



# GUÍA NORMATIVA PARA EL DESARROLLO DE INFRAESTRUCTURAS LINGÜÍSTICAS

Agustí Cerrillo Martínez

Raquel Xalabarder

Mayo 2017



Este estudio ha sido realizado con financiación del convenio de colaboración entre la Secretaría de Estado para la Sociedad de la Información y la Agenda Digital y la entidad pública empresarial Red.es para la ejecución del Plan de impulso de las Tecnologías del Lenguaje, que no comparten necesariamente los contenidos expresados en el mismo. Dichos contenidos son responsabilidad exclusiva de sus autores.

Reservados todos los derechos. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de las mismas.

## Índice

1	Introducción .....	5
2	El desarrollo de recursos lingüísticos .....	7
2.1	Los recursos públicos de interés lingüístico .....	8
2.1.1	¿Qué se puede hacer con los recursos públicos de interés lingüístico? .....	9
2.1.2	¿Qué son los datos abiertos de interés lingüístico? .....	12
2.1.3	¿Cómo se relacionan los recursos lingüísticos? .....	13
2.2	Los recursos privados de interés lingüístico.....	15
2.3	Regímenes que pueden restringir el uso de los recursos (públicos y privados) .....	16
2.3.1	La protección de la propiedad intelectual.....	16
2.3.2	¿Se pueden utilizar recursos lingüísticos que incorporen datos personales? .....	30
2.3.3	¿Existen límites al uso de los recursos lingüísticos?.....	34
3	El desarrollo de procesadores lingüísticos .....	36
3.1	La protección jurídica de las aplicaciones .....	36
3.2	¿Qué papel pueden tener las administraciones públicas en el desarrollo de procesadores del lenguaje? .....	38
3.3	¿Pueden las Administraciones públicas contratar el desarrollo de los procesadores? .....	40
3.4	¿Cómo se puede fomentar los procesadores del lenguaje en lenguas cooficiales? .....	41
3.5	¿Se pueden reutilizar los procesadores del lenguaje? .....	41
4	El resultado.....	42
4.1	¿Qué se puede hacer con los resultados del procesamiento del lenguaje? .....	42
4.2	Las licencias de propiedad intelectual.....	44
5	Referencias .....	45
6	Glosario de siglas y acrónimos .....	46
7	Referencias normativas .....	47



## GUÍA NORMATIVA PARA EL DESARROLLO DE INFRAESTRUCTURAS LINGÜÍSTICAS

Esta guía constituye la segunda entrega en el marco del proyecto “Análisis normativo para el desarrollo de infraestructuras lingüísticas y la apertura de recursos lingüísticos en el marco del Plan de Impulso a las Tecnologías del Lenguaje” impulsado por la Secretaría de Estado para la Sociedad de la Información y Agenda Digital en colaboración con Red.es.

De acuerdo con los requisitos definidos en el proyecto, esta guía está destinada a los gestores públicos responsables de la implementación del Plan de Impulso de las Tecnologías del lenguaje en la que se describen las cuestiones jurídicas que emergen en el desarrollo de infraestructuras lingüísticas (recursos, procesadores y campañas de evaluación) que pretende el Plan y exponen las opciones que el marco legal permite.

Esta guía normativa para el desarrollo de infraestructuras lingüísticas ha sido elaborado por el profesor Agustí Cerrillo i Martínez, catedrático de Derecho administrativo de la Universitat Oberta de Catalunya y la profesora Raquel Xalabarder, catedrática de propiedad intelectual de la Universitat Oberta de Catalunya.

## 1 Introducción

---

Las tecnologías del lenguaje son aquellas tecnologías utilizadas para la comprensión automática del lenguaje humano. Generalmente, bajo esta expresión se incluyen las tecnologías de procesamiento de lenguaje natural y las tecnologías de traducción automática.

Las tecnologías del lenguaje se basan en el uso de recursos lingüísticos a través de procesadores que llevan a cabo distintas operaciones. Para evaluar la calidad de los recursos lingüísticos y los procesadores se llevan a cabo campañas de evaluación que permiten conocer el grado de calidad y de precisión de los resultados obtenidos. Los recursos lingüísticos, los procesadores y las campañas de evaluación conforman lo que ese conoce como infraestructuras lingüísticas.

El desarrollo de las infraestructuras lingüísticas genera diversas cuestiones desde el punto de vista jurídico a las que se debe dar respuesta para facilitar su desarrollo. Asimismo, la normativa también puede facilitar el impulso de las infraestructuras lingüísticas.

En las próximas páginas se describen las normas que deben ser tenidas en cuenta en el desarrollo de infraestructuras lingüísticas (recursos, procesadores y campañas de evaluación). Estas normas definen el marco en el que los desarrolladores y usuarios de las tecnologías lingüísticas pueden llevar a cabo su actividad. Asimismo, estas normas determinan el papel de los desarrolladores de infraestructuras lingüísticas en el marco del Plan de Impulso de las Tecnologías del Lenguaje.

**El desarrollo de las infraestructuras lingüísticas genera diversas cuestiones desde el punto de vista jurídico**

A lo largo de las próximas páginas se analizarán los aspectos jurídicos que deben tenerse en cuenta a fin de poder desarrollar las infraestructuras lingüísticas, impulsar la industria de las tecnologías del lenguaje y fortalecer el papel de la Administración pública como impulsora de la industria del lenguaje.

Para ello, se identificarán, sistematizarán y analizarán las distintas normas que de manera directa o indirecta inciden en el desarrollo de las infraestructuras lingüísticas.

La complejidad de las infraestructuras lingüísticas determina la diversidad de normas jurídicas que deben ser tenidas en cuenta a fin de conocer los diferentes desarrollos así como la variedad de



instrumentos necesarios para poder conseguir los objetivos del Plan de Impulso de las Tecnologías del Lenguaje.

Si bien las distintas normas inciden en el conjunto de las infraestructuras lingüísticas, a efectos de sistematización, este informe está organizado en tres partes en las que, respectivamente, se analiza la regulación que afecta a los recursos, los procesos y los resultados. En relación a cada una de estas partes se estudia en particular el papel de la Administración pública.

## 2 El desarrollo de recursos lingüísticos

---

Cualquier procesamiento del lenguaje se basa en el uso de recursos diversas procedencias. En particular, las tecnologías de procesamiento de lenguaje requieren el uso de grandes volúmenes de recursos lingüísticos como corpus de textos (textos científicos, documentos traducidos y memorias de traducción, contratos, sentencias, patentes, textos anotados, etc.), listas de nombres (de personas, organizaciones, marcas, topónimos, etc.), taxonomías y clasificaciones o terminologías y diccionarios.

Asimismo, los recursos no solo pueden estar en formato de texto sino que eventualmente también se pueden utilizar como recursos vídeos, imágenes o fotografías respecto a las que serán de aplicación las mismas consideraciones que se hagan a continuación.

Las tecnologías del lenguaje se basan en el uso y la reutilización de información de muy distinto origen y tipología. Con frecuencia, los recursos utilizados no están en posesión o son de titularidad del organismo o entidad que los procesa.

Estos recursos deben estar a disposición de desarrolladores no sólo en unos formatos técnicos sino también en unas condiciones jurídicas que faciliten su utilización. Ello implica la necesidad de tener en cuenta un amplio abanico de normas para poder determinar si un determinado recurso lingüístico puede ser utilizado por las tecnologías del lenguaje, con que finalidad y bajo qué condiciones.

**La titularidad de los recursos lingüísticos incide en el régimen jurídico aplicable**

Las Administraciones públicas y los organismos y entidades que conforman el sector público son grandes productoras de datos de gran utilidad en el procesamiento del lenguaje. Asimismo, las entidades privadas, con o sin ánimo de lucro, también ponen a disposición de desarrolladores y usuarios recursos necesarios para su actividad.

A la hora de valorar si se pueden utilizar determinados recursos, en la mayoría de los casos se deberá responder a cuestiones relativas a la propiedad intelectual que eventualmente puedan existir. Asimismo, también se deberá tener presente la regulación de la reutilización de la información pública y de la apertura de datos públicos cuando la información tenga su origen en las Administraciones públicas. Finalmente, en ocasiones será necesario tener en cuenta las normas

previstas para proteger otros bienes y derechos que puedan verse afectados por el uso de determinados recursos y, en particular, la protección de los datos personales y la confidencialidad de la información prevista para proteger cuestiones como el secreto comercial o industrial.

La regulación vigente determina el tipo de acceso que se puede realizar a los distintos recursos así como los usos que se pueden hacer. De este modo, en determinadas ocasiones, el acceso es libre para cualquier desarrollador y para llevar a cabo cualquier uso. En otras ocasiones, si bien se facilita el acceso a los recursos por parte de cualquier usuario, se limitan los usos que se puede realizar a determinadas condiciones. Finalmente, ante determinadas circunstancias, el acceso a los recursos puede estar limitado a un sujeto y los usos que pueda hacer de dichos recursos también puede ser ampliamente limitado ante la existencia de acuerdos de confidencialidad o de preceptos que persigan proteger otros intereses.

RECURSOS PÚBLICOS	RECURSOS PRIVADOS
Reutilización de información pública Datos abiertos Interoperabilidad de datos públicos	Recursos en formato digital (texto, imagen, sonoro, audiovisual): abiertos en Internet, licenciados (p.ej. bases de datos) Recursos en formato analógico: escaneado de libros, documentos impresos, grabaciones sonoras o audiovisuales, etc
REGÍMENES QUE PUEDEN RESTRINGIR EL USO DE LOS RECURSOS	
Protección de la propiedad intelectual Protección de los datos personales Confidencialidad de la información	

## 2.1 Los recursos públicos de interés lingüístico

Las Administraciones públicas generan grandes volúmenes de información en el desarrollo de su actividad que constituyen un recurso de gran utilidad para la industria del procesamiento del lenguaje. En efecto, las Administraciones públicas disponen de extensos corpus (sentencias, historias clínicas, traducciones de documentos oficiales, etc.), glosarios o listados de personas y empresas, marcas u organizaciones para el ejercicio de las funciones públicas que tienen atribuidas.

Sin embargo, es necesario que esta información sea puesta a disposición de todos aquellos que puedan requerirlo en unas condiciones y unos formatos que faciliten su utilización.





Como reconoce el Plan de Impulso de las Tecnologías del Lenguaje, “el gran valor potencial que tiene buena parte de la información que genera el sector público como recurso lingüístico representa una extraordinaria oportunidad para el desarrollo de la industria del procesamiento del lenguaje natural.”

A través de la apertura de estos datos y el fomento de su reutilización, las Administraciones públicas pueden poner a disposición de desarrolladores de las tecnologías de procesamiento del lenguaje, ingentes cantidades de recursos que pueden ser de gran utilidad. Sin embargo, para que los recursos públicos puedan ser reutilizados es necesario que se tenga en cuenta si pueden verse afectados otros derechos o bienes que el ordenamiento jurídico prevé que deben ser protegidos como los datos personales o la confidencialidad de los secretos comerciales. Asimismo, es necesario garantizar la interoperabilidad entre los recursos a fin y efecto de poder extraer el máximo provecho posible del procesamiento del lenguaje.

### *2.1.1 ¿Qué se puede hacer con los recursos públicos de interés lingüístico?*

La información que producen o que está en poder de las administraciones públicas en el ejercicio de sus funciones puede constituir un recurso básico utilizado por los desarrolladores de las tecnologías del lenguaje con el fin de ofrecer bienes y servicios al mercado. Por ello, el Plan de Impulso de las Tecnologías del Lenguaje se propone entre otros objetivos asegurar la disponibilidad pública gratuita o a bajo coste de infraestructuras lingüísticas de calidad en español y lenguas cooficiales apoyando la creación y explotación de datos abiertos de interés lingüístico. Para ello, se deben adaptar los recursos a formatos reutilizables por procesadores lingüísticos y facilitar su interoperabilidad.

La reutilización consiste en el **uso de documentos que están en poder de las administraciones y organismos del sector público, por parte de personas físicas o jurídicas, con finalidades comerciales o no comerciales**, siempre que este uso no constituya una actividad administrativa pública, ni tampoco el intercambio de documentos entre administraciones públicas en el ejercicio de las funciones públicas que tengan atribuidas.

La reutilización de la información pública está regulada en la Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público (en adelante, LRISP). La LRISP es de aplicación a las administraciones y organismos del sector público.

Sin embargo, la lectura del ámbito subjetivo de aplicación de la LRISP permite observar que no es de aplicación a algunos organismos y entidades que están en posesión de importantes corpus o recursos que tienen o pueden tener una utilidad significativa en el desarrollo de la industria de la tecnología del lenguaje. A pesar de ello, **los organismos o entidades que están excluidos de la aplicación de la LRISP puedan facilitar la reutilización de la información que producen o está en su poder.**<sup>1</sup>

**Los documentos de las Administraciones y organismos del sector público serán reutilizables en los términos previstos en la LRISP**

Como punto de partida se debe observar que la LRISP recoge la **obligación para las Administraciones y organismos del sector público de autorizar la reutilización de los documentos**. Sin embargo, una lectura atenta de dicha norma permite observar ciertas limitaciones en el alcance de dicha obligación. En efecto, la pretendida obligación de las Administraciones públicas de facilitar la reutilización de la información que está en su poder puede quedar ciertamente limitada al hecho de que cada entidad pública finalmente facilite la reutilización de la información que está en su poder.<sup>2</sup>

De todos modos, en el ámbito de la Administración General del Estado debe tenerse presente que el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP, para el ámbito del sector público estatal afirma que los órganos de la Administración General del Estado y los demás organismos y entidades del sector público estatal autorizarán la reutilización de los documentos elaborados o custodiados por ellos y únicamente podrá denegarse motivadamente la reutilización de los documentos si concurre alguno de los supuestos establecidos en la LRISP.

La LRISP reconoce que pueden ser objeto de reutilización los documentos, es decir, toda información, sea cual sea su soporte y forma de expresión, elaborada o custodiada por las administraciones públicas y organismos públicos. De este modo, todos aquellos datos que hayan producido o estén en poder de las administraciones u organismos públicos son susceptibles de convertirse en recursos al servicio de los desarrolladores y usuarios de las tecnologías del lenguaje.<sup>3</sup>

Sin embargo, la LRISP excluye del régimen de reutilización previsto determinados documentos como, por ejemplo, aquellos que no sean accesibles, aquellos cuyo conocimiento requiera una legitimación especial, cuando existan derechos de propiedad intelectual o industrial por parte de terceros y los contenidos públicos creados, por ejemplo, por entidades de radiodifusión o entidades educativas y de investigación (con las salvedades que se realizarán posteriormente) así como producidos o conservados por determinadas instituciones culturales.<sup>4</sup>



Sin embargo, debe quedar claro que ello no quiere decir que dicha información no pueda ser reutilizada. Será necesario que se adopten previamente algunas decisiones como por ejemplo que el autor de una obra protegida por la propiedad intelectual o quien ostente los derechos de explotación autorice su reutilización o que el documento que contenga datos personales haya sido previamente sometido a un proceso de anonimización o que el reutilizador haya obtenido el consentimiento del titular de los datos para poder llevar a cabo dicho tratamiento tal y como se indicará posteriormente.

Cada Administración pública puede concretar cómo se facilita la reutilización de la información que está en su poder. En esta dirección se puede destacar el papel que pueden tener los portales de datos abiertos como canal para facilitar la reutilización de los datos en poder de las administraciones públicas.

Junto la creación de portales de datos, también puede ser de utilidad poner a disposición de desarrolladores de las tecnologías del lenguaje catálogos que integren la datos o recursos públicos reutilizable tal y como prevé la propia LRISP al disponer que en la Administración General del Estado se mantendrá un catálogo de información pública reutilizable que permita acceder, desde un único punto, a los distintos recursos de información pública reutilizable disponibles.<sup>5</sup>

En esta dirección no se debe olvidar que el Plan de Impulso de las Tecnologías del Lenguaje contempla que las Administraciones públicas pueden contribuir a mejorar el conocimiento profundo y actualizado de la información gestionada por los distintos sectores para la dirección de políticas públicas. Para ello, entre otras medidas, se prevé catalogar los recursos lingüísticos abiertos dentro del portal de datos abiertos introduciendo una experiencia de usuario avanzada y crear una plataforma común de procesamiento del lenguaje y de traducción automática para las Administraciones Públicas a través de la que se ponga a disposición recursos lingüísticos de propósito general y específicos de dominio.

Por otro lado, no se puede desconocer que la LRISP establece la posibilidad de que las Administraciones públicas puedan aplicar una tasa o un precio público por el suministro de documentos para su reutilización. Las tarifas que se dispongan por la reutilización deben corresponderse con los costes marginales en que se incurra para la reproducción, puesta a disposición y difusión de la información pero no los costes de producción de la información. Las tasas podrán ser distintas según si la reutilización tenga carácter comercial o no comercial. En cualquier caso, las Administraciones públicas deben informar adecuadamente de las tasas o precios públicos que eventualmente se fijen.<sup>6</sup>

En cualquier caso, en la determinación de las tasas o precios públicos que eventualmente se puedan establecer en relación a los recursos lingüísticos se debe tener en cuenta que el Plan de Impulso de las Tecnologías del Lenguaje ha fijado como objetivo “asegurar la disponibilidad pública gratuita o a bajo coste de los recursos lingüísticos generados a partir de la información del sector público”.

Finalmente, tanto las Administraciones públicas como los desarrolladores deberían informar adecuadamente a los usuarios de las condiciones a las que se somete la reutilización. En esta dirección, las Administraciones públicas y los desarrolladores deberían difundir los recursos junto con un aviso legal que informe de manera adecuada y transparente de dichas condiciones. El aviso legal no tiene un valor jurídico específico más allá del meramente informativo pero resulta un instrumento útil para dar transparencia y seguridad jurídica al proceso de reutilización de los recursos lingüísticos ante la complejidad que ha adquirido la regulación y las lagunas que con frecuencia puede tener en este ámbito. El aviso legal se debe adaptar a las condiciones específicas de cada tipología de recursos lingüísticos que se difundan.

### 2.1.2 ¿Qué son los datos abiertos de interés lingüístico?

Los datos abiertos son todos aquellos datos en poder de las administraciones públicas que se difunden en formatos que facilitan su reutilización.

De acuerdo con la *Sunlight Foundation* los datos abiertos son aquellos datos completos, primarios, oportunos, accesibles de forma física y electrónica, procesables electrónicamente, no discriminatorios, que usan estándares abiertos en la difusión, sin licencia, permanentes y sin costes.<sup>7</sup>

De este modo, los datos abiertos se caracterizan desde un punto de vista técnico por ser procesables electrónicamente y, desde el punto de vista jurídico, por la ausencia de restricciones que limiten su reutilización (propiedad intelectual, datos personales, etc.).

Las administraciones públicas deben facilitar sus documentos en cualquier formato o lengua preexistente, pero también procurarán, siempre que ello sea posible y apropiado, proporcionarlos en formato abierto y legible por máquina y conjuntamente con sus metadatos, con los niveles más elevados de precisión y desagregación cumpliendo en la medida de lo posible los estándares y

**Los datos abiertos se caracterizan por ser procesables electrónicamente y por la ausencia de restricciones que limiten su reutilización**

normas formales abiertas. De todos modos, es importante tener presente que si bien, las



Administraciones públicas deben tratar de difundir sus datos en formato abierto y promover que la puesta a disposición de los datos para su reutilización se realice por medios electrónicos y mediante plataforma multicanal, no tienen la obligación de difundir los datos en dicho formato más allá de lo que con carácter genérico se desprende de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno que dispone que cuando las administraciones públicas difunden datos de acuerdo con las obligaciones de transparencia debe ser difundida, preferiblemente, en formatos reutilizables.<sup>8</sup>

Uno de los actores que producen mayores cantidades de datos reutilizables por las tecnologías del lenguaje son las universidades y los centros de investigación. Respecto a estos datos, cabe tener presente que la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación dispone su difusión en acceso abierto. En particular, en esta norma se prevé que los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación impulsarán el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación, y establecerán sistemas que permitan conectarlos con iniciativas similares de ámbito nacional e internacional. En particular, la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadas o periódicas, serán hechos públicos en formato electrónico tan pronto como resulte posible, antes de doce meses desde la publicación. Esta versión electrónica debe difundirse a través de repositorios de acceso abierto.<sup>9</sup>

Finalmente, se puede observar que los datos abiertos no solo tienen su origen en las propias Administraciones públicas sino que estas también pueden abrir datos que tengan otro origen como, por ejemplo, contratos, convenios, acuerdos, conciertos o subvenciones. Así puede estar previsto en el contrato que se haya firmado entre la Administración pública y el proveedor de los datos que puede incluir una cláusula específica en este sentido.

### 2.1.3 ¿Cómo se relacionan los recursos lingüísticos?

Los recursos lingüísticos pueden tener orígenes diversos y pueden encontrarse en formatos diferentes. Para facilitar su reutilización es recomendable que los formatos que se utilicen sean interoperables.

La **interoperabilidad** es la capacidad de los sistemas de información, y por ende de los procedimientos a los que estos dan soporte, de compartir datos y posibilitar el

**La interoperabilidad tiene una importancia para facilitar la reutilización**



**intercambio de información y conocimiento entre ellos.** La interoperabilidad tiene diferentes manifestaciones en relación al desarrollo y uso de las infraestructuras lingüísticas. En esta dirección, el Plan de Impulso de las Tecnologías del Lenguaje dispone como objetivo vinculado a la línea relativa a los recursos lingüísticos de las Administraciones Públicas y política de Reutilización de la información, adoptar normas técnicas de interoperabilidad, una política de licencias adecuada, y mecanismos de protección de datos personales.

Para garantizar la interoperabilidad es necesario seguir las normas, criterios y recomendaciones que permitan a las aplicaciones utilizadas por las administraciones públicas intercambiar información y utilizar mutuamente la información intercambiada.

Desde un punto de vista general, el Esquema Nacional de Interoperabilidad aprobado por el Real Decreto 4/2010, de 8 de enero define los criterios y recomendaciones en materia de seguridad, conservación y normalización de la información, de los formatos y de las aplicaciones que deberán ser tenidos en cuenta por las Administraciones Públicas. El Esquema Nacional de Interoperabilidad es aplicable a los recursos lingüísticos generados por las Administraciones públicas.<sup>10</sup>

En desarrollo del Esquema Nacional de Interoperabilidad, la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información dispone el conjunto de pautas básicas para la reutilización de documentos y recursos de información elaborados o custodiados por el sector público y las condiciones comunes sobre selección, identificación, descripción, formato, condiciones de uso y puesta a disposición de los documentos y recursos de información elaborados o custodiados por el sector público. De este modo, los recursos que las Administraciones públicas pongan a disposición de los desarrolladores y usuarios de las tecnologías del lenguaje deben cumplir con estas condiciones.

La lectura de la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información permite identificar los siguientes elementos que deben tenerse en cuenta ante la reutilización de recursos lingüísticos por parte de las Administraciones públicas:<sup>11</sup>

- los recursos de mayor relevancia y potencial social y económico se considerarán prioritarios al objeto de seleccionar los documentos y recursos de información aptos para la reutilización;
- los recursos serán primarios, evitando las modificaciones o alteraciones de la información existente en la fuente;
- los recursos tendrán el nivel granular más fino posible, evitando agregaciones adicionales;



- los recursos tendrán asociada información estructurada que permita su procesamiento automatizado;
- los recursos estarán actualizados a sus últimas versiones y se indicará la fecha de última actualización, así como el periodo de la misma;
- los recursos estarán identificados mediante referencias únicas y unívocas, basadas en identificadores de recursos uniformes. La norma técnica define los requisitos que deben tenerse en cuenta para la construcción de los identificadores de recursos uniformes;
- los recursos se asociarán a metadatos mínimos;
- los recursos, los metadatos y los servicios asociados a los mismos utilizarán estándares abiertos, así como, en su caso y de forma complementaria, estándares que sean de uso generalizado por la ciudadanía de acuerdo con lo previsto en el artículo 11 del Real Decreto 4/2010, de 8 de enero y en la Norma técnica de interoperabilidad de catálogo de estándares, aprobada por Resolución de 3 de octubre de 2012, de la Secretaría de Estado de Administraciones Públicas;
- los recursos podrán ser puestos a disposición pública a través de una o varias distribuciones en varios formatos distintos, con el objeto de facilitar la reutilización a agentes con distintos perfiles;
- los recursos deben atender al principio de accesibilidad a la información y a los servicios por medios electrónicos en los términos establecidos por la normativa vigente;
- los recursos reutilizables puestos a disposición pública llevarán asociada la información necesaria que permita su interpretación.

## **2.2 Los recursos privados de interés lingüístico**

El desarrollo de infraestructuras lingüísticas y de proyectos de tecnologías del lenguaje no son posibles sin la utilización de múltiples y variados recursos, de naturaleza diversa (textual, visual, sonora, audiovisual, etc); recursos que, con frecuencia, serán de titularidad privada.

Estos recursos pueden provenir de fuentes distintas. Por ejemplo, pueden utilizarse textos provenientes de una base de datos a la cual se tiene acceso mediante suscripción (o cambio del pago de un precio) o gratuitamente pero mediante clave de acceso; pueden utilizarse textos, grabaciones (sonoras o audiovisuales) que se hallan disponibles en abierto en Internet; o incluso pueden utilizarse textos que han sido digitalizados a partir de ejemplares publicados (por ejemplo, el escaneado de libros y documentos en las colecciones de bibliotecas y archivos). Todos estos recursos



pueden ser de interés lingüístico y, en la medida que se hallen protegidos por el régimen de la propiedad intelectual, o contengan datos personales o sean de carácter confidencial, su uso a los efectos de tratamiento lingüístico se verá sometido a las restricciones que se puedan derivar de tales regímenes de protección y que serán objeto de estudio en los siguientes capítulos.

El simple hecho de que un recurso se halle disponible en abierto en Internet (páginas web de acceso público y no restringido), o se tenga acceso al mismo lícitamente (por haber abonado un precio de suscripción o realizado un acceso mediante clave) o se tenga la propiedad del soporte tangible que lo contiene (por ejemplo, un libro), no confiere automáticamente autorización para tratar estos recursos. Para ello, deberemos analizar tales regímenes para ver en qué medida pueden suponer una restricción a tal tratamiento, dependiendo de su naturaleza y fuente de acceso.

Por otra parte, no examinaremos aquí la protección de la propiedad industrial, en especial, de las patentes, marcas y signos distintivos, por cuanto -más allá de su posible incidencia en relación con la protección de la aplicación informática (software)- su incidencia en el tratamiento lingüístico es mínima.

## 2.3 Regímenes que pueden restringir el uso de los recursos (públicos y privados)

### 2.3.1 *La protección de la propiedad intelectual*

El régimen de la propiedad intelectual **afecta a las acciones de tecnologías del lenguaje en todas sus diferentes etapas**: tanto en el uso de recursos, como en la protección del procesador (programa de ordenador), y en la reutilización de los resultados.

El **régimen de la propiedad intelectual** confiere al **autor** un conjunto de **derechos** sobre su **obra**, por un **plazo** de protección determinado y sujeto a algunas **excepciones** o límites legalmente previstos, tales como la copia privada o el límite de copias técnicas.

**El régimen de la propiedad intelectual confiere al autor un conjunto de derechos sobre su obra, por un plazo de protección determinado**

La **normativa actualmente vigente** en España para regular el derecho de autor es el Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual (TRLPI).<sup>12</sup> Esta norma se ha visto altamente influenciada por diferentes convenios



internacionales de protección de la propiedad intelectual,<sup>13</sup> así como por diversas normas dictadas por la Unión Europea.<sup>14</sup>

Sin embargo, a pesar de que estas leyes nacionales han ido confluyendo en los últimos años, a través de los convenios internacionales (y en Europa, de las Directivas) es importantísimo tener presente que las soluciones que podemos derivar de la ley española (TRLPI) no necesariamente serían las mismas bajo otro derecho nacional aplicable.

### 2.3.1.1 Recursos que quedan sujetos al régimen de propiedad intelectual

A efectos del TRLPI, obra es toda **creación original expresada**.<sup>15</sup> Por lo tanto, se trata éste de un concepto jurídico indeterminado, que sólo puede apreciarse en cada caso concreto y a la vista de las circunstancias concretas que rodean la creación de la obra.

**Los conceptos de autor y de obra son los pilares fundamentales e indisociables del régimen de PI: no hay obra sin autor, ni autor sin obra**

Así pues, , el régimen de la propiedad intelectual se aplicará a todos aquellos recursos que sean “creación original”, con independencia de su origen: recursos generados por la Administración pública, por una empresa privada, o por personas ya sea a título individual o a través de plataformas colaborativas (por ejemplo, wikis); tampoco importa (a los efectos de determinar su condición o no de obra protegida) su forma de explotación o la manera como se ha accedido a ellos: tanto si los recursos utilizados están disponibles en un web abierto al público o se han obtenido a través de una licencia de acceso a una base de datos (de acceso restringido), como si se han digitalizado de formatos impresos o analógicos. En tanto el recurso sea una creación original, con independencia de su origen y fuente, será de aplicación el TRLPI.

No importa el mérito, ni la finalidad, ni el grado de elaboración de una obra; solo la originalidad – y no es fácil definirla. En todo caso, incluso cuando el margen para la creación es escaso (obras de carácter factico, como mapas, u obras breves) **es necesario que haya un mínimo (un *quantum módico*) de creación original.**

Ahora bien, aunque a veces es muy difícil distinguirlas (e incluso tienen más valor), **el derecho de autor no protege las ideas ni la información**; sólo protege la expresión de la creación original. Esta primera distinción, fundamental, a la vez que difícil, será determinante a la hora de ver en qué medida el uso de recursos lingüísticos puede verse restringido por la propiedad intelectual.



A pesar de que con frecuencia se trata de obras protegidas (creaciones originales), el TRLPI excluye expresamente del ámbito de la PI la normativa, las resoluciones judiciales y los actos de organismos públicos, así como las traducciones oficiales de todos ellos (las traducciones no oficiales estarán, pues protegidas como obra).<sup>16</sup>

Así pues, de entre los diferentes recursos que pueden utilizarse para el procesamiento lingüístico (**para realizar el primer procesamiento**) habrá algunos protegidos por la PI como obra o prestación protegidas (por ejemplo, una novela escaneada, ya sea en versión original o traducida o ambas, una grabación de voz) pero habrá otros que no lo serán (por ejemplo, los data-sets de resultados científicos o las sentencias y normativa). Ahora bien, cuidado con éstos últimos porque a pesar de no estar protegidos como obra, su reutilización puede quedar indirectamente protegida a través del derecho *sui generis* sobre la base de datos que los contiene, tal como veremos más adelante.

Una vez se ha identificado que el recurso es una “obra”, cabe plantearse la segunda cuestión: **quién es el titular de derechos sobre la misma.**

El autor es el titular inicial de todos los derechos que la ley confiere sobre la obra. A partir de este momento, el autor puede explotar él mismo la obra o ceder sus derechos de explotación a otra persona (natural o jurídica). **Autor es la persona natural que crea la obra.**<sup>17</sup> Cuando en la creación participan varias personas, se está ante una coautoría. Sin embargo, tanto la condición de autor como la de coautor exige que la contribución personal sea una creación original. Hay muchas personas que participan en la creación de una obra, pero no todas con una aportación creativa; hay quien participa con una contribución de carácter técnico (fundamental y, a veces, de gran exigencia) que no le aportará la condición de autor ni de coautor.

Cuando la obra ha sido creada en virtud de una **relación laboral o funcional**, la ley prevé directamente una presunción de cesión de los derechos de explotación a favor del empresario para realizar su actividad habitual.<sup>18</sup> Se trata, sin embargo, de una presunción de cesión y, como tal, cabe pacto en contra: las partes pueden acordar otra cosa (que no hay cesión alguna o ampliar la cesión). Esta presunción de cesión no será de aplicación para aquellas obras que han sido encargadas por la Administración (por ejemplo, a través de un contrato de obra o servicio).

Además, la ley prevé dos situaciones excepcionales donde una persona jurídica puede ser titular inicial *ab origine* de los derechos de autor: la obra colectiva en cuya creación participan un gran número de autores y los programas de ordenador.<sup>19</sup> En ambos casos una persona física o jurídica

podrá ser el titular inicial de los derechos de propiedad intelectual de la obra colectiva que haya coordinado<sup>20</sup>.

No siempre se crea *ex novo*. También se puede crear utilizando total o parcialmente obras de otros. Crear una "obra nueva que incorpora una obra preexistente sin la colaboración del autor de esta última" es realizar una **obra compuesta**.<sup>21</sup>

Uno ejemplo de **obra compuesta/derivada** lo constituye la traducción. La traducción normalmente será una creación original y el traductor será autor (de la traducción) y ello a pesar de que la obra original (traducida) subyazca en la misma. Al realizar un procesamiento de una obra traducida será necesaria la autorización doble: del titular de la obra original y del titular de la traducción.

Por otra parte, el procesamiento HLT dará lugar a la creación de obras compuestas. El resultado del primer (y de subsiguientes) procesamiento/s puede ser también objeto de protección por la PI (si la tarea de HLT supone una creación original que da lugar a una obra nueva: derivada o compuesta), con lo cual a través de las diferentes cadenas o árboles de procesamiento (pipelines) se pueden estar generando nuevas obras o prestaciones protegidas.<sup>22</sup> Ello no quiere decir que todas las anotaciones sean obra; para determinar cuando estamos ante una nueva obra compuesta/derivada se debe aplicar el mismo criterio: creación original expresada. Así pues, por ejemplo, a lo mejor la segmentación de un texto puede no constituir una creación original que dé lugar a una obra nueva (obra derivada), pero alguna de las restantes tareas (i.e. la desambiguación o la caracterización temática) podrían conllevar un input de originalidad suficiente para entender que el resultado del procesamiento es un "nuevo" objeto de protección intelectual.

Además del "derecho de autor", la ley de PI confiere también una protección muy similar (aunque menor) a otras personas, naturales o jurídicas, que de alguna manera contribuyen a la creación o a la explotación de tales obras (intérpretes y ejecutantes, productor de fonogramas, productor audiovisual, entidades de radiodifusión, ).

Por otra parte, el régimen de **protección "sui generis" sobre las bases de datos** puede tener un peso importante en el ámbito de las tecnologías del lenguaje y ello por partida doble: en tanto los recursos utilizados (entradas) pueden estar contenidos en una base de datos (a los efectos de la LPI) -con el consiguiente derecho de su

**El derecho *sui generis* puede ser de aplicación tanto a bases de datos que contienen obras o prestaciones protegidas, como a base de datos que contengan datos e información no protegida por el derecho de autor**

fabricante a poder autorizar o prohibir la extracción y reutilización de la totalidad o una parte sustancial de su contenido-, y que los recursos resultantes del tratamiento (salidas) pueden también contenerse en una base de datos, con el mismo derecho respecto a terceros. Veamos la primera parte más en detalle.<sup>23</sup> El derecho *sui generis* puede ser de aplicación tanto a bases de datos que contienen obras o prestaciones protegidas (artículos científicos, grabaciones sonoras), como a base de datos que contengan datos e información no protegida por el derecho de autor (bases de datos de investigación científica, datos médicos, bases de datos de webs de dominio -es) o bases que contengan datos y obras (por ejemplo, bases de datos de patentes o de ayudas públicas).

Tener acceso legal a una BBDD no necesariamente permite extraer, reutilizar o explotar las obras protegidas/prestaciones ahí contenidas, ni a realizar el procesamiento de las mismas. Para ello se debe tener en cuenta:

- a- en primer lugar, si el procesamiento de los datos/obras constituye un acto de explotación o de reutilización a los efectos del TRLPI;
- b- Ante una respuesta positiva, si el procesamiento queda amparado por algún límite (excepción) legalmente previsto – por ejemplo, la copia privada o el tal vez inminente límite para la “minería de textos y datos” (*Text & Data Mining*, en adelante TDM);
- c- Ante una respuesta negativa, si existe alguna autorización o licencia que lo autorice. Esta licencia puede venir en cualquier formato y momento.

Si la respuesta a la pregunta (a) es negativa, ya no será necesario ni la aplicación de un límite legal que lo permite ni siquiera obtener licencia alguna del titular de tales contenidos. Igualmente, ante una respuesta positiva (a) y (b), no será necesaria la obtención de licencia alguna al quedar el procesamiento amparado por algún límite legalmente previsto.

#### 2.3.1.2 Actos (usos) que quedan sujetos al derecho exclusivo del autor/titular

**La primera cuestión a resolver es en qué medida el uso de tales recursos para técnicas de procesamiento del lenguaje supone un acto de explotación o de extracción/reutilización que la ley reserva al autor/titular/fabricante**

A los efectos de procesamiento del lenguaje, de entre todos los derechos los que centrarán la atención son los derechos exclusivos -en especial, los de reproducción, comunicación pública y transformación- ya que según como se definan (y el alcance que se les otorgue), será necesario el amparo de un límite o la obtención de una licencia (o no) para poder procesar los recursos

(entradas). Este mismo análisis servirá también para analizar la procedencia de autorizar a terceros la reutilización de los elementos (recursos) resultantes del procesamiento.

**A efectos del derecho de autor y conexos.** ¿Qué actos de explotación pueden verse afectados por las tareas de procesamiento lingüístico? En primer lugar, el **derecho de reproducción**. Este derecho alcanza a cualquier acto de “fijación directa o indirecta, provisional o permanente, por cualquier medio y en cualquier forma, de toda la obra o de parte de ella, que permita su comunicación o la obtención de copias.”<sup>24</sup>

Es decir, la amplia definición del derecho de reproducción queda reducida, en cierta medida, por la exclusión de las “reproducciones provisionales” que cumplan los requisitos ahí previstos:

- 1) que no tengan una significación económica independiente,
- 2) que sean transitorios o accesorios,
- 3) que formen parte integrante y esencial de un proceso tecnológico,
- 4) cuya única finalidad consista en
  - facilitar una transmisión en red (entre terceras partes) por un intermediario, o
  - facilitar una utilización lícita: la autorizada por el autor/titular o por la ley.

Se trata de requisitos cumulativos (salvo las dos alternativas previstas para el último) que deben cumplirse para que la copia realizada quede excluida del ámbito de exclusividad que la ley confiere al autor. Aunque es arriesgado aventurarse a una conclusión sin analizar las circunstancias concretas de cada tipo de utilización, es previsible que no todas las copias que se realizan para el procesamiento lingüístico cumplirán todas las condiciones citadas.

En segundo lugar, la **comunicación pública**, es decir, “todo acto por el cual una pluralidad de personas pueda tener acceso a la obra sin previa distribución de ejemplares a cada una de ellas.”<sup>25</sup> Este derecho se configura como la antítesis del derecho de distribución (que opera a través de copias tangibles de la obra y que, en principio, no parece que afecte demasiado los formatos de HLT) e incluye, pues, todo tipo de actos de transmisión o puesta a disposición del público en internet.

Y, para acabar, el **derecho de transformación**.<sup>26</sup> Es necesario distinguir también entre aquellos actos que se realizan en tareas HLT que son actos de reproducción (copias) de los recursos utilizados, de aquellos actos que constituyen actos de transformación por las diferentes implicaciones que tiene a varios niveles (límites aplicables, licencia o autorización necesaria o titularidad de los resultados).

La ley se refiere expresamente a la traducción, la adaptación y “**cualquier otra modificación en su forma de la que se derive una obra diferente...**” pero determinar cuando la modificación de una obra da lugar a una obra nueva o cuando simplemente la obra es transformada sin que derive una obra nueva es una de las cuestiones más escurridizas del derecho de autor. En particular, resulta de interés conocer si a resultas de las diferentes tareas de procesamiento lingüístico de un recurso (obra protegida) se puede obtener una obra nueva, distinta que generará un nuevo “set” de titulares y de derechos sobre tal resultado.

Con frecuencia, los recursos lingüísticos utilizados para el procesamiento procederán de una base de datos – con independencia de que se trate de datos no protegidos por la propiedad intelectual o que se trate de obras y prestaciones protegidas. El fabricante de una base de datos tiene un derecho a autorizar y prohibir la extracción y reutilización del todo o de una parte substancial del contenido de la base de datos. En otras palabras, cuando se utilizan recursos extraídos de una BBDD de jurisprudencia o normativa (y los recursos extraídos son sentencias o normativa en su versión original o en su traducción oficial) el fabricante de la misma podrá autorizar o prohibir tanto la copia (extracción) como la explotación (reutilización) de la totalidad o una parte sustancial de los contenidos (a pesar de no estar protegidos por el derecho de autor).

### 2.3.1.3 Uso de recursos sin necesidad de licencia (actos autorizados por ley)

Además de los recursos que no son obra ni prestación protegida, o aquellos que quedan excluidos de protección, ni tampoco están contenidos en una BBDD sujeta al derecho *sui generis*, también podrán ser libremente reutilizables aquellos recursos que se hallen en el **dominio público por haber expirado ya su plazo de protección**.

La protección de los derechos de propiedad intelectual no es eterna, sino que se protege únicamente durante un plazo. La regla general es de **70 años *post mortem auctoris***.

**Los derechos de propiedad intelectual se protegen únicamente durante un plazo, generalmente, 70 años *post mortem auctoris*.**

Los plazos de protección de los **derechos conexos son más cortos** y no se calculan en base a la vida del autor sino a partir de su publicación o divulgación: 50 años para las grabaciones sonoras o 25 años para las publicaciones editoriales.



El **derecho “sui generis”** sobre la base de datos se protege durante 15 años a partir de su fabricación – pero con cada nueva inversión substancial realizada en la obtención, verificación o presentación del contenido, empieza a correr un nuevo plazo de 15 años.

Y es que una vez agotado estos plazos (según corresponda) la obra / prestación / BBDD entra en el “dominio público” y podrá ser explotada (reutilizada) por cualquiera. La **distinción entre recursos protegidos y recursos en el “dominio público”** es, pues, fundamental a los efectos del tratamiento lingüístico. Todo aquello que está en el dominio público no requiere autorización alguna ni, de hecho, permite licencia de reutilización alguna (al menos, a los efectos de la propiedad intelectual).

Más allá del plazo de protección y el dominio público, la ley recoge un listado de “**excepciones” o límites** que permiten (directamente, por imperativo legal) el uso de una obra/prestación mientras todavía está protegida.

Entre estos límites se encuentran el llamado “derecho de cita,” la parodia y la copia privada – límites que benefician a todos los ciudadanos-, pero también otros límites más específicos como límites para fines de enseñanza y de investigación, para ciertos usos de obras huérfanas, el préstamo público, y otros límites dirigidos a facilitar la difusión de la información de actualidad, etc.

Para poderse amparar en un límite legal, el acto de explotación de la obra debe quedar circunscrito en las condiciones concretas dibujadas para cada límite (y siempre en armonía con el criterio interpretativo de la “regla de los tres pasos”);<sup>27</sup> de no cumplirse tales condiciones, el acto de explotación no autorizado se convierte en una infracción.

A efectos del procesamiento lingüístico, de entre los límites previstos en el TRLPI cabría destacar la copia privada del y, en cierta medida, la copia que pueden realizar las bibliotecas (y similares) para fines de investigación.<sup>28</sup> Ninguno ofrece suficiente amparo legal para poder prescindir de licencia para realizar actos de procesamiento del lenguaje. El primero porque sólo permitiría a los ciudadanos (no a las empresas ni a la Administración) realizar copias de obras, pero no transformarlas ni reutilizarlas de forma colectiva; y el segundo porque, además de solo permitir la reproducción, solo beneficiaría a bibliotecas, hemerotecas o archivos, y en tanto no existiera finalidad lucrativa. Por otra parte, la finalidad “de investigación o de conservación” exigida para este límite tampoco engranaría bien con el procesamiento lingüístico.

Pero mucho más interesante que los límites actualmente vigentes en el TRLPI, es la propuesta de límite a favor de la minería de datos (TDM) que se incluye en la propuesta de Directiva sobre Derecho

de Autor en el Mercado Único Digital COM(2016)593final que actualmente está siendo objeto de estudio por el Parlamento y la Comisión.

Este límite autorizaría la reproducción y extracción de recursos protegidos para fines de minería de datos, realizados por organismos de investigación únicamente, y siempre y cuando se trate de contenidos a los que se tenga acceso legal y la minería se realice para fines de investigación. Es decir, se trata de un límite con un alcance muy reducido que dejaría a gran parte de los procesos de tratamiento lingüístico desprovistos de amparo legal.

#### 2.3.1.4 La obtención de licencias / autorización para realizar HLT

En la medida en que el tratamiento lingüístico de obras y prestaciones protegidas no quede amparado por un límite legal, será necesario contar con la autorización del titular de tales obras y prestaciones.

Esta licencia puede ser de múltiples formas y a través de canales diversos, pero es importante tener presente que sólo pueden ser objeto de cesión o licencia los derechos de explotación; los derechos morales y los de simple remuneración son, normalmente, intransmisibles.

**Los derechos de explotación pueden ser objeto de cesión o licencia mientras que los derechos morales y los de simple remuneración son, normalmente, intransmisibles**

La autorización puede darse en el momento mismo de dar acceso a una BBDD o puede darse posteriormente; la autorización puede incluirse en las condiciones de uso (aviso legal) de una página web, etc. En todo caso, el alcance y las condiciones de la licencia son las que guiarán qué actos se pueden realizar y cuáles no; aunque ante la duda, la interpretación de una cláusula contractual siempre se realizará de forma restrictiva, en el sentido de no entender autorizado el uso.<sup>29</sup> No hay formalidad alguna a seguir, basta con que la existencia de la autorización sea clara y explícita.

Para proteger al autor, se establecen reglas tales como la nulidad de ceder derechos sobre el conjunto de la obra futura o sobre modalidades de explotación desconocidas en el momento de efectuar la cesión; o la interpretación restrictiva de las cláusulas contractuales (en caso de duda o silencio, siempre se realizará una interpretación *pro auctoris*), etc. Será importante tener presentes estas reglas tanto para conocer las condiciones en que los recursos pueden ser licenciados como las condiciones para licenciar los resultados para su reutilización en subsiguientes procesamientos.

La autorización o licencia para realizar actos de explotación HLT puede venir directamente del titular de derechos o a través de **las entidades de gestión colectiva**.












A menudo, los autores y los titulares encargan la gestión de sus derechos a entidades de gestión colectiva (art. 147 y siguientes TRLPI). Las entidades de gestión (sociedades de gestión colectiva de derechos de propiedad intelectual) son asociaciones formadas por los propios autores o titulares de derechos: artistas, productores, editores, etc. Estas entidades son capaces de actuar con eficacia en el ámbito mundial en virtud de los acuerdos de cooperación recíproca que tienen suscritos con las entidades de otros países. Normalmente, están más o menos especializadas según el tipo de obras (musicales, coreográficas, plásticas) y titulares (autores, artistas, productores). En principio, la gestión colectiva es voluntaria: el autor puede decidir si encomendar la gestión de sus derechos a una entidad de gestión o gestionarlos directamente. Sin embargo, en algunos supuestos, la propia ley impone la gestión colectiva de forma obligatoria (por ejemplo, la compensación por copia privada, o la compensación por la agregación de noticias, introducida por la Ley 21/2014). Aun a riesgo de una simplificación excesiva, las entidades actualmente existentes en España podrían ordenarse en tres grupos:

- Sociedades de autores: SGAE (Sociedad General de Autores y Editores: [www.sgae.es](http://www.sgae.es))
- VEGAP (Visual, Entidad de Gestión de Artistas Plásticos: [www.vegap.es](http://www.vegap.es))
- CEDRO (Centro Español de Derechos Reprográficos: [www.cedro.org](http://www.cedro.org) )
- DAMA (Asociación Derechos de Autor de Medios Audiovisuales: [www.damautor.es](http://www.damautor.es))
- AIE (Sociedad de Artistas Intérpretes o Ejecutantes de España: [www.aie.es](http://www.aie.es) )
- AISGE (Asociación de Actores Intérpretes, Sociedad de Gestión de España: [www.aisge.es](http://www.aisge.es))
- AGEDI (Asociación de Gestión de Derechos Intelectuales: [www.agedi.es](http://www.agedi.es) )
- EGEDA (Entidad de Gestión de Derechos de Productores Audiovisuales: [www.egeda.es](http://www.egeda.es) ).

Una vez legalmente constituidas y reconocidas por el Ministerio de Cultura, las entidades de gestión están obligadas a aceptar la administración de los derechos de autor y otros derechos de propiedad intelectual que les sean encomendados de acuerdo con su objeto o fines, y con sujeción a sus estatutos y demás normas aplicables al efecto.

Un formato de licencias de propiedad intelectual que ha obtenido gran aceptación pública, especialmente en internet y en el ámbito de la reutilización de información pública, son las **licencias Creative Commons**, con las que el autor (o titular) autoriza al público en general a realizar actos de

explotación de su obra, sujeto a unas condiciones que el autor escoja (i.e., finalidades comerciales y/o transformación). De nuevo, estas licencias pueden ser un instrumento fundamental tanto para identificar que recursos pueden ser lícitamente tratados (entradas) como para gestionar la reutilización de los recursos resultantes (salidas).

		<b>Reconocimiento (by):</b> Autoriza cualquier uso o acto de explotación que se haga de la obra (incluidos los usos comerciales y las transformaciones).
	=	<b>Reconocimiento (by) - Sin obra derivada (nd):</b> Excluye la transformación de la obra para crear otra nueva (obra derivada).
	€	<b>Reconocimiento (by) - No comercial (nc):</b> Excluye la posibilidad de utilizar la obra con fines comerciales.
	= €	<b>Reconocimiento (by) - Sin obra derivada (nd) - No comercial (nc):</b> No se permite ni hacer obras derivadas, ni los usos comerciales.
		<b>Reconocimiento (by) - Compartir igual (sa):</b> La cláusula de <i>copyleft</i> obliga al usuario a aplicar la misma licencia o similar a las obras derivadas creadas al amparo de la licencia.
	€ 	<b>Reconocimiento (by) - No comercial (nc) - Compartir igual (sa):</b> No se permiten los usos comerciales y se exige <i>copyleft</i> respecto de las obras derivadas en tanto no tengan fines no comerciales.

(by): attribution / (nc): non-commercial / (nd): no derivatives / (sa): share alike

Las licencias más permisivas BY y BY-SA permitirían claramente la utilización de los recursos para fines de HLT (teniendo en cuenta que la SA exige que los resultados del procesamiento lingüístico sean puestos a disposición del público para su reutilización bajo la misma licencia, es decir, con carácter gratuito y en abierto). En cambio, la licencia BY-NC podría restringirla en demasía al excluir el procesamiento que se realizara con fines directa o indirectamente comerciales (por ejemplo, obteniendo subvenciones o algún otro tipo de ayuda económica).

#### 2.3.1.5 Licencias *Creative Commons*, en detalle

Un formato de licencias de propiedad intelectual que ha obtenido gran aceptación pública, especialmente en internet y en el ámbito de la RISP, son **las licencias *Creative Commons***, con las que el autor (o titular) autoriza al “público” en general a realizar actos de explotación de su obra, sujeto a unas condiciones que el autor escoja (i.e., finalidades comerciales y/o transformación). De nuevo, estas licencias pueden ser un instrumento fundamental tanto para identificar que recursos pueden ser lícitamente tratados (recursos) como para gestionar la reutilización de los recursos resultantes (resultados).

Mediante estas licencias estandarizadas, el autor autoriza al público, en general, a utilizar y explotar su obra, sujeto a unas condiciones contractuales. Cuando un usuario decide utilizar una obra bajo una licencia CC se convierte en licenciatarario y se compromete a aceptar y respetar las condiciones de la licencia establecida por el autor.














Todas las licencias CC se expresan en tres niveles de lectura:

- *Commons deed*: es el nivel más visible (y característico) de la licencia.
- *Legal code*: es la licencia que redactaría un abogado (norteamericano).
- *Digital code*: permite la identificación digital de las obras licenciadas (y las condiciones) por motores de búsqueda en Internet y demás programas.

Las cláusulas de la licencia vienen prefijadas. La licencia básica (by) autoriza la reproducción, distribución, transformación y comunicación pública de la obra, para cualquier finalidad y para todas las modalidades de explotación, con carácter gratuito y para todo el plazo de protección, bajo una única condición: que se mantenga el crédito a su autor. A partir de esta licencia "básica", el autor puede reducir el alcance de su autorización de tres maneras:

- excluyendo los usos comerciales de su obra y/o
- excluyendo la modificación de su obra; o,
- permitiendo la modificación, a condición de que la nueva obra quede sujeta a la misma licencia (esta es la cláusula "copyleft").

Estas opciones dan lugar a **seis licencias** diferentes que se identifican con unos iconos característicos:

		<b>Reconocimiento (by)</b> : Autoriza cualquier uso o acto de explotación que se haga de la obra (incluidos los usos comerciales y las transformaciones).
		<b>Reconocimiento (by) - Sin obra derivada (nd)</b> : Excluye la transformación de la obra para crear otra nueva (obra derivada).
		<b>Reconocimiento (by) - No comercial (nc)</b> : Excluye la posibilidad de utilizar la obra con fines comerciales.
	 	<b>Reconocimiento (by) - Sin obra derivada (nd) - No comercial (nc)</b> : No se permite ni hacer obras derivadas, ni los usos comerciales.
		<b>Reconocimiento (by) - Compartir igual (sa)</b> : La cláusula de <i>copyleft</i> obliga al usuario a aplicar la misma licencia o similar a las obras derivadas creadas al amparo de la licencia.
	 	<b>Reconocimiento (by) - No comercial (nc) - Compartir igual (sa)</b> : No se permiten los usos comerciales y se exige <i>copyleft</i> respecto de las obras derivadas en tanto no tengan fines no comerciales.

(by): attribution / (nc): non-commercial / (nd): no derivatives / (sa): share alike

Con esta combinación de opciones, el autor puede calibrar el nivel de control que quiere mantener sobre su obra y ejercer fácilmente los derechos de explotación que la ley le confiere, a favor del público en general.

Las principales características de las licencias CC son:

a) Las licencias CC pueden ser utilizadas por cualquier titular de propiedad intelectual sobre cualquier tipo de obra o prestación:

- por autores (respecto de sus obras literarias, musicales o de arte),
- por artistas (respecto de sus interpretaciones o ejecuciones musicales),
- por productores de fonogramas (respecto de los fonogramas producidos por ellos),
- por entidades de radiodifusión (respecto de sus emisiones radiofónicas),

Y en la v.4, también por el titular del derecho sui generis sobre una base de datos.

b) Todas las licencias CC obligan al cumplimiento de los derechos morales de atribución e integridad, al exigir el reconocimiento del autor de la obra (su nombre o pseudónimo y, si así lo establece el autor, también la fuente de publicación) y establecer que el usuario "no puede distorsionar, mutilar, modificar o realizar cualquier otra acción denigratoria en relación con la obra que perjudique el honor o la reputación de su autor".

c) Las licencias CC permiten autorizar todos los derechos de explotación reconocidos por la ley española; en concreto, los derechos de reproducción, distribución, comunicación pública (y transformación, cuando el autor así lo haya decidido). Con sujeción a tres condiciones de aplicación imperativa (que no pueden ser excluidas ni modificadas por el autor):

- **Carácter gratuito:** el carácter gratuito se limita a los actos/ usos autorizados bajo la licencia. El autor puede autorizar (y cobrar por) cualquier acto de explotación que autorice más allá de la licencia CC (especialmente, si ha excluido los usos comerciales y las obras derivadas) o incluso obtener remuneración por usos incluidos en la licencia (si es que alguien está dispuesto a pagar por un uso que otros realizarán gratuitamente). Además, a pesar del carácter gratuito de las licencias, el autor se puede beneficiar de las remuneraciones que la ley pueda establecer a su favor, con carácter irrenunciable y de gestión colectiva obligatoria (por ejemplo, la remuneración compensatoria por copia privada). Los autores pueden negarse a percibir las, pero no hay duda alguna de que la entidad de gestión está legitimada para cobrar el canon por copia privada, también por obras licenciadas bajo CC. En cambio, quedan comprendidas en la licencia (y por tanto, el autor renuncia a ellas) las

remuneraciones que, a pesar de venir establecidas por la ley, son renunciables (el autor puede renunciar a cobrarlas) o aquellas que sean de gestión colectiva "voluntaria".

- Todas las modalidades de explotación. Las licencias CC cubren expresamente "todos los medios y formatos, tangibles o intangibles, conocidos o por conocer". Las modalidades de explotación son transversales a los derechos de explotación y corresponden a los medios técnicos y a los mercados disponibles en cada momento. Por tanto, los actos de explotación autorizados por las licencias CC no se limitan a Internet sino que también alcanzan al mundo "real" (ediciones en papel, en CD y DVD, radiodifusión, etc.).

- A perpetuidad: para todo el plazo de protección de la obra, de acuerdo con la ley de cada país donde se explote y proteja la obra. El autor puede, en cualquier momento, explotar la obra bajo otra licencia (ya sea CC o no) o, incluso, retirarla, pero no puede resolverla. La resolución de la licencia sólo se prevé en caso de incumplimiento por parte de un usuario: en tal caso, la resolución será automática, pero sólo será efectiva respecto de ese usuario.

d) Las licencias CC no son exclusivas; por tanto, el autor puede otorgar otras licencias sobre la misma obra con diferentes condiciones. Sin embargo, las subsiguientes licencias (CC u otras) sólo se podrán otorgar en régimen de no exclusividad y, muy a menudo, lo que tiene valor en el mercado es precisamente la exclusividad

e) Las licencias CC no afectan a los usos autorizados por los límites legales

f) Las licencias CC prohíben el uso de DRM y TPM; Con el objetivo de asegurar el uso y los actos de explotación licenciados, las licencias CC prohíben utilizar ningún tipo de medida tecnológica de control de acceso y/o anticopia (DRM) que restrinja el uso de la obra licenciada. Indirectamente, esta prohibición añade otra ventaja: favorecer que las obras estén al alcance del público sin trabas tecnológicas para que puedan ser utilizadas, lícitamente, también más allá de los usos cubiertos por las licencias CC.

g) Las licencias CC aportan una apariencia de legalidad (se presume que ha sido el autor quien ha puesto la obra a disposición del público con la licencia que libremente ha escogido); su valor depende directamente de la confianza que tenga el público y ello depende directamente del buen uso que se haga de ellas.

Sirva esta explicación para ayudar a identificar las condiciones para utilizar recursos que se hallen disponibles en internet (o en otros formatos) sujetos a una licencia CC. La consideración de cual de estas licencias es la que mejor se ajusta a las finalidades de reutilización se analizará en otra guía. Las aplicaciones informáticas (software) tienen también a su disposición licencias específicas que pueden facilitar su reutilización y que examinaremos en el siguiente capítulo 3 (Proceso).

### 2.3.2 ¿Se pueden utilizar recursos lingüísticos que incorporen datos personales?

Las administraciones públicas en el desarrollo de sus funciones recaban ingentes cantidades de datos personales de la ciudadanía. Para hacer frente a las posibles amenazas y riesgos, en cualquier procesamiento de lenguaje debe valorarse la existencia de datos personales y, en su caso, adoptar las medidas previstas respecto a su tratamiento adecuado y seguridad.

Los datos personales son **cualquier información concerniente a una persona física identificada o identificable**, es decir, cuya identidad pueda determinarse, directamente o indirectamente, en particular, a través de un

número de identificación o uno o diversos elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social. Algunos datos personales son considerados sensibles y cuentan con una protección legal más elevada (por ejemplo, los datos relativos a la ideología la afiliación sindical, la religión, el origen racial o étnico, la salud o la vida sexual).

Desde el punto de vista de los recursos utilizados por las tecnologías del lenguaje, es frecuente que recursos de gran utilidad contengan datos personales como es el caso de las sentencias judiciales, los directorios o las historias clínicas.

El uso de datos personales como recurso para **el procesamiento del lenguaje supone realizar un tratamiento**, es decir, una operación o procedimiento técnico de carácter automatizado, que permite la recogida, grabación, conservación, elaboración, modificación, bloqueo y cancelación, así como la cesión de datos que resulte de comunicaciones, consultas, interconexiones y transferencias.

**El procesamiento de recursos lingüísticos que contengan datos personales puede generar amenazas para la privacidad de las personas y poner en riesgo dichos datos**

Como punto de partida, es importante tener presente que la obligación de difundir información pública para que pueda ser reutilizada en el marco de un procesamiento del lenguaje no incluye una obligación de difusión de información personal.

La LOPD establece una serie de principios que deben guiar cualquier tratamiento de datos personales:

- Calidad de los datos: los datos personales deben **ser adecuados, pertinentes y no excesivos respecto a la finalidad para la que se obtienen y que sean exactos y actualizados**. En este punto la principal dificultad estriba en conocer cuando una finalidad es compatible con otra para lo que pueden resultar de utilidad diversos criterios que se han ido proponiendo [1].
- Consentimiento: **el tratamiento de los datos personales requiere el consentimiento inequívoco del interesado** que, en determinados, deberá ser expreso y por escrito. Sin embargo, el principio del consentimiento tiene algunas excepciones, por ejemplo, cuando los datos de carácter personal se recojan para el ejercicio de las funciones propias de las Administraciones públicas en el ámbito de sus competencias atribuidas por una norma con rango de ley o de derecho comunitario o cuando se refieran a las partes de una relación administrativa.
- Seguridad de los datos: para ello se deben adoptar las medidas de índole técnica y organizativas necesarias para **evitar la alteración de los datos o su pérdida, tratamiento o acceso no autorizado**.
- Secreto en el tratamiento de los datos: el responsable del fichero y del tratamiento están obligados al **secreto profesional**.
- Congruencia y racionalidad en su utilización: los datos de carácter personal **sólo podrán ser comunicados a un tercero para el cumplimiento de fines directamente relacionados con las funciones legítimas** del cedente y del cesionario con el previo consentimiento del interesado.

Cuando una Administración pública permite la reutilización de recursos que contengan datos de carácter personal, está autorizando llevar a cabo una cesión o comunicación de datos personales, es decir, una revelación de datos realizada a una persona distinta del interesado lo que requiere el consentimiento del interesado. Sin embargo, este

**La cesión de datos personales sólo se puede realizada para el cumplimiento de fines directamente relacionados con las funciones legítimas de cedente y cesionario con el consentimiento del interesado**



consentimiento no será preciso cuando la cesión está autorizada en una ley, cuando se trate de datos recogidos de fuentes accesibles al público, cuando la cesión se produzca entre Administraciones públicas y tenga por objeto el tratamiento posterior de los datos con fines históricos, estadísticos o científicos. Tampoco será necesario el consentimiento, entre otras situaciones, cuando la administración obtenga o elabore los datos personales con destino a otra.

El uso de datos personales exige la implantación de las medidas de seguridad previstas en la LOPD para garantizar la confidencialidad, integridad y disponibilidad de los datos. Las medidas deben ser impulsadas por los responsables de los ficheros y los encargados de tratamiento o, en su caso, los responsables de seguridad a quienes hayan asignado formalmente la función de coordinar y controlar las medidas de seguridad aplicables. Las medidas de seguridad serán diferentes en función de los datos (básico, medio y alto).

Las normas anteriores pueden complementarse a través de **códigos de conducta** que establezcan las condiciones de organización, régimen de funcionamiento, procedimientos aplicables, normas de seguridad del entorno, programas o equipos, obligaciones de los implicados en el tratamiento y uso de la información personal. Estos códigos de conducta deben ser depositados e inscritos en el Registro General de Protección de Datos o en el registro autonómico correspondiente que dará traslado al registro para su inclusión en el registro gestionado por la AEPD.

Cuando los datos personales estén en poder de las administraciones públicas, es importante recordar que la LRISP dispone que el régimen de reutilización garantiza el pleno respeto de los principios que consagran la protección de datos personales. Por ello, la LRISP no es de aplicación a los documentos o a las partes de los mismos que obren en poder de las Administraciones públicas cuyo acceso esté limitado por motivos de protección de los datos personales.

De este modo, cuando en aplicación de la normativa sobre transparencia prevalezca la protección de los datos personales sobre el acceso a la información, no podrán reutilizarse dichos datos excepto que así lo prevea una ley, se recabe el consentimiento de los interesados o se proceda a la anonimización de los datos personales.

Por lo que respecta al consentimiento, este deber ser informado de forma que el afectado conozca inequívocamente la finalidad a la que se destinarán los datos respecto de cuya comunicación se solicita el

**Los interesados podrán en todo momento ejercer sus derechos de acceso, rectificación, cancelación y oposición**



consentimiento y el tipo de actividad desarrollada por el cesionario. Los interesados podrán en todo momento ejercer sus **derechos de acceso, rectificación, cancelación y oposición** (derechos ARCO).

Por lo que respecta a la anonimización o disociación de los datos, esta debe impedir la identificación de las personas de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable, es decir, que permita obtener un dato disociado. De acuerdo con la AEPD, hay que **garantizar la irreversibilidad de la anonimización**, y para ello deberán valorar tanto las fuentes de información disponibles en los diferentes medios y, en particular, a través de internet, como la tecnología aplicable en los procesos de anonimización y en los de reidentificación [2]. La anonimización se puede llevar a cabo utilizando diferentes técnicas como la aleatorización, la generalización y la pseudonimización [3].

Sin embargo, es importante ser conscientes de que cuando se faciliten datos para que puedan ser reutilizados como recursos para el procesamiento del lenguaje, la Administración pública pierde el control sobre los quien los reutiliza y los usos que hace con ellos. En efecto, a pesar de la anonimización de los datos, las técnicas del *big data* permiten el cruce de datos de diferente origen que pueden llevar a que finalmente se pueda llegar identificar a una persona [3].

En cualquier caso, cuando una Administración pública se plantee la posibilidad de facilitar la reutilización de recursos lingüísticos que contengan datos personales debería llevar a cabo con carácter previo una **evaluación del impacto sobre la privacidad** (*Privacy Impact Assessments*). Este trámite no es preceptivo en la actualidad pero sí que se contempla en el Reglamento General de Protección de Datos, que entrará en vigor en 2018, que se propone impulsar esta medida en aquellos casos en los que sea probable que las operaciones de tratamiento entrañen un alto riesgo para los derechos y libertades de las personas físicas para evaluar el origen, la naturaleza, la particularidad y la gravedad de dicho riesgo y adoptar las medidas adecuadas o en su caso consultarse a la autoridad de control antes del tratamiento. Entre otros aspectos, la evaluación deberá incluir una evaluación de los riesgos para los derechos y libertades de los interesados y las medidas previstas para hacer frente a dichos riesgos [4].

Más allá de la anonimización técnica de los datos, se puede garantizar la anonimización a través del **establecimiento de condiciones en las licencias de reutilización**. En efecto, las Administraciones públicas pueden someter la reutilización a determinadas condiciones entre las que destaca cuando la información contenga datos de carácter personal, la finalidad o finalidades concretas para las que es posible la reutilización futura de los datos.

Además de la inclusión de estas condiciones o en el caso de que la reutilización no esté sometida a la obtención de una autorización, las Administraciones públicas pueden informar de estas circunstancias a los reutilizadores a través avisos legales que se difundan en los portales de datos abiertos o en los sitios web donde estén disponibles los datos.

Finalmente, un último aspecto a destacar es el relativo al responsable de la anonimización que será el responsable del tratamiento por lo que las Administraciones públicas no deben facilitar los recursos con los datos personales a los reutilizadores bajo el compromiso de que estos lleven a cabo el proceso de anonimización. Este es un aspecto importante por las consecuencias que se pueden derivar del incumplimiento de lo previsto (multa de 40.001 a 600.000 euros por cesión de datos personales sin contar con legitimación para ello o multa de 10.001 a 50.000 euros por el incumplimiento de las condiciones impuestas en la correspondiente licencia o en la normativa reguladora de la reutilización).

### 2.3.3 ¿Existen límites al uso de los recursos lingüísticos?

La LTAIBG ha incluido diferentes límites que son aplicables tanto a las obligaciones de difusión activa de la información como a las solicitudes de acceso a la información. En particular, el derecho de acceso podrá ser limitado cuando acceder a la información suponga un perjuicio para: la seguridad nacional; la defensa; las relaciones exteriores; la seguridad pública; la prevención, investigación y sanción de los ilícitos penales, administrativos o disciplinarios; la igualdad de las partes en los procesos judiciales y la tutela judicial efectiva; las funciones administrativas de vigilancia, inspección y control; los intereses económicos y comerciales; la política económica y monetaria; el secreto profesional y la propiedad intelectual e industrial; la garantía de la confidencialidad o el secreto requerido en procesos de toma de decisión; la protección del medio ambiente.

**La confidencialidad de la información constituye una excepción al principio general de reutilización de los recursos públicos**

Tal y como se desprende de la lectura de la LTAIBG, la concurrencia de uno de estos límites no supone la imposibilidad ni de acceder a la información ni de reutilizarla. Es necesario que el acceso a la información (o su reutilización) pueda suponer un perjuicio para los bienes o derechos que dichos límites persiguen proteger. La aplicación de los límites debe ser justificada y proporcionada a su



objeto y finalidad de protección y también deberá atender a las circunstancias del caso concreto, especialmente a la concurrencia de un interés público o privado superior que justifique el acceso.

Asimismo, es importante tener presente que tal y como dispone la LTAIBG, en los casos en que “la aplicación de alguno de los límites previstos no afecte a la totalidad de la información, se concederá el acceso parcial previa omisión de la información afectada por el límite salvo que de ello resulte una información distorsionada o que carezca de sentido. En este caso, deberá indicarse al solicitante que parte de la información ha sido omitida”.

Además de los límites a la transparencia pública que deben trasladarse a la reutilización de información previstas en la LTAIBG, otras normas disponen preceptos que persiguen garantizar la confidencialidad de información facilitada por las empresas a las administraciones públicas y que, eventualmente, pueden ser un recurso utilidad para las tecnologías de procesamiento del lenguaje, por ejemplo, en materia de defensa de la competencia o de contratación pública.

### 3 El desarrollo de procesadores lingüísticos

---

Las tecnologías de procesamiento del lenguaje permiten realizar multitud de procesos como el análisis, la traducción, el resumen, la relación o la explotación de textos. El procesamiento de lenguaje se basa en el uso de procesadores lingüísticos [5, 21].

#### 3.1 La protección jurídica de las aplicaciones

Más allá de examinar como el régimen de la propiedad intelectual puede afectar a los recursos objeto de tratamiento (entrada) y a los resultados del mismo (salida), corresponde analizar ahora la protección del procesador (el programa informático), ya que tiene un régimen especial de protección dentro de la propia ley de propiedad intelectual.

Un **programa** es un conjunto de instrucciones que hacen que una máquina –el ordenador– funcione y ejecute una tarea o proporcione un resultado. El programa de ordenador puede quedar protegido por el TRLPI siempre y

**La posibilidad de patentar programas es uno de los caballos de batalla actuales para la industria del software. En Europa no es posible (al menos formalmente) patentar un programa de ordenador, si bien es posible que éste forme parte de una invención patentable.**

cuando sea una “creación intelectual propia de su autor” (criterio armonizado en toda la UE). El derecho de autor protege el programa en cualquiera de sus expresiones posibles (**código fuente y código objeto**) y en su totalidad, incluidas las interfaces. Como es lógico, no se protegen las ideas ni los principios que conforman el programa. En cambio, sí que se protegen la documentación preparatoria, la documentación técnica y los manuales de uso del programa de ordenador.

Además de los casos de autoría individual (autor es quien crea el programa) y de coautoría (obra en colaboración), se establecen dos reglas específicas para los programas de ordenador:

- cuando el programa de ordenador haya sido creado en régimen de **obra colectiva**, la persona natural o jurídica que la edita y divulga tendrá -salvo pacto en contra- la consideración de autor;
- cuando el programa de ordenador haya sido **creado en virtud de una relación laboral** (trabajador asalariado), la presunción de cesión de derechos a favor del empleador -salvo pacto en contra- tiene un alcance muy amplio y es en exclusiva.



Puesto que no se dice nada de los **programas creados por encargo a personas (o empresas) externas**, si la Administración que encarga el programa quiere obtener la titularidad de los derechos de explotación, deberá obtenerlos por cesión expresa (contrato) del autor / titular original.

Esta cesión de derechos puede pactarse tanto en las cláusulas de la licitación o concurso, como en el contrato de encargo; sin esta cesión de derechos, la Administración podrá utilizar la obra o prestación encargada (en su actividad interna) pero no podrá realizar propiamente una “explotación” de la PI de la obra encargada (a pesar de haber pagado por ella) ni tampoco sujetarla a la reutilización. Previendo la importancia de este aspecto, la Ley de Contratos del Sector Público (Real Decreto Legislativo 3/2011, de 14 noviembre) establece una presunción de cesión (también sujeta a pacto en contra) a favor de la Administración contratante.<sup>30</sup>

En este mismo sentido, véanse también las consideraciones realizadas ya en la Reutilización de activos. Guía del Ministerio de Hacienda y Administraciones Públicas (2015), apartados 46 a 52.

Por otra parte, véase el modelo cláusula para incluir en los pliegos de contratación administrativa para la adquisición de programas de ordenador que propone el ENI en “Reutilización de Activos. Guía de publicación y licenciamiento de activos” (2015) #52, pag.18-19.

El **plazo de protección** de un programa de ordenador se rige por las normas generales:: 70 años *post mortem auctoris* o, cuando el autor es persona jurídica, 70 años desde la divulgación.<sup>31</sup>

El autor (persona física o jurídica) del programa de ordenador tiene derechos de explotación y derechos morales sobre el mismo. Pero la ley prevé unos **límites específicos** a estos derechos, a favor del “usuario legítimo” (la ley autoriza al usuario legítimo a **transformar y corregir errores** en la medida necesaria para utilizar el programa "de acuerdo con su finalidad propuesta"; hacer una **copia de seguridad**; y analizar el programa, mientras lo utiliza). También autoriza la ley a realizar acciones de ingeniería inversa y descompilación para asegurar **la interoperabilidad con otros programas** (que los programas sean capaces de intercambiar información y funcionar con los ya existentes).

La explotación de un programa de ordenador puede hacerse a través de una gran variedad de licencias. Además de la posibilidad de utilizar licencias (**End User License Agreements**) con los “usuarios legítimos”, la explotación de un programa de ordenador puede efectuarse a través de las llamadas **licencias de software libre** (hay diversas y no todas son iguales). Entre ellas, merece especial atención la licencia GPL (*General Public License*) de la Free Software Foundation, no sólo por ser la primera, sino por ser la que asegura el llamado *copyleft*.



Estas licencias nos afectarán en la medida en que utilicemos software pre-existente en la creación del procesador (por ejemplo, la GPL obliga al siempre *copyleft* y por tanto, cualquier software que se produzca utilizando código de un programa de ordenador sujeto a una GPL deberá también explotarse bajo la misma licencia -es el llamado "efecto viral").

Mediante la GPL, el titular del programa permite expresamente a sus usuarios usar, modificar y redistribuir el programa (con o sin modificaciones), con dos únicas condiciones: dar atribución al autor original y asegurar que cualquier nueva versión del mismo será distribuida bajo la misma licencia GPL; De esta manera se aseguran las "libertades de uso, modificación y redistribución" y se evita la "apropiación" de los programas realizados con software libre. La licencia se adjunta directamente al código fuente (licencia y obra son inseparables) y debe reproducirse siempre, en todo el software redistribuido.

Más allá de este **mínimo común a todas las licencias de software libre**, existen variedades importantes. Por ejemplo, el Open Source renuncia a la exigencia del *copyleft* y se evitó utilizar la expresión "libre", superando así las dudas sobre la gratuidad (que siempre había generado la expresión software "libre" / *free software*). Así pues, el adjetivo "libre" se refiere a estas libertades fundamentales que otorgan las licencias, y no a la gratuidad del software (el software comercial es una parte importante del software libre).

En el caso del software "propietario", el usuario legítimo no tiene acceso al código fuente, de manera que su ámbito de actuación se limita al simple "uso" del mismo.

### **3.2 ¿Qué papel pueden tener las administraciones públicas en el desarrollo de procesadores del lenguaje?**

Las Administraciones públicas utilizan procesadores de lenguaje en el desarrollo de sus funciones. En ocasiones, estos procesadores han sido desarrollados por los propios servicios técnicos de las Administraciones públicas. En otras ocasiones, como se verá en el siguiente punto, estas encargan su desarrollo a empresas privadas. Más allá de sus propias necesidades, las Administraciones públicas pueden fomentar el desarrollo de procesadores del lenguaje por parte de la industria.

Desde un punto de vista general, el Plan de Impulso de las Tecnologías del Lenguaje prevé como objetivo general desarrollar la industria del procesamiento del lenguaje y la traducción automática en España. Asimismo, se fija como objetivo la creación de las plataformas comunes de procesamiento de lenguaje y de traducción automática de las Administraciones Públicas.

**Las Administraciones públicas deben fomentar el desarrollo de infraestructuras lingüísticas por parte de la industria**

En el marco del Plan de Impulso de las Tecnologías del Lenguaje, las Administraciones públicas deben fomentar el desarrollo de estas infraestructuras por parte de la industria. Asimismo, deben desarrollar infraestructuras para el procesamiento del lenguaje que utilizarán en el ejercicio de sus funciones y actividades. Finalmente, las Administraciones públicas también pueden plantearse la posibilidad de ofrecer en el mercado las tecnologías que hayan desarrollado para tal cometido con la finalidad de contribuir a los objetivos fijados por el Plan de Impulso de las Tecnologías del Lenguaje.

Tanto el fomento del desarrollo de la industria de las tecnologías del lenguaje como la oferta de servicios vinculados a dichas tecnologías por parte de las propias Administraciones públicas, puede generar problemas desde la perspectiva del derecho de la competencia. En efecto, si bien el ordenamiento jurídico prevé tanto la actividad de fomento como la propia actividad económica de las Administraciones públicas, es necesario que el desarrollo de estas actividades no distorsione las reglas del mercado perjudicando a otros actores.

**Las ayudas o subvenciones que se faciliten a la industria para el desarrollo de las tecnologías del lenguaje no puedan ser contrarias a las normas de la competencia**

Por lo que se refiere a la actividad de fomento de la industria de las tecnologías del lenguaje por parte de las Administraciones públicas, es necesario garantizar que las ayudas o subvenciones que se faciliten a la industria para el desarrollo de las tecnologías del lenguaje no puedan ser contrarias a las normas de la competencia ni a nivel estatal ni tampoco comunitario [6].

Respecto a la oferta en el mercado de las aplicaciones desarrolladas por las Administraciones públicas, es necesario diferenciar esta actividad de las Administraciones públicas vinculadas a las tecnologías del lenguaje de aquellas estrictamente relacionadas con el ejercicio de funciones públicas. Cuando las Administraciones públicas decidan ofrecer servicios en el mercado basados en

las aplicaciones desarrolladas por ellas debe garantizarse la libre competencia respecto a otros desarrolladores del lenguaje evitando de este modo el abuso de posición dominante, los acuerdos exclusivos o las ayudas cruzadas u otras prácticas que puedan ser consideradas contrarias a la competencia.

En cualquier caso, la dificultad en este punto radica en delimitar adecuadamente estas funciones públicas. Desde un punto de vista general esta delimitación se vincula al alcance de la noción de lo público y, por ende, del papel de la Administración pública en la sociedad [7]. De hecho, no se puede desconocer que cuando más amplio sea el papel de la Administración pública menor será el espacio que se dejará a la industria para desarrollar las tecnologías del lenguaje. De este modo, cuando las Administraciones públicas ofrezcan unos servicios basados en aplicaciones de las tecnologías del lenguaje que hayan desarrollado para el ejercicio de sus funciones o actividades pero que vayan más allá de la satisfacción de sus funciones propias ofreciéndolos como un servicio en el mercado, deben cumplir con los mismos principios exigidos a la industria por las normas de la competencia. De no ser así, las Administraciones públicas podrían estar abusando de su posición infringiendo de este modo las normas de la competencia.

### **3.3 ¿Pueden las Administraciones públicas contratar el desarrollo de los procesadores?**

Buena parte de los procesadores de lenguaje que utilizan las Administraciones públicas han sido desarrollados por la industria sobre la base de contratos.

Desde un punto de vista general, en este punto debe hacerse una remisión a lo previsto en el Real Decreto Legislativo 3/2011, de 14 de noviembre, por el que se aprueba el texto refundido de la Ley de Contratos del Sector Público.

Sin embargo, es necesario centrar la atención en dos aspectos estrechamente relacionados con la contratación del desarrollo de aplicaciones de procesamiento del lenguaje.

En primer lugar, la relativa a los derechos de propiedad intelectual o industrial vinculados a dichas aplicaciones. En estos casos, el TRLCSP prevé que los contratos de servicios que tengan por objeto el desarrollo y la puesta a disposición de productos protegidos por un derecho de propiedad intelectual o industrial llevarán aparejada la cesión de éste a la Administración pública contratante. Sin embargo, los pliegos de cláusulas administrativas o en el documento contractual pueden disponer otra cosa.



En segundo lugar, la relativa a la innovación que puede acompañar el desarrollo de estas aplicaciones y en la que las PYMES pueden tener un papel significativo. En la actualidad, la innovación se puede canalizar a través de la compra pública pre-comercial o la compra pública de tecnología innovadora [8]. En el futuro deberá tenerse en cuenta el nuevo procedimiento de asociación para la innovación que prevé el proyecto de ley de contratos del sector público.

### **3.4 ¿Cómo se puede fomentar los procesadores del lenguaje en lenguas cooficiales?**

El Plan de Impulso de las Tecnologías del Lenguaje define como objetivo general fomentar el desarrollo del procesamiento del lenguaje y la traducción automática en lengua española y lenguas cooficiales. En particular, se propone aumentar el número, calidad y disponibilidad de las infraestructuras lingüísticas en español y lenguas cooficiales.

En particular, es necesario destacar el papel que en este ámbito deben jugar las administraciones autonómicas en aquellas Comunidades Autónomas que disponen de lengua propia para favorecer, estimular y fomentar la investigación, producción y comercialización de todo tipo de productos en las lenguas cooficiales relacionados con las industrias de la lengua.

### **3.5 ¿Se pueden reutilizar los procesadores del lenguaje?**

Con el fin de facilitar la interoperabilidad y promover la economía y la eficiencia en las administraciones públicas, la legislación prevé la reutilización de sistemas y aplicaciones propiedad de la Administración pública. De este modo, gracias a las aplicaciones de procesamiento del lenguaje que haya podido desarrollar una Administración pública se pueden beneficiar otras Administraciones públicas con los beneficios que ello puede suponer no sólo por el ahorro de costes económicos y temporales sino también por la experiencia compartida que la reutilización de las aplicaciones puede generar.

La reutilización de sistemas y aplicaciones se articula como un instrumento mediante el cual las administraciones titulares de los derechos de propiedad intelectual de aplicaciones, bien hayan sido desarrolladas por ellas, bien por un tercero, podrán ponerlas a disposición de cualquier administración pública sin contraprestación y sin necesidad de convenio. Estas aplicaciones podrán ser declaradas fuentes abiertas.

En esta dirección, se debe tener presente que el Plan de Impulso de las Tecnologías del Lenguaje dispone como

**Para facilitar la reutilización de las aplicaciones, las aplicaciones pueden ser declaradas de fuentes abiertas**

objetivo que “asegurar la coordinación en el desarrollo de infraestructuras lingüísticas, evitando duplicidades y buscando sinergias” y “disponer de herramientas comunes para la generación así como campañas de evaluación de infraestructuras lingüísticas”.

Para facilitar la reutilización de las aplicaciones la Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público (LRJSP) dispone que las aplicaciones pueden ser declaradas de fuentes abiertas.

Asimismo se prevé que la Administración General del Estado debe crear un directorio de aplicaciones, que debe integrar los directorios de aplicaciones del resto de Administraciones públicas, que deberán consultar las Administraciones públicas antes de desarrollar una aplicación propia para ver si existe alguna aplicación que pueda ser reutilizada y que satisfaga total o parcialmente sus necesidades. En el caso de que exista, las Administraciones públicas están obligadas a su uso salvo que justifiquen su decisión de no utilizarla y desarrollar una aplicación propia en términos de eficiencia de acuerdo con lo que dispone la Ley Orgánica 2/2012, de 27 de abril, de Estabilidad Presupuestaria y Sostenibilidad Financiera.

## 4 El resultado

---

La última etapa del procesamiento del lenguaje es la obtención de un resultado que puede ser de muy diverso tipo como las campañas de evaluación o textos anotados o traducidos. Estos resultados pueden ser reutilizados por terceras personas para enriquecerlos con nuevos procesamientos o para producir servicios o nuevas informaciones basados en ellos.

### 4.1 ¿Qué se puede hacer con los resultados del procesamiento del lenguaje?

Más allá de las aplicaciones (programas de ordenador), los resultados obtenidos por el procesamiento del lenguaje pueden ser objeto de reutilización para llevar a cabo nuevos procesamientos o simplemente para ofrecer informaciones o servicios basados en ellos.

En la medida en que estos resultados sean obtenidos por las Administraciones públicas u organismos sometidos al ámbito de aplicación de la LRISP que se ha expuesto en la primera parte, podrán ser objeto de reutilización de acuerdo con lo previsto en dicha norma.

Al respecto, la primera cuestión a destacar es que, tal y como se ha avanzado, la obligación de facilitar la reutilización de estos resultados viene definida en la LRISP de forma un tanto ambigua de

manera que, en última instancia, cada Administración pública que haya llevado el procesamiento podrá decidir si facilita o no la reutilización y en qué condiciones. En todo caso, como ya apuntábamos en el capítulo 2 (Recursos), la reutilización deberá salvaguardar la protección de la propiedad intelectual e industrial de terceros, así como los datos personales y los acuerdos de confidencialidad.

En el caso de que se facilite la reutilización, **la LRISP define tres modalidades de reutilización**, entre las cuales puede optar cada administración pública al definir el régimen jurídico específico de la reutilización de la información:

- Reutilización sin sujeción a condiciones.
- Reutilización con sujeción a condiciones establecidas en licencias tipo, que tienen que estar disponibles en formato digital y deben ser procesables de manera electrónica.
- Reutilización con solicitud previa, que puede incorporar condiciones en una licencia; esta licencia tiene que reflejar, al menos, la información relativa a la finalidad concreta, comercial o no comercial, por la cual se concede la reutilización, la duración de la licencia, las obligaciones del beneficiario y del organismo concedente, las responsabilidades de uso y las modalidades financieras, por lo cual se tiene que indicar el carácter gratuito o, si procede, la contraprestación económica aplicable.

Las condiciones que pueden tener las licencias deben ser claras, justas y transparentes; no restringir las posibilidades de reutilización ni limiten la competencia; no ser discriminatorias para categorías comparables de reutilización y no contener acuerdos exclusivos.

Estas condiciones se pueden referir a que:

- el contenido de la información, incluyendo sus metadatos, no sea alterado.
- no se desnaturalice el sentido de la información.
- se cite la fuente.
- se mencione la fecha de la última actualización.
- la finalidad o finalidades concretas para las que es posible la reutilización futura de los datos cuando la información contenga datos de carácter

**Las condiciones deben ser claras, justas y transparentes; no restringir las posibilidades de reutilización ni limitar la competencia ni ser discriminatorias o contener acuerdos exclusivos**

personal.

- la prohibición de revertir el procedimiento de disociación mediante la adición de nuevos datos obtenidos de otras fuentes cuando la información, aún siendo facilitada de forma disociada, contuviera elementos suficientes que pudieran permitir la identificación de los interesados en el proceso de reutilización.

## 4.2 Las licencias de propiedad intelectual

A parte de tener que cumplir con las condiciones de reutilización previstas en la ley, los resultados del procesamiento lingüístico (así como el propio programa) puede ponerse a disposición del público para su reutilización. En la medida en que todo ello puede ser objeto de protección por la propiedad intelectual, es conveniente establecer una licencia de PI que cubra tales aspectos.

Pueden utilizarse licencias como las *Creative Commons*, teniendo en cuenta que las más permisivas son las que mejor facilitan la reutilización (BY) mientras que las que no permiten los fines comerciales o la adaptación son las que menos contribuyen a la reutilización. También pueden utilizarse licencias creadas específicamente para el tipo de contenido y uso previsto; como contrapartida, una licencia “específica” tendría el inconveniente de no ser tan conocida (y visible) como las licencias CC. El titular deberá decidir cuál es la licencia que mejor se ajusta a la reutilización que quiere permitir.

- ¿Cuáles son las fronteras hasta las cuales se puede discriminar el acceso a infraestructuras lingüísticas generadas en el marco del Plan a empresas extranjeras?

En principio, es perfectamente posible discriminar entre diferentes tipos de usuarios (por ejemplo, solo permitir los usos para fines de investigación no comerciales, etc) y discriminar a empresas (con ánimo de lucro). En el ámbito de la UE, sin embargo, no será posible discriminar a otras empresas europeas y en la medida que casi todas las empresas extranjeras (no-UE) tienen alguna sede en la UE, sería de difícil aplicación.



## 5 Referencias

---

- [1] Grupo de trabajo sobre protección de datos del artículo 29, "Dictamen de 3/2013 sobre limitación de la finalidad," 2003.
- [2] Agencia Española de Protección de Datos, "Orientaciones y garantías en los procedimientos de anonimización de datos personales," 2016.
- [3] Grupo de Trabajo sobre protección de datos del artículo 29, "Dictamen 05/2014 sobre técnicas de anonimización," 2014.
- [4] Agencia Española de Protección de Datos, "Guía para una evaluación de impacto en la protección de datos," 2014.
- [5] Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información, "Plan de Impulso de las Tecnologías del Lenguaje," 2015.
- [6] Comisión Nacional de los Mercados y la Competencia, "Informe de la CNMC sobre las ayudas públicas en España -2016 IAP/CNMC/001/16," 2016.
- [7] B. Lundqvist, Y. Forsberg, M. de Vries, and M. Maggolino, "Open Data and Competition Law. Some issues regarding access and pricing of raw data," *Masaryk University Journal of Law and Technology*, vol. 9, 2015.
- [8] Ministerio de Economía y Competitividad, "Guía 2.0 para la compra pública de innovación," 2015.



## 6 Glosario de siglas y acrónimos

---

BBDD	Bases de datos
ENI	Real decreto 4/2010 de 8 de enero por el que se regula el esquema nacional de interoperabilidad
LM	Ley 17/2001, de 7 diciembre, de Marcas
LOPD	Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal
LP	Ley 24/2015, de 24 julio, de Patentes
LRISP	Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público
LTAIBG	Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno
NTI	Norma Técnica de Interoperabilidad de Reutilización de recursos de la información
PI	Propiedad intelectual
TDM	Text & Data Mining
TRLCSP	Real Decreto Legislativo 3/2011, de 14 de noviembre, por el que se aprueba el texto refundido de la Ley de Contratos del Sector Público
TRLPI	Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual; última modificación por Ley 21/2014 de 5 de noviembre
UE	Unión Europea

## 7 Referencias normativas

---

<sup>1</sup> Artículo 2 LRISP.

<sup>2</sup> Artículo 4 LRISP.

<sup>3</sup> Artículo 3 LRISP.

<sup>4</sup> Artículo 3 LRISP.

<sup>5</sup> Artículo 4 LRISP. Véase al respecto <http://datos.gob.es/>

<sup>6</sup> Artículo 7 LRISP.

<sup>7</sup> Véase en: <https://sunlightfoundation.com/wp-content/uploads/2016/11/Ten-Principles-for-Opening-Up-Government-Data.pdf> (última consulta: marzo 2017).

<sup>8</sup> Artículo 5.4 Ley 19/2013, de 9 de diciembre.

<sup>9</sup> Artículo 37 Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación.

<sup>10</sup> Artículo 3 Real Decreto 4/2010, de 8 de enero.

<sup>11</sup> Resolución de 19 de febrero de 2013 de la Secretaría de Estado de Administraciones Públicas.

<sup>12</sup> Última modificación por Ley 21/2014 de 5 de noviembre. Puede consultarse el texto normativo actualizado en <https://www.boe.es/legislacion/codigos/>

<sup>13</sup> Convenio de Berna para la protección de las obras literarias y artísticas (1886), el Tratado OMPI sobre Derecho de autor (Ginebra, 1996) y el ADPIC-Acuerdo sobre los Aspectos de la Propiedad Intelectual Relacionados con el Comercio (1994). Todos ellos son accesibles a través del web de la Organización Mundial de la Propiedad Intelectual (OMPI/WIPO) y de la Organización Mundial del Comercio (OMC/WTO): <http://www.wipo.int> y <http://www.wto.org/>.

<sup>14</sup> Directiva 91/250/CEE (derogada y codificada por la Directiva 2009/24/CE), sobre la protección de los programas de ordenador; Directiva 93/98/CEE (derogada y codificada por Directiva 2006/116/CE), sobre plazo de protección y modificada por la Directiva 2011/77/UE, sobre plazo de protección de los fonogramas; Directiva 96/9/CE, sobre bases de datos; Directiva 2001/29/CE, sobre derechos de autor y conexos en la sociedad de la información; Directiva 2012/28/UE, sobre obras huérfanas;

<sup>15</sup> Artículo 10 TRLPI.

<sup>16</sup> Artículo 13 TRLPI.

<sup>17</sup> Artículo 5 TRLPI.

<sup>18</sup> Artículo 51 TRLPI.

<sup>19</sup> Artículo 8 TRLPI.

<sup>20</sup> Artículo 8 TRLPI.

<sup>21</sup> Artículo 9 TRLPI.

<sup>22</sup> Artículo 11 TRLPI.

<sup>23</sup> Artículo 133 y ss. TRLPI.

<sup>24</sup> Artículo 18 TRLPI.

<sup>25</sup> Artículo 20 TRLPI.

<sup>26</sup> Artículo 21 TRLPI.

<sup>27</sup> Artículo 40 bis TRLPI.

<sup>28</sup> Artículo 31.2 TRLPI y artículo 37.1 TRLPI, respectivamente.

<sup>29</sup> Artículo 43 TRLPI.

<sup>30</sup> Artículo 301.2 TRLCSP.

<sup>31</sup> Artículo 98 TRLPI.