

Máster Universitario: Análisis de Datos Masivos - UIB

Aspectos legales del análisis de datos

Jorge Morell Ramos

29 de Noviembre de 2017

¿Qué es Términos y Condiciones?



Prestamos servicios legales en nuevas tecnologías. Por ejemplo, comercio electrónico, protección de datos, marcas y la vertiente legal de cualquier proyecto tecnológico.

In English too!

¿Qué es Términos y Condiciones?



¡También tenemos un Legal LAB! Es el departamento de I+D del despacho

¿Qué es Términos y Condiciones?

El Boletín de Términos y Condiciones

Controlamos cuándo las Condiciones de Uso, las Políticas de Privacidad, los Avisos de Cookies u otros términos y condiciones cambian.

Poniendo fin al clásico de: *"Te recomendamos consultar estas condiciones de forma regular ya que pueden cambiar sin previo aviso"*.

Mediante tu lector de feeds RSS puedes estar al día de todos los cambios, o sólo los de una empresa o servicio en particular. Para ello sigue las migas de pan > > > en un cambio concreto.

[Más información »](#)

7127

Documentos

Condiciones de uso, privacidad y más

5230

Servicios

Grandes y pequeños

116

Cambios

Durante la última semana

Jorge Morell Ramos



Email: hola@terminosycondiciones.es

Web: www.terminosycondiciones.es

Twitter: [@jorge_morell](https://twitter.com/jorge_morell)

Story time

2010, **Pete Warden**, ingeniero de software, web scraping y Facebook

Story time

Su conclusión:

*“Big data? Cheap. Lawyers?
Not so much.”*

Índice

- ❖ **Scraping** y legalidad
- ❖ Propiedad intelectual y **data sets**
- ❖ Análisis de **datos personales**
 - ❖ Transparencia y nube
 - ❖ Consentimiento
 - ❖ Decisiones automatizadas
 - ❖ Origen de los datos
 - ❖ Conservación y reutilización
 - ❖ Anonimización y cifrado
- ❖ Otras cuestiones y **conclusiones**

Scraping

¿Es legal un web scraper?

Obviamente sí, no es más que un instrumento.
Lo que puede ser ilegal son sus usos

Scraping

En EE.UU., eBay en el año 2000

Caso Bidder's Edge: 100.000 accesos al día,
1,53% del tráfico diario

Scraping

En España, Ryanair en 2008-2009

Su campaña contra las agencias online o, como las denominó, *“los parásitos del sector”*

Scraping

¿Cómo nos podemos meter en un lío legal por hacer web scraping?

Scraping

Planteado desde la perspectiva del
“**scraper**”, pero se le puede dar la vuelta
fácilmente

Scraping

¿Qué dicen las **condiciones legales** del sitio que voy a scrapear?

Scraping

Cláusula habitual

“Se prohíbe el uso de cualquier sistema automatizado o software para extraer datos de este sitio web para fines comerciales.”

Scraping

¿Por tanto, usar un web scraper vulnera esas condiciones?

¿Debería responder por un incumplimiento contractual?

Scraping

Uso personal o no comercial

Debe verse el caso concreto, sobretodo en usos no comerciales, pero no debería dar problema

Scraping

Uso comercial

Definitivamente debe verse el caso concreto, ya que podría haber vulneración, **pero hay algunas pautas básicas...**

Scraping

Uso comercial

- ❖ Necesitas una “*aceptación*” de las condiciones
- ❖ El acceso y uso de la web por el bot no basta
- ❖ No equivale a un “*derecho de admisión digital*”

Scraping

¿Cómo puede el servicio protegerse y convertir el scraping en vulneración?

- ❖ Adoptar medidas tecnológicas apropiadas
- ❖ Exigir la aceptación de condiciones para acceder al sitio o los datos

Scraping

¿Hay una **base de datos** en el sitio web
scrapeado y / o puede la web ser considerada
como tal?

Scraping

Entonces puede que estemos
extrayendo algo que resulta una
creación intelectual protegible

Scraping

Por tanto, un scraping de una **base de datos protegida por la propiedad intelectual** sí puede suponer un **lío legal**

Scraping

En todo caso, luego habrá que ver el **carácter sustancial o no** de la utilización de los datos scrapeados

Scraping

Otros aspectos legales:

- ❖ Competencia desleal
- ❖ Datos de carácter personal
- ❖ Intrusión informática no autorizada
- ❖ Consumo de recursos
- ❖ Daños informáticos

Scraping - Conclusiones

- ❖ Importancia del **uso** que se hará de los datos scrapeados
- ❖ Los **términos y condiciones** deben leerse; muchas veces no supondrán incumplimiento contractual
- ❖ Se puede estar scrapeando una **base de datos** protegida por la propiedad intelectual
- ❖ Mucho ojo cuando se extraen y gestionan **datos de carácter personal**
- ❖ No conviene nada hacer web scraping sin autorización de **datos no públicos**
- ❖ El scraping no debe **sobrecargar técnicamente los servicios y equipos** de quien lo sufre

Bases de datos - Data sets

La **protección legal** de un conjunto de
datos estructurado

Bases de datos - Data sets

Una base de datos puede ser una **creación intelectual protegible**, ya sea por el derecho de autor o por el derecho “*sui generis*” de las BBDD

Bases de datos - Data sets

Es decir, **el dato en sí** (un número por ejemplo) muchas veces no está amparado por la propiedad intelectual

Bases de datos - Data sets

Cuestión distinta sería con **otro tipo de contenido**, como una fotografía, un vídeo o un determinado texto

Bases de datos - Data sets

Pero de los datos puede surgir una **estructura original** que haya implicado una inversión sustancial

Bases de datos - Data sets

Generando una BBDD que por los recursos invertidos para su **obtención, verificación o presentación** sea protegible

Bases de datos - Data sets

En todo caso, el criterio jurisprudencial es **bastante estricto** a la hora de entender que una base de datos es protegible legalmente

Bases de datos - Data sets

En el Tribunal Europeo de Justicia, caso
Football Dataco vs Yahoo UK

Bases de datos - Data sets

Además de lo dicho, también deberá tenerse en consideración el **carácter sustancial o no** de los datos scrapeados

Conclusiones - Data sets

- ❖ Un data set - base de datos puede ser objeto de protección vía la **propiedad intelectual**
- ❖ Bien el dato en sí o por la **estructura** y esfuerzo que ha llevado su creación
- ❖ El criterio para reconocer una base de datos legalmente es **estricto**
- ❖ Resulta *más fácil* legalmente argumentar que una BBDD no tiene protección que sí la tiene

Datos personales

Qué ocurre cuando manejamos datos
que contienen **información personal**
según la ley | Concepto

Datos personales

Sobretudo cuando la **LOPD** está a punto de desaparecer y en su lugar llega el **RGPD** (el 25 de mayo 2018)

Datos personales

Tratamientos masivos de datos personales
buscan una cosa: **el profiling** | Lo que genera
necesidades y obligaciones legales

Datos personales

Algunas cuestiones legales a valorar:

- ❖ Transparencia y nube
- ❖ Consentimiento
- ❖ Decisiones automatizadas
- ❖ Origen de los datos
- ❖ Conservación y reutilización
- ❖ Anonimización y cifrado

Datos personales

Transparencia y nube: información previa, clara y precisa, control de los datos, su localización física y seguridad

Datos personales

Consentimiento: libre, expreso y que pueda ser revocado,
ya no cabe consentimiento tácito | Hay excepciones

Datos personales

Consentimiento: interés legítimo, estadísticas, usos científicos y el principio de finalidad

Datos personales

Consentimiento: el problema del alcance del proyecto de análisis al inicio, análisis adicionales y el “*por si acaso*”

Datos personales

Decisiones automatizadas: transparencia,
consentimiento explícito, derecho a rechazarlo si
tiene consecuencias jurídicas

Cuando el software me da o no el préstamo

Datos personales

Decisiones automatizadas: principios de minimización, evaluaciones de impacto y conservación de los datos

Datos personales

Decisiones automatizadas: el papel del responsable y del encargado del tratamiento

Datos personales

Origen de los datos: externos o internos, los metadatos pueden facilitar la clasificación y garantizar cierta calidad

Datos personales

Origen de los datos: pero a más metadatos recogidos, mayor riesgo en materia de privacidad en el proyecto

Datos personales

Origen de los datos: entra en juego la privacidad desde el diseño y por defecto; además del principio de minimización

Datos personales

Conservación y reutilización: que sea lícita, por el tiempo legalmente correspondiente, actualizando los datos y no incompatible

Datos personales

Conservación y reutilización: análisis al inicio del proyecto, verificación periódica de los datos y que no sean excesivos

Datos personales

Conservación y reutilización: reidentificar datos previamente anonimizados que por tanto vuelven a afectar la privacidad

Datos personales

Anonimización y cifrado: convertir el dato en no personal de forma irreversible sería el objetivo

Datos personales

Anonimización y cifrado: por ej. anonimizar para una agregación de datos o cifrar para no hacerlos visibles a usuarios

Datos personales

Anonimización y cifrado: *Hash* como una de las opciones de anonimizar datos; revisar el carácter reversible y su trazabilidad

Otras cuestiones

- ❖ Licencias de **APIs** y librerías
- ❖ La “**caja negra**” de los algoritmos
- ❖ Valor, flujo y propiedad de **datos no personales**

Conclusiones

- ❖ Más “*sencillo*” legalmente trabajar con **datos no personales**
- ❖ La **anonimización no reversible** evita tener en cuenta normas sobre privacidad (pero reduce utilidad al análisis)
- ❖ El **scraping** no es un juego, puede tener efectos legales serios
- ❖ La propiedad de los **data sets** será cada vez mayor
- ❖ Cuando se trabaja con datos personales, sobretodo a partir de **2018**, los elementos legales a valorar se disparan

Jorge Morell Ramos



Email: hola@terminosycondiciones.es

Web: www.terminosycondiciones.es

Twitter: [@jorge_morell](https://twitter.com/jorge_morell)