



UNIVERSITAT DE
BARCELONA

HerStory (2024-27)



Revitalización de datos en silos mediante IA neurosimbólica y grafos de conocimiento: una propuesta para la sostenibilidad de los datasets científicos

Miquel Centelles; Núria Ferran-Ferrer; Marina Salse y Elena Gómez
Barcelona, 19 de febrero de 2026



CoDi

Congreso de Divulgación,
Transferencia e Impacto Social
de la Ciencia



<http://creativecommons.org/licenses/by-nc-sa/3.0/es>

Sumario

Planteamiento

Objetivos

Solución

Metodología

Conclusiones

Planteamiento:

Cuando el dataset se apaga

- *Año 1: publicamos el portal. Año 3: el servidor se apaga.*
- El dataset no se pierde del todo, pero desaparece para la comunidad científica.
- Eso es un silo: difícil de encontrar, frágil, poco interoperable.
- Resultado: alto coste → bajo retorno público.



Objetivos: Revitalizar datos para que sean explicables y reutilizables

Proponer y ejemplificar una solución de IA neurosimbólica, explicable y replicable, para revitalizar datasets ya existentes.

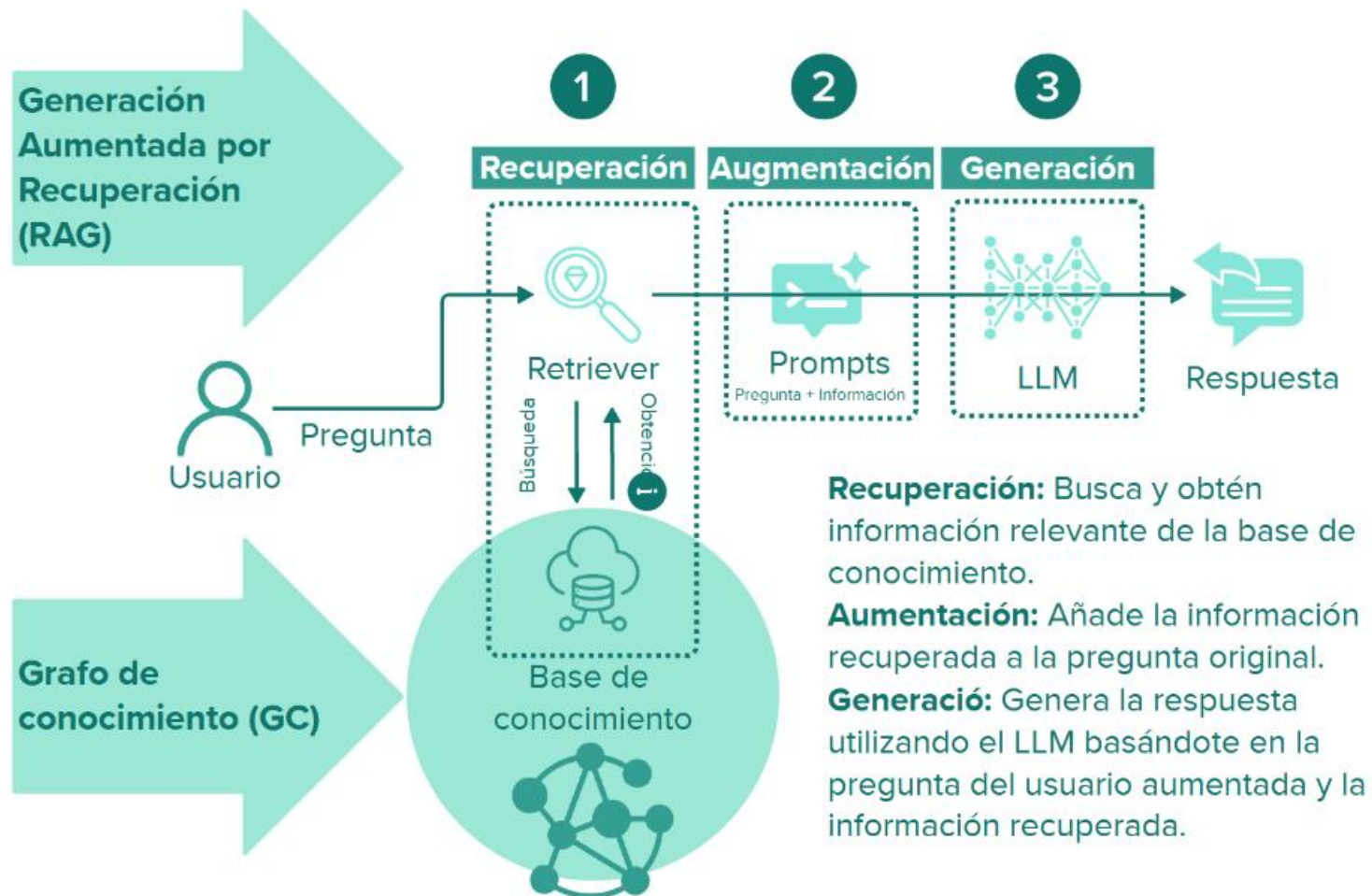


Definir un método para identificar y describir datasets en riesgo de convertirse en silos, con evidencias FAIR y CARE.

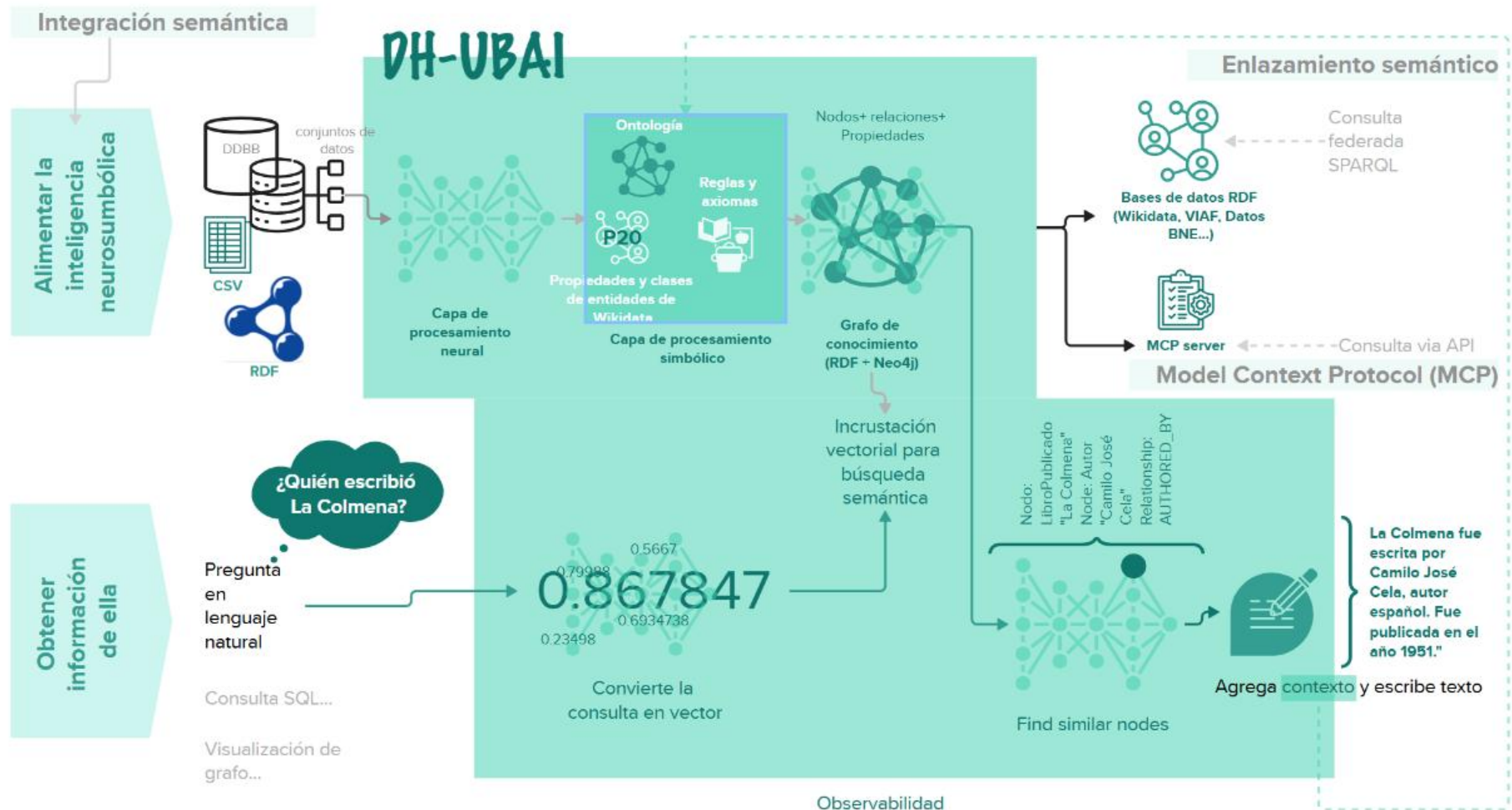


Mostrar cómo integrar un dataset ya creado en un grafo de conocimiento, con trazabilidad y validación.

Solución: IA neurosimbólica con RAG y grafo de conocimiento



Conclusión: Tres vías de conexión según el estado de salud del dataset



Metodología: Identificación y descripción de datasets

¿Cuál es el alcance de la identificación de datasets?

- A medio plazo identificamos datasets generados en proyectos de Humanidades UB desde 2016.
 - *Número de proyectos de investigación: 731.*
- A corto plazo analizamos datasets generados en proyectos del Hub de Humanidades Digitales.
 - *En el Hub ya hemos identificado 85 datasets.*

¿Cómo llegamos a los datasets?

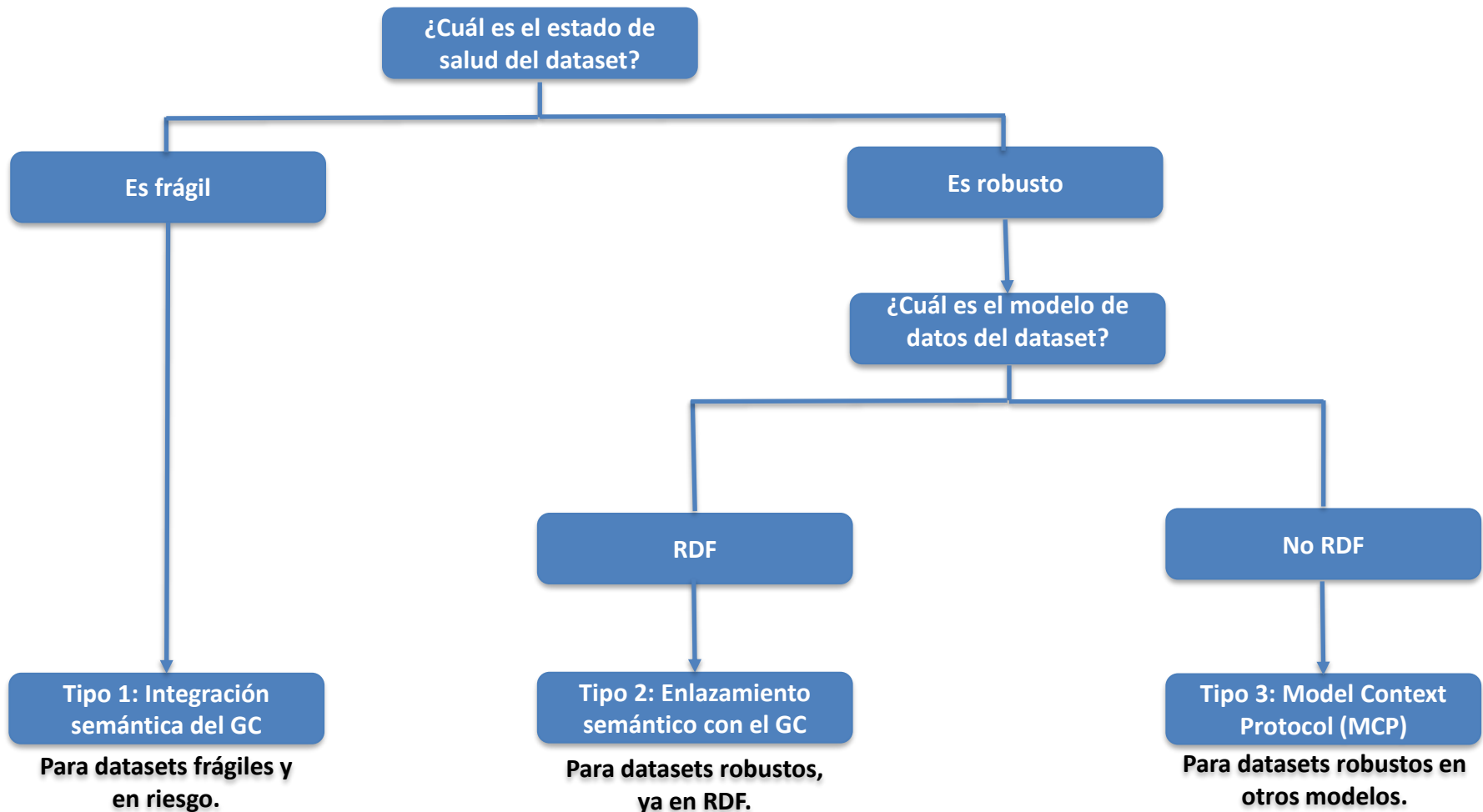
- La guía principal es el plan de gestión de datos (PGD)* y documentos equivalentes en los proyectos de investigación.
- Localizamos estas guías en memorias y auditorías de proyectos.
- También consultamos repositorios, plataformas y responsables de proyectos.

**Importante: La exigencia de PGD es reciente y no siempre se verifica su ejecución.*

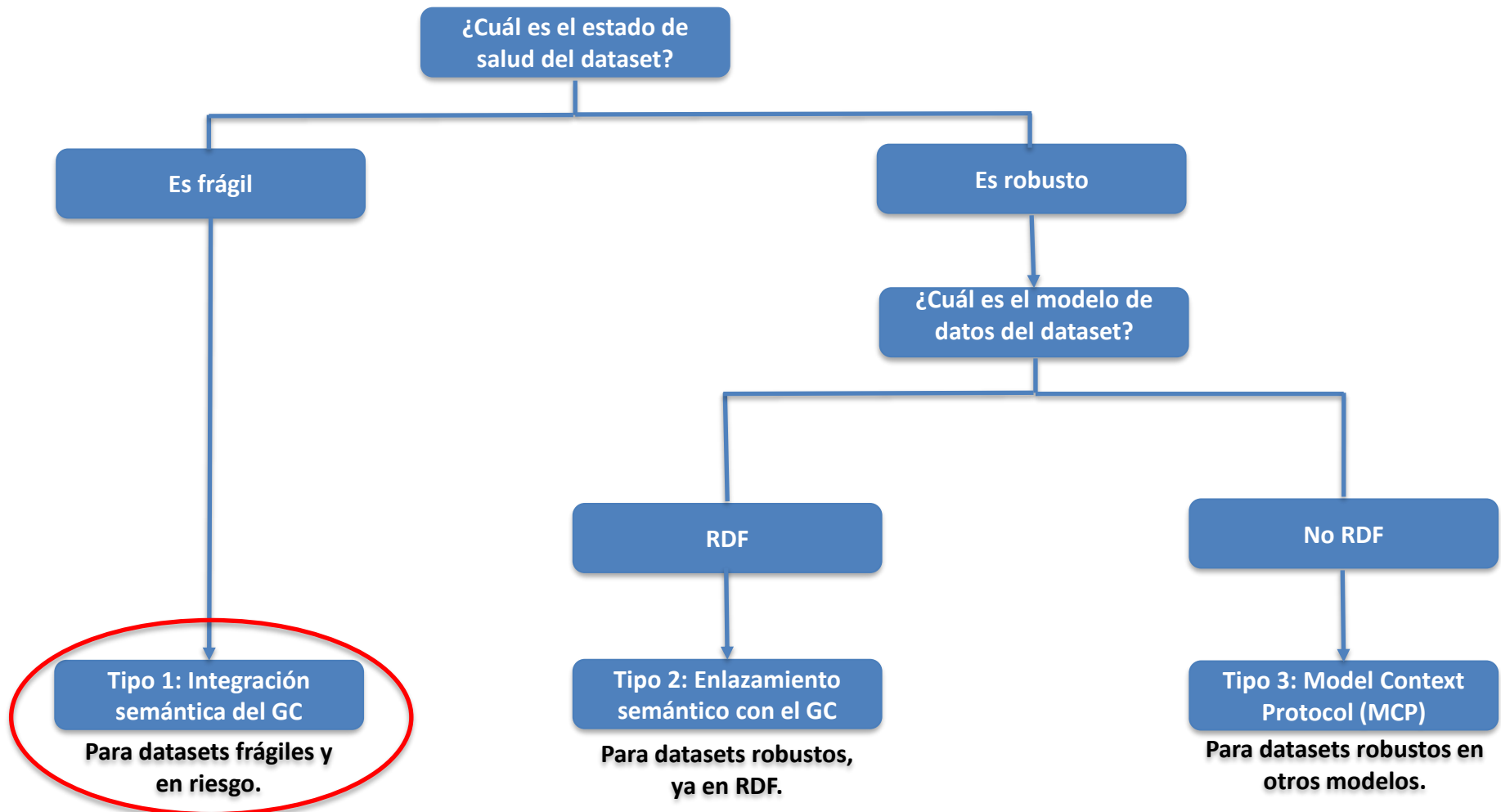
¿Qué nos interesa de los datasets?

- Describimos cada dataset con evidencias FAIR y CARE verificables.
- Comprobamos si hay depósito en repositorio o plataforma estable.
- Comprobamos si existe un identificador persistente para citar.
- Comprobamos metadatos y documentación que expliquen los datos.
- Comprobamos la licencia y las condiciones de acceso y reutilización.

Metodología: Clasificación del dataset y tipo de conexión con DH-UBAI



Metodología: Clasificación del dataset y tipo de conexión con DH-UBAI



Metodología: Integración semántica trazable en seis pasos



1. Analizamos el dataset original, su estructura, acceso y licencias.



2. Modelamos entidades y relaciones como componentes RDF del grafo de conocimiento.



3. Transformamos el dataset original a RDF, con identificadores estables y alineación con la ontología del grafo de conocimiento (basada en Wikidata).



4. Aplicamos RAG sobre los datos RDF integrados para recuperar información de una misma entidad desde distintos datasets y generar respuestas basadas en nodos y relaciones, con fuentes trazables.



5. Validamos los datos RDF generados con cocreación humana y procedimientos automáticos.



6. Publicamos los datos en RDF de forma sostenible (repositorios/plataformas de datos abiertos) con cumplimiento de principios FAIR y CARE.

Conclusiones: De silo frágil a activo preservable y conectable



Revitalizar no es “poner IA”; es **preservar, conectar y explicar con evidencias**.



Primero **diagnosticamos** y elegimos vía A, B o C, y si es frágil aplicamos **integración trazable en seis pasos**.



Es conocimiento transferible porque deja un resultado **auditable**: evidencias FAIR y CARE, clasificación de riesgo y trazabilidad en las respuestas.



Cuando el portal se apaga, el dato muere; cuando el dataset se preserva y se conecta, los datos se revitalizan.

Gràcies!

¡Gracias!

Thank you!

HerStory (2024-27)



**Your comments,
suggestions,
proposals...**

