Control de nuestros datos personales en la era del *big data*: el caso del rastreo web de terceros

Control of Our Personal Data in the Big Data Era: The Case of Third Party Web Tracking

Controle de nossos dados pessoais na era do *big data*: o caso do rastreamento web de terceiros

Laura Daniela González Guerrero*

Fecha de recepción: 28 de junio de 2018. Fecha de aprobación: 20 de agosto de 2018

Doi: http://dx.doi.org/10.12804/revistas.urosario.edu.co/sociojuridicos/a.6941

Para citar este artículo: González Guerrero, L. D. (2019). Control de nuestros datos personales en la era del *big data*: el caso del rastreo web de terceros. *Estudios Socio-Jurídicos, 21*(1), 209-244. Doi: http://dx.doi.org/10.12804/revistas.urosario.edu.co/sociojuridicos/a.6941

RESUMEN

Cientos de corredores de datos instalan tecnologías de rastreo en la web para monitorear la navegación de las personas, lo que se conoce como rastreo web de terceros. Este artículo se divide en dos partes, la primera estudia el rastreo web en el contexto de la gran industria de datos que recoge información de millones de fuentes para perfilar a las personas según sus condiciones demográficas, sus prácticas de consumo, sus intereses políticos, entre otros. Se exponen los principales productos comerciales y riesgos de discriminación que surgen del perfilamiento de datos. En la segunda parte del artículo se revisan diferentes regulaciones que buscan dar control a las personas sobre los datos recolectados con tecnologías de rastreo. En particular, se evalúa la efectividad del modelo de consentimiento informado, para lo que se examinaron las políticas de privacidad de las diez páginas de noticias más visitadas desde Colombia. El estudio reveló un bajo cumplimiento de los estándares mínimos de información y de libertad necesarios para que las personas controlen el uso de los datos recolectados por medio de tecnologías de rastreo.

Palabras clave: rastreo web, políticas de privacidad, consentimiento informado, protección de datos, *big data*.

* Abogada y candidata a magíster en Políticas Públicas de la Universidad de Virginia, Estados Unidos. Investigación financiada por el programa Jóvenes Investigadores de Colciencias, convocatoria 761. Agradezco al Centro de Internet y Sociedad de la Universidad del Rosario, a Julio César Gaitán por su guía y a Jimena Hurtado y César Rodríguez por su asistencia en la investigación. orcid: https://orcid.org/0000-0002-4872-8450. Correo electrónico: laurad.gonzalez@urosario.edu.co

ABSTRACT

Hundreds of data brokers install technologies to track the browsing activities of people, this is called third-party web tracking. This paper is divided into two parts, the first part studies the web tracking in the context of the huge data industry which collects information from millions of sources to make profiles according to people's demographics, consumption practices, and their political interests, among others. We address the main commercial products and discriminatory risks arising from data profiling. In the second part, this paper reviews different regulations that aim at giving control to people over the data they produce through tracking technologies. In particular, the paper evaluates the effectiveness of the informed consent model. For this purpose, we examined the privacy policies of the ten most visited news web pages. This study revealed a low compliance of the minimum liberty and information standards needed to allow people to control the usage over the data collected through tracking technologies.

Keywords: Web tracking, privacy policies, informed consent, data protection, big data.

RESUMO

Centenas de corredores de dados instalam tecnologias de rastreamento na web para monitorar a navegação das pessoas, isto se conhece como rastreamento web de terceiros. Este artigo divide-se em duas partes, a primeira estuda o rastreamento web no contexto da grande indústria de dados que recolhe informação de milhares de fontes para perfilar as pessoas segundo suas condições demográficas, suas práticas de consumo, seus interesses políticos, entre outros. Expõem-se os principais produtos comerciais e riscos de discriminação que surgem do perfilhamento de dados. Na segunda parte do artigo revisam-se diferentes regulações que buscam dar controle as pessoas sobre os dados recoletados com tecnologias de rastreamento. Particularmente, avalia a efetividade do modelo de consentimento informado, para o qual se examinaram as políticas de privacidade das 10 páginas de notícias mais visitadas desde a Colômbia. O estudo revelou um baixo cumprimento dos standards mínimos de informação e de liberdade necessários para que as pessoas controlem o uso dos dados recoletados por meio de tecnologias de rastreamento.

Palavras-chave: rastreamento web, políticas de privacidade, consentimento informado, proteção de dados, *big data*.

Introducción

El *big data* es el agregado de avances tecnológicos que ha permitido recolectar, almacenar y analizar grandes cantidades de datos de todo tipo. Cada interacción con la tecnología deja un rastro susceptible de ser recolectado, almacenado, analizado y correlacionado con otros datos. Por ejemplo, las tecnologías de rastreo (TR) producen récords de las páginas web que se visitan, los clics en las páginas, las búsquedas en internet, las interacciones en redes sociales, las compras en línea y la ubicación de las personas. El internet de las cosas hace posible recolectar información, en tiempo real, sobre funciones corporales como el ritmo cardíaco, la temperatura corporal y los impulsos cerebrales; todo esto se almacena y se transmite por internet. En este contexto, las regulaciones sobre protección de datos tienen grandes desafíos para asegurar que la información detallada de los seres humanos, su entorno y sus actividades se use con los estándares más altos posibles.

Este artículo evalúa los retos del modelo de consentimiento informado en Colombia frente a una de las formas de recolección de información más expandida: el rastreo web. En la primera parte del texto se expone el contexto de la industria de datos y sus fuentes de información. Luego, se explica el rastreo web, su uso comercial y las tecnologías que se usan. En seguida, se expone la discusión sobre la anonimización y, finalmente, se explican los riesgos de algunos productos de la industria de datos. En la segunda parte, primero, se plantean las propuestas regulatorias que buscan dar control a las personas sobre los datos que se producen con tecnologías de rastreo. Luego, se describe el modelo colombiano de consentimiento informado y algunas de sus falencias. Finalmente, se estudian las prácticas en torno a las TR de los diez portales de noticias web más visitados por los colombianos. En esta última sección, se evalúa la calidad del consentimiento adoptado por los portales y se muestra el uso efectivo de tecnologías de rastreo.

Parte I

Industria de datos

1. Corredores de datos

Los corredores de datos, o *data brokers*, se refieren a empresas dedicadas a recolectar datos personales de múltiples fuentes para ofrecer productos que se derivan de la venta directa de datos o de su análisis. Este tipo de negocio no es nuevo; las empresas de *marketing* recolectan información demográfica y sobre los intereses de grupos poblacionales para ofrecer publicidad dirigida por medios impresos o telefónicos. Los burós financieros, como Datacrédito y Procrédito, en Colombia, recolectan información sobre transacciones financieras para generar calificaciones de crédito (Datacrédito Experian, s. f.). Lo que ha cambiado de forma radical es la capacidad tecnológica para recolectar, almacenar, cruzar y analizar volúmenes inmensos de datos de todo tipo, con diferentes formatos y en tiempo real.

Los corredores agregan información de múltiples fuentes. Según el estudio de la Comisión Federal del Comercio de Estados Unidos —FTC, por sus siglas en inglés— (2014), las fuentes pueden ser gubernamentales, del sector central y del sector descentralizado, de las agencias de censo, de empresas de seguridad social, tanto públicas como privadas, de los juzgados y los tribunales. También es posible recolectar datos relativos a las licencias profesionales y a los récords de conducción de vehículos. Otro origen es la recolección automática de datos con *webcrawlers* de la parte pública de sitios como LinkedIn, Facebook, Twitter, Bebo y páginas institucionales.

En este contexto, también se agregan datos de las actividades comerciales de las personas. Las tiendas venden e intercambian información sobre prácticas de consumo, como las ventas por catálogo, las suscripciones a revistas, ventas de automóviles, encuestas de *marketing*, tarjetas de clientes frecuentes, entre otras. Por ejemplo, los corredores de datos Oracle y Datalogix (2014) afirman "agregar y proveer información sobre

más de 2 billones de dólares en consumo de 110 millones de hogares, proporcionada por 1500 socios de datos"[1].

La práctica de intercambio comercial de datos es común en Colombia. La mayoría de políticas de privacidad incluyen una autorización para suministrar datos personales a terceras partes que los usuarios deben aceptar para acceder a los servicios. Por ejemplo, la política de datos personales del portal Eltiempo.com (2017) establece en el numeral 9 que la información anonimizada podrá "ser utilizada por las Entidades Autorizadas y/o terceros, quienes podrán disponer discrecionalmente de la totalidad de la información". La política de privacidad de Falabella (2016) incluye en el numeral 6 una autorización para transferir o transmitir los datos a sus filiales, comercios, entidades afiliadas y aliados estratégicos que operen o no en otra jurisdicción.

Finalmente, mediante las TR se recolecta información personal crucial para la industria de datos. Las TR, como las *cookies*, permiten recolectar información en tiempo real sobre cómo actúan las personas en internet, es decir, cuáles son las páginas que visitan, los contenidos que consumen, la hora, fecha y lugar de las actividades, la frecuencia de las visitas, qué compran en línea, entre otros. Esta información hace parte del gran agregado de datos que analizan y cruzan los corredores para crear perfiles de conducta, de consumo y de hábitos personales. Estos perfiles se comercializan con las industrias que necesitan conocer y calificar el comportamiento humano, como la industria publicitaria, las calificadoras crediticias y de riesgo y las empresas de selección de personal entre otras.

2. Rastreo en línea

La capacidad tecnológica para recordar lo que una persona hace en la web surgió en 1994 con el desarrollo de las HTTP *cookies*. Antes del desarrollo de las *cookies*, cada visita o clic en una página web no se podía recordar ni relacionar. Las *cookies* solucionan este problema, ya que son pequeños archivos de texto que guardan información sobre las acciones del usuario; los servidores leen esta información para recordar lo que se hizo en cada clic anterior. El primer navegador que implementó las

[1] Traducción del inglés de la autora.

cookies necesitaba recordar los productos seleccionados por el usuario en un portal de compras en línea (Angwin, 2010). Estas hicieron posible que los productos seleccionados en cada clic quedaran registrados al momento de pagar.

Las *cookies* se almacenan en el equipo terminal y guardan cualquier información que sea necesario recordar para una navegación fluida (Kristol, 2001, p. 5). Por ejemplo, almacenan la información de inicio de sesión de tal forma que, en un nuevo acceso, no es necesario proporcionar usuarios y contraseñas. Al comprar en internet y elegir un producto, esta información se almacena en una *cookie* a la que el servidor tiene acceso para recordar las selecciones. Las *cookies* también pueden registrar las secciones que una persona visita en un sitio web, lo que permite medir el contenido más popular, el menos visitado y cómo llegan, usualmente, las personas al portal. Esta información sirve para mejorar el rendimiento de los portales.

El uso más controversial de las *cookies* y de otras tecnologías similares es el rastreo. Dentro del texto que estas almacenan, por lo general se incluye una forma de identificación única o ID como este: *oeu15254886 48985r0.9022323644713817*. La *cookie* que contiene este identificador es instalada por una empresa de rastreo, como Oracle, en el *browser* del usuario. Una vez establecida, los servidores de Oracle pueden detectar si el ID visita otras páginas en las que Oracle tiene presencia. Mientras más presencia tenga Oracle en la web, está en mejor capacidad de seguir la actividad del ID único que instaló en el *browser* de la persona por medio de la *cookie*.

Por esta razón, al visitar portales en la web, se instalan dos tipos de *cookies*. Las primeras son de los servidores de las páginas que el usuario visita directamente. Las segundas son las que instalan servidores ajenos, llamadas *cookies* de terceros, como las usadas por Oracle. Por ejemplo, en el portal Eltiempo.com, se instalan las que esta empresa usa para medir la efectividad del sitio web o para recordar usuarios y contraseñas, y las *cookies* instaladas por terceros como Oracle, para rastrear. De esta forma, el usuario transmite información, no solo a los servidores de la página que visita directamente, como en el caso Eltiempo.com, sino también a otros servidores de empresas que el usuario no visita, como Oracle (Kristol, 2001, p. 31).

Otros corredores como Acxiom, DoubleClick y Rubicon están presentes en miles de páginas web instalando *cookies* y otras TR para rastrear la navegación de las personas en internet. El rastreo en línea por parte de terceros ha generado gran debate porque permite conocer sobre los intereses reales de una persona, sus prácticas de consumo, los lugares que visita, los que planea visitar en las vacaciones, el banco que utiliza, los periódicos que lee, las noticias que le interesan, las páginas de salud que visita y otro sinfín de información. Incluso, desde el proceso de estandarización de las *cookies*, el Grupo de Ingeniería de Internet —ITF— (1997) reconoció que la capacidad de recordar lo que las personas hacen en internet puede ser una intromisión en la privacidad, por lo que el usuario debe controlar los datos transmitidos por medio de *cookies*.

2.1 Uso comercial de las tecnologías de rastreo

Las *cookies* se volvieron fundamentales en el negocio de la publicidad y el *marketing online*. Antes, se pautaba para una audiencia amplia y se pagaba por estar en aquellos espacios en los que era más probable encontrar clientes (Sorjen, 2014). Sin embargo, las empresas empezaron a pagar más si alguien hacía clic en la publicidad. Esto incentivó el uso masivo de *cookies* y otras TR para identificar los intereses y los gustos que las personas revelan en su navegación web. Así, es posible mostrar anuncios más relevantes e incluso personalizados con mayor probabilidad de enganche (Angwin, 2010).

Los dividendos recibidos por la publicidad permiten no generar cobros al consumidor por el uso de contenidos y servicios en línea. Como se pueden cobrar precios más altos por la publicidad personalizada que por la publicidad generalizada, se incentiva recolectar mayor información de las personas para hacer perfiles comerciales lo más detallados posibles. Para 2010, el costo promedio de un anuncio personalizado o dirigido era de US$ 4,12 por cada mil espectadores, en comparación con US$ 1,98 por cada mil espectadores de un anuncio no personalizado (Angwin, 2010). Por esta razón, Scheneider (2015) afirma que "la vigilancia es el modelo de negocio de internet"[2] ("Free and convinient", párr. 1). La publicidad

[2] Traducción del inglés de la autora.

genera mayores ingresos mientras más efectiva sea y, para ello, se necesitan más detalles sobre los intereses de las personas.

El mercado de la publicidad en línea creció exponencialmente y, para el final de 2018, se espera que alcance los US$ 578 billones de dólares (Nanji, 2017). El modelo *Real Time Bidding* es uno de los mayores propulsores de la industria. Al visitar una página web que tiene espacios publicitarios, las *cookies* almacenadas y el conjunto de información que tienen los corredores de datos permite elaborar un perfil de consumo. Estos perfiles son subastados en tiempo real a empresas interesadas en publicitar a perfiles específicos. Se realiza la subasta en una plataforma llamada Ad Exchange en donde se encuentra la demanda y la oferta de anuncios, se escoge la mejor oferta y el anuncio ganador aparece en la página web que el usuario está visitando. Todo este proceso ocurre en décimas de segundo (Sorjen, 2014).

2.2 Estudios sobre rastreo web de terceros, crecimiento y tecnologías

El crecimiento exponencial de las TR en internet empezó a ser objeto de preocupación por los impactos en la privacidad y el desconocimiento de los usuarios. El primer trabajo sobre rastreo web estudió las páginas más populares de acuerdo con la clasificación de Alexa. Para 2005, los diez dominios de rastreo con mayor penetración estaban en el 40% de las páginas y en 2008 se habían expandido al 70% de las páginas estudiadas (Krishnamurthy & Wills, 2009). Estudios posteriores dan cuenta de su creciente penetración en la web (Leal-Taixé et al., 2017) y de su concentración en pocas compañías.

Englehardt & Narayanan (2016) estudiaron un millón de páginas clasificadas como las más visitadas según la clasificación de Alexa y encontraron que Google está presente en más del 70 % de las páginas analizadas con cinco dominios de rastreo: googleanalitycs.com, gstatic.com, doubleclick.net, google.com y fontsgoogleapis.com. En seguida está Facebook, con presencia en más del 20 % de las páginas, luego Twitter, Adnexus, Oracle y MediaMath.

La expansión del rastreo web no es la única preocupación reciente; las nuevas herramientas de rastreo resultan problemáticas porque son más difíciles de controlar y de eliminar por el usuario. Las TR se pueden

clasificar en dos tipos. Las primeras, se guardan en el equipo terminal como HTTP *cookies*, *E-tags*, y *Flash cookies*. El segundo tipo identifica las particularidades técnicas de los dispositivos para individualizarlos y se denomina "huella digital de dispositivo" o *fingerprinting* (Hoofnagle, Soltani, Good & Wambach, 2012).

Los *E-tags* se guardan en la memoria caché y agilizan la navegación al almacenar ciertos datos para no tener que cargarlos cada vez que se visita una página. Las empresas de rastreo utilizan los *E-tags* para guardar números únicos de identificación en la memoria caché. Con este método, se puede rastrear a un usuario incluso cuando todas las HTTP *cookies* están inhabilitadas y cuando se navega en modo incógnito. Además, si los usuarios bloquean los *E-tags*, la búsqueda en la web se hace mucho más lenta (Ayenson, Wambach, Soltani, Good & Hoofnagle, 2011). Las *Flash cookies* tienen dos características problemáticas: primero, permiten recuperar la información almacenada por las *cookies* estándar si estas se borran; y, segundo, se instalan fuera del navegador, por lo que permiten hacer rastreo incluso si se cambia de navegador (Hoofnagle, Soltani, Good & Wambach, 2012).

Las *E-tags* y las *Flash cookies* se usan para pasar por alto las elecciones de privacidad de los usuarios que eliminan o bloquean las *cookies* estándar, hecho que ha suscitado controversias judiciales. En 2012, Quantcast pagó 2,4 millones de dólares por usar *Flash cookies* para recuperar las *cookies* estándar eliminadas por los usuarios (Singel, 2012). En 2017, Google pagó 5,5 millones de dólares por instalar *cookies* en Safari e Internet Explorer, pasando por alto la opción de bloqueo de *cookies* predeterminada en estos buscadores (Privacy Litigation Google, 2017).

El *fingerprinting*, o huella de dispositivo, es más difícil de controlar, ya que no se instala nada en el equipo terminal. Consiste en recolectar características de los dispositivos tales como los detalles de los dibujos que produce, la forma de reproducción del sonido, información sobre el funcionamiento de la batería, entre otros. Estos datos permiten aislar un equipo e identificar la navegación que realiza en internet. Para detener el rastreo con *fingerprinting*, es necesario desinstalar partes funcionales de las páginas web como el Java Script y el Adobe Flash (Hoofnagle, Soltani, Good & Wambach, 2012).

Otro aspecto tecnológico importante es la capacidad de compartir y sincronizar los ID o cualquier identificador de dispositivo entre empresas de rastreo para unificar la información. Se ha identificado que 45 de las 50 empresas más grandes de rastreo comparten sus ID únicos con otras empresas de rastreo (Englehardt & Narayanan, 2016). También es posible seguir a los usuarios en diferentes dispositivos, lo que se conoce como *cross-device matching*. Esto se hace con tecnologías de *machine learning* para "determinar qué dispositivos y huellas digitales pertenecen a la misma persona"[3] (Christl, 2017, p. 69).

En Colombia, son pocos los estudios sobre prácticas de rastreo, lo que dificulta la medición del fenómeno y de sus impactos. Este año, la Fundación Karisma (2018) realizó una primera medición en las páginas web de candidatos a la Presidencia de Colombia y encontró *cookies* de empresas de rastreo relacionadas con la campaña presidencial de Trump. Aunque el estudio no profundiza las posibles consecuencias, es clara la necesidad de estudiar más a fondo el rastreo y sus impactos en la personalización de la propaganda política.

3. El rastreo en línea y la acumulación de datos no es anónimo

El parte de tranquilidad más común en las políticas de protección de datos es que la información recolectada es anónima o pseudoanónima, es decir que está ligada a un identificador numérico. Si no es posible enlazar una base de datos a personas concretas, en teoría son solo datos numéricos, no datos personales. Sin embargo, existen múltiples mecanismos que permiten asociar información anónima o pseudoanónima con personas concretas. Narayanan (2011) describe algunos mecanismos directos de identificación: (i) cuando la tercera parte que instala mecanismos de rastreo es también un proveedor de servicios. Por ejemplo, Google y Facebook instalan mecanismos de rastreo en la web y, a su vez, las personas tienen cuentas con estas empresas y dan información como nombres reales, teléfonos, direcciones, fotos, localizaciones, etcétera. No existen barreras legales ni tecnológicas que impidan enlazar la información recolectada mediante rastreo web con la información que

[3] Traducción del inglés de la autora.

tienen de sus usuarios; y (ii) las empresas que tienen relación directa con las personas filtran información personal a terceras partes. Por ejemplo, cuando se revela el nombre o el correo electrónico del usuario en la URL.

Mayer J. (octubre de 2011) ejemplifica otros casos en los que se filtra la información del usuario a terceras partes. Por ejemplo, al ver una propaganda en la página de Home Depot, se enviaba el primer nombre del usuario y el correo electrónico a trece compañías diferentes. Al cambiar la configuración de video en el sitio Metacafe, se enviaba el primer y segundo nombre del usuario, su fecha de nacimiento el correo electrónico y el teléfono a dos compañías diferentes.

No es necesario que se filtren o intercambien datos personales, como en los ejemplos anteriores, ya que es posible identificar personas al correlacionar datos anónimos. Con solo cuatro puntos espaciotemporales aleatorios se pudo identificar, de forma única, el 95 % de los individuos en un estudio sobre la movilidad de un millón y medio de personas. (Montjoye, Hidalgo, Verleysen & Blondel, 2013). Narayanan & Shmatikov (2007) identificaron personas al comparar clasificaciones y marcas de tiempo anónimas de Netflix con clasificaciones y marcas de tiempo de *rankings* públicos en internet. Los investigadores explican que, incluso si se eliminan los nombres y números de seguridad social, se puede utilizar "el conocimiento contextual y de fondo, así como la correlación cruzada con las bases de datos disponibles públicamente, para volver a identificar los registros de datos individuales"[4] (Narayanan & Shmatikov, 2007, p. 1).

El documento CONPES 3920 sobre la Politica Nacional de Explotación de Datos (2018) determinó que es necesario expedir estándares de anonimización y minimización de datos de los que carece el país y se espera tener los estándares requeridos en 2019.

4. Productos que ofrece la industria de datos

Con la información recolectada en medios *offline*, *online* y por el rastreo web, los corredores de datos ofrecen múltiples productos.

[4] Traducción del inglés de la autora.

4.1 *Marketing*

Dentro de los servicios de *marketing*, el estudio de la FTC (2014, pp. 24-25) documentó varios productos como el anexo de datos, en donde las compañías compran información sobre sus clientes para saber más sobre sus preferencias y hacer publicidad dirigida. Con algún dato inicial que la compañía provea sobre su cliente, el corredor de datos ofrece información sobre la edad, la afiliación religiosa, intereses en tecnología, género, afiliación política, hábitos vacacionales, ocupación, peso, entre otra extensa lista de detalles. Otro producto son las listas de *marketing*: una compañía identifica las características de las personas a las que quiere llegar y el corredor de datos le ofrece una lista de personas que cumplen dichas características. Para esto, los corredores no solo usan datos fácticos, sino también inferidos. Por ejemplo, si una persona tiene una licencia para manejar bote, es fácil inferir que tiene un "interés en la navegación náutica". Los corredores de datos también ofrecen el servicio de "incorporación", esto es mostrar publicidad en línea con base en las actividades y compras que los usuarios hacen por fuera de internet.

4.2 Servicios de calificación

Los corredores de datos ofrecen servicios de calificación financiera. Lenddo es una empresa filipina dedicada a generar calificaciones crediticias de datos no convencionales. La compañía agrega información de redes sociales, de navegación en línea, la geolocalización del celular y otros datos almacenados por los dispositivos (Lenddo, 2016). La empresa afirma usar más de 12 000 variables que se procesan por medio de *machine learning* para entender las conductas del consumidor y la fortaleza de sus relaciones para generar un puntaje crediticio. Existen empresas como BomPra Credit, en Brasil, que utilizan el reconocimiento facial por medio de *selfies* para disminuir riesgos de fraude de identidad. Además, agregan los datos del dispositivo desde el que se toma la foto, como las ubicaciones y compras recientes de las personas para medir la capacidad de consumo (Dias & Natusch, 2016).

La principal preocupación de este tipo de producto es la falta de transparencia al momento de establecer un puntaje crediticio. Con miles

de datos y variables de diferentes tipos y fuentes, es difícil entender los pesos que se asignaron a cada variable. Además, muchos de los algoritmos que procesan la información están bajo secreto comercial, lo que dificulta la evaluación de los mecanismos de puntaje y conocer los criterios utilizados y el peso de cada uno. No se puede verificar que la calificación no incluya datos sensibles como etnia, género, religión e influencias políticas; en general, que la calificación no esté basada en prejuicios que empeoren el acceso al crédito y otros servicios. De continuar esta tendencia, es crucial que se adopten medidas para que la gente pueda elegir no estar sometida a este tipo de calificaciones con datos no convencionales y que exista completa transparencia y entendimiento de los procesos utilizados.

4.3 Personalización y manipulación

Grandes compañías tienen la capacidad de ofrecer experiencias altamente personalizadas por la cantidad de información que poseen sobre las personas. Es el caso del *news feed*, de Facebook, o de los resultados de búsqueda de Google y las recomendaciones de Netflix. Los corredores de datos y otros actores también ofrecen experiencias personalizadas de internet. De acuerdo con Christl (2017), los corredores de datos ofrecen desplegar una página a la medida de usuarios, aparentemente anónimos, al elegir qué contenido mostrar con relevancia, dados los intereses del usuario, y mostrar precios diferenciados de productos y servicios de acuerdo con su disposición a pagar (Christl, 2017, pp. 75-76).

La industria de datos conoce cada vez más las inclinaciones psicológicas de las personas. Algunos son más perceptivos a las recomendaciones de amigos, otros a recomendaciones de expertos. Unas personas responden a los incentivos positivos, otras a los negativos. El análisis masivo de datos permite develar estas características para personalizar los mensajes, tener efectos directos y generar comportamientos (Helberger, 2016). La personalización en campañas electorales, por ejemplo, pretende persuadir con mensajes adecuados para determinados segmentos. Así, "gente que gana más de $ 100.000 al año, dueños de armas, personas que han leído noticias sobre una posición de determinados temas, veteranos

desempleados… y cualquier cosa que se pueda imaginar"[5] (Scheneider, 2015). Esta práctica se generalizó para eventos políticos como el Brexit, las elecciones presidenciales de Estados Unidos y el plebiscito en Colombia.

Existe una preocupación por la transición de la personalización a la manipulación. Sosaina Zuboff (2015) describe los nuevos mercados de *behavioral control* cuya mercancía es predecir y modificar el comportamiento humano. Por esto, Schneider (2015) afirma que "la manipulación psicológica, basada en la información personal y en el control de los sistemas, solo mejorará. Peor aún, será tan buena, que no nos daremos cuenta de su presencia"[6].

4.4 Tratamiento de datos sensibles, perfilamiento y discriminación

Los datos sensibles no se obtienen directamente de los titulares; se predicen a partir de sus comportamientos en línea. En 2011, el laboratorio de seguridad digital de Stanford descubrió que la red de publicidad Epic Marketplace clasificaba a los usuarios de internet por su historial de búsqueda en segmentos como "menopausia", "buscando un embarazo" y "mejorando historial crediticio" (Mayer J., julio de 2011). El laboratorio también encontró que la aplicación de citas OkCupid vendía información sobre sus suscriptores a los corredores de datos BlueKai y Lotame. La información incluía la frecuencia del consumo de drogas y alcohol (Mayer J., octubre de 2011).

También se han identificado corredores de datos que usan segmentos o perfiles basados en características étnicas, nivel de ingresos o nivel de educación. Por ejemplo, un corredor usaba las categorías *urban scrambler* y *mobile mixers* que incluían, en su mayoría, personas latinas y afroamericanas de bajos recursos (Federal Trade Commission, 2014). Los perfiles publicitarios de Facebook también usan categorías sensibles como afiliaciones políticas, religiosas e intereses sociales.

Otro ejemplo de discriminación con tecnologías de rastreo es el sitio web que Wells Fargo creo en 2000 para promocionar sus préstamos de vivienda. El sitio tenía una herramienta para encontrar habitación llamada

[5] Traducción del inglés de la autora.

[6] Traducción del inglés de la autora.

community calculator. Esta identificaba el código postal del visitante y la población predominante en ese sector; posteriormente, lo dirigía a barrios que tuvieran la misma población. De esta forma, la herramienta dirigía a personas blancas a barrios de blancos y a personas negras a barrios de negros (Scheneider, 2015).

Los efectos negativos de la agregación de datos para crear perfiles basados en categorías sensibles han generado diversas medidas regulatorias. El nuevo Reglamento de Protección de Datos Personales de la Unión Europea (GDPR, por sus siglas en inglés) (2016) y los Estándares Iberoamericanos de Protección de datos Personales (2017) incluyen:

- Informar si los datos son usados para perfilamiento y sus posibles consecuencias.
- Incluir el derecho a oponerse al perfilamiento para el *marketing* directo y brindar un mecanismo efectivo para oponerse.
- El titular tiene el derecho a no ser objeto de decisiones que se basen únicamente en un tratamiento automatizado de datos que evalúe aspectos personales y produzca efectos jurídicos o que le afecten significativamente de modo similar, como la denegación automática de un crédito.
- Esto aplica en particular al tratamiento que analiza o predice aspectos relacionados con el rendimiento en el trabajo, la situación económica, la salud, las preferencias o intereses personales o la fiabilidad de las personas, en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar.
- Todo tratamiento debe estar sujeto a garantías apropiadas que incluyan información específica, el derecho a obtener intervención humana, a expresar su punto de vista, a recibir una explicación de la decisión tomada y a impugnar la decisión.
- Deben impedirse efectos discriminatorios por motivos de raza u origen étnico, opiniones políticas, religión o creencias, afiliación sindical, condiciones genéticas, estado de salud u orientación sexual.

En Colombia, la Corte Constitucional ha reconocido explícitamente los peligros de la acumulación desproporcionada de información personal por los perfiles y categorías que se pueden crear para tomar decisiones

sobre las personas (C-748, 2011). Sin embargo, esta preocupación no quedó consignada en la ley ni en la regulación de protección de datos. El documento CONPES sobre la Politica de Explotación de Datos (2018) tocó el tema en dos pies de página, pero no abordó medidas específicas para prevenir las consecuencias discriminatorias del perfilamiento y la calificación.

Parte II

Regulación

5. Regulación de tecnologías de rastreo web

La falta de control y de información sobre la recolección de datos personales por medio de tecnologías de rastreo, sus usos e implicaciones, ha suscitado discusiones sobre qué tipo de soluciones desarrollar. Un mecanismo que restrinja por completo las TR afecta el interés comercial de las empresas que acumulan millones de perfiles y ofrecen variedad de servicios y productos. Por otra parte, un mecanismo con nulo poder de elección para las personas afecta gravemente la privacidad y la libertad en internet. Esta sección expone el nivel de información y capacidad de elección efectiva que brindan distintos modelos de protección de datos.

Los modelos pueden clasificarse en dos: 1) soluciones descentralizadas: cada página o sitio web debe suministrar alguna forma de administrar los datos recolectados por medio de TR. 2) soluciones centralizadas: concentran la administración de los datos recolectados con TR en los buscadores de internet o extensiones.

5.1. Las soluciones descentralizadas

Este modelo exige a cada portal informar y obtener una autorización para el tratamiento de los datos que se recogen por medio de TR. Las regulaciones europeas tienen un largo historial de discusión sobre cómo hacer efectivo este modelo, por lo que se explican algunas de sus disposiciones.

5.1.1 Regulación europea de 2002

La directiva sobre la privacidad y las comunicaciones electrónicas (e-privacy, 2002) consagró la necesidad de obtener el consentimiento informado para recolectar datos por medio de TR. Este debe ser una manifestación de voluntad inequívoca, libre, específica e informada. Sin embargo, el modelo no logró dar control a las personas sobre sus datos porque no se establecieron parámetros o guías específicas. De esta manera, cada país adoptó distintos niveles de información y de control sobre los datos personales recolectados con TR. Por esta razón, el grupo de trabajo Artículo 29 (2013) emitió una guía básica sobre las condiciones para obtener un consentimiento válido al recolectar datos personales con TR:

- Información específica, condición que incluye:

  (i) Un aviso visible, comprensible y claro que informe sobre el uso de TR.
  (ii) Mostrar todos los tipos de TR y *cookies* que hay en el sitio y sus finalidades.
  (iii) Qué tipo de datos se recopilarán y utilizarán.
  (iv) Presencia de *cookies* de terceras partes.
  (v) Quién tiene acceso a la información recolectada en la página por medio de TR.
  (vi) Formas de negar el consentimiento para el uso de TR.

- El momento: el consentimiento debe darse antes del procesamiento de los datos, es decir, antes de instalar o leer las *cookies*.
- Comportamiento inequívoco: cada sitio web es libre de establecer la conducta que signifique consentimiento, es decir hacer clic en un botón o marcar una casilla. Sin embargo, solo un clic en un enlace "más información sobre *cookies*" no puede considerarse consentimiento. Si este se entiende como positivo con la simple entrada de la persona a la plataforma o al sitio web, no se puede considerar un comportamiento inequívoco.

– Libre: deben darle a la persona condiciones reales para hacer una elección. Por tanto, se recomienda no usar formatos en que solo exista la posibilidad de aceptar *cookies* sin una opción para negar su uso.

5.1.2 Nuevo Reglamento General de Protección de Datos

El nuevo Reglamento General de Protección de Datos (GDPR, 2016) tiene dos impactos en la actividad de rastreo web. Primero, reconoce que los identificadores creados con TR son datos personales y deben ser protegidos. Segundo, establece requisitos más estrictos para adquirir un consentimiento válido. El grupo de trabajo Artículo 29 (2018) elaboró una nueva guía para identificar cuándo el consentimiento se da de forma libre, es informado y es una manifestación inequívoca de voluntad:

– Libre: implica darle real control de elección al titular de los datos. El consentimiento no se considera libre si la persona no puede rehusarse o retirarlo en cualquier momento sin perjuicio alguno. También se debe evaluar si la ejecución de un contrato o la prestación de un servicio se supeditan al consentimiento del tratamiento de datos personales no necesarios para la ejecución de dicho contrato. Por ejemplo, una aplicación de edición de fotos pide recolectar datos del GPS. La localización no es una condición necesaria para el funcionamiento de la edición de fotos, por tanto, si el uso de la aplicación está condicionada a tener prendido el GPS, el consentimiento no se puede entender como libre.
  La granularidad es otra condición fundamental. Si se recolecta información para distintos propósitos, la persona debe poder aceptarlos y rechazarlos selectivamente. Si no se busca un consentimiento separado para cada propósito, existe una falta de libertad.
– Informado: a los requisitos mínimos que se establecieron en 2002 se adicionan:

  (i) Información sobre el uso de los datos para la toma automatizada de decisiones.
  (ii) Informar sobre el derecho de oposición al *marketing* directo.

(iii) Informar sobre posibles riesgos por transferencias de datos a países que no cuentan con salvaguardias apropiadas.

– Inequívoco: se establecieron lineamientos útiles para evaluar que el consentimiento sea inequívoco:

(i) El consentimiento debe ser un acto afirmativo y claro.
(ii) El uso de casillas premarcadas no es válido.
(iii) No se permiten mecanismos *opt-out* en los que el tratamiento ocurre sin previa autorización y requieren una intervención del interesado para evitar un acuerdo.
(iv) La acción de consentir debe ser distinguible de otras acciones. El mero uso del servicio, mover el cursor o dar clic en cualquier parte del sitio web no puede significar consentimiento.

– Consentimiento especial: se estableció un estándar más alto para obtener el consentimiento cuando el tratamiento de datos acarrea un riesgo considerable para la persona:

(i) Cuando se trata de datos sensibles que revelen el origen étnico o racial, las convicciones religiosas o filosóficas, las opiniones políticas, la afiliación sindical o datos sobre la vida sexual de una persona, entre otros.
(ii) Cuando existe transferencia de datos a terceros países que no cuentan con niveles adecuados de protección de datos.
(iii) Cuando los datos se usan en toma de decisiones automatizadas y en perfilamiento.

Para un estándar más alto, el grupo de trabajo Artículo 29 (2018, p. 18) propone el doble consentimiento. Por ejemplo, una persona recibe una notificación pidiendo el consentimiento para tratar datos médicos. El controlador de los datos indica que para dar el consentimiento es necesario enviar un correo con la frase "acepto". En seguida, la persona recibe un enlace, que debe ser abierto, o un código de confirmación.

– Retiro del consentimiento: retirar el consentimiento debe ser tan fácil como darlo y debe poder retirarse en cualquier momento.

5.2 Soluciones centralizadas

Las soluciones centralizadas organizan en un solo lugar las autorizaciones para el tratamiento de datos personales recolectados por TR.

5.2.1 Propuesta de reforma en Europa

En 2016 se hizo una evaluación de la directiva *e-privacy*, de 2002, que arrojó bajas eficiencia y eficacia del modelo descentralizado, principalmente por la excesiva carga que se impone a empresas y a consumidores, y porque ofrece un bajo control a los usuarios sobre sus datos. Además, no abarca con claridad otras tecnologías de seguimiento como la huella digital de dispositivo (Propuesta de reforma de la directiva *e-privacy*, 2002; 2017).

Por estas razones, se propuso, en 2017, una reforma para que se administren las TR desde los navegadores o *software* similares. El usuario no tendría que aceptar el uso de *cookies* de cada página que visita y bastaría con que eligiera una configuración de privacidad en el navegador, que aplique a toda su navegación en internet, y que sea vinculante para todos los sitios web. La reforma también plantea que se deben ofrecer distintos niveles de protección, desde el bloqueo general de las TR, pasando por la elección de lugares autorizados, hasta permitir TR en todos los sitios. Además, debe ser posible cambiar la configuración de forma sencilla durante la navegación en internet (Propuesta de reforma de la directiva *e-privacy*, 2002; 2017, consideraciones 22, 23 y 24).

Uno de los aspectos más debatidos es la adopción del principio de privacidad por diseño y hacer que los navegadores configuren, de forma predeterminada, el bloqueo de las TR de terceras partes. Esta idea ha tenido un vehemente rechazo por parte de la industria publicitaria (Kristol, 2001) puesto que sería mucho más difícil recolectar información de navegación y crear perfiles detallados.

5.2.2 Iniciativa *Do Not Track* en Estados Unidos

La FTC impulsó la iniciativa *Do Not Track* para implementar una herramienta sencilla y uniforme que diera control a los usuarios sobre las herramientas de rastreo web (Federal Trade Comission, 2010). El proceso fue saboteado por grandes jugadores de la publicidad en internet, como Facebook y Google, las propuestas recibieron amplio rechazo y poco a poco los actores fueron abandonando el proyecto (Chmielewski, 2016). En 2015, la idea se retomó como un proyecto de ley de los senadores Ed Markey y Richard Blumenthal, llamado *Do Not Track Online Act* (2015). Dicho proyecto instaba a la FTC a adoptar un mecanismo uniforme, efectivo y fácil de usar, que permitiera administrar el rastreo web en servicios de internet móvil y fijo. Sin embargo, el proyecto de ley tampoco fue aprobado.

5.2.3 Autorregulación en Estados Unidos

Ante la creciente presión por regular el uso de TR, algunas organizaciones de la industria publicitaria, como The Digital Advertising Alliance (DAA), desarrollaron plataformas centralizadas para que el usuario pudiera controlar la publicidad comportamental basada en el rastreo web (YourAdChoices, s. f.). Una de las principales críticas a estas plataformas es que solo se da la opción de rechazar la publicidad, pero no se puede decidir sobre la recolección de datos en sí misma, ni sobre otros usos (Toner, 2017).

El sistema funciona por medio de *cookies* que señalan la elección de la persona a las compañías de rastreo. Estas son frágiles: si la persona las borra del navegador, debe volver a hacer el procedimiento en la herramienta; además, son fáciles de eliminar por terceras partes (Mayer & Mitchell, 2012). Adicionalmente, la participación de las empresas no es generalizada, ya que las redes sociales y las pequeñas empresas de publicidad no están integradas. El sistema no es obligatorio y carece de un cuerpo sancionatorio efectivo que castigue a las empresas que pasen por alto la elección del usuario.

Finalmente, estas plataformas tienen serios inconvenientes de usabilidad e información. Al utilizar los mecanismos, las personas entienden que optan por salir de la publicidad dirigida, pero no entienden que el

rastreo continúa (McDonald & Cranor, 2010). En otro estudio, de cinco personas que usaron la herramienta del DAA, dos no pudieron entrar a la herramienta, dos solo pudieron salirse del rastreador de Yahoo y solo una persona pudo utilizar la herramienta sin guía (Leon et al., 2011).

5.1.4 Extensiones de navegador y privacidad en navegadores

Las extensiones de navegador como Ghostery, Disconnect y Privacy Badget permiten administrar las TR desde plataformas centralizadas. Aunque prácticas, la efectividad de estas herramientas requiere experticia. No basta con instalarlas en el navegador; es necesario explorarlas y entender la forma de configurarlas y usarlas. Por esto, autores como Mayer & Mitchell (2012) y Narayanan (2018) sostienen que son herramientas para "usuarios avanzados".

En el mercado de navegadores, uno de los criterios con los que se compite es la privacidad. Navegadores como Brave, Firefox y Safari han adoptado diferentes mecanismos de *machine learning* o de listas negras para bloquear el rastreo web de terceros (Toner, 2017). En estos navegadores, el rastreo de terceros está bloqueado de forma predeterminada y si el usuario quiere permitirlo, debe cambiar la configuración. El navegador Cliqz, lanzado en 2015, muestra los rastreadores en cada página web y los bloquea de forma predeterminada. Estas alternativas son efectivas, pero carecen de un público informado que adopte como criterio de elección la protección de la privacidad.

Google Chrome, que tiene el 60 % del mercado (Statistic, s. f.), no ha adoptado medidas de privacidad por diseño en el navegador. La configuración predeterminada es "permitir que todos los sitios guarden y lean datos de *cookies*" y la opción para "bloquear *cookies* de terceros" está apagada. El usuario es rastreado de forma predeterminada y debe cambiar la configuración si desea detenerlo.

6. Regulación de tecnologías de rastreo en Colombia: modelo descentralizado

En Colombia, el tema de las TR ha tenido escaso debate regulatorio y social. La Superintendencia de Industria y Comercio (2016) solo ha especificado que la información obtenida por medio de *cookies* son datos

personales sujetos a la Ley 1581 de 2012. En consecuencia, se entiende que el titular debe otorgar una autorización previa, expresa, informada y libre para su tratamiento. Sin embargo, no se han abordado las particularidades técnicas que acarrea la recolección de datos personales por medio de TR. De la Ley 1581 de 2012 y del Decreto Regulatorio 1377 de 2013, se extraen las siguientes condiciones:

- Previa: la autorización debe suministrarse antes de la incorporación del dato. Es decir, antes de la instalación del mecanismo de rastreo en los equipos terminales.
- Expresa: en primer lugar, de acuerdo con la Corte Constitucional (C-748, 2011) en el ordenamiento colombiano no existe la figura de consentimiento tácito o implícito, por lo que la autorización debe ser directa. Sin embargo, el Decreto 1377 contempló la posibilidad de obtener el consentimiento mediante conductas inequívocas "que permitan concluir de forma razonable que se otorgó la autorización" (Artículo 7). El problema es que en Colombia no existe una guía para determinar cuándo una conducta se considera inequívoca. Como se describió en apartados anteriores, el grupo de trabajo Artículo 29 (2018) desarrolló parámetros para solucionar este vacío.

En segundo lugar, en Colombia no es obligatoria la granularidad del consentimiento; basta con que las personas consientan en un documento general. Solo hasta 2017 la Superintendencia de Industria y Comercio publicó un formato que contempla autorizaciones separadas para cada finalidad de tratamiento[7], pero no es obligatorio adoptar este formato. Por tanto, la mayoría de portales usan una autorización general y no existe poder de decisión sobre tratamientos específicos.

- Informada: el responsable del tratamiento debe informar como mínimo:
  (i) Los datos personales que serán recolectados.

[7] EL formato está disponible en la página de la Superintendencia de Industria y Comercio en el siguiente enlace: http://www.sic.gov.co/sites/default/files/files/Nuestra_Entidad/Publicaciones/Cartilla_formatos_datos_Personales_nov22.pdf

(ii) Todas las finalidades específicas del tratamiento.
(iii) El tratamiento al que serán sometidos los datos.
(iv) El carácter facultativo de la respuesta a preguntas sobre datos sensibles o sobre niñas, niños o adolescentes.
(v) Los derechos del titular.
(vi) La identificación y datos del contacto del responsable del tratamiento de datos.

Aunque en Colombia no existen especificaciones sobre la recolección de datos con TR, las personas deben ser informadas al respecto, así que es necesario contar con la siguiente información:

(i) Un aviso que informe sobre el uso de tecnologías de rastreo y qué tipos se utilizan.
(ii) Qué tipo de información se recolecta por medio TR y su finalidad.
(iii) Presencia de rastreo web de terceros.
(iv) Quién tiene el control de la información recolectada con TR
(v) Formas de negar el consentimiento para la recolección de datos con TR.

– Libre: ni la ley ni el reglamento en Colombia establecieron la obligación de hacer que dar y retirar el consentimiento fuera igual de fácil. Sin esta paridad, en la práctica solo se facilita dar el consentiminto, pero se dificulta retirarlo en cualquier momento.

7. Crítica a las políticas de protección de datos personales

Los modelos descritos tienen como epicentro el consentimiento informado, los datos solo pueden tratarse bajo las condiciones y para los usos que se informaron al titular. Además, si el titular no está de acuerdo con el tratamiento, puede abstenerse libremente de aceptar las condiciones y no contratar con el prestador del servicio. Sin embargo, estas premisas de información y libertad de elección no se cumplen.

En primer lugar, la presunción básica sobre la adecuada información es difícil de materializar. Varios estudios empíricos demuestran que las personas no visitan los términos y condiciones de los portales web (Bakos,

Wurgler & Trossen, 2014) y, si lo hacen, no le dan el tiempo requerido a leer y entender. Obar & Oeldorf-Hirsch (2016), por ejemplo, utilizaron una red social ficticia y revelaron que de 543 participantes el 74 % pasó por alto la política de privacidad. De aquellos que la leyeron, el 80 % gastó menos de un minuto leyendo el documento y el 14 % menos de cinco minutos. Las guías de la OCDE sobre privacidad (2013) reconocieron que es cada vez más difícil para los individuos entender y hacer elecciones con respecto a los usos de sus datos personales ya que estos se hacen más complejos, son poco transparentes para los individuos y están en continuo cambio.

La segunda presunción equivocada del modelo es el poder de negociación del usuario. La vasta mayoría de políticas de privacidad adopta un formato en el que se acepta por completo o se declina del servicio, sin que el usuario pueda entrar a hacer negociaciones. Por esto, Belli, Schwartz & Louzada (2017) han definido el modelo de consentimiento informado, no como un sistema de protección de datos, sino como un sistema de pago en el que se intercambia el consentimiento de explotación de los datos a cambio de la posibilidad de utilizar servicios.

8. Evaluación de páginas web en Colombia

En esta sección se presentan las prácticas concretas de uso de TR de los diez portales de noticias web más visitados por los colombianos de acuerdo con la clasificación de Alexa. Se presentan los resultados de dos estudios. En el primero, se evalúa la forma de adquirir el consentimiento para el tratamiento de datos recolectados con TR. En el segundo, se midió la presencia efectiva de tecnologías de rastreo por medio de la herramienta Ghostery.

– Aspectos metodológicos: de acuerdo con la clasificación Alexa[8], las diez páginas de noticias más visitadas desde Colombia son: Eltiempo.com, Elespectador.com, Semana.com, Canalcaracoltv.com, Canalrcn.

[8] Las clasificaciones de Alexa son variables; se utilizaron los resultados del mes de abril. Top Sites by Category-Alexa. (2018), noticias y medios. Recuperado de: https://www.alexa.com/topsites/category/Top/World/Espa%C3%B1ol/Regional/Am%C3%A9rica/Colombia/Noticias_y_medios

com, Elcolombiano.com, Elheraldo.com, Eluniversal.com.co, Caracol.com.co y Dinero.com.

Semana.com y Dinero.com tienen las mismas políticas de tratamiento de datos personales, por tanto solo se analiza la primera. Para completar diez portales, se incluye la página Minuto30.com por ser de noticias y estar en el décimo puesto de las páginas más visitadas desde Colombia[9].

8.1 Resultados

8.1.1 Evaluación del consentimiento para tratar datos personales por medio de TR

Se revisaron las políticas de privacidad, las políticas de *cookies*, cuando existían, y los términos y condiciones de uso de cada página. Se identificó la forma de adquirir el consentimiento para el tratamiento de datos personales con TR y se evaluó, a la luz de la regulación colombiana y de los parámetros de la nueva regulación europea, por compartir el mismo modelo descentralizado de consentimiento expreso, libre, informado y previo.

– Consentimiento expreso:

Tabla 1. **Formas de establecer consentimiento**

Portales que establecen acciones ambiguas para significar consentimiento. El usuario acepta el uso de TR al navegar en la página.	***Portales que no identifican la forma de otorgar consentimiento.***	***Portales que no informan sobre el uso de cookies y otras TR.***
*Eltiempo.com *Semana.com *Minuto30.com *Caracol.com.co	*Elcolombiano.com *Elheraldo.com *Eluniversal.com.co	*Caracoltv.com *Canalrcn.com *Elespectador.com

Fuente: elaboración propia.

[9] Resultados del mes de abril. Top Sites in Colombia-Alexa. (2018). Recuperado de: https://www.alexa.com/topsites/countries/CO

- Consentimiento previo: todos los portales instalan mecanismos de rastreo desde el primer momento en el que se accede a ellos. La persona no tiene información o capacidad de negar el consentimiento antes de la instalación de las TR. Aunque Eltiempo.com es el único que despliega un aviso informativo sobre el uso de *cookies*, las instala de forma automática, ya que la forma de consentir es "navegar el portal", con lo cual el usuario no tiene opción para negar de forma previa la instalación de TR.
- Consentimiento libre: Ningún portal analizado ofrece condiciones reales para hacer una elección. Ninguno establece un mecanismo directo para aceptar o rechazar la instalación de mecanismos de rastreo. La aceptación del uso de TR es implícita al entrar en las páginas, pero el usuario debe ir a la configuración del navegador para detener el rastreo. De esta forma, se facilita y se automatiza el otorgar el consentimiento pero se dificulta negarlo.
- Consentimiento informado:

  (i) Aviso de presencia de tecnologías de rastreo: solo Eltiempo.com presenta un aviso visible que anuncia la presencia de TR, en particular *cookies*. En las otras nueve páginas no existe aviso visible; es necesario dirigirse al final de la página y buscar las políticas de privacidad de cada sitio para encontrar información relativa a TR. En el portal Elheraldo.com, no se encontró la política de privacidad; fue necesario hacer una búsqueda externa en Google para encontrarla.

  (ii) Informar sobre los mecanismos de rastreo que usa la página principal: las únicas páginas que no informan sobre la presencia de TR son Elespectador.com, Canalcaracoltv.com y Canalrcn.com

  (iii) Informar sobre quién tiene el control de los datos recolectados con TR: solo Eltiempo.com y Minuto30.com especifican las empresas que ellos contratan para recolectar datos por medio de TR en su página web.

  (iv) Informar sobre el tipo de datos que se recogen por medio de TR: todas las políticas de privacidad informan que se realiza tratamiento de los datos suministrados al inscribirse en los portales, tales como nombres, direcciones IP, teléfonos, correos, entre

otros. Sin embargo, ninguna página informa el tipo de datos que se recopilan específicamente con TR.

(v) Informar sobre la finalidad de los datos recolectados:

a) Los portales Elespectador.com, Canalcaracoltv.com y Canalrcn.com no dan información sobre la finalidad de los datos recolectados con TR en específico, pero informan sobre las finalidades generales de los datos personales recolectados. Los otros siete portales sí informan de forma separada las finalidades de los datos recolectados con mecanismos de rastreo.

b) Cada portal tiene una forma particular de describir la finalidad de la información recolectada con TR, lo que dificulta que el usuario identifique rápida y claramente los usos que tendrán sus datos. El grado de especificidad sobre las finalidades varía en cada página. Por ejemplo, Elcolombiano.com se limita a informar que los datos se usarán para "para personalizar y facilitar al máximo la navegación del USUARIO por su *site*". Eltiempo.com brinda información más detallada sobre distintos usos para facilitar el registro, para desplegar publicidad de interés, para personalizar los servicios del portal y de las *cookies* de desempeño que sirven para mejorar los servicios.

(vi) Informar sobre la instalación de TR de terceros como redes publicitarias y corredores de datos: solo los portales de Minuto30.com y Semana.com informan sobre la presencia de TR de terceras partes. Sin embargo, no detallan las consecuencias que esto acarrea.

8.1.2 Presencia de tecnologías de rastreo

La herramienta Ghostery identifica las TR que se instalan en el equipo terminal cuando se visita una página web. La herramienta clasifica las tecnologías de rastreo según su funcionalidad; estas pueden ser entre otras (Ghostery, s. f.):

– De publicidad: suministran publicidad o realizan acciones necesarias con este fin, como recolección de datos y análisis comportamental.

Estas TR son, en su mayoría, de terceros, tales como redes publicitarias y corredores de datos que no tienen relación con la página visitada ni con los usuarios.

- Esenciales: incluye administradores de *tags*, avisos de privacidad y otros elementos críticos para la funcionalidad de la página.
- Analítica o estadísticas de sitios: colectan información relacionada con el uso y funcionamiento del sitio.

Tabla 2. **Presencia de herramientas de rastreo reveladas por Ghostery entre abril y mayo de 2018**

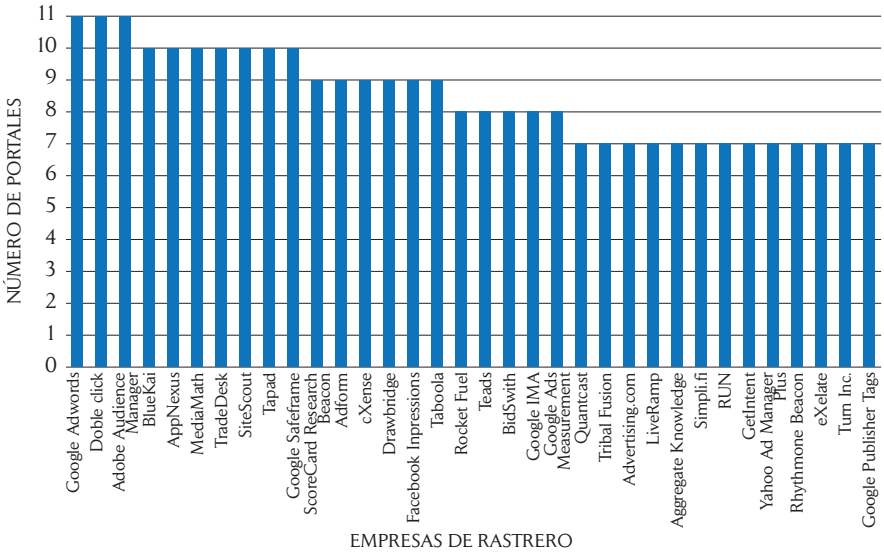
Sitio / tipo de rastreo	*Esenciales*	*Estadísticas de sitios*	*Publicidad*
Eluniversal.com.co	2	8	97
Minito30.com	2	5	86
Caracol.com.co	2	4	80
Eltiempo.com	1	6	69
Elcolombiano.com	1	4	55
Canalrcn.com	1	2	53
Dinero.com	1	4	44
Elheraldo.com	5	0	36
Semana.com	2	5	31
Canalcaracoltv.com	1	4	25
Elespectador.com	2	3	22

Fuente: elaboración propia con los datos arrojados por Ghostery. Estas mediciones cambian, ya que la presencia de rastreadores no es fija.

Si los usuarios no instalan mecanismos adicionales de protección en sus navegadores, desconocen totalmente las terceras empresas que recolectan datos sobre su navegación, los usos que le dan a esos datos y las posibles consecuencias que puede tener su perfilamiento y análisis. Blukei, por ejemplo, está presente en diez de las páginas analizadas y pertenece a Oracle, una empresa dedicada a la recolección masiva de información *online* y *offline* para generar campañas dirigidas[10]. Los datos agregados

[10] Política de privacidad de Oracle. Recuperado de: https://www.oracle.com/legal/privacy/marketing-cloud-data-cloud-privacy-policy.html

Figura 1. **Presencia de terceras empresas de rastreo hasta en siete de los portales analizados**



Fuente: elaboración propia con los datos arrojados por Ghostery.

incluyen información de las tarjetas de fidelización de tiendas y compras por catálogo; la información que se recoge en línea incluye identificadores únicos como las *cookies*, la dirección IP y la presencia de aplicaciones. La empresa recolecta información demográfica como el género, la edad y el rango de ingreso, también datos sobre el comportamiento de los usuarios en los sitios web, la publicidad que abren y la fecha y hora de las actividades. Además, la política de Oracle incluye la posibilidad de compartir e intercambiar datos con terceros. Así como Blukei y Oracle, la empresa Mediamath[11], presente en nueve de los portales analizados, recolecta información sobre el dispositivo, el historial de navegación y sobre la interacción en las páginas para ofrecer servicios de audiencias publicitarias con más de un billón de récords de consumidores.

[11] Política de privacidad de Mediamath. Audiences. Recuperado de http://www.mediamath.com/audiences/#partners

8.2 Conclusiones

(i) Ninguno de los portales analizados cumple con la totalidad de los requisitos para un consentimiento libre, expreso, informado y previo. Todos instalan TR antes de darle una opción real, sencilla e informada a la persona para aceptar o rechazar la instalación TR. Solo dos páginas informan quién controla la información recolectada con TR por encargo de la página principal. De los diez portales, solo tres informan sobre la presencia de TR instaladas por terceras empresas. En conclusión, es bajo el nivel de información y la capacidad de elección de las personas para prevenir ser rastreadas durante su visita a estos portales. Ninguna página analizada cumple con el estándar de la regulación colombiana.

(ii) Los portales analizados no brindan la información necesaria para que los usuarios comprendan la magnitud y el alcance del rastreo de terceros en sus páginas. Estos resultados muestran la falta de cumplimiento del estándar de consentimiento informado y la baja efectividad de la actual regulación para controlar los posible efectos negativos del tratamiento de datos en Colombia.

(iii) El rastreo web de terceros permite recolectar información personal que se agrega a otros datos que los corredores recolectan en otros medios. La agregación y cruce de esta información permite diseñar perfiles y categorizar a las personas. Sin una regulación apropiada, estos perfiles pueden tener consecuencias discriminatorias indeseadas en la sociedad.

(iv) Aunque en la mayoría de políticas de privacidad revisadas se da un parte de tranquilidad respecto a la anonimidad de los datos, existen diversos estudios que demuestran la facilidad de correlacionar datos para particularizar una persona y conocer diversos aspectos de su vida privada, como opiniones y creencias políticas. Al obtener esta información es posible personalizar diversos productos desde la propaganda comercial hasta la propaganda política y los mensajes sugestivos.

(v) La regulación de consentimiento informado en Colombia y el documento CONPES del año 2018 no proponen soluciones efectivas

para el riesgo que representa la recolección automatizada de datos humanos.

Referencias

Angwin, J. (30 de julio de 2010). The web's new gold mine: Your secrets. Retrieved from *The Wall Street Journal*: https://www.wsj.com/articles/SB10001424052748703940904575395073512989404

Ayenson, M. D., Wambach, D. J., Soltani, A., Good, N. & Hoofnagle, C. J. (2011). Flash cookies and privacy II: Now with HTML5 and ETag respawning. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390

Bakos, Y., Wurgler, F. M. & Trossen, D. R. (2014). Does anyone read the fine print? Consumer attention to standard form contracts. *Journal of Legal Studies, 43*(1), 1-35.

Belli, L., Schwartz, M. & Louzada, L. (2017). Selling your soul while negotiating the conditions: From notice. *Health and Technology, 7*(4), 453-467. Doi:10.1007/s12553-017-0185-3

Chmielewski, D. (4 de enero de 2016). *How "Do Not Track" ended up going nowhere*. Retrieved from Recode: https://www.recode.net/2016/1/4/11588418/how-do-not-track-ended-up-going-nowhere

Christl, W. (2017). *Corporate survelillance in every day life*. Viena: Cracked Labs.

Comisión Europea. (2017). *Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre el respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas y por el que se deroga la Directiva 2002/58/CE*. Bruselas. Recuperado de https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52017PC0010

Corte Constitucional de Colombia. (2011). Control constitucional al Proyecto de Ley Estatutaria No. 184 de 2010 "Por la cual se dictan disposiciones generales para la protección de datos personales", C-748/11.

Datacrédito Experian. (s. f.). *¿Quiénes somos?* Recuperado el 11 de mayo de 2018, de https://www.datacredito.com.co/empresas/index.jsp#/empresas/quienessomos

Departamento Nacional de Planeación. (2018). *Documento CONPES 3920 sobre la Política Nacional de Explotación de Datos*. Bogotá.

Dias, T. & Natusch, I. (2016). Te están stalkeando para darte un valor. Chupa datos de CODINGRIGHTS. Recuperado de https://chupadados.codingrights.org/es/te-estan-stalkeando/

Directiva 2002/58/CE del Parlamento Europeo y del Consejo relativa al tratamiento de los datos personales y a la protección de la intimidad en el sector de las. (12 de julio de 2002).

Do Not Track Online Act. (15 de diciembre de 2015). S. 2404-114th Congress. Retrieved from https://www.congress.gov/bill/114th-congress/senate-bill/2404/text

El Tiempo. (2017). Política de privacidad. Recuperado de http://www.eltiempo.com/politica-privacidad

Englehardt, S. & Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. *ACM Conference on Computer and Communications Security.*

Fallabella. (2 de julio de 2016). Política de tratamiento de datos personales. Recuperado de https://www.falabella.com.co/falabella-co/static/staticContent1.jsp?active=9&id=cat4840961

Federal Trade Comission. (2010). Protecting consumer privacy in an era of rapid change. Preliminary ftc staff report. Retrieved from https://www.ftc.gov/news-events/press-releases/2010/12/ftc-staff-issues-privacy-report-offers-framework-consumers

Federal Trade Commission. (2014). Data brokers. A call for transparency and accountability. Recuperado el 5 de junio de 2018, de https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Fundación Karisma. (6 de abril de 2018). Análisis confirma relación de Nation Builder, empresa que ayudó a Trump a llegar a la presidencia, con dos campañas presidenciales en Colombia. Recuperado de https://karisma.org.co/analisis-confirma-relacion-de-nation-builder-empresa-que-ayudo-a-trump-a-llegar-a-la-presidencia-con-dos-campanas-presidenciales-en-colombia/

Ghostery. (s. f.). What are the new tracker categories? Retrieved from https://ghostery.zendesk.com/hc/en-us/articles/115000740394-What-are-the-new-tracker-categories-

Grupo de trabajo Artículo 29. (2013). Working document 02/2013 providing guidance on obtaining consent for cookies.

Grupo de trabajo Artículo 29. (2018). Guidelines on consent under Regulation 2016/679.

Helberger, N. (6 de febrero de 2016). Profiling and targeting consumers in the Internet of Things - A new challenge for consumer law. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2728717

Hoofnagle, C. J., Soltani, A., Good, N. & Wambach, D. J. (2012). Behavioral advertising: The offer you can't refuse. *Harvard Law and Policy Review, 6*, 273.

In Re: Google Inc. Cookie Placement Consumer Privacy Litigation, Civ. No. 12-MD-2358 (SLR) (United States District Court For the District of Delaware 2 de 2 de 2017). Retrieved from http://www.ded.uscourts.gov/sites/default/files/opinions/slr/2017/february/12-2358.pdf

Internet Engineering Task Force. (Febrero de 1997). HTTP State Management Mechanism. *Request for Comments: 2109*. Retrieved from https://www.ietf.org/rfc/rfc2109.txt

Krishnamurthy, B. & Wills, C. (2009). Privacy diffusion on the web: A longitudinal perspective. *Proceedings of the 18th international conference on World wide web*, 541-550.

Kristol, D. M. (2001). HTTP cookies: Standards, privacy and politics. *ACM Transactions on Internet Technology*, 151-198.

Leal-Taixé, L., Milan, A., Schindler, K., Cremers, D., Reid, I. & Roth, S. (2017). Tracking the trackers: An analysis of the state of the art in multiple object tracking. *arXiv:1704.02781v1*.

Lenddo. (Junio 20 de 2016). How does lendo work? Retrieved from https://www.lenddo.com/

Leon, P. G., Ur, B., Balebako, R., Cranor, L. F., Shay, R. & Wang, Y. (2011). Why Johnny can't opt out: A usability evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 589-598.

Ley 1266 de 2008. (2008). "Por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países". *Diario Oficial* No. 47.219 de diciembre 31 de 2008.

Ley 1581. (2012). "Por la cual se dictan disposiciones generales para la protección de datos personales". *Diario Oficial* No. 48.587 de 18 de octubre de 2012.

Mayer, J. (Julio de 2011). Tracking the trackers: To catch a history thief. Retrieved from The Center for Internet and Society at Stanford Law School: http://cyberlaw.stanford.edu/blog/2011/07/tracking-trackers-catch-history-thief

Mayer, J. (Octubre de 2011). Tracking the trackers: Where everybody knows your username. Retrieved from The Center for Internet and Society at Stanford Law School: http://cyberlaw.stanford.edu/blog/2011/10/tracking-trackers-where-everybody-knows-your-username

Mayer, J. R. & Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. *Security and Privacy (SP), 2012 IEEE Symposium on*. Doi:10.1109/SP.2012.47

McDonald, A. & Cranor, L. F. (2010). Beliefs and behaviors: Internet users' understanding of behavioral advertising. *TPRC*.

Montjoye, Y.-A. D., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*. Doi:10.1038/srep01376

Nanji, A. (13 de diciembre de 2017). Global ad spend forecast for 2018. Retrieved from Marketingprofs: https://www.marketingprofs.com/charts/2017/33282/global-ad-spend-forecast-for-2018

Narayanan, A. (28 de julio de 2011). There is no such thing as anonymous online tracking. Retrieved from The Center for Internet and Society at Stanford Law School: http://cyberlaw.stanford.edu/blog/2011/07/there-no-such-thing-anonymous-online-tracking

Narayanan, A. (16 de enero de 2018). *The web tracking arms race: Past, present, and future*. Santa Clara: USENIX Association. Retrieved from https://www.usenix.org/node/208180

Narayanan, A. & Shmatikov, V. (2007). Robust de-anonymization of large datasets (how to break anonymity of the Netflix Prize dataset). *08 Proceedings of the 2008 IEEE Symposium on Security and Privacy*, 111-125. Doi:10.1109/SP.2008.33

Obar, J. A. & Oeldorf-Hirsch, A. (2016). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy 2016*.

OECD. (2013). The OECD privacy framework. Retrieved from http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

Oracle y Datalogix. (22 de diciembre de 2014). Oracle. Retrieved from http://www.oracle.com/us/corporate/press/2395487

Parlamento y Consejo Europeo. (12 de julio de 2002). Directiva 2002/58/ce del 12 de julio de 2002 relativa al tratamiento de los datos personales y a la protección de la intimidad en el sector de las comunicaciones electrónicas (Directiva sobre la privacidad y las comunicaciones electrónicas). *Diario Oficial* No. L 201.

Parlamento y Consejo Europeo. (27 de abril de 2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

Red Iberoamericana de Protección de Datos. (20 de junio de 2017). Estándares de protección de datos.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales (27 de abril de 2016).

Scheneider, B. (2015). *Data and Goliath. The hidden battles to collect your data and control your world.* Nueva York: W. W. Norton & Company.

Singel, R. (10 de mayo de 2012). Online tracking firm settles suit over undeletable cookies. Retrieved from Wired: https://www.wired.com/2010/12/zombie-cookie-settlement/

Sorjen. (10 de marzo de 2014). The history of online ad targeting. Retrieved from https://www.sojern.com/blog/history-online-ad-targeting/

Statistic. (s. f.). Desktop internet browser market share 2017. Retrieved from https://www.statista.com/statistics/544400/market-share-of-internet-browsers-desktop/

Superintendencia de Indsutria y Comercio. (2016). Concepto radicado No. 16-172268 del 9 de agosto de 2016. Bogota. Recuperado de http://www.sic.gov.co/recursos_user/boletin-juridico-sep2016/articulo/datos/tratamiento-datos-personales-a-traves-de-cookies.html

Toner, A. (7 de junio de 2017). With new browser tech, Apple preserves privacy and Google preserves trackers. Retrieved from Electronic Frontier Foundation: https://www.eff.org/deeplinks/2017/06/with-new-browser-tech-apple-preserves-privacy-google-preserves-trackers

YourAdChoices. (s. f.). YourAdChoices gives you control. Retrieved may 8th 2018 from http://youradchoices.com/control

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 75-89.