



BIG DATA

MARÍA PARRA AMAT
Almería, 5 junio 2015



BIG DATA

- "Petróleo del XXI"
- Nuevo enfoque en el entendimiento y la toma de decisiones
- Conjunto de técnicas y herramientas

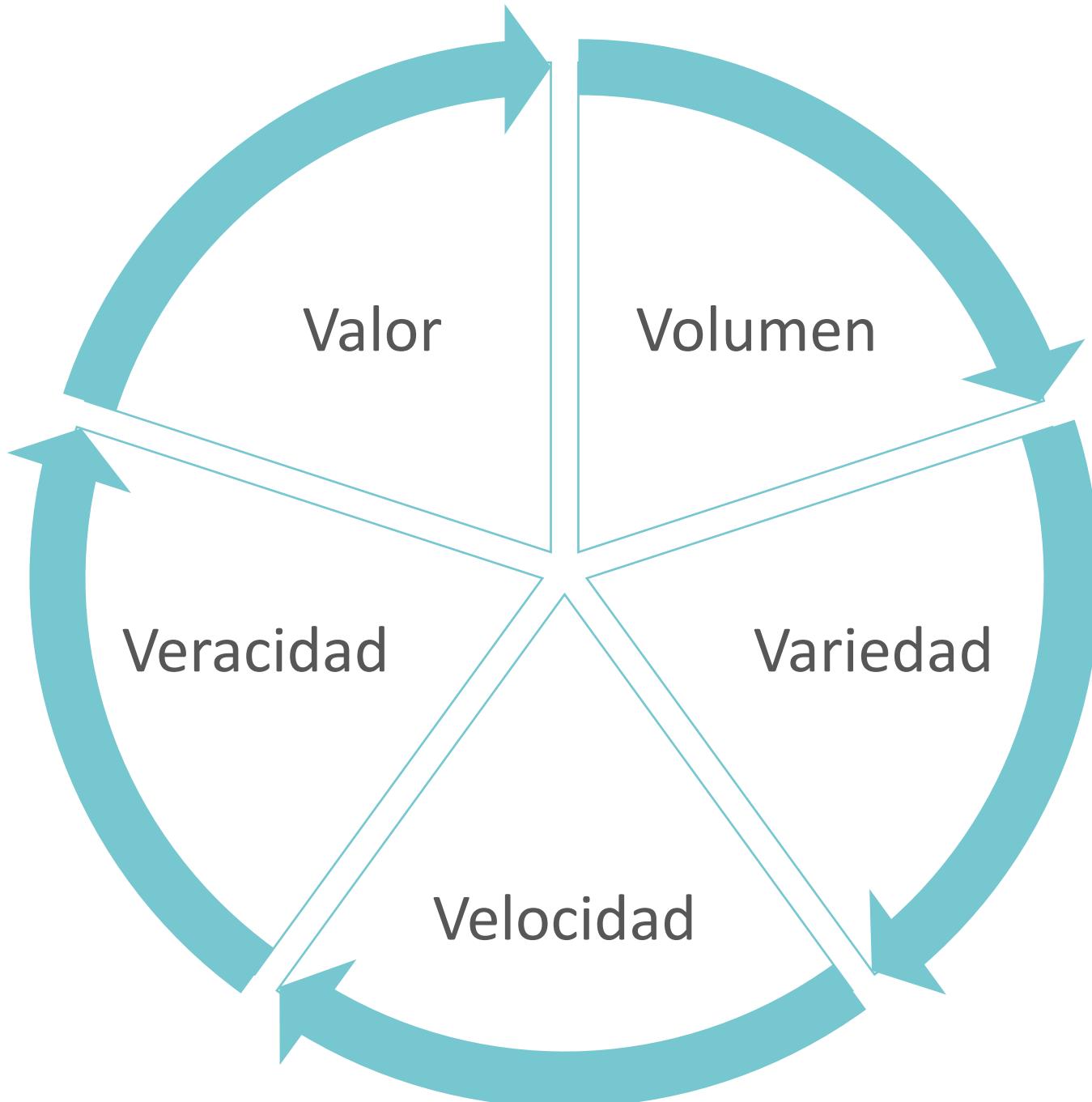
ORIGEN

- Estadística
- Metodologías existentes aplicadas a nuevos problemas



POPULARIZACIÓN

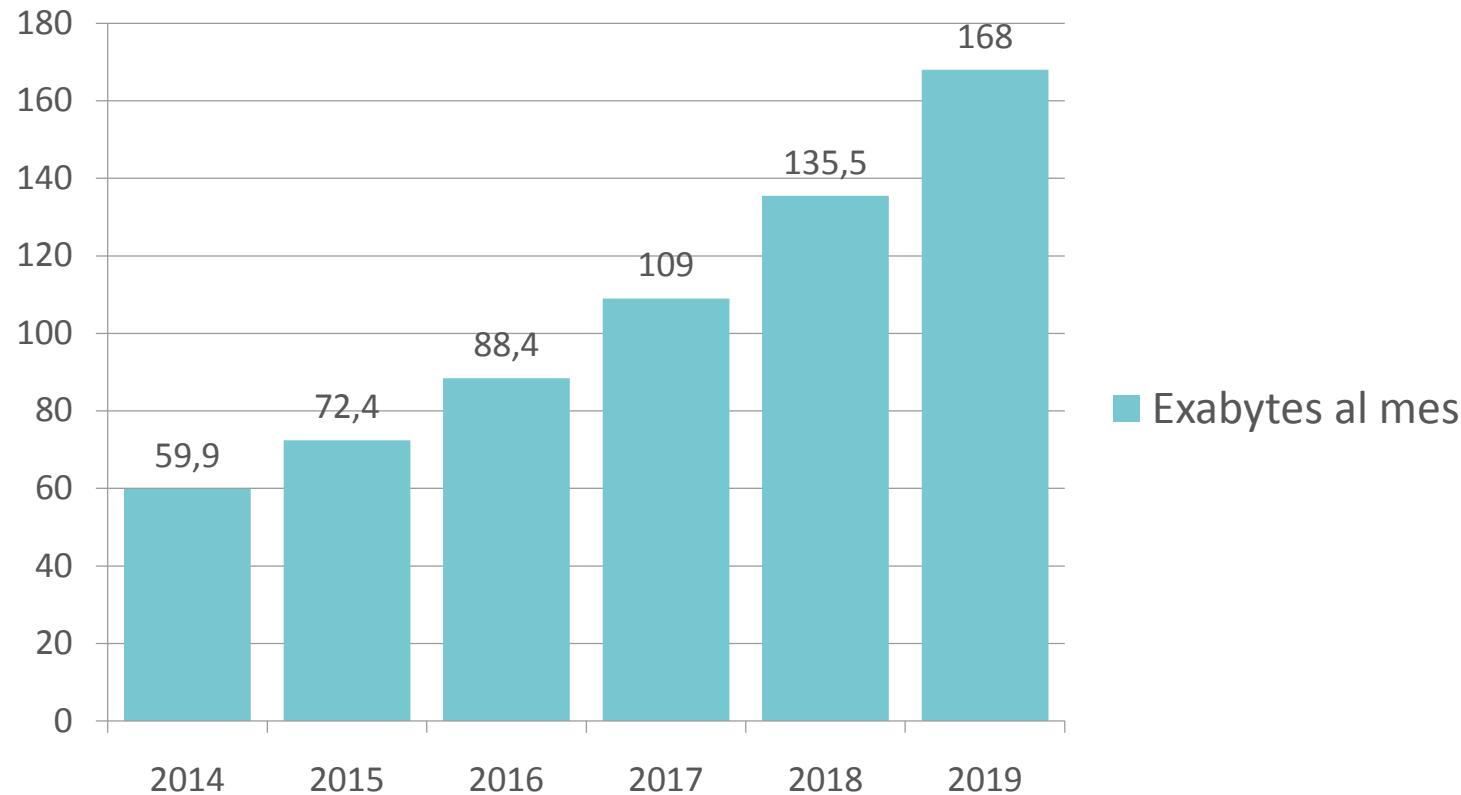
- Ultraconectividad
- Más sensores y fuentes de datos
- Desarrollo tecnológico
 - Almacenamiento: capacidad y coste
 - Procesamiento: HDFS y Mapreduce



CARACTERÍSTICAS

- 5 V's

Estimación tráfico de datos



VOLUMEN

- La cantidad de datos.
- Desconocimiento del volumen real.
- Incremento del tráfico en internet será exponencial - Era del Zettabyte.



VARIEDAD

- Procedencia diversa
- Multitud de formatos



VELOCIDAD

- Los datos en movimiento
- Obtener la información correcta en el momento preciso



VERACIDAD

- Confianza o fiabilidad de los datos
- Evitar sesgos o ruidos
- Limpieza de datos apropiados



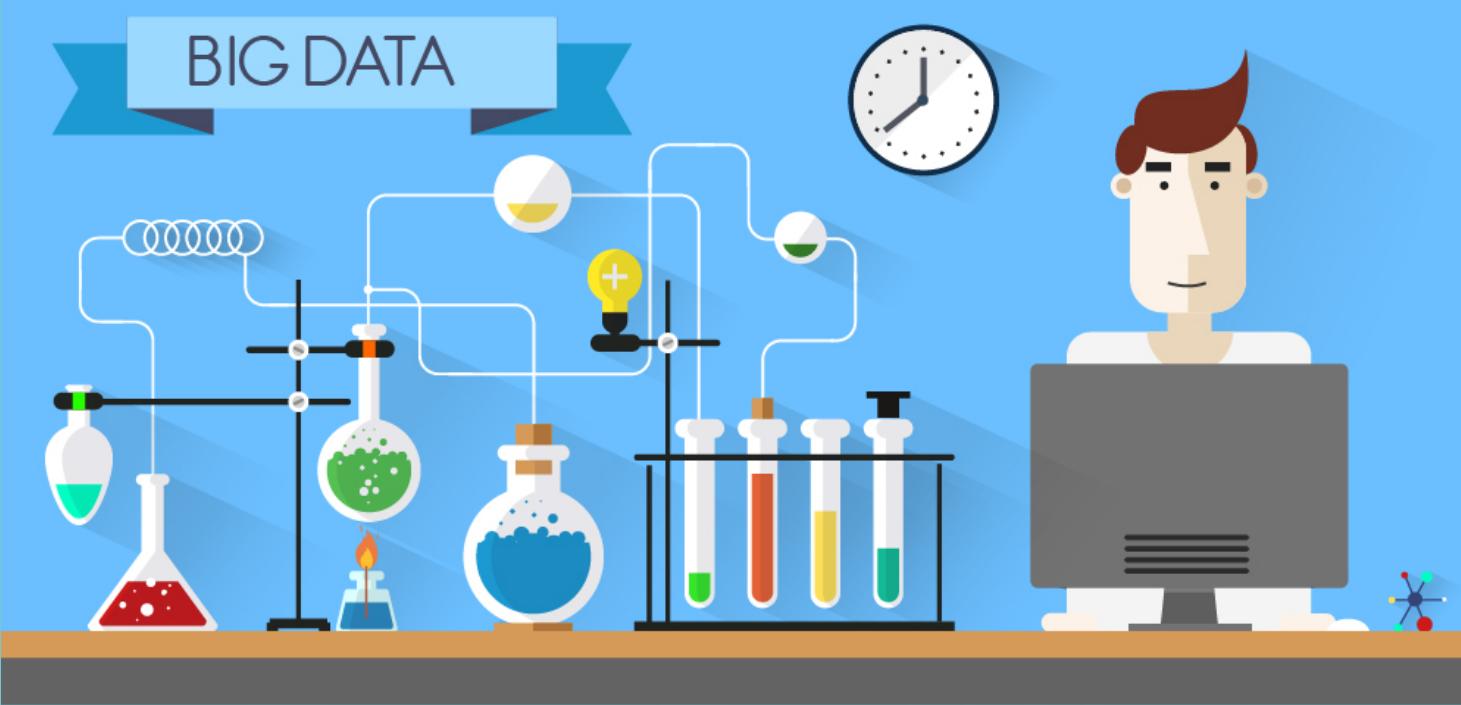
VALOR DEL DATO

Es la importancia que puede tener el dato para la toma de decisiones en las empresas y los gobiernos.



DISCIPLINAS

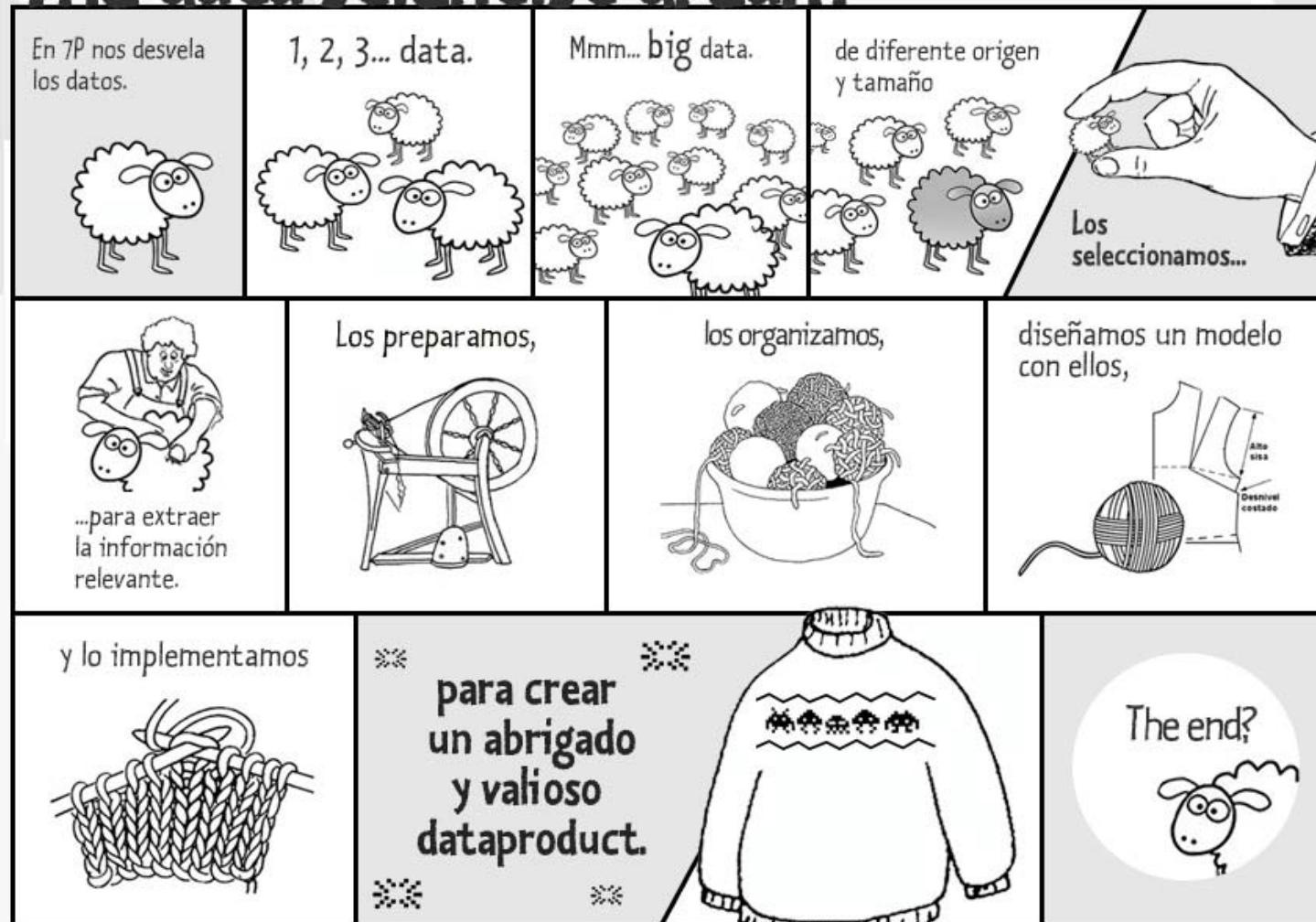
- Estadística
- Informática
- Campo de aplicación



LAS CATEGORÍAS DE “CIENTÍFICO DE DATOS”

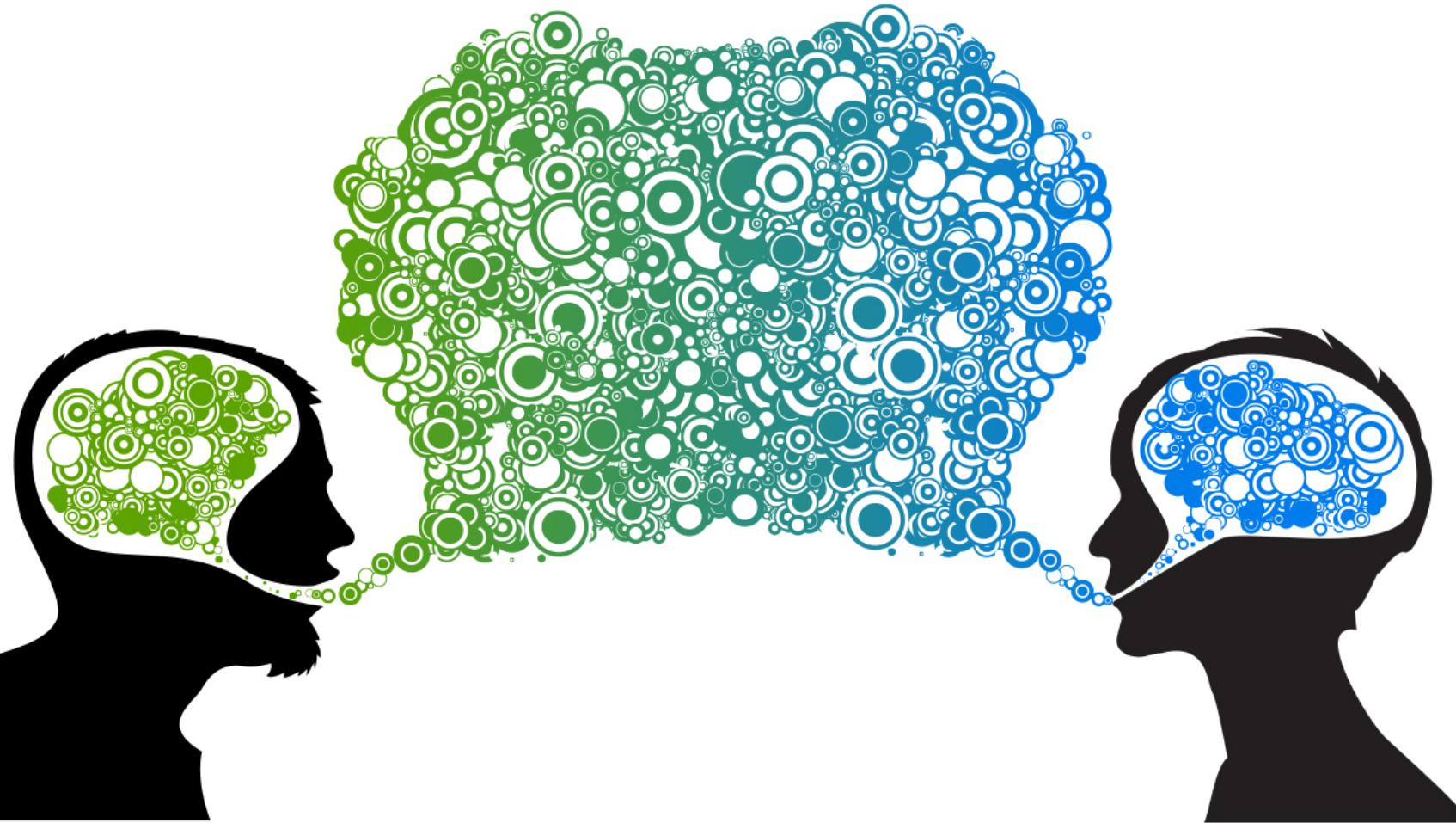
- Expertos en:
 - Estadística y matemáticas
 - Ingeniería de datos
 - Aprendizaje científico
 - Negocio
 - Desarrollo de producción de código
 - Visualización
 - GIS

The data scientist dream



TAREAS PRINCIPALES EN UN PROYECTO DE BIG DATA

- Diálogo con expertos en la materia
- Captura y almacenamiento de la información
- Procesado masivo de datos
- Análisis con técnicas de inteligencia artificial - Minería de datos
- Visualización



DIÁLOGO CON EXPERTOS EN LA MATERIA

- Fundamental para tener éxito
- Conocer inquietudes del experto
- Informar de las soluciones reales que aporta Big Data



{ JSON }



CAPTURA Y ALMACENAMIENTO DE LA INFORMACIÓN

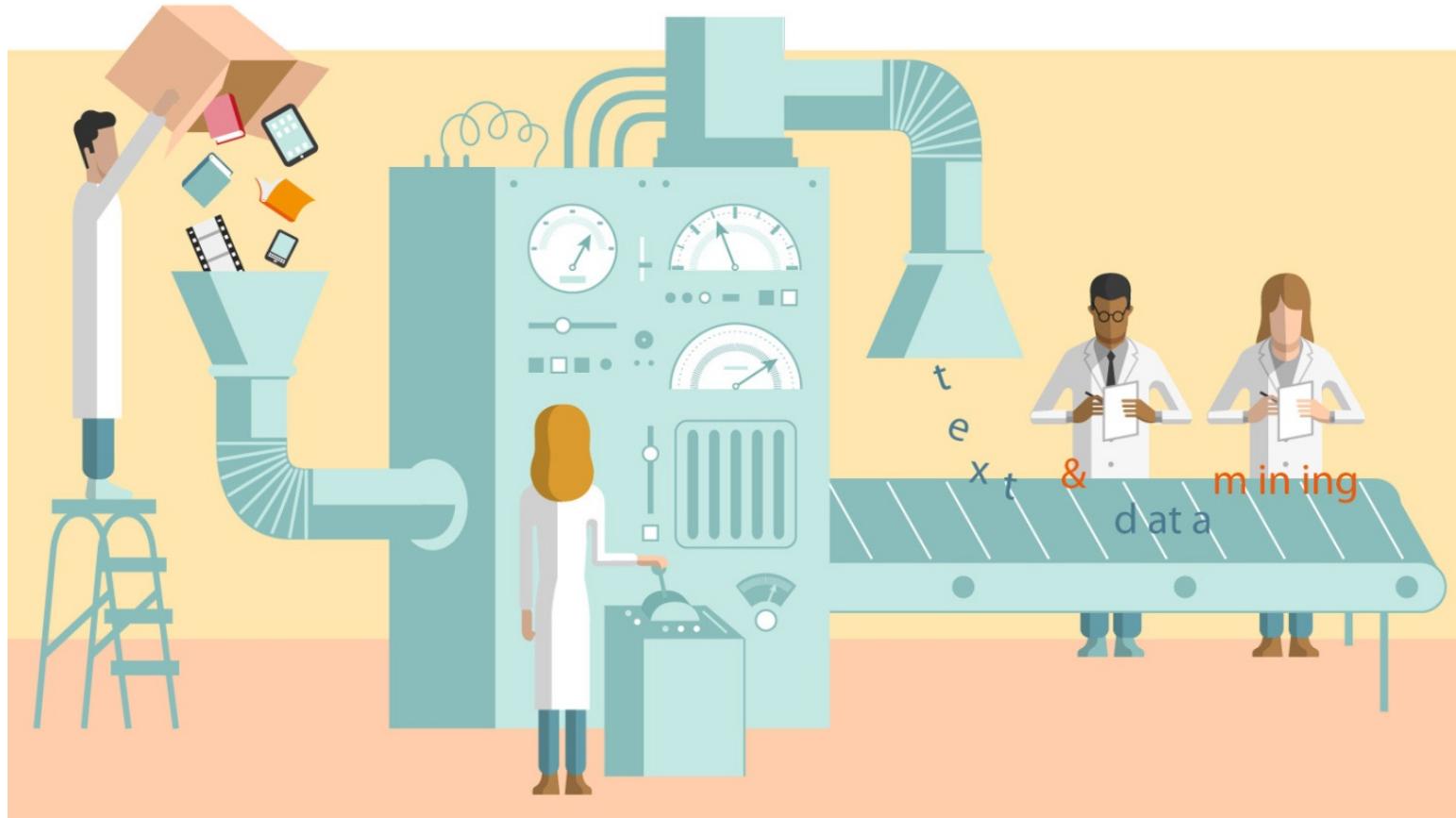
- Datos existentes:
 - Documentos
 - Web
 - Bases de datos Sql
 - noSql
- Crear nuevas infraestructuras



PROCESADO MASIVO DE DATOS

- Framework Hadoop
 - HDFS
 - MapReduce
- Otras utilidades
 - Hive, Pig, integración con R

@mariaparraamat #jpd2015



TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS

- Minería de datos
 - Utiliza técnicas de aprendizaje automático
- Minería de texto



TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS

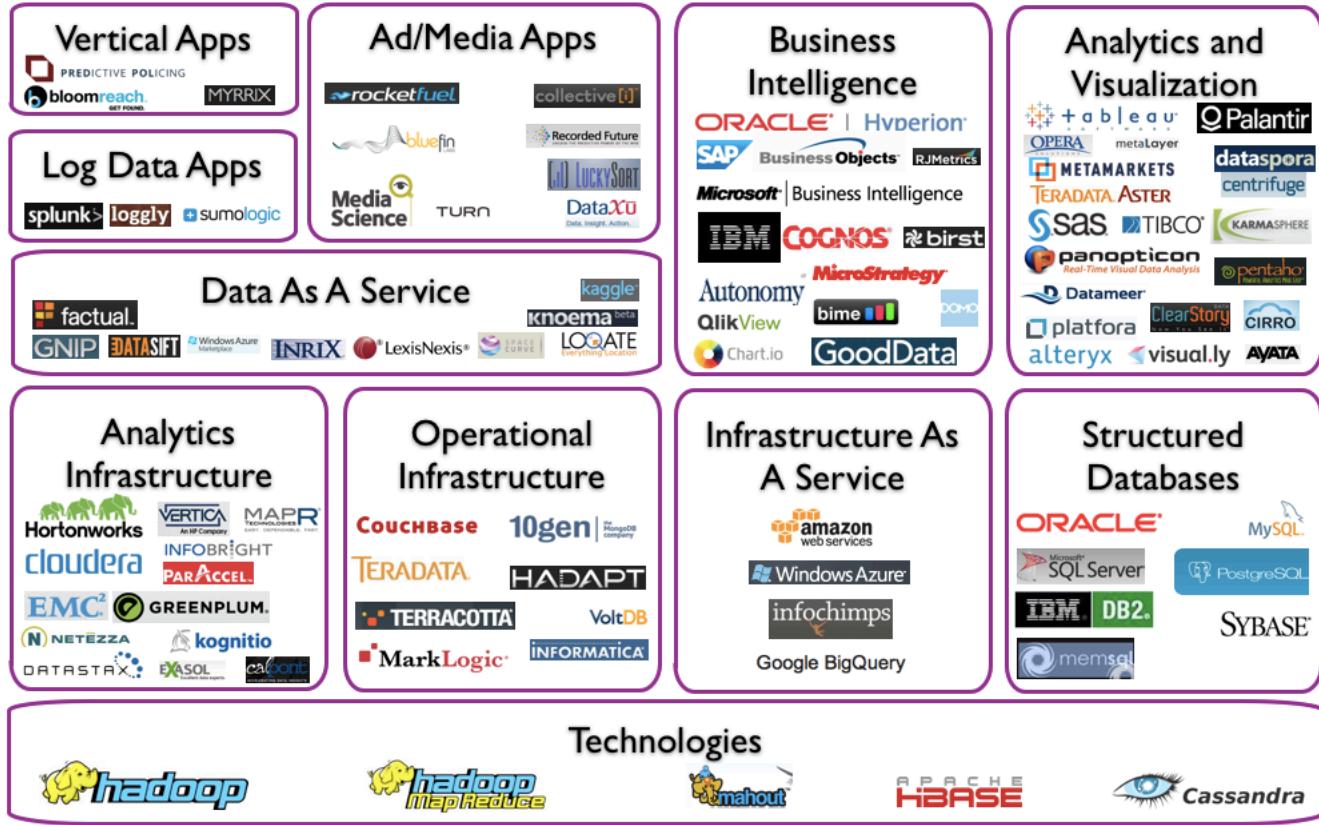
- Minería de datos
- **Minería de texto**
 - Encontrar términos frecuentes
 - Buscar términos relacionados con otros
 - Agrupamiento
 - Clasificación
- **MOA, Weka, R**

VISUALIZACIÓN



- Representación gráfica mediante:
 - Estadísticas
 - Mapas
 - Diagramas
 - Esquemas
- Dos finalidades
 - Presentar datos para una audiencia determinada
 - Cómo herramienta de análisis
- Herramientas: Tableau, infogram, cartoDB.
- Librerías: Google Charts, JQuery plug-ins, D3.js

Big Data Landscape



Copyright © 2012 Dave Feinleib

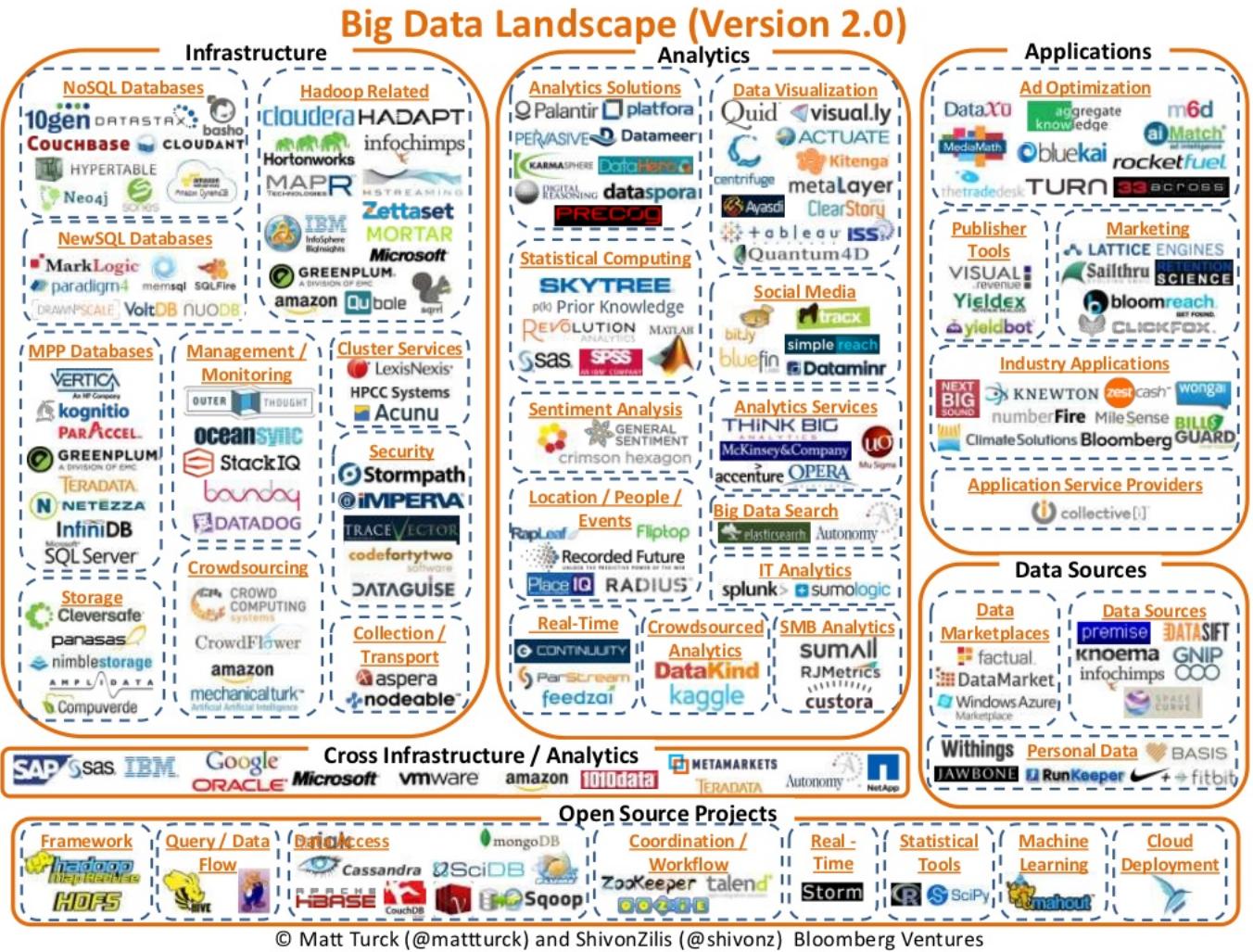
dave@vcDave.com

blogs.forbes.com/davefeinleib

PAISAJE BIG DATA

- **Big Data Landscape 1.0**
- **Big Data Landscape 2.0**
- **Big Data Landscape 3.0**

@mariaparraamat #jpd2015

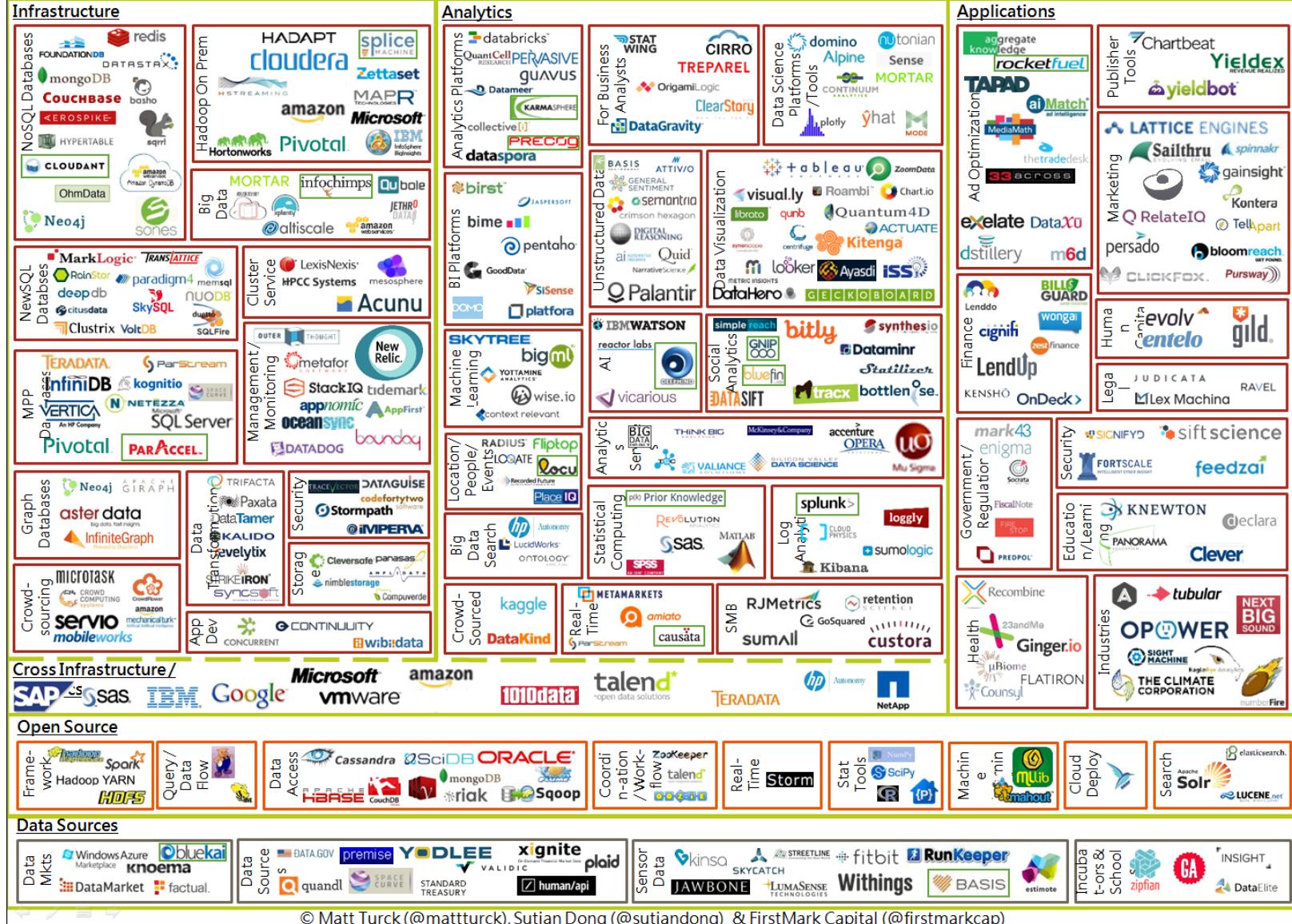


PAISAJE BIG DATA

- Big Data Landscape 1.0
- **Big Data Landscape 2.0**
- Big Data Landscape 3.0

BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or
IPO



PAISAJE BIG DATA

- Big Data Landscape 1.0
 - Big Data Landscape 2.0
 - **Big Data Landscape 3.0**

@mariaparraamat #jpd2015



LIMITACIONES

- No permite detectar comportamientos individuales.
- Puede detectar correlación, pero no causalidad.
- Gran cantidad de ruido.
- Big Data puede sugerir hipótesis.
- Período de validez determinado.



APLICACIONES

- **Marketing**
- Análisis de sentimientos
- En la industria del entretenimiento
- Seguimiento de epidemias
- Educación
- Crimen organizado
- Salud
- Periodismo de datos



APLICACIONES

- Marketing
- **Análisis de sentimientos**
- En la industria del entretenimiento
- Seguimiento de epidemias
- Educación
- Crimen organizado
- Salud
- Periodismo de datos



APLICACIONES

- Marketing
- Análisis de sentimientos
- **Industria del entretenimiento**
- Seguimiento de epidemias
- Educación
- Crimen organizado
- Salud
- Periodismo de datos

[Google.org home](#)

[Dengue Trends](#)

[Flu Trends](#)

[Home](#)

[Select country/region](#)

[How does this work?](#)

[FAQ](#)

Flu activity

Intense

High

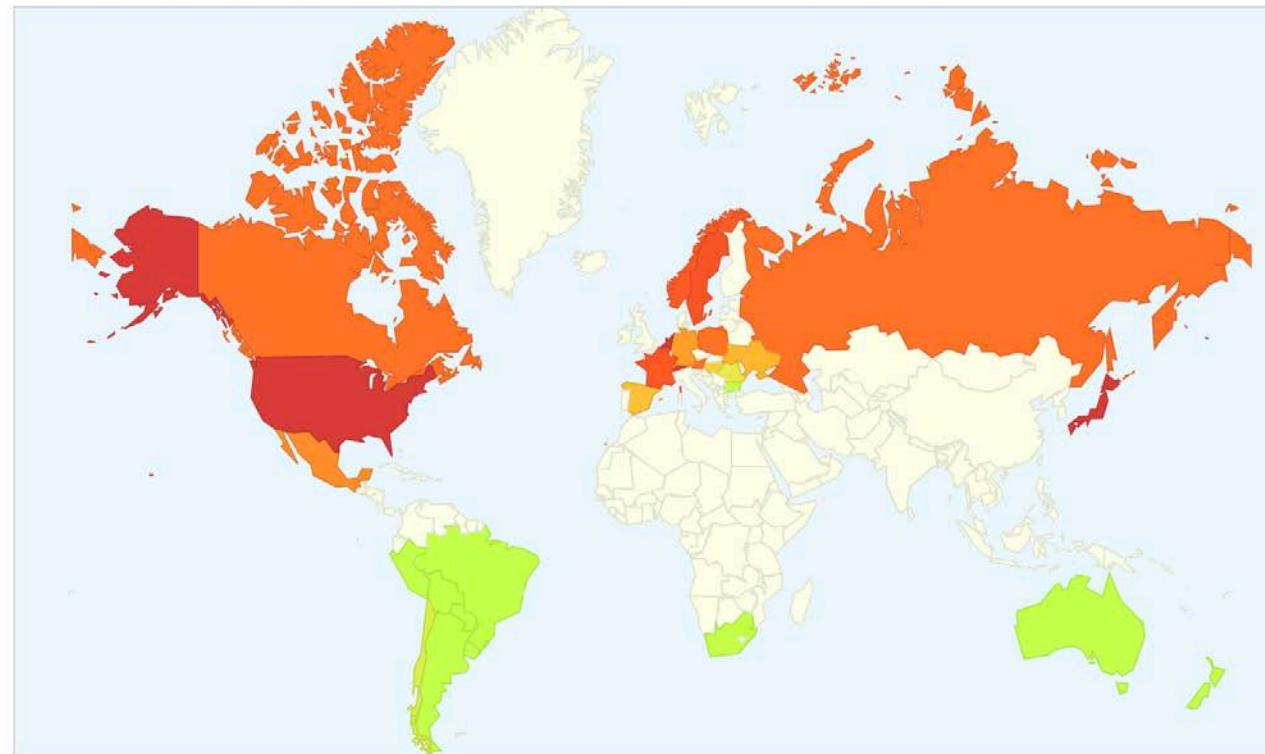
Moderate

Low

Minimal

Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)



APLICACIONES

- Marketing
- Análisis de sentimientos
- En la industria del entretenimiento
- **Seguimiento de epidemias**
- Educación
- Crimen organizado
- Salud
- Periodismo de datos



APLICACIONES

- Marketing
 - Análisis de sentimientos
 - En la industria del entretenimiento
 - Seguimiento de epidemias
 - **Educación**
 - Crimen organizado
 - Salud
 - Periodismo de datos



APLICACIONES

- Marketing
- Análisis de sentimientos
- En la industria del entretenimiento
- Seguimiento de epidemias
- Educación
- **Crimen organizado**
- Salud
- Periodismo de datos



APLICACIONES

- Marketing
- Análisis de sentimientos
- En la industria del entretenimiento
- Seguimiento de epidemias
- Educación
- Crimen organizado
- **Salud**
- Periodismo de datos

Los derechos de las personas homosexuales en el mundo

- Legislación en diferentes países: del matrimonio a la pena de muerte
- [Irlanda vota sobre el matrimonio gay pendiente del voto oculto](#)

EL PAÍS | Madrid | 22 MAY 2015 - 19:17 CEST

Archivado en: Activismo lgtb Matrimonio Homosexualidad LGBT Derechos civiles Activismo Sexualidad Derechos humanos Familia Grupos sociales
Sociedad

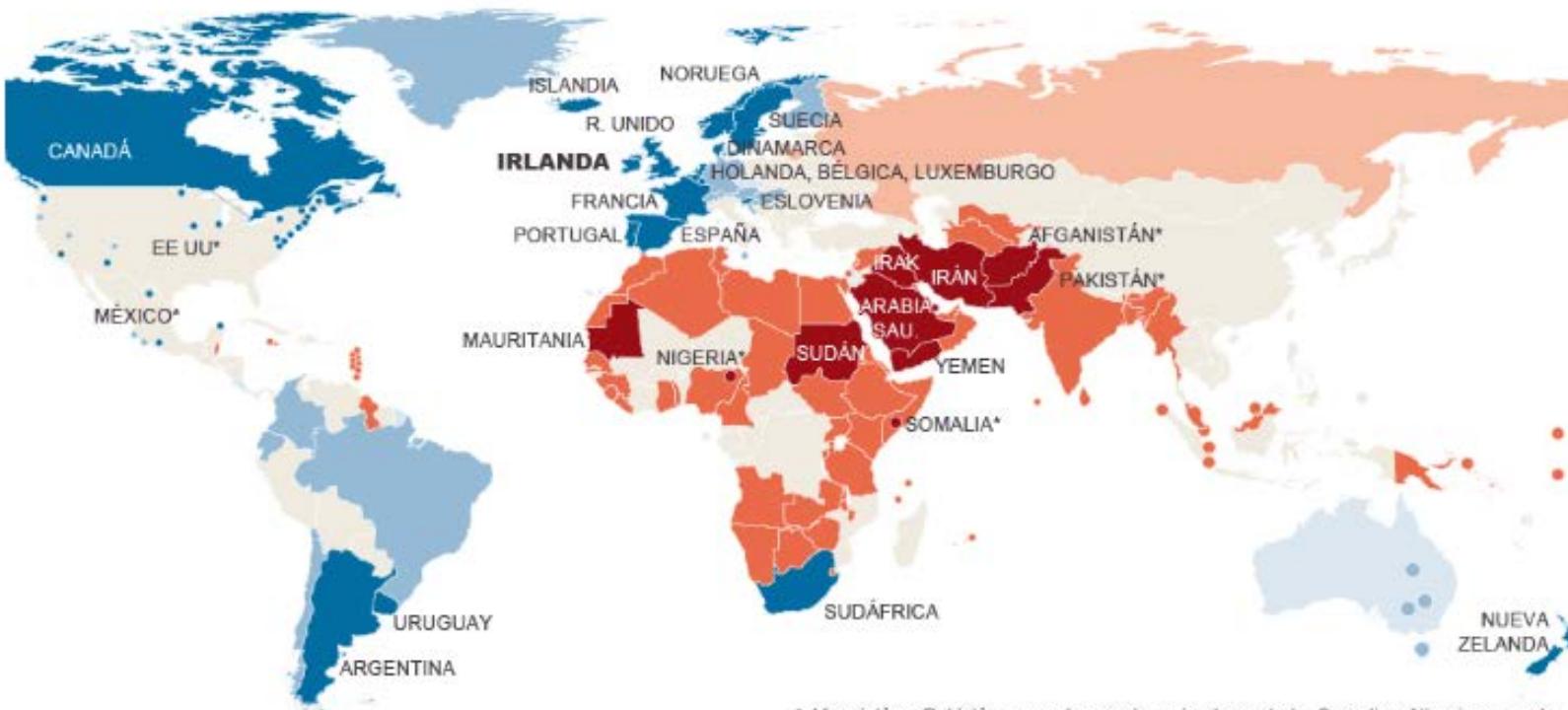


Enviar Guardar

PERSECUCIÓN ■ Pena de muerte ■ Prisión ■ Ley anti propaganda homosexual

RECONOCIMIENTO ■ Matrimonio ■ Unión con derechos similares ■ Inferior al matrimonio

● ● ● Leyes válidas para ciertas partes del país
● ● ● Inferior al matrimonio



Fuente: The International Lesbian, Gay, Bisexual, Trans and Intersex Association (ILGA).

EL PAÍS

APLICACIONES

- Marketing
- Análisis de sentimientos
- En la industria del entretenimiento
- Seguimiento de epidemias
- Educación
- Crimen organizado
- Salud
- Periodismo de datos**

@mariaparraamat #jpd2015

MUCHAS GRACIAS

ALMERÍA, 5 DE JUNIO DE 2015

MARÍA PARRA AMAT

Email: mariaparraamat@gmail.com

Twitter: @mariaparraamat

Facebook:

www.facebook.com/mariaparraamat

@mariaparraamat #jpd2015