

The logo for inetum, featuring the word "inetum" in a white, lowercase, sans-serif font, followed by a small teal square icon with a white arrow pointing up and to the right.

inetum

Joseba Ruiz
Data & AI Director
Inetum Norte

Calidad de los datos

en Espacios de datos

Calidad del dato. Dimensiones clave

PRECISIÓN

¿Los datos reflejan la realidad y están libres de errores?

COMPLETITUD

¿Está presente toda la información necesaria?

CONSISTENCIA

¿Se persiste un mismo dato en diferentes sistemas?

SINGULARIDAD

¿Existen datos duplicados?

ACTUALIDAD

¿Se reflejan los datos necesarios más recientes?

VALIDEZ

¿El negocio se ajusta a los datos?

Retos. **Precisión**

Verificar exactitud y origen del dato en entornos federados y distribuidos.



Los datos imprecisos son como Loki: aparentan ser correctos, pero esconden engaños y distorsionan la realidad.

CONFIANZA EN EL DATO COMPARTIDO

- En un ecosistema federado, garantizar que el dato refleje la realidad de forma exacta es complejo.
- Se requiere verificar métodos de origen, contexto de captura y errores sistemáticos.
- Se plantean mecanismos como validaciones automáticas, smart contracts de calidad y certificaciones de precisión.
- La interoperabilidad semántica es clave para que el significado de “preciso” se mantenga entre dominios.

Retos. **Complejidad**

Unificar datos fragmentados
asegurando cobertura
suficiente para cada uso.

Como Slender Man, sin rostro, los datos incompletos parecen humanos pero les falta lo esencial.

HETEROGENEIDAD FUENTES Y FORMATOS

- La integración de datos fragmentados, con estructuras distintas y niveles de detalle variables, genera lagunas en la información.
- Es crítico definir esquemas mínimos comunes y plantillas de datos federadas.
- La completitud relativa al caso de uso debe ser evaluada dinámicamente, adaptando las exigencias al contexto.

Retos. Consistencia

Mantener coherencia de datos distribuidos sin control centralizado.



Un dato puede tener dos versiones contradictorias, como la moneda de Dos Caras.

GOBERNANZA DISTRIBUIDA

- Sin un control centralizado, asegurar que los datos mantengan coherencia entre diferentes nodos es difícil.
- Se requieren mecanismos de sincronización, reglas de validación compartidas y acuerdos de interoperabilidad entre participantes.
- Las políticas de consistencia deben respetar la soberanía del dato, alineando validaciones entre dominios con semántica compartida.

Retos. Singularidad

Evitar duplicidades a gran escala con identificadores únicos y resoluciones automáticas.



Como Smith, los datos duplicados invaden el sistema y crean caos en los procesos.

ESCALABILIDAD CONTROLES DE CALIDAD

- En espacios con millones de registros de múltiples fuentes, evitar duplicados o identificaciones incorrectas es crítico.
- Se necesitan algoritmos de entidad resolutiva automática, uso de identificadores persistentes y controles on-the-fly.
- Las soluciones deben ser escalables y ejecutables en tiempo real, especialmente en contextos edge/cloud distribuidos.

Retos. Actualidad

Asegurar que los datos estén actualizados según contexto y frecuencia requerida.



Los datos desactualizados son como Drácula: parecen relevantes, pero están anclados en el pasado.

EVALUACIÓN DINÁMICA DE LA CALIDAD

- La calidad temporal depende del tipo de dato y su caso de uso (real-time, daily, batch).
- Es necesario implementar indicadores de frescura y políticas de actualización, junto con alertas ante obsolescencia.
- La evaluación dinámica permite ajustar la vigencia esperada del dato según proveedor y contexto.

Retos. Validez

Verificar que el dato cumple reglas, normativas y trazabilidad en todo su ciclo.



Como el Demogorgon, los datos inválidos vienen de otra dimensión que rompe con las reglas del modelo.

TRAZABILIDAD Y CERTIFICACIÓN

- Validar que un dato cumple con normativas, esquemas y contextos específicos requiere metadatos enriquecidos, trazabilidad y mecanismos de certificación.
- Se plantean sistemas de etiquetado de calidad, logs auditables y uso de verificadores distribuidos.
- Los smart contracts y catálogos de datos interoperables ayudan a gestionar la validez de forma automatizada y transparente.

De la teoría a la práctica ...

Desarrollo propio. Data Observability

Funcionalidad desarrollada por Inetum para ser integrada dentro de Data Spaces y que pretende garantizar la Calidad de datos mediante el uso de GEN-AI.

inetum.
FIN

Joseba Ruiz
Data & AI Director
Inetum Norte



inetum.com