



Datos no son petróleo. Son lenguaje.

Paul Fervoy

Club de Investigación Tecnológico · Costa Rica



Paul Fervoy

[Verify now](#)

CEO, Siftia | Vicepresidente ALETI | Docente, ULEAD | Presidente Honorario, CAMTIC

San Pedro, San Jose, Costa Rica · [Contact info](#)

[5,222 followers](#) · [500+ connections](#)



siftia



University of Wisconsin-Madison

<https://www.linkedin.com/in/paulfervoy/>



Comisión Nacional de Datos Abiertos

Gobierno · Industria · Datos públicos

4 horas de reunión.

3 horas de reclamos.

“Años pasan y aún los datos no se comparten.”

“Ciudadanos se quejan de respuestas que llegan tarde con excusas...”

“...y de datos que no se pueden usar.”



“ Los datos son el
nuevo petróleo.
Lo que necesitamos
es más datos. ”

01100101 11010101 10010101 11010101
11101112 -00111 -2 3 1000111101 715 01010
100100010 0100100 Σ 11001000110 101 1111 00101

Se solicitó el índice de cáncer de mama al Ministerio de Salud.

Casos	Indice	Base	Acumulado	Ubicación
4823	1.2	0.87	3401	001
1204	0.9	-0.43	8820	047
7710	2.1	0.62	2190	003

“¿Alguien entiende qué miden estas columnas?” nos preguntó

Los datos llegaron.

El significado no.

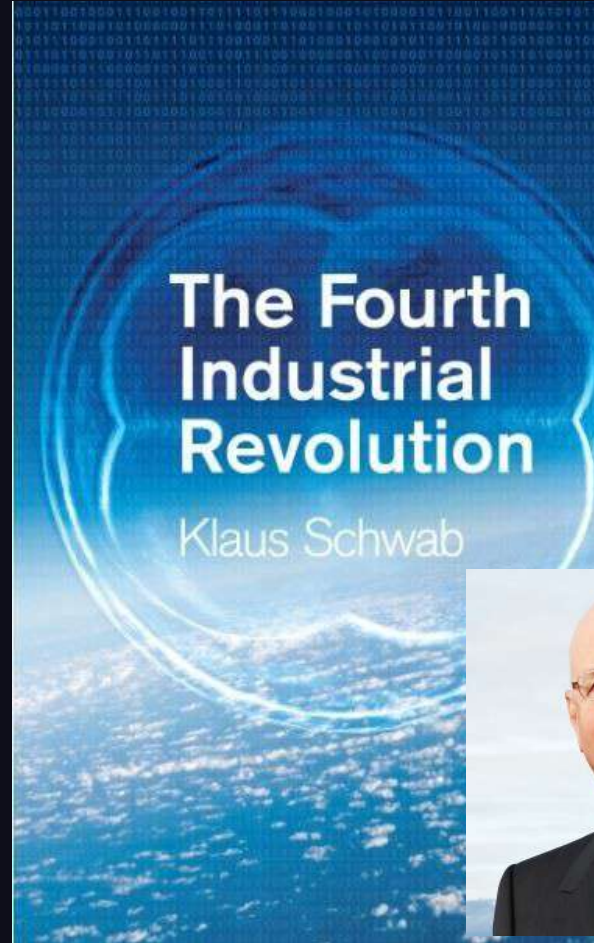
**Datos no son
petróleo.
Son lenguaje.**

2016

Klaus Schwab

Foro Económico Mundial

Cuarta Revolución Industrial



La idea era simple:

tres tecnologías convergían al mismo tiempo.

 **Datos** +  **Nube** +  **IA**

Y fue en ese momento cuando empezó a popularizarse: ***"Los datos son el nuevo petróleo."***

Los datos no son petróleo. Son lenguaje.



Libro: Wizard of Earthsea

La magia no consiste
en lanzar hechizos.

**Consiste en conocer
el nombre verdadero
de las cosas.**

Los datos funcionan igual.

*Si no acordás el nombre verdadero,
dependés del nombre que otro decidió por vos.*

Y como sociólogo...

Me dejó una preocupación:

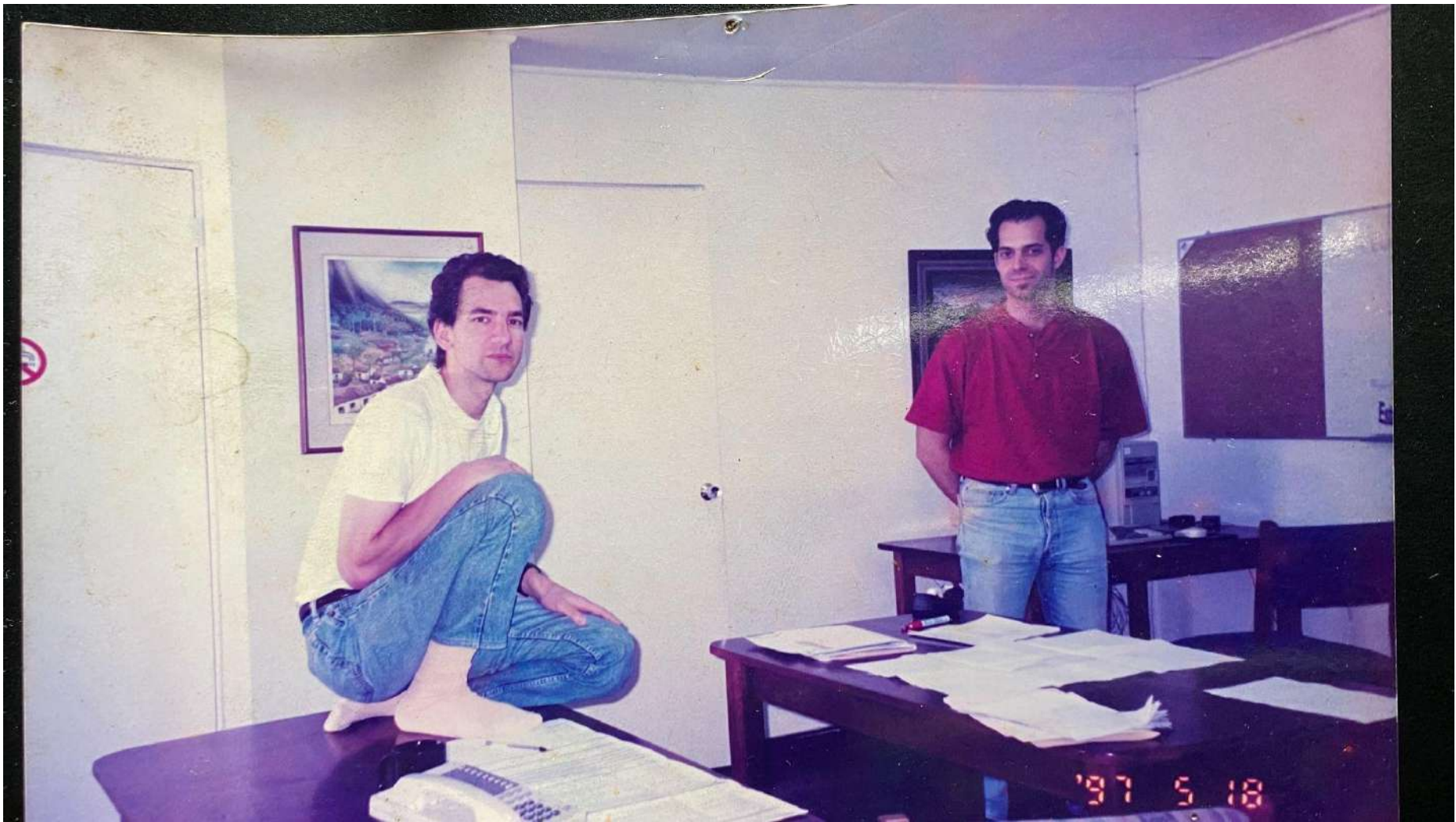
**cómo los humanos nos relacionamos
y cómo la tecnología y datos son una representación.**

Llevo 30 años haciendo el mismo trabajo con distintos nombres. Tres momentos:

**Arquitectura de
información - 1990s**

Análisis de Datos - 2000s

Gobernanza de datos · hoy

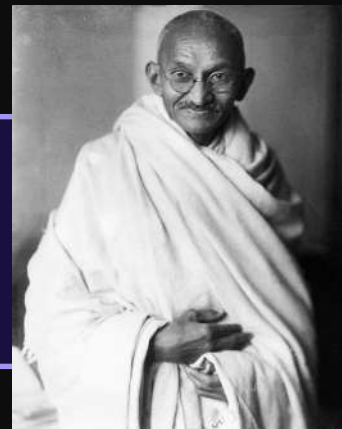


11 de septiembre, 2001.

En MetaBase.net revisé los logs.

La búsqueda más frecuente:

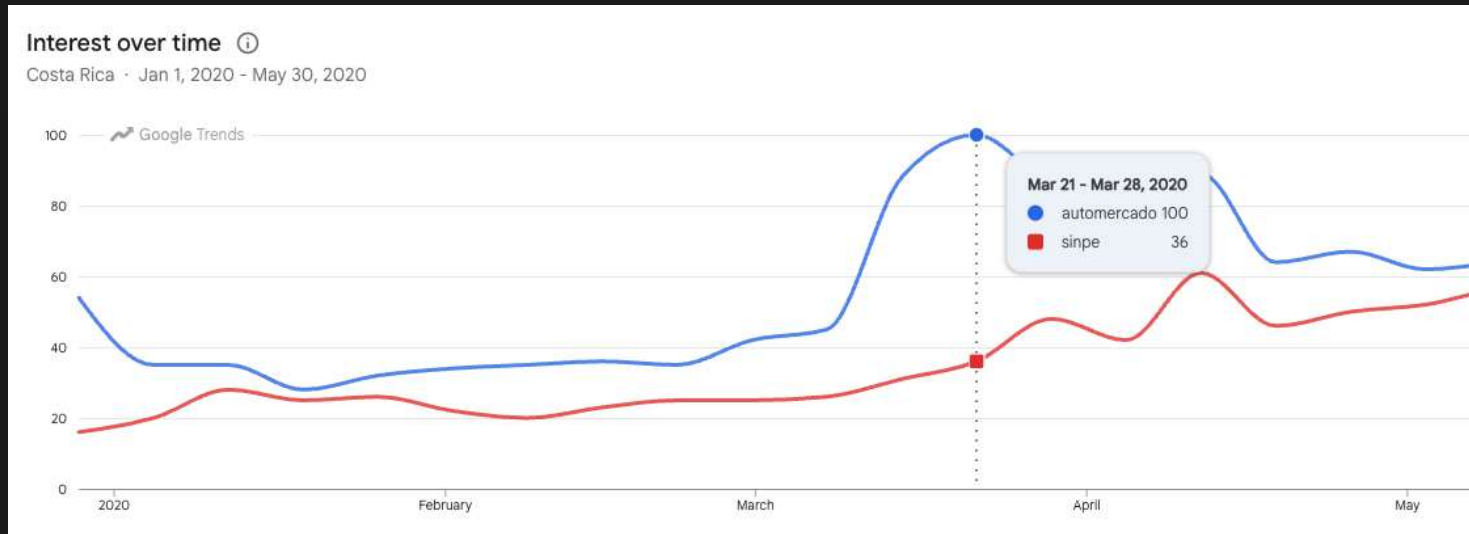
Gandhi



Un dato sin contexto es una palabra. Con contexto, es un síntoma.

Veinte años después, marzo 2020

Google Trends. Semanas antes del lockdown de COVID-19.



La sociedad se preparaba para el lockdown antes de que nadie lo declarara.

Hoy “*Customer Journey*” de un cliente del sector automotriz.

Marketing:

Audiencia / Segmento

Ventas:

Prospecto calificado

Postventa:

Propietario de vehículo

CLIENTE

Finanzas:

Quien firma el contrato

CRM:

Combinación de
términos de la config.

Dirección Ejecutiva:

KPI de conversión

No faltaban datos. Faltaba acuerdo.

16,000 sistemas de información.

Muchos con un campo relacionado con: "fumador"

Hospital A	12	cigarrillos/día	Clínica B	1.5	paquetes/día
Centro C	84	cigarrillos/ semana	Sistema D	1	(binario: sí/no)

Mismo campo. Mismo nombre. Cuatro *idiomas* distintos.

¿Cuántos veces celebramos una integración exitosa...
y el negocio siguió pidiendo el Excel igual?

Integrar

Los bits llegaron.
Formato correcto.
Sin errores técnicos.

Interoperabilidad

El significado llegó.
Ambos extremos
entienden lo mismo.

Eso no es interoperabilidad. Es integración. Y no son lo mismo.

¿La solución?

Meter IA y cerrar los ojos.

La IA no resuelve el problema semántico.

Lo automatiza.

Y si falla, falla a escala.



Un ejemplo concreto.

"femenino"

1

Históricamente: cuando RH contrataba, 'femenino' gana menos puestos.

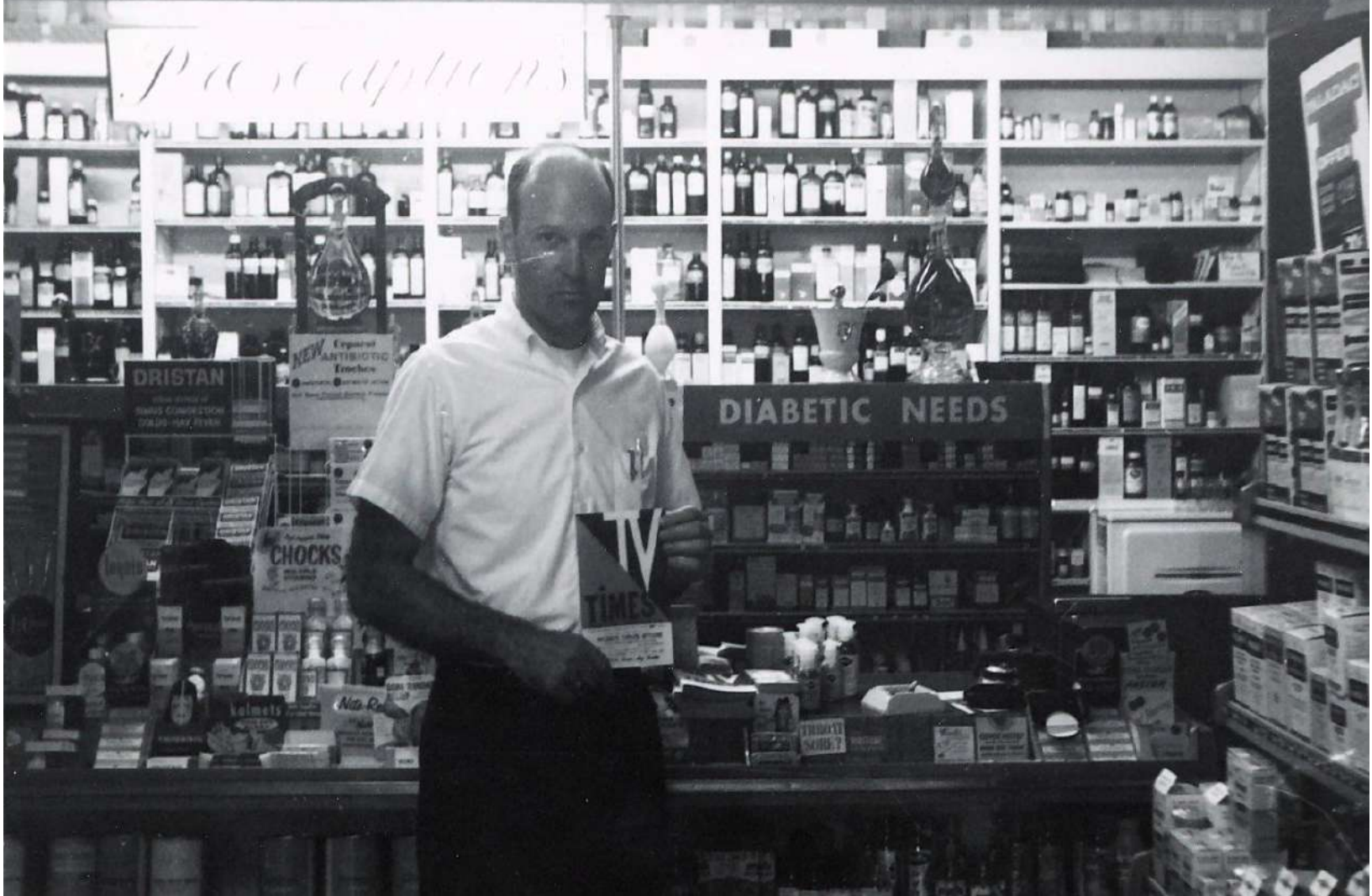
2

La IA entrena sobre ese historial. Aprende: 'femenino' = menor probabilidad.

3

No es sexismo por diseño. Es semántica heredada, automatizada a escala.

Prescriptions



DRISTAN

NEW Cephal
ANTIBIOTIC
Emches

CHOCKS

DIABETIC NEEDS

¿Qué pasa cuando el lenguaje sí está acordado?

FHIR + SDOH

Farmacias independientes (EEUU) acordaron capturar campos SDOH — determinantes sociales de salud — en sus conversaciones con pacientes.

La hipótesis era clara: los determinantes sociales predicen la adherencia mejor que los clínicos.

El hallazgo diseñado

Personas de bajos ingresos con mascotas en casa cumplían peor sus regímenes de medicamentos.

Priorizaban la salud de su mascota sobre la propia.

Solución: extender coberturas de seguro para incluir salud animal.

El dato no estaba escondido. Estaba esperando que el lenguaje lo dejara aparecer.

Hay organizaciones que decidieron algo diferente.

01

Nombraron el problema

Lo pusieron en agenda como tema de definiciones, no de tecnología.

02

Asignaron responsabilidad

Alguien responde por el significado, no solo por la infraestructura.

03

Crearon proceso de arbitraje

Cuando hay conflicto de definiciones, se resuelve formalmente.

04

Midieron lo correcto

Confianza. Menos reportes paralelos. Menos discusiones sobre números.

Crearon “guías semánticas” del negocio. Eso no es burocracia. Es madurez organizacional.

Cuando la industria le pone nombre a algo,

el problema es real.

OSI — Open Semantic Interchange

open-semantic-interchange.org · 2025

"A vendor-neutral single source of truth for your semantic data."

- Snowflake
- Databricks
- Salesforce
- Mistral AI
- BlackRock
- Collibra
- Informatica
- dbt Labs
- Qlik
- + 30 más

Costa Rica hoy.

FHIR Core Guides · 2024

El Ministerio de Salud publicó las primeras guías formales de interoperabilidad para Costa Rica.

Primer intento de acordar el lenguaje antes de conectar los sistemas.

Comité Técnico Salud Digital

El debate más difícil no es técnico.

Es semántico: ¿qué significa 'paciente' en la CCSS vs. el sistema privado? ¿Qué es un 'episodio de atención'?

Misma tensión que Cataluña — diferente escala.

La infraestructura está avanzando. El acuerdo semántico va detrás.

Lo mismo pasa cuando trabajamos con IA.

El spec es el lenguaje.

El problema

La IA produce código que compila sin errores — pero no hace lo que el negocio necesita.

Nunca hubo un error técnico. Hubo un error semántico en el spec.

La causa

Spec vago = lenguaje del dominio no acordado.

La máquina hizo exactamente lo que le dijiste. Pero no era lo que querías decir.

La solución: spec-driven design

Antes de escribir código, acordamos el lenguaje del dominio:

qué significa cada término, qué relaciones existen, qué está en scope.

Después le pasamos eso a la IA.

Es arquitectura de información. La herramienta cambió. El problema, no.

El cuello de botella ya no es la tecnología.

Es el significado.

Antes de enseñarle a una máquina a hablar,

**tenemos que ponernos de acuerdo
sobre qué queremos decir.**

La próxima vez que alguien les proponga integrar sistemas, háganle una pregunta:

“¿ya acordamos qué significa esto que vamos a intercambiar?”

Datos no son petróleo.

Son lenguaje.