

La Privacidad en la Era de Internet

Autor: Miguel Toro

Catedrático de la Universidad de Sevilla

<http://blogmigueltoro.blogspot.com.es/>

Introducción

Hoy día todo está en Internet. Se suele decir el que no aparece en Internet no existe. La persona, idea, producto o negocio que no aparece en una búsqueda en Internet se convierte en algo irrelevante. Esto es debido a la globalización, a la expansión de las comunicaciones y últimamente a la explosión de los *Smartphone*.

Junto a esa realidad se está produciendo otra los datos están siendo almacenados en la *nube (the Cloud)*. Es decir nuestros datos tienden a no estar en nuestro disco duro. Más bien están guardados en almacenes externos gestionados usualmente por empresas privadas. Es lo que se llama la nube. Ejemplos son *DropBox*, etc. Pero también están en la nube los datos de nuestras aplicaciones: *Facebook*, *WhatsApp*, correos electrónicos de tipo *Gmail*, conversaciones telefónicas, las búsquedas que hace en *Google*, etc.

Además últimamente están apareciendo dispositivos que incorporan más datos nuestros a la red: cámaras de vigilancia caseras y otros sistemas de seguridad, cámaras públicas en centros comerciales y en espacios públicos, etc.

También otros dispositivos que inspeccionan nuestras actividades y pasan nuestros datos a la nube: datos de geolocalización tomados en el móvil, pulseras *fitbit* que durante el día monitoriza la distancia recorrida, las calorías quemadas, etc. y sube los datos a la nube, los relojes inteligentes, sensores para personas mayores, etc.

Se están incorporando dispositivos a los automóviles para seguimiento de su localización, consumo de combustible, incidencias, etc.

Junto con los datos anteriores hay grandes volúmenes de datos gestionados por administraciones públicas o grandes empresas privadas donde aparecemos:

- Historiales médicos
- Transacciones bancarias
- Archivos de radio y televisión

- UAV (o *drones*) como el *Predator* recogen una ingente cantidad de videos para el reconocimiento de imágenes en escenarios hostiles.
- Imágenes y videos procedentes de los vehículos de exploración y reconocimiento terrestres, bien en el espectro visible o bien en el infrarrojo.
- Imágenes obtenidas de satélites de vigilancia.
- Cámaras de vigilancia en lugares públicos gestionados por las diferentes administraciones estatales y locales.
- Cámaras de vigilancia en los entornos de lugares privados como hospitales, colegios, y empresas.
- Videos publicados y compartidos en las diferentes redes sociales, tales como *YouTube, Facebook, Twitter*, blogs u otros lugares del ciberespacio.
- Datos de los sistemas *LIDAR* (o *LADAR*). Datos sobre entornos geográficos, bosques, etc.

En definitiva estamos conectados e identificados casi de forma permanente y automática, enviado y recibiendo datos las 24 horas del día. Unos datos que, en la mayor parte de las ocasiones, esconden nuestros gustos, intereses, opiniones, una información tremendamente valiosa a nivel empresarial e institucional

El tamaño de los datos

Los datos que están en Internet ya se miden en *Petabytes* (un *Petabyte* = Un millón de GB)

Google procesa todos los días alrededor de 24 *Petabytes*. Se espera que para 2020 el volumen de datos almacenado en Internet sea 44 veces mayor al actual

Filmar la vida de una persona (100 años) ocuparía 0,5 *Petabytes*.

Facebook tiene 60 mil millones de imágenes, lo que supone 1,5 *Petabytes* de almacenamiento.

Actualmente es posible almacenar todos esos datos de forma efectiva. Mientras que en el año 2000 la instalación de un sistema de almacenamiento de capacidad 10 *Terabytes* (1.000 Gigabytes) era un pequeño reto técnico y económico, en 2013 la instalación de un sistema de almacenamiento de 1 *Petabyte* (1000 TB), que ocupa menos de un rack estándar, *no conlleva ninguna dificultad*. La razón hay que buscarla por un lado en la evolución de la capacidad y precio de los discos magnéticos (actualmente alcanzan 3TB en formato SAS 3.5) y por otra parte en la mejora de los sistemas de conexión y ficheros distribuidos.

La explotación de los datos: Big Data

Actualmente todos esos datos pueden explotarse. Los procedimientos para explotar ese ingente volumen de datos se designan con el nombre *Big Data*. Es una línea de investigación de mucha actualidad en Informática.

Hoy en día el término *Big Bata* se ha utilizado para trasladar al público todo tipo de conceptos entre los que se incluyen grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar los datos, datos en tiempo real y mucho más. Independientemente de la etiqueta, las empresas comienzan a comprender y explorar cómo procesar y analizar de nuevas formas toda esa información.

Pero también los gobiernos están comprendiendo la importancia del análisis de esos datos.

Algunos ejemplos de aplicación de este análisis son los siguientes:

- Investigadores de la universidad de Birmingham presentaron un algoritmo con el que es posible predecir delitos gracias a la monitorización de teléfonos móviles
- Google Trends predice, por ejemplo, cómo se comportará el virus de la gripe en cualquier parte del mundo. Se basa en las tendencias de búsqueda, en este caso en las palabras relacionadas con la gripe.
- El equipo de *Obama* utiliza Big data para analizar las encuestas y las redes sociales para afinar el discurso presidencial. Según su equipo electoral, llegaron a tomar 70 decisiones diarias basadas en Big data.
- Búsquedas de fotos de una persona sin estar previamente etiquetada
- Deducción de la relación de unas personas con otras sobre la base de la ubicación de las etiquetas en las fotos publicadas en los medios sociales
- Descubrimiento de expertos para una determinada tarea en las redes sociales
- Análisis de la influencia social. Aplicación de marketing viral en la que utilizamos los mensajes entre los participantes interconectados en una red social con el fin de propagar la información a través de las diferentes partes de la red.
- Análisis de video en tiempo real. Búsqueda y recuperación rápida en librerías de video o audio
- Detección de intrusión física en grandes espacios o infraestructuras abiertas

Los intereses en juego

Cada vez más empresas están interesadas en el Big Data para poder tomar decisiones que marquen la diferencia en el producto o servicio ofrecido, para ajustarse a los gustos del cliente, para ofrecerle un tipo determinado de producto o servicio.

Por ejemplo, el *Ford Focus* eléctrico produce ingentes cantidades de datos mientras está siendo conducido y cuando está aparcado. Mientras se encuentra en movimiento el conductor recibe constantemente información actualizada acerca de la aceleración, la frenada, la carga de la batería y la ubicación del vehículo. Esto resulta útil para el conductor, pero esos mismos datos también llegan a los ingenieros de Ford, quienes reciben información acerca de los hábitos de conducción de los clientes, incluido cómo, cuándo y dónde cargan sus automóviles. Y mientras el vehículo se encuentra detenido continúa enviando datos acerca de la presión de los neumáticos, el estado de la batería, etc.

Los diferentes gobiernos y servicios de inteligencia recopilan y almacenan una creciente cantidad de datos relativos a delitos, fichas criminales, atentados terroristas, ataques de ciberseguridad, redes criminales, etc. Al mismo tiempo y dado el carácter global de una parte de las actividades criminales, en los últimos años se han realizado importantes esfuerzos en la estandarización e interconexión de mucha de esta información.

La existencia y disponibilidad de todos estos conjuntos de datos supone una oportunidad, para aquellos usuarios autorizados, como fuente de nuevos datos y evidencias derivadas del cruce de datos entre estas bases de datos y/o con otras externas no asociadas directamente con la actividad criminal (acceso a datos bancarios, de compañías aéreas, etc..).

Los gobiernos están, por lo tanto, muy interesados en la explotación del gran volumen de datos que aparecen en Internet.

Todo lo anterior está abriendo un negocio enorme alrededor de la obtención de los datos en Internet, su almacenamiento y su posterior explotación. De acuerdo con los datos de *Gartner*, en 2015 serán necesarios 4,4 millones de personas formadas en esta disciplina. Por cada puesto de trabajo creado se generarán tres puestos adicionales. Se crearán 15 millones de trabajos nuevos relacionados con *Big Data* en tres años, a nivel mundial.

Big Data tuvo en 2013 con un volumen de negocio asociado de 34.000 millones de dólares, 6.000 millones más que en 2012.

Algunos hechos publicados

Los intereses que se mueven sobre los datos en Internet (Gobiernos y Empresas) están produciendo una invasión en la privacidad de las personas. En los periódicos van apareciendo muchos de ellos:

- El programa PRISM norteamericano, que durante años ha estado espiando correos electrónicos de miles de usuarios con el supuesto objetivo de combatir el terrorismo
- Huella dactilar Electrónica. La NSA (National Security Agency) dispone de un sofisticado programa de vigilancia con el que controla todas las comunicaciones electrónicas que Merkel realiza con sus asesores más directos.
- El espionaje británico pinchó las webcams de millones de usuarios de *Yahoo*
- Dentro de la NSA existía una unidad llamada ANT. Su objetivo era: encontrar la forma de *hackear* cualquier dispositivo, equipamiento de comunicaciones o firewall.
- La Sanidad Británica venderá datos de sus pacientes

Regulación del uso de la Información

Hay muchas lagunas legales en el uso de la información.

El uso ético y legal de la información tendrá que ser regulado por las autoridades pertinentes. El análisis de toda la información que generamos, ya sea en las redes sociales, en nuestros dispositivos móviles, o en nuestras cuentas de correo, puede ser útil para nosotros mismos, pero se debe garantizar el consentimiento y el uso lícito de estos datos.

Aunque las soluciones a muchos problemas económicos y sociales modernos se pueden encontrar en una mejor comprensión de datos, el aumento dramático en la cantidad y variedad de la recogida de datos plantea serias preocupaciones sobre las infracciones en materia de privacidad.

En las mejores universidades del mundo se están planteando como abordar muchos aspectos legales de estos problemas. Un ejemplo es uno de los últimos números de la [Stanford Law Review](#).

Otros problemas importantes en la regulación la de red

Hay otros problemas muy relevantes que están apareciendo y que afectarán al ciudadano: la *neutralidad de la red* y los problemas legales derivados de la negociación automática de los *acuerdo de nivel de servicio*.

Una **red neutral** es aquella que está libre de restricciones en las clases de equipamiento que pueden ser usadas y los modos de comunicación permitidos, que no restringe el contenido,

sitios y plataformas, y donde la comunicación no está irrazonablemente degradada por otras comunicaciones.

Los que proponen la neutralidad aseguran asimismo que las compañías de telecomunicaciones intentan imponer su modelo de servicio para conseguir beneficios aprovechándose del control del servicio, más que por demanda de sus servicios. Otros aseguran que creen que la neutralidad de la red es fundamental para preservar nuestras libertades. Así lo creemos nosotros. Este tema está siendo muy debatido en EEUU y empieza a ser debatido en Europa. Este concepto de neutralidad es ampliable a otros productos informáticos como buscadores, sistemas operativos, etc.

Los gobiernos deberían mantener la neutralidad de la red. En otros sectores, como el sector eléctrico español, la no neutralidad de la red eléctrica está ocasionando graves perjuicios al ciudadano y sobre todo impide el desarrollo de nuevas tecnologías como es el caso de las renovables.

Google está siendo investigado en Europa por el uso monopolístico de su buscador. Tras la investigación, y bajo la amenaza de sanción, accede a presentar los logos de sus rivales y sus enlaces en los resultados de las búsquedas para evitar la multa de Competencia.

Como hemos comentado los datos están cada vez más en la nube. Pero también las aplicaciones están pasando a la nube. Cada día aparecen nuevos servicios en la nube y la posibilidad de construir nuevos servicios a partir de los mismos. Aparecerán problemas relacionados con la responsabilidad ante el usuario de la calidad de los mismos y relacionado con eso cada vez serán más importantes los acuerdos de nivel de servicio.

Los **Acuerdos de Niveles de Servicio (SLA)** son un factor crítico de éxito para la contratación de servicios proporcionados en la nube por un tercero

El propósito primario de los SLA es especificar y clarificar las expectativas del funcionamiento, establecer la responsabilidad, y detallar las alternativas y consecuencias si el funcionamiento o la calidad del servicio no son los acordados por ambas partes.

Los SLA deben poder ser comprobados automáticamente y en el futuro negociados automáticamente. Todo ello necesitará de nuevas leyes y regulaciones.

Gobierno Abierto

Junto al gran volumen de datos existentes en Internet tenemos una gran carestía de datos públicos de relevancia sobre la acción de los gobiernos, las decisiones que toman, las razones sobre las mismas, etc.

Datos abiertos (open data en inglés) es una filosofía y práctica que persigue que determinados datos estén disponibles de forma libre a todo el mundo, sin restricciones de copyright, patentes u otros mecanismos de control.

Un **Gobierno Abierto** asume el compromiso de garantizar que la administración y operación de todos los servicios públicos que el Estado brinda puedan ser supervisadas por la comunidad, es decir, que estén abiertos al escrutinio de la ciudadanía.

Necesitamos Gobiernos Abiertos.

La apertura de datos públicos (**Open Data**): que implica la publicación de información del sector público en formatos que permitan su reutilización por terceros para la generación de nuevo valor