



ANÁLISIS CRIMINAL

EL CICLO DE LA INFORMACIÓN EN
LA ERA DEL BIG DATA Y LA
INTELIGENCIA ARTIFICIAL

Mg. Lic. Gaston Pezzuchi, Msc.

AGENDA (TÓPICOS)

- © Análisis Criminal (Ciencia del crimen)
- © Ciencia de Datos (Big Data, Inteligencia Artificial)
- © Pronósticos (Anticipación)
- © Complejidades
- © Buena Ciencia

El 12 de Febrero de 2002, al hablar sobre las armas de destrucción masiva en Irak, el Secretario de Defensa de los EEUU empleó la siguiente clasificación:

- *(i) “known knowns” (things we know we know),*
- *(ii) “known unknowns” (things we know we don’t know), and*
- *(iii) “unknown unknowns” (things we don’t know we don’t know).*

Donald Rumsfeld

¡GESTIÓN EN BASE A CONOCIMIENTO! - ¡GESTIÓN EN BASE A EVIDENCIA!

- © Es necesario reemplazar el “trabajar con más empeño” por el **“trabajar más inteligentemente”**.

Frederick W. Taylor (1856-1915)

- © La gestión es **imposible** en ausencia de conocimiento.

© Políticas Públicas Basadas en Evidencia

- Promover el uso del conocimiento científico como fuente de información para la toma de decisiones.
- Producción continua y progresiva de conocimiento para retroalimentar el ciclo de gestión de las políticas públicas.



Tomada del Banco de Evidencia en Seguridad y Justicia del BID

“El **análisis criminal** es la aplicación de métodos analíticos particulares a datos recolectados con el propósito de una investigación criminal, ya sea en el contexto de un proceso penal o en el ámbito de un trabajo de investigación científico-académico... Este tipo de análisis se realiza en agencias policiales, en agencias de defensa, en el ejército e inclusive en agencias de seguridad privada... En algunas organizaciones se lo emplea informalmente como parte del trabajo de los investigadores o de los fiscales, mientras que en otras se realiza únicamente por personal especialmente capacitado que suele denominarse “analistas de inteligencia”, “analistas del delito”, o “analistas de información criminal” ”...

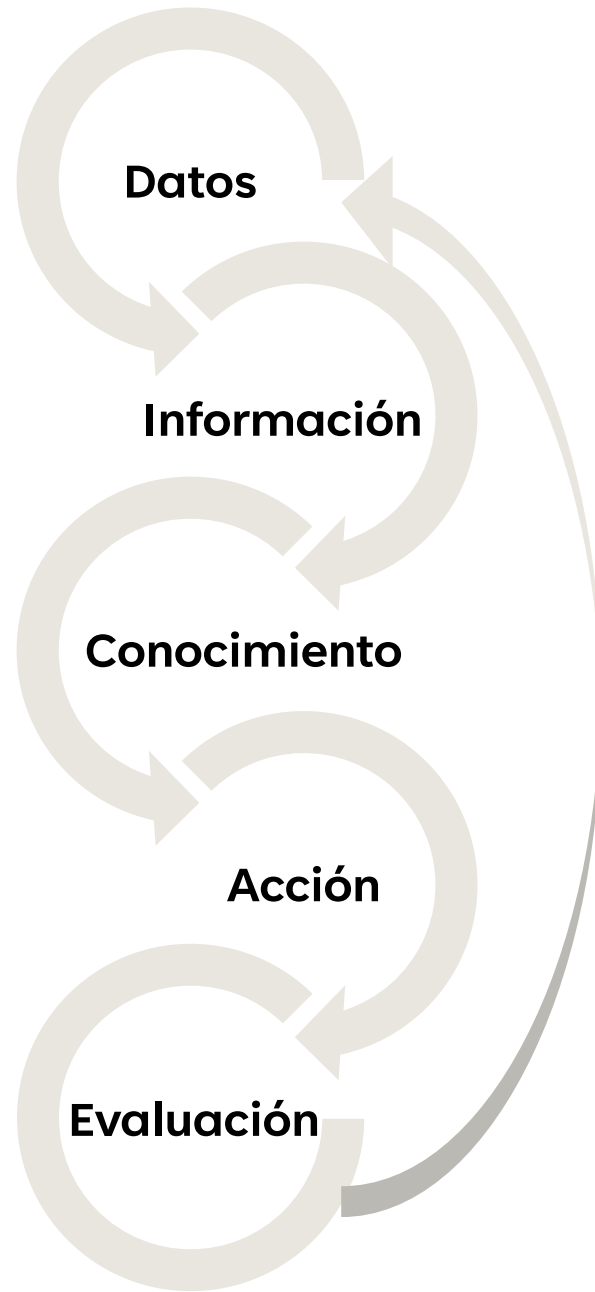
Marilyn B. Peterson, Applications in Criminal Analysis, a sourcebook (1994)

Análisis Criminal no es tan común como análisis de inteligencia.

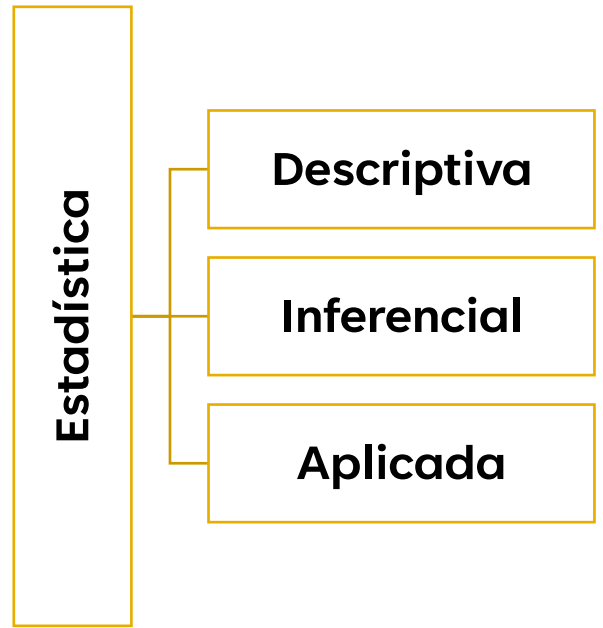
“La mayor parte de lo que se enseña como análisis de inteligencia es de hecho análisis para ser empleado como ayuda en una investigación penal y, consecuentemente, también se lo ha denominado “análisis investigativo”...

Marilyn B. Peterson, Applications in Criminal Analysis, a sourcebook (1994)

Prevencion del Delito



Investigacion Criminal



- En los últimos años la **ciencia de datos** (*data science*) ha emergido como una nueva e importante disciplina:
 - Puede ser considerada como una amalgama de disciplinas clásicas:
 - - Estadística
 - - Minería de Datos (Data mining)
 - - SGBD (DBMS)
 - - Sistemas Distribuidos
- **Convertir datos en valor** para individuos, organizaciones y la sociedad.
- Nuevos desafíos (“*Big Data*”) y nuevas preguntas a responder.

INTERNET DE EVENTOS (INTERNET OF EVENTS)

- La Sociedad ha cambiado de ser predominantemente “analógica” a predominantemente “digital” (Hilbert y López (2011)) en muy pocos años.
- Actividades de Rutina
- Impacto en la forma de hacer negocios y de comunicar (Manyika, 2011)
- La Sociedad, las organizaciones y las personas están “*Always On*”.
- Los datos se reúnen *sobre cualquier cosa, a cualquier momento y en cualquier lugar.*
 - BIG DATA...

M. Hilbert and P. Lopez. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, **332**(6025):60–65, 2011.

J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011.

INTERNET DE EVENTOS (INTERNET OF EVENTS)

- El estudio “*IDC Digital Universe Study*” (Abril 2014) confirma el crecimiento espectacular de los datos:
 - Ese estudio estima que la cantidad de datos digitales (en PC, cámaras digitales, servidores y sensores) almacenados en 2014 excedió 4 Zettabytes y predijo que el “*universo digital*” iba a crecer a 44 Zettabytes en el 2020.
- La tan anticipada *explosión de datos* ya ha ocurrido...

1 TB = 1.000 GB

1 PB = 1.000 TB

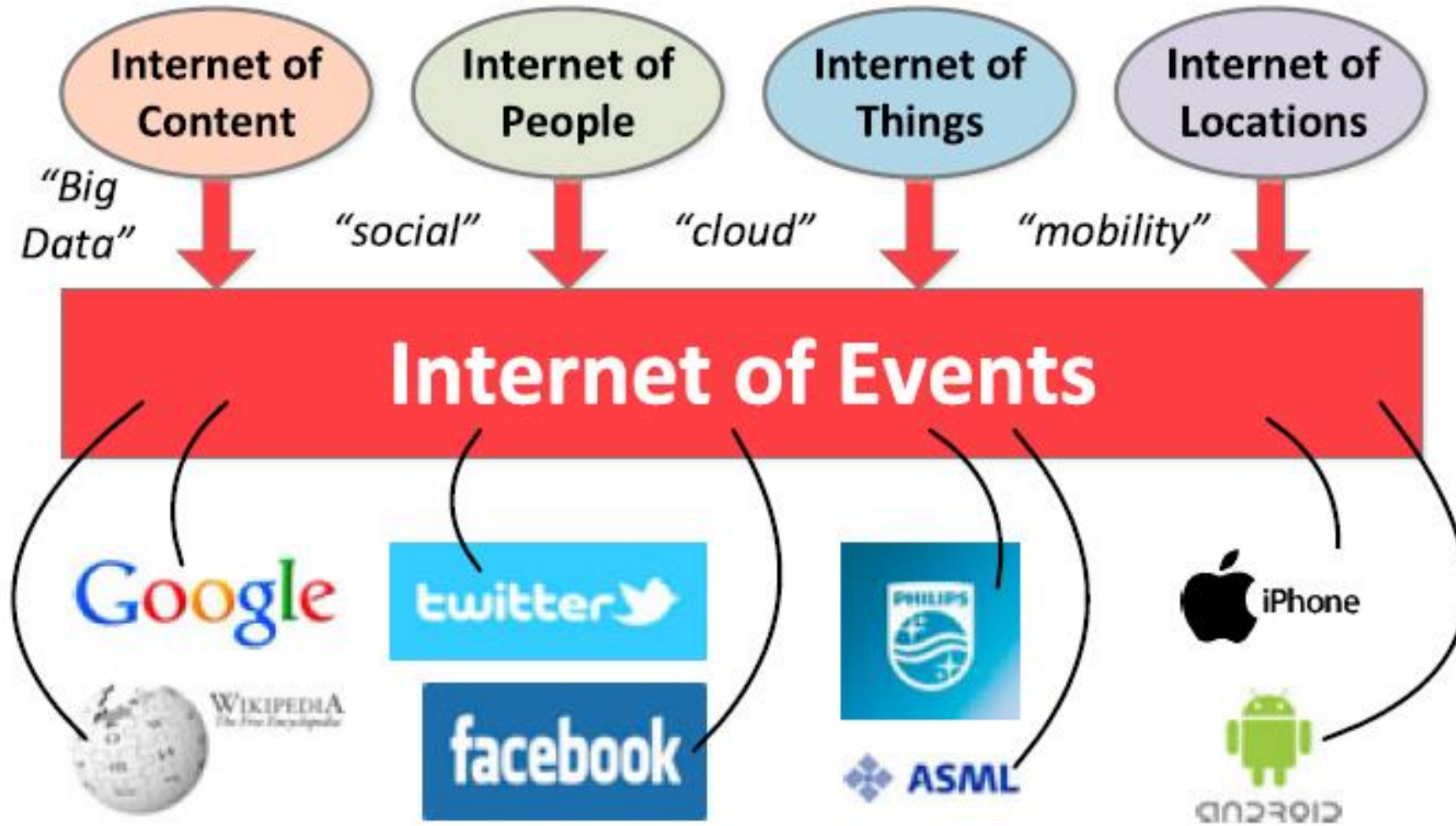
1 EB = 1.000 PB

1 **ZB** = 1.000 EB = 1.000.000 PB = 1.000.000.000 TB

DATOS NO ESTRUCTURADOS

- La gran mayoría de los datos almacenados en el “*universo digital*” es no-estructurada, y las organizaciones tienen problemas para tratar con estas grandes cantidades de datos.
- Uno de los principales desafíos actuales es *extraer información y valor a partir de los datos* almacenados en los sistemas de información.

INTERNET OF EVENTS



INTERNET DE LOS EVENTOS (IOE)

El termino *Internet of Events* (IoE), fue acuñado en 2014 y se refiere a todos los datos de eventos disponibles. En esencia la IoE está compuesta por:

La ***Internet del Contenido*** (IoC), se refiere a toda la información creada por los seres humanos para incrementar el conocimiento sobre un tema particular. (incluye las páginas web tradicionales, los artículos de enciclopedia como Wikipedia, YouTube, e-books, newsfeeds, etc.)

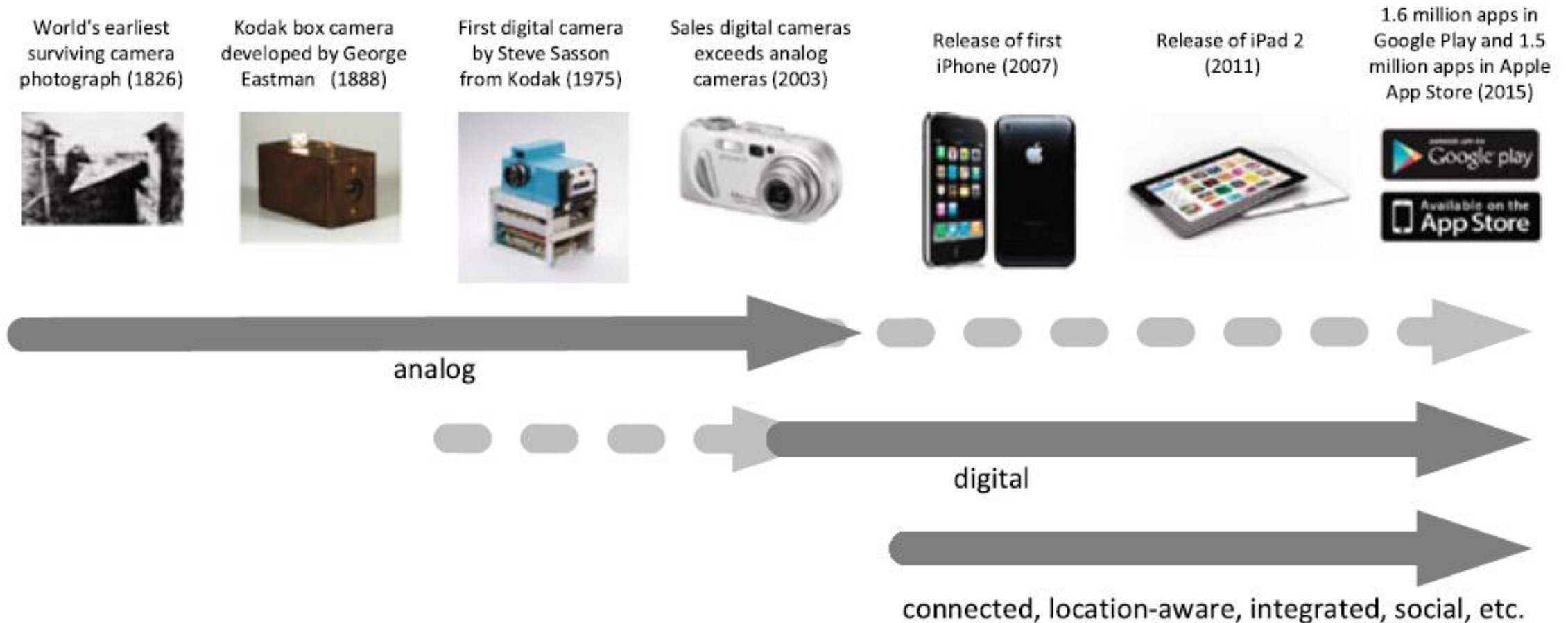
La ***Internet de las Personas*** (IoP), se refiere a todos los datos relacionados con la interacción social (incluye los e-mails, Facebook, Twitter, foros, LinkedIn, etc.)

La ***Internet de las Cosas*** (IoT), se refiere a todos los objetos físicos conectados a la red. Incluye todas las cosas que tienen un único ID y una presencia en una estructura similar a la Internet.

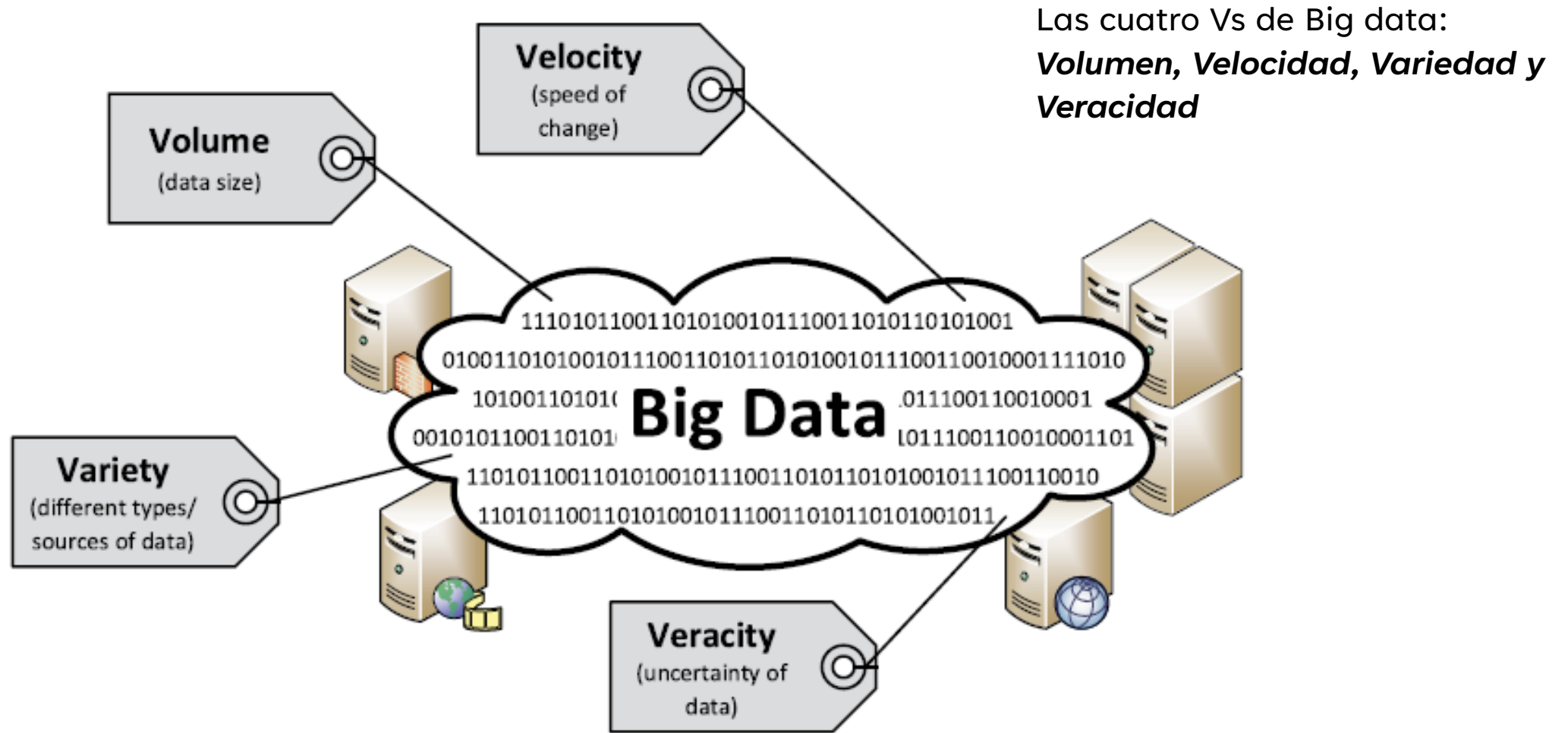
La ***Internet de las Localizaciones*** (IoL) se refiere a todos los datos que tienen una dimensión geográfica o geoespacial. Algo que con la masividad de los dispositivos móviles (por ej. Smartphones) más y más eventos tienen atributos de localización o de movimiento.

Es importante notar que IoC, IoP, IoT e IoL tienen muchas superposiciones. Por ejemplo, el nombre en una página web o la localización de la que un tweet se envía.

TRANSICIÓN DE ANALÓGICO A DIGITAL



Cambios dramáticos en la forma en que creamos y compartimos fotografías.



Las primeras tres V corresponden al trabajo de (Laney, 2001), luego se han propuesto varias otras Vs: *Variabilidad, Valor, Validez, etc.*

CIENCIA DE DATOS

- *La Ciencia de Datos* ha emergido como una nueva disciplina en los últimos años. De hecho, hay muchas definiciones que se han sugerido para ella. Ver por ejemplo (Donoho, 2015 y Forbes, 2013)
- ***“Data science is an interdisciplinary field aiming to turn data into real value. Data may be structured or unstructured, big or small, static or streaming. Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and predictions, and the exploitation of results taking into account ethical, social, legal, and business aspects.”***

CIENCIA DE DATOS

- Esa definición implica que la *ciencia de datos* es en principio un concepto más amplio que el de *estadística aplicada* y el de *minería de datos*.
- En líneas generales los “*cientistas*” de datos asisten a la organización para **transformar datos en valor**. Y se espera que respondan ciertas preguntas en base a los datos:
 - (Reporting) *What happened?*
 - (Diagnosis) *Why did it happen?*
 - (Prediction) *What will happen?*
 - (Recommendation) *What is the best that can happen?*

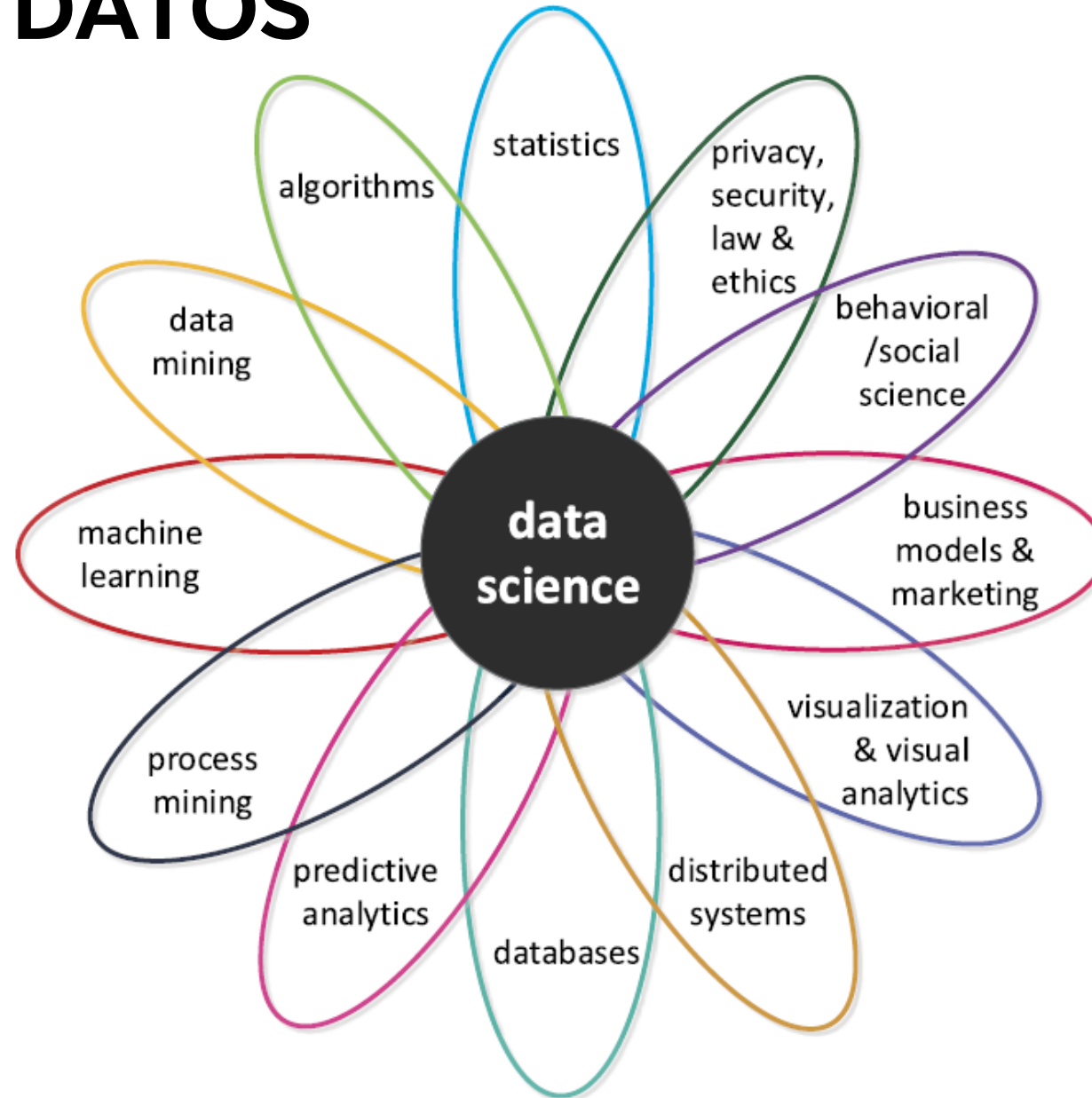
ANALISIS DE DATOS

- John Tukey (1915–2000), conocido por la transformada rápida de Fourier y los “*box-plots*” entre otras cuestiones, pero en 1962 escribió:
 - “For a long time, I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. . . . I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.”

ANALISIS DE DATOS

- Leo Breiman (1928–2005), otro eminente estadístico, escribió en 2001:
 - “This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics.”

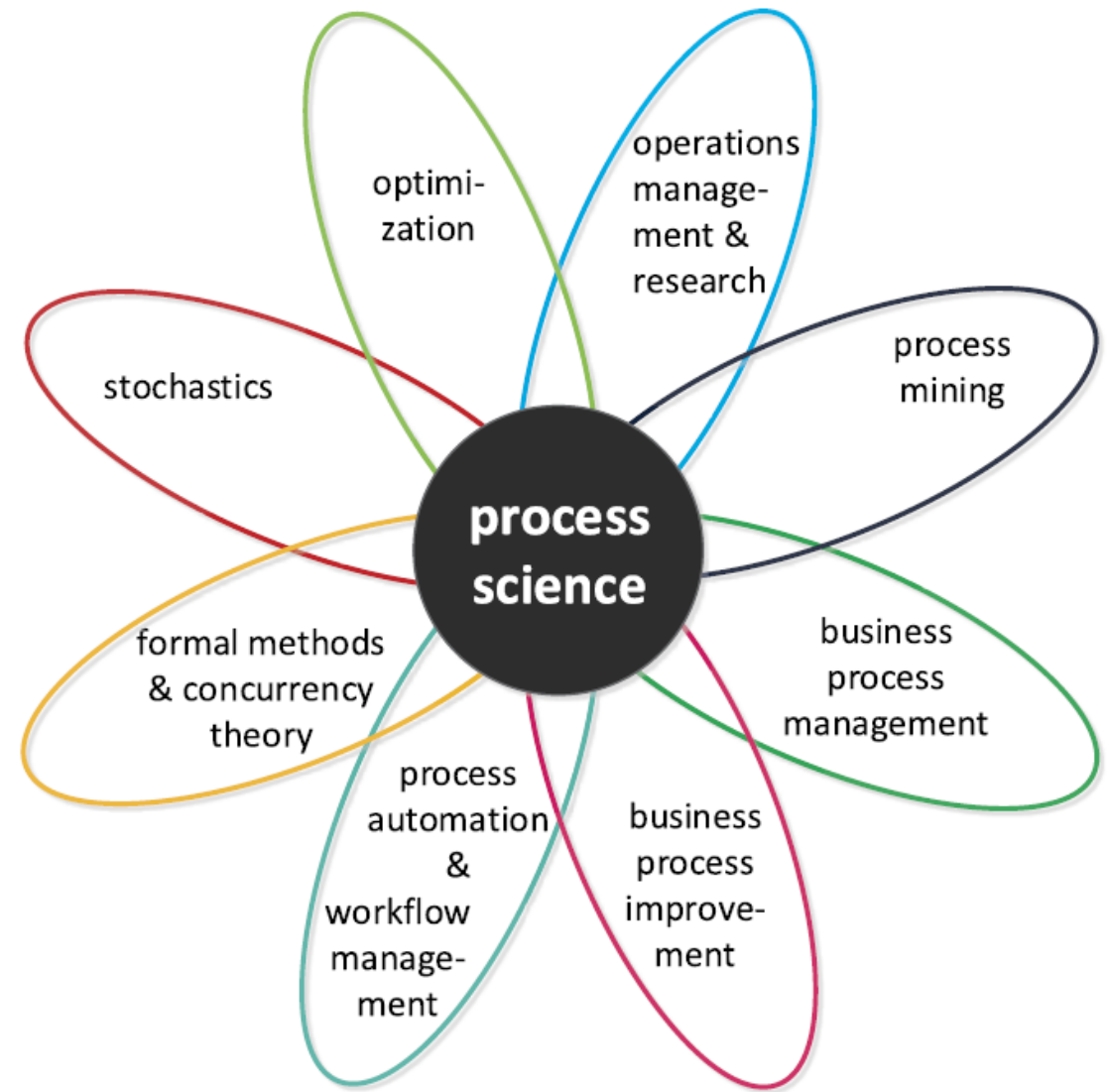
CIENCIA DE DATOS



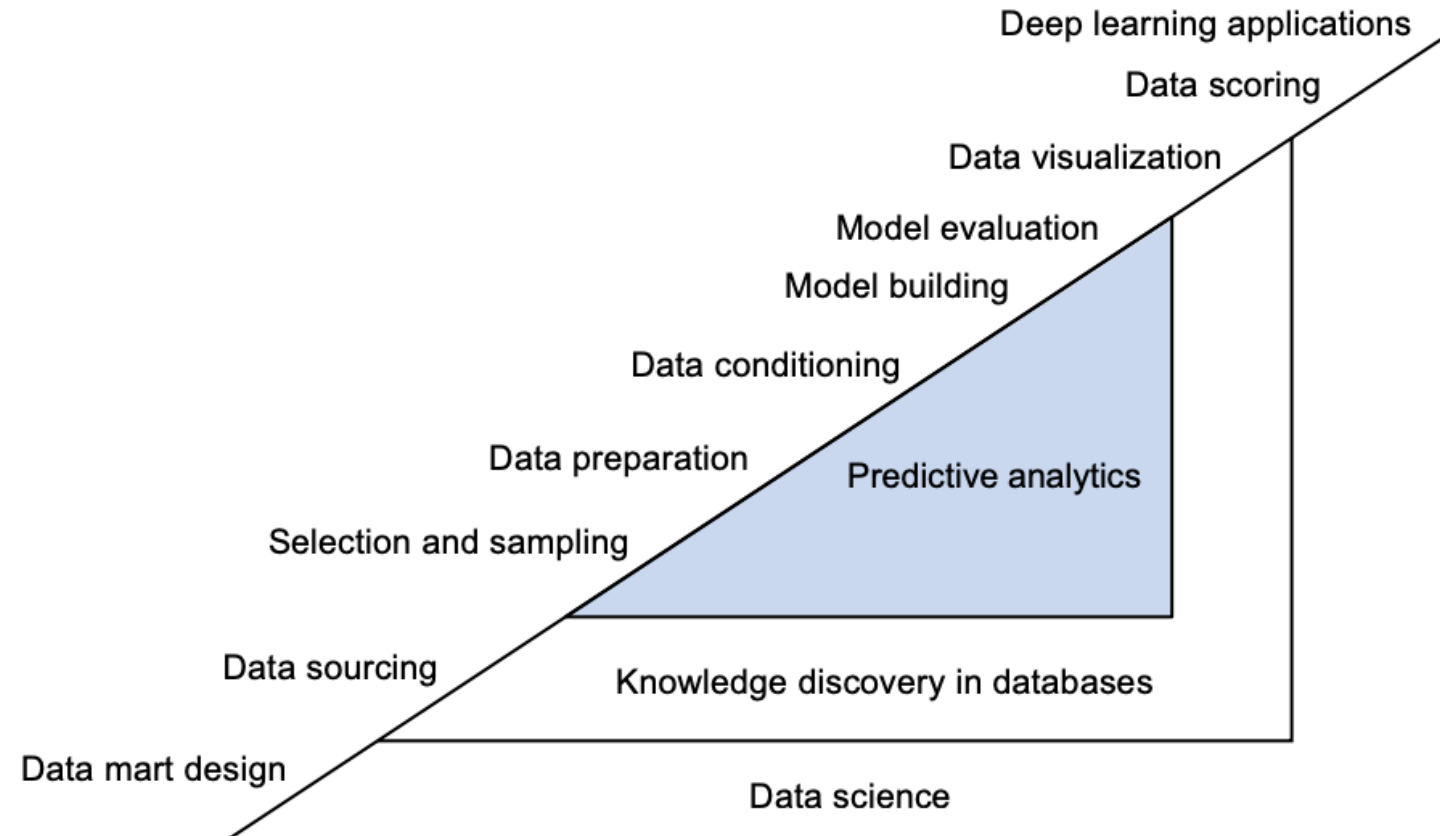
Ingredientes

PROCESS SCIENCE

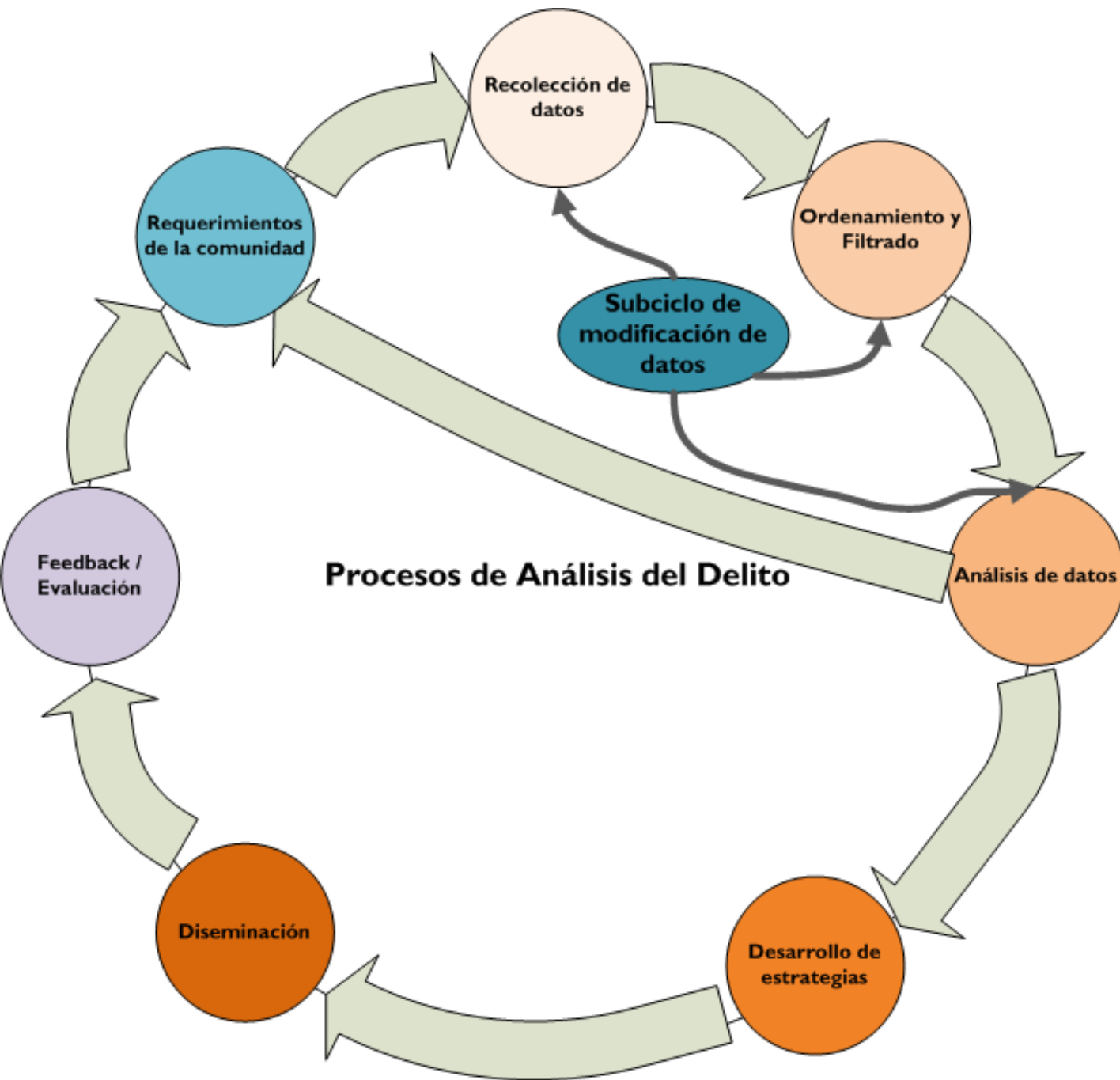
Como concepto envolvente para la disciplina que combina el conocimiento de la tecnología de información y el conocimiento de las ciencias de la administración para mejorar y realizar procesos operacionales.



WHAT IS DATA MINING?



Tomada de Nisbet et al. (2019) Handbook of Statistical Analysis and Data Mining Applications



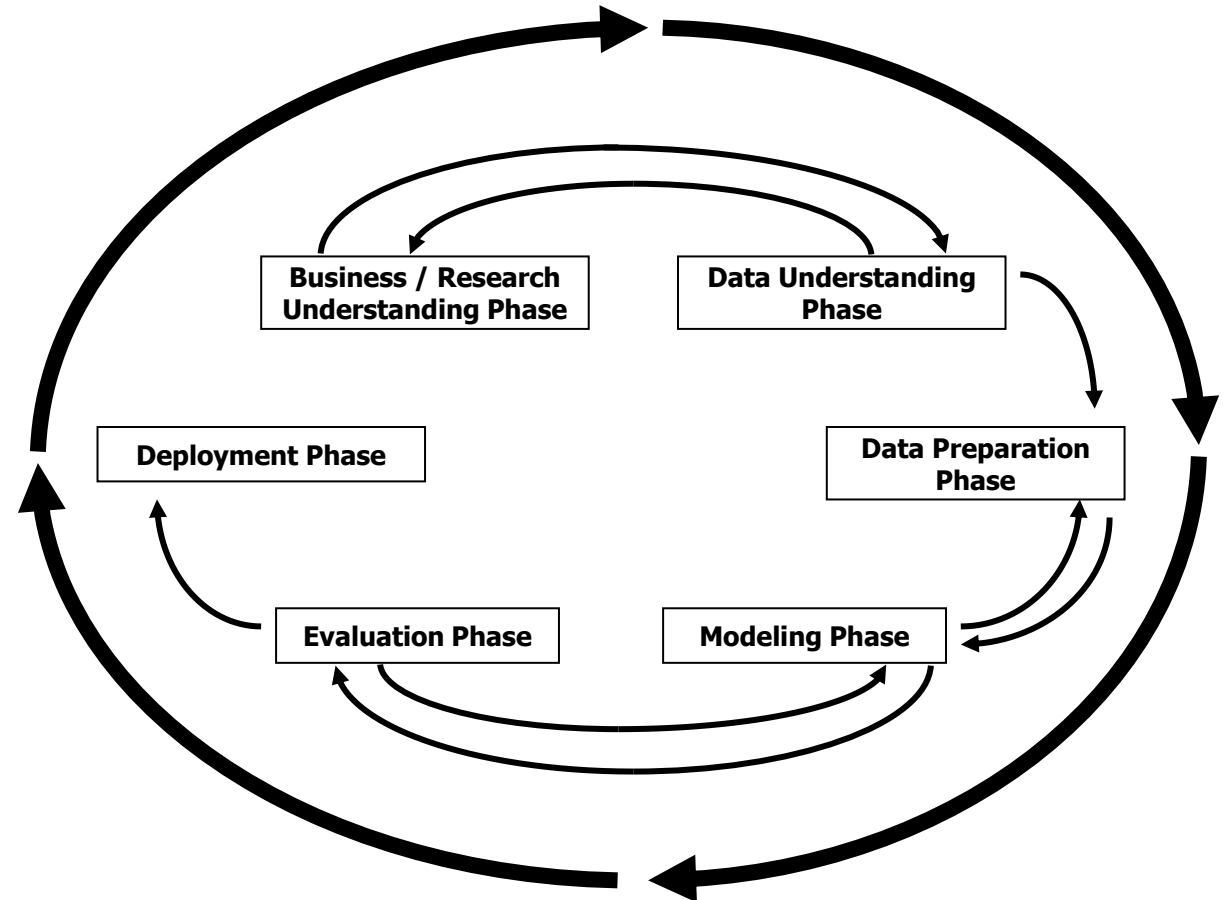
El **método científico** puede resumirse en:

1. Definir el problema.
2. Recopilar información existente sobre un fenómeno.
3. Formular una o más hipótesis.
4. Recopilar nuevos datos experimentales.
5. Analizar la información en el nuevo conjunto de datos.
6. Interpretar los resultados.
7. Sintetizar conclusiones, basadas en los datos antiguos, nuevos datos e intuición.
8. Formular nuevas hipótesis para pruebas adicionales.
9. Repetir el proceso (iteración).



CROSS INDUSTRY STANDARD PROCESS: CRISP-DM

- El Cross-Industry Standard Process for Data Mining (CRISP-DM) fue desarrollado en 1996
 - Considera a la minería de datos dentro de una estrategia general de resolución de problemas para una unidad de investigación o para un negocio.
 - En su desarrollo contribuyeron entre otros: DaimlerChrysler, SPSS, y NCR
 - Se lo considera un proceso neutro en las dimensiones
 - Campo disciplinar (Industria, etc.)
 - Herramienta
 - Aplicación
 - Es un proceso **NO** propietario y disponible en forma abierta.
 - Los proyectos de minería de datos siguen un ciclo iterativo, adaptativo de 6 fases.
 - Las fases secuenciales son **adaptativas**.



Ciclo de Vida de CRISP-DM

FALACIAS DE LA MINERÍA DE DATOS. - IA

	Falacias	Realidad
1	<ul style="list-style-type: none">• Es posible desplegar sin planificación un conjunto de herramientas en repositorios de datos• Encuentra soluciones a todos los problemas de negocios	<ul style="list-style-type: none">• No hay herramientas automáticas de minería de datos que resuelvan problemas.• Si hay diferentes procesos de minería de datos.• Se integra a los objetivos generales de negocios.
2	<ul style="list-style-type: none">• El proceso de minería de datos es autónomo• Requiere poca supervisión	<ul style="list-style-type: none">• Requiere una intervención humana significativa en cada fase del proceso.• Luego de que el modelo se despliega, nuevos modelos requieren actualizaciones.• Los analistas monitorean y evalúan continuamente
3	<ul style="list-style-type: none">• La minería de datos se paga sola rápidamente	<ul style="list-style-type: none">• Las tasas de retorno varían.• Dependen del punto de inicio, el personal, los costos de preparación de datos, etc.
4	<ul style="list-style-type: none">• El software de minería de datos es fácil de usar.	<ul style="list-style-type: none">• La facilidad de uso varía entre proyectos.• Los analistas deben combinar el conocimiento del tema con conocimiento específico del dominio del problema.

	Falacia	Realidad
5	<ul style="list-style-type: none"> La minería de datos identifica la cauda de los problemas de negocios. 	<ul style="list-style-type: none"> El proceso de descubrimiento de conocimiento es el que encuentra patrones de comportamiento Los seres humanos interpretan los resultados e identifican las causas
6	<ul style="list-style-type: none"> La minería de datos automáticamente limpia los datos en las bases de datos. 	<ul style="list-style-type: none"> La minería de datos muchas veces utiliza datos de sistemas heredados (<i>legacy</i>) Es posible que esos datos no hayan sido examinados en años Las organizaciones suelen iniciar tareas de minería de datos confrontándose con gigantescas tareas de preprocesamiento de datos.
7	<ul style="list-style-type: none"> La minería de datos siempre provee resultados positivos. 	<ul style="list-style-type: none"> No hay garantías de resultados positivos. Pero si la minería de datos se realiza adecuadamente, se obtendrán siempre resultados “accionables” y muchas veces redituables.

EJEMPLOS DE POLICIAMIENTO PROACTIVO

POLICIAMIENTO PREDICTIVO

Como pronosticar el delito?

POP

Requiere Analistas Criminales altamente capacitados

Definir PROBLEMA con claridad.

HSP

Focalizacion de recursos en areas geograficas especificas
(concentrando esfuerzos)

«El **Policiamiento Predictivo** es la aplicación de técnicas analíticas – particularmente técnicas analíticas cuantitativas – para identificar objetivos probables para la intervención policial y prevención del delito; o resolver delitos pasados mediante la realización de predicciones estadísticas.»

Perry et al. (2013)

POLICIAMIENTO PREDITIVO

1. ***Métodos de predicción de delitos:*** Se emplean para predecir lugares y momentos con mayor riesgo de ocurrencia de eventos delictivos.
2. ***Métodos para predecir delincuentes:*** Identifican a las personas en riesgo de ser futuros delincuentes.
3. ***Métodos para predecir la identidad de los perpetradores:*** Crean perfiles que relacionan con precisión a los posibles delincuentes con delitos específicos pasados (ya ocurridos).
4. ***Métodos para predecir víctimas de delitos:*** Se emplean para identificar grupos o, en algunos casos, individuos que probablemente se conviertan en víctimas de delitos.

MITOS SOBRE POLICIAMIENTO PREDITIVO

- *Mito 1: La computadora realmente conoce el futuro.*
- *Mito 2: La computadora hace todo.*
- *Mito 3: Se necesita un modelo potente (y caro).*
- *Mito 4: Las predicciones precisas conducen automáticamente a grandes reducciones del delito.*

TRAMPAS (OBSTACULOS)

Para ser útiles para la policía, los métodos predictivos deben aplicarse como parte de **una estrategia general de prevención del delito**.

Para garantizar que los métodos predictivos contribuyan de manera significativa, es necesario evitar ciertos escollos (peligros / trampas):

- 1: Centrarse en la precisión del pronóstico en lugar de su utilidad táctica.*
- 2: Depender de datos de baja calidad (Garbage In, Garbage Out –GIGO–)*
- 3: Incomprensión de los factores detrás del pronóstico.*
- 4: Subestimar los procesos de evaluación (no evaluar la efectividad de los pronósticos)*
- 5: Ignorar los derechos civiles y la privacidad de las personas.*

ANALISIS CRIMINAL

- Construcción de conocimiento sobre la realidad criminal con el objetivo de **PREVENIR EL DELITO**.
 - No es criminalística, no es criminología, implica conocer el delito, al delinciente y a la víctima para prevenir el crimen.
 - Requiere de BUENA CIENCIA, marcos éticos, transparencia y rendición de cuentas.
- El Analisis Criminal es más que DESCRIPCION, es fundamentalmente **PRONOSTICO**.
- “No se puede prevenir lo que no se conoce”...

POLICIAMIENTO FOCALIZADO EN ZONAS CALIENTES (*HOT SPOTS*)

- El HSP es una estrategia eficaz de reducción del delito.
- Lo que la policía hace (tácticas policiales) en las zonas calientes, es importante.
 - *Curva de Koper*
 - Detenciones proactivas en las zonas calientes de entre 10 a 16 minutos, maximizan la disuasión.
 - Las visitas a las zonas calientes deben ser aleatorias e intermitentes.
- Existen varios beneficios adicionales al considerar la curva de Koper además de la Ley de Weisburd de concentración del delito en los lugares.

PRONÓSTICOS ESPACIALES DEL DELITO

- La criminología ambiental provee una base teórica importante para entender las distribuciones espaciales del delito.
 - Atractores del delito
 - Generadores del delito
- } El delito **no es aleatorio** en el espacio ni en el tiempo
- **Teoría de los Patrones Criminales** (Brantingham y Brantingham, 1984)
 - Explica el motivo por el que el delito ocurre en zonas específicas (no es aleatorio)
 - Organizado u Oportunista
 - Nodos de Actividad

PRONÓSTICOS ESPACIALES DEL DELITO

- La mayoría de las técnicas se emplean para pronósticos **retrospectivos**, es decir, predecir el futuro a través de datos históricos.
 - Datos históricos
 - Datos históricos + atractores / generadores
- Datos estáticos (demográficos / socioeconómicos)
- Datos dinámicos (redes sociales, flujos de pasajeros en taxis, etc.)

TODOS LOS PRONÓSTICOS SIGUEN ESTE PRINCIPIO

Sea:

- D_t el delito observado en el tiempo t
- E_{t+1} el delito estimado en el tiempo $t + 1$
- $V_{estática}$ es una variable explicativa o variables que no cambian entre t y $t+1$
- $V_{dinámica_lag}$ es una variable explicativa o variables que cambian entre t y $t+1$ y *lag* es un periodo de tiempo anterior al periodo de la variable dependiente.
 - Se modela D_t para derivar E_{t+1} que es evaluado contra D_{t+1}

Este principio puede ser aplicado a cuatro aproximaciones al pronóstico:

- 1- D_t se modela para derivar E_{t+1} que se evalúa contra D_{t+1}
- 2- D_t y $V_{estática}$ se modelan para derivar E_{t+1} que es evaluado contra D_{t+1}
- 3- D_t y $V_{dinámica_lag}$ se modelan para derivar E_{t+1} que es evaluado contra D_{t+1}
- 4- D_t , $V_{estática}$ y $V_{dinámica_lag}$ se modelan para derivar E_{t+1} que es evaluado contra D_{t+1}

CLASES DE TÉCNICAS PREDICTIVAS

Categoría Analítica de su Aplicación Primaria	Técnica Predictiva	Clase			
		Clásica	Simple	Compleja	Adaptada
Análisis de Zonas Calientes Predecir <i>DONDE</i> solo en base a datos retrospectivos	Mapeo en Grillas	X			X
	Elipses de Cobertura (NNHC)	X			
	KDE	X			
	Heurísticas		X		X
Métodos de Regresión Predecir <i>DONDE</i> usando rangos de datos	Lineal	X	X		
	A pasos (<i>stepwise</i>)	X		X	
	<i>Splines</i>			X	X
	Indicadores Principales (<i>Leading Indicators</i>)	X			X
Minería de datos Predecir <i>DONDE</i> usando rangos de datos	Conglomeración (<i>Clustering</i>)	X		X	
	Clasificación	X		X	
Cuasi-Repetida (Near Repeat) Predecir <i>DONDE</i> para los próximos días usando solo datos delictivos	<i>Self-Exciting Point Process</i>			X	
	<i>ProMap</i>			X	
	Heurísticas		X		
Análisis Espacio-Temporal Predecir <i>DONDE</i> usando datos delictivos y temporales	Mapas de Calor	X	X		X
	Modelos Aditivos			X	
	Estacionalidad	X			
Modelos de Riesgo de Terreno Predecir <i>DONDE</i> usando la geografía del riesgo	Análisis predictivo Geo-Espacial			X	X
	<i>RTM</i>		X		X

MÉTODOS PROPUESTOS USUALMENTE EN LA LITERATURA ACADÉMICA

Top 4 Propuestos	Top 4 Mejores	Top 4 Linea Base
Random Forest	Random Forest	Autoregressive model-based
Multilayer Perceptron	Multilayer Perceptron	Logistic Regression
Kernel Density Estimation-based (KDE) Support Vector Machines (SVM)	Kernel Density Estimation-based (KDE) Risk Terrain Modelling (RTM)	Autoregressive Integrated Moving Average (ARIMA), Multilayer Perceptron, Linear Regression, KDE-Based, KNN

No*	Authors and date	Space		Time		Crime Data		Forecasting			
		Study area	Scale	Sampling period	Months	Type	Sample	Inference	Task	Spatial unit	Temporal unit
1	Araujo Junior et al. (2017)	Natal, Brazil	City	2006–2016	132	<i>U</i>	<i>U</i>	# of crimes	Regression	Rectangular grid (<i>U</i>), districts	Week
2	Araújo et al. (2018)	Natal, Brazil	City	2006–2016	132	<i>U</i>	<i>U</i>	Hotspots	Binary classification	k-means cells of varying size (<i>U</i>)	Week

Análisis criminal

- ⊙ Existen diferentes tipos, niveles y actores involucrados en las respuestas que surjan como resultado de los análisis y el volumen actual de datos no hace sino aumentar las complejidades que enfrentan los analistas.
- ⊙ El proceso de análisis criminal supone una clara diferenciación entre qué tipo de información se requiere y con qué fines. Entender el motivo del análisis es el primer paso para producir análisis de calidad aceptable.
- ⊙ Definir cuáles son los datos necesarios, y su calidad mínima que sirva a estos objetivos junto con la capacitación para su procesamiento, lectura y utilización es un punto usualmente no atendido en el contexto del Análisis Criminal. (Una falla en cualquiera de estos componentes supone un fracaso en los objetivos planteados).
- ⊙ El apuro con el que se suele enfrentar el problema de la criminalidad en las agencias encargadas de su prevención y represión penal nos confronta cotidianamente con varios pasos en falso en estos aspectos.

GRACIAS POR SU TIEMPO
¿PREGUNTAS?

gpezzuchi@austral.edu.ar