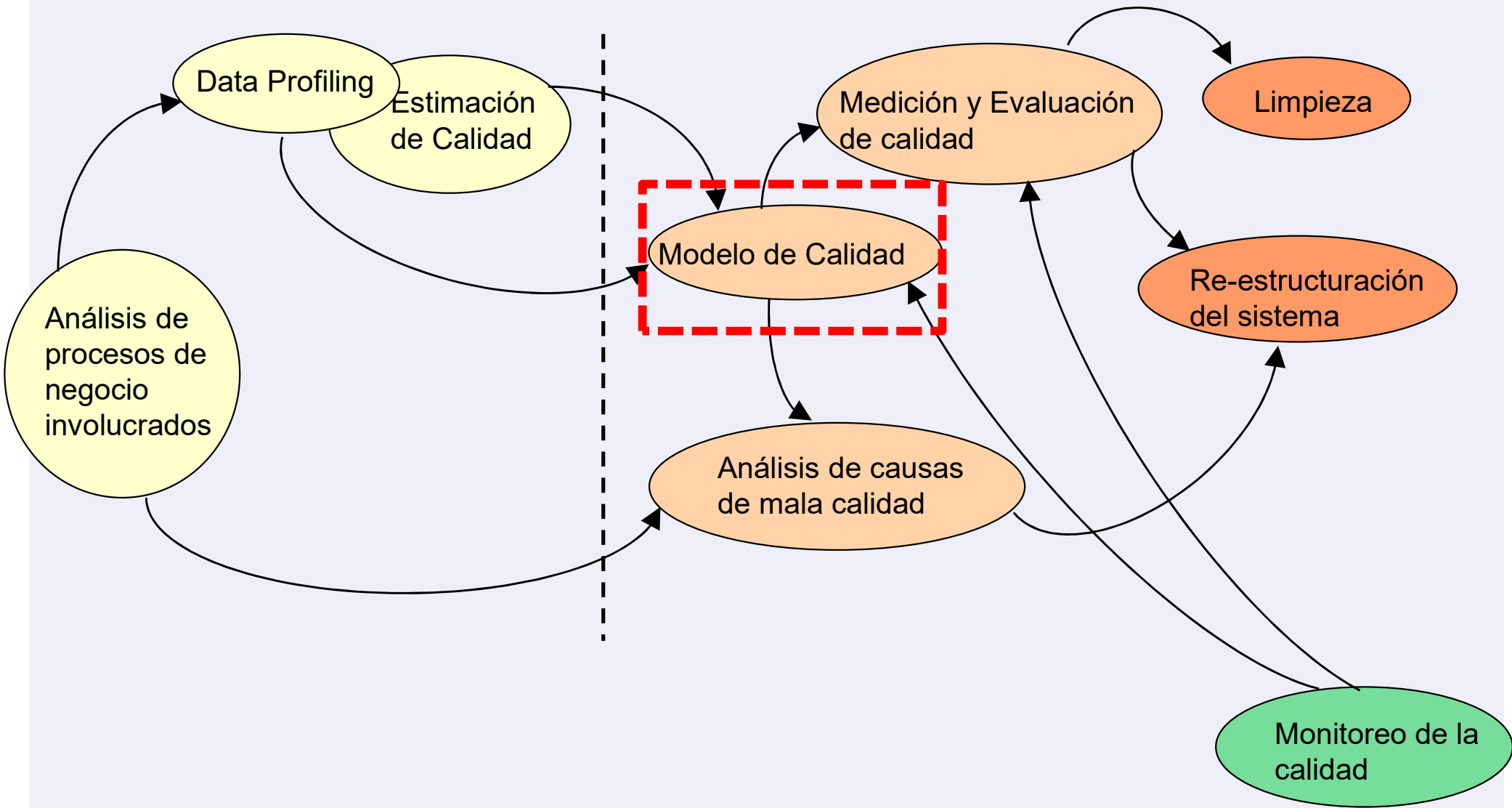


---

# Calidad de Datos e Información

## Dimensiones de Calidad

# Gestión de la calidad en SI



# Modelo de Calidad de datos

---

- Define:
  - qué características de calidad se manejan
  - sobre qué datos aplican
  - cómo se miden esas características
- Para poder construir el modelo de calidad es necesario estudiar cómo clasificar los problemas de calidad de datos

# Dimensiones de Calidad

---

- Multi-dimensionalidad de la calidad
  - Concepto de dimensión de calidad
  - Jerarquía de conceptos de calidad: factores, métricas y métodos de medición
- Estudio de algunas dimensiones
  - Exactitud, completitud, frescura, consistencia, unicidad
- Relaciones entre dimensiones

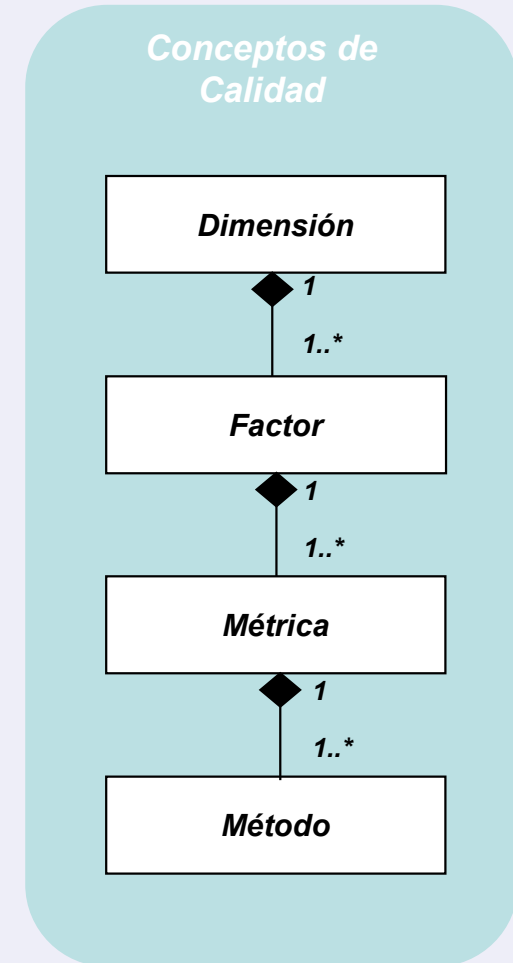
# Multi-dimensionalidad de la calidad

---

- La calidad se caracteriza vía múltiples dimensiones o atributos que ayudan a calificar los datos.
- Conceptos de Calidad
  - Dimensión
  - Factor
  - Métrica
  - Método de Medición

# Jerarquía de conceptos de calidad

- Las dimensiones representan las facetas de la calidad a alto nivel.
- Cada dimensión puede refinarse en un conjunto de factores que representan aspectos particulares.
- Cada factor puede medirse con varias métricas.
- Cada métrica puede implementarse con varios métodos de medición.



# Multi-dimensionalidad de la calidad

---

- **Dimensión de calidad:**

- Una dimensión captura una faceta (a alto nivel) de la calidad.
- Ejemplos:
  - *Frescura*: los datos son recientes/actualizados.
  - *Exactitud*: los datos son exactos/correctos.
  - *Compleitud*: disponemos de todos los datos.

- **Factor de calidad:**

- Un factor representa un aspecto particular de una dimensión de calidad.
- Ejemplo: Varios aspectos de la dimensión *Exactitud* son:
  - *Exactitud semántica*: si los datos representan entidades/estados del mundo real.
  - *Exactitud sintáctica*: si los datos no tienen errores sintácticos.
  - *Precisión*: si los datos tienen el suficiente nivel de detalle.

# Medición de la calidad

- **Métrica de calidad:**
  - Define la forma de medir un factor de calidad
  - Se define con:
    - Un *nombre*
    - Una *descripción* (qué se mide)
      - Ej.: cantidad de valores nulos, cantidad de tuplas, tiempo transcurrido desde la última actualización
    - Las *unidades* de medición
      - Ej.: *tiempo* de respuesta en ms, volumen en GB, un valor entre 0 y 1, etc.
    - La *granularidad* de la medida
      - Fuertemente dependiente del modelo de datos
      - Modelo relacional: celda, tupla, columna, tabla, grupo de tablas, base de datos

## **Modelo Relacional**

Tipo	Producto	Cant	PrecioUnit
Lacteos	Leche	5	1
Lacteos	Yogur	7	1.5
Bebidas	Agua Min	9	0.8



# Medición de la calidad

---

- **Método de medición:**
  - Un método es un proceso que implementa una métrica.
  - Es el encargado de tomar una serie de medidas (correspondientes a una métrica) para una BD concreta.
  - Ejemplo: para medir el tiempo transcurrido desde la última actualización, se puede:
    - Usar timestamps de la BD
    - Acceder a los logs de actualización
    - Comparar versiones de la BD
    - ...
- Una misma métrica puede ser medida por diferentes métodos.

# Ejemplo de conceptos de calidad

---

- Dimensión:
  - **Exactitud**: Conciernen la correctitud y la precisión con que los datos del mundo real son representados en un sistema de información
- Factor:
  - **Exactitud sintáctica**: Indica qué tan libre de errores sintácticos están los datos
- Métricas:
  - **Exact. Sint. Booleana**: Un booleano que indica si un dato es sintácticamente correcto o no. (*Ej. un teléfono es correcto o no*)
  - **Desviación de exact. sint.**: La distancia a un dato considerado como sintácticamente válido (*Ej. Montevideo, Mtdo*)
- Métodos:
  - **CheckRule**: Chequea si un dato satisface una regla de formato.
  - **CheckDictionary**: Chequea si un dato se encuentra en un diccionario.
  - **ComputeDistance**: Calcula la distancia entre un dato y el valor más cercano en un diccionario.

# Agregación de medidas

---

- **Medida de calidad**

- Es el valor obtenido por la aplicación de una métrica de calidad durante una medición (Ej., 0 o 1)

- **Granularidad**

- En Modelo Relacional puede ser: Celda, Tupla, Atributo (columna), Tabla (o relación)
- Análogamente en otros modelos
  - En XML/JSON podría ser: cada dato de determinados tags, o un documento.

- **Agregar**

- Obtener a partir de la medida a determinado nivel de granularidad, la medida a un menor nivel de granularidad
- Ej.: A partir de las medidas de las celdas, una medida para toda la tabla.

# Multitud de dimensiones

	WandWang 1996	WangStrong 1996	Redman 1996	Jarke 1999	Bovee 2001	Naumann 2002
Accuracy	S	S	S	S	S	S
Completeness	S	D	S	D%	S	S
Consistency	S		D%	S	S	
Representational Consistency		S	S			S
Timeliness	S	S		S	S	S
Currency	S		S	S	S	
Volatility	S			S	S	
Interpretability		S	D%	D%	S	S
Ease of Understanding/ Understandability		S				S
Reliability	D			D		
Credibility				D	D	
Believability		S				S
Reputation		S				S
Objectivity		S				S
Relevancy/ Relevance		S	S		D%	S
Accessibility		S		S	S	
Security/ Access Security		S		S		S
Value-added		S				S
Concise representation		S				S
Appropriate amount of data/amount of data		D	D			D
Availability				S		S
Portability			D	D		
Responsiveness/ Response Time				S		S

S- iguales  
D- diferentes  
D%- similares

# Algunas propuestas

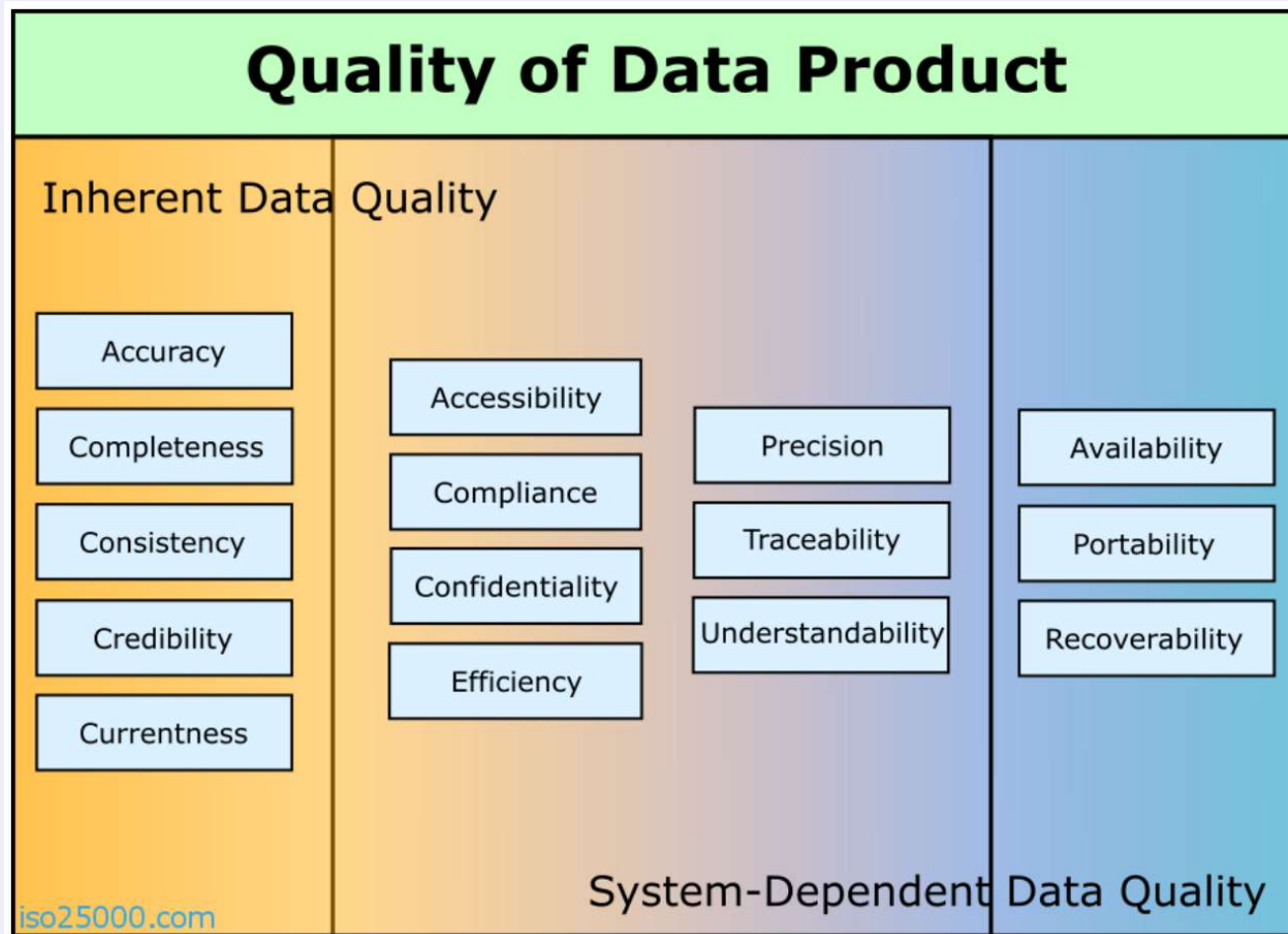
---

- “Data quality in context”, D. M. Strong, Y. W. Lee, and R. Y. Wang, *Commun. ACM*, vol. 40, no. 5, pp. 103–110, May 1997.

<b>Categoría</b>	<b>Dimensiones</b>
Intrinsic	Accuracy, Objectivity, Believability, Reputation
Accessibility	Accessibility, Access security
Contextual	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational	Interpretability, Ease of understanding, Concise representation, Consistent representation

# Algunas propuestas

- ISO/IEC 25012
  - Presenta las dimensiones de calidad de datos para los Sistemas de Información (SI).



# Algunas propuestas

- Data and Information Quality: Dimensions, Principles and Techniques, C. Batini, M. Scannapieco. 2016.

<b>Cluster</b>	<b>Tipo</b>	<b>Aspecto</b>
Accuracy	Structural Accuracy	Syntactic accuracy
		Semantic accuracy
	Time-Related Accuracy	Currency
		Volatility
	Timeliness	
Completeness	Completeness of Relational Data	Presence/absence and meaning of null values in an open/closed world assumption
	Completeness of Web Data	Completeness
Accessibility	Accessibility	Accessibility
Consistency	Integrity Constraints	Intrarelational constraints
		Interrelational constraints
	Data Edits	Data editing

# Estudio de algunas dimensiones

---

- Exactitud
- Completitud
- Frescura
- Consistencia
- Unicidad



# Dimensiones - factores

DIMENSION	FACTOR
Exactitud	Exactitud Semántica
	Exactitud Sintáctica
	Precisión
Compleitud	Densidad
	Cobertura
Frescura	Actualidad
	Oportunidad
	Volatilidad
Consistencia	Integridad de Dominio
	Integridad Intra-relación
	Integridad Inter-relación
Unicidad	No-duplicación
	No-contradicción

# Estudio de algunas dimensiones

---

- **Exactitud**
- Completitud
- Frescura
- Consistencia
- Unicidad

# Exactitud (Accuracy)

---

- Intuitivamente, la exactitud indica qué tan precisos, válidos y libres de errores están los datos:
  - ¿Estos datos son lo suficientemente precisos para nuestras necesidades?
  - ¿El nivel de detalle de los datos es adecuado?
  - ¿Estos datos se corresponden con el mundo real?
  - ¿Estos datos tienen errores? Y en tal caso, ¿los errores son tolerables?
  - ¿El formato de presentación de los datos es correcto? ¿Es estándar?
- La exactitud se relaciona con la correctitud y la precisión con la que están representados los datos en un SI. Por tanto, abarca aspectos:
  - de correctitud que son intrínsecos de los datos y
  - de representación (formato, precisión, etc.).

# Factores de exactitud

---

- **Exactitud semántica** (semantic accuracy):
  - ¿Los datos de mi SI se corresponden con la realidad?
  - **Interesa medir qué tan bien se representan los estados del mundo real en el SI.**
  - Varios problemas de exactitud semántica:
    - Datos que no corresponden a ningún estado del mundo real (mismembers).
    - Datos que corresponden a un estado equivocado del mundo real.
    - Datos con errores en algunos atributos.
  - Ejemplo: Datos de un estudiante pueden referenciar
    - a una persona inexistente,
    - a una persona equivocada, o
    - a la persona correcta pero con algunos errores (ej., su dirección)

# Factores de exactitud

---

- **Exactitud sintáctica** (syntactic accuracy):
  - ¿Los datos de mi SI tienen errores sintácticos o de formato?
    - Valores mal escritos son difíciles de interpretar por un proceso
  - **Interesa medir si los valores del SI corresponden a valores válidos del dominio (no importa si son los valores reales)**
  - **Varios problemas de exactitud sintáctica:**
    - Errores de valores: Valores fuera de rango, errores ortográficos y de tipo.
      - Apellido: “Marínez” en lugar de “Martínez”
      - Edad: 338 años
    - Errores de estandarización: Valores que no tienen el formato esperado.
      - Sexo: “0” y “1” en lugar de “F” y “M”.
      - Precios: en moneda extranjera en lugar de pesos
      - Pesos: en gramos en lugar de kilos
    - Valores embebidos: Valores que corresponden a múltiples atributos
      - Dirección: embebe *calle número apto CP ciudad*.

# Factores de exactitud

---

- **Precisión** (precision):
  - ¿Los datos de mi SI brindan el suficiente detalle?
  - **Interesa medir qué tan detallados son los datos del SI.**
  - Ejemplos:
    - Salario: “\$10.000” vs. “\$10.014” vs. “\$10.013,88”
    - Fecha: “1977” vs. “julio de 1977” vs. “14/7/1977” vs. “14/7/1977 10:55:32.4”
    - Color: “Rojo” vs. “204R-51G-0B”
    - Cabello: “Castaño” vs. “Castaño claro cobrizo n° 5”
    - Dirección: “J.Herrera y Reissig 565, 11300, Montevideo” vs. “Montevideo”

# Comparación factores exactitud

stid	name	address	telephone	interview	test
21	María Roca	Carrasco	6001104	low	1.0
22	Juan Pérez	Coloniaa 1280/403	9023365	medium	.5
43	Emilio Gutiérrez	Irigoitia 3843	3364244	high	.8
56	Gabriel García	Propios 2145/101		low	.5
57	Laura Torres	Maldonado & Yaro	099628734	medium	.7
58	Raúl González	Rbla Rca Chile 1280/1102	4112533	medium	.9
101	Carlos Schnider	Copacabana 1210	094432528	high	.9701
102	Miriam Revoir		9001029	medium	.7945
103	A. Benedetti	Charrúa 1284/1	7091232 (tía)	low	.9146
104	Luis López	Sixtina s/n		high	.822

■ Errores semánticos

■ Errores sintácticos

■ Faltas de precisión

# Estudio de algunas dimensiones

---

- Exactitud
- Completitud
- Frescura
- Consistencia
- Unicidad



# Compleitud (Completeness)

---

- Intuitivamente, la completitud indica si el SI contiene toda la información de interés:
  - ¿El SI representa todos los objetos de nuestra realidad?
  - ¿Qué porción de la realidad está representada en el SI?
  - ¿Tenemos todos los datos que describen a nuestros objetos?
  - ¿Tenemos muchos valores nulos?
- La completitud recubre los siguientes aspectos del SI:
  - Extensionales: La cantidad de entidades/estados de la realidad, representados en el SI
  - Intensionales: La cantidad de datos sobre cada entidad/estado del SI

# Factores de completitud

- **Cobertura** (coverage):

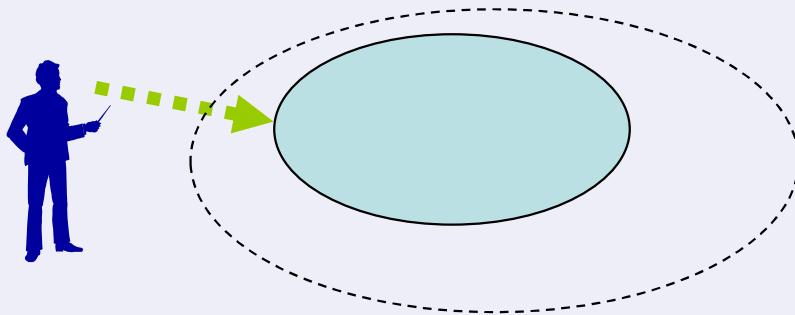
- ¿Cuántas entidades de la realidad contiene mi SI?

- Mundo cerrado (close world): Una tabla contiene todos los estados de la realidad que ella describe.
- Mundo abierto (open world): Una tabla puede contener sólo una parte de los estados de la realidad que ella describe.

- Interesa medir la porción de los datos de la realidad contenidos en el SI.

- Ejemplos:

- De los clientes potenciales, ¿cuántos conozco?
- ¿Qué porcentaje de las empresas están registradas en la DGI?



# Factores de completitud

---

- **Densidad** (density):
  - ¿Cuánta info tengo sobre las entidades de mi SI?
  - Interesa medir cuánta info tengo y cuánta me falta sobre las entidades del SI.
- Varias interpretaciones de la falta de valores (nulos):
  - Existen pero no los conozco (ej. No conozco el teléfono de Raquel).
  - Porque no existe (ej. Raquel no tiene teléfono).
  - No se si existe (ej. No se si Raquel tiene teléfono).

# Estudio de algunas dimensiones

---

- Exactitud
- Completitud
- Frescura
- Consistencia
- Unicidad

# Frescura (freshness)

---

- Intuitivamente, la frescura indica qué tan viejos son los datos:
  - ¿Estos datos son lo suficientemente frescos para nuestras necesidades?
  - ¿Estos datos son todavía vigentes o están obsoletos?
  - ¿Tenemos los datos más recientes?
  - ¿Actualizamos últimamente los datos?
- La frescura embebe la perspectiva temporal de los datos:
  - ¿Cuándo se crearon/modificaron estos datos?
  - ¿Cuándo se actualizaron estos datos?
  - ¿Cuándo se corroboraron estos datos?

# Factores de frescura

- **Actualidad** (currency):

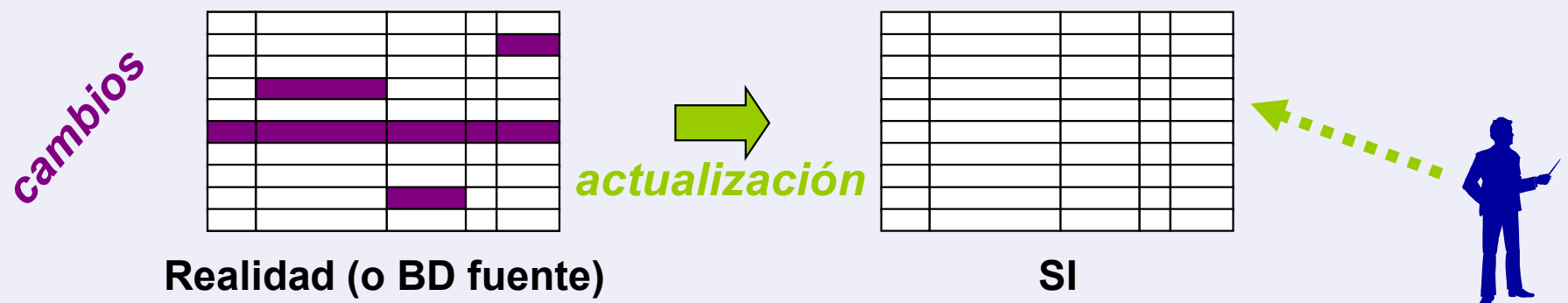
- ¿Qué tan vigentes son los datos de mi SI?

- Un SI es una vista de entidades/estados de la realidad (o de BD fuentes o externas) en un momento dado.
- Cuando la realidad cambia, el SI puede quedar desactualizado.

- Ejemplos:

- Datos de clientes (direcciones, teléfonos,...)
- Saldos bancarios (casa central, sucursales, est. de cuentas, ...)

- Interesa medir el desfase entre los datos del SI y los datos reales o de una fuente de datos



# Factores de frescura

---

- **Oportunidad** (timeliness):
  - ¿Qué tan vigentes/oportunos son los datos de mi SI?
    - Cuán actuales son los datos para la tarea en mano.
    - Los datos de un SI pueden ser actuales pero inútiles por no llegar a tiempo para un uso específico.
  - Ejemplos:
    - Cartelera de cursos universitarios que se publica luego de comenzados los cursos.
    - Stock que se actualiza luego que se sacaron las órdenes de compra en base a cantidades en stock.
  - Interesa medir la oportunidad de los datos del SI, teniendo en cuenta la actualidad y chequeando si el dato estuvo a tiempo.

# Factores de frescura

---

- **Volatilidad** (volatility):
  - ¿Qué tan inestables son los datos de mi SI?
    - Caracteriza la frecuencia con que los datos cambian en el tiempo.
    - Es una característica inherente a la naturaleza del dato.
  - Ejemplos:
    - Fecha de nacimiento debería tener volatilidad 0.
    - Cantidad en stock seguramente tendrá alta volatilidad, es válido por intervalos de tiempo muy cortos.
  - Interesa medir el lapso durante el cual los datos se mantienen válidos.



# Estudio de algunas dimensiones

---

- Exactitud
- Completitud
- Frescura
- **Consistencia**
- Unicidad

# Consistencia (consistency)

---

- Intuitivamente, la consistencia captura la satisfacción de reglas semánticas definidas sobre los datos:
  - ¿Los datos satisfacen las reglas de dominio?
  - ¿Las dependencias funcionales y referenciales se satisfacen?
  - ¿Hay contradicciones entre los datos?
- Pueden ser reglas de integridad para una BD o reglas de los usuarios
  - Reglas de integridad: son propiedades que deben satisfacer todas las instancias de una BD.
  - Reglas de usuarios: no implementadas en la BD pero necesarias para una aplicación.

# Factores de consistencia

---

- **Integridad de dominio**
  - Satisfacción de reglas sobre el contenido de un atributo.
    - Ej. edad entre 0 y 120 años.
- **Integridad intra-relación**
  - Satisfacción de reglas entre atributos de una misma tabla.
  - Reglas más típicas:
    - Dependencias de clave y de unicidad
    - Dependencias funcionales
    - Dependencias de valores. Ej. Edad = Year (now()) – FechaNacimiento)
    - Expresiones condicionales (edits). Ej. EstadoCivil = “casado” → Edad ≥ 14
- **Integridad inter-relación**
  - Satisfacción de reglas entre atributos de varias tablas.
  - Reglas más típicas:
    - Dependencias de inclusión (clave foránea, integridad referencial)
- **Interesa medir qué tan bien se satisfacen las reglas de integridad**

# Estudio de algunas dimensiones

---

- Exactitud
- Completitud
- Frescura
- Consistencia
- Unicidad

# Unicidad (Uniqueness)

---

- Intuitivamente, la unicidad indica el nivel de duplicación entre los datos.
  - ¿Los mismos datos están repetidos en el SI?
  - ¿Hay datos contradictorios?
- La duplicación ocurren cuando una misma entidad está representada dos o más veces en un SI.
- La duplicación va más allá de un chequeo de unicidad de clave.
  - Una misma entidad puede identificarse de diferentes formas.
    - Ej. Un docente se identifica por un su NumeroCobro; un estudiante se identifica por su CI. Pero hay datos repetidos de docentes que son también estudiantes.
  - Una misma entidad puede repetirse por errores en la clave.
    - Ej. una CI mal digitada.
  - Una misma entidad puede repetirse con claves diferentes.
    - Ej. un cliente se identifica por su cuenta bancaria. Pero una persona puede tener varias cuentas.
    - Ej. Un usuario se identifica por su email. Pero una persona puede tener varios emails.

# Factores de unicidad

---

- **No-duplicación** (duplication-free):
  - Hay duplicación si la misma entidad aparece repetida en forma exacta.
    - Los valores de la clave y los atributos coinciden (o son nulos en algunas tuplas).
    - Ej. <1.234.567-8, “A. Sosa”, 25 años, casado>, .  
<1.234.567-8, “A. Sosa”, NULL, NULL>
- **No-contradicción** (contradiction-free):
  - Hay contradicción si la misma entidad aparece repetida con contradicciones.
    - Los valores de la clave pueden coincidir o no.
    - Hay diferencias en valores de algunos atributos (no nulos)
    - Ej. <1.234.567-8, “A. Sosa”, 25 años, casado>, .  
<1.234.567-6, “Andrés Sosa”, 24 años, NULL>
- En ambos casos interesa medir la cantidad de repetidos.

# EJERCICIO 4

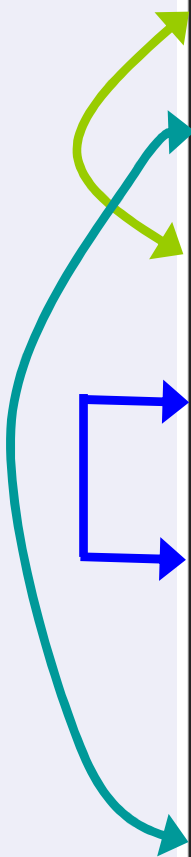
---

- Identificar las dimensiones y los factores de calidad donde clasificaría cada uno de los problemas descritos en el Ejercicio 1.

# Comparación entre definiciones

- Dimensiones del tiempo.

Reference	Definition
Wand 1996	<u>Timeliness</u> refers only to the delay between a change of a real world state and the resulting modification of the information system state
Wang 1996	<u>Timeliness</u> is the extent to which age of the data is appropriate for the task at hand
Redman 1996	<u>Currency</u> is the degree to which a datum is up-to-date. A datum value is up-to-date if it is correct in spite of possible discrepancies caused by time-related changes to the correct value
Jarke 1999	<u>Currency</u> describes when the information was entered in the sources and/or the data warehouse. <u>Volatility</u> describes the time period for which information is valid in the real world
Bovee 2001	<u>Timeliness</u> has two components: age and volatility. Age or <u>currency</u> is a measure of how old the information is, based on how long ago it was recorded. <u>Volatility</u> is a measure of information instability-the frequency of change of the value for an entity attribute
Naumann 2002	<u>Timeliness</u> is the average age of the data in a source
Liu 2002	<u>Timeliness</u> is the extent to which data are sufficiently up-to-date for a task





# Comparación entre definiciones

- Dimensiones de completitud.

Reference	Definition
Wand 1996	The ability of <u>an information system</u> to represent every meaningful state of the represented real world system.
Wang 1996	The extent to which data are of sufficient breadth, depth and scope for the task at hand
Redman 1996	The degree to which values are present in a <u>data collection</u>
Jarke 1999	Percentage of the real-world information entered in the <u>sources and/or the data warehouse</u>
Bovee 2001	Deals with <u>information</u> having all required parts of <u>an entity's information</u> present
Naumann 2002	It is the quotient of the number of non-null values in <u>a source</u> and the size of the <u>universal relation</u>
Liu 2002	All values that are supposed to be collected as per a collection theory

# Dimensión que está cobrando importancia

---

- **Credibilidad**

- Las fuentes de datos externos no nos suelen dar tanta confianza como las fuentes internas a la organización
- Si la fuente es desconocida, cómo entendemos su relevancia y credibilidad?
- **Source credibility.** *En social media data:*
  - Identity
  - Expertise
  - Reputation
- Provenance (data lineage): La procedencia y traza de los datos

# Dimensiones tradicionales - BigData

	Datos Transaccionales	Big Data o Social Media Data
ACCURACY	Utilizamos un referencial para medirla.	En general no es posible tener un referencial. Crowd-sourcing?
CONSISTENCY	Controles en el ingreso de datos. Restricciones de integridad y triggers para reglas de negocio (BDs relacionales)	Imposible controlar ingreso de datos. BDs NoSQL no tienen funcionalidades para eso, y muy costoso implementarlo
TIMELINESS	Tiempo transcurrido entre generación del dato y uso del mismo	Típicamente describe eventos en tiempo real. Esta información es muy sensible al momento en que se genera / lee.
COMPLETENESS	De gran relevancia. Cobertura y Densidad.	No tenemos esquema para medir densidad. Cobertura depende del contexto de uso, para qué y con respecto a qué.

# EJERCICIO 5

---

- Identificar pares de dimensiones/factores de calidad que cumplan: cuando mejora uno automáticamente mejora el otro.
- Identificar pares de dimensiones/factores de calidad que cumplan: cuando mejora uno automáticamente empeora el otro.

# Relaciones entre dimensiones

---

- Hasta ahora hemos estudiado los factores de calidad por separado, pero... **los factores no son independientes entre sí.**
- Algunos ejemplos de relaciones:
  - Errores de tipeo (**correctitud sintáctica**) pueden provocar que un dato sea considerado como **semánticamente incorrecto** por no poder encontrarlo en un referencial.
  - Datos **desactualizados** pueden causar **incorrectitudes semánticas**.
  - Los valores nulos (problemas de **densidad**) pueden ocultar **duplicados** e **inconsistencias**.
- En consecuencia, algunos factores pueden estimarse a partir de otros. Por ejemplo:
  - en contextos donde hay pocos errores, en gral. de tipeo, la **correctitud semántica** puede estimarse a partir de la **correctitud sintáctica**.
  - las **inconsistencias** suelen ser indicios de problemas de **correctitud semántica** o de datos **desactualizados**.

# Relaciones entre dimensiones

---

- Análogamente, las acciones que puedan realizarse para mejorar un factor de calidad pueden impactar en otros factores.
- Dos formas de influencia:
  - Correlaciones positivas: se mejoran ambos factores
  - **Negativa**: se mejora un factor mientras se empeora el otro
- Dichas correlaciones **dependen del contexto** de aplicación.
  - Correlaciones válidas en un contexto pueden no serlo en otro.
  - Ejemplo:
    - En una aplicación bancaria, contar con datos **actualizados** generalmente se asocia a un incremento de la **completitud** (tengo info de más transacciones)
    - En un sistema de sensores en tiempo real, requerir datos **actualizados** puede significar trabajar con información **incompleta** ya que algunos sensores pueden no transmitir a tiempo.

# Relaciones entre dimensiones

---

- Las correlaciones positivas permiten aplicar acciones de mejora conjunta.
  - Ejemplos:
    - Si se logran obtener datos más **actualizados**, se podría mejorar la correctitud semántica (más datos corresponderían a la realidad).
    - Si se completan los valores nulos (**densidad**) también se podría mejorar la correctitud semántica.
  - Correlaciones consideradas típicamente:
    - Correctitud sintáctica y semántica
    - Consistencia y exactitud
    - Actualización y correctitud semántica

# Relaciones entre dimensiones

---

- Las correlaciones negativas implican poner en balanza algunos factores.
  - Ejemplos:
    - Si se realizan tareas costosas para corregir errores de **correctitud semántica** (por ej. de control manual), se puede penalizar al sistema en sus **tiempos de respuesta** o en la **frescura** de los datos.
    - Si se integran datos externos para mejorar la **completitud**, se pueden introducir **duplicados** o degradar la **consistencia** (si los nuevos datos no satisfacen las reglas de integridad).
    - Si se eliminan datos **inconsistentes** se pierde en **completitud**.
  - Correlaciones negativas consideradas típicamente:
    - Completitud vs. Consistencia
    - Completitud vs. Exactitud
    - Frescura vs. Exactitud / Completitud / Consistencia



# Relaciones entre dimensiones

---

- A veces optamos por favorecer una dimensión en perjuicio de otras.
- Ejemplos:
  - Aplicaciones web
    - Restricciones de tiempo.
    - Se prefiere frescura a exactitud, completitud o consistencia.
    - Ej.: Lista de cursos publicados en una universidad.
  - Aplicaciones administrativas
    - Es más importante la exactitud, completitud y consistencia de los datos, que el tiempo de respuesta .
    - También son mas importantes que la actualidad de los datos, si se considera que son datos que no cambian con una gran frecuencia.

# Bibliografía

---

- **Carlo Batini, Monica Scannapieco. Data and Information Quality. Springer. ISBN: 978-3-319-24104-3. 2016.**
- **Thomas C. Redman. Data Quality for the Information Age. 1996 Artech House Inc., ISBN 0-89006-883-6**
- **G. Shankaranarayanan y R. Blake. From Content to Context: The Evolution and Growth of Data Quality Research. J. Data and Information Quality, vol. 8, n.º 2, p. 9:1–9:28, 2017.**
- **Jack E. Olson. Data Quality. The Accuracy Dimension. Morgan Kaufmann Publishers, Elsevier. 2003. ISBN-10 1-55860-891-5**
- **W. Eckerson. Data Warehouse Institute Survey on Data Quality. Proceedings of the Seventh International Conference on Information Quality (ICIQ-02).**
- **Larry English. The TIQM® Quality System for Total Information Quality Management: Business Excellence through Information Excellence. MIT Information Quality Industry Symposium, 2009.**
- **Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, “AIMQ: a methodology for information quality assessment,” Information & management, vol. 40, no. 2, pp. 133–146, 2002.**

# Bibliografía

---

- **S. E. Madnick, R. Y. Wang, Y. W. Lee, and H. Zhu, “Overview and Framework for Data and Information Quality Research,” J. Data and Information Quality, vol. 1, no. 1, pp. 2:1–2:22, Jun. 2009.**
- **Felix Naumann, Kai-Uwe Sattler. Information Quality: Fundamentals, Techniques and Use. EDBT Tutorial, Munich, 2006.**
- **D. M. Strong, Y. W. Lee, and R. Y. Wang, “Data quality in context,” Commun. ACM, vol. 40, no. 5, pp. 103–110, May 1997.**
- **R. Y. Wang and D. M. Strong, “Beyond accuracy: What data quality means to data consumers,” Journal of management information systems, pp. 5–33, 1996.**
- **M. Scannapieco and T. Catarci, “Data quality under a computer science perspective,” Archivi & Computer, vol. 2, pp. 1–15, 2002.**
- **B. Otto, K. M. Huner, and H. Osterle, “Identification of Business Oriented Data Quality Metrics,” presented at the ICIQ, 2009, pp. 122–134.**
- **Y. Lee, S. Madnick, R. Wang, F. Wang, H. Zhang. A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data. MIS Quarterly Executive, 2014.**