Agencia Española de Protección de Datos

Premio Nacional Emilio Aced

Convocatoria de 2017

Trackula

Autores:
Santiago Saavedra
Sofia Prósper
Guidiana Landívar
Alba Martín

Colaboran:
Pablo Martín
Carla Tortul
Adolfo Antón Bravo
Medialab-Prado

MEDIALAB
PRADO

7 de diciembre de 2017

Índice

Resumen

Trackula es un proyecto creado para concienciar al individuo medio sobre la protección de datos y los problemas de privacidad derivados del uso de las tecnologías de la sociedad de la información. El proyecto se centra alrededor de una extensión para navegadores, de forma que el usuario pueda experimentar su propia versión de lo que ocurre en base a sus interacciones con la web. Todos los datos que Trackula recopila se conservan exclusivamente en el dispositivo del interesado y no son cedidos de ninguna forma. Esta memoria describe el proyecto, enmarcado en los ámbitos legislativo y tecnológico.

En si mismo, el proyecto consta, además de la extensión del navegador, de una página web con la que comenzar una narrativa alrededor que permita al individuo medio interesarse paulatinamente por los distintos conceptos de protección de datos. Además, se recopilan herramientas útiles para el propósito de limitar los datos cedidos a terceros a la hora de visitar páginas web.

Este proyecto, que se enmarca dentro del espacio de trabajo de Visualizar'17 de Medialab-Prado, está formado por un grupo multidisciplinar en el que se aúnan diversas disciplinas para llevar la narrativa a cabo.

1. Introducción

Trackula es un proyecto que nace para crear concienciación sobre privacidad de datos y que plantea la cuestión de qué nivel de protección merecen datos aparentemente inocuos que, al representarse en su conjunto pueden describir perfiles de personas identificables. En un marco legislativo adaptándose de la Directiva 95/46/CE a la entrada en vigor del Reglamento 2016/679, este texto pretende explicar las pretensiones de Trackula y su labor en cuanto a la temática de la protección de datos.

Trackula se origina a partir de un grupo de trabajo multidisciplinar enmarcado dentro del taller Visualizar'17 de Medialab-Prado, en Madrid. Este es un taller organizado por el Departamento de Periodismo de Datos de esta entidad. Después de la conclusión del marco de trabajo organizado, los participantes en el proyecto continúan con el mismo tratando de mejorarlo y de darlo a conocer a la ciudadanía. Uno de los resultados del proyecto es un plugin interactivo que, tras ser instalado y utilizado por el usuario, le permite al mismo obtener información sobre la exposición que está haciendo a terceros de sus hábitos de navegación.

Con esto, de Trackula emana una actividad de concienciación y la capacidad de proporcionar un marco de trabajo más amplio sobre el que construir herramientas que permitan que los usuarios de la web conozcan mejor la huella que dejan al comunicarse con otros sitios web.

A lo largo del texto se utiliza vocabulario relacionado con la protección de datos. En donde sea aplicable, serán tenidas en cuenta las definiciones de estos términos según lo establecido en el Reglamento 2916/679 del Parlamento Europeo y del Consejo.

En la próxima sección se introduce la materia, tanto el marco legislativo como el tecnológico. En la siguiente se consideran las cesiones de datos personales en el conjunto de ambos marcos de la realidad. A continuación se explica el proyecto realizado y su rol en la concienciación sobre la protección de datos. En una sección posterior se mencionan trabajos relacionados con la protección de datos y con la privacidad, tanto desde el ámbito institucional como desde otras partes interesadas, y por último se extienden unas conclusiones del trabajo y de la situación, así como posibles líneas futuras para el proyecto.

2. Estado del Arte

En esta sección hablaremos del *status quo*. La primera sección comenta el estado legislativo, y de como ello afecta al problema en cuestión que se plantea en la sección siguiente. A continuación, una segunda subsección describe cuestiones técnicas relevantes que pueden influir o verse influidas por el marco legislativo.

2.1. Legislación

En el marco legislativo ha de considerarse la actual Directiva 95/46/CE, desarrollada en España en la Ley Orgánica de Protección de Datos (LOPD), vigente actualmente, así como también en el Reglamento 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 y a la libre circulación de estos datos (en adelante, RGPD) que será directamente aplicable en cada Estado miembro y entrará en vigor el 25 de mayo de 2018. En la LSSI-CE se establecen los criterios por los cuales un prestador de servicios puede incluir en los dispositivos de comunicación de personas físicas identificadores únicos u otro tipo de información que puede permitir discernir comunicaciones con diferentes interlocutores.

En particular, el nuevo Reglamento de Protección de Datos (RGPD) define más detalladamente que la actual LOPD lo que es información personal, y en particular, detalla que será «toda información sobre una persona física identificada o identificable [...]; se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo [...] un número de identificación, datos de localización, un identificador en línea o [otras características de la persona]».[1] En el preámbulo del Reglamento se detalla en la Consideración 30ª explícitamente a las cookies y se deja abierta a interpretación el uso de otros identificadores, explícitos o no, pudiendo interpretar que se incluyen técnicas como Browser Fingerprinting (al incluir cualesquiera "datos recibidos por los servidores"). Esto implica un marco regulador para las interacciones en sitios web que incluyen contenido de terceros (a través de tecnologías como, por ejemplo, los *iframe*) más estricto que el anterior, ya que se extiende el

[1] Reglamento 2016/679 Artículo 4, apartado 1.

requisito de informar del consentimiento de datos personales incluso más allá de la utilización de cookies u otros elementos identificativos *proporcionados por el proveedor* (para lo que era de aplicación la Ley de Servicios de la Sociedad de Información y Comercio Electrónico), para incluir aquellos *proporcionados por la plataforma del usuario* que anteriormente podrían no estar siendo considerados datos sobre una persona física identificable.

De este modo, debe de pasar a un segundo plano la propuesta sobre el uso de cookies y cobrar importancia la misma comunicación inicial, debido a que en ésta el usuario ya puede estar proporcionando información identificativa.

Los distintos Dictámenes del Grupo de Trabajo del Artículo 29 de la Directiva 95/46/CE (WP29), a pesar de no tener carácter regulatorio, resultan muy relevantes, y son bastante clarificadores en la forma en la que instan a la Comisión a que se interprete lo escrito en la ley, y a entender aquellas cuestiones que no habían sido explicitadas en marco legislativo, como las formas en las que es razonable realizar un consentimiento informado, o si las técnicas de Browser Fingerprinting son consideradas en el ámbito legislativo. Algunos de estos Dictámenes son finalmente adecuados al Reglamento General de Protección de Datos.

2.2. Consideraciones técnicas

Por otro lado, el funcionamiento técnico de la web no tuvo en cuenta estas consideraciones de privacidad en un inicio. Además, toda iteración en la tecnología se ha hecho pensando en la retrocompatibilidad y en los principios de *graceful degradation*, es decir, que los comportamientos que ya estaban disponibles en versiones anteriores sigan estándolo. Esto complica extremadamente la posibilidad de incluir de forma efectiva métodos tecnológicos para garantizar la privacidad de un modo que no sea fácilmente esquivable.

En esencia, la tecnología de rastreo más sencilla, el *Browser Fingerprinting*, es prácticamente inevitable de forma técnica (está en manos del usuario tratar de enviar peticiones de la forma más genérica posible, pero ello evitará al servidor de contenidos conocer información sobre el agente de usuario relevante para mostrar la disposición más adecuada a su dispositivo). Y las formas de identificación directa como las cookies o el almacenamiento de identificadores en otros lugares del navegador como el

localStorage o la IndexedDB están preparados para no solicitar permiso previo al usuario, para garantizar la experiencia de usuario más fluida posible. Esto obliga a que los propietarios de los sitios web tengan que tomar medidas específicas en cuestiones de protección de datos para poner límites a lo que puede hacerse técnicamente desde sus sitios, y actuar acorde al arreglo jurídico. Por otra parte, las innovaciones que ha habido en la web han venido más motivadas por la extracción de la información y la recopilación de datos para su análisis posterior que por las vistas en la privacidad. Prueba de ello es la inexistencia de grupos específicos dentro del World Wide Web Consortium (W3C.org) o del Web Hypertext Application Technology Working Group (WHATWG) comprometidos directamente con la privacidad.

En contraste, el W3C sí ha emitido recomendaciones respecto del uso de Extensiones de Cifrado de Medios (Encrypted Media Extensions, o EME), lo que muestra los intereses del W3C en proteger cierta información sensible, en este caso, una recomendación destinada a mejorar la protegida con derechos de autor.

Es más, la recomendación actual del WHATWG (el Grupo de Trabajo que se encarga de estandarizar HTML5) en su sección §1.10.1 (Introduction/Privacy Concerns/Cross-site communication) dice que «Fundamentally, users that do not trust a site to treat their information with respect have to avoid visiting that site at all.», dejando el grupo de trabajo en el usuario (o, al menos, fuera del apartado técnico) cualquier solución al respecto del tratamiento de la información.

Fundamentalmente esto es inevitable, ya que los estándares definidos actualmente en torno a la web prefieren que la experiencia del usuario sea sin interrupciones antes que apostar por la privacidad del mismo. Aún así, un mejor diseño permitiría que el usuario pudiese elegir en su agente de usuario cuáles de estas características quiere habilitar a cada sitio web.

Por ejemplo, el mecanismo de los navegadores para permitir a un sitio web mostrar notificaciones es un API en el que todas las funciones están deshabilitadas salvo la que permite pedir permiso, hasta que éste se otorgue. De este modo, hasta haber pedido permiso y que este haya sido concedido, no puede utilizarse la funcionalidad. De un modo similar podrían haberse diseñado las características que permiten a un sitio web interactuar con otros y enviar información a terceros, así como decidir qué información proporcionar incluso al sitio web visitado en primera

instancia.

3. Las cesiones inadvertidas de datos en la web e implicaciones

Este trabajo se pretende enfocar desde la perspectiva de los datos generados y cedidos por los usuarios, la mayor parte de las veces inadvertidamente, al navegar por la web. Sobre todo, de las interacciones con sitios de terceros incrustados en un sitio web. Pese a la LOPD y la LSSI, que sientan bases respecto a la forma en la interacción que se da entre un cliente y un sitio web principal, no quedan muchas veces claros los límites en cuanto a terceras partes que tienen contenidos incrustados y que por tanto también son conocedoras de tus datos y que consideramos que no están explícitamente regulados.

El problema subyace en que, por fundamentos tecnológicos, la forma en la que se experimenta la web implica que el navegador cargue ciertos recursos de una página antes de que el usuario pueda aceptar términos algunos de tratamiento de información, puesto que para ello, al menos, dichos términos tendrían que ser transmitidos al usuario y, por tanto, ya habría una transmisión previa que consideraría ser tratada. En particular, si este sitio web solicita recursos a terceras partes por razones técnicas (una Red de Distribución de Contenido para poder servir su contenido más rápidamente a sus usuarios, por ejemplo), éstas terceras partes son sujeto de un tratamiento de datos con el que el individuo podría no estar de acuerdo, pero no puede saberlo antes de que éste se produzca.

La cuestión en esencia radica en que un mismo proveedor de contenidos (por ejemplo, una web de alojamiento de vídeos, un complemento para compartir el contenido en una red social, o un proveedor de analíticas) puede estar incrustado en páginas web de muy diversa índole, y si varias de ellas son visitadas por un mismo usuario, cabe la posibilidad de identificarle. En particular, ha de tenerse en cuenta también que, en el caso de utilizar la tecnología de iframes, entre los datos proporcionados a estos terceros, se incluye, salvo que el sitio web haya dispuesto otra cosa, la propia referencia completa a la página web visitada (no solamente el dominio, sino la URL completa) como campo *referer*.[2] en las cabeceras de

[2]La dirección referer se refiere al origen en el cual esta tercera parte se encuentra

la petición que se hace al tercero. Por consiguiente, si esta tercera parte está incrustada en muchas de las webs que visitas, podrá obtener, después de un tratamiento de los datos, una parte del historial de navegación de un usuario, en base a consolidar los *referer* obtenidos de cada interacción anterior, y discriminarlo del historial de otros usuarios, en base a identificadores de sesión que haya, o bien podido establecer previamente en el navegador, o mediante técnicas de *fingerprinting*.

Ejemplo 1. *Un usuario visita tres noticias diferentes, cada una en un periódico digital diferente. En todos ellos, su contenido incluye un vídeo, y el proveedor del mismo es YouTube.*

Supuesto 1. Si el usuario ha previamente iniciado sesión en YouTube, y aceptado los términos en los que sus datos sean tratados (a pesar de lo laxo de su redacción), YouTube habrá identificado al usuario por su cookie y el vídeo, aunque no sea reproducido ni considerado en el historial visible del usuario, podrá ser tenido en cuenta para la elaboración de un perfil, que pueda, entre otras, alterar el servicio de recomendación de YouTube.

Supuesto 2. Si el usuario no ha iniciado previamente sesión en YouTube, cabe esperar que YouTube no establezca ninguna cookie en su navegador. En este caso, YouTube todavía podría identificar al usuario por su IP y su agente de usuario, en combinación con anteriores accesos del usuario habiendo iniciado sesión. Si el usuario, al haber iniciado sesión había aceptado en los términos de YouTube la posibilidad de identificarle aún sin haber iniciado sesión, entonces no estaremos encontrando en un caso similar al del supuesto 1. En caso contrario, al supuesto 3.

Supuesto 3. El usuario no es usuario de YouTube y no ha iniciado sesión previamente en él. Aún así, YouTube podría estar recopilando información para elaborar un perfil de esta persona seudónima y recomendarle vídeos como si hubiese iniciado sesión, pero por no haberlo hecho, resultaría complicado identificarle para que pueda ejercer sus derechos, por lo que no podrá oponerse a los objetos del tratamiento que realice YouTube posteriormente.

incrustada. Por ejemplo, incrustar un plug-in de contenido social en una página web podrá permitir recibir al proveedor del contenido social la dirección completa de dicha página web, sin previa interacción del usuario, debido al funcionamiento técnico de los iframe, según el apartado 4.8.5 del estándar HTML del WHATWG, salvo que se establezca lo contrario mediante una "referrer policy" apropiada.

4. Trackula

El objetivo de Trackula es mostrar al usuario las interacciones latentes con terceras partes cuando navega por la web. En esta sección se describe en detalle los componentes del proyecto, su funcionamiento y las bases sobre las que se sustenta.

4.1. Introducción

La web ha sido ideada alrededor del hiperenlace y, en la actualidad, de lo que se ha venido llamando *hypermedia*, de forma que es muy sencillo técnicamente incluir recursos de terceros en una página. Esto, que tiene muchas ventajas técnicas, pues al incluir recursos de distintos servidores podemos, entre otras, acelerar la carga de una página, también presenta una serie de complicaciones desde el punto de vista de la protección de datos, ya que cuando estos terceros servidores no son propiedad de la misma entidad de la interacción principal, cabrían múltiples interpretaciones respecto de sobre quién descansa la responsabilidad del flujo de datos que ocurre entre el navegador del usuario y estas terceras partes. Según la legislación actual y la que entrará en vigor, esta responsabilidad ha de caer en el propietario de la página web que el usuario visita, ya que solamente él puede anticipar las necesidades que nuestro navegador requerirá de terceros para ver su página (puesto que es su recurso, y por tanto son su responsabilidad catalogar los enlaces que son automáticamente incrustados en su web desde el punto de vista de cesión de datos).

El proyecto trata de mostrar, concretamente, cómo estas interacciones, inadvertidas en muchos casos, e incluso técnicamente imprescindibles para mostrar lo que el propietario desea, pueden poner en peligro Datos Personales, e incluso Datos Personales de Especial nivel de protección, por la mera circunstancia de ser incrustadas en más de un sitio web explícitamente visitado.

La regulación de la LOPD, y los Dictámenes del Grupo de Trabajo del Artículo 29 describen varios de los límites con los que una web ha de *informar* al usuario de la cesión de datos que está suponiendo implícitamente visitar la web, pero la seguridad jurídica que rodea a la cesión misma en el momento de producirse es bastante escasa.

Por un lado, en muchos casos ocurre una "aceptación implícita" de estas cesiones, a pesar de los Dictámenes del WP29 sobre la definición

de consentimiento y sobre la exención de la necesidad de consentimiento en ciertas circunstancias, y tal y como traslada la Agencia en su Guía de Cookies y otros documentos, en la mayoría de casos una aceptación implícita es inadmisible como consentimiento. En varias webs consultadas, las cookies que hayan de establecerse en el navegador lo hacen con la primera petición, antes de que el usuario pueda dar su consentimiento informado, y al utilizar el protocolo HTTP a la hora de enviar estas cookies es complicado evitar tal efecto. Por lo tanto, aún negándose el usuario a cualquier tratamiento ulterior, resulta descabellado pensar que, en la actualidad, el usuario pueda ejercer sus derechos ARCO[3] a todas las terceras partes que han recabado datos en estas interacciones.

A pesar de que podría caber pensar que tales derechos todavía no han nacido, pues por la mera interacción con una web el navegador no habría de proporcionar datos considerados como personales en sí mismos, la interacción repetida con un mismo tercero puede, si éste tuviese interés, generar datos suficientes para producir un perfil del visitante, de un modo que podría interpretarse, en el nuevo marco del RGPD, como datos seudonimizados. Con la tecnología actual de perfilado, es razonable considerar que el esfuerzo para recomponer la identidad de la persona tratada no sería desproporcionado, en términos legales (Consideración 62ª), y por tanto tendrían la consideración de datos de personas identificables.

Por otro lado, otra cosa que ocurre debido a cómo es la propia tecnología, es que se necesitan cargar recursos ya antes de comenzar la interacción con el sitio web; recursos incluso de terceros, antes de que se haya podido informar al usuario de los posibles tratamientos de datos y, por tanto, imposibilitando la oposición a ciertos tratamientos antes de que ocurran por la propia razón de ser de la tecnología utilizada. En estos casos además, a pesar de que un usuario haya podido haber leído la política de privacidad de un tercero y estar de acuerdo con los objetivos de los tratamientos de datos, ésta política puede, en muchos casos, cambiar sin aviso previo.

[3]Los derechos ARCO son los de Acceso, Rectificación, Cancelación y Oposición, según se definen en la Ley Orgánica de Protección de Datos o en el Reglamento General de Protección de Datos.

4.2. Intención y objetivos del proyecto

El Grupo de Protección de Datos del Art.29 de la Directiva 95/46/CE en su Dictamen 16/2011 reconoce que, «numerosas encuestas públicas han mostrado que el usuario medio de Internet no sabe que su comportamiento está siendo seguido mediante cookies u otros identificadores únicos, ni tiene conocimiento de quién lo hace y con qué finalidad. Este desconocimiento contrasta enormemente con la creciente dependencia de muchos ciudadanos europeos del acceso a Internet para actividades cotidianas como compras, lectura, comunicación con los amigos y búsqueda de información.» por ello Trackula trata de hacer ver las implicaciones de cesión de datos personales a terceros, y cómo éstos pueden recomponer el historial de tus visitas, mostrando lo sencillo que resultaría seguir este patrón de comportamiento simplemente guardando las interacciones en un grafo.

Existe la necesidad de informar a los usuarios sobre lo que está ocurriendo en la web, y en general con los datos que generan, y con cómo estos son tratados, ya que en la mayor parte de casos, en las situaciones que se están considerando tienen el potencial de convertirse en datos de alto riesgo tras el tratamiento. Las personas no deben continuar adquiriendo más dependencia al acceso a internet sin antes ser conscientes de las implicaciones que esto tiene. El proyecto de Trackula lo que intenta es realizar este data awareness y hacer crecer la cultura de la privacidad, para que todos los individuos puedan relacionarse con los medios digitales de manera informada y siendo dueños de sus datos o por lo menos siendo conscientes de las consecuencias que tiene el cederlos.

5. Significado del nombre

La lógica del nombre viene derivada de lo que se entiende por "trackers" en este proyecto: aquellas terceras partes que se encuentran en varias páginas web visitadas y que, por tanto, tienen el potencial de estar rastreando, al menos parcialmente, la navegación web del usuario. El concepto detrás de la palabra 'Trackula' vendría a ser el tracker de los trackers, el ente que los visibiliza, los saca del secretismo y la oscuridad para que cualquier ciudadano sepa de su existencia y pueda actuar para defenderse de sus efectos.

Siendo la realidad del problema tan compleja y también tan desconocida para el usuario medio, el proyecto se apoya en diferentes recursos que van permitiendo que el individuo pueda acercarse paso a paso a ella.

6. Narrativa del proyecto

Como punto de entrada para poder entender y dar a conocer el proyecto aparece el sitio web https://trackula.org en donde se dan unas primeras pinceladas sobre situación actual y un breve resumen del mismo: «Queremos que conozcas el mundo digital en el que vivimos, el alcance de tus datos y lo que conlleva para tu libertad personal que seas rastreado por la web. Defendemos tu derecho a la privacidad y trabajamos para concienciar acerca del recorrido de tu información. Creemos que esto te interesa seas quien seas.

El plugin de Trackula.org muestra en una visualización interactiva cómo tus datos migran por la red y qué empresas de seguimiento web, o trackers, se conectan a los contenidos que visitas cada día. Lo hacemos sin guardar, vender o enviar tu información a nadie, todo queda en tu ordenador. Nuestra única intención es que seas consciente de esta información oculta y que puedas adquirir una visión crítica sobre esta cuestión. Somos activistas y nacimos en MediaLab Prado.»

En la web también se muestra la dirección del repositorio público en el que se encuentran sus fuentes y una política de privacidad que no es más que la propia explicación del funcionamiento técnico de la privacidad en la web y motivaciones del proyecto en sí: «Trackula no solo le plantea al usuario el problema de la privacidad, sino que activamente trata de respetarles en este aspecto. Por ello, los humanos tras esta web no recopilan ningún dato personalmente identificativo de forma directa o indirecta. Sin embargo, la privacidad es un asunto complicado. Por el mero hecho de entrar en esta web, existen una serie de trazas técnicas que es imposible evitar, entre ellas: habrás realizado una petición DNS a algún servidor fuera de nuestro control (probablemente de tu proveedor de servicios, o un DNS público como Google u OpenDNS). En caso de que tu proveedor no tenga cacheada la respuesta DNS, éste además habrá hecho una petición a CloudFlare, que es la proveedora de los servicios de DNS para este dominio. El servidor que está poniendo esta página a disposición de Internet es, actualmente, GitHub Pages, con lo que GitHub Inc.

podría estar registrando los accesos a esta página, incluyendo la dirección IP del usuario, y aquellos datos que su navegador proporciona, como la información sobre el agente de usuario, idiomas aceptables o sistema operativo.

Teniendo en cuenta las dificultades técnicas, se ha tratado de minimizar la exposición de información a terceras partes. En particular, utilizamos un dominio propio, y no cargamos ningún recurso de terceras partes. Esto impide que se utilicen la mayoría de métodos de rastreo de navegadores, salvo aquellos mencionados arriba por razones técnicas.

Téngase en cuenta que para que esta página llegue a su destino, múltiples entidades más la han hecho llegar hasta el dispositivo final. En particular, el proveedor de acceso a Internet que el usuario utilice, así como todos los Sistemas Autónomos de Internet que hayan sido involucrados en la interconexión con el/los proveedor/es de acceso a Internet en el centro de datos desde el que se esté sirviendo esto y que es potestad de GitHub. GitHub y CloudFlare ofrecen estos servicios de forma gratuita y automática, y Trackula no recibe ningún tipo de información ni contraprestación, ni tiene vigente ningún trato con GitHub diferente a lo disponible libremente en sus respectivas páginas web.» La web también aloja una infografía que llama a la acción aludiendo a la pérdida de privacidad que el usuario puede sufrir de seguir relacionándose con los medios digitales como hasta ahora lo venía haciendo, y que va en línea con las consideraciones del RGPD respecto a estos problemas, y la libertad que se pierde al ceder el control sobre datos personales.

Trackula.org incluye también una serie de proyectos relacionados para que el usuario pueda ejercer un mayor control sobre los datos que genera con sus navegaciones, los cuales se mencionan en la sección siguiente.

Una vez que el visitante se encuentra contextualizado, puede dar un paso más instalando el plug-in (disponible en el repositorio público de add-ons de Mozilla https://addons.mozilla.org/en-US/firefox/addon/trackula/) para visualizar el problema a partir de su propia navegación. Para realizar esta visualización, Trackula reaprovecha el trabajo anterior de Mozilla Lightbeam en el componente técnico que recaba las interacciones de los usuarios (almacenándolas exclusivamente en su propio navegador). A diferencia de Lightbeam, Trackula genera una visión lineal, como si se tratase del historial de páginas visitadas (y sus enlaces con terceros) en lugar de la visión holística de mostrar todo el panorama de páginas y terceras partes conectadas en un solo vistazo.

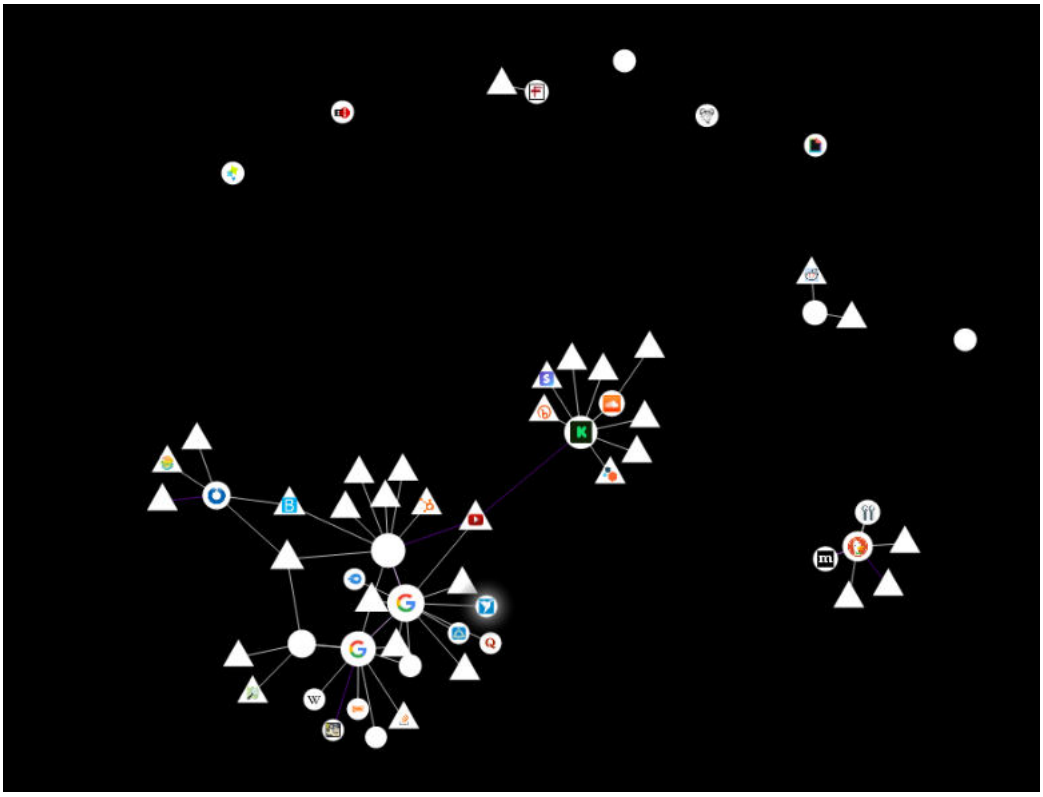

Figura 1: Visualización realizada con Mozilla Lightbeam.

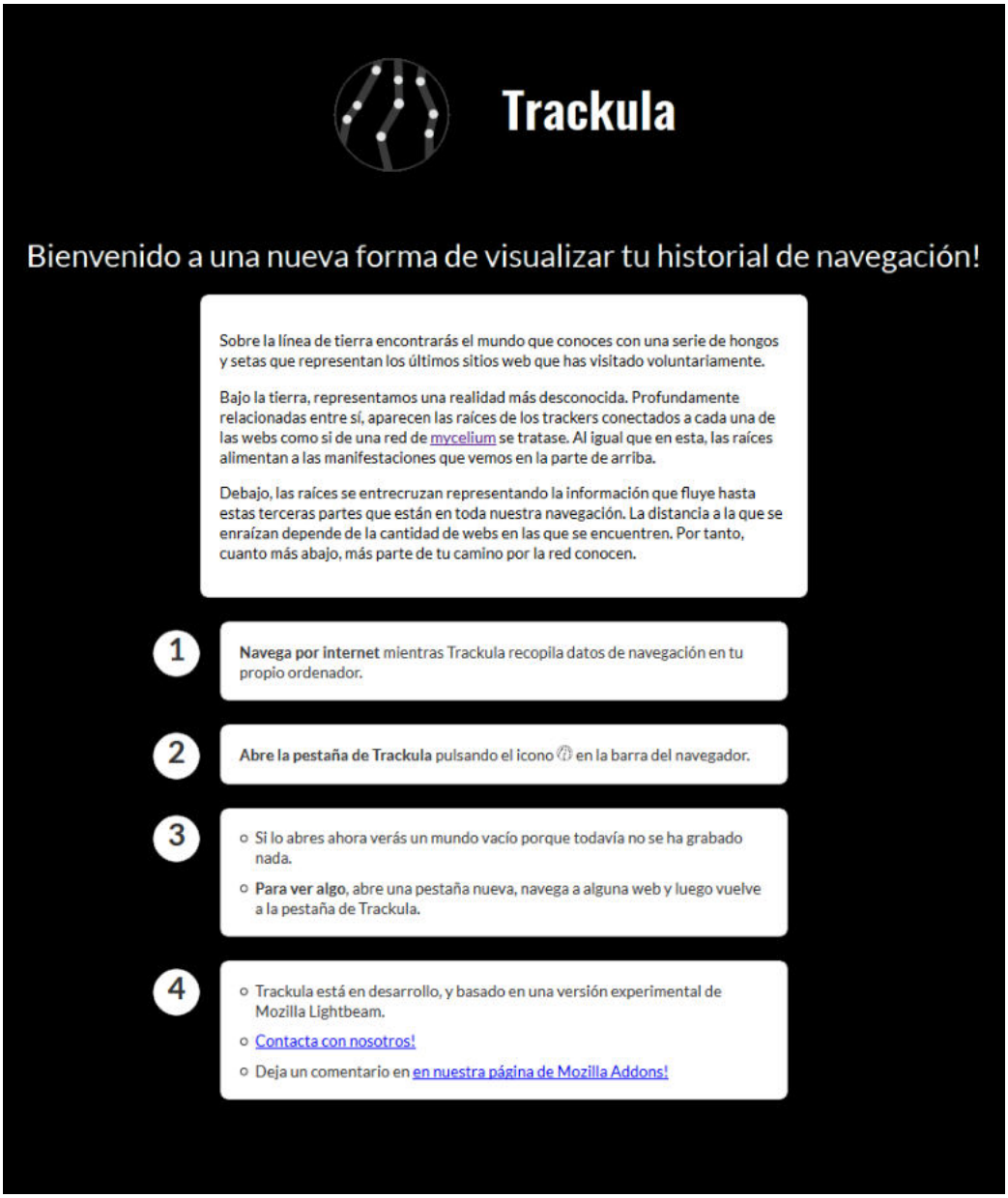

Figura 2: Mensaje de entrada al instalar el plug-in.

Además, la tecnología utilizada para la visualización de Trackula está basada en el dibujo, pensada como una infografía personalizada (utilizando p5.js y la tecnología canvas de HTML) en lugar de utilizar imágenes vectoriales sintetizadas. Cada visualización, por tanto, es única para cada persona, ya que depende de las webs que ha visitado y del nivel de privacidad que tiene configurado en su navegador. Por ello, en la medida que el usuario esté más concienciado y utilice más barreras de protección de sus datos, la visualización será más limpia, con menos nodos y menos profunda. La singularidad para cada usuario en cada momento define a la perfección uno de los objetivos clave del proyecto, que es que el sujeto tenga una experiencia personal, con sus propios datos, en su propio dispositivo y sin almacenamiento alguno de datos en ningún lugar fuera de su alcance. Todo se basa en la información que él mismo ha generado con su terminal y no sale de este.

Esta presentación mejora el entendimiento para aquellos que no están familiarizados con la forma en la que suceden estas interacciones en la web y que podrían dar una mayor visibilidad a la situación que la presentación completa de Lightbeam. A pesar de ello, se incluyen los componentes de visualización de Lightbeam en el proyecto para que puedan también utilizarse en un futuro ambas visualizaciones.

La visualización que la persona se encuentra después de una breve navegación se compone de varias partes. En un primer lugar hay una barra que recuenta el número de lugares visitados, voluntariamente, comparado con el número de terceros a los que se ha conectado el navegador y la fecha desde la que se recopilan estos datos. Observando simplemente estos tres valores se puede llegar a percibir la magnitud del problema que se intenta poner de manifiesto ya que el número de terceras partes resulta ser diez veces mayor que el de webs visitadas.

En la parte derecha aparece un desplegable señalado con un icono de un matraz que muestra los trackers más persistentes que la persona tiene. Como elementos principales se encuentran las webs visitadas y las terceras partes a las que estas se conectan. En este punto cabe destacar que la representación misma de la visualización así como toda la estética del proyecto responden a una metáfora que simboliza dos mundos, él que el usuario percibe, conoce y accede voluntariamente, que serían las webs que visita, representadas como diferentes tipos de setas y hongos y, bajo este, el mundo de lo desconocido, de las terceras partes implicadas, y las conexiones que estas tienen con los sitios web que el usuario visita sin

que desde el exterior esto se pueda percibir. Se hace referencia al reino fungi ya que el micelio de este reino funciona de manera análoga, son las raíces que se entrecruzan interconectando entes que están en la superficie y haciendo que la información fluya hasta estas terceras partes.

La distancia a la que se enraízan los trackers dependerá de la cantidad de webs en las que se conecten. Por tanto, cuanto más abajo aparece el tracker, más parte del historial por la red conocen y por tanto más datos sobre el usuario pueden haber recabado. En cierto modo podría decirse que cuanto más scroll tenga que hacer el usuario hacia abajo y más trackers aparezcan en lo más profundo, más conectados están con todas las páginas que se han visitado por lo que indirectamente más información se ha cedido a estas terceras partes.

Como se ha puesto de manifiesto anteriormente, pese a que esta información que se cede pueda parecer inofensiva durante la navegación, los sitios web acceden a datos que permiten obtener una huella del usuario bastante precisa (datos como el tamaño de la pantalla, el sistema operativo, la dirección IP, logs, parámetros de la URL, preferencias y características del navegador, clicks realizados y otros, como cookies identificativas). Este conjunto de referencias, junto con la información temporal relativa a las peticiones realizadas puede permitir a un posible tracker recomponer un perfil del usuario y poder tratar datos diseminados de un modo coherente y que pueda representar información de una persona identificable. Dependiendo de la variedad de datos en el perfil, incluyendo el tipo de páginas web visitadas, o información que puede desprenderse de la forma en la que se haya incluido el contenido del tracker en la página (por ejemplo, proporcionando la dirección de la página principal al tracker, mediante técnicas como la propagación de la dirección origen mediante el campo referer, que como se menciona en la sección anterior, es el comportamiento por defecto), este perfil puede incluir datos personales considerados de alto riesgo, como podría ser información médica o de orientación sexual o religiosa, por ejemplo, a partir del comportamiento de un usuario al leer noticias o buscar información sobre determinados fármacos o noticias.

La visualización que el usuario recibe de su navegación con el plug-in instalado es interactiva por lo que puede ir inspeccionando los diferentes nodos que aparecen en ambos mundos (las webs visitadas voluntariamente y el mundo oscuro de los trackers) para que las conexiones que cada uno tiene con el resto aparezcan resaltadas y facilite el entendimien-

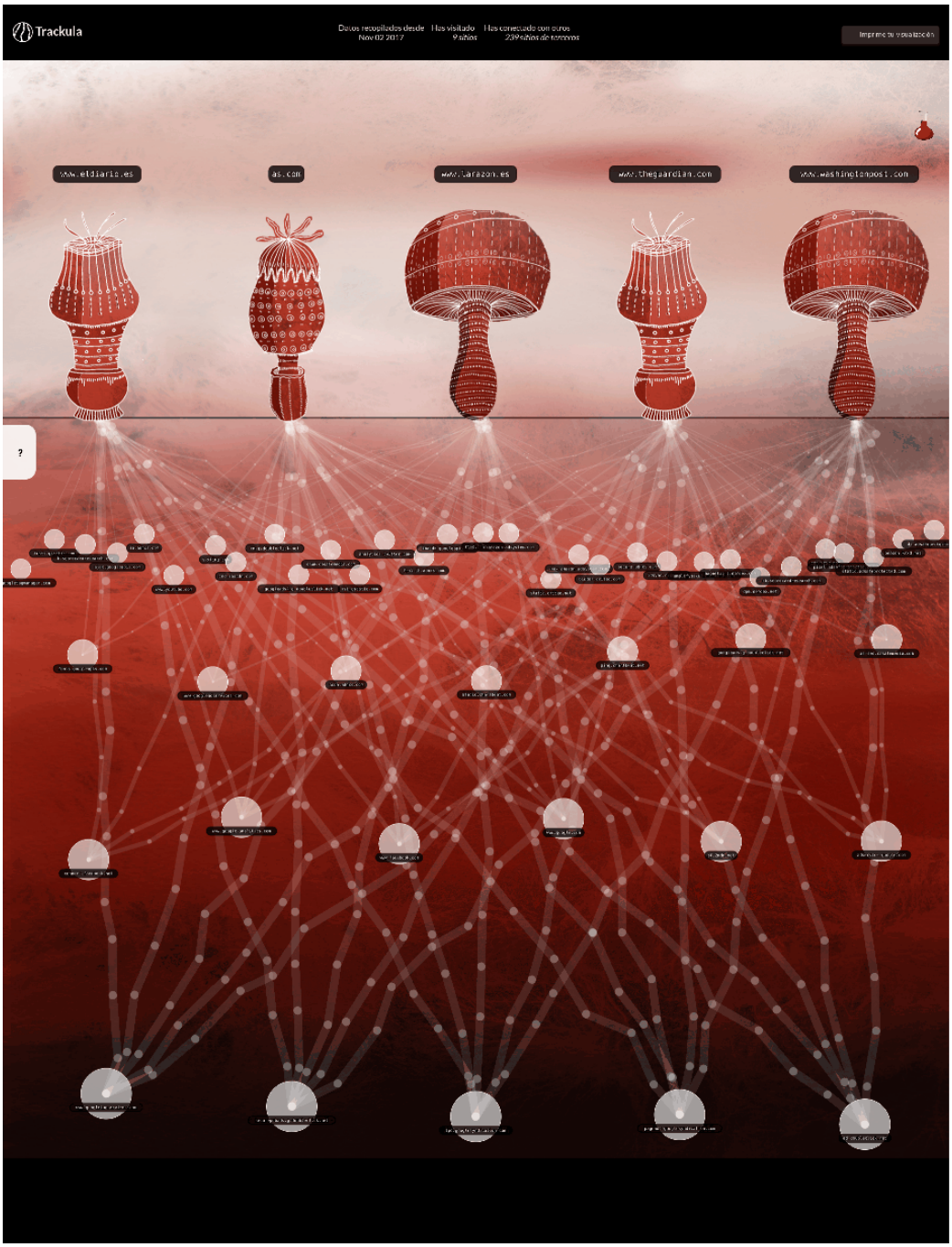

Figura 3: Vista generada por Trackula tras una navegación de ejemplo.

to de la información que se está representando.

Para una mayor inmersión en el contexto se ha realizado una investigación periodística cuyo resultado son píldoras de información de máximo 140 caracteres que se van desplegando y cambiando cada vez que el usuario acerca el puntero al símbolo '?'. Estos 'Sabías que?' son de naturaleza diversa y varían desde contenido didáctico que explica conceptos relacionados con el de los datos personales hasta titulares de noticias, pasando por datos curiosos sobre el tema.

7. Trabajo relacionado

En materia de Protección de Datos, ciudadanos, instituciones, organizaciones de consumidores, entes privados y colectivos se han manifestado sobre la necesidad de garantizar la efectiva tutela de la privacidad. Hay numerosos trabajos, tanto previos como posteriores al estado actual de la legislación, en los diferentes países de la Tierra que trabajan en pro de la protección de la privacidad y la libertad de los usuarios. En este apartado se desgranan las diferentes actividades y proyectos desarrollados por diversos entes.

7.1. Labor institucional

En España, existen por un lado las iniciativas estatales llevadas a cabo por la Agencia Española de Protección de Datos, como el Premio Anual Emilio Aced, o el Premio a la Comunicación de Datos Personales. También son de destacar otras iniciativas europeas como las incluidas dentro del programa de financiamiento Horizonte 2020 sobre protección de datos.[4]

La Agencia Española de Protección de Datos ha editado una gran cantidad de proyectos y guías educativas enfocadas de manera didáctica para adultos y niños. Algunos de ellos se componen de vídeos que hablan de manera cercana sobre temas relacionados con la privacidad en la Red. Con estos vídeos se le plantean al espectador supuestos en los que podría perder su privacidad y poner en riesgo de alguna manera

[4]Entre otras, las identificadas como REC-RDAT-TRAI-AG-2017, SU-DS03-2019-2020 o DS-08-2017.

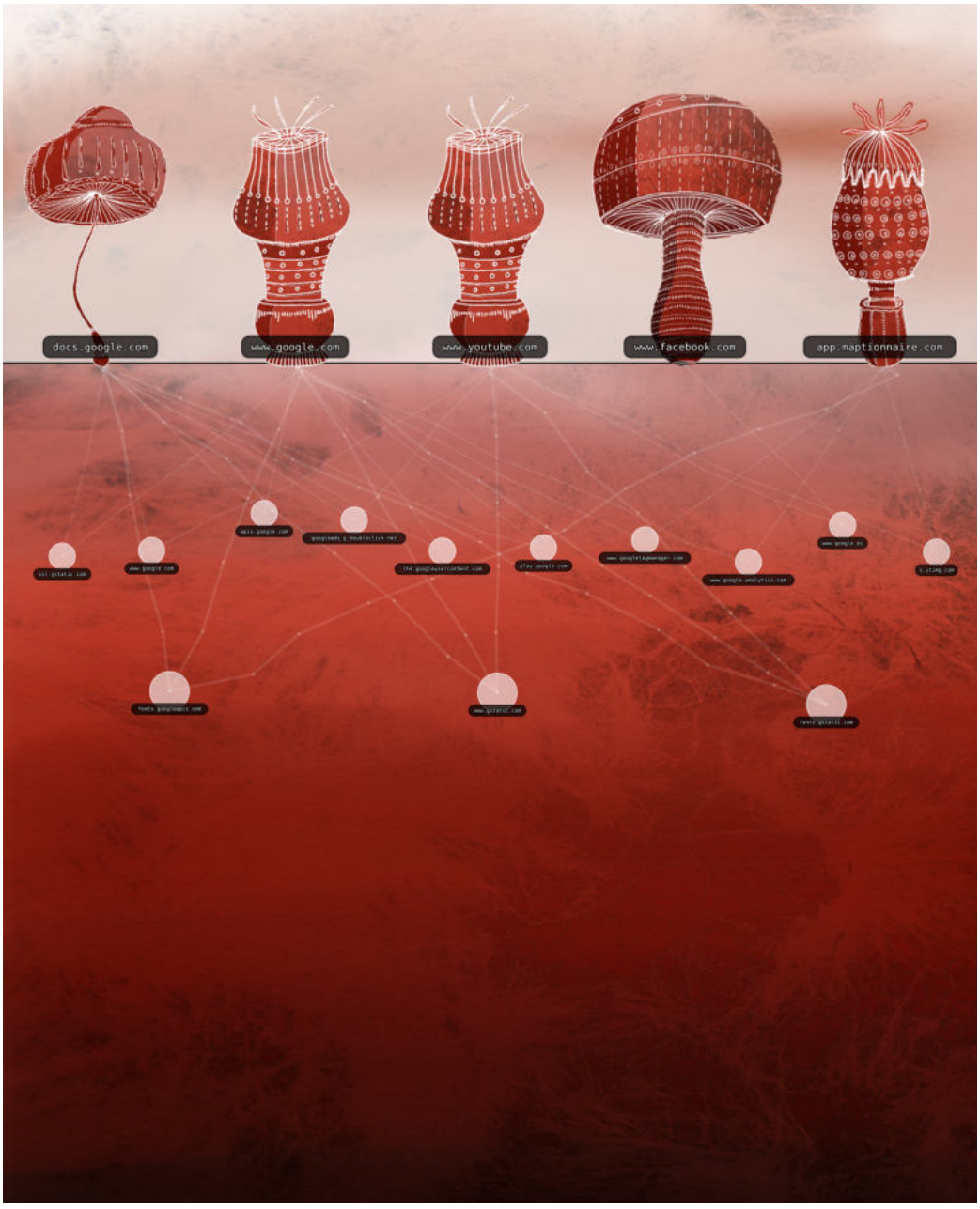

Figura 4: Ejemplo de visualización más limpia después de instalar Private Badger y uBlock.

su integridad, y cómo este puede reaccionar y prevenir estas situaciones. Un ejemplo es Ṕrotege tus Datos en Internetq́ue se compone de una serie de vídeos que muestran cómo configurar la privacidad en los distintos navegadores. Uno enfocado hacia menores es Ú decides en internet(www.tudecideseninternet.es), un proyecto constituido por guías para centros educativos, talleres para familias y niños de diferentes edades y en general una gran cantidad de material didáctico puesto a disposición de los ciudadanos.

7.2. Colectivos privados sin ánimo de lucro

También asociaciones y agrupaciones de personas sin ánimo de lucro han realizado diversas actuaciones en este sentido. A nivel global, la Free Software Foundation, la Free Software Foundation Europe, la Electronic Frontier Foundation, la Wikimedia Foundation o Guardian Project son ejemplos de entidades dedicadas a la protección de datos, a la conservación de la privacidad y a exponer defectos a la hora de implementar procedimientos.

La Electronic Frontier Foundation, en adelante EFF, tiene como propósito de existencia la defensa de los derechos civiles en el mundo digital. Esta consigna tan amplia se aborda desde diferentes puntos de vista que van desde proteger los derechos y las libertades de los usuarios hasta asegurar el avance de la tecnología en este sentido. EFF utiliza su voz independiente y libre para abrir el camino al software de código abierto, la encriptación, la búsqueda segura y las tecnologías para compartir contenido, todo ello libre de terceras partes.

La perspectiva de la Free Software Foundation es promover la libertad del usuario para que tenga control sobre la tecnología que utiliza día a día sin que una mayor dependencia a esta suponga una mayor monitorización o restricción de su comportamiento.

A nivel más local, existen distintas agrupaciones de usuarios (como GPUL o GALPon en Galicia) y otros organismos como las Oficinas de Software Libre de algunas universidades españolas, como la de la Universidad de Granada, que colaboran en favor de mejorar la situación e informar a los ciudadanos.

En particular, el esfuerzo de la Electronic Frontier Foundation y de voluntarios ha dado como resultado herramientas como Privacy Badger.

Block Origin y uMatrix de Raymond Hill son otros buenos ejemplos de proyectos extremadamente útiles que funcionan en favor de la privacidad de las personas. Estas herramientas permiten al usuario no sólo conocer las terceras partes que entran en juego secretamente cada vez que se conecta a una web si no que le otorgan al individuo la capacidad de bloquearlas selectivamente.

uBO-Scope es otro plugin que proporciona una métrica de "exposición" a sitios de terceros con los que también mostrar al usuario el rastro digital que deja su navegación, de una forma más sucinta.

Además, otras herramientas como Decentraleyes nos permiten no solo mejorar el rendimiento y disminuir el tráfico de red, sino incluso evitar conexiones a redes de distribución de contenido de recursos que hayan sido licenciados bajo licencias permisivas (y que, por tanto, puedan incluirse en el plugin: componentes como jQuery).

7.3. Colectivos privados con ánimo de lucro

Empresas cuyos modelos de negocio se mueven alrededor de la privacidad de los datos personales de individuos también llevan realizando diversas labores, tanto de concienciación como de exploración de posibles áreas de negocio. Por un lado, tal vez una de las extensiones más conocidas al respecto, Ghostery, es propiedad de una de estas empresas. Hasta hace poco, era un producto de Evidon, una empresa dedicada a la trazabilidad de los datos en el mundo digital.[5] Recientemente, el equipo de Ghostery y su propiedad intelectual han sido adquiridos por la empresa alemana Cliqz GmbH.[6] Ésta, además de Ghostery, también comercializa un navegador (también llamado Cliqz) en el que componentes relativos a los datos que son cedidos a terceros también son mucho más preeminentemente visibles y aparentemente sencillos de controlar.

Otros actores, como Proton Technologies AG proporcionan servicios como ProtonMail y ProtonVPN que tratan de hacer frente a los servicios proporcionados de forma gratuita a usuarios a cambio de la explotación de datos personales, a cambio de utilizar otro modelo de negocio: el de pago por contraprestación de un servicio. Este modelo permite a empresas como Proton Technologies mantener la calidad de sus servicios sin

[5]https://www.evidon.com/company/about-evidon/

[6]https://cliqz.com/en/about

necesidad de invadir la privacidad de sus usuarios ni de solicitar de ellos tratamientos de sus datos diferentes a los imprescindibles para proporcionar el servicio que se les solicita.

7.4. Activismo y actividades de concienciación

Una iniciativa nacida en el núcleo del periodismo es 'Personaldata',[7] una plataforma desarrollada con el fin de generar conciencia en los usuarios para que ejerzan sus derechos sobre sus datos personales, con mínimo esfuerzo. Su consigna es 'Your data, Your control, Your options'. Ellos se proponen como intermediarios entre las compañías y los usuarios para facilitar de manera legal que estos puedan reclamar una copia de sus datos, y eliminar o corregir los datos que cualquier empresa pueda almacenar sobre el individuo. Entre sus campañas más relevantes se encuentran las relacionadas con Tinder, Uber y Facebook.

'Europa contra Facebook'[8] es el nombre de un grupo de activistas liderados por Max Shrems, quienes luchan por una mayor transparencia en dicha red social. Su principal objetivo es reglamentar la cantidad de información que Facebook almacena de sus usuarios, así como el uso que se hace de la misma. Denuncian lo vagamente definidas y contradictorias que son las políticas de privacidad de esta empresa y lo opacos que son los fines para los que se recaban los datos y los usos que se hacen de estos.

El grupo de trabajo del Artículo 29 Ese interés por la privacidad ha quedado reflejado en diversos documentos hechos públicos como las Opiniones del Grupo de Trabajo del artículo 29, en adelante GT29. Este Grupo, creado por la Directiva 95/46/CE, es un órgano consultivo independiente integrado por las Autoridades de Protección de Datos de todos los Estados miembros, el Supervisor Europeo de Protección de Datos y la Comisión Europea. «Las funciones del GT29 reconocidas por la Directiva incluyen estudiar toda cuestión relativa a la aplicación de las disposiciones nacionales tomadas para la aplicación de la Directiva, emitir dictámenes sobre el nivel de protección existente dentro de la Comunidad y en países terceros, asesorar a la Comisión sobre cualquier proyecto de modificación de la Directiva, y formular recomendaciones so-

[7] https://personaldata.io/

[8] http://www.europe-v-facebook.org/ES/es.html

bre cualquier asunto relacionado con la protección de datos en la Unión Europea. El GT 29 se pronuncia a través de Dictámenes, Documentos de Trabajo, Informes o Recomendaciones, aunque también manifiesta su posición en cartas o comunicados de prensa. Las decisiones del Grupo no son jurídicamente vinculantes, pero tienen un importante valor doctrinal y son frecuentemente utilizados y citados por los legisladores y los tribunales nacionales y europeos.»

8. Conclusiones

Trackula es un proyecto que, a partir del taller Visualizar'17 de Medialab-Prado (http://medialab-prado.es/article/visualizar17-migraciones-proyectos-seleccionados) ha seguido creciendo y cuyo propósito es mejorar la concienciación del usuario medio respecto a la cesión de datos personales. El Dictamen 16/2011 del Grupo de Trabajo del Artículo 29 respecto a la concienciación del usuario medio deja entrever la preocupación del grupo por este tipo de cuestiones que deberían de ser mejor trasladadas a la sociedad en general.

Con el sitio web creado y la información elaborada, un individuo poco informado podría tener un mayor interés en obtener información adicional. Si así fuere, el plugin que se ha elaborado y que está disponible para la descarga, permitirá al individuo observar su interacción con terceros respecto a sus hábitos de navegación.

En un futuro, integrar esta visión con la proporcionada por Mozilla Lightbeam podría permitir a los interesados obtener información más detallada de la que ahora mismo el plugin puede proporcionar, y sería un añadido relevante.

Este proyecto se centra en la cesión inadvertida de datos a través de la navegación en la web, pero indudablemente esto constituye una pequeña fracción de las posibles formas de ceder datos personales y de la creación de perfiles. En particular, no se han abordado cuestiones relacionadas con aplicaciones de ordenador o móviles, ni con otras formas de rastreo de personas potencialmente identificables como las utilizadas por los diversos mecanismos colectivamente conocidos como "Internet de las Cosas", como podrían ser sensores utilizados en Ciudades Inteligentes o vallas de publicidad con detección de presencia y de las características biométricas de su público.

Por otro lado, parece bastante evidente en base al trabajo relacionado, que la protección de datos personales en un ambiente de alta tecnología en la sociedad de la información y de la explotación de datos no puede entenderse de forma aislada y que la cooperación con la tecnología que habilita tales cesiones y tratamientos ha de ser entendida como un conjunto. De este modo, podría verse a la protección de datos personales como una parte más de un interés general en llegar hacia una mayor soberanía e independencia tecnológicas, no solo respecto a la transmisión de información y de elección de protocolos estándar, sino también a la hora de que los interesados tengan una libertad de elección respecto a las formas en las que sus datos sean tratados y a no alimentar modelos de negocio que sean contrarios incluso a sus intereses particulares. En este respecto, parece que las visiones ofrecidas por colectivos afines a los movimientos a favor del Software Libre y de la protección de datos personales encajan perfectamente en cuanto a los objetivos de esta soberanía tecnológica.

Otra cuestión que queda por explorar es la capacidad de incluir enlaces a las políticas de privacidad de distintos sitios web, e incluso promover un estándar para poder mecanizar la visualización de las partes más relevantes de las políticas de privacidad de los distintos agentes involucrados en el tratamiento de datos.

En resumen, creemos que Trackula aporta una forma original y novedosa de exponer las relaciones entre terceras partes al visitar sitios web y que ésta puede proporcionar una forma efectiva de comunicarse con personas que todavía no tengan una opinión formada al respecto de la privacidad de sus datos personales para que puedan tomar decisiones mejor informadas sobre la cesión de sus datos a diversas páginas web y a las interacciones que en ellas ocurren.

A. Enlaces de interés

Debido a que este proyecto tiene un carácter eminentemente práctico, todo el trabajo se encuentra disponible en la web. A continuación se reúnen los distintos enlaces desde los que acceder a cada una de las facetas desarrolladas.

Web https://trackula.org

Fuentes de la web https://github.com/medialab-prado/trackula

Extensión de Firefox para su instalación https://addons.mozilla.org/en-US/firefox/addon/trackula/versions/[9]

Fuentes de la extensión de Firefox https://github.com/ssaavedra/trackula-we

[9] La última versión disponible en el momento de edición de este documento es la 0.1.3.