

Sociedad digital

Tema 1.

Introducción a la Sociología y la sociedad de la información

Susana Rodríguez Díaz

susana.rodriguez@u-tad.com

1.4. PRIVACIDAD PERSONAL Y BIG DATA. ECONOMÍA DE LA ATENCIÓN

▪ ¿Qué es el Big Data?

- El Big Data es **el análisis masivo de datos**. Se trata de una cantidad de datos tan sumamente grande que las aplicaciones de software de procesamiento de datos que tradicionalmente se venían utilizando no pueden capturar, tratar y ponerla en valor en un tiempo razonable.
- Este término se utiliza también para referirse a las **nuevas tecnologías que hacen posible su almacenamiento y procesamiento** (además del uso que se hace de la información obtenida a través de dichas tecnologías).



▪ ¿De dónde vienen los datos?

- **Producidos por personas.** Mandar un email, hacer un comentario en Facebook, responder a un WhatsApp o hacer clic en un enlace de Internet son algunas de las acciones que generan datos.
- **Entre máquinas.** Entre máquinas se comparten datos en lo que se conoce como M2M (*machine to machine*). Por ejemplo, parquímetros, GPS o máquinas expendedoras de bebidas y alimentos, se comunican con otros aparatos a los que transmiten los datos que van recogiendo.
- **Biométricas.** Son datos que proceden de sensores de huellas dactilares, escáneres de retina, lectores de ADN, sensores de reconocimiento facial o reconocimiento de voz. Se usan en seguridad (privada, corporativa, militar, policíaca o de servicios de inteligencia).



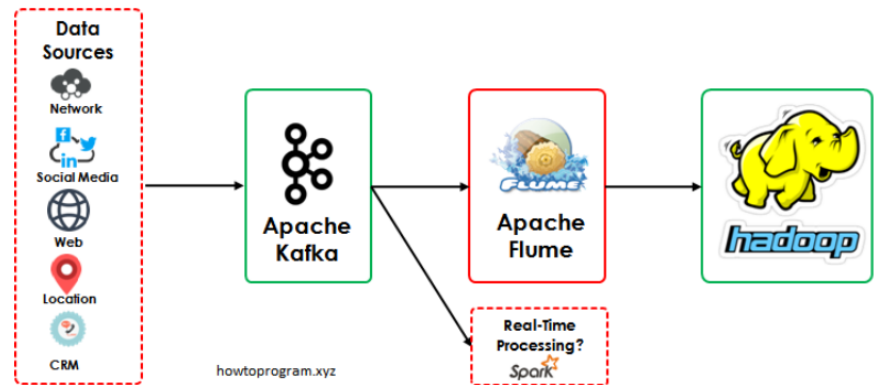
- **Marketing web.** Los movimientos en la Red están sujetos a todo tipo de mediciones que tienen como objeto estudios de marketing y análisis de comportamiento. Por ejemplo, cuando se realizan mapas de calor basados en el rastreo del movimiento del cursor por parte de los usuarios de una web, en la detección de la posición de la página, o en el seguimiento de desplazamiento vertical a lo largo de esta. Con esos datos se llega a conclusiones tales como qué partes de una página atraen más al usuario, dónde hace clic o en qué zona de esta pasa más tiempo.
- **Transacciones de datos.** Por ejemplo, el traspaso de dinero de una cuenta bancaria a otra, la reserva de un billete de avión o añadir un artículo a un carrito de compra virtual de un portal de comercio electrónico.





■ Tipos de datos según su estructura

- **Datos estructurados.** Son los que tradicionalmente se han utilizado en el tratamiento de datos. Se pueden almacenar en tablas y tienen una clara definición de longitud y formato (por ejemplo, números, cadenas de caracteres y fechas).
- **Datos no estructurados.** No poseen un formato específico que permita almacenarlos de forma tradicional (por ejemplo, emails, presentaciones multimedia, archivos de procesadores de textos o en formato PDF).
- **Datos semiestructurados.** Siguen una especie de estructura y poseen ciertos patrones comunes, pero esta no es suficientemente regular. Por ejemplo, el HTML (lenguaje para la elaboración de páginas web), ya que su sistema de etiquetas permite detectar pautas comunes.



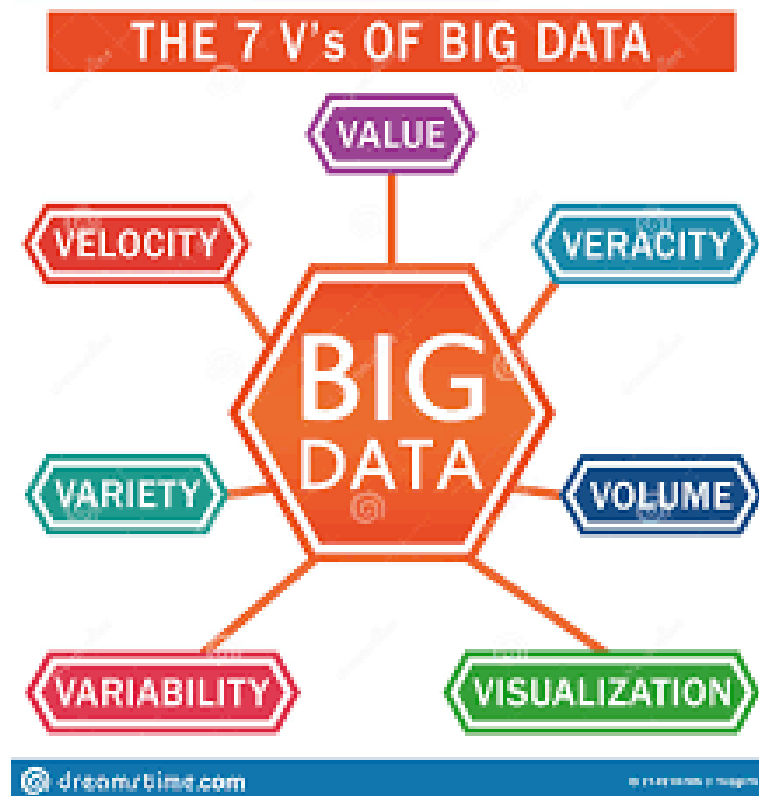
■ ¿Cómo se gestionan?

- **Captura de información.** Existen varios métodos como el Web Scraping (técnica que mediante programas de software extrae información en sitios web), gestión de la información a través de API creadas para ello, o servicios como Apache Flume (diseñado para recopilar grandes volúmenes de datos).
- **Almacenamiento.** Dependiendo del tipo de información y del uso se puede optar por hojas de cálculo para información estructurada, o por sistemas NoSQL (que permiten almacenar información no estructurada de forma rápida y flexible).
- **Tratamiento.** Desde tratamientos sencillos a sistemas predictivos complejos. Se puede buscar patrones repetitivos a través de la estadística y el *machine learning*.
- **Puesta en valor.** El valor no está en los propios datos, sino en la relación que se establezca entre ellos, ya que nos permiten extraer patrones en muchos campos.

■ Características

Sus características se acuñan popularmente como las «V» del Big Data. No hay un consenso sobre cuántas “V” han de ser tomadas en consideración, pero las 7 más extendidas son:

- **Velocidad.** La velocidad de análisis es una de las características fundamentales que tienen los datos a gran escala.
- **Volumen.** La cantidad de datos generados está aumentando.
- **Veracidad.** La fiabilidad de la información recogida es importante.
- **Visualización.** Convertir cientos de hojas de información en un único gráfico que muestre claramente unas conclusiones predictivas.
- **Variedad.** Los datos son diversos, ya que proceden de numerosas fuentes y se encuentran en distintos formatos.
- **Valor.** Una gran cantidad de datos frecuentemente extrae pequeñas informaciones de valor.
- **Variabilidad.** En un entorno tan cambiante como el de los macrodatos, la información varía mucho.



- [Mare Nostrum.](#)
- [Archivo Mundial del Ártico.](#)

▪ Algunas aplicaciones

- **Amazon.** Basa buena parte de su éxito en conocer de antemano lo que necesitan sus clientes; a través de la inteligencia de datos se analizan factores de carácter individual del usuario como hábitos de compra o intereses, y otros más generales como tendencias del momento o pautas de conducta de otros usuarios que adquirieron productos similares, sugiriendo productos.
- **Netflix.** Utiliza los macrodatos para crear nuevos contenidos para sus usuarios, cuyos hábitos de consumo y preferencias son observados para descubrir qué es lo que van a querer ver a continuación en base a patrones predictivos.



- Dentro de los procesos electorales se puede la percepción que tienen los votantes sobre los candidatos. De hecho, la forma de hacer campaña cambia: ya no son tan decisivos los grandes mítines, sino un acercamiento mucho más segmentado que tiende a individualizar a los votantes para **conseguir llegar con un mensaje más personal**, cercano y directo a cada uno de ellos. Uno de los grandes ejemplos del uso de la inteligencia de datos en campaña electoral con éxito fue el de Barack Obama para lograr su segundo mandato como presidente de los Estados Unidos de América.

- [Cambridge Analytica.](#)



- En el campo médico, hay múltiples avances que pueden ser posibles gracias a los datos a gran escala. Por ejemplo, el estudio del material genético de los organismos permite conocer cómo funcionan los genes y su influencia en las enfermedades. Analizando los genes de una persona a través de los macrodatos, se puede **aplicar de forma individual una medicina más eficaz**, con un diseño de tratamientos a la carta adaptados a cada paciente.



- Big Data de las redes sociales para predecir el comportamiento ciudadano.
- Big Data en el fútbol.
- Big Data para acabar con los atascos.

- Cloud computing.



Para ampliar información: “[Big Data](#)”. Facultad de Estudios Estadísticos. UCM.

Ver, analizar y debatir:

- [Datos masivos, datos mucho mejores](#). Kenneth Cukier.
- [Cómo pirateé el sistema de citas de Internet](#). Amy Webb.

▪ El concepto de privacidad

- La **privacidad** puede ser definida como el ámbito de la vida **personal** de un individuo que se desarrolla en un espacio reservado que tiene como propósito principal mantenerlo confidencial.
- **Privacidad digital** es el derecho que tiene cualquier usuario de la web a decidir cuáles datos personales desea compartir y cuáles desea mantener resguardados para proteger su intimidad.
- El concepto de **privacidad digital** surgió al mismo tiempo que **internet** y su capacidad para recopilar y compartir datos.



▪ Características de la privacidad digital

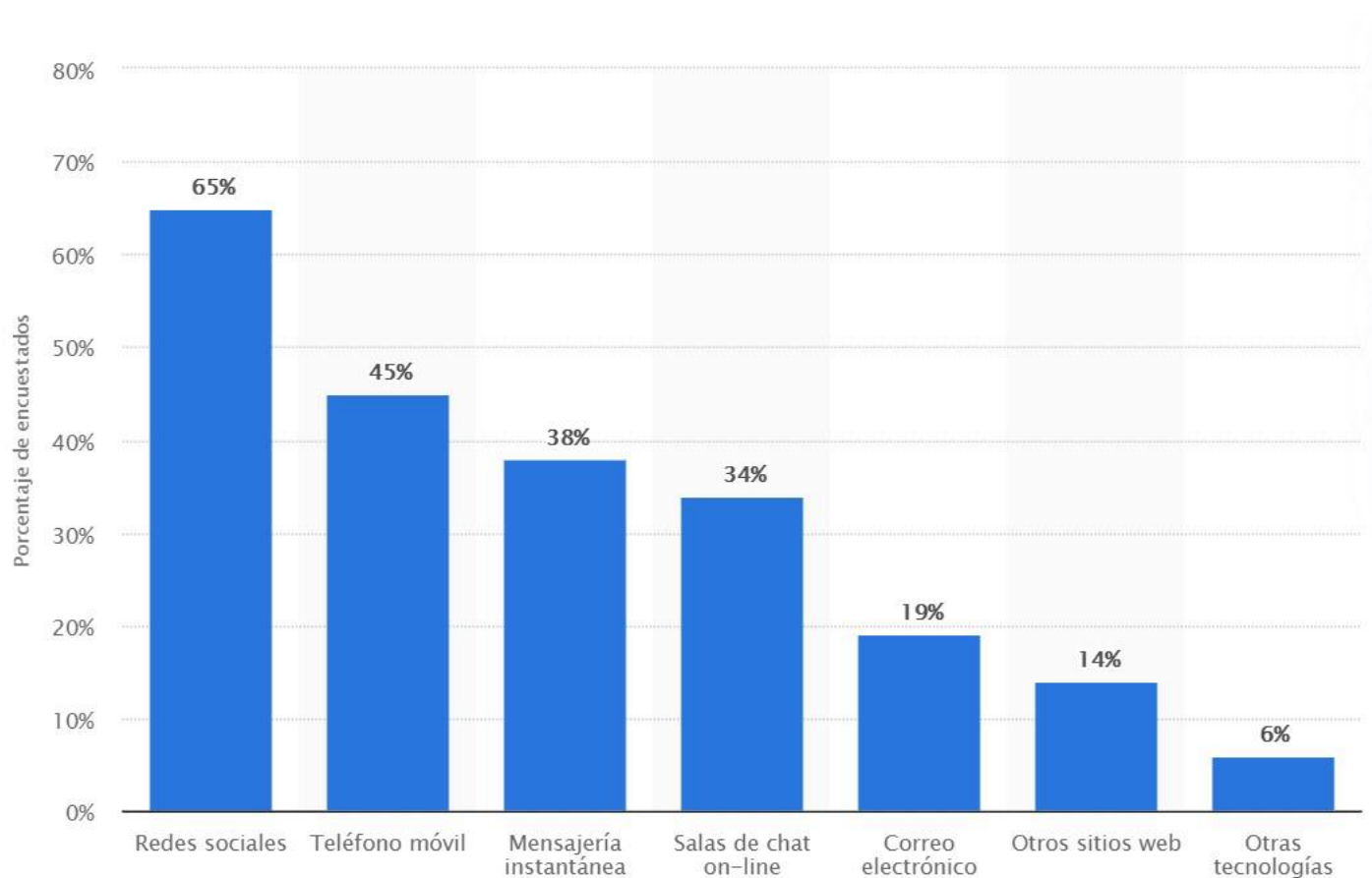
- Se refiere a toda la información de un usuario que circula por internet, además de datos personales como nombre, DNI, teléfono o domicilio. Por ejemplo, imágenes, vídeos, correo electrónico, geolocalización, historial de navegación, IP o cualquier otro dato que permita la identificación de un usuario en la red.
- También implica el uso de páginas web o redes sociales, y transmisión de datos a través de tiendas online, aplicaciones o servicios de mensajería.
- La **privacidad digital en España** está regulada por la Ley Orgánica de Protección de Datos y Garantía de Derechos Digitales (LOPDGDD), la cual adapta a la normativa española el Reglamento General de Protección de Datos (RGPD), que se aplica a nivel europeo.



- ¿Nos preocupa la privacidad de nuestros datos?
- ¿Qué hacemos para protegerla?

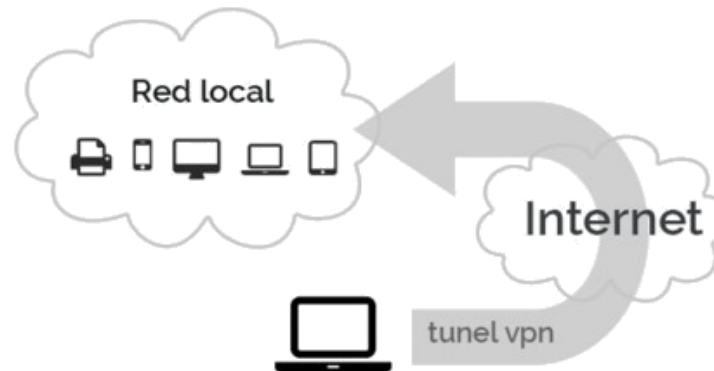
- **Phishing:** es un método de abuso informático que consiste en conseguir información confidencial de manera fraudulenta para suplantar la identidad de un usuario.
- **Spam:** es la recepción masiva de correos electrónicos no deseados.
- **Virus y troyanos:** son programas que se cuelan en el ordenador del usuario para llevar a cabo acciones no solicitadas como robo de información o borrado de datos.
- Chema Alonso:
 - Tu iPhone es tan inseguro como tu Windows.
 - You are where you are.
 - Tus datos mueven el mundo.
- Everyday cybercrime- and what you can do about it. James Lyne.

- **Ciberacoso:** el cyberbullying o acoso virtual consiste en el uso de internet o las redes sociales para amenazar, acosar o chantajear a una persona.

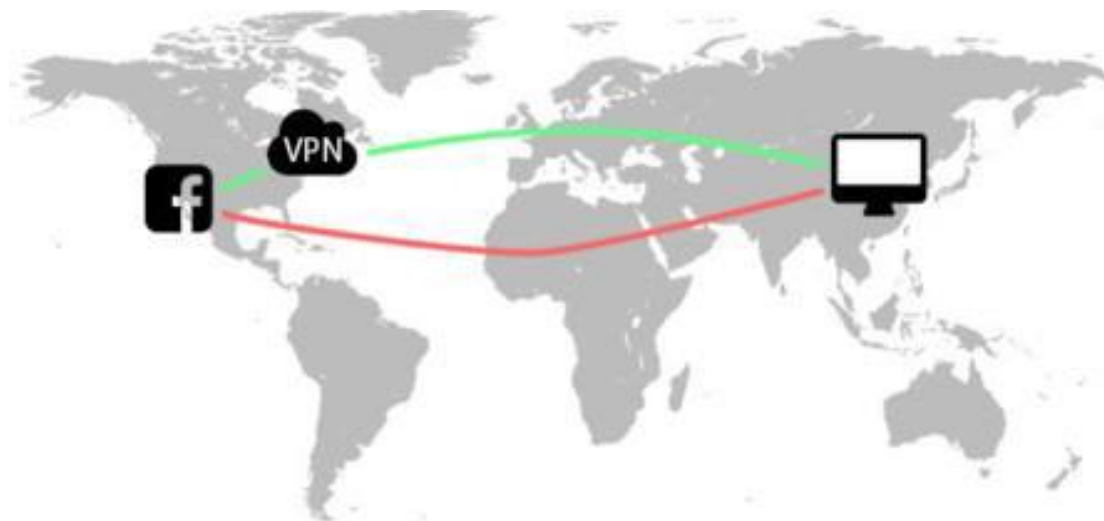


■ VPN

- **Virtual Private Network.** Permite una extensión segura de la red de área local sobre una red pública o no controlada como Internet.
- Permite crear una red local **sin necesidad que sus integrantes estén físicamente conectados entre sí**. Se obtienen las ventajas de la red local con una mayor flexibilidad, pues la conexión es a través de Internet y puede ser de una punta del mundo a la otra.



- Las conexiones VPN suelen tener **cifrado** de los paquetes que se transmiten con ellas, por lo que si se necesita conectarse a un Wi-Fi público, se pueden utilizar. También permite falsear la ubicación.
- Además, los paquetes varían de cifrado.
- El inconveniente es que el servidor de VPN debe ser confiable, ya que puede capturar el tráfico de datos y a veces la navegación es más lenta.
- Qué es una VPN y para qué sirve.



- **Economía de la atención**
 - La economía de la atención en la era digital.