



OPINIÓN LEGAL SOBRE CONSULTA PARA DETERMINAR SI LA TÉCNICA DE ESCRAPEADO Y, EN SU CASO, LA DIFUSIÓN DE LOS DATOS ASÍ OBTENIDOS PARA FINES ESTADÍSTICOS, PUDIERA VULNERAR EL ORDENAMIENTO JURÍDICO.

REF.: 15/ 2020 DDLCN-OL

I. INTRODUCCIÓN

El día 17 de febrero de 2020 se nos traslada por el Director General del EUSTAT/Instituto Vasco de Estadística, solicitud de consulta legal sobre la cuestión reseñada en el encabezamiento.

Junto con la petición se incluye la siguiente documentación:

- Carta explicativa del Director General del EUSTAT en la que se detallan los antecedentes de la consulta y las circunstancias más relevantes para contextualizar la petición de opinión legal sobre si la técnica de escrapeado (raspado) y, en su caso, la difusión de los datos así obtenidos para fines estadísticos, pudiera vulnerar el ordenamiento jurídico vigente.

Se emite la presente opinión jurídica en virtud de lo dispuesto en el artículo 5.1.b) de la Ley 7/2016, de 2 de junio, de Ordenación del Servicio Jurídico del Gobierno Vasco y en el artículo 7.1.b) y 7.3 del Decreto 144/2017, de 25 de abril, del Servicio Jurídico del Gobierno Vasco, en relación, ambos, con el artículo 6.1 h) del Decreto 24/2016, de 26 de noviembre, del Lehendakari, de creación, supresión y modificación de los Departamentos de la Administración de la Comunidad Autónoma del País Vasco y de determinación de funciones y áreas de actuación de los mismos, y con el artículo 12.1.a) y 14.1 del Decreto 71/2017, de 11 de abril, por el que se establece la estructura orgánica y funcional del Departamento de Gobernanza Pública y Autogobierno.

II. ANTECEDENTES Y OBJETO DE LA CONSULTA.

El Eustat elabora desde el año 1992 la “**Encuesta de Establecimientos Turísticos Receptores**”. Dicha encuesta ofrece **información estadística** relativa a los establecimientos hoteleros de todo tipo y con los objetivos específicos que se indican en la carta reseñada.

Por otro lado, a comienzos del año 2014 y liderados por Eurostat (Oficina Europea de Estadística), los Institutos de Estadística Oficial de los Estados de la Unión Europea empezaron a valorar la posibilidad de calcular estadísticas oficiales a partir de fuentes de “Big Data”; bien para complementar; bien para sustituir, la forma clásica de su cálculo basada en encuestas.



En este contexto, por “Big Data” se entiende toda **fuentes de datos no creada por el propio Instituto de Estadística**. La idea subyacente no es otra que el de modernizar el área de la estadística oficial a partir de la irrupción y creciente protagonismo social de estas novedosas fuentes de datos.

Para tal fin, Eustat puso en marcha en el año 2017 un grupo de trabajo para explorar las posibilidades del Big Data en las **operaciones estadísticas de su ámbito de responsabilidad**. Y así, entre las alternativas que se consideraron, se decidió que la opción más recomendable para llevar a cabo la citada encuesta de **Establecimientos Turísticos Receptores** era la de obtener la tarifa media mensual por el procedimiento de “escrapeado” **de alguna plataforma web de reserva de habitaciones**. El objetivo no es otro, y así se nos indica, que el de **obtener de manera eficiente y automatizada los precios de los alojamientos turísticos de Euskadi a efectos meramente estadísticos**.

Es tal sentido, una vez que se establecieron los criterios, la metodología y la infraestructura estadística necesaria, por el EUSTAT **se contrató los servicios de una empresa externa** para que llevara a cabo exclusivamente la labor informática de escrapeo y el almacenamiento temporal de la información.

Se nos informa igualmente que la difusión de los resultados de los precios ADR **en ningún caso afecta a registros individuales de establecimientos hoteleros**, que se tratan siempre de manera agregada; bien a nivel de categoría, bien a nivel de zona geográfica. Y que existen también otras variables que forman parte del modelo matemático como son el mes y el grado de ocupación.

Por lo tanto, **la información obtenida y su tratamiento, en función de lo que se nos transmite, no permite singularizar a ninguna empresa en particular ni identificar la información como específica, singular y propia de una sola fuente**.

III. WEB SCRAPING: CONCEPTO, USOS Y SU VALOR AÑADIDO.

Es indudable que el desarrollo de las nuevas tecnologías está suponiendo una revolución absoluta en el campo de la creación de contenidos. En tal sentido, se puede decir que cada usuario de las tecnologías se convierte, consciente o inconscientemente, en una fuente constante de información y por múltiples motivos.

Dichos avances están propiciando el que cada vez haya más y más datos de acceso público y de un innegable interés, esto es, un potencial de información a explorar para distintos fines.

Datos que se pueden analizar y de donde, obviamente, se pueden sacar conclusiones que apoyen o lleguen a convertirse en elementos importantes de las estrategias empresariales e institucionales en el mundo actual, y que pueden dar a las empresas y a las organizaciones en general un valor añadido en su desarrollo empresarial e institucional.

En ese sentido, **una de las técnicas con más peso en este proceso de obtención de datos es la conocida como web scraping.**

El Web Scraping es el conjunto de acciones ejecutadas por programas de software dirigidas a la extracción de grandes cantidades de datos de sitios web. Gracias a estos programas se puede automatizar la obtención de información, y hacerlo de un modo rápido, seguro y fiable.

El web scraping utiliza para ello sus spiders para extraer la información, y almacenarla en el formato que se precise para su posterior análisis.

Los spiders son robots automáticos que utilizan la mayoría de los grandes buscadores internacionales y conocidos para indexar los sitios en sus motores de búsqueda (Esto es, ordenar una serie de datos o informaciones de acuerdo a un criterio común a todos ellos, para facilitar su consulta y análisis). Los mismos realizan las siguientes acciones: **Recorren las páginas recopilando información sobre los contenidos de éstas.**

La verdadera utilidad o valor añadido del web scraping proviene sin duda de la unión de los datos extraídos, **en muchas ocasiones de varias y distintas fuentes**, el posterior análisis y la obtención final de conclusiones. Unas conclusiones que son la verdadera inteligencia de cara a cualquier estrategia de negocio, al desarrollo de políticas públicas, a los esfuerzos en marketing, al desarrollo de nuevos productos al mercado **o simplemente a la transmisión de datos a efectos estadísticos, como es el presente caso.**

Así, algunos otros usos del web scraping son la comparación de precios en tiendas y establecimientos, la monitorización de datos relacionados con el clima de cierta región, la detección de cambios en sitios webs o la integración de datos en sitios webs.

El uso, cada vez más extendido de la web scraping, abarca a muchos y diversos ámbitos, así como por ejemplo el de los Recursos Humanos (selección), donde es utilizada para analizar perfiles en redes sociales y así poder dar con los candidatos idóneos para los puestos disponibles.

IV.- OPINIÓN LEGAL

Adentrándonos ya en el tema de la consulta, esto es, en la legalidad de la técnica del web scraping en relación con los fines para los que es usada por parte del EUSTAT, de entrada, diremos que **su uso es legal y no contraviene el ordenamiento jurídico, si bien, con sus matices**, en los que nos adentraremos a continuación.

En tal sentido, es importante que entendamos que el problema no está tanto en el uso de los crawlers (rastreadores) para rastrear la red, **como en el uso que podemos dar de los datos conseguidos. Y ahí está la clave de los límites legales. Es decir, resulta necesario**

comprobar la legitimidad del tratamiento de la información que aparece en las bases de datos objeto de uso a fines estadísticos.

De tal manera que, a mi juicio, la técnica del scraping **no sería aceptable en su uso en situaciones tales como las tres siguientes:**

1. A nivel de competencia, si con el scrapeo incurrimos en **competencia desleal** para con la web scrapeada, generando confusión en el consumidor. **Este supuesto no nos afectaría.**
2. Si estamos incurriendo en un delito desde el punto de vista de la propiedad intelectual. **Tampoco estaríamos afectados puesto que en general hablamos de información de datos menores y no sometidos a copyright.**
3. Dentro del marco de la **Ley Orgánica 3/2018 de Protección de Datos (LOPD)**, cuando **accedemos a datos personales de terceros sobre los que no se tiene consentimiento para su almacenamiento o tratamiento. Este supuesto merece ser matizado.**

En efecto, en relación a este apartado, tal y como ya indicó la **Agencia Española de Protección de Datos en su Informe 342/ 2008**, las páginas web no podrán ser consideradas en ningún caso fuentes accesibles al público, al no figurar en el listado exhaustivo y limitativo del artículo 7 del Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de Desarrollo de la LOPD.

Esto significa que, aunque los datos personales se encuentren publicados en websites, de tal forma que cualquiera pueda acceder a los mismos, **esto no significa que puedan ser utilizados de forma indiscriminada para cualquier finalidad, yendo contra los intereses del propietario original y sin contar con el consentimiento previo de los titulares de los datos.**

Por otro lado, **para poder utilizar la información contenida en las mismas debe de justificarse la finalidad**, principio esencial en materia de protección de datos que se contempla en el **artículo 4 de la Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales (LOPD)**, la cual viene a decir que **los datos de carácter personal sólo se podrán recoger para su tratamiento, cuando sean adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades determinadas explícitas y legítimas para las que se haya obtenido.**

En tal sentido, la citada Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, establece lo siguiente:

“Artículo 8. Tratamiento de datos por obligación legal, interés público o ejercicio de poderes públicos.



2. El tratamiento **de datos personales** solo podrá considerarse fundado en el cumplimiento de una **misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable**, en los términos previstos en el artículo 6.1 e) del Reglamento (UE) 2016/679, cuando derive de una competencia atribuida por una norma con rango de ley.”

“Artículo 25. Tratamiento de datos en el ámbito de la función estadística pública.

1. El tratamiento de datos personales llevado a cabo por los organismos que tengan atribuidas las competencias relacionadas **con el ejercicio de la función estadística pública** se someterá a lo dispuesto en su **legislación específica, así como en el Reglamento (UE) 2016/679 y en la presente ley orgánica.**

2. La comunicación de los datos a los órganos competentes en materia estadística solo se entenderá amparada en el artículo 6.1 e) del Reglamento (UE) 2016/679 en los casos en que la estadística para la que se requiera la información venga exigida por una norma de Derecho de la Unión Europea o se encuentre incluida en los instrumentos de programación estadística legalmente previstos.”

De conformidad con lo dispuesto en el artículo 11.2 de la Ley 12/1989, de 9 de mayo, de la Función Estadística Pública, serán de aportación estrictamente voluntaria y, en consecuencia, solo podrán recogerse previo consentimiento expreso de los afectados los datos a los que se refieren los artículos 9 y 10 del Reglamento (UE) 2016/679.

3. Los organismos competentes para el ejercicio de la función estadística pública podrán denegar las solicitudes de ejercicio por los afectados de los derechos establecidos en los artículos 15 a 22 del Reglamento (UE) 2016/679 cuando los datos se encuentren amparados por las garantías del secreto estadístico previstas en la legislación estatal o autonómica.

El artículo 6.1.e) de la citado Reglamento 2016/679 establece que el tratamiento será lícito cuando sea necesario **“para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento”**

La Ley 12/1989, de 9 de mayo, de la Función Estadística Pública, señala al respecto lo siguiente:

“Artículo 10.

2. *Todas las personas físicas y jurídicas que suministren datos, tanto si su colaboración es obligatoria como voluntaria, deben contestar de forma veraz, exacta, completa y dentro del plazo a las preguntas ordenadas en la debida forma por parte de los servicios estadísticos.*

3. *La misma obligación incumbe a todas las instituciones y entidades públicas de la Administración del Estado, las Comunidades Autónomas y las Corporaciones Locales. Cuando para la realización de estadísticas sea precisa la utilización de datos obrantes en*

fuentes administrativas, los órganos, autoridades y funcionarios encargados de su custodia prestarán la más rápida y ágil colaboración a los servicios estadísticos.”

Es obvio que, si con el escraqueo estuvieran obteniendo datos personales, el escenario que plantea la consulta **no estaría contemplado en este supuesto del artículo 10**, dado que la obtención de datos se está haciendo en este caso sin previa petición ni consentimiento del emisor de los datos y, de alguna manera, deberíamos decir que **esta fuera del marco legal de lo que dicha disposición establece**.

Con respecto al País Vasco, partiremos de los límites que establece el artículo 5 de la **Ley 4/1986 de Estadística de la Comunidad Autónoma de Euskadi**. Y, por consiguiente, las estadísticas se pueden elaborar a partir de datos estadísticos y de datos administrativos, teniendo la consideración de datos estadísticos los obtenidos por el Eustat a su nivel operativo **para uso exclusivamente estadístico**.

Por otro lado, debemos también advertir que el web scraping pudiera ir en contra de los términos de uso de algunos sitios webs, desde otros puntos de vista. En tal sentido, el cumplimiento de estos términos no está totalmente claro. Mientras que la duplicación de expresiones originales puede ser en muchos casos ilegal, en Estados Unidos la Corte dictó en el caso Feist Publications v. Rural Telephone Service que la duplicación de hechos es permitida. Sin embargo, las Cortes de Estados Unidos en ciertas ocasiones han reconocido que ciertos usos de los scrapers no deberían estar permitidos. Podría considerarse una computadora como una propiedad personal y, de esta forma, el scraper estaría entrando sin autorización en esta propiedad. En el caso más conocido, eBay vs Bidder's Edge, la segunda empresa tuvo que parar de realizar peticiones automáticas al sitio de eBay. En este caso, Bidder's Edge pujaba automáticamente por ciertos productos en este sitio.

Una de las principales pruebas de scraping involucró a American Airlines y a una empresa llamada FareChase. American Airlines ganó esta batalla, haciendo que FareChase parara de vender un software que les permitía a los usuarios comparar tarifas en línea si el sitio de American Airlines era incluido. La aerolínea dijo que las búsquedas de FareChase entraban sin autorización en los servidores cuando recopilaban la información públicamente disponible.

Aunque las decisiones judiciales hasta ahora tomadas no son uniformes, es difícil ignorar que un patrón está emergiendo, en el cual podemos ver que los tribunales en general están preparándose para proteger el contenido propietario en sitios webs comerciales, previendo de esta forma que este sea utilizado sin el consentimiento de los propietarios de los sitios. Sin embargo, **el grado de protección de estos contenidos aún no está establecido, y dependerá del tipo de acceso realizado por los scrapers, de la cantidad de información recopilada, del uso que se haga de la misma y del grado en el que afecten estos factores al propietario del sitio web**.

No debemos olvidar el hecho cierto y constatado de que **cada vez hay más datos que son de acceso público, esto es, subidos por los propios usuarios, por lo que un uso inteligente y legal del web scraping ofrece sin duda una ventaja competitiva sustancial.**

Para garantizar las actividades de los miembros del Sistema Estadístico Europeo (Eurostat y los institutos nacionales de estadística), **entre las cuales se encuentra la obtención de datos**, se ha establecido una política de buenas prácticas, la cual fue adoptada por el Comité del Sistema Estadístico Europeo el 28 de septiembre de 2011.

Y así, los principios que deben regir dichas actividades son: transparencia, coherencia y ética, respetando la legislación nacional y comunitaria. Los datos web deben extraerse y utilizarse de una manera adecuada y ética, que limite la carga de los propietarios de sitios web y los encuestados en la mayor medida posible.

En tal sentido Eurostat establece que el escraqueo, definido como la extracción automática de datos disponibles en la Word Wide WEB, **es un método legítimo de obtención de datos que complementa la recogida de información por encuesta y de registros administrativos como fuente de información para la elaboración de las estadísticas europeas.**

Por otra parte, es importante resaltar que **la LOPD extiende su ámbito de aplicación exclusivamente, a los datos de carácter personal sometidos a tratamiento**; por lo tanto, a lo que sería la **información concerniente a personas físicas identificadas o identificables**, de tal manera que sólo las personas físicas son titulares del derecho fundamental a la protección de datos de carácter personal.

La protección que ofrece la normativa sobre protección de datos, con carácter general, no se extiende a las personas jurídicas ni a los ficheros que se limiten a incorporar los datos de las personas físicas que presten sus servicios en aquéllas, consistentes únicamente en nombre y apellidos, las funciones o puestos desempeñados, la dirección postal o electrónica, teléfono y número de fax profesionales.

Por lo tanto, quedarían excluidos de su ámbito de aplicación los datos relativos a empresarios individuales, cuando hagan referencia a ellos en su calidad de comerciantes, industriales o navieros; esto es, se vincula la exclusión a la utilización de los datos en el ámbito estrictamente empresarial.

V.- CONCLUSIÓN.

En definitiva, a la vista de la legislación reseñada, tanto europea como estatal y vasca, teniendo en cuenta los criterios de actuación que promueve el Eurostat en este tema, el uso de datos a efectos meramente estadísticos y, a la vista de las distintas resoluciones judiciales reseñadas sobre el asunto en cuestión, podemos concluir que, en relación a la consulta planteada, y en los términos en los que ésta ha sido realizada, nos lleva a entender que la forma de obtener los datos por parte del Eustat, mediante la web scraping, y su posterior tratamiento a efectos meramente estadísticos, es lícita y acorde con la normativa vigente.

No parece razonable dar la consideración jurídica a los datos así obtenidos de datos de carácter personal y, por tanto, el control de los mismos tiene más que ver con la política de protección (términos de uso) que establezca la propietaria de la página web de la que se pretende obtener la información por el método de web scraping.

Por último, y dado que en esta Comunidad contamos con la presencia institucional de la Agencia Vasca de Protección de Datos y con la figura, desde mayo de 2018, de la **Delegada de Protección de Datos de la Administración Pública de la Comunidad Autónoma de Euskadi**, el propio EUSTAT, si lo considera oportuno, puede elevar una consulta complementaria sobre el tema, dirigiéndose para ello a dicha Delegada, toda vez que entraría en el ámbito de sus funciones el emitir informes en materia de protección de datos de carácter personal.

Esta es mi opinión jurídica a la consulta planteada y que emito y someto de buen grado a cualquier otra mejor fundada en derecho.