
GESTIÓN Y USO DE COLECCIONES BIOLÓGICAS EN CHILE:

EL CASO DEL CONSORCIO DEL PATRIMONIO BOTÁNICO DE LA UNIVERSIDAD DE CHILE (PBUCH)

Victoria Alejandra Arévalo Rocha
Licenciada en Ciencias Forestales
Universidad de Chile

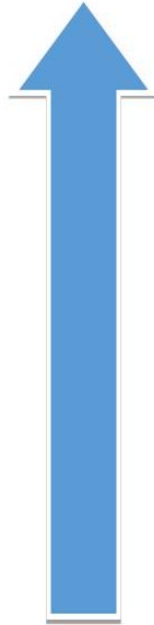
Diciembre, 2018

GESTIÓN DE BASES DE DATOS DE REPOSITARIOS DE HISTORIA NATURAL: RELEVANCIA DE LA CALIDAD DE DATOS EN LAS COLECCIONES BOTÁNICAS

— Victoria Alejandra Arévalo Rocha —
(victoria.Arevalo.r@ug.uchile.cl)

Licenciada en Ciencias Forestales
Universidad de Chile

De los datos a la toma de decisiones



DECISIONES



Conocimiento



Información



DATOS



PERO...EXISTEN VACÍOS DE CONOCIMIENTO



- ✓ Cuántas y cuáles especies existen?
- ✓Cuál es el tamaño de sus poblaciones y dinámicas?
- ✓Cuál es su distribución temporal y espacial?
- ✓Cuántas están siendo afectadas por condiciones bióticas y abióticas?



KEEP
CALM
AND
COLLECT
MORE DATA

1. Necesitamos más datos

- ✓ Regiones pocas estudiadas o representadas
- ✓ Trabajo de campo y Laboratorio
- ✓ Apoyo y financiación científica

2. Necesitamos usar mejor los datos existentes

Una gran cantidad de datos no están disponible para su uso

- No digitalizados, no compartidos
- No fácilmente accesibles
- Problemas de calidad





PRINCIPIO #1

Basura entra → Basura Sale

BASURA ENTRA → BASURA SALE

✓ Problemas de Calidad

Conllevan a resultados de mala calidad, análisis, decisiones, etc.

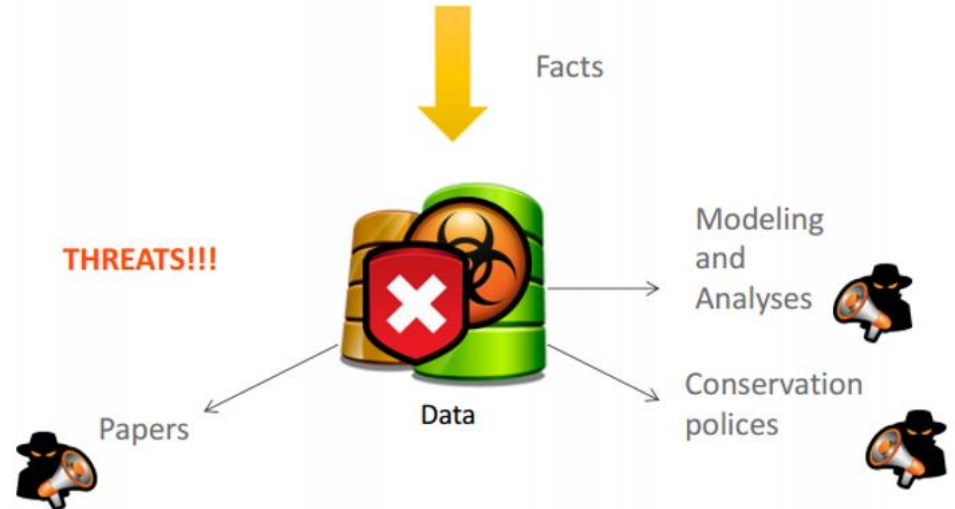


✓ Los problemas surgen de

Toma de datos, digitalización, falta de metadatos, ausencia de estándares.

✓ Hay mucho por hacer

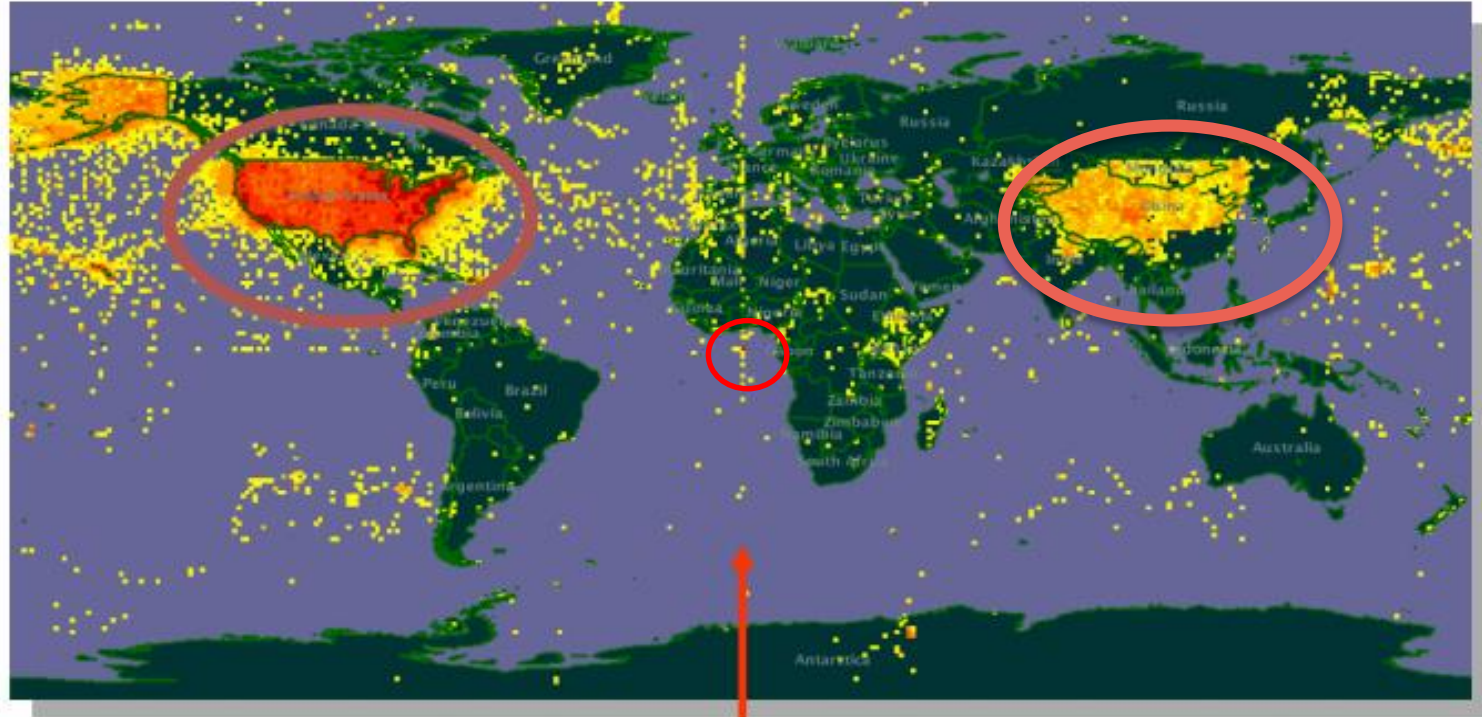
Limpieza de datos (corrección), prevención y políticas de calidad de datos



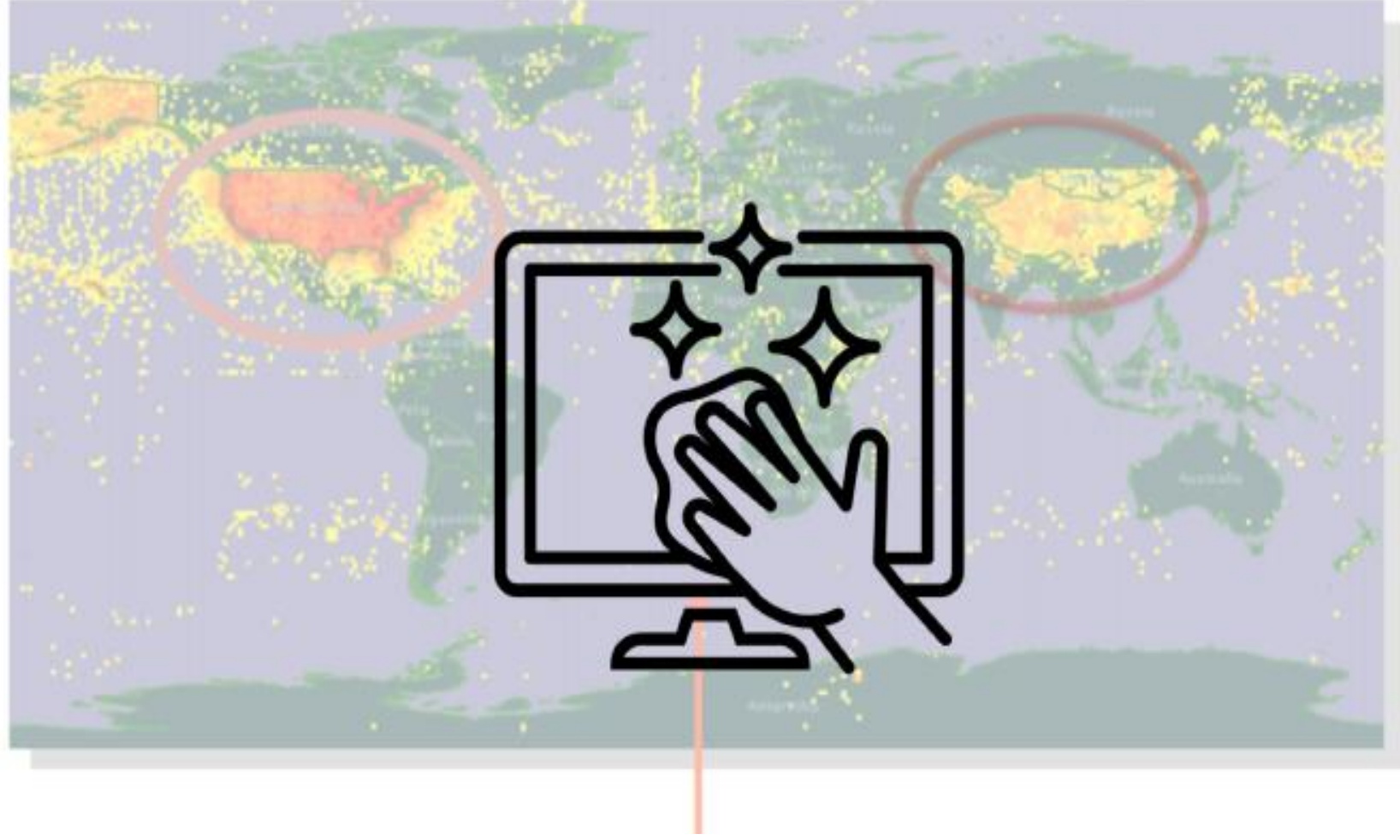
Ejemplo: Problema Calidad de Datos

- USA occurrence data in GBIF database

CALIDAD DE DATOS



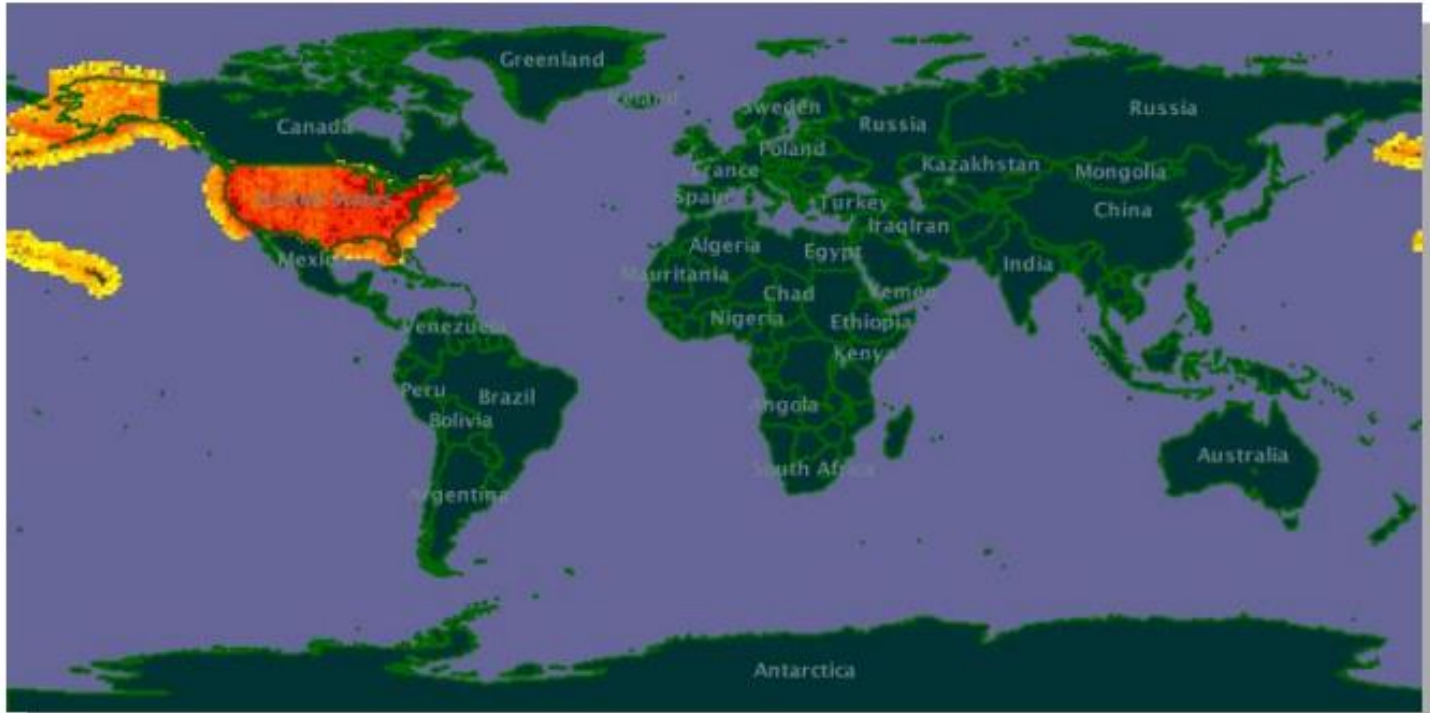
CALIDAD DE DATOS



...luego de la Limpieza de Datos

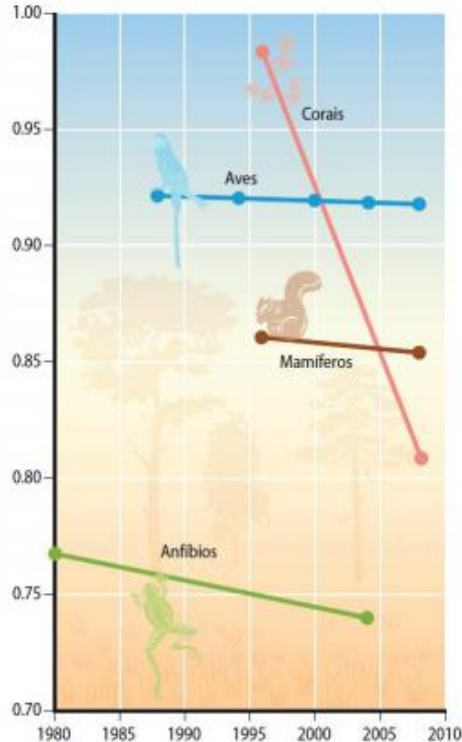
- Wrong longitude (missing minus signal)

CALIDAD DE DATOS

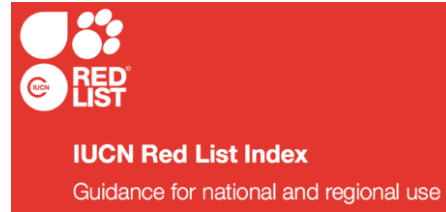
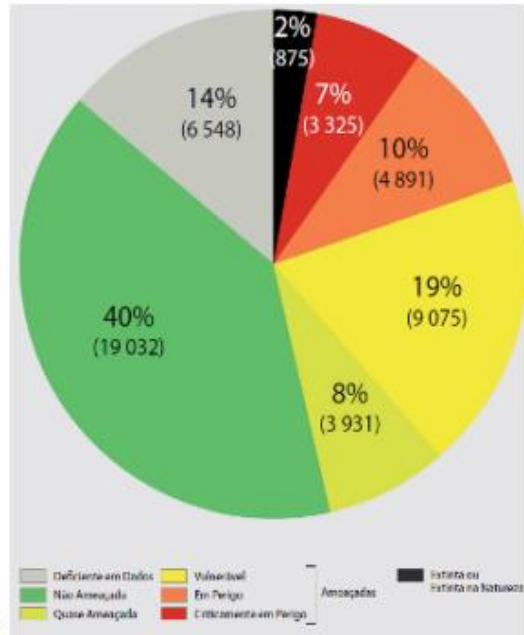


La calidad de datos puede afectar los indicadores, análisis, toma de decisiones y políticas

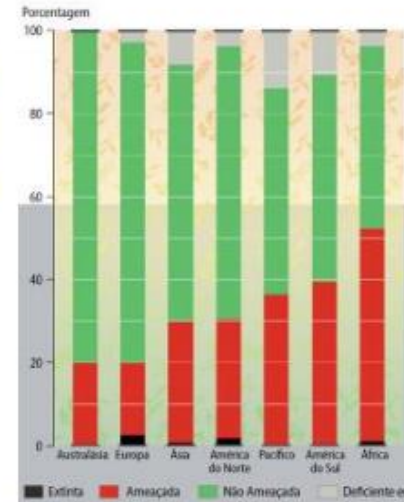
Red List Index



Threatened species - IUCN



Conservation status of medicinal plants



Algunas precisiones

- ✓ **Información:** un DATO con significado
- ✓ La **información** es una “REPRESENTACIÓN de la realidad”
- ✓ Existe un **vacío** entre la **representación de la realidad** y la **realidad**
- ✓ Este **vacío** puede ser medido a partir de algunos atributos (llamados dimensiones), tales como:

Compleitud

Precisión

Consistencia

Exactitud

etc.



CD Definición #1: Conformidad con los hechos

Los **DATOS** tienen **CALIDAD** si
la **INFORMACIÓN** derivada de
los datos representa
correctamente el **MUNDO
REAL** (HECHOS)



Información

Datos

Mundo real



Ejemplo



DATO 1

Nombre: *Saguinus*

DATO 2

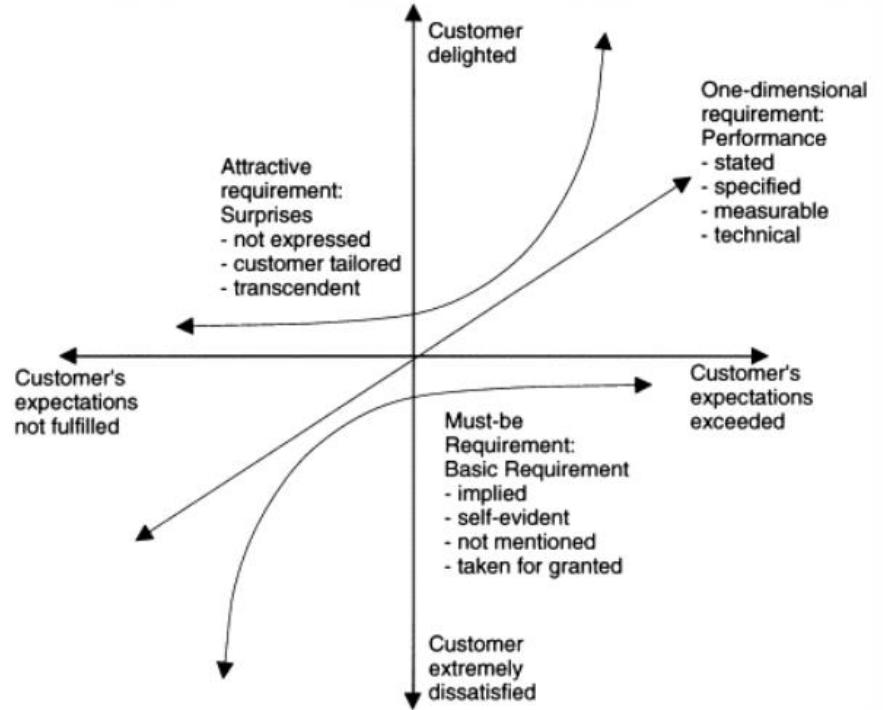
Nombre: **Mico tití**

Los datos 1 y 2
REPRESENTAN LA
REALIDAD / tienen la
misma calidad ??

CD Definición #2: Conformidad con los requisitos

Satisfacción del consumidor

SI un consumidor esta satisfecho con un servicio/ producto, este servicio o producto tiene calidad para este consumidor



Requisito:

Los datos deben tener un nombre científico y debe ser suministrado a nivel de especie.

DATO 1:

Nombre científico: *Saguinus*

Categoría taxonómica: Género

Este dato tiene calidad ??



¿Puede ser usado en el estudio de la distribución de primates en Sudamérica?

CD Definición #3: Los Datos deben ser adecuados al USO

[*Fitness-for-USE*]

Usabilidad. Un dato tiene calidad si es adecuado para ser usado. Si el dato no sirve para el propósito del que lo usa.



Pueden ser **ÚTILES** para otros!

CD Definición #3: Los Datos deben ser adecuados al USO

[Fitness-for-USE]

Aptitud depende de que los datos sean:

- Accesibles
- Oportunos
- Fáciles de leer
- Fáciles de interpretar
- Relevantes
- Consistentes
- Completos
- Exactos
- Detallados
- Exhaustivos



PRINCIPIO #2

Calidad de Datos, concepto idiosincrático

PRINCIPIO #2. La Calidad de Datos es un concepto IDIOSINCRÁTICO

Definición: “Idiosincrático es algo distintivo y propio de un individuo”

Definir qué es **Calidad de Datos** es como definir qué es Bonito, Bueno, Alegre o Valioso, por ejemplo.



ADECUACIÓN AL USO de los datos

Fitness-for-Use of Data



KEYWORDS

La palabra clave y la definición mas aceptada
Usabilidad de los datos

USO: La Calidad se relaciona con un propósito

“Ej. Modelo de distribución, Lista nacional de especies, Monitoreo de la biodiversidad, etc”

DATO: Para cada propósito específico existe un conjunto de tipos de datos que deben tener calidad para ser usados.

“Ej. Para la construcción de Modelos de distribución las coordenadas y los nombres de las especies deben tener calidad”

ADECUACIÓN: Para cada tipo de dato existe un conjunto de atributos de calidad que los datos deben tener para responder a un propósito específico.

“Ej. Completitud, consistencia, exactitud, precisión y credibilidad”

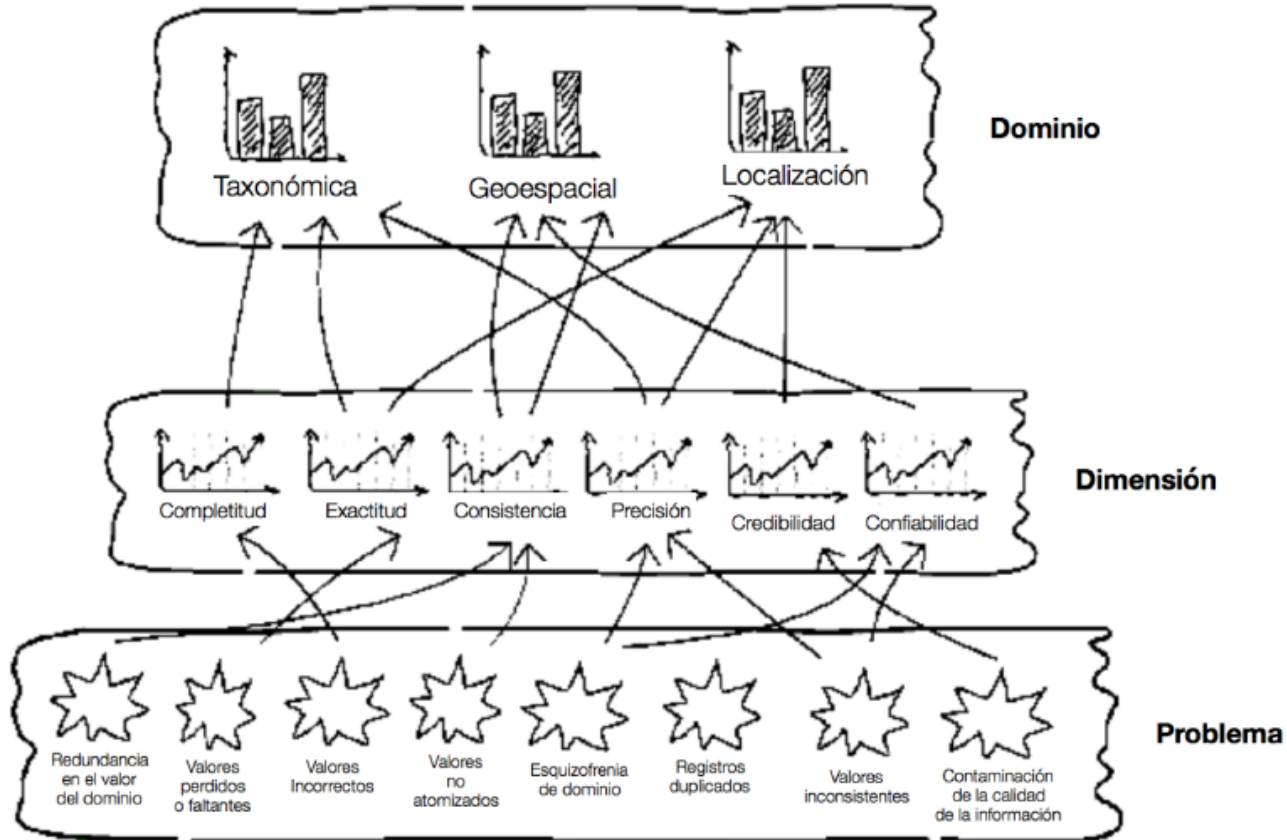


PRINCIPIO #3

Los datos tienen calidad si no existe ningún problema que la degrade

Evaluación de la Calidad de Datos

CALIDAD DE DATOS



Principales Dominios de Calidad de Datos



Geo-espacial: Datos con coordenadas o georreferenciados y regiones político-administrativas documentadas.



Taxonomía (nomenclatura): Nombres científicos, niveles taxonómicos



Formato (Estandarización): Errores de tipeo, formatos de fecha, formatos de coordenadas, caracteres especiales y demás.

Problemas

- ❖ **Redundancia:** Sin estándares, sinónimos
- ❖ **Valor Faltante:** Valor ausente
- ❖ **Valor Incorrecto:** Valor no representa los hechos
- ❖ **Valor no atomizado:** Más de un valor en un campo atomizado
- ❖ **Esquizofrenia:** Campo utilizado con un fin incorrecto
- ❖ **Registros duplicados:** Mas de un registro representando el mismo hecho
- ❖ **Valores Inconsistentes:** Incongruencia en la representación de un hecho
- ❖ **Contaminación:** Creación de nuevos datos a partir de datos existentes incorrectos

Cadena de la información

Flujo de Publicación Datos de Biodiversidad

COSTO DE LA CORRECCIÓN DE ERRORES



1



2



3



4



5

Mecanismos de mejora en la CD



PREVENCIÓN: Evitar que se presenten errores previo a la creación de los datos.



DETECCIÓN Y LIMPIEZA: Detectar errores en el conjunto de datos y corregirlos.



DETECCIÓN Y RECOMENDACIONES: Detectar errores en el conjunto de datos y generar recomendaciones de limpieza.

GESTIÓN Y USO DE COLECCIONES BIOLÓGICAS EN CHILE:

**EL CASO DEL CONSORCIO DEL PATRIMONIO BOTÁNICO DE LA
UNIVERSIDAD DE CHILE (PBUCH)**

Victoria Alejandra Arévalo Rocha
Licenciada en Ciencias Forestales
Universidad de Chile

Muchas gracias por su atención