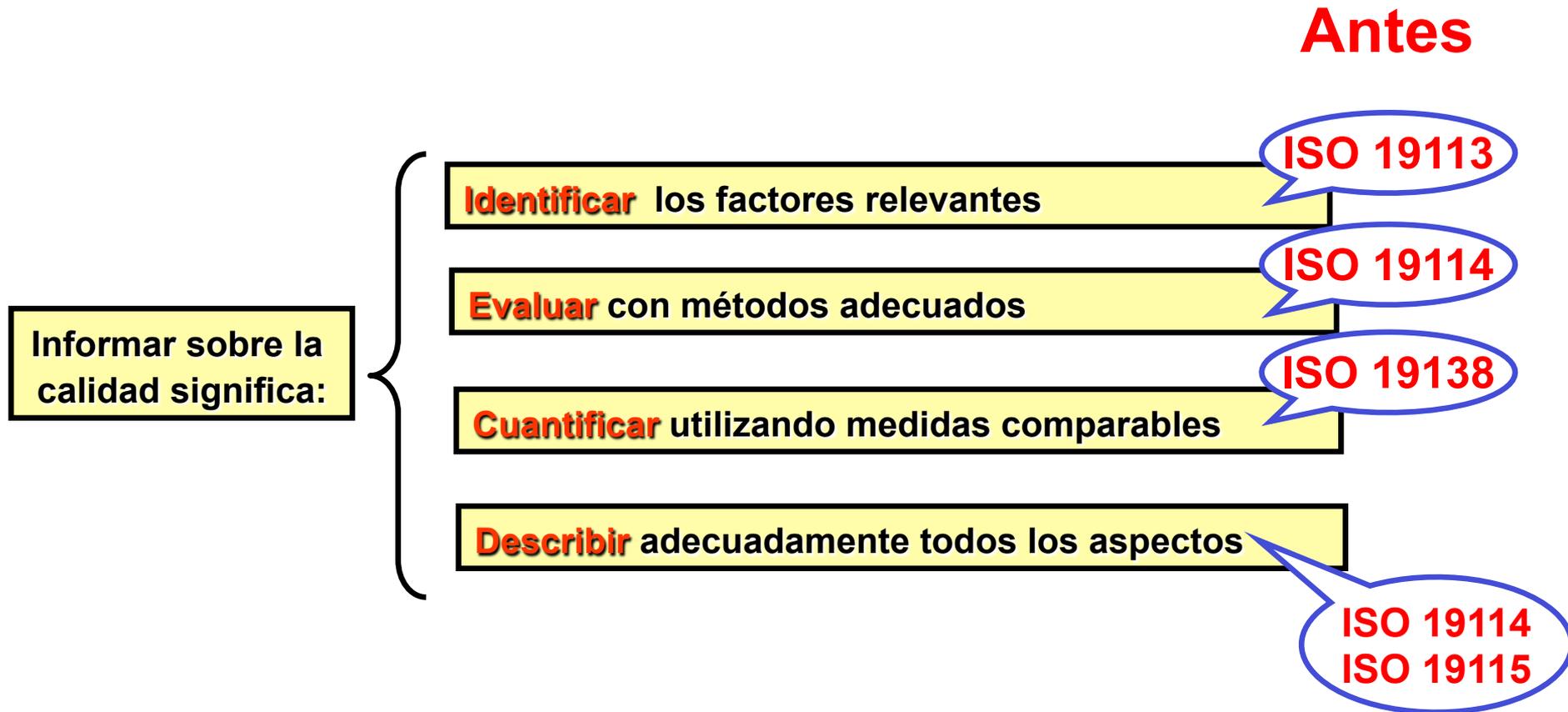
Número de normas desarrolladas relativas a la IG por cada institución. Tomado de Ariza-López (2011)



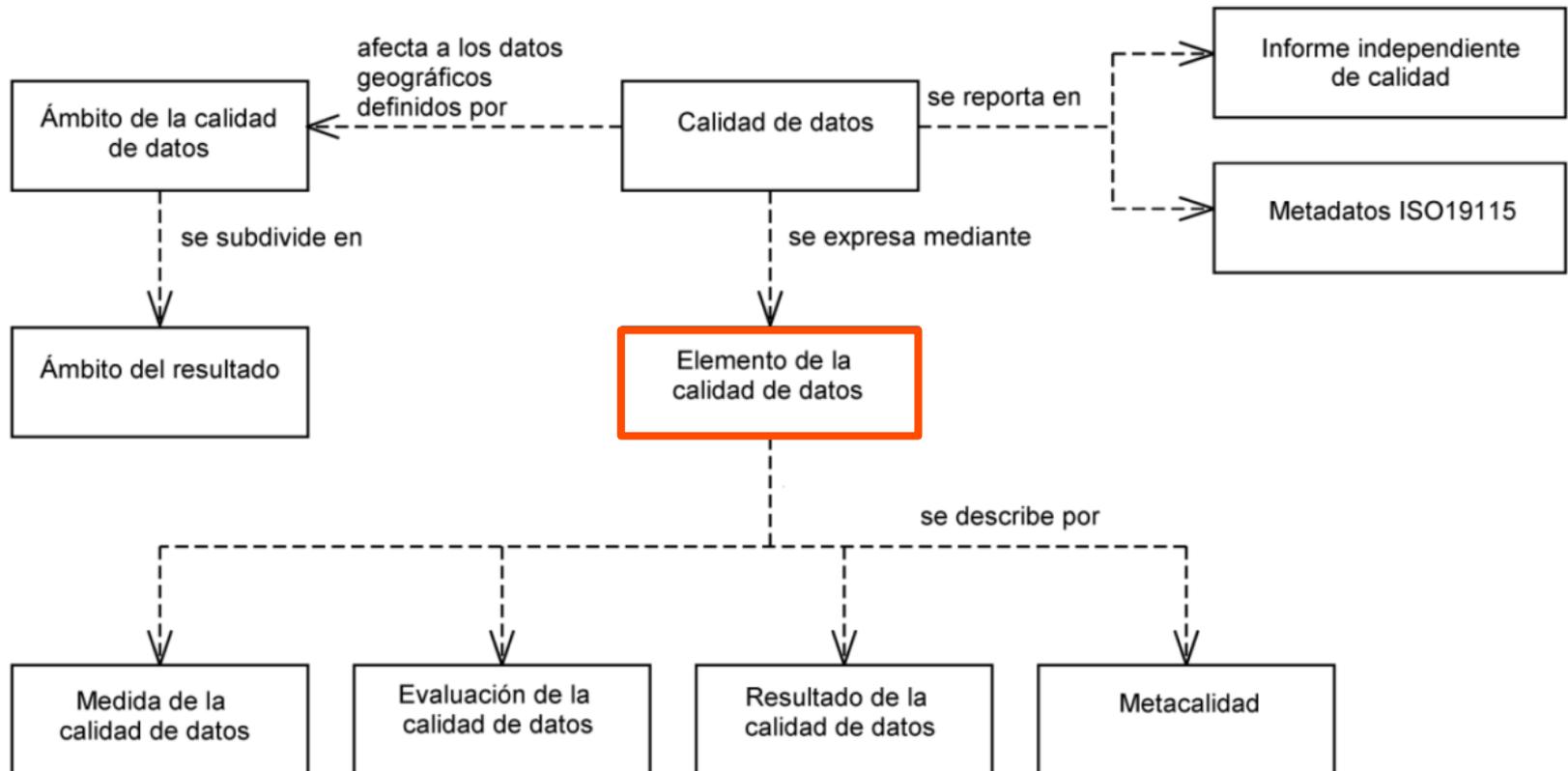
OBJETIVOS:

- Dar transparencia al mercado
- Favorecer la buena elección y uso de los productos.
- Incorporar la voz del cliente.
- Competitividad.

ISO 19157 – Calidad de datos



Visión general

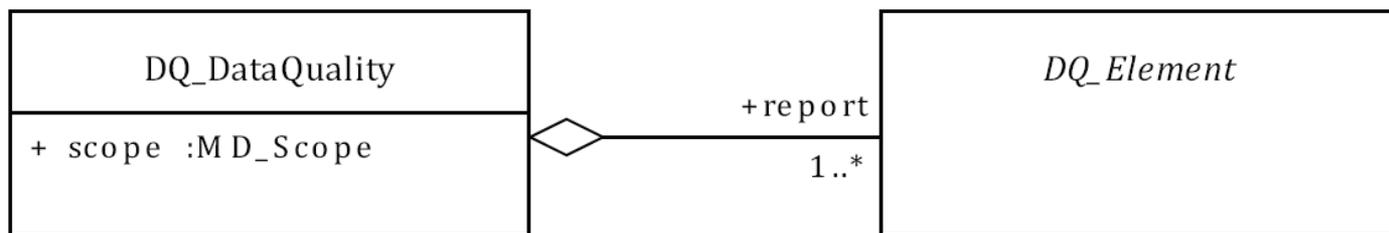


ISO 19157

Unidad de calidad de datos

Data Quality Unit (DQU)

- Novedad de ISO 19157.
- Establece una relación entre un elemento de la calidad y un ámbito.
- SIEMPRE se evalúa e informa sobre un DQU concreto.
- Se debe identificar, al menos, un ámbito de la calidad por cada elemento aplicable.
- El ámbito especifica la extensión espacial, temporal, lógica...



Idea

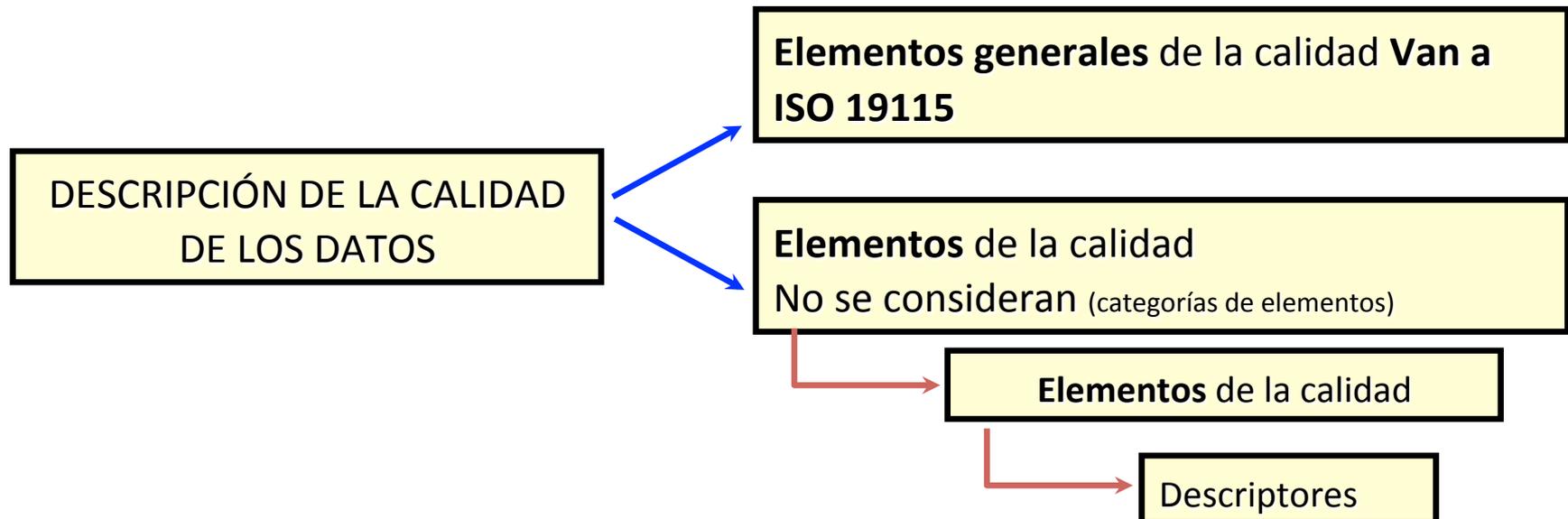
En un CDE en la que se aplica una metodología de evaluación dada,
Tal que el **resultado** se expresa según una medida dada.
El **resultado** puede ser interpretado de manera muy distinta, si el
aspecto de interés no está claramente definido

→ **Ambigüedad**

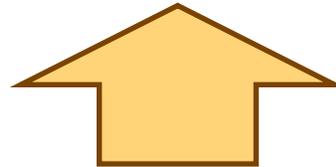
ISO 19157 – Calidad de datos

Objetivo

Establecer los **principios para describir** la calidad y la manera de informar sobre ella. NO establece niveles de calidad aceptables.



Con ISO 19157 ahora van a metadatos (ISO 19115)



ELEMENTOS GENERALES

Propósito

Describe la justificación para la creación del conjunto de datos y contiene información sobre su utilización prevista.

Uso

Describe las aplicaciones para las que se ha empleado un conjunto de datos, tanto del productor como de otros usuarios.

Linaje

Describe la historia del conjunto de datos, referir el ciclo de vida, desde su captura y adquisición, pasando por su compilación y derivación, hasta su forma actual.

Categorías de elementos o “componentes”

Compleción	Presencia y ausencia de fenómenos, sus atributos y relaciones.
Consistencia lógica	Grado de adherencia a las reglas lógicas de la estructura de los datos, atributos y relaciones.
Exactitud posicional	Exactitud de la posición de los fenómenos.
Calidad temporal	Exactitud de los atributos temporales y de las relaciones temporales de los fenómenos.
Exactitud temática	Exactitud de atributos cuantitativos y corrección de no cuantitativos, y de las clasificaciones de fenómenos y sus relaciones.
Usabilidad	Cumplimiento de unos requisitos.

Categoría

Compleción

Consistencia lógica

Exactitud posicional

Calidad temporal

Exactitud temática

Elementos

- **Comisión:** Datos excedentes.
- **Omisión:** Datos ausentes.

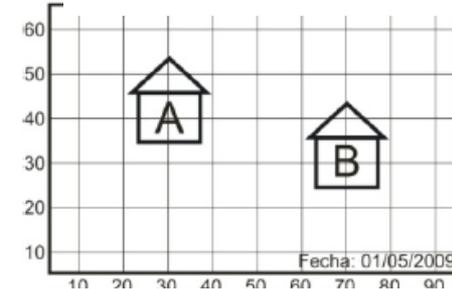
Categoría

Compleción

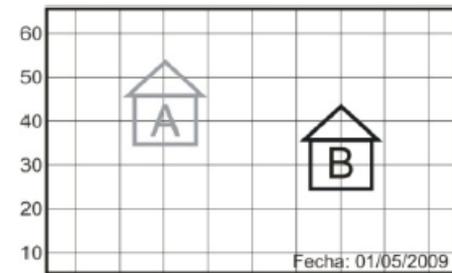
- Se requieren especificaciones muy buenas.
- Se requiere personal entrenado y métodos bien especificados.
- Métodos directos, externos.
- Necesidad de trabajo de campo (muestreo).
- Variables de conteo.
- Se determina la presencia/ausencia.
- La omisión $T_n \rightarrow$ Producto
- La comisión Producto $\rightarrow T_n$
- Muy afectada por el tiempo.

De objetos (C,O)
De atributos (O)

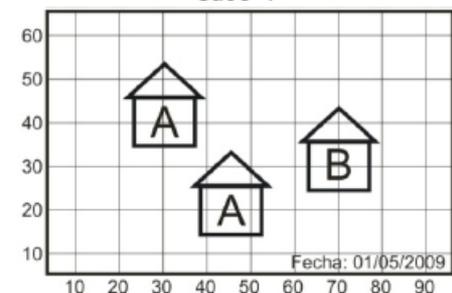
Terreno nominal



Caso 3



Caso 4



Categoría

Compleción

Consistencia lógica

Exactitud posicional

Calidad temporal

Exactitud temática

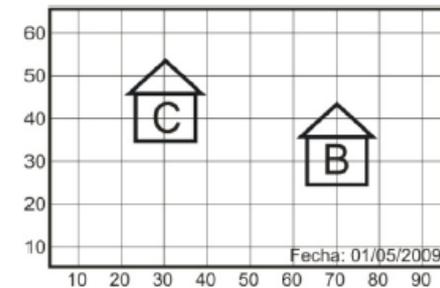
Elementos

- **Consistencia conceptual:** Adherencia a las reglas del modelo conceptual.
- **Consistencia de dominio:** Adherencia de los valores a su dominio.
- **Consistencia de formato:** Grado en que los datos se almacenan de acuerdo con la estructura física.
- **Consistencia topológica:** Corrección de las características topológicas codificadas explícitamente.

Categoría

Consistencia lógica

- Se requieren especificaciones explícitas (modelo).
- Se requieren rutinas verificadas.
- Métodos directos, internos.
- La inspección 100% es posible.
- Se trabaja con la base de datos.
- Variables de conteo.
- Se determina la presencia de fallos.
- Las rutinas avisan de los fallos.
- Algunos fallos se corrigen automáticamente otros no.



Categoría

Consistencia lógica

CL Dominio

Ejemplo enumerada:

C = Cereal

CT = Trigo

CC = Cebada

CM = Maíz

F = Frutal

FC = Cítricos

....
 Todo valor que no está en la lista es una violación

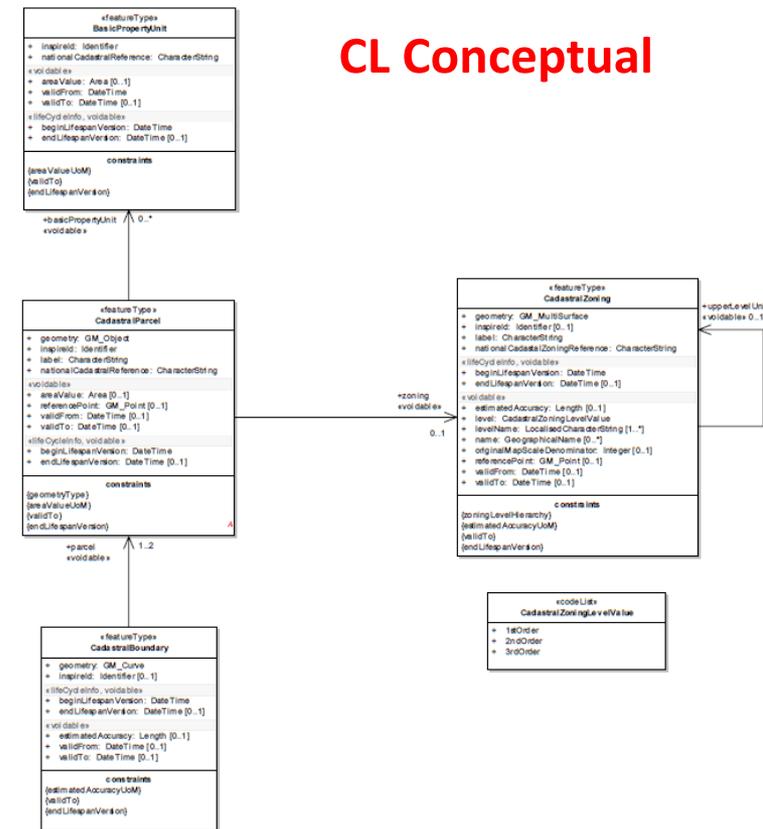
Ejemplo intervalos:

Valores enteros [0, 100]

Valores reales [0.5, 2,85]

Todo valor fuera de los intervalos es una violación

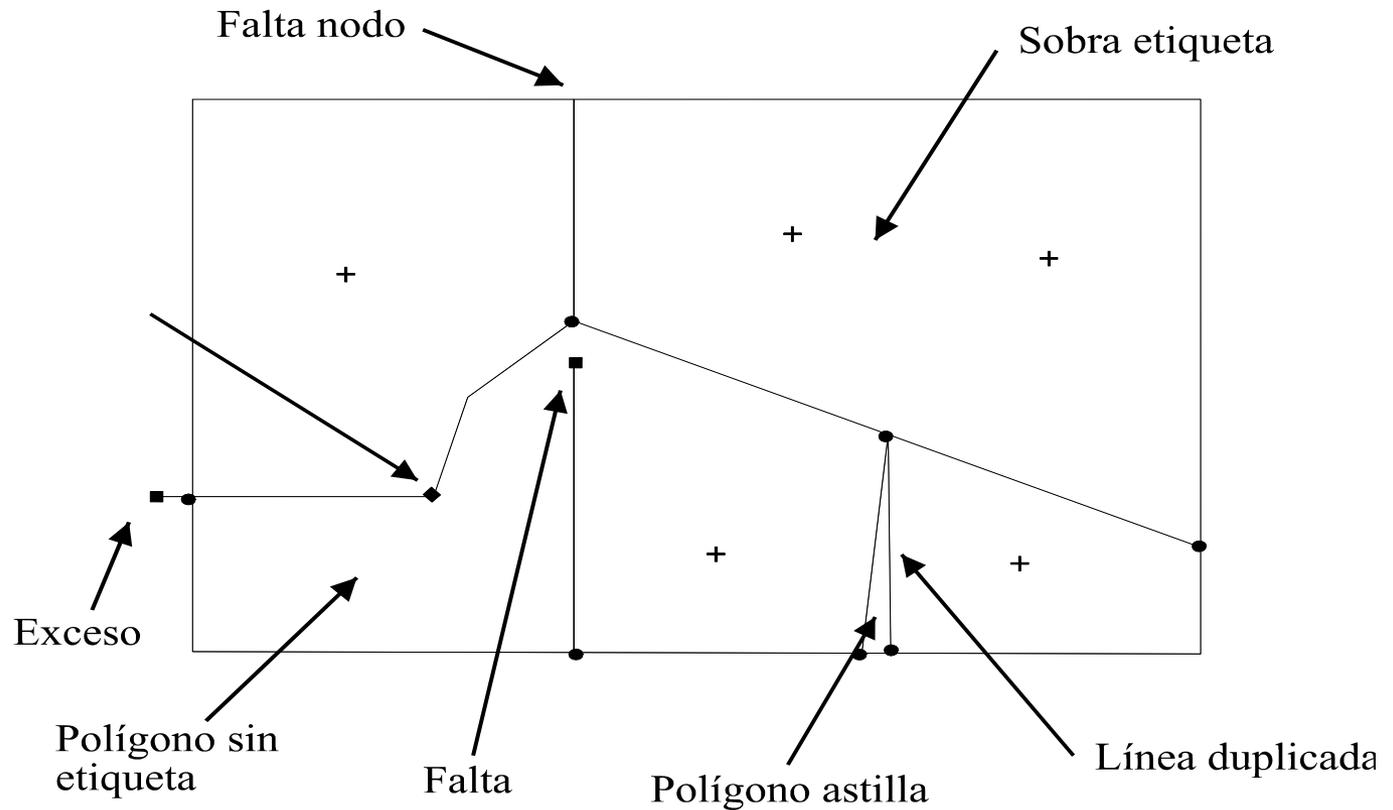
CL Conceptual



Categoría

Consistencia lógica

CL Topológica



Categoría

Compleción

Consistencia lógica

Exactitud posicional

Calidad temporal

Exactitud temática

Elementos

- **Exactitud absoluta o externa:** Proximidad entre los valores de coordenadas reportados y los valores verdaderos o aceptados como tales.
- **Exactitud relativa o interna:** Proximidad entre las posiciones relativas de los fenómenos y sus respectivas posiciones relativas verdaderas o aceptadas como tales.
- **Exactitud posicional de datos en malla:** Proximidad de los valores de posición de los datos en estructura de malla regular a los valores verdaderos o aceptados como tales.

ERROR e INCERTIDUMBRE son dos conceptos clave.

Error: Es la diferencia entre un valor medido y un valor de referencia (valor convencional o valor verdadero) [VIM 2007]

Incertidumbre: La discrepancia entre dos valores que se supone debían ser iguales.

$$e_{x_i} = x_{t_i} - x_{m_i}$$

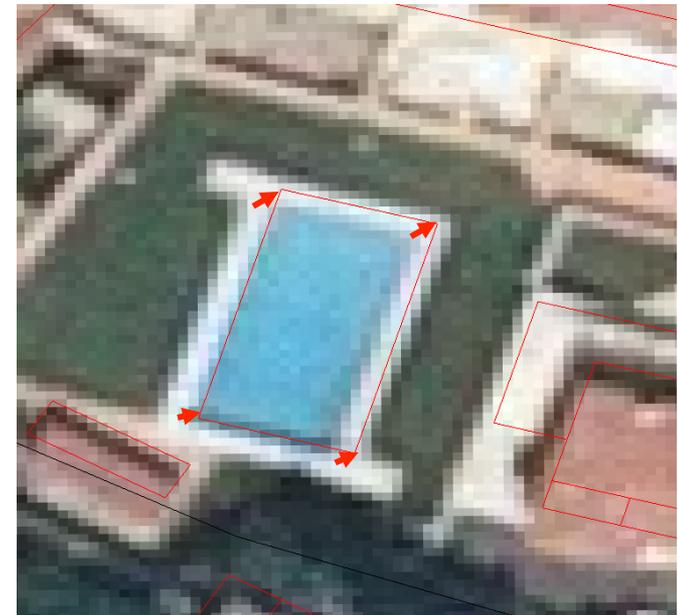
$$e_{y_i} = y_{t_i} - y_{m_i}$$

$$e_{z_i} = z_{t_i} - z_{m_i}$$

Tipos de los posibles errores:

- Groseros (equivocaciones).
- **Sistemáticos (sesgos)** (constantes o variables)
- Aleatorios
- **Atípicos**

Atípicos → Son señales, no errores

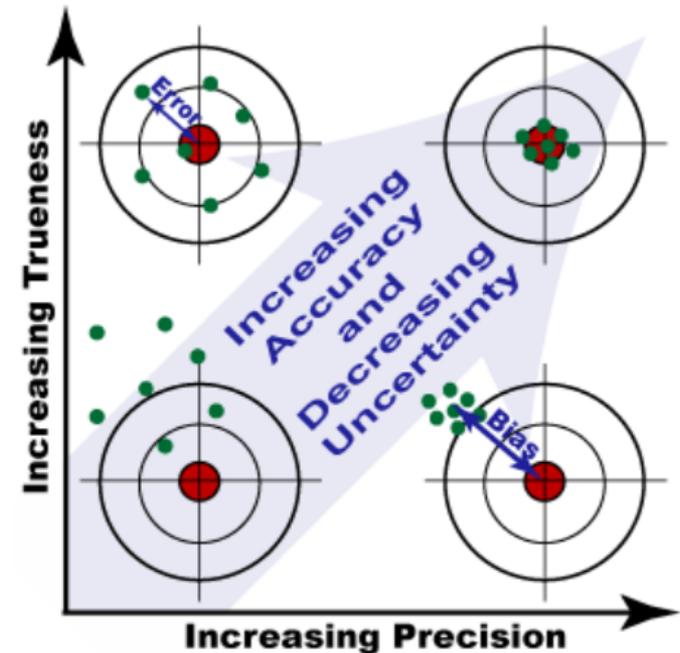


Exactitud: Grado de acuerdo entre el resultado de una prueba y el valor de referencia aceptado [ISO 3534-1]

Exactitud = veracidad + precisión

Veracidad: Proximidad entre la media de un número finito de **valores medidos repetidos** y un valor de referencia. → se puede mejorar (eliminar/reducir el sesgo)

Precisión: Proximidad entre los valores medidos obtenidos en mediciones repetidas de un mismo objeto o de objetos similares bajo condiciones específicas. → **no se puede mejorar (es connatural a cada tipología de dato), sólo se puede evitar su incremento en los procesados.**



En Geomática se usan modelos matemáticos para simplificar la gestión analítica de los errores (incertidumbres)

En la mayoría de los casos: Es el modelo **Gaussiano o Normal el modelo asumido**

Muchos estudios indican que esto no es correcto. Esta hipótesis no es siempre cierta.

Otros modelos base :

- LIDAR (Maune, 2007): No paramétrico (distribution free)
- Digitalización manual (Bolstad et al 1990): Bimodal
- Digitalización (Tong & Liu, 2004): p-norm (Normal + Laplace)
- Geocodificación (Cayo and Talbot 2003; Karimi and Durcik 2004, Whitsel et al. 2004): Log normal
- Observaciones GNSS (Wilson, 2006; Logsdon, 1995): Raleigh, Weibull
- Otros modelos mencionados: Folded Normal, Half Normal, Gamma

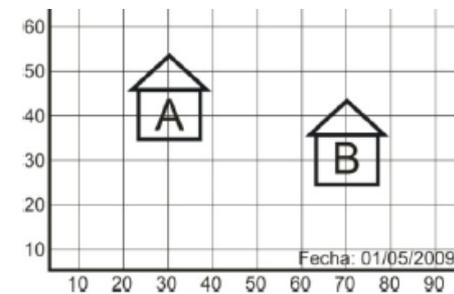


Categoría

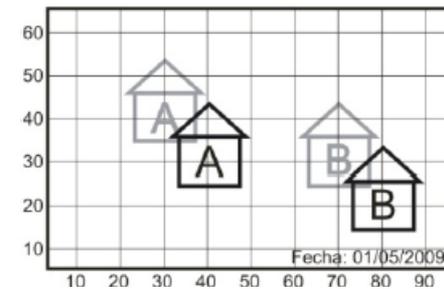
Exactitud posicional

- Se requiere personal entrenado y equipos GNSS.
- Métodos directos, externos.
- Necesidad de trabajo de campo (muestreo).
- Valoración de la exactitud.
- Se determina la dimensión de los errores.
- Hay numerosos estándares (NMAS, EMAS, NSSDA, ASPRS, ISO 3951, UNE 148002, STANAG 2215, etc.)

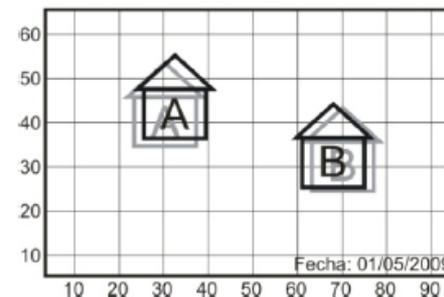
Terreno nominal



Caso 1



Caso 2



Categoría

Compleción

Consistencia lógica

Exactitud posicional

Calidad temporal

Exactitud temática

Elementos

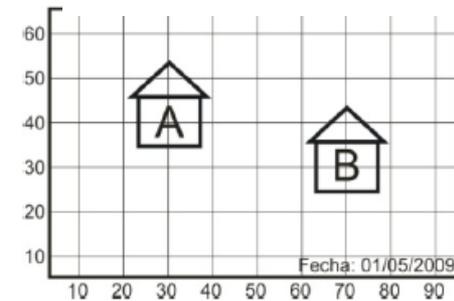
- **Exactitud de una medida de tiempo:** Corrección de las referencias temporales.
- **Consistencia temporal:** Corrección de eventos ordenados o secuencias.
- **Validez temporal:** Validez de los datos con respecto al tiempo.

Categoría

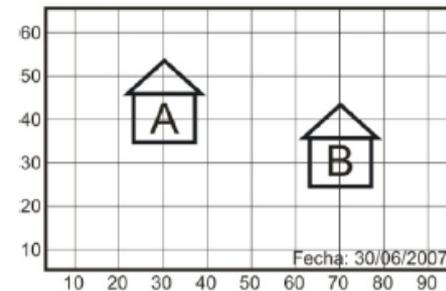
Calidad temporal

- Se requieren especificaciones muy buenas.
- Se requiere personal entrenado y métodos bien especificados.
- Algunos aspectos son automatizables
- Métodos directos, externos e internos.
- Necesidad de trabajo de campo (muestreo), en algunos casos sobre fotografías.
- Variables de conteo para aspectos cualitativos
- Variables de dimensión para aspectos cuantitativos
- Para una gestión adecuada se requiere de un modelo temporal (ciclos de vida, tasas de cambio, etc.)

Terreno nominal



Caso 8



$$Q_T = Q_{T-1} - \downarrow Q_{\Delta T} + \uparrow Q_{\Delta T}$$

Categoría

Compleción

Consistencia lógica

Exactitud posicional

Calidad temporal

Exactitud temática

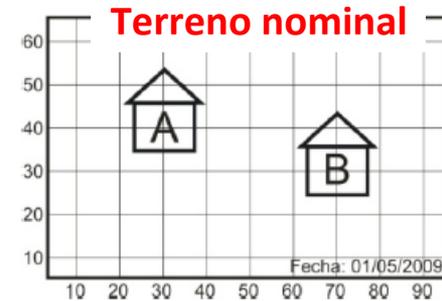
Elementos

- **Corrección de la clasificación:** Comparación de las clases asignadas a fenómenos o a sus atributos, con respecto a las que le corresponden en el universo de discurso.
- **Corrección de atributo no cuantitativo.**
- **Exactitud de atributo cuantitativo.**

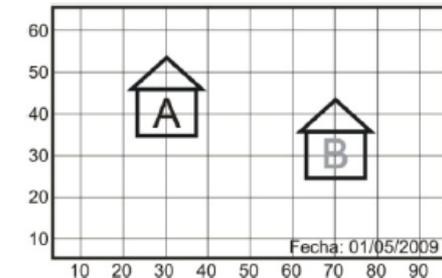
Categoría

Exactitud temática

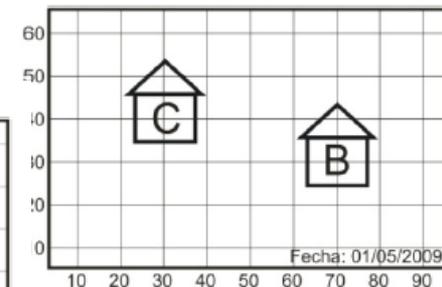
- Se requieren especificaciones muy buenas.
- Se requiere personal entrenado y métodos bien especificados.
- Métodos directos, externos.
- Necesidad de trabajo de campo (muestreo), en algunos casos sobre fotografías.
- Variables de conteo para aspectos cualitativos
- Variables de dimensión para aspectos cuantitativos
- Muy afectada por el tiempo.
- Se chequea junto con la compleción.



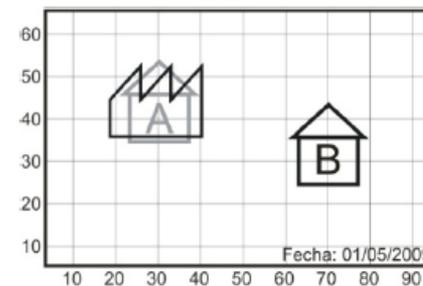
Caso 5



Caso 6



Caso 7



¡ Novedad ISO 19157 !

Categoría

Usabilidad

Elementos

Grado de adherencia a un conjunto específico de requisitos de calidad de datos.

Todos los elementos de calidad pueden usarse para evaluar la usabilidad

- Cualquier aspecto.
- Uso en la **agregación** de resultados.
- **Cajón de sastre:** Cualquier aspecto.
- ¡un mismo producto puede presentar usabilidades diferentes!

¡ Novedad ISO 19157 !

USABILIDAD

Es una definición no del todo acorde con otras definiciones de ISO:

- ISO/IEC 9126: "La usabilidad se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso"
- ISO/IEC 9241: "Usabilidad es la eficacia, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico"

Categoría

Usabilidad

Un Resultado Agregado de la Calidad de datos (RACD) combina los resultados de la calidad a partir de evaluaciones de la calidad de datos basadas en diferentes elementos de la calidad de datos o diferentes ámbitos de la calidad de datos.

$$RACD = v_1 \times v_2 \times v_3 \times \dots \times v_n,$$

$$RACD = v_1 \times w_1 + v_2 \times w_2 + v_3 \times w_3 + \dots + v_n \times w_n$$

$$RACD = \max. (v_i, i_n = 1 \dots n)$$

$$RACD = \min. (v_i, i_n = 1 \dots n)$$

ELEMENTOS

Subelementos

En ISO 19113 se podían incorporar nuevos elementos y subelementos.

**EN ISO 19157 NO SE CONSIDERA
Problema de retro-compatibilidad**

Elementos no considerados

Integridad

Incluida en las especificaciones de CDG para la navegación aérea (OACI Anexo 15).

Se define como:

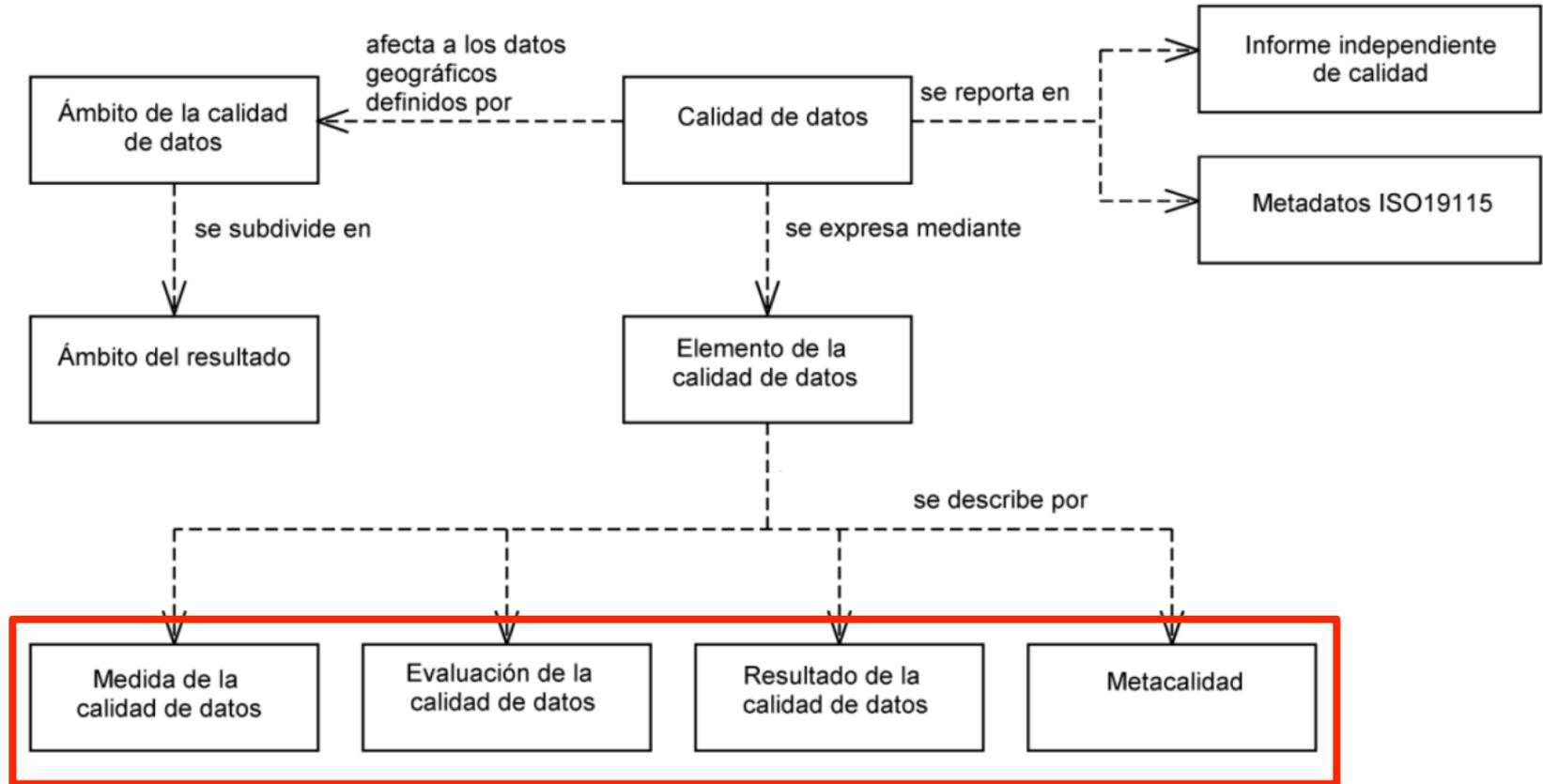
“Grado de garantía de que no se han perdido o alterado ninguna de las referencias ni sus valores después de la obtención del original o de una enmienda autorizada”

Elementos no considerados

Interoperabilidad/Confrontabilidad/capacidad de integración

Diferentes nombres:

- Consistencia entre CDG
- Confrontabilidad
- Interoperabilidad espacial
- Aspecto muy importante pero no plenamente formalizado
- Punto clave del futuro de las IDE.
- La idea es: Nivel al que es posible fundir 2 BDG correspondientes a la misma área geográfica, pero de fuentes distintas.



Medida

Un elemento de la calidad debe hacer referencia a una única medida.

Método de Evaluación

Describe aquellos procedimientos y métodos que se aplican para obtener un resultado de la calidad de datos

- Para cada medida de la calidad → un método de evaluación.
- Debería incluirse una fecha o un rango de fechas.

Resultados

Proporcionar, al menos, un resultado de la calidad para cada elemento de la calidad.

Cuantitativo
De conformidad
Descriptivo
En cobertura

¡ Novedad ISO 19157 !

Metacalidad

- Definición: Es la información sobre la calidad de la información de la calidad de un CDG (ISO 19157).
- Se describe por medio de:
 - **Confianza:** Exactitud del resultado de la calidad.
 - **Representatividad:** Grado de representatividad de la muestra utilizada
 - **Homogeneidad:** Uniformidad comprobada o esperada de los resultados obtenidos.

Principales características y cambios

- Se han eliminado los elementos generales de la calidad
 - Se ha eliminado la relación jerárquica entre elementos y subelementos.
 - Se ha suprimido la posibilidad de crear nuevos elementos y subelementos de la calidad.
-
- Hay mayor formalización (UML).
 - Se ha incluido la usabilidad.
 - Se ha incluido la metacalidad.
 - Se ha incluido la unidad de calidad de datos (DQU).
 - Mayor integración.

Idea

En un CDE en el que se analiza un **aspecto** (elemento) concreto, y tal que el resultado se expresa con un **medida** dada.

Se pueden obtener **resultados** distintos según la **metodología** de evaluación utilizada

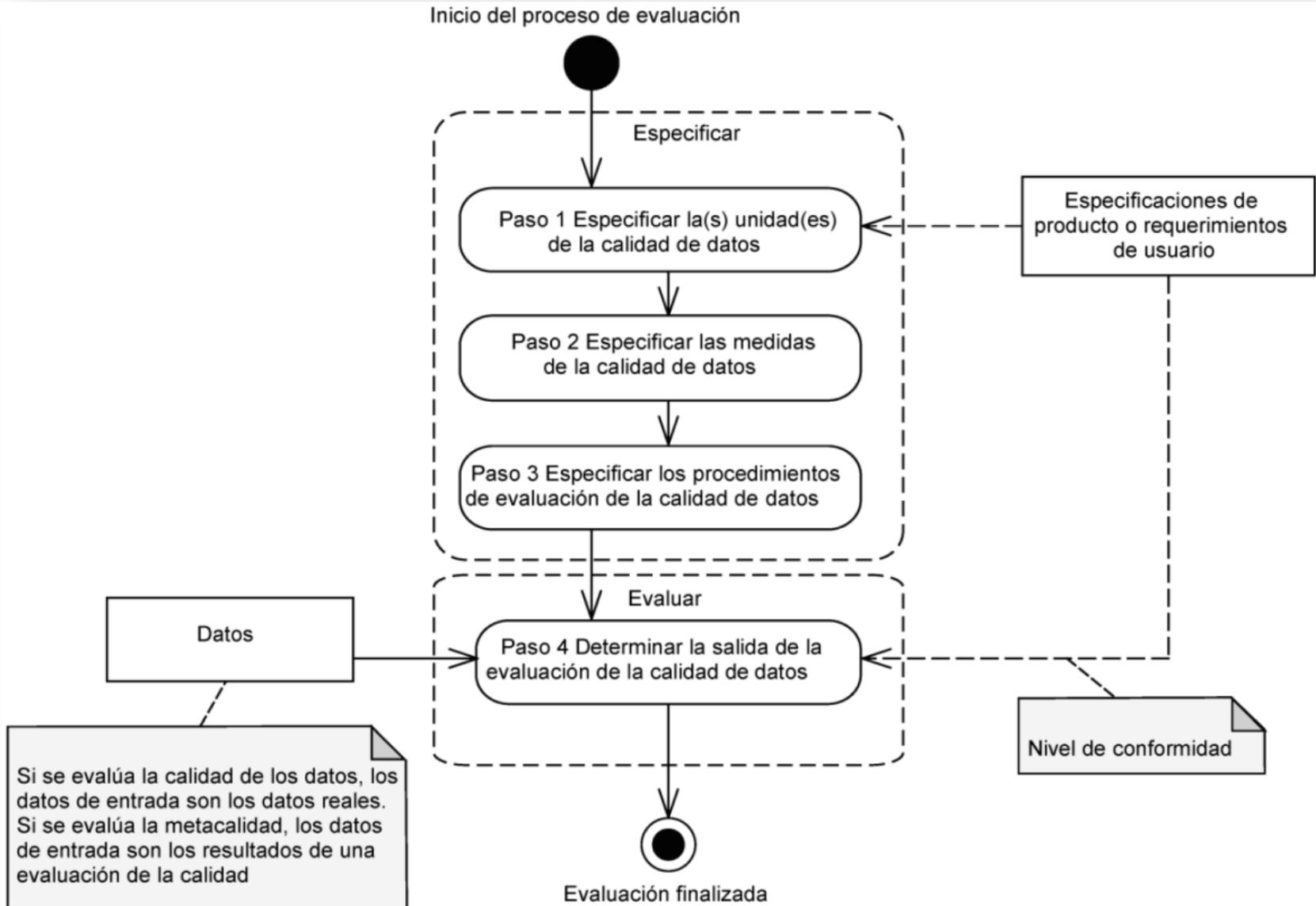
→ **Ambigüedad**

Objetivo

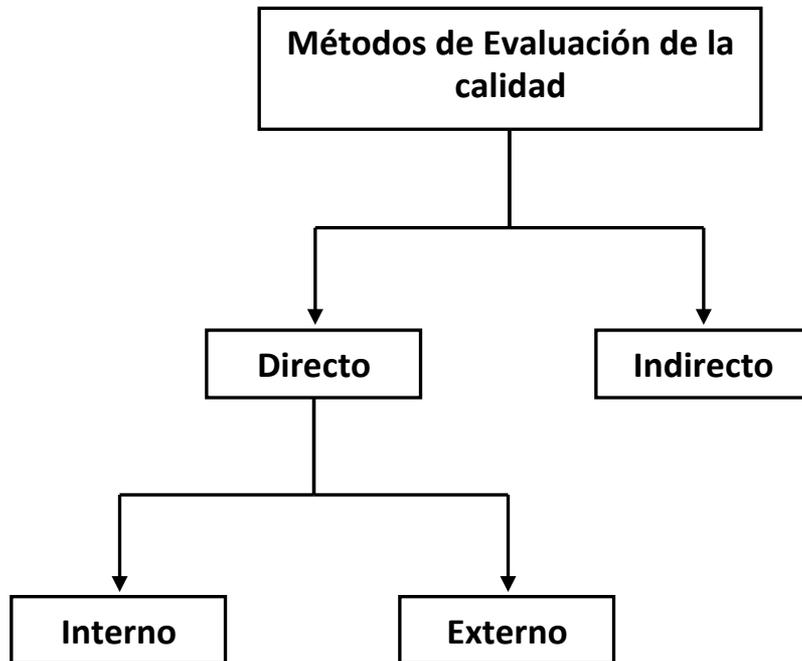
Facilitar un **marco de procedimientos** para determinar y evaluar la calidad de BDG consistente con los principios definidos en ISO 19113, como parte de los metadatos o del informe de evaluación de la calidad.

¡ Hay cambios en el proceso !

¡ Hay cambios en la manera de informar !

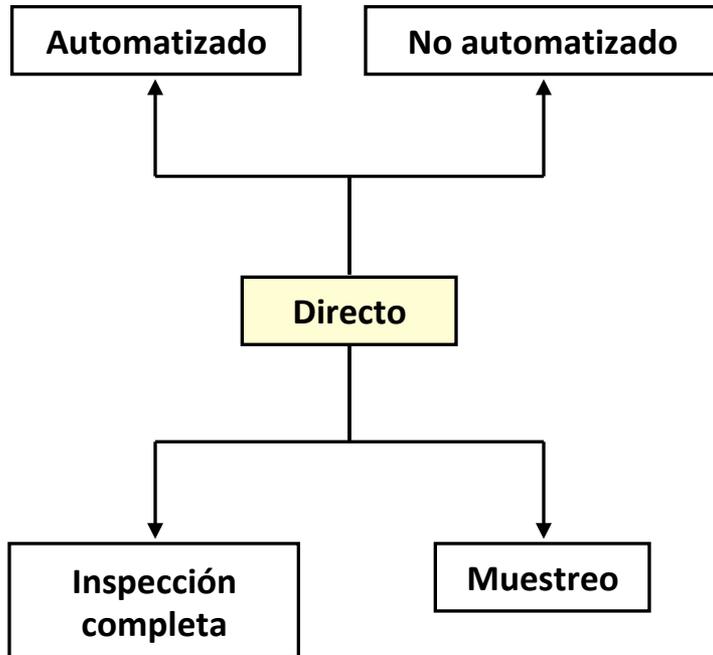


Métodos



- Método directo: comparación de los datos con información de referencia interna y/o externa.
- Método indirecto: Basados en el conocimiento externo. Se usan sólo si no se pueden usar métodos directos
- Método interno: usa datos internos al propio conjunto que evaluado.
Ejemplo: prueba de consistencia lógica sobre la consistencia topológica del cierre de límites.
- Método externo: usa datos de referencia externos al conjunto evaluado.
Ejemplo: prueba de completación de los nombres de carreteras requiere otra fuente de información.

Métodos



Elementos fáciles de automatizar:

- Consistencia lógica / consistencia topológica.
Ej: Cierre de polígonos
- Compleción / omisión.
Ej: Comparación de nombres de calles con otro archivo.
- Etc.

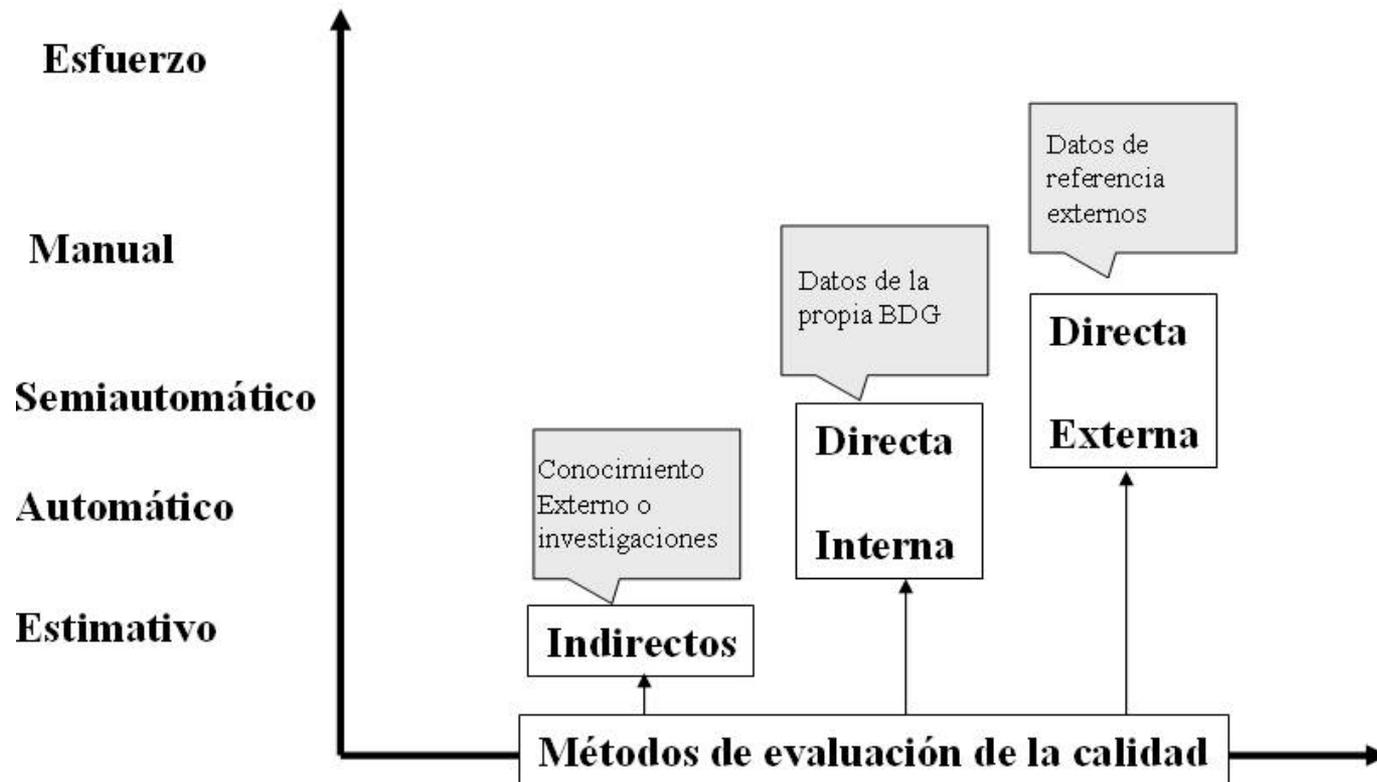
Inspección completa:

- Se comprueba cada ítem de la población especificada por el ámbito.
- Apropia para poblaciones pequeñas o para pruebas que pueden ser automatizadas

Muestreo:

- Se comprueban suficientes ítems de la población para alcanzar un resultado representativo

Métodos



Resultado

Resultado agregado de la calidad de datos (RACD)

- Combina resultados de diferentes elementos o ámbitos de la calidad.
- Puede ser cuantitativo o cualitativo. Puede representarse por un valor numérico o booleano.
- Se puede considerar que un conjunto de datos tiene una calidad agregada aceptable a pesar de que uno o más de los resultados individuales no sean aceptables.
- El significado **del RACD debe quedar claro**, antes de extraer cualquier tipo de conclusión.

Ejemplos de agregación

- Conforme / no conforme al 100%. A cada resultado de la calidad implicado en el cálculo se le da un valor Booleano v de uno (1) si ha pasado la prueba y de cero (0) si ha fallado. La calidad agregada se determina por la ecuación:

$$\text{RACD} = v_1 \times v_2 \times v_3 \times \dots \times v_n$$

Resultado

Derivación de resultados

- Se crea un nuevo resultado a partir de los ya existentes.
- Por ejemplo, se dispone de un nivel de incertidumbre y con una nueva tolerancia se considera un nuevo resultado de aceptación/rechazo.

Idea

En un CDE en el que se analiza un aspecto (elemento/
subelemento) concreto,
y tal que se aplica una **metodología** de evaluación
dada,
pueden obtenerse **resultados** distintos según que la
medida utilizada no sea exactamente igual

→ **Ambigüedad**

Objetivo

Normalizar las medidas de la calidad de datos.

Generalidades

- Todo resultado será interpretable si se sabe la medida utilizada.
- La normalización de las medidas permite guiar al productor en la elección de las adecuadas, y al usuario en la adecuación de la BDG.
- Cada medida se define con unos “componentes”
- Catálogo de 78 medidas (anexo D).
- Se permite crear nuevas medidas.
- **Obliga a disponer de un catálogo (nuevo en ISO 19157).**

Medidas básicas

Evitan la repetición en la definición de conceptos. Son medidas base para el desarrollo de las medidas concretas del Anexo D.

Se recogen en el Anexo G (normativo). Hay dos tipos:

- **Medidas relacionadas con el recuento:** Se basan en el número de errores (defectos) o de valores correctos.
- **Medidas relacionadas con la incertidumbre:** Se basan en modelos estadísticos que tratan de cuantificar la incertidumbre de las medidas.

Medidas

Para facilitar las comparaciones entre conjuntos de datos, es necesario que los resultados de los informes de la calidad de datos se expresen de forma comparable y que exista consenso en las medidas de calidad que se han usado.

Medidas normalizadas de la calidad de datos

En el anexo D se ofrece una lista de medidas normalizadas de la calidad. Cada medida de esta lista contiene todos los componentes requeridos.

Cualquier registro que se establezca para gestionar medidas normalizadas de la calidad de datos debe ser conforme con la **Norma ISO 19135:2005**.

Medidas

Medidas de calidad de datos definidas por el usuario

Debido a la naturaleza de los datos geográficos y de su calidad, la lista de medidas normalizadas no puede estar completa. El usuario puede añadir nuevas medidas según necesidad.

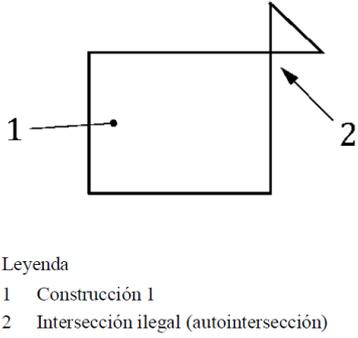
Catálogo de medidas de la calidad de datos

Pueden proporcionarse catálogos de medidas de calidad asociados con los metadatos o disponibles en línea, para describir de forma completa las medidas referenciadas en el informe de la calidad de datos evaluados.

Lista de componentes: Cada medida de la calidad de datos se describe mediante los siguientes componentes:

Ejemplo

Medida 26

Línea	Componente	Descripción
1	Nombre	número de errores por autointersecciones no válidas
2	Alias	bucles (<i>loops</i>)
3	Nombre del elemento	consistencia topológica
4	Medida básica	recuento de errores
5	Definición	recuento de todos los ítems de los datos que ilegalmente se intersectan con ellos mismos
6	Descripción	–
7	Parámetro	–
8	Tipo de valor	entero
9	Estructura del valor	–
10	Fuente de referencia	–
11	Ejemplo	 <p>Leyenda 1 Construcción 1 2 Intersección ilegal (autointersección)</p>
12	Identificador	26

Idea

La calidad es importante, pero si no se conoce no sirve para nada.

Deben existir metadatos sobre la calidad de la IG.

Los metadatos deben tener una estructura mínima estandarizada.

Centrada en los resultados más que en detalles del proceso.



Opciones

ISO 19157

Reporte (¿Cómo informar?)

Como **METADATOS**.

- ¡Obligatorio!
- Conforme a ISO 19115

Como **informe de la calidad independiente**:

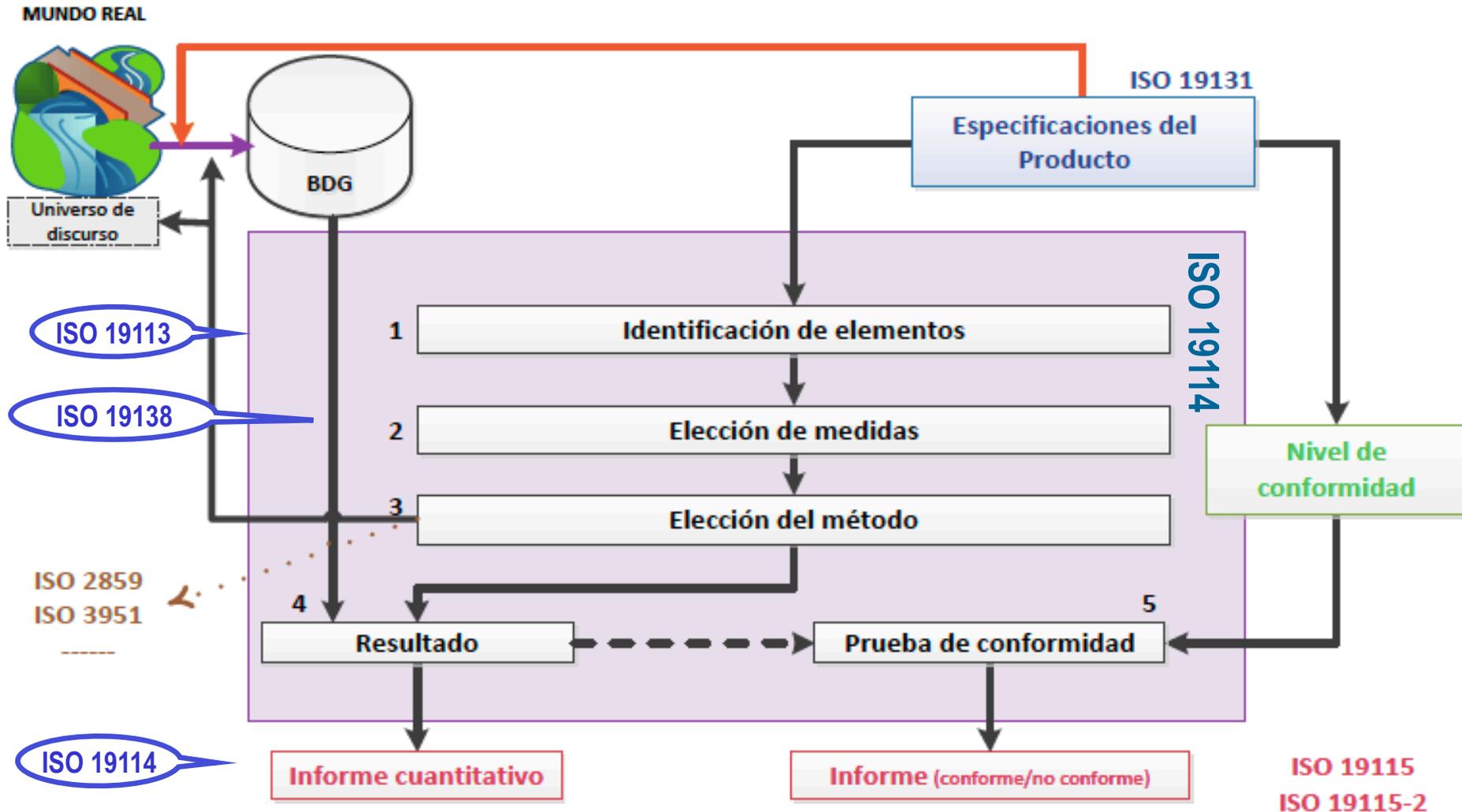
- NO SUSTITUYE NUNCA a los metadatos.
- Proporciona más detalle sobre los resultados y procedimientos. A utilizar cuando se desee.
- **Estructura libre (punto débil de la norma).**
- Se debe generar cuando:
 - **Resultados en metadatos como conforme / no conforme**
 - **Cuando se generan resultados agregados (¿cómo se ha agregado?, ¿cómo interpretar el resultado?)**



Opciones

Se debe informar sobre la calidad de datos con metadatos en conformidad con el capítulo 7, capítulo 10, anexo C, ISO 19115-1:2014 y ISO 19115-2:2009.

- Para proporcionar **más detalles** que los ofrecidos en los metadatos, se puede, adicionalmente, crear un informe de calidad independiente. Su estructura es libre.
- No obstante, el informe de calidad independiente **no debe reemplazar a los metadatos**.
- Los metadatos deberían proporcionar una referencia al informe de calidad independiente cuando exista.



Relación entre los procesos de ISO 19157 y normas relativas a la calidad de la IG. Tomado de Ariza-López y Rodríguez-Pascual (2007)

Conclusiones

- Disponemos de un marco para la calidad de la IG
- ISO 19157 está mejor definida y realiza avances respecto al sistema anterior (19113+14+38)
- Se va a completar con un esquema XML (19157-2).
- Se propone un informe pero sin estandarizarlo.

- Se abarca la definición de los elementos de la calidad, la evaluación de la calidad, las medidas de la calidad y el informe de la calidad.

Conclusiones

Limitaciones

- No es retrocompatible.
- No permite crear nuevos elementos de la calidad.
- No ha añadido nuevas medidas.
- No es un modelo adecuado para los servicios (ni los trata)
- No contempla otros tipos de datos (BIM, enlazados, de producto).
- No es directamente integrable con otros modelos de ISO
(p.ej. datos de producto).
- Modelo poco operativo para flujos de datos.

Referencias

- AENOR (2014). UNE-EN ISO 19157:2014 Información geográfica - calidad de datos. AENOR.
- Ariza-López, Francisco Javier (2011). *Apuntes de clase: Introducción a la Calidad de la IG.*
- Ariza-López, FJ et al. (2013). *Fundamentos de evaluación de la calidad de la información geográfica.*
- Ariza-López, Francisco Javier (2015). *Calidad de la IG: Situación actual y perspectivas de futuro.* En: Seminario SIGTIERRAS-Quito.
- Ariza-López, Francisco Javier (2017). *Taller sobre la gestión y evaluación de la calidad.* En: Taller-Santa Fe de Bogotá.

¡Gracias por su atención!