

BIG DATA: Aplicaciones a Inteligencia Artificial, Ciberseguridad y Knowledge discovery

CAMPUS PRAKTIKUM

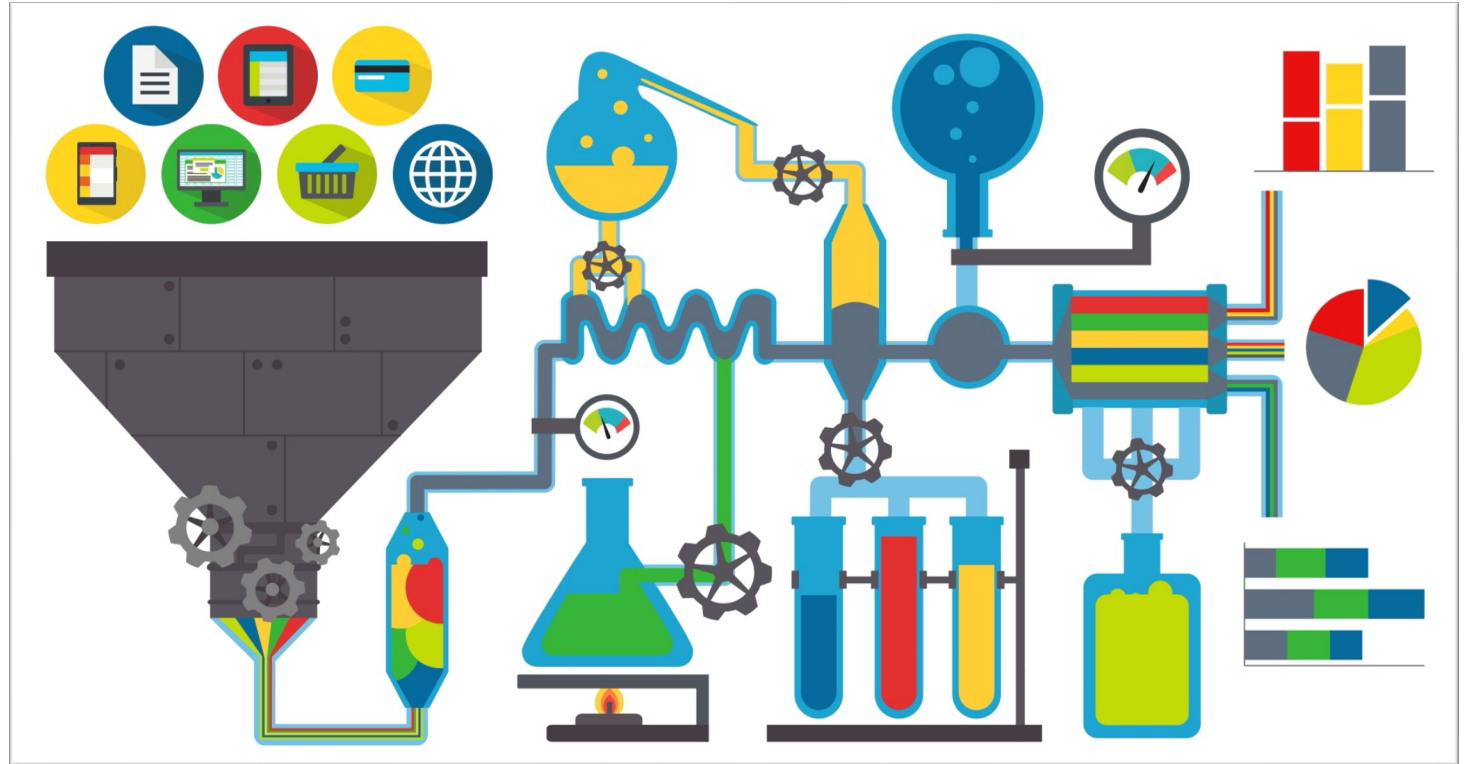
Datos, Información y Herramientas

R. Alcover | V. Chirivella | E. Vázquez
Dto. de Estadística e Investigación Operativa Aplicadas y Calidad

UPV 2024

¿Y si lo analizamos?

Datos,
Información y
Herramientas



R. Alcover | V. Chirivella | E. Vázquez

Dto. de Estadística e Investigación Operativa Aplicadas y Calidad

Convertir datos en información: *extracción de conocimiento*



- **Economía y sector financiero**
 - Datos de clientes de un **banco** → ¿Se debe conceder un crédito?
 - Datos de clientes de un **supermercado** → ¿qué previsión de stock se necesita? ¿Qué productos de venden?
 - ...
- **Sociedad y salud**
 - Datos de temperatura, presión atmosférica, etc. de **AEMET** → Predicciones meteorológicas y prevención de catástrofes (riadas, sequía, etc)
 - Datos de pacientes y usuarios de SPS → organización de recursos sanitarios (materiales y humanos), campañas de prevención de enfermedades
 - ...
- **Smart Cities y energía**
 - Datos de EMT → ¿Qué capacidad deben tener los autobuses? ¿Qué frecuencia de autobuses y a qué hora? ¿Cuáles son las rutas óptimas? ¿Qué modelos adquirir?
 - Datos socio – económicos de empadronados en VLC → ¿Cuántos colegios se necesitan?, ¿dónde y con qué capacidad?
 - ...

Un ejemplo: sociedades del futuro

“Una Ciudad **Inteligente y Sostenible** es una ciudad **innovadora** que **aprovecha las Tecnologías de la Información y la Comunicación (TIC)** y otros medios para **mejorar la calidad de vida, la competitividad, la eficiencia del funcionamiento y los servicios urbanos**, al tiempo que se asegura de que responde a las **necesidades de las generaciones presente y futuras** en lo que respecta a los **aspectos económicos, sociales, medioambientales y culturales.**”

UNE 178201:2016

Ciudades inteligentes. Definición, atributos y requisitos.



Smart Cities

Aplicaciones del

BIG DATA

¿QUÉ ES BIG DATA?

Big Data son grandes volúmenes de datos que surgen a enorme velocidad desde múltiples orígenes y con distintos formatos.

¿PARA QUÉ SIRVE EL ANÁLISIS DEL BIG DATA?

RECABAR TODA LA INFORMACIÓN DISPONIBLE Y CONVERTIRLA EN CONOCIMIENTO

PROFUNDIZAR EN LAS PREFERENCIAS DEL CONSUMIDOR

PERSONALIZAR LAS ACCIONES DE MARKETING



1 ECONOMÍA Y SECTOR FINANCIERO

■ **CAPTAR TENDENCIAS ECONÓMICAS:**
el sector bancario puede inferir a través del análisis de los movimientos de las tarjetas de crédito cómo va a evolucionar la economía

■ **PREVENCIÓN Y DETECCIÓN DEL FRAUDE:**
plataformas de pago online identifican patrones de actividad fraudulenta en las operaciones que registran para crear alertas y permitir una detección precoz



2 SOCIEDAD Y SALUD



■ **PREVENCIÓN DE CATÁSTROFES NATURALES:**
los científicos pueden analizar datos históricos de variables como la temperatura, la presión atmosférica o la nubosidad para predecir con mayor rapidez y exactitud la magnitud de un huracán y el rumbo que va a tomar.

■ **PREVENCIÓN SANITARIA:**
la secuenciación del genoma humano produce gran cantidad de datos que los científicos pueden analizar para hacer más fácil, barata y rápida la investigación de enfermedades

3 SMART CITIES Y ENERGÍA

■ **EFICIENCIA Y AHORRO ENERGÉTICO:**
con datos procedentes de contadores inteligentes de los usuarios y con datos sobre variables meteorológicas se puede optimizar la producción de energía y maximizar el uso de energías renovables.

■ **MEDICIÓN DEL TRÁFICO:**
proporcionar a los usuarios información sobre dónde tienen más plazas libres para aparcar analizando los datos que emiten los sensores de aparcamiento instalados en una ciudad para reducir tiempos y contaminación.



Big Data

LAS 4 V'S DEL BIG DATA



VOLUMEN

Gran cantidad de datos

VELOCIDAD

Alta velocidad de flujo de datos.
Procesamiento en tiempo real

VARIEDAD

Recolectados de diferentes fuentes
(redes, mobile, estructurado, no
estructurado, etc)

VERACIDAD

Calidad y confiabilidad de
los datos



Datos, Información y Herramientas



1 Datos

Recopilación

Generación

Reutilización



2 Herramientas informáticas

Excel

Statgraphics

Lenguaje R

Python

...



3 Modelos matemáticos

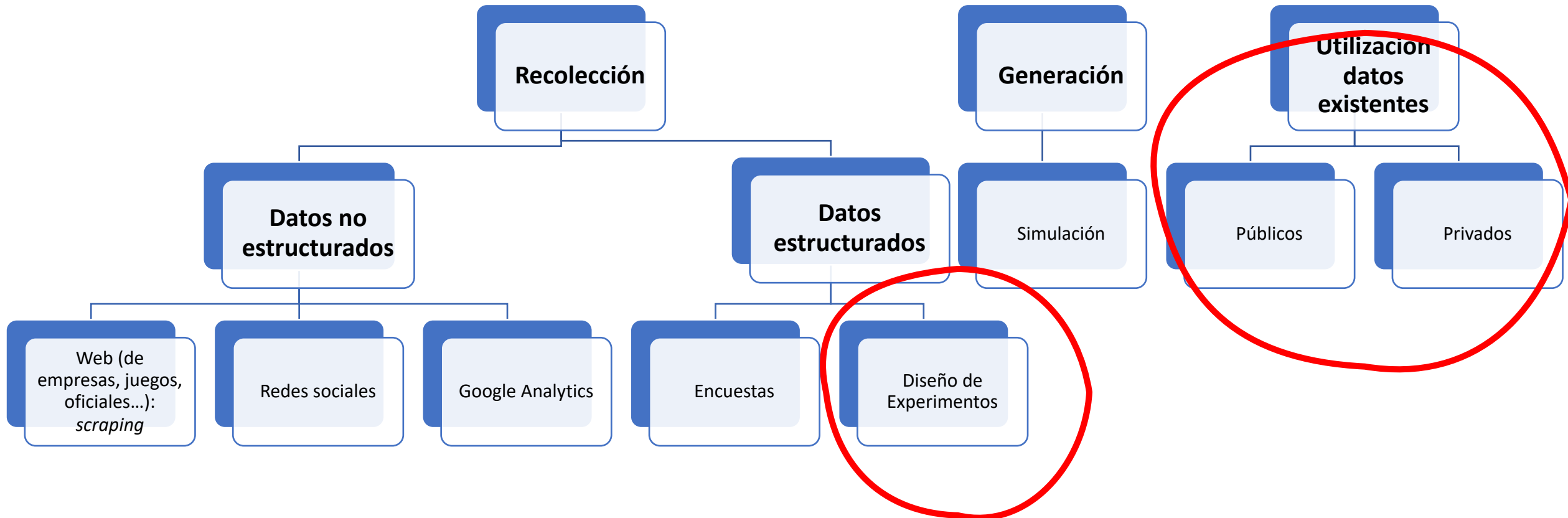
Técnicas estadísticas

Machine Learning

etc

...

Datos e Información



Bases de Datos Públicas España



Instituto nacional de Estadística - INE



Ministerio de Asuntos Económicos y Transformación Digital

Realiza operaciones estadísticas de gran envergadura (censos demográficos y económicos, cuentas nacionales, estadísticas demográficas y sociales, indicadores económicos y sociales, coordinación y mantenimiento de los directorios de empresas, formación del Censo Electoral...)

En coordinación con **EUROSTAT**



<https://ine.es>

Aplicaciones



Actualizar una
renta con el IPC



Apellidos y
nombres



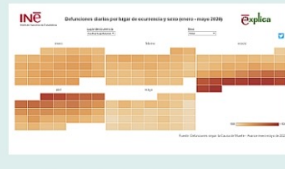
Estadísticas
territoriales



ODS.
Indicadores
Agenda 2030



Cuenta de flujos
de materiales



Defunciones
diarias 2020



Panel de
indicadores de
coyuntura



Gente en
movimiento. UE



TIC en los
hogares



¿Cuántos
cumplen años?



Indicadores de
turismo



Demografía de
Europa

⊕ Ver más aplicaciones

Pincha en
alguna



INE



Divulgación



Revista de estadística y sociedad

https://www.ine.es/explica/explica_quees.htm



Inicio / **Primeros pasos**

¿Qué es **Explica**?

Primeros pasos

Estadísticas oficiales

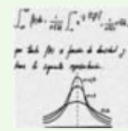
Estadística y mucho más

Juega con nosotros

Un poco de historia

Competición Estadística

Introducción a la Estadística



Una guía básica para acercarte a la Estadística:

Primera parte
Segunda parte

Píldoras estadísticas



Colección de vídeos breves con explicaciones de conceptos estadísticos

Información más atractiva



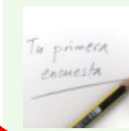
Te proponemos técnicas para captar la atención del lector

Gráficos de ayer y hoy



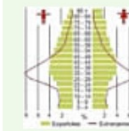
Representar datos mediante mapas y gráficos tiene su historia ...

Tu primera encuesta



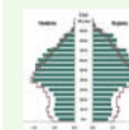
Cómo recoger los datos y presentar los resultados de una encuesta

Tipos de gráficos ¿cuál uso?



Una imagen bien utilizada vale más que mil palabras

Pirámides de población empadronada



Puedes ver las pirámides de ccaa, provincias y todos los municipios



¿Qué es **Explica**?

Primeros pasos

Estadísticas oficiales

Estadística y mucho más

Juega con nosotros

Un poco de historia

Competición Estadística

Infografías



Te mostramos otra forma más visual de presentar los datos estadísticos

Estadística y sellos



La estadística da la vuelta al mundo a través de los sellos

¿Sabías que ...?



Hay muchos datos interesantes que a lo mejor no conoces

Si España fuese un pueblo



de 100 habitantes habría 49 hombres y 51 mujeres, ...

Un día en cifras



Todos iguales todos distintos: cada país se identifica con sus datos

El 1 es rojo, el 8 es azul, ...



¿De qué color son los números? ¿Cómo suena un color?

Aplicaciones prácticas



Cuántos somos en mi pueblo ...
Apellidos y nombres ...
Actualizar renta ...



mié. 22 ●
jue. 23 ●
vie. 24 ●

Mi web Atención al ciudadano Atención al ciudadano sociales Meteoalerta: Avisos Efectos altas temperaturas Tiempo espacial APP Datos y estadísticas



<https://www.aemet.es/es/portada>

Meteoalerta

Península y Baleares Canarias

Nivel de aviso máximo:

mié. 22	●
jue. 23	●
vie. 24	●

Municipios es...

Buscar municipio...

Introduce el municipio o geolocalización de interés

Ver más municipios

El tiempo en tu web

Ciudades del Mundo



Acceso Web al Banco de Datos del CIS

El CIS tiene por finalidad el estudio científico de la sociedad española y la lleva a cabo mediante estudios, diseñados por su propio personal y en su mayoría encuestas, que tratan de captar las opiniones y actitudes de la población. Dentro de la estrategia general de transparencia y de servicio a la comunidad, académica, medios de comunicación y ciudadanos en general, y como continuación de la política de difusión emprendida años atrás, se abrió un nuevo canal de entrada a toda la colección de sus encuestas que ahora presentamos modernizado y ampliado.

El objetivo es posibilitar una vía de acceso a los datos de un modo sencillo y eficiente. Desde esta aplicación Web se puede acceder a la información completa de los estudios, de las preguntas que los componen y a las series temporales del CIS.

Algunas bases de datos públicas internacionales



- [Data.gov](#) : La página web Data.gov tiene 210.912 conjuntos de datos que están abiertos y disponibles gratuitamente para su descarga y uso. Muchos de los conjuntos de datos son visibles a través de mapas interactivos.
- [ARDA](#) – Association of Religion Data Archives: Este sitio web tiene también su propia sección de Mapas de GIS donde los usuarios pueden trazar conjuntos de datos religiosos sobre un barrio y / o mapas mundo.
- [Fuentes de datos públicos y abiertos en España y en el mundo.](#) : sitio web La Oficina del Censo de EE.UU. tiene la versión más reciente del censo de EE.UU., que se puede descargar gratuitamente. También tienen una galería de visualizaciones de datos donde se pueden encontrar infografías y mapas en los que se han utilizado los conjuntos de datos de la Oficina del Censo.
- [CIA World Factbook](#) : The World Factbook, es elaborado por la Agencia Central de Inteligencia y proporciona información sobre la historia, la gente, gobierno, economía, geografía, comunicaciones, transporte, militar y cuestiones transnacionales de 267 entidades del mundo.
- [Eurostat](#) : La misión de Eurostat es ser el principal proveedor de estadísticas de alta calidad a la Unión Europea y los países candidatos. https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&display=list&sort=category
- [Observatorio Mundial de la Salud](#) : Esta colección cuenta con más de 50 conjuntos de datos sobre temas prioritarios de salud como la mortalidad y la carga de las enfermedades, los Objetivos de Desarrollo del Milenio (nutrición infantil, salud infantil, la salud materna y reproductiva, la inmunización, el VIH / SIDA, la tuberculosis, el paludismo, las enfermedades olvidadas, agua y saneamiento), las enfermedades no transmisibles y factores de riesgo, enfermedades epidémicas, sistemas de salud, la salud ambiental, la violencia y las lesiones, la equidad, entre otros.

Algunas bases de datos públicas internacionales



- [Harvard Dataverse Red](#) : Esto es un repositorio para compartir, citando y la preservación de los datos de la investigación, abierta a todos los datos científicos de todas las disciplinas en todo el mundo. Incluye la mayor colección del mundo de los datos de la investigación en ciencias sociales.
- [Fondo Monetario Internacional \(FMI\) de datos](#) : El FMI (Fondo Monetario Internacional) publica una serie de datos de series de tiempo sobre los préstamos del FMI, de cambio y otros indicadores económicos y financieros.
- [NOAA: La Administración Nacional Oceánica y Atmosférica](#) : National Climatic Data Center de la NOAA (CNDC) es el mayor proveedor mundial de datos meteorológicos y climáticos. -Base en tierra, mar, modelo, radar, globo de tiempo, satélite y paleoclimáticos son sólo algunos de los tipos de conjuntos de datos disponibles.
- [Colección de datos del Banco Mundial](#) : El Catálogo de datos proporciona acceso a más de 8.000 descargas indicadores a partir de conjuntos de datos del Banco Mundial, una búsqueda por país, los indicadores, o tema.
- [NYC Open Data](#) : Esta colección cuenta con más de 800 conjuntos de datos que pertenecen a la ciudad de Nueva York, la mayoría de los cuales se puede ver como un mapa interactivo. Incluyen ubicaciones de graffiti, la ubicación de aseos en los parques públicos, lugares wifi, entradas de metro, y mucho más.

¿Qué es *Open Data*?

El concepto **datos abiertos** (*open data*, en inglés) es una [filosofía](#) y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de [derechos de autor](#), de [patentes](#) o de otros mecanismos de control. Tiene una ética similar a otros movimientos y comunidades abiertos, como el [software libre](#), el [código abierto](#) (*open source*, en inglés) y el [acceso libre](#) (*open access*, en inglés).



Iniciativa de *datos abiertos* del Gobierno de España

datos.gob.es es plataforma que organiza y gestiona el **Catálogo Nacional de Datos Abiertos**.

datos.gob.es ofrece un punto de acceso único a los conjuntos de datos de los diferentes niveles de administración (General, Autonómica, Local, Universidades, etc.).

Asimismo, desde *datos.gob.es* se proporciona información general, materiales formativos y noticias de actualidad sobre la reutilización de la información del sector público.



datos.gob.es
reutiliza la información pública

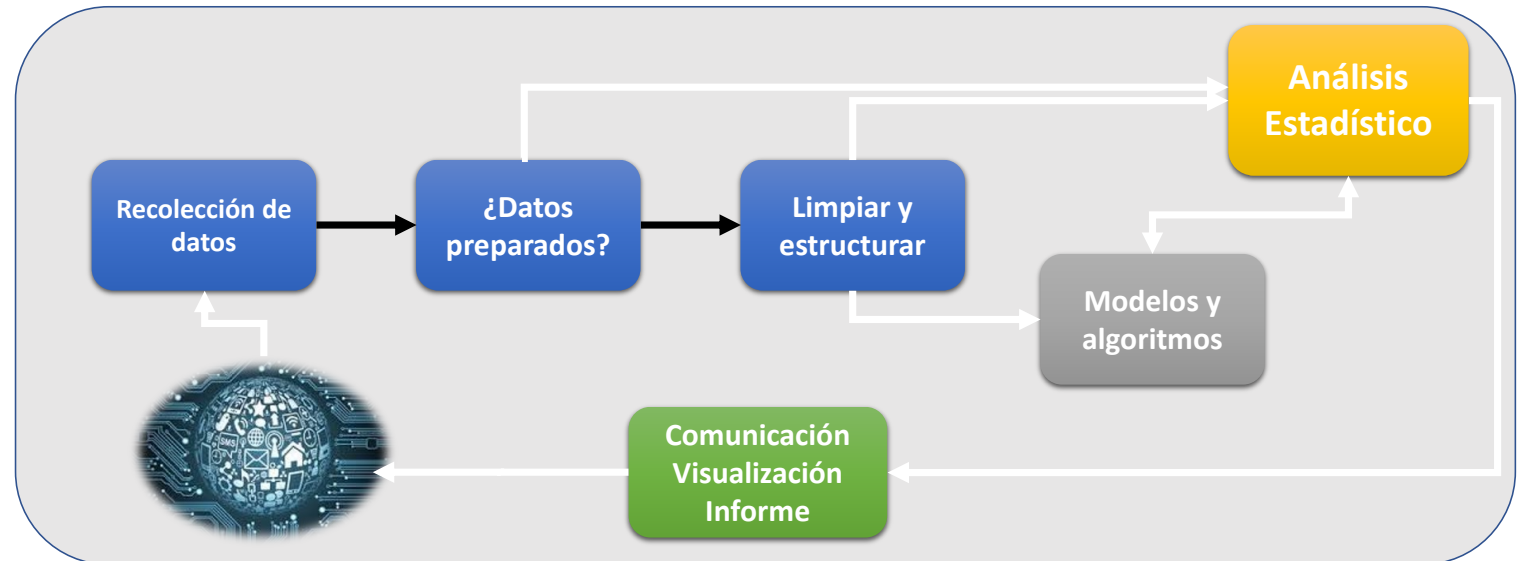
DATOS VS INFORMACIÓN

- Lo más importante en la ciencia de los datos es la pregunta, y la segunda cosa más importante son los datos
- Los datos determinan el tipo de preguntas que podemos hacer, pero son inútiles si no se formula la pregunta adecuada.
- **Los datos por sí mismos no contienen la respuesta, necesitamos (Big o Small) los datos y procedimientos adecuados.**

Herramientas informáticas



¿Qué hacemos con tanto dato?



¡QUE ALGUIEN ME AYUDE!

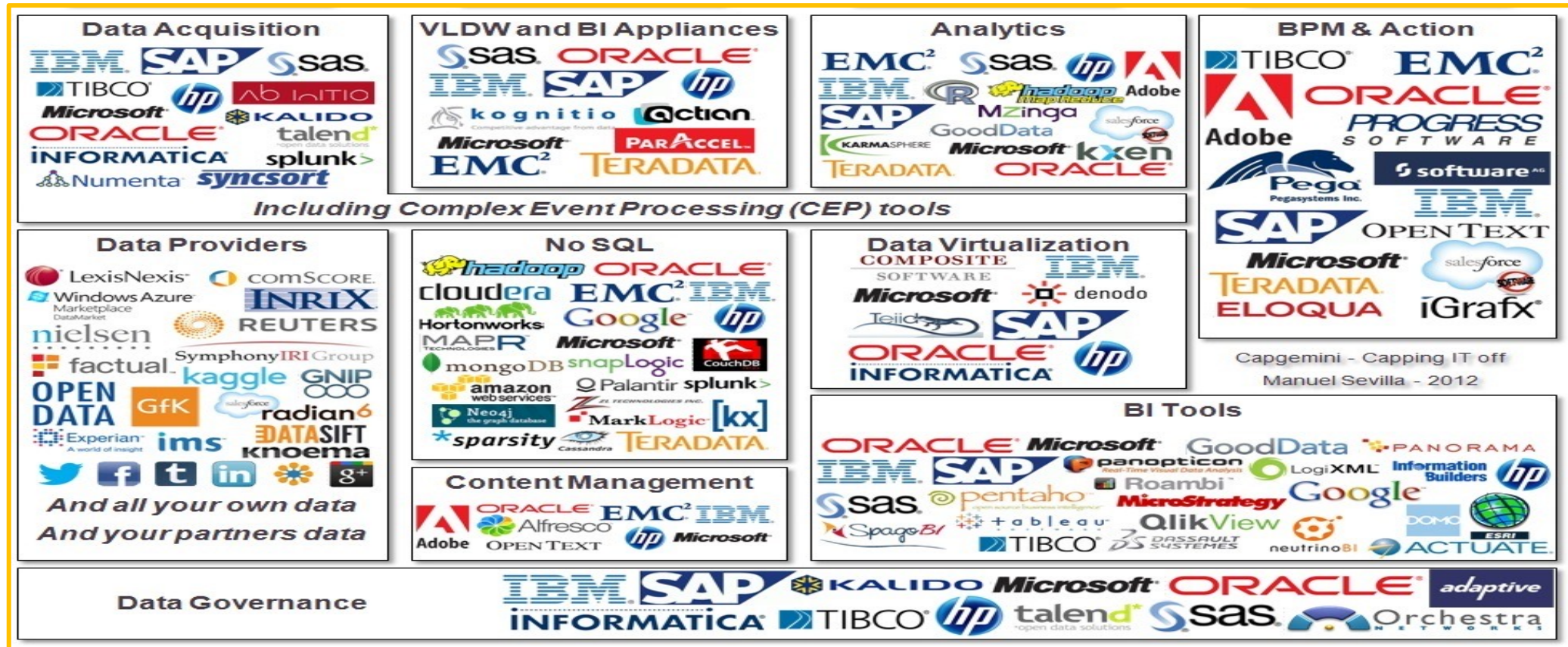
SOS

Adquisición

Preparación

Análisis

Resultados



CASO A: ESTUDIOS CIENTÍFICOS

(SALUD, BIOLOGÍA, SOCIOLOGÍA, ESTUDIOS DE MERCADOS...)

- Relativamente pocos datos
- Análisis simples sin necesidad de modelos personalizados
- No es necesario la iteración o replicar el cálculo muchas veces
- La mayoría de artículos científicos

Software Estadístico a base de clics



CASO GENERAL: REPORTING rápido y resumen de datos

- Representar gráficos
- Crear tablas resumen
- Pequeños cálculos
- Informes y reporting rápidos

HOJA DE CÁLCULO -
HERRAMIENTAS DE REPORTING



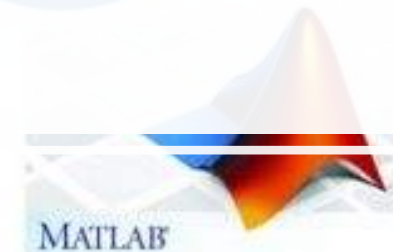
CASO B: CIENCIA DE DATOS


- Puede trabajar con grandes volúmenes de datos
- lectura de datos sea robusta y automatizada(no se introduce a mano)
- Análisis personalizados
- Necesidad de automatizarlos y robustez en el cálculo
- Flexibilidad

Software Estadístico a base de comandos (programación)



Software





Software para manejo y análisis de datos

- **Comerciales:**
 - Específicos:
 - SAS
 - Statgraphics
 - SPSS
 - Minitab, etc.
 - Generales:
 - MATLAB
 - Microsoft Excel también realiza muchas funciones estadísticas
 - ...
- **Software libre:** R, PSPP, etc

Herramientas estadísticas



Generar o recopilar datos



Analizar

Descriptivo

- Resumir y organizar

Inferencia

- Extrapolar conclusiones



Interpretar



Visualizar

Datos, Información y Herramientas

Fin

PRAKTIKUM 2024

**BIG DATA: Aplicaciones a Inteligencia Artificial, Ciberseguridad y
Knowledge discovery**