

Ciencia de Datos para la Toma de Decisiones II

Mtro. René Rosado González
Director de Programa LTP

¿Qué es un dato?

data
data
data
data
data
data
SMALL ~~big~~ data
data bandwidth ~~QUALITY~~
IMPERFECT ~~infallible~~ data
SUBJECTIVE ~~impartial~~ data
INSPIRING ~~descriptive~~ data
SERENDIPITOUS ~~predictive~~ data
data ~~conventions~~ POSSIBILITIES
data to simplify complexity / ~~DEPICT~~
data processing ~~DRAWING~~
data driven design
SPEND ~~save~~ time with data
data is numbers ~~PEOPLE~~
data will make us more efficient ~~HUMAN.~~

a manifesto for a DATA HUMANISM

a new renaissance where we
can question the impersonality
of a merely technological approach
to DATA, where we are ready to
REconnect numbers to what
they really stand for: which
ARE - MORE and MORE - OUR
UNIQUE LIVES.

giorgia lupi

Una expresión de deseo por
almacenar una vivencia



Fuente: World Press Photo (2020)

World Press Photo 2020



Fuente: World Press Photo (2020)

Metadatos

Título: *Straight Voice*

Autor: *Yasuyoshi Chiba*

Fecha: 19 de junio de 2019

Lugar: Khartoum, Sudan

Publicación: Agence France-Presse

World Press Photo 2020



Fuente: World Press Photo (2020)

Datos Cuantitativos

¿Cuántas personas hay?

¿Cuántas están sonriendo?

¿Cuántos son hombres y cuántas son mujeres?

¿Cuántas cámaras hay?

World Press Photo 2020



Fuente: World Press Photo (2020)

Datos Cualitativos

¿Qué están diciendo?

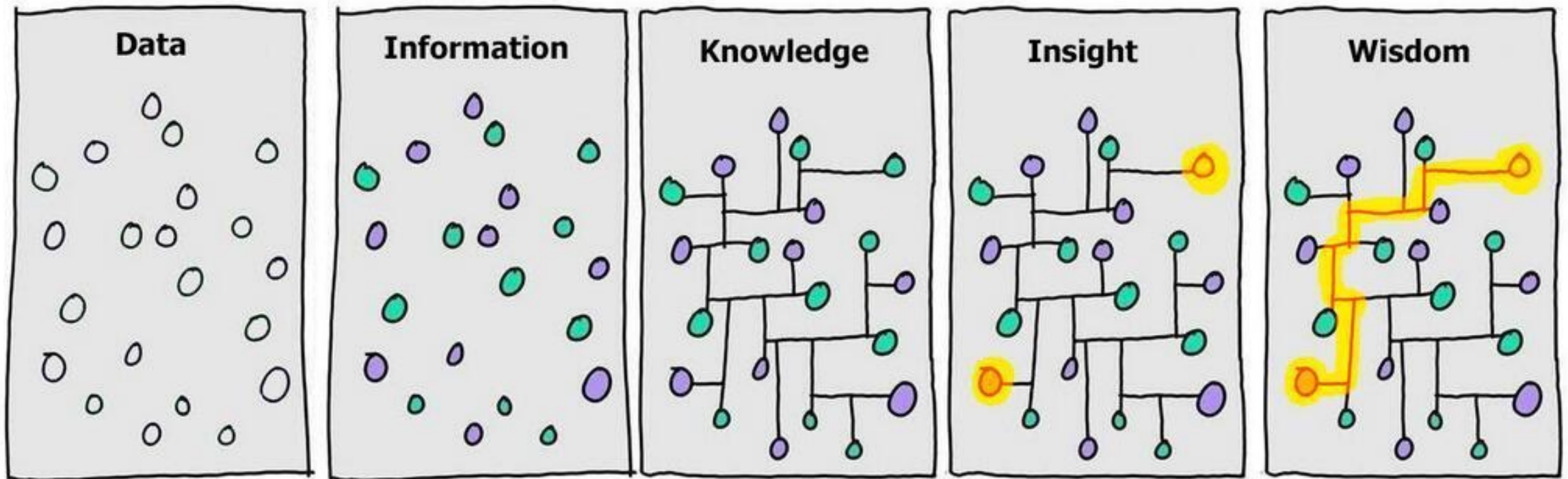
¿Quién lo está diciendo?

¿Cómo lo está diciendo?

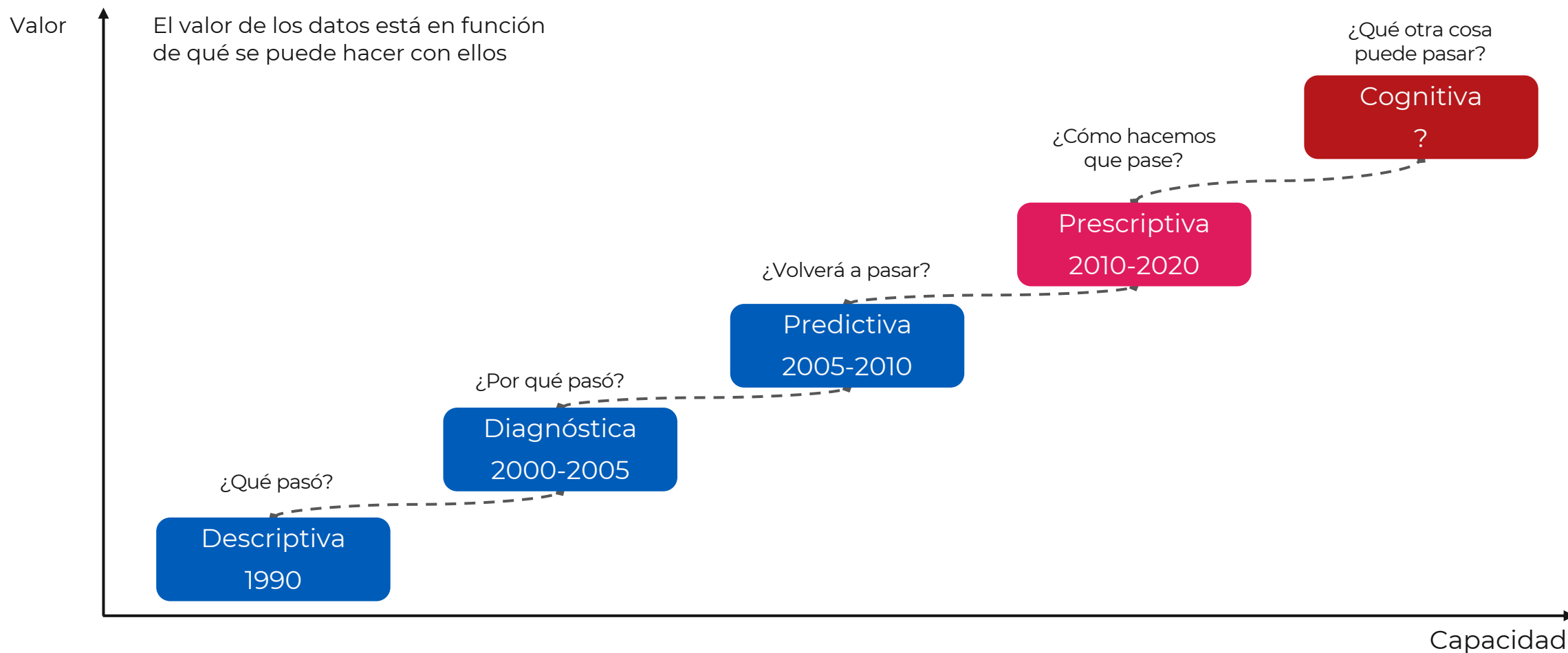
¿Por qué lo está diciendo?

¿Dónde lo está diciendo?

¿Qué hace la ciencia de datos?

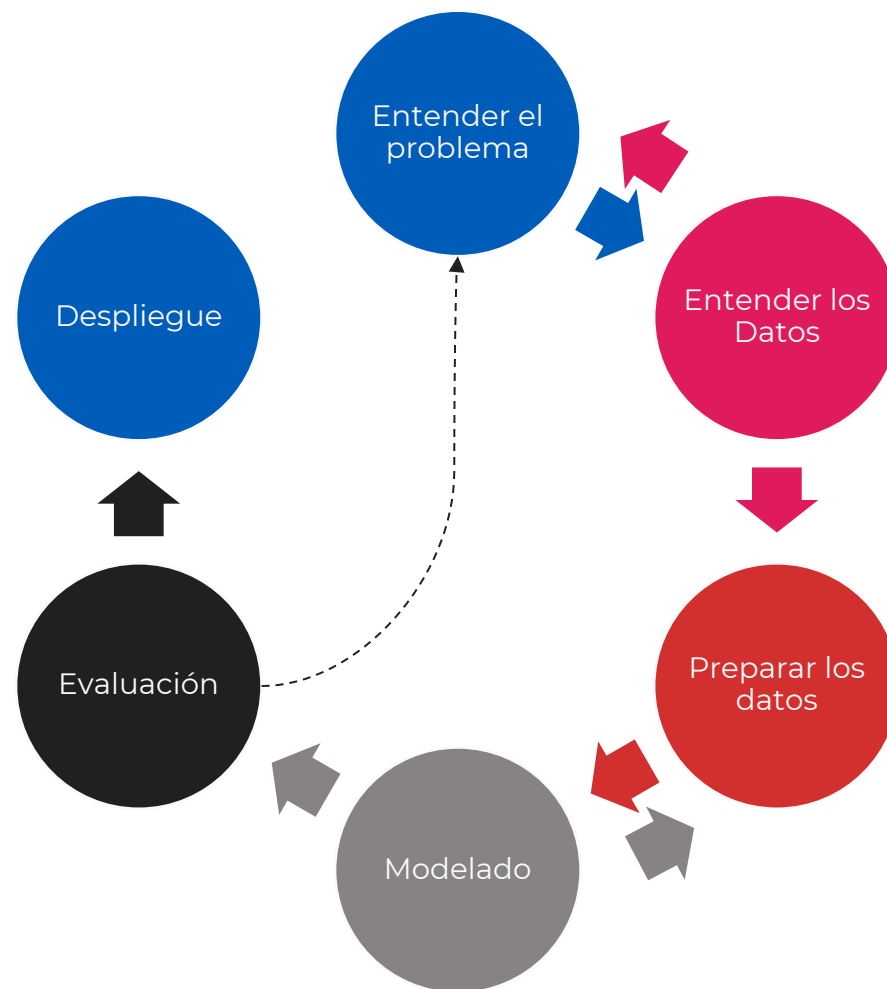


El valor de los datos en la política pública



CRISP - DM

Cross-industry Standard Process for Data Mining



Datos estructurados y limpios

"Tidy datasets are all alike, but every messy dataset is messy in its own way [...]".

-Hadley Wickham

The diagram illustrates a tidy dataset structure. It features a table with 5 rows and 4 columns. The columns are labeled 'Name', 'Score', 'Attempts', and 'Qualify'. The rows are indexed 0 to 4. The data is as follows:

	Name	Score	Attempts	Qualify
0	Anastasia	12.5	1	yes
1	Dima	9.0	3	no
2	Katherine	16.5	2	yes
3	James	NaN	3	no
4	Emily	9.0	2	no

Labels and annotations: 'Columns' points to the header row. 'Rows' points to the row indices. 'Data' points to the individual cells. Specific cells (Dima, 16.5, 3, and no) are highlighted with yellow boxes and connected by lines to the 'Data' label.

Datos limpios (*tidy data*):

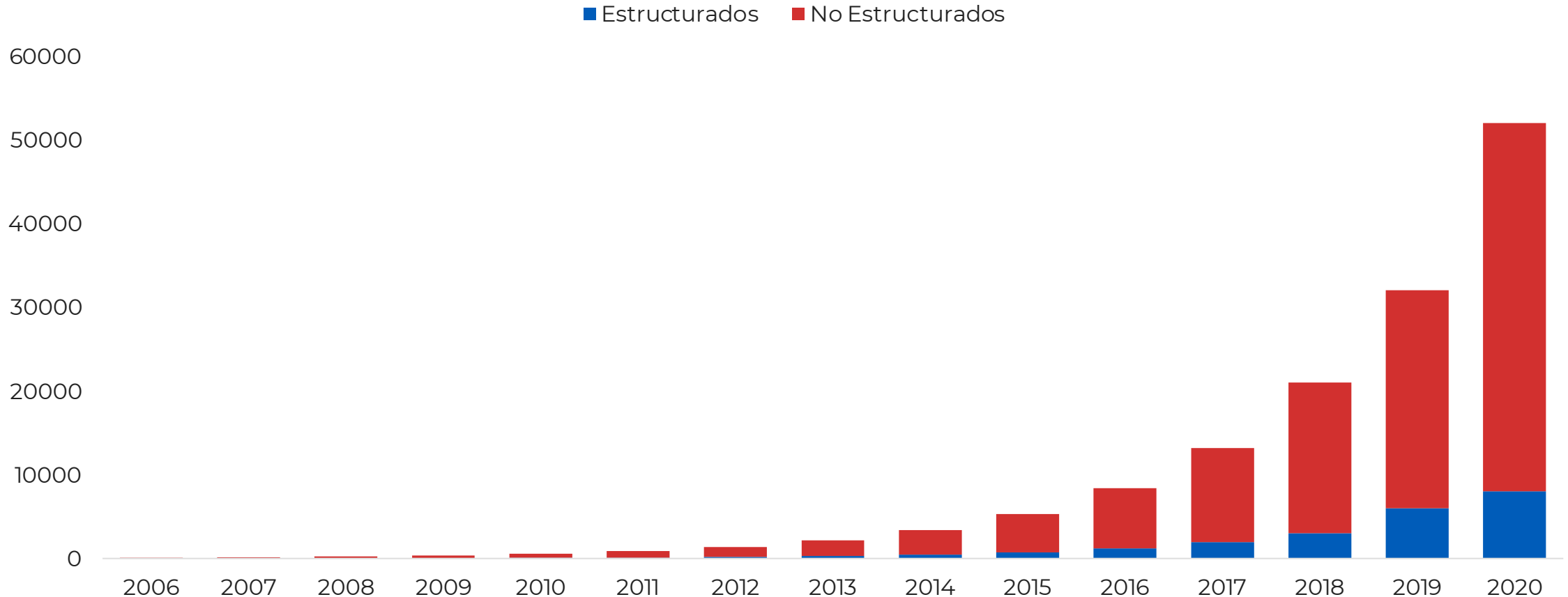
- Cada variable está representada por una columna.
- Cada observación está representada por una fila
- Cada celda es una medida única

Fuente: W3resource

Wickham, H. (2014). Tidy data. *Journal of statistical software*, 59(1), 1-23.

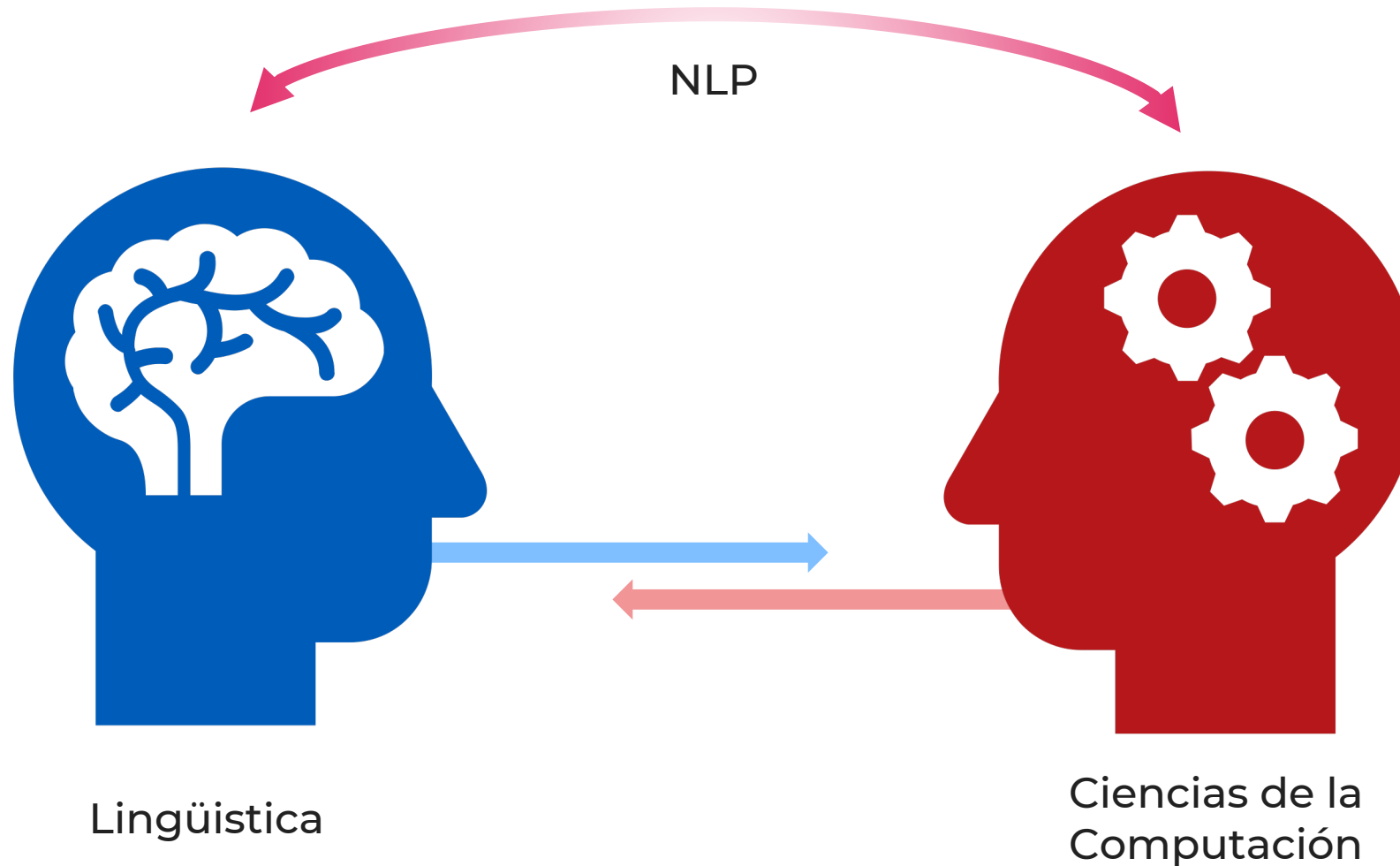
La Explosión Cámbrica de los Datos

Volumen de datos generado en Exabytes (Miles de millones de Gigabytes)



Azad et al (2020). The role of structured and unstructured data managing mechanisms in the Internet of things.

Procesamiento del Lenguaje Natural (NLP)



¿Por qué aprender sobre NLP?

Why natural language processing is AI's jewel in the crown

As we enter the intelligent era, smart devices and data volumes will skyrocket, as will the demand for human-computer interaction through speech and language - that's why it will be a US\$16 billion market by 2021.

By Linda Xu, Xue Hua



Fuente: Xue Hua (2020)



Procesamiento del Lenguaje Natural (NLP)

Generación del Lenguaje Natural (NLG)



Entendimiento del Lenguaje Natural (NLU)



¿Por qué enfocarnos en NLU?

 MILENIO

AMLO exhibe a presidente de CMDH por lanzar insultos durante protesta en embajada de Cuba

"Entonces uno de los participantes le dice a un funcionario de la embajada de Cuba, bueno no lo voy a repetir, el de Derechos Humanos, pero ...

5 days ago



 Infobae

AMLO criticó a defensor de DDHH por incidente en la Embajada de Cuba

AMLO criticó a defensor de DDHH por incidente en la Embajada de Cuba.

"Hay una confrontación, afortunadamente menor, pero si hay insultos, ...

5 days ago



 Proceso

AMLO exhibe racismo del defensor de DDHH René Bolio en protesta panista frente a embajada de Cuba

El presidente Andrés Manuel López Obrador utilizó el caso de la protesta que hizo ayer un grupo de panistas frente a la embajada de Cuba en ...

In-Depth · 5 days ago



¿Por qué enfocarnos en NLU?


 MILENIO

AMLO exhibe a presidente de CMDH por lanzar insultos durante protesta en embajada de Cuba

"Entonces uno de los participantes le dice a un funcionario de la embajada de Cuba, bueno no lo voy a repetir, el de Derechos Humanos, pero ...

5 days ago



 Infobae

AMLO criticó a defensor de DDHH por incidente en la Embajada de Cuba

AMLO criticó a defensor de DDHH por incidente en la Embajada de Cuba.

"Hay una confrontación, afortunadamente menor, pero si hay insultos, ...

5 days ago



 Proceso

AMLO exhibe racismo del defensor de DDHH René Bolio en protesta panista frente a embajada de Cuba

El presidente Andrés Manuel López Obrador utilizó el caso de la protesta que hizo ayer un grupo de panistas frente a la embajada de Cuba en ...

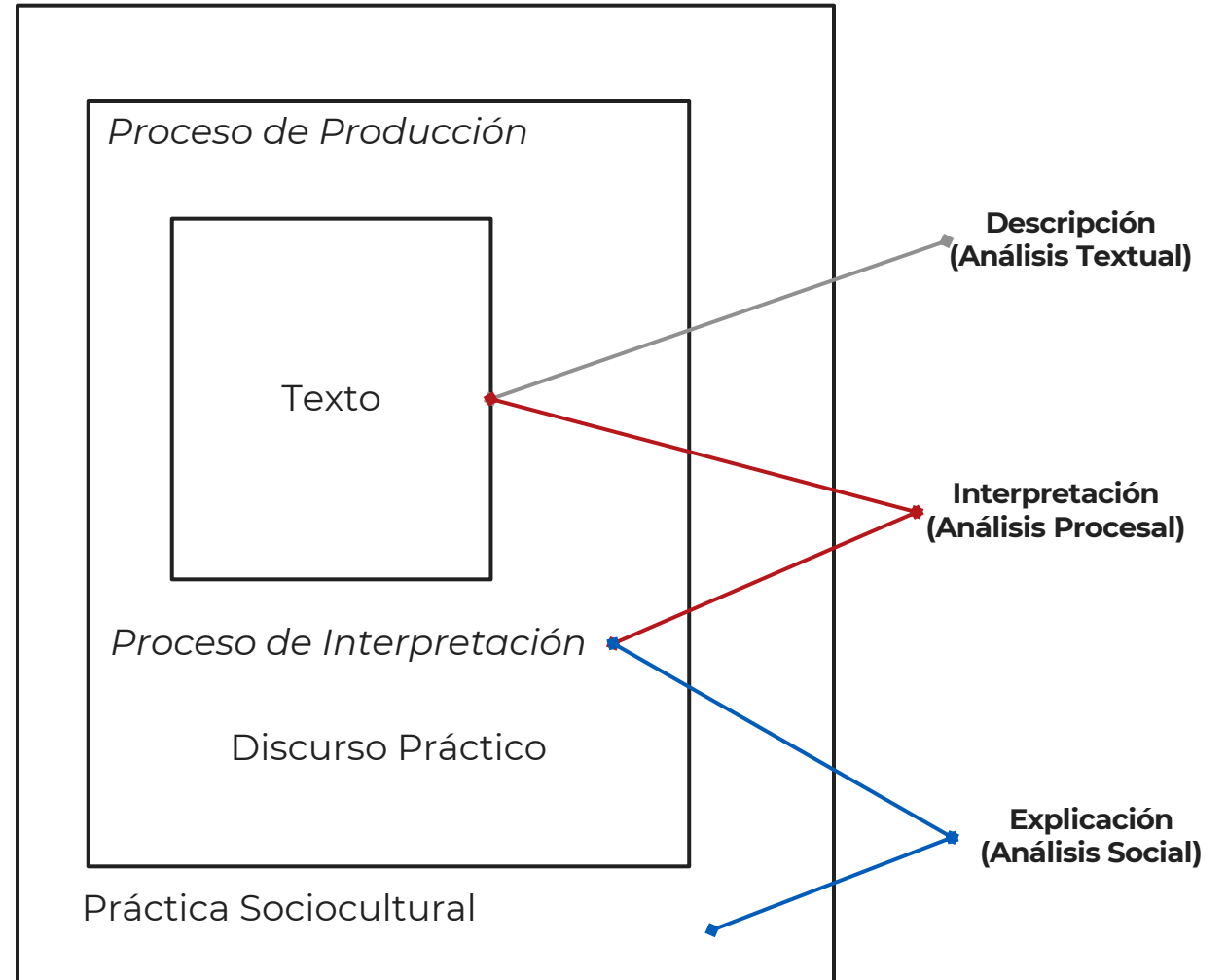
In-Depth · 5 days ago



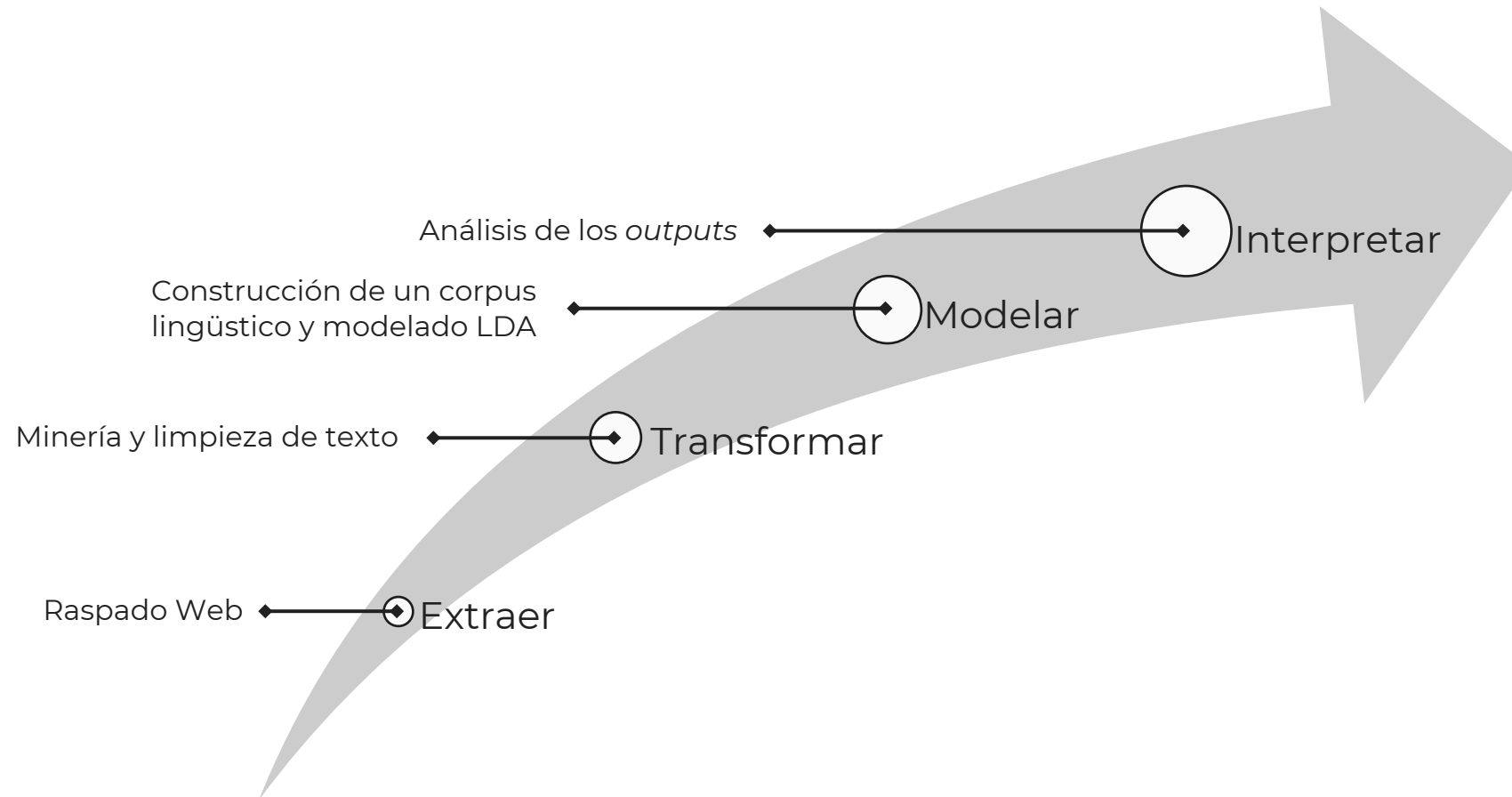
Análisis Crítico del Discurso

CDA sees discourse – language use in speech and writing – as a form of ‘social practice’. Describing discourse as social practice implies a dialectical relationship between a particular discursive event and the situation(s), institution(s) and social structure(s) which frame it: the discursive event is shaped by them, but it also shapes them.

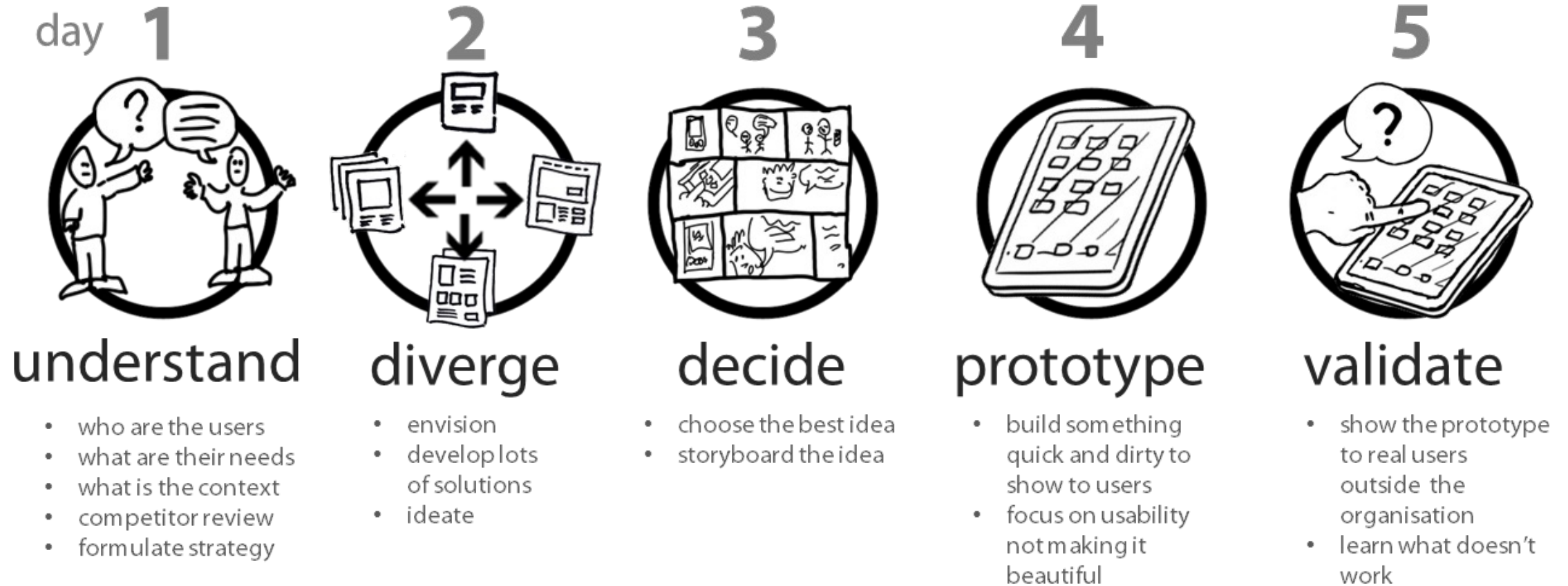
Fairclough & Wodak, 1997



Nuestro Flujo de Trabajo



Design Sprint



¡Gracias por tu atención!

Escríbeme: r.rosado@tec.mx