

MARCO NORMATIVO PARA LA IMPLEMENTACIÓN DEL PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE

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	Los recursos	9
2.1	Los recursos públicos de interés lingüístico	11
2.1.1	La reutilización de los recursos públicos de interés lingüístico.....	11
2.1.2	Los datos abiertos de interés lingüístico	21
2.1.3	La interoperabilidad de los recursos lingüísticos	25
2.2.4.	Condiciones para la reutilización de los recursos generados por el sector público.....	28
2.2	Los recursos privados de interés lingüístico.....	31
2.3	Regímenes que pueden restringir el uso de los recursos (públicos y privados)	32
2.3.1	La protección de la propiedad intelectual.....	32
2.3.2	La protección de los datos personales y la anonimización	57
2.3.3	La confidencialidad de la información.....	68
3	El proceso	72
3.1	El papel de las administraciones públicas en el desarrollo de procesadores del lenguaje natural72	
3.2	La contratación del desarrollo de los procesadores	75
3.3	La normativa lingüística.....	76
3.4	La interoperabilidad de las aplicaciones	77
3.5	La protección jurídica de las aplicaciones	79
4	El resultado.....	87
4.1	La reutilización de los resultados	87
4.2	Las licencias de propiedad intelectual.....	89
5	Referencias	94
6	Glosario de siglas y acrónimos	95



MARCO NORMATIVO PARA LA IMPLEMENTACIÓN DEL PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE

El presente estudio constituye la primera 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, la finalidad de este estudio es identificar las normas jurídicas que deben ser tenidas en cuenta en la aplicación de las medidas del Plan de Impulso de las Tecnologías del Lenguaje en relación a la creación, el uso de recursos lingüísticos y a su reutilización por parte de la industria de las tecnologías del lenguaje y de la propia Administración pública.

El marco normativo para la implementación del Plan de Impulso de las Tecnologías del Lenguaje 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 uso de las infraestructuras lingüísticas genera diversas cuestiones desde el punto de vista jurídico a las que se debe dar respuesta para facilitar su desarrollo y uso. Asimismo, la normativa también puede facilitar el impulso de las infraestructuras lingüísticas.

En las próximas páginas se analizarán las normas que deben ser tenidas en cuenta al impulsar el diseño, desarrollo y uso de las infraestructuras lingüísticas. 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 diferentes actores en el marco del Plan de Impulso de las Tecnologías del Lenguaje. En efecto, más allá de concretar qué y cómo se pueden utilizar las tecnologías del lenguaje, el marco jurídico identifica el rol de los diferentes actores y, en particular, de las administraciones públicas.

El marco normativo determina el alcance de las aplicaciones de lenguaje natural y define los límites de lo que se puede hacer y lo que no. Sin embargo, también puede ser un instrumento que, más allá del desarrollo tecnológico, promueva el impulso de las tecnologías del lenguaje en España.

En esta dirección, el conocimiento y la interpretación del ordenamiento jurídico constituye un elemento primordial en el desarrollo del Plan de Impulso de las Tecnologías del Lenguaje. Por ello, a lo largo de las próximas páginas se analizará los aspectos jurídicos que deben tenerse en cuenta a fin de poder desarrollar las infraestructuras lingüísticas (eje I), impulsar la industria de las tecnologías del lenguaje (eje II) y fortalecer el papel de la Administración pública como impulsora de la industria del lenguaje (eje III).

Para ello, se identificarán, sistematizarán y analizarán las distintas normas que de manera directa o indirecta inciden en el desarrollo de las diferentes acciones y en la consecución de los distintos objetivos previstos en el Plan de Impulso de las Tecnologías del Lenguaje. Para ello también se tendrán en cuenta las directrices, recomendaciones y buenas prácticas que puedan ser de aplicación en la ejecución del Plan. Finalmente, se tendrán en cuenta los informes que a nivel europeo se han impulsado en esta materia.

La complejidad de las infraestructuras lingüísticas determina la diversidad de normas jurídicas que deben ser tenidas en cuenta a fin de conocer los diferentes diseños y usos 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.

A modo de introducción, se pueden identificar las siguientes normas que han sido analizadas a lo largo de las siguientes páginas.

RECURSOS		
PÚBLICOS	REUTILIZACIÓN Y DATOS ABIERTOS	<p>Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público (LRISP)</p> <p>Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP, sobre reutilización de la información del sector público, para el ámbito del sector público estatal</p> <p>Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación</p> <p>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 (TRLPI)</p>
	INTEROPERABILIDAD	<p>Real decreto 4/2010 de 8 de enero por el que se regula el esquema nacional de interoperabilidad (ENI)</p> <p>Norma Técnica de Interoperabilidad de Reutilización de recursos de la información (NTI)</p>

PRIVADOS	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	
RESTRICCIÓN DEL USO DE RECURSOS		
Protección propiedad intelectual	Protección de datos personales	Confidencialidad
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	Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal (LOPD) Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo	Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG) Ley 15/2007, de 3 de julio, de Defensa de la Competencia 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
PROCESO		
CONTRATACION	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	
NORMATIVA LINGÜÍSTICA	Ley 1/1998, de 7 de enero, de Política Lingüística Catalán Ley 3/1983, de 15 de junio, de normalización lingüística del Gallego Ley 3/1986, de 29 de abril, de normalización lingüística en las Illes Balears Ley 10/1982, de 24 de noviembre, básica de normalización del uso del euskera	
INTEROPERABILIDAD	Ley 40/2015, de 1 de octubre, de Régimen Jurídico 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	
PROTECCIÓN DE LAS APLICACIONES (software)	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 Ley 24/2015, de 24 julio, de Patentes (LP) Ley 17/2001, de 7 diciembre, de Marcas (LM)	
ACTOS DE TRATAMIENTO	Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el	



LINGÜÍSTICO	texto refundido de la Ley de Propiedad Intelectual; última modificación por Ley 21/2014 de 5 de noviembre
RESULTADOS	
REUTILIZACIÓN ISP	Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público
REUTILIZACIÓN PI	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
OTRAS RESTRICCIONES	Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal (LOPD) Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG) Ley 15/2007, de 3 de julio, de Defensa de la Competencia 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

2 Los recursos

Cualquier procesamiento del lenguaje natural se basa en el uso de recursos diversas procedencias. En particular, las tecnologías de procesamiento de lenguaje natural 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 [1, 8].

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 y usuarios 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 qué finalidad y bajo qué condiciones.

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 natural. Asimismo, las entidades privadas, con o sin ánimo de lucro, también ponen a disposición de desarrolladores y usuarios recursos necesario 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. Sin embargo, no se puede desconocer el hecho que en la medida que en la mayoría de las ocasiones la información tenga su origen en las Administraciones públicas también se deberá tener presente la regulación de la reutilización de la información pública y de la apertura de datos públicos. Finalmente, como se observará, 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.

Despejar las incógnitas sobre cómo se pueden utilizar o reutilizar determinados recursos tiene una importancia significativa a fin de poder disponer de grandes volúmenes de datos que permitan llevar a cabo adecuadamente los procesamientos de lenguaje natural. De este modo, se podrá dar respuesta a la limitación señalada por el Plan, “la carencia de recursos para el procesamiento del lenguaje natural en una lengua particular son la barrera más importante para el desarrollo en Europa de aplicaciones de la industria del lenguaje” [2, 12].

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 será libre para cualquier usuario y para llevar a cabo cualquier uso. En otras ocasiones, si bien se facilitará 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 limitados 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 natural.

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.”

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.

A través la apertura de estos datos y el fomento de su reutilización, las Administraciones públicas pueden poner a disposición de desarrolladores y usuarios de las tecnologías de procesamiento del lenguaje natural, 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 *La reutilización de 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 y los usuarios de las tecnologías del lenguaje con el fin de ofrecer bienes y servicios al mercado. En efecto, tal y como señala el Plan de Impulso de las Tecnologías del Lenguaje la información que genera o está en poder de las Administraciones públicas como los textos legales y contractuales, las traducciones juradas, los diccionarios terminológicos especializados en dominios específicos, los nombres de entidades (personas, topónimos, organizaciones, empresas, publicaciones, ...), las terminologías, los corpus paralelos de traducción, etc. tienen un valor muy importante para el desarrollo de la industria del procesamiento del lenguaje. 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). Esta norma, modificada por la Ley 18/2015, de 9 de julio, transpone al ordenamiento jurídico español la Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre, de reutilización de la información del sector público modificada por la Directiva 2013/37/UE. Asimismo, en el ámbito de la Administración General del Estado, la reutilización de la información está regulada por el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP, sobre reutilización de la información del sector público, para el ámbito del sector público estatal.

La LRISP es de aplicación a las administraciones y organismos del sector público. En particular, se entiende por administraciones y organismos del sector público las siguientes:

- a) La Administración General del Estado, las Administraciones de las Comunidades Autónomas y las Entidades que integran la Administración local.
- b) Las Entidades gestoras y los servicios comunes de la Seguridad Social.
- c) Los Organismos autónomos, las Agencias estatales y cualesquiera entidades de derecho público que, con independencia funcional o con una especial autonomía reconocida por la ley, tengan atribuidas funciones de regulación o control de carácter externo sobre un determinado sector o actividad.
- d) Las entidades de derecho público con personalidad jurídica propia, vinculadas o dependientes de cualquiera de las Administraciones públicas que hayan sido creadas para satisfacer específicamente necesidades de interés general que no tengan carácter industrial o mercantil o que se trate de entidades cuya actividad esté mayoritariamente



financiada por las Administraciones públicas u otras entidades de derecho público, o bien, cuya gestión se halle sometida a un control por parte de estas últimas, o cuyos órganos de administración, de dirección o de vigilancia estén compuestos por miembros de los que más de la mitad sean nombrados por las Administraciones públicas y otras entidades de derecho público.

e) Los Consorcios dotados de personalidad jurídica.

f) Las fundaciones del sector público.

g) Las asociaciones constituidas por las Administraciones, organismos y entidades mencionados en las letras anteriores.

Sin embargo, la lectura del ámbito subjetivo de aplicación de la LRISP permite observar que no es de aplicación a 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. Así, por ejemplo, la LRISP no es de aplicación a entidades como la Real Academia de la Lengua que produce grandes cantidades de recursos de gran importancia para el procesamiento del lenguaje natural, respecto a la que el Real Decreto 1109/1993, de 9 de julio, por el que se aprueba los Estatutos de la Real Academia Española define como institución con personalidad jurídica propia. Tampoco, lo es al Centro de Documentación Judicial (CENDOJ) respecto a la reutilización de la información judicial y, en particular, de las sentencias judiciales.

Sin embargo, ello no quiere decir que estos organismos o entidades que están excluidos de la aplicación de la Ley 37/2007, de 16 de noviembre no puedan facilitar la reutilización de la información que producen o está en su poder. De hecho, en el caso de las sentencias y resoluciones judiciales, la propia LRISP dispone que las previsiones contenidas en ella son de aplicación a las sentencias y resoluciones judiciales, sin perjuicio de lo previsto en la legislación relativa al poder judicial.

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 (en relación, con lo previsto en el considerando 8 Directiva 2013/37/UE). Sin embargo, una lectura atenta del artículo 4 permite observar ciertas limitaciones en el alcance de dicha obligación. En efecto, en dicho artículo se dispone que "los documentos de las administraciones públicas y organismos del sector público serán reutilizables en los términos que prevé esta Ley", y se remite a los regímenes establecidos en el mismo artículo, que son determinados por cada Administración pública. Así pues, el alcance de esta



obligación de facilitar la reutilización viene determinado en el mismo precepto que prevé que “las Administraciones Públicas y organismos del sector público velarán porque los documentos a los que se aplica esta normativa puedan ser reutilizados para fines comerciales o no comerciales”. De este modo, la pretendida obligación de las Administraciones públicas de facilitar la reutilización de la información que está en su poder puede quedar ciertamente limitada a pesar de que la finalidad tanto de la Directiva 2013/37/UE como de la propia LRISP.

De todos modos, debe tenerse en cuenta la regulación de la reutilización de cada Administración pública. En esta dirección, 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 a que se hace referencia en el artículo 1.2 autorizarán la reutilización de los documentos elaborados o custodiados por ellos e incluidos en el ámbito de aplicación de este real decreto” y “únicamente podrá denegarse motivadamente la reutilización de los documentos si concurre alguno de los supuestos establecidos en el apartado 3 del artículo 3 de la Ley 37/2007, de 16 de noviembre”.

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. Este concepto incluye asimismo los recursos que puedan contener el archivo electrónico único previsto en el artículo 17 de la Ley 39/2015, de 1 de octubre, de procedimiento administrativo común de las Administraciones públicas y el artículo 46 Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público en el que se incluyen los documentos electrónicos que correspondan a procedimientos finalizados siempre y cuando no sean de aplicación ninguna de las limitaciones o exclusiones a las que se hace referencia a lo largo de este análisis y se cumplan con los requisitos de accesibilidad que fije la normativa.

Sin embargo, la LRISP excluye del régimen de reutilización previsto los documentos que no sean accesibles, aquellos cuyo conocimiento requiera una legitimación especial y los contenidos públicos creados, por ejemplo, por entidades de radiodifusión o entidades educativas.

En principio, toda la información en poder de estas Administraciones y organismos públicos es susceptible de ser reutilizada. En efecto, la LRISP se aplica a los documentos elaborados o custodiados por las Administraciones y organismos del sector público, cuya reutilización no esté expresamente limitada por ellos. No obstante, no a toda la información que tienen en su poder estas



Administraciones u organismos públicos se aplica la LRISP y, por lo tanto, podrán ser objeto de reutilización. En particular, no es de aplicación esta norma a:

- a) Los documentos sobre los que existan prohibiciones o limitaciones en el derecho de acceso en virtud de lo previsto en la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno y las demás normas que regulan el derecho de acceso o la publicidad registral con carácter específico.
- b) los documentos que afecten a la defensa nacional, la seguridad del Estado, la protección de la seguridad pública, así como los obtenidos por la Administración Tributaria y la Administración de la Seguridad Social en el desempeño de sus funciones, los sometidos al secreto estadístico, a la confidencialidad comercial, tales como secretos comerciales, profesionales o empresariales y, en general, los documentos relacionados con actuaciones sometidas por una norma al deber de reserva, secreto o confidencialidad.
- c) Los documentos para cuyo acceso se requiera ser titular de un derecho o interés legítimo.
- d) Los documentos que obran en las Administraciones y organismos del sector público para finalidades ajenas a las funciones de servicio público, de acuerdo con la legislación aplicable y en particular, con la normativa de creación del servicio público de que se trate
- e) Los documentos sobre los que existan derechos de propiedad intelectual o industrial por parte de terceros
- f) Los documentos conservados por las entidades que gestionen los servicios esenciales de radiodifusión sonora y televisiva y sus filiales
- g) Los documentos producidos o conservados por instituciones educativas y de investigación (incluidas las organizaciones para la transferencia de los resultados de la investigación, centros escolares y universidades, exceptuando las bibliotecas universitarias) así como los museos y archivos estatales como agentes de ejecución del Sistema Español de Ciencia, Tecnología e Innovación siempre que sean resultado de una investigación.



- h) Los documentos producidos o conservados por instituciones culturales que no sean bibliotecas, incluidas las universitarias, museos y archivos
- i) Las partes de documentos que solo incluyan logotipos, divisas e insignias
- j) Los documentos a los que no pueda accederse o cuyo acceso esté limitado en virtud de regímenes de acceso por motivos de protección de los datos personales, de conformidad con la normativa vigente y las partes de documentos accesibles en virtud de dichos regímenes que contengan datos personales cuya reutilización se haya definido por ley como incompatible con la legislación relativa a la protección de las personas físicas con respecto al tratamiento de los datos personales
- k) Los documentos elaborados por entidades del sector público empresarial y fundacional en el ejercicio de las funciones atribuidas legalmente y los de carácter comercial, industrial o mercantil elaborado en ejecución del objeto social previsto en sus Estatutos
- l) Los estudios realizados por entidades del sector público en colaboración con el sector privado, mediante convenios o cualquier otro tipo de instrumento, como fórmula de financiación de los mismos.

Por lo tanto, en términos generales se debe tener en cuenta que la LRISP no es de aplicación cuando la documentación esté incurso en los límites del derecho de acceso a la información pública, y en particular, a la protección de datos personales o a la propiedad intelectual. 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. Al margen de los mecanismos previstos en la legislación para la reutilización a los que se hará referencia en la última parte de este informe, en este punto se puede avanzar que la LRISP no dispone con carácter básico a través de que canal se deben facilitar los datos.



Sí lo hace el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP, para el ámbito del sector público estatal. En particular, el reglamento de desarrollo de la LRISP prevé que “los órganos de la Administración General del Estado y los demás organismos y entidades a que se hace referencia en el artículo 1.2 informarán de manera estructurada y usable, preferentemente a través de un espacio dedicado de su sede electrónica con la ubicación «sede.gob.es/datosabiertos», sobre qué documentación es susceptible de ser reutilizada, los formatos en que se encuentra disponible, las condiciones aplicables a su reutilización, indicando la fecha de la última actualización de los documentos reutilizables, proporcionando, cuando esté disponible, la información complementaria precisa para su comprensión y procesamiento automatizado y facilitando al máximo la identificación, búsqueda y recuperación de los documentos disponibles para su reutilización mediante mecanismos tales como listados, bases de datos o índices de información reutilizable”.

La lectura de este precepto permite 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. Esta previsión definida con carácter general únicamente para los datos abiertos de la Administración General del Estado puede ser trasladada a otras Administraciones públicas a través de las respectivas normas reguladoras o de la adopción de planes o políticas de datos abiertos que contemplen esta opción como ya sucede en algunas Comunidades Autónomas. Asimismo, esta declaración general no es obstáculo para que se promueva la creación de portales específicos para desarrolladores o usuarios de las tecnologías del lenguaje.

Junto la creación de portales de datos, también puede ser de utilidad poner a disposición de desarrolladores y usuarios de las tecnologías del lenguaje catálogos que integren los datos o recursos públicos reutilizables. En esta dirección, la propia LRISP ya dispone que “la Secretaría de Estado para la Función Pública del Ministerio de Política Territorial y Administración Pública y la Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información del Ministerio de Industria, Turismo y Comercio mantendrán un catálogo de información pública reutilizable correspondiente, al menos, a la Administración General del Estado y a los demás organismos y entidades a que se refiere el artículo 1.2, que permita acceder, desde un único punto, a los distintos recursos de información pública reutilizable disponibles”. Nuevamente, cabe entender que esta previsión genérica no es obstáculo para que se puedan crear catálogos específicos que recojan y sistematicen los recursos a disposición de desarrolladores y usuarios de las tecnologías del lenguaje.

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 natural 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.

Finalmente, y al margen del régimen de la reutilización al que se hará mención en la última parte del informe, en este punto cabe señalar que la LRISP establece la posibilidad de que las administraciones puedan aplicar una tasa o un precio público por el suministro de documentos para su reutilización. Las tarifas por la reutilización deben corresponderse con los costes marginales en que se incurra para la reproducción, puesta a disposición y difusión de la información.

Las Administraciones públicas deben informar adecuadamente de las tasas o precios públicos que eventualmente se fijen. Así lo dispone la LRISP al prever que “las Administraciones y organismos del sector público publicarán por medios electrónicos, siempre que sea posible y apropiado, las tarifas fijadas para la reutilización de documentos que estén en poder de organismos del sector público, así como las condiciones aplicables y el importe real de los mismos, incluida la base de cálculo utilizada”.

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 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”.

En esta dirección, se debe tener en cuenta que la LRISP prevé la posibilidad de que los precios que deban satisfacer los reutilizadores sean distintos según si la reutilización tenga carácter comercial o no comercial. De este modo, las Administraciones públicas al ofrecer sus recursos lingüísticos podrían definir una política de precios que, en el marco de los criterios y condiciones fijados en la LRISP, facilitase su uso por parte de otras Administraciones públicas o de centros de investigación al establecerse la gratuidad de la reutilización. En cambio, fijar un precio para la reutilización de la misma información con finalidades comerciales por parte de la industria de las tecnologías del lenguaje.

Más allá de lo previsto en la LRISP, resulta útil, sino necesario, que las Administraciones públicas informen adecuadamente a los reutilizadores de las condiciones a las que se somete la reutilización. En esta dirección, las Administraciones públicas 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. Gracias al aviso legal los usuarios de los recursos lingüísticos pueden tener certeza del alcance de sus derechos y obligaciones sobre los mismos y, en última instancia, la reutilización se haga con la mayor seguridad jurídica. En última instancia, el uso de un aviso legal para informar de estos extremos puede generar confianza en los reutilizadores.

En la actualidad, es frecuente que los portales de transparencia y los portales de datos abiertos contengan avisos legales que, con mayor o menor detalle, informan sobre las condiciones de la reutilización.

En esta dirección, el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP, para el ámbito del sector público estatal, dispone que las “condiciones generales serán accesibles mediante un aviso legal por medios electrónicos, de forma permanente, fácil y directa, preferentemente dentro de la ubicación «sede.gob.es/datosabiertos» de la sede electrónica del órgano de la Administración General del Estado, organismo o entidad correspondiente, y vincularán a cualquier agente reutilizador por el mero hecho de hacer uso de los documentos sometidos a ellas”.

En particular, el Real Decreto 1495/2011 recoge en su anexo el siguiente aviso legal:

Obligatoriedad de las condiciones generales.

Las presentes condiciones generales, disponibles con carácter permanente bajo “www.datos.gob.es/avisolegal”, vincularán a cualquier agente reutilizador por el mero hecho de hacer uso de los documentos sometidos a ellas.

Autorización de reutilización y cesión no exclusiva de derechos de propiedad intelectual.

Las presentes condiciones generales permiten la reutilización de los documentos sometidos a ellas para fines comerciales y no comerciales. Se entiende por reutilización el uso de documentos que obran en poder de los órganos de la Administración General del Estado y los demás organismos y entidades del sector público estatal referidos en el artículo 1.2 del Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público estatal, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública. La reutilización autorizada incluye, a modo ilustrativo, actividades como la copia, difusión, modificación, adaptación, extracción, reordenación y combinación de la información.

El concepto de documento es el establecido en el apartado 2 del artículo 3 de la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, por lo que comprende toda

información cualquiera que sea su soporte material o electrónico así como su forma de expresión gráfica, sonora o en imagen utilizada, incluyendo, en consecuencia, también los datos en sus niveles más desagregados o “en bruto”.

Esta autorización conlleva, asimismo, la cesión gratuita y no exclusiva de los derechos de propiedad intelectual, en su caso, correspondientes a tales documentos, autorizándose la realización de actividades de reproducción, distribución, comunicación pública o transformación, necesarias para desarrollar la actividad de reutilización autorizada, en cualquier modalidad y bajo cualquier formato, para todo el mundo y por el plazo máximo permitido por la Ley.

Condiciones generales para la reutilización.

Son de aplicación las siguientes condiciones generales para la reutilización de los documentos sometidos a ellas:

1. Está prohibido desnaturalizar el sentido de la información.

2. Debe citarse la fuente de los documentos objeto de la reutilización. Esta cita podrá realizarse de la siguiente manera: “Origen de los datos: *[órgano administrativo, organismo o entidad del sector público estatal de que se trate]*”.

3. Debe mencionarse la fecha de la última actualización de los documentos objeto de la reutilización, siempre cuando estuviera incluida en el documento original.

4. No se podrá indicar, insinuar o sugerir que la *[órgano administrativo, organismo o entidad del sector público estatal de que se trate]* titular de la información reutilizada participa, patrocina o apoya la reutilización que se lleve a cabo con ella.

5. Deben conservarse, no alterarse ni suprimirse los metadatos sobre la fecha de actualización y las condiciones de reutilización aplicables incluidos, en su caso, en el documento puesto a disposición para su reutilización.

Exclusión de responsabilidad.

La utilización de los conjuntos de datos se realizará por parte de los usuarios o agentes de la reutilización bajo su propia cuenta y riesgo, correspondiéndoles en exclusiva a ellos responder frente a terceros por daños que pudieran derivarse de ella.

[El órgano administrativo, organismo o entidad del sector público estatal de que se trate] no será responsable del uso que de su información hagan los agentes reutilizadores ni tampoco de los daños sufridos o pérdidas económicas que, de forma directa o indirecta, produzcan o puedan producir perjuicios económicos, materiales o sobre datos, provocados por el uso de la información reutilizada.

[El órgano administrativo, organismo o entidad del sector público estatal de que se trate] no garantiza la continuidad en la puesta a disposición de los documentos reutilizables, ni en contenido ni en forma, ni asume responsabilidades por cualquier error u omisión contenido en ellos.

Responsabilidad del agente reutilizador

El agente reutilizador se halla sometido a la normativa aplicable en materia de reutilización de la información del sector público, incluyendo el régimen sancionador previsto en el artículo 11 de la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.»

2. Con el objetivo de informar a los motores y sistemas automatizados de búsqueda en Internet, se incorporarán además en la codificación de la citada ubicación los mecanismos de localización de información pública reutilizable que se estimen oportunos. Para ello, si bien se podrán utilizar otras modalidades técnicas, se propone el siguiente comando básico, que enlaza con las condiciones generales de reutilización:

```
<a href="http://www.datos.gob.es/avisolegal" rel="license">Aviso legal</a>
```

o bien el comando

```
<a href="http://www.datos.gob.es/avisolegal">Aviso legal</a>
```

El aviso legal se debe adaptar a las condiciones específicas de cada tipología de recursos lingüísticos que se difundan.

2.1.2 *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.

La apertura de los datos públicos tiene un impacto significativo en la rendición de cuentas de las Administraciones públicas y en el incremento de la transparencia pública. Asimismo, los datos abiertos tienen un impacto importante para favorecer la creación de servicios e informaciones basados en la reutilización. En esta dirección se ha manifestado claramente la Comisión europea en su Comunicación al Parlamento, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre *Datos abiertos. Un motor para la innovación, el crecimiento y la gobernanza transparente* [COM(2011), 882 final] donde se destaca que "[...] el impulso hacia la utilización de datos abiertos está cobrando fuerza en varios Estados miembros, los cuales están adoptando este concepto por razones de transparencia, eficiencia administrativa y potencial económico de la 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.¹

¹ Véase en: <https://sunlightfoundation.com/wp-content/uploads/2016/11/Ten-Principles-for-Opening-Up-Government-Data.pdf> (última consulta: marzo 2017).

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.).

El artículo 5 de la Directiva 2003/98/CE relativa a la reutilización de la información del sector público –modificada en este punto por la Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013– prevé que "[...] los organismos del sector público facilitarán sus documentos en cualquier formato o lengua en que existan previamente y, siempre que sea posible y apropiado, en formato legible por máquina y conjuntamente con sus metadatos"

En la actualidad, la LRISP después de la modificación operada por la Ley 18/2015, de 9 de julio dispone que

“2. Las Administraciones y organismos del sector público facilitarán 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 conforme a lo previsto en el apartado anterior y conjuntamente con sus metadatos, con los niveles más elevados de precisión y desagregación. Tanto el formato como los metadatos, en la medida de lo posible, deben cumplir estándares y normas formales abiertas”.

De este modo, la LRISP prevé que cuando sea posible las Administraciones públicas deben facilitar los datos abiertos para facilitar su reutilización. De todos modos, a pesar de la dicción de este precepto 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.

Además de la ausencia de una obligación en esta dirección, la propia LRISP establece que las Administraciones y organismos del sector público no están obligados a crear documentos, adaptarlos o facilitar extractos de documentos, cuando ello suponga un esfuerzo desproporcionado que conlleve algo más que una simple manipulación (artículo 5.2).

De todos modos, sí que debe tenerse presente que cuando las administraciones públicas difunden datos en aplicación de lo previsto en la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, lo deben hacer en formatos reutilizables. Esta norma de carácter básico dispone que “la información sujeta a las obligaciones de transparencia será publicada



en las correspondientes sedes electrónicas o páginas web y de una manera clara, estructurada y entendible para los interesados y, preferiblemente, en formatos reutilizables. Se establecerán los mecanismos adecuados para facilitar la accesibilidad, la interoperabilidad, la calidad y la reutilización de la información publicada así como su identificación y localización” (artículo 5.4). En particular, por lo que respecta al portal de transparencia de la Administración General del Estado, se fomentará que la información sea publicada en formatos que permitan su reutilización (artículo 11.c).

En este marco encajan a la perfección las medidas previstas en el Plan de Impulso de las Tecnologías de la Información al prever que en el marco de la política de reutilización de la información del sector público deben introducir el concepto de datos abiertos de interés lingüístico y difundir este concepto en la Administración pública.

Las normas autonómicas de transparencia han sido más explícitas en relación al impulso de los datos abiertos lo que puede tener un impacto positivo a la hora de promover los datos abiertos lingüísticos.

En efecto, diversas normas autonómicas prevén que las administraciones públicas deberán promover la difusión de los datos públicos en formatos abiertos (Ley 4/2016, de 15 de diciembre, de Transparencia y Buen Gobierno de Castilla-La Mancha, Ley 1/2016, de 18 de enero, de transparencia y buen gobierno de Galicia, Ley 8/2015, de 25 de marzo, de Transparencia de la Actividad Pública y Participación Ciudadana de Aragón, Ley 12/2014, de 16 de diciembre, de Transparencia y Participación Ciudadana de la Comunidad Autónoma de la Región de Murcia, Ley 1/2014, de 24 de junio, de Transparencia Pública de Andalucía, Ley 4/2013, de 21 de mayo, de Gobierno Abierto de Extremadura). Asimismo, disponen la creación de portales de datos abiertos o de secciones específicas al efecto en los portales de transparencia con la finalidad de favorecer la reutilización (Ley 4/2016, de 15 de diciembre, de Transparencia y Buen Gobierno de Castilla-La Mancha, Ley 8/2015, de 25 de marzo, de Transparencia de la Actividad Pública y Participación Ciudadana de Aragón). También prevén la creación de catálogos de información reutilizable (Ley 1/2016, de 18 de enero, de transparencia y buen gobierno de Galicia, Ley 3/2014, de 11 de septiembre, de Transparencia y Buen Gobierno de La Rioja).

Tal y como se expondrá en la última parte de este informe, las Administraciones públicas pueden establecer determinadas condiciones que deben regir la reutilización de la información. Estas condiciones deben ser claras, justas y transparentes; no pueden restringir las posibilidades de reutilización ni limitar la competencia y además no pueden ser discriminatorias para categorías comparables de reutilización. En esta dirección, la LRISP recoge algunas condiciones que se pueden establecer (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 personal o la prohibición de revertir el procedimiento de disociación).

En esta dirección, las Administraciones públicas podrían también establecer como condición que los recursos que se han ofrecido en abierto mantengan dicho carácter una vez hayan sido procesados.

Las universidades y los centros de investigación producen grandes cantidades de recursos que pueden ser de utilidad para el procesamiento del lenguaje natural. 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.

Artículo 37. Difusión en acceso abierto.

1. 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.

2. El personal de investigación cuya actividad investigadora esté financiada mayoritariamente con fondos de los Presupuestos Generales del Estado hará pública una versión digital de la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadadas o periódicas, tan pronto como resulte posible, pero no más tarde de doce meses después de la fecha oficial de publicación.

3. La versión electrónica se hará pública en repositorios de acceso abierto reconocidos en el campo de conocimiento en el que se ha desarrollado la investigación, o en repositorios institucionales de acceso abierto.

En efecto, como se desprende de la lectura de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación, los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación deben facilitar el acceso abierto a las publicaciones de su personal de investigación.

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. 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.

En esta dirección, resulta de interés traer a colación lo previsto en la Ley 8/2015, de 25 de marzo, de Transparencia de la Actividad Pública y Participación Ciudadana de Aragón “los nuevos conjuntos de datos generados por la formalización de contratos, convenios, acuerdos, conciertos, subvenciones u otras figuras jurídicas en las que se plasme la relación del Gobierno de Aragón y otras entidades se dispondrán como conjunto de datos abiertos, siempre que sea posible. Para ello se fomentará que dichos instrumentos jurídicos contengan una cláusula “open data” que garantice la efectiva liberación de los conjuntos de datos”.

2.1.3 La interoperabilidad de 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 intercambio de información y conocimiento entre ellos (Real decreto 4/2010 de 8 de enero por el que se regula el esquema nacional de interoperabilidad). La interoperabilidad tiene diferentes manifestaciones en relación al desarrollo y uso de las infraestructuras lingüísticas.

Al margen de que posteriormente se exponga como la interoperabilidad se traslada a la reutilización de los procesadores lingüísticos que puedan desarrollar las administraciones públicas, la interoperabilidad de los recursos lingüísticos es un aspecto de gran relevancia para facilitar el desarrollo del procesamiento del lenguaje natural. 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 establecer unas normas, criterios y recomendaciones que permitan a las aplicaciones utilizadas por las administraciones públicas intercambiar información y utilizar mutuamente la información intercambiada.



La interoperabilidad tiene una importancia para facilitar la reutilización. Así, lo ha puesto de manifiesto la Directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 que ha modificado en este extremo la Directiva de 2003 para prever que los organismos del sector público facilitarán sus documentos en cualquier formato o lengua en que existan previamente que en la medida de lo posible, deben cumplir normas formales abiertas, es decir, que especifique los criterios de interoperabilidad de la aplicación informática (artículo 5).

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.

El Esquema Nacional de Interoperabilidad prevé que su desarrollo se debe realizar a través de la adopción de normas técnicas de interoperabilidad que serán de obligado cumplimiento por parte de las Administraciones públicas. En particular, el Esquema Nacional de Interoperabilidad prevé la aprobación de una norma técnica relativa a la reutilización de recursos de información que tratará de las normas comunes sobre la localización, descripción e identificación unívoca de los recursos de información puestos a disposición del público por medios electrónicos para su reutilización.

En la actualidad, la Resolución de 19 de febrero de 2013 de la Secretaría de Estado de Administraciones Públicas aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información. Esta Norma técnica tiene por objeto establecer 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. Tal y como recoge el preámbulo de dicha norma técnica, “estas condiciones tienen el objetivo de facilitar y garantizar el proceso de reutilización de la información de carácter público procedente de las Administraciones públicas, asegurando la persistencia de la información, el uso de formatos así como los términos y condiciones de uso adecuados”.

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:



- 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.

2.2.4. Condiciones para la reutilización de los recursos generados por el sector público.

Los recursos de interés lingüístico que pueden ser objeto de utilización para realizar el tratamiento pueden ser recursos generados o custodiados por el sector público (Administración y organismos públicos) que han sido objeto de puesta a disposición del público en condiciones de reutilización. Por ejemplo, imaginemos que se utilizan sentencias y resoluciones judiciales, normativa (leyes, reglamentos, órdenes, etc), memorias publicadas por alguna Administración pública, contenidos disponibles en las páginas web de tales instituciones, grabaciones de sesiones parlamentarias o plenos de la administración local, etc. Otro ámbito de recursos de interés lingüístico puede encontrarse en aquellos documentos (libros, cartas, etc) custodiados por instituciones culturales, archivos, bibliotecas, museos de titularidad pública y que son objeto de difusión ya sea en abierto a través de su página web, o lícitamente obtenidos de alguna otra manera. Para su uso, deberemos estar a las condiciones de reutilización impuestas sobre tales recursos ya que, todos ellos quedan sujetos a la normativa de RISP (reutilización de la información del sector público) y a las condiciones que allí se establecen.

El ámbito de aplicación de la normativa RISP es amplio y alcanza a todos los contenidos (documentos) “que **obran en poder** de las Administraciones y organismos del sector público” (Art.3.1 LRISP) incluidos “los **documentos elaborados o custodiados** por las Administraciones y organismos del sector público, cuya reutilización no esté expresamente limitada por estos.” (Art.3.2 LRISP), y con independencia del formato: escrito en papel, almacenado en forma electrónica o como grabación sonora, visual o audiovisual.

También es amplia la definición de reutilización: el uso de tales documentos “por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública.” Quedando excluido “el intercambio de documentos entre Administraciones y organismos del sector público en el ejercicio de las funciones públicas que tengan atribuidas” (Art.3.1 in fine LRISP).

Pero estas amplias definiciones se completan con un listado de exclusiones en el Art.3.3 LRISP, entre las que se incluyen: “los documentos (que obren en poder de las Administraciones y organismos del sector público) sobre los que existan derechos de **propiedad intelectual o industrial por parte de terceros**” (Art.3.3e) LRISP). Sin embargo, este artículo continúa:

No obstante, la presente Ley no afecta a la existencia de derechos de propiedad intelectual de las Administraciones y organismos del sector público ni a su posesión por éstos, ni restringe el ejercicio de esos derechos fuera de los límites establecidos por la

presente Ley. El ejercicio de los derechos de propiedad intelectual de las Administraciones y organismos del sector público deberá realizarse de forma que se facilite su reutilización.

Lo previsto en el párrafo anterior será de aplicación, asimismo, a los documentos respecto de los que las bibliotecas, incluidas las universitarias, los museos y los archivos sean titulares originarios de los derechos de propiedad intelectual como creadores de la misma conforme a lo establecido en la legislación de propiedad intelectual, así como cuando sean titulares porque se les haya transmitido la titularidad de los derechos sobre dicha obra según lo dispuesto en la citada legislación, debiendo en este caso respetar lo establecido en los términos de la cesión.”.

En otras palabras, la Administración puede ser titular de derechos de propiedad intelectual y, como tal, podrá decidir si sujetar tales contenidos (de su titularidad) a la reutilización, estableciendo las condiciones de tal reutilización. Para saber qué contenidos están protegidos por el régimen de propiedad intelectual (y cuales no), y con qué alcance, nos remitiremos a tal normativa (TRLPI). Ahora bien, la normativa de RISP impone a la Administración la obligación de ejercer sus derechos de propiedad intelectual de manera que favorezca su reutilización.

La LRISP establece diversas condiciones para la reutilización de contenidos elaborados (o custodiados) por la Administración pública. Además de las condiciones concretas de reutilización previstas en el Art.8 LRISP (y que ya hemos tenido ocasión de ver), se exige que la reutilización sea abierta a todos los agentes potenciales del mercado, se establece la prohibición de autorizar la reutilización mediante licencias exclusivas - salvo que ello sea necesario para la prestación de un servicio público- y que las tarifas aplicadas por el suministro de documentos para su reutilización se limiten a los costes marginales incurridos para hacerla posible. El Real Decreto de desarrollo LRISP viene a recoger tales obligaciones:

Artículo 10. Ejercicio de los derechos de propiedad intelectual de titularidad de los órganos administrativos, organismos o entidades del sector público estatal.

1. De acuerdo con lo establecido en el artículo 3.3.e) de la Ley 37/2007, de 16 de noviembre, los órganos de la Administración General del Estado y los demás organismos y entidades del sector público estatal referidos en el artículo 1.2 deben *ejercer sus derechos de propiedad intelectual sobre sus documentos de forma que se facilite su reutilización.*

2. A tal efecto, la puesta a disposición de dichos documentos para su reutilización realizada conforme a lo dispuesto en el artículo 8.1 conllevará la *cesión gratuita y no exclusiva de los derechos de propiedad intelectual* correspondientes necesarios para desarrollar la actividad de reutilización autorizada, *en cualquier modalidad y bajo cualquier formato, para todo el mundo y por el plazo máximo permitido por la Ley.*

*No obstante, lo dispuesto en el párrafo anterior **podrá ser excepcionado**, en todo lo no referente a la no exclusividad de la cesión, mediante el establecimiento de condiciones*

específicas de acuerdo con lo dispuesto en los apartados 2 a 4 del artículo 8 cuando se empleen las modalidades de puesta a disposición previstas en los mismos, siempre dentro de los límites establecidos en la Ley 37/2007, de 16 de noviembre, y, en particular, en su artículo 4.3 y en su artículo 6.

Por otra parte, la Administración puede tener en su **custodia contenidos (documentos) que quedan también sujetos a los mandatos de la RISP**; Ello incluye también los contenidos custodiados en bibliotecas museos y archivos. Al poner tales contenidos a disposición del público para su reutilización, la Administración deberá salvaguardar siempre los derechos de propiedad intelectual de terceros (no titularidad de la Administración) que pudieran existir sobre tales contenidos. De nuevo, la decisión de que contenidos custodiados por la Administración son objeto de propiedad intelectual y cuáles no, corresponde a la normativa de propiedad intelectual (TRLPI). Una vez confirmado que se trata de contenidos de **propiedad intelectual de titularidad ajena**, su reutilización sólo será posible si se cuenta con la debida autorización del titular de derechos. La misma conclusión es predicable respecto de contenidos sobre los que recaigan derechos de **propiedad industrial** (patentes, marcas o diseños industriales) que sean titularidad de terceros. El Real Decreto de desarrollo de LRISP así nos lo recuerda:

*Artículo 9. Documentos e información objeto de **derechos de propiedad intelectual o industrial de terceros**.*

*La reutilización de los documentos que custodian los órganos de la Administración General del Estado y los demás organismos y entidades del sector público estatal referidos en el artículo 1.2 sobre los que existan derechos de propiedad intelectual o industrial de terceros sólo podrá ser autorizada si tales órganos, organismos y entidades disponen u obtienen, cuando la reutilización concreta que se vaya a hacer lo exija y en los términos en que sea necesaria, **la preceptiva y suficiente cesión de los derechos de explotación por parte de sus titulares**.*

En cualquier caso, tanto si se trata de contenidos propios como custodiados, la reutilización deberá salvaguardar siempre la protección de los **datos personales y la confidencialidad** que pudieran ser de aplicación a tales contenidos.

La salvaguarda, en la reutilización de la ISP, de los derechos de propiedad intelectual e industrial, los datos personales y la confidencialidad, serán objeto de estudio en los siguientes subcapítulos. Las condiciones concretas de reutilización que la Administración deberá respetar y asegurar al autorizar

la reutilización de los contenidos públicos -que aquí nos hemos limitado a esbozar- serán objeto de estudio más detallado en el capítulo 4 (Resultados).

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.

En este momento, es importante tener presente que 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)

Más allá de las condiciones específicas previstas en la normativa de RISP que deberán ser respetadas por la Administración al poner contenidos a disposición del público para su reutilización tres son los regímenes que pueden afectar mayoritariamente el uso de recursos para fines de tratamiento lingüístico: la propiedad intelectual, los datos personales y la confidencialidad.

2.3.1 *La protección de la propiedad intelectual*²

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 para su tratamiento, como en la protección de la aplicación con la cual se realizará el tratamiento (programa de ordenador), y en la reutilización de los resultados que se deriven de tal tratamiento. Por ello, a pesar de que analizaremos en este subcapítulo los pilares básicos de este régimen, los mismos serán también aplicables, en diferente medida y perspectiva, en los siguientes capítulos 3 (Proceso) y 4 (Resultados). Además, como veremos, la normativa de propiedad intelectual rige para los contenidos tanto de titularidad pública como privada.

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. Además, la ley de propiedad intelectual contiene algunas reglas relativas a la **cesión y licencia** de los derechos que deben ser unas, observadas y otras, tenidas en cuenta.

Estos 5 elementos (y la normativa que los regula) serán objeto de estudio y análisis detallado en las diferentes etapas de este informe, aunque desde posiciones distintas (la Administración como usuario o como titular) y variando los contenidos sobre los que se realiza el análisis (recursos de entrada y de salida, y procesador). Dicho de otra manera, además de convertirse en una restricción al uso de recursos lingüísticos, el mismo régimen de propiedad intelectual será de aplicación tanto en relación con la protección del programa de ordenador (procesador) y los actos de explotación que se llevan a cabo al realizar el tratamiento lingüístico, como en relación con la reutilización de los resultados de tal tratamiento.

² NOTA: Es importante tener presente que a pesar de que la ley española utilice el nombre de “propiedad intelectual”, lo que está protegiendo es, en concreto, los **derechos de autor y derechos conexos (es decir, el copyright)**.



Por ello, en esta primera parte, al analizar cómo el régimen de protección de la propiedad intelectual puede imponer restricciones al uso de los recursos para fines de tratamiento lingüístico, aprovecharemos también para realizar una **primera aproximación** (introdutoria) a los elementos básicos de este régimen.

2.3.1.1 Normativa aplicable

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).³

Es importante tener presente que esta normativa se encuentra altamente influenciada por el proceso de armonización que ha tenido lugar en la Unión Europea en los últimos años, así como por los diferentes **convenios internacionales** de protección de la propiedad intelectual. En concreto, vale la pena mencionar el **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).⁴

En la **Unión Europea**, se han adoptado principios básicos que afectan directamente al derecho de autor:

- no se puede discriminar a los autores europeos por razón de su nacionalidad (principio de no discriminación por razón de la nacionalidad): todos los ciudadanos de la UE deben ser tratados como si fueran nacionales;
- y una vez han sido lícitamente distribuidos en un Estado miembro, los ejemplares deben poder circular (ser revendidos) libremente por toda la UE (principio de agotamiento del derecho de distribución después de la primera venta).

Además, a lo largo de los años han sido adoptadas diferentes **directivas** -que los Estados miembros deben "transponer" a la ley nacional- para armonizar las leyes nacionales y evitar, así, que el ejercicio

³ Última modificación por Ley 21/2014 de 5 de noviembre. Puede consultarse el texto normativo actualizado en <https://www.boe.es/legislacion/codigos/>

⁴ 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/>.

de los derechos nacionales no obstaculice el buen funcionamiento del "mercado interior." Las Directivas más interesantes por el tema que nos ocupa son:

- 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**;

No existe un derecho de autor internacional, sino un conjunto de leyes nacionales que protegen el derecho de autor (y conexos) en los diferentes territorios (países).

Por lo tanto, 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. Es cierto que en el ámbito de la UE (y el EEE), las leyes de los Estados miembros son mucho más próximas que si comparamos la ley española con, por ejemplo, la *Copyright Act* de los EUA. Sin embargo, a pesar de ser más próximas, las diferencias entre las leyes europeas pueden ser importantes – especialmente, cuando se trata de temas (como el que nos ocupa) que todavía no se hallan completamente resueltos por las leyes de PI. A lo largo del informe, se apuntará aquellos supuestos en que las soluciones nacionales pueden ser más divergentes.

2.3.1.2 Obra y Autor: ¿qué recursos quedan sujetos al régimen de propiedad intelectual?

Los conceptos de autor y de obra son, pues, los pilares fundamentales e indisolubles del régimen de PI: no hay obra sin autor, ni autor sin obra. Ambos conceptos son, pues, conceptos jurídicos y es en este sentido que los utilizaremos en este informe y dejando de lado posibles significados que estos conceptos puedan tener (de hecho, tienen) en el lenguaje “común”. En otras palabras: aquí nos referimos a “obra” y a “autor” como aquel contenido y aquella persona que quedan protegidos por el régimen de la PI.



La clarificación de ambos conceptos (obra protegida y titular) es fundamental tanto para determinar cuáles de los recursos utilizados para el procesamiento (entradas) pueden verse sujetos al régimen de PI y quien sería el titular que podría autorizar o prohibir su procesamiento. Como veremos, ambos conceptos son también fundamentales para conocer el alcance de propiedad intelectual que podrá ejercitarse sobre los recursos resultantes (salidas).

A efectos del TRLPI, obra es toda **creación original expresada (artículo 10 TRLPI)**. 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. El artículo 10 TRLPI contiene una lista de ejemplos, únicamente a título enunciativo: no todo lo que está listado será obra protegida (sólo lo serán aquellas que sean “creaciones originales”) ni tampoco aquello que no está listado no podrá ser obra protegida (cualquier creación original estará protegida). Así pues, y con independencia de su origen, 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. Podemos hablar de originalidad en el sentido de “huella de la persona del autor” o en el sentido de “no haber sido copiado” o, como hacen las Directivas de la UE en el sentido de “creación original propia de su autor” (que parece englobar ambos aspectos). Según de qué tipo de obra se trate nos decantaremos más por uno o por el otro; por ejemplo, para los programas de ordenador, bases de datos o las meras fotografías será más difícil exigir originalidad en el sentido de creación “personal” de su autor. 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. A los efectos que nos ocupan, será pues fundamental distinguir qué recursos son “obra” protegida y qué



recursos son “datos” o información no protegida. 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 artículo 13 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).

Artículo 13 TRLPI – Exclusiones

“No son objeto de propiedad intelectual las disposiciones legales o reglamentarias y sus correspondientes proyectos, las resoluciones de los órganos jurisdiccionales y los actos, acuerdos, deliberaciones y dictámenes de los organismos públicos, así como las traducciones oficiales de todos los textos anteriores.”

Ello significa que, por lo que se refiere al derecho de autor, la Administración no podrá ejercer sobre tales contenidos (normativa, resoluciones judiciales, actos de organismos públicos, así como las traducciones oficiales de todos ellos) ningún derecho de propiedad intelectual. No poder ejercer derechos de propiedad intelectual no significa que la utilización de tales contenidos no venga sujeta a condiciones concretas de reutilización que deberán ser respetadas por el usuario; tal como vimos en el anterior subcapítulo, la normativa sobre reutilización de la información del sector público no sólo alcanza a contenidos protegidos por la propiedad intelectual sino que incluye cualesquiera contenidos, documentos, datos e información generados o custodiados por la Administración pública.

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, contenidos obtenidos de páginas web, artículos en publicaciones académicas, novelas que han sido escaneadas por una biblioteca, correspondencia custodiada por un archivo o museo... grabaciones sonoras o audiovisuales) pero habrá otros que no lo serán (por ejemplo, los data-sets de resultados científicos o las sentencias y normativa excluidas por el artículo 13 TRLPI). 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.

2.3.1.3 Titularidad

Una vez hemos identificado que el recurso es una “obra” y existe un “autor”, podemos plantearnos 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** – artículo 5 TRLPI. Cuando en la creación participan varias personas, hablaremos de coautoría; la **obra en colaboración** - artículo 7 TRLPI- pertenece a todos los coautores, según las proporciones que pacten entre ellos (si no pactan nada, se entiende que participan a partes iguales - artículo 393 CC- a menos que se pueda demostrar que sus contribuciones fueron en proporción diferente - artículo 1689 CC-.

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. Para ser autor **hay que “crear” no basta con “hacer”** (la precisión técnica –por compleja, excelente y valiosa que sea– no es suficiente para otorgar la condición de autor).

Cuando la obra ha sido creada en virtud de una **relación laboral (empleados) o funcional**, la ley (artículo 51 TRLPI) prevé directamente una **presunción de cesión de los derechos de explotación a favor del empresario para realizar su actividad habitual** (el empresario es, al fin y al cabo, es quien ha solicitado la creación de la obra, ha sufragado los costes de creación y ha remunerado al autor por la misma). Se trata, sin embargo, de una presunción de cesión y, como tal, **cabe pacto en contra**: las partes pueden acordar otra cosa (que no haya cesión alguna o ampliar la cesión a favor del empleador).

Ahora bien, esta presunción de cesión no será de aplicación para aquellas obras y prestaciones que han sido encargadas a un tercero por la empresa o Administración (por ejemplo, a través de un contrato de obra o servicio o a través de una licitación pública). En tales casos, el autor (y titular original de la obra resultante del encargo) será quien la haya creado; si la empresa o la Administración que la encarga quiere obtener los derechos de propiedad intelectual sobre ella, deberá pactar una cesión de derechos a su favor. En el caso de los encargos realizados por la Administración a personas externas, 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:

Art.301.2 TRLCSP (RDL 3/2011):

Salvo que se disponga otra cosa en los pliegos de cláusulas administrativas o en el documento contractual, 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 contratante.

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.

- *Una organización que contrata servicios de traducción, ¿puede reclamar a la empresa traductora las memorias de traducción? ¿Quién es el propietario de las memorias de traducción, la organización contratante o la empresa traductora? ¿Qué derechos tiene el contratista sobre esas memorias de traducción?*

La organización que contrata servicios de traducción sólo tendrá aquellos derechos de PI que haya obtenido por cesión de la empresa de traducción / traductor que la ha realizado. La organización puede reclamar las memorias de traducción, pero ello no supondrá cesión alguna de PI. Debe distinguirse entre el propietario de las memorias de traducción y el titular de la propiedad intelectual de estas memorias. La distinción entre soporte (sujeto a reglas de propiedad civil / posesión) y la obra (sujeto a la propiedad intelectual) es fundamental para responder a la pregunta. La propiedad de las memorias debe estipularse en el contrato y la titularidad de la PI (cesión de derecho a favor de quien encarga) también – si no se pacta nada, los derechos de PI permanecerán con el traductor/empresa traductora. Recordemos que traductor es autor de la traducción (suponiendo que no sea una traducción automática) y que la empresa traductora será titular de los derechos sobre la traducción (suponiendo que el traductor es trabajador de la misma).

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 y los programas de ordenador. El artículo 8 TRLPI define el concepto de **obra colectiva** y establece una presunción de titularidad inicial a favor de la persona que la ha coordinado y divulgado:

Art.8 TRLPI .- Obra colectiva:

Se considera obra colectiva la creada por la iniciativa y bajo la coordinación de una persona natural o jurídica que la edita o divulga bajo su nombre y está constituida por la reunión de aportaciones



de diferentes autores cuya contribución personal se funde en una creación única y autónoma, para la cual haya sido concebida sin que sea posible atribuir separadamente a cualquiera de ellos un derecho sobre el conjunto de la obra realizada.

Salvo pacto en contrario, los derechos sobre la obra colectiva corresponderán a la persona que la edite y divulgue bajo su nombre.

Este concepto está pensado para aquellas obras en cuya creación participan un gran número de autores (que serán autores de su contribución, pero no de la obra colectiva en su conjunto) y que ha sido coordinada por una persona o empresa que será quien la divulgará. Estamos hablando de periódicos y enciclopedias, de bases de datos, de programas de ordenador o de obras multimedia, de algunos mapas; pero no, de obras colectivas en el sentido editorial de obras con capítulos de diferentes autores. La empresa que la coordina, edita y divulga será el titular inicial de los derechos de propiedad intelectual de una obra colectiva, creada por varios autores bajo su coordinación y con independencia de la relación contractual que tuvieran con ella (laboral, contractual, encargo, etc), siempre y cuando no se hubiere pactado lo contrario con todos o alguno de los autores que han realizado aportaciones a la misma. .

Por otra parte, la protección de los **programas de ordenador** (el procesador de HTL) la examinaremos en el capítulo siguiente, por el momento basta con decir que el autor (y titular inicial de los derechos de PI) del programa de ordenador puede ser no sólo la persona natural (o personas naturales) que lo han creado sino también la persona jurídica que haya coordinado su creación como obra colectiva (artículo 8).

Ahora bien, 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" (artículo 9 TRLPI) es realizar una **obra compuesta**. El autor y titular de derechos sobre la obra compuesta es aquél que la ha llevado a cabo, pero para explotarla (quizás no, para crearla) necesitará la autorización del autor de la obra preexistente (el cual podrá decidir cómo autorizarla y en qué condiciones). La incorporación de la obra preexistente en la obra compuesta puede realizarse de varias maneras: modificando su forma ya sea anotando, arreglando, revisando, resumiendo, actualizando, etc. (artículo 11 TRLPI) o transformándola traduciéndola o adaptándola.

De hecho, uno de los ejemplos de **obra compuesta/derivada** que nos afecta directamente (en este momento) es 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 necesitaremos autorización doble: del titular de la obra original y del titular de la traducción.



Por otra parte, **el procesamiento HLT puede dar lugar a la creación de obras compuestas**. Como veremos más adelante (y trataremos especialmente en el capítulo 4 - Resultados) 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 del recurso utilizado); con lo cual a través de las diferentes cadenas o árboles de procesamiento (pipelines) se pueden estar generando nuevas obras o prestaciones protegidas. De hecho, el artículo 11 TRLPI dice expresamente que “también son objeto de propiedad intelectual... las anotaciones.” Ello no quiere decir que todas las anotaciones sean obra; para determinar cuando estamos ante una nueva obra compuesta/derivada (con el consiguiente titular y derechos sobre la misma) deberemos aplicar el mismo criterio del art.10 TRLPI: 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. Para ello será necesario analizar en más detalle en que consiste cada una de las tareas que conforman un HLT.

2.3.1.4 Derechos conexos sobre prestaciones y derecho *sui generis* sobre bases de datos

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: es el caso de los artistas (intérpretes y ejecutantes), el productor de fonogramas, el productor audiovisual, las entidades de radiodifusión, el realizador de meras fotografías y a algunas producciones editoriales; en estos casos, hablamos de “**derechos conexos**”. Estos derechos conexos no deberían tener un peso especial en el ámbito que nos ocupa, más allá de los casos en que el recurso entrante (y/o saliente) sea una **producción editorial** especialmente protegida (en concreto, la edición de una obra inédita en el dominio público o las ediciones que puedan ser individualizadas por su composición tipográfica, su presentación o sus características editoriales – artículo 129 TRLPI) o una **grabación sonora** que contenga expresiones lingüísticas sujetas al procesamiento.

Por otra parte, el régimen de **protección “sui generis” sobre las bases de datos** (artículo 133 y ss. TRLPI) 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 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. Empecemos apuntando que este 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). Es decir, aunque el contenido de la BBDD no esté protegido por la PI ello no impediría a su fabricante (si ha realizado una inversión sustancial en la “obtención, verificación o presentación” del contenido) a poder autorizar o prohibir la extracción y reutilización de tal contenido.

A continuación, analizaremos el significado de esta “extracción y reutilización” del contenido de la BBDD, así como el significado de “explotación” de una obra o prestación, en relación con el procesamiento del lenguaje.

Por ejemplo, tener acceso legal a una BBDD no necesariamente nos permite extraer, reutilizar o explotar las obras protegidas/prestaciones ahí contenidas, ni a realizar el procesamiento de las mismas. Para ello tendremos que 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 (y, por lo tanto, conferidos al ámbito de exclusividad del autor o fabricante de la BBDD);
- 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) propuesto en la Directiva sobre derechos de autor en el Mercado Único Digital actualmente en trámite parlamentario en la UE;
- c- Ante una respuesta negativa, si existe alguna autorización o licencia que lo autorice. Esta licencia puede venir en cualquier formato y momento; por ejemplo, podría estar contenida en la licencia de acceso a la BBDD (cada vez son más las editoriales que incluyen la autorización para hacer TDM en sus licencias de acceso) o darse en un momento posterior (por ejemplo, previa solicitud), o podría estar contenida en las condiciones de uso (o Aviso Legal) establecidas en un web de acceso abierto, del que se extraen los recursos, etc.



Si la respuesta a la pregunta (a) es negativa (en el sentido que el procesamiento de las obras/prestaciones protegidas y/o datos contenidos en una BBDD sujeta al *sui generis* no constituye acto de explotación y/o de extracción y reutilización), 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), ya no será necesaria la obtención de licencia alguna al quedar el procesamiento amparado por algún límite legalmente previsto.

De todo ello nos ocuparemos en los siguientes apartados.

2.3.1.5 ¿Qué actos/ usos quedan sujetos al derecho exclusivo del autor/titular?

Una vez identificado que el recurso a procesar es una obra o prestación protegida o información / datos contenidos en una BBDD protegida por el derecho *sui generis*, compete realizar la siguiente pregunta: **¿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?**

Parece una pregunta fácil pero no lo es. Veamos cuales son los derechos que la ley confiere al autor (y, en menor medida, a los titulares de derechos conexos) para ver en qué medida los actos de procesamiento son subsumibles en ellos. Los derechos de autor son de naturaleza diversa:

- **derechos morales** (divulgación, atribución, integridad, acceso y retirada) intransmisibles e irrenunciables destinados a proteger los intereses personalísimos del autor respecto a su obra;
- **derechos de explotación** (reproducción, distribución, comunicación pública y transformación) que ofrecen al autor la posibilidad de obtener un rendimiento económico de su esfuerzo creativo y de controlar “en exclusiva” la explotación de su obra (la facultad absoluta de autorizar y prohibir los actos de explotación);
 - Reproducir es realizar copias (tangibles o intangibles) de la obra o parte de ella.
 - Distribuir consiste en poner ejemplares tangibles de la obra a disposición del público, mediante la venta, alquiler, préstamo – en tanto este derecho sólo se ejercita a través de soportes físicos, no será especialmente relevante a los efectos del tratamiento lingüístico.
 - Comunicar al público engloba cualquier acto mediante el cual la obra pueda llegar al público sin necesidad de ejemplares tangibles – por lo tanto, este derecho alcanza la



exhibición de la obra en un museo, la proyección de una película en una sala de cine, la representación (teatral, musical, etc), la radiodifusión, la transmisión (y retransmisión) con o sin cable, la emisión por satélite, así como la puesta a disposición del público a través de internet.

- Transformar consiste en realizar una modificación de la obra, de la que derive una diferente; ello incluye la traducción o la adaptación de la obra
- y otros **derechos llamados “de simple remuneración”** que, si bien no le permiten ejercer ningún tipo de “exclusividad” sobre su obra, le permiten obtener un rendimiento económico (estos derechos son normalmente intransmisibles e irrenunciables y de gestión colectiva obligatoria).

A los efectos de procesamiento del lenguaje, de entre todos estos derechos los que centrarán nuestra atención serán 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), necesitaremos el amparo de un límite o la obtención de una licencia (o no) para poder procesar los recursos (entradas). Este mismo análisis nos 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? Empecemos por el **derecho de reproducción**. Tal como viene definido en el art.18 TRLPI (y de forma armonizada, todas las leyes de la UE – ex art.2 Directiva 2001/29) 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.” Con tan amplia definición, es innegable que el procesamiento lingüístico (de obras/prestaciones protegidas) comporta la realización de copias (acto de reproducción) de los recursos protegidos.

De acuerdo con el Art.20 TRLPI (armonizado en la UE -ex art.3 Directiva 2001/29/CE), es **comunicación pública** “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.” 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. En la medida en que el procesamiento de datos comportará, antes o después del procesamiento, su transmisión o puesta a disposición de

las personas que van a realizarlo o que posteriormente van a reutilizar los resultados del mismo, estaremos indiscutiblemente en el alcance de este derecho exclusivo.

Y, para acabar, el **derecho de transformación**, definido en el art.21 TRLPI como

Art.21 TRLPI: Transformación

1.La transformación de una obra comprende su traducción, adaptación y cualquier otra modificación en su forma de la que se derive una obra diferente.

Cuando se trate de una base de datos a la que hace referencia el artículo 12 de la presente Ley se considerará también transformación, la reordenación de la misma.

2. Los derechos de propiedad intelectual de la obra resultado de la transformación corresponderán al autor de esta última, sin perjuicio del derecho del autor de la obra preexistente de autorizar, durante todo el plazo de protección de sus derechos sobre ésta, la explotación de esos resultados en cualquier forma y en especial mediante su reproducción, distribución, comunicación pública o nueva transformación.

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. La distinción es necesaria por las diferentes implicaciones que tiene a varios niveles: a efectos de límites aplicables (por ejemplo, vimos que el art.31.1 TRLPI sólo excluye la reproducción y no alcanzaría a los actos de transformación del recurso utilizado haciendo preceptiva la correspondiente autorización); a efectos de la licencia o autorización necesaria (y quien puede otorgarla: ¿una entidad de gestión o el titular únicamente?); así como a los efectos de titularidad de los resultados (lo veremos más en detalle en el capítulo correspondiente).

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 el caso que nos ocupa, la pregunta a resolver es si a resultas de las diferentes tareas de procesamiento lingüístico de un recurso (obra protegida) podemos obtener una obra nueva, distinta. En cuyo caso, tendremos un nuevo “set” de titulares y de derechos sobre tal resultado.

- ¿De qué manera los derechos de propiedad sobre una memoria de traducción afecta a un modelo de traducción estadístico entrenado con ese recurso? Los modelos estadísticos de traducción lo que hacen, a grandes rasgos, es aprender a base de mostrarles una cantidad significativa de ejemplos de traducción. Una vez entrenado con esos ejemplos, ya es capaz de generalizar y traducir nuevos textos (por supuesto, con un cierto grado de cobertura y de precisión).



- En general ¿De qué manera los derechos de propiedad sobre un recurso lingüístico afecta a un procesador lingüístico estadístico entrenado con ese recurso? Por ejemplo, mediante técnicas de *machine learning*, a partir de una muestra suficiente de páginas web de empresas de ciberseguridad, se puede entrenar un procesador que sea capaz de detectar automáticamente páginas web de empresas de ciberseguridad.
- ¿Existe algún nivel de atomización de un texto a partir del cual no esté protegido por la propiedad intelectual? Para algunos entrenamientos bastan secuencias de un número no muy largo de palabras.

Dejemos ahora por un momento el derecho de autor para centrarnos en el alcance de la exclusividad que la ley confiere al fabricante de una base de datos a través del **derecho *sui generis***.

¿Qué significa extracción y reutilización del contenido de una BBDD? La respuesta simple pasa por realizar una aproximación al mismo alcance que se da a los autores sobre sus obras. En la Directiva sobre bases de datos, el legislador de la UE evitó claramente utilizar la terminología del derecho de autor (derechos de reproducción, distribución, comunicación pública y transformación) pero claramente pretendía con este régimen que el fabricante de una BBDD protegida por este derecho *sui generis* pudiera hacer los mismos actos que el titular de derechos. Por tanto, extraer puede interpretarse como “reproducir” (realizar una copia) y reutilizar como “distribuir, comunicar y – ¿cómo no?- transformar”). 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).

Veremos a continuación si existe algún límite en el TRLPI que permita tanto el uso de recursos protegidos (obra o prestación) como el uso de contenidos de una BBDD con derecho *sui generis*.

2.3.1.6 Límites: ¿qué recursos se pueden utilizar sin necesidad de autorización o licencia?

Además de los recursos que no son obra ni prestación protegida, o – aun siéndolo- quedan excluidos de protección por el artículo 13 TRLPI, 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*** (en el caso de obra en colaboración, se calculará a partir de la muerte del último coautor). Para las obras colectivas, el plazo



de 70 años *pma* se calcula a partir de la publicación. Estos plazos han sido objeto de armonización en la UE. Sin embargo, en España, es de aplicación transitoria el plazo de 80 años PMA previsto en la Ley de 1879 para aquellas obras de autores fallecidos antes del 7 Dic. 1987 -fecha de entrada en vigor la LPI'1987).

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 protegidas por el artículo 129 TRLPI.

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. Esto significa que sería perfectamente lícito (sin necesidad de autorización alguna) el uso (extracción y reutilización) de datos o recursos no protegidos contenidos en una BBDD publicada hace más 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. Se trata de supuestos en los que la ley directamente autoriza ciertos actos de explotación (negando al autor la posibilidad de ejercer sus derechos exclusivos), confiriéndole -en algunos casos- un derecho a ser remunerado o compensado (e.g. la compensación por copia privada o por préstamo público). Los límites normalmente responden a la necesidad de atender a otros derechos fundamentales o intereses públicos que se consideran tan o más importantes que el derecho de autor (piénsese, por ejemplo, en la libertad de expresión, la libertad de creación artística, la libertad de dar y recibir información, el derecho a la educación e investigación, etc). En otras palabras, los límites al derecho de autor son la manifestación de la función social que, como toda propiedad, tiene la propiedad intelectual. 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. Identificar claramente el alcance de estos límites legales es



fundamental para saber cuándo el tratamiento lingüístico de un recurso protegido queda amparado ya por ley y cuando exigirá la preceptiva autorización o licencia, así como para saber en qué condiciones los resultados del tratamiento lingüístico podrán ser reutilizados (¿y por quien?) sin necesidad de licencia y cuando su uso requerirá autorización.

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” del artículo 40 bis TRLPI); de no cumplirse tales condiciones, el acto de explotación no autorizado se convierte en una infracción. En internet (y en contextos de redes digitales) la amplia definición (e interpretación) de los derechos exclusivos supone una mayor presión sobre la interpretación del alcance de estos límites legales, con frecuencia mal equipados para adaptarse a los cambios tecnológicos constantes y a la rápida evolución de los mercados.

A efectos del procesamiento lingüístico, de entre **los límites previstos en el TRLPI** cabría destacar la **copia privada** del artículo 31.2 TRLPI y, en cierta medida, la copia que pueden realizar las bibliotecas (y similares) para **finés de investigación** del artículo 37.1 TRLPI. Como veremos ninguno nos 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 a 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 (con lo cual no alcanzaría a dar cobertura al potencial comercial de las HLT) – y ello, sin mencionar que la finalidad “de investigación o de conservación” exigida para este límite tampoco engranaría bien con el procesamiento lingüístico (no todo proyecto de procesamiento lingüístico lo es para fines de investigación).

Mucho más interesante a los efectos del tratamiento lingüístico es el límite de copias técnicas previsto frente al derecho de reproducción. Como si de una definición negativa se tratara, el art.31.1 TRLPI (y de forma armonizada, todas las leyes de la UE – ex art.5.1 Directiva 2001/29) establece:

Artículo 31. Reproducciones provisionales.

1. No requerirán autorización del autor los actos de reproducción provisional a los que se refiere el artículo 18 que, además de carecer por sí mismos de una significación económica independiente, sean transitorios o accesorios y formen parte integrante y esencial de un proceso tecnológico y cuya única finalidad consista en facilitar bien una transmisión en

red entre terceras partes por un intermediario, bien una utilización lícita, entendiéndose por tal la autorizada por el autor o por la ley.

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. A los efectos de entender mejor el alcance de este límite, es interesante tener en consideración las sentencias del TJUE en los casos *Infopaq I y II* C-5/08 C-302/10 (de 16.07.2009 y 17.01.2012, respectivamente) donde el tribunal analizó el significado de estos requisitos en relación con el escaneado y tratamiento (indexación) de contenidos periodísticos, concluyendo que:

- 1) *El artículo 5, apartado 1, de la Directiva 2001/29/CE ...debe interpretarse en el sentido de que los actos de reproducción provisionales que se efectúan en el transcurso de un procedimiento de recopilación de datos, como los de que se trata en el litigio principal,*
 - *cumplen el requisito de formar parte integrante y esencial de un proceso técnico, aunque inicien y pongan fin a dicho proceso e impliquen una intervención humana;*
 - *cumplen el requisito de perseguir una finalidad única, a saber, permitir una utilización lícita de una obra o prestación protegidas;*
 - *cumplen el requisito de no tener una significación económica independiente siempre que, por un lado, la realización de dichos actos no permita obtener un beneficio adicional, que vaya más allá del derivado de la utilización lícita de la obra protegida y que, por otro, los actos de reproducción provisionales no tengan como consecuencia una modificación de dicha obra.*
- 2) *El artículo 5, apartado 1, de la Directiva 2001/29/CE debe interpretarse en el sentido de que ha de considerarse que los actos de reproducción provisionales que se efectúan en el transcurso de un procedimiento de recopilación de datos, como los de que se trata en el litigio principal, no entran en conflicto con la explotación normal de la obra ni perjudican injustificadamente los intereses legítimos del titular del derecho.*

Será necesario, pues, identificar si las copias de recursos realizadas para su procesamiento lingüístico cumplen todos estos requisitos, de manera que pudieran quedar excluidas del ámbito del derecho de reproducción. 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, pero posiblemente algunas de las acciones de tratamiento lingüístico sí que las cumplirían y quedarían, pues, exentas del derecho de reproducción. En cualquier caso, el alcance de este límite solo abarcaría los actos de reproducción, pero no los subsiguientes actos de comunicación pública o de transformación que se hiciera de los recursos utilizados para el tratamiento lingüístico – con lo cual, su alcance es del todo insuficiente para cubrir el uso de recursos protegidos por la propiedad intelectual en actuaciones de HTL.

Estrechamente relacionado con este límite, existe una propuesta de límite específico a favor de la minería de datos (*text and data mining* - 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 considerada por la Comisión y el Parlamento de la Unión Europea.

Artículo 3 -Minería de textos y datos

1. *Los Estados miembros establecerán una excepción a los derechos previstos en el artículo 2 de la Directiva 2001/29/CE, el artículo 5, letra a), y el artículo 7, apartado 1, de la Directiva 96/9/CE, y el artículo 11, apartado 1, de la presente Directiva con respecto a las reproducciones y extracciones realizadas por organismos de investigación con el fin de proceder a la minería de textos y datos de obras u otras prestaciones a las que tengan acceso legítimo con fines de investigación científica.*
2. *Será inaplicable toda disposición contractual contraria a la excepción prevista en el apartado 1.*
3. *Los titulares de derechos estarán autorizados a aplicar medidas para garantizar la seguridad e integridad de las redes y bases de datos en que estén almacenadas las obras u otras prestaciones. Dichas medidas no irán más allá de lo necesario para lograr ese objetivo.*
4. *Los Estados miembros alentarán a los titulares de derechos y organismos de investigación a establecer las mejores prácticas comunes para la aplicación de las medidas contempladas en el apartado 3.*

Los considerandos relativos a este límite son especialmente relevantes:

- (8) Las nuevas tecnologías hacen posible el análisis computacional automatizado de información en formato digital, por ejemplo de textos, sonidos, imágenes o datos, al que generalmente se denomina «minería de textos y datos». Estas tecnologías permiten a los investigadores tratar grandes cantidades de información para obtener nuevos conocimientos y descubrir nuevas tendencias. Aun cuando las tecnologías de este tipo estén muy extendidas en la economía digital, se reconoce generalmente que la minería de textos y datos puede beneficiar especialmente a la comunidad de investigadores, impulsando de este modo la innovación. No obstante, los organismos de investigación de la Unión, como

las universidades y los institutos de investigación, se enfrentan a cierta inseguridad jurídica a la hora de determinar hasta qué punto pueden llevar a cabo actividades de minería de textos y datos de contenidos. En determinados casos, la minería de textos y datos puede comportar actos protegidos por derechos de autor o por el derecho *sui generis* sobre las bases de datos, en particular la reproducción de obras u otras prestaciones o la extracción de contenidos de una base de datos. Cuando no existe ninguna excepción o limitación aplicable, debe solicitarse una autorización a los titulares de derechos para llevar a cabo tales actos. La minería de textos y datos también puede tener por objeto meros hechos o datos que no están protegidos por derechos de autor y, en tales casos, no ha de ser necesaria una autorización.

- (9) El Derecho de la Unión ya prevé determinadas excepciones y limitaciones con respecto a los usos para fines de investigación científica que pueden aplicarse a los actos de minería de textos y datos. Con todo, esas excepciones y limitaciones tienen carácter optativo y no están plenamente adaptadas al uso de las tecnologías en la investigación científica. Por otra parte, en los casos en que los investigadores pueden acceder lícitamente a los contenidos —por ejemplo, a través de suscripciones a publicaciones o licencias de acceso abierto—, las condiciones de las licencias pueden excluir la minería de textos y datos. La creciente utilización de las tecnologías digitales en las actividades de investigación puede afectar a la posición competitiva de la Unión como espacio de investigación, a menos que se tomen medidas para eliminar la inseguridad jurídica existente en materia de minería de textos y datos.
- (10) Esa inseguridad jurídica debe subsanarse estableciendo una excepción obligatoria respecto del derecho de reproducción, así como del derecho de prohibir la extracción de una base de datos. La nueva excepción debe entenderse sin perjuicio de la actual excepción obligatoria aplicable a los actos de reproducción provisional, prevista en el artículo 5, apartado 1, de la Directiva 2001/29/CE, que debe seguir aplicándose a las técnicas de minería de textos y datos que no requieran la realización de copias más allá del alcance de dicha excepción. Los organismos de investigación también han de poder acogerse a la excepción cuando participen en asociaciones público-privadas.
- (11) Los organismos de investigación de la Unión abarcan una amplia variedad de entidades cuyo principal objetivo es la investigación científica, combinada o no con la prestación de servicios educativos. Debido a la diversidad de tales entidades, es importante disponer de una definición común de los beneficiarios de la excepción. A pesar de las diversas formas y estructuras jurídicas, los organismos de investigación de los Estados miembros suelen llevar a cabo sus actividades sobre una base no lucrativa o en el contexto de una misión de interés público reconocida por el Estado. Dicha misión puede quedar reflejada, por ejemplo, a través de la financiación pública o de disposiciones de la legislación nacional o de los contratos públicos. Al mismo tiempo, no han de considerarse organismos de investigación a los efectos de la presente Directiva aquellos organismos sobre los que las empresas comerciales tienen una influencia decisiva que les permite ejercer el control debido a situaciones estructurales tales como la calidad de sus accionistas o socios, lo cual puede dar lugar a un acceso preferente a los resultados de la investigación.
- (12) Habida cuenta del número potencialmente elevado de solicitudes de acceso y descargas de sus obras u otras prestaciones, es conveniente autorizar a los titulares de derechos a aplicar medidas cuando exista el riesgo de que la seguridad e integridad de la red o las bases de datos que contengan las obras u otras prestaciones puedan verse comprometidas. Estas medidas no deben exceder de lo necesario para alcanzar el objetivo de garantizar la seguridad e integridad de la red y no han de menoscabar la aplicación efectiva de la excepción.
- (13) No es necesario prever una indemnización para los titulares de derechos por los usos en el marco de la excepción aplicable a la minería de textos y datos establecida en la presente

Directiva por cuanto el perjuicio debería ser mínimo, habida cuenta de la naturaleza y el ámbito de aplicación de la excepción.

De aprobarse, este límite autorizaría la reproducción y extracción (nótese la doble referencia: a obras y al derecho *sui generis* sobre BBDD) 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 científica. Es decir, se trata de un límite muy reducido que difícilmente permitiría entender autorizados gran parte de los actos de procesamiento lingüístico analizados. Más aún, lo alarmante de esta propuesta es lo que implica: que cualquier acto de procesamiento digital de recursos cae dentro del derecho exclusivo que la ley confiere al autor (al menos, bajo el amplio derecho de reproducción) y que el límite de las copias temporales (artículo 31.1 TRLPI) no es suficiente para permitirlo. En otras palabras, de aprobarse este límite (que sería de aplicación obligatoria en todos los países de la UE) el procesamiento lingüístico para fines que no sean propiamente de investigación científica quedará *de facto* sujeto al ámbito de exclusividad del autor sin que pueda recurrirse a la interpretación del límite de las copias técnicas del art.31.1 TRLPI. El potencial de las nuevas tecnologías para tratar grandes cantidades de información no se limita a su uso para los fines de investigación sino también al desarrollo de plataformas como la que ahora nos ocupa que ciertamente favorecen el desarrollo social y económico de una comunidad. Parecería más sensato, dado tal potencial, ampliar más el límite de TDM (para cubrir todo tipo de procesamiento automático de texto o imagen con fines puramente instrumentales), o restringir el alcance del derecho de reproducción mediante la interpretación del límite de copias técnicas del art.31.1 TRLPI, para que en lugar de alcanzar a todo tipo de copia, se limitara a cubrir aquellas copias de la obra que la utilizan como tal obra (en su expresión y forma, así como en su mercado), y no a cualquier copia que se haga de la información que la conforma; al fin y al cabo, las palabras y las letras, procesadas de forma individual, no son más que datos e información ajenos al ámbito de protección de la propiedad intelectual.

Ante la incertidumbre normativa, los editores de publicaciones científicas y académicas (y fabricantes de bases de datos) han estado licenciando a las universidades e instituciones científicas, la actividad de TDM de los contenidos accesibles a través de sus bases de datos. Sobre el TDM de publicaciones académicas, véase Jonathan Clark (2012) *Text Mining and Scholarly Publishing*, Publishing Research Consortium. Pero el TDM va más allá de las publicaciones científicas (y obras textuales) y de los fines de investigación científica.



Además, en este punto merece la pena realizar un comentario de derecho comparado. El único país de la UE que actualmente tiene un límite para el TDM (para fines no comerciales de investigación) es el Reino Unido – los demás países todavía no se han atrevido a pronunciarse al respecto. En cambio, en los USA el TDM se considera amparado por la doctrina del *fair use*; en este sentido, diferentes sentencias han confirmado que la minería de datos y textos es lícita: *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014), *Authors Guild v. Google*, 770 F.3d 666 (2d Cir. 2015); en este último caso, la sentencia firme de la *US Court of Appeal, 2nd Circuit*, de 16 Oct. 2015 rechazó que Google estuviera infringiendo el derecho de autor de las obras que escaneaba e indexaba en su totalidad y mostraba parcialmente (*snippets*) en respuesta a las búsquedas de los usuarios, dentro de su proyecto *Google Books*. Las conclusiones de esta sentencia son fácilmente extensibles al procesamiento lingüístico, también con fines comerciales, en tanto tal uso aporte un valor añadido y no entre en conflicto con la explotación de las obras procesadas. De aprobarse el límite propuesto por la Comisión, Europa andaría justo en el sentido opuesto a los USA, dificultando el desarrollo de nuevos servicios y productos en el mercado que exigieran el previo procesamiento de obras / prestaciones protegidas, desaprovechando el potencial tecnológico digital y, en última instancia, poniendo a las empresas y a la sociedad europea en una desventaja competitiva con la norteamericana y demás países.

2.3.1.7 Licencias: ¿Cómo puedo obtener 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.

La autorización para realizar HLT puede venir de diversas formas, pero en todo caso, deberá proceder del titular de los derechos objeto de licencia. 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 (artículo 43 TRLPI).



No hay formalidad alguna a seguir, basta con que la autorización sea clara y explícita (en teoría, es posible demostrar la existencia de una autorización implícita pero siempre será mucho más difícil de probar).

Para proteger al autor (normalmente, la parte contratante más débil), el TRLPI establece 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 (para realizar el procesamiento lingüístico) como las condiciones para licenciar los resultados para su reutilización en subsiguientes procesamientos.

La autorización o licencia para realizar actos de explotación HTL 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)
- VEGAP (Visual, Entidad de Gestión de Artistas Plásticos: www.vegap.es)
- CEDRO (Centro Español de Derechos Reprográficos: www.cedro.org)
- DAMA (Asociación Derechos de Autor de Medios Audiovisuales: www.damautor.es)



- AIE (Sociedad de Artistas Intérpretes o Ejecutantes de España: www.aie.es)
- AISGE (Asociación de Actores Intérpretes, Sociedad de Gestión de España: www.aisge.es)
- AGEDI (Asociación de Gestión de Derechos Intelectuales: www.agedi.es)
- EGEDA (Entidad de Gestión de Derechos de Productores Audiovisuales: 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 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 licenciatario 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:

	Reconocimiento (by): Autoriza cualquier uso o acto de explotación que se haga de la obra (incluidos los usos comerciales y las transformaciones).
	Reconocimiento (by) - Sin obra derivada (nd): Excluye la transformación de la obra para crear otra nueva (obra derivada).
	Reconocimiento (by) - No comercial (nc): Excluye la posibilidad de utilizar la obra con fines comerciales.
	Reconocimiento (by) - Sin obra derivada (nd) - No comercial (nc): No se permite ni hacer obras derivadas, ni los usos comerciales.
	Reconocimiento (by) - Compartir igual (sa): 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.
	Reconocimiento (by) - No comercial (nc) - Compartir igual (sa): 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 el capítulo 4 (Resultados). 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 La protección de los datos personales y la anonimización

Las administraciones públicas en el desarrollo de sus funciones recaban ingentes cantidades de datos personales de la ciudadanía. El procesamiento de recursos lingüísticos que contengan datos personales puede generar amenazas para la privacidad de las personas y poner en riesgo dichos datos. Para hacer frente a las posibles amenazas y riesgos, en cualquier procesamiento de lenguaje



natural debe valorarse la existencia de datos personales y, en su caso, adoptar las medidas previstas respecto a su tratamiento adecuado y seguridad. En cualquier caso, se debe tener en cuenta que, como señala la Comisión Europea, “los objetivos de promover los datos abiertos de la administración pública y proteger los datos personales pueden reforzarse mutuamente si se basan en una gestión proactiva y consciente de la información por parte del sector público” [3, 6].

En las próximas páginas, la atención se debe centrar en el análisis de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal (en adelante, LOPD) así como el Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo y, en determinados aspectos, la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información. Asimismo, debe tenerse en cuenta el Real Decreto 3/2010, de 8 de enero, por el que se aprueba el Esquema Nacional de Seguridad. También se debe tener presente lo previsto en la LRISP. Finalmente, no se puede desconocer el impacto que en el futuro cercano tendrá el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE que ha unificado y modernizará la regulación europea de la protección de datos dando respuesta a los nuevos retos generados en los últimos años.

Como punto de partida, se debe tener presente que 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).

De este modo, un dato personal no solo es una información numérica o alfanumérica sino también tienen esta consideración las imágenes, la voz, las huellas digitales o los datos biométricos en la medida en se refieran a una persona física siempre y cuando no se requieran plazos o actividades desproporcionadas.

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 natural 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.

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 natural no incluye una obligación de difusión de información personal. En efecto, como recuerda el Grupo de Trabajo del Artículo 29, “el «principio de reutilización» no es automático cuando está en juego el derecho a la protección de los datos personales, y no anula las disposiciones aplicables de la normativa sobre protección de datos” [4].

De este modo, es necesario conocer los principios y normas que guían cualquier tratamiento de carácter personal para identificar las condiciones que deben concurrir para que se puedan reutilizar datos que contengan datos personales como recurso para el procesamiento del lenguaje natural.

Desde un punto de vista general, 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 esta dirección, la LOPD dispone que “los datos de carácter personal objeto de tratamiento no podrán usarse para finalidades incompatibles con aquellas para las que los datos hubieran sido recogidos. No se considerará incompatible el tratamiento posterior de éstos con fines históricos, estadísticos o científicos” (artículo 4.2). En este punto la principal dificultad estriba en conocer cuando una finalidad es compatible con otra. Al respecto, como recuerda el Grupo de Trabajo del Artículo 29 “es difícil aplicar el principio de limitación de la finalidad de forma efectiva en caso de reutilización de la ISP. Por una parte, la propia idea y la fuerza impulsora de la innovación que se esconde tras el concepto de «datos abiertos» y reutilización de la ISP es que la información debe estar disponible para su reutilización para nuevos productos y servicios innovadores y, de este modo, para fines que no estén previamente definidos y no puedan preverse claramente” [4]. Para llevar a cabo dicho análisis, el propio Grupo de Trabajo en un dictamen anterior, afirmaba que para poder conocer si un tratamiento es compatible o incompatible con los fines para los que se recogieron los datos se deben tener en cuenta diferentes factores como la relación entre los fines para los que se recogieron los datos personales y los fines de su tratamiento ulterior; el



contexto en el que se recogieron los datos personales y las expectativas razonables de los interesados en cuanto a su uso ulterior; la naturaleza de los datos personales y el impacto del tratamiento ulterior en los interesados; o las medidas de salvaguardia aplicadas por el responsable del tratamiento para garantizar el tratamiento leal y evitar que se produzcan repercusiones indebidas sobre los interesados [5].

- **Consentimiento:** el tratamiento de los datos personales requiere el consentimiento inequívoco del interesado. Existen datos como los relativos a la ideología, la afiliación sindical, la religión o las creencias que tienen una protección reforzada y requieren un consentimiento expreso y por escrito. Otros datos como los que se refieren al origen racial, la salud o la vida sexual que únicamente pueden ser recogidos, tratados, cedidos cuando, por razones de interés general, así lo establezca una ley o el afectado lo consienta expresamente.

Sin embargo, el principio del consentimiento tiene algunas excepciones. En particular, la LOPD prevé que no será preciso el consentimiento cuando, por ejemplo, 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.** Tal y como establece el art.9 LOPD "[e]l responsable del fichero, y, en su caso, el encargado del tratamiento deberán adoptar las medidas de índole técnica y organizativas necesarias que garanticen la seguridad de los datos de carácter personal y eviten su alteración, pérdida, tratamiento o acceso no autorizado, habida cuenta del estado de la tecnología, la naturaleza de los datos almacenados y los riesgos a que están expuestos, ya provengan de la acción humana o del medio físico o natural".
- **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.** Según el art.11 LOPD "Los datos de carácter personal objeto del tratamiento 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" excepto en los casos previstos en la ley.

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. La cesión o



comunicación de datos personales sólo se puede realizada para el cumplimiento de fines directamente relacionados con las funciones legítimas del cedente y del cesionario con el previo consentimiento del interesado. Sin embargo, este 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. Las medidas de seguridad deben impulsarse por los responsables de los ficheros y los encargados de tratamiento. En efecto, los responsables de los ficheros tienen la obligación de garantizar la seguridad y guardar el secreto profesional de los datos a los que tienen acceso. Para ello tienen el deber, junto a los responsables de tratamiento, de adoptar las medidas de índole técnica y organizativa necesarias para garantizar la seguridad de los datos de carácter personal y evitar su alteración, pérdida, tratamiento o acceso no autorizado, según el estado de la tecnología, la naturaleza de los datos y los riesgos a los que estén expuestos. A su vez, pueden nombrar un responsable de seguridad que será la persona o personas a las que el responsable del fichero haya asignado formalmente la función de coordinar y controlar las medidas de seguridad aplicables.

Las medidas de seguridad persiguen garantizar la confidencialidad (permitir el conocimiento de los datos sólo por los usuarios autorizados), integridad (impedir la alteración de la información) y disponibilidad de los datos (de manera que la información sea utilizada por los usuarios autorizados).

En función de los datos personales, las medidas de seguridad que se deberán adoptar serán diferentes. En efecto, el reglamento de desarrollo de la LOPD prevé que las medidas de seguridad exigibles a los ficheros y tratamientos se clasifican en tres niveles: básico, medio y alto.

- nivel básico. Todos aquellos datos personales que no estén incluidos en otros niveles de seguridad.
- nivel medio: ficheros o tratamientos de datos relativos a la comisión de infracciones administrativas; que sean responsables Administraciones tributarias y se relacionen con el ejercicio de sus potestades tributarias; que sean responsables las Entidades Gestoras y Servicios Comunes de la Seguridad Social y se relacionen con el ejercicio de sus competencias; que contengan un conjunto de datos de carácter personal que ofrezcan una



definición de las características o de la personalidad de los ciudadanos y que permitan evaluar determinados aspectos de la personalidad o del comportamiento de los mismos.

- nivel alto: ficheros o tratamientos de datos de carácter personal que se refieran a datos de ideología, afiliación sindical, religión, creencias, origen racial, salud o vida sexual; que contengan o se refieran a datos recabados para fines policiales sin consentimiento de las personas afectadas o que contengan datos derivados de actos de violencia de género.

En función del nivel, la normativa de desarrollo de la LOPD fija las medidas de seguridad que se deben adoptar en función del nivel de los ficheros o tratamientos de datos. Asimismo, es necesario tener en cuenta las medidas de seguridad que deban adoptar las administraciones públicas en aplicación de lo previsto en el Real Decreto 3/2010, de 8 de enero, por el que se aprueba el Esquema Nacional de Seguridad.

Además de estas normas, debe tenerse presente que algunas entidades han adoptado códigos de conducta. La LOPD prevé la posibilidad de formular códigos tipos a los responsables de los ficheros, a través de acuerdos sectoriales, convenios administrativos o mediante decisiones de empresa, 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 Ley 37/2007, de 16 de noviembre dispone que el régimen de reutilización garantiza el pleno respeto de los principios que consagran la protección de datos personales. De hecho, es necesario tener presente cuando se lleve a cabo el procesamiento de textos que contengan datos personales de los riesgos que pueden entrañar para ellos. En esta dirección, se manifiesta el Grupo de trabajo del Artículo 29 al afirmar que “teniendo en cuenta los riesgos potenciales de la reutilización de la ISP, y, en particular, el hecho de que, una vez que los datos personales se han puesto a disposición del público para su reutilización, será muy difícil controlar eficazmente la utilización de estos datos”. Por ello, el Grupo recomienda “la realización de un análisis de impacto exhaustivo sobre la protección de datos por el organismo del sector público antes de facilitar datos personales para su reutilización” [4].



Como punto de partida cabe recordar que la Ley 37/2007, de 16 de noviembre no es de aplicación a los documentos o a las partes de los mismos que obren en poder de las Administraciones públicas a los que no se pueda accederse o 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, tal y como se ha señalado anteriormente, así lo prevea una ley, se recabe el consentimiento de los interesados o se procesa a la anonimización de los datos personales.

En primer lugar, por lo que respecta al consentimiento de los interesados, la LOPD prevé que cuando la información que se facilite al interesado no le permita conocer la finalidad a que destinarán los datos cuya comunicación se autoriza o el tipo de actividad de aquel a quien se pretenden comunicar, será nulo el consentimiento para la comunicación de los datos de carácter personal a un tercero (artículo 11.3 LOPD). Asimismo, el reglamento LOPD dispone que “cuando se solicite el consentimiento del afectado para la cesión de sus datos, éste deberá ser informado de forma que conozca inequívocamente la finalidad a la que se destinarán los datos respecto de cuya comunicación se solicita el consentimiento y el tipo de actividad desarrollada por el cesionario. En caso contrario, el consentimiento será nulo” (artículo 12.2 Reglamento LOPD).

Sin embargo, en ocasiones recabar el consentimiento de los interesados puede entrañar dificultades. Cuando el consentimiento se haya obtenido *a priori* no representará problema alguno (por ejemplo, cuando los interesados hayan dado su consentimiento para participar en otras fases de un estudio). Sin embargo, cuando se tenga que identificar, contactar y obtener el consentimiento de un gran número de interesados es evidente que puede ser una operación compleja y costosa.

Relacionado con lo anterior está la posibilidad de los interesados de ejercer sus derechos de acceso, rectificación, cancelación y oposición (derechos ARCO) en cualquier momento. En relación a esta cuestión, resulta de interés traer a colación el Dictamen en relación al Proyecto para dar valor a la información del sistema sanitario catalán en el marco de las políticas públicas, VISC + (versión marzo de 2015) de 16 de abril de 2015 de la Autoridad Catalana de Protección de Datos en el que se señala que “más allá de la posibilidad de ejercicio del derecho de oposición, **hay que valorar muy positivamente que se explicita, en la documentación aportada, la posibilidad de que los afectados ejerciten la opt-out** para mantenerse al margen del proyecto VISC + , en el sentido de que sus datos no sean anonimizados ni incorporados a VISC +. Es decir, se prevé que los ciudadanos podrán decidir libremente si quieren que sus datos se excluyan de este proyecto. Hay que tener en cuenta que el



derecho de oposición previsto en la normativa de protección de datos y que se puede ejercer ante el responsable del tratamiento, requiere que la persona que lo ejercita alegue una causa justificada basada en el su situación personal. En cambio, con la incorporación del mecanismo de opt-out, sería posible que aquellas personas que a pesar de las garantías que ofrezca el sistema, no deseen participar en el proyecto, puedan hacerlo sin tener que alegar una justificación específica” [negrita en el original].

En segundo lugar, la reutilización se puede llevar a cabo si se produce la disociación de los datos que impida la identificación de las personas. En este caso, estaríamos hablando de una cesión de datos anonimizados que no requiere el consentimiento de los interesados tal y como dispone la LOPD (artículo 11.6 y, en particular en el ámbito de la salud el artículo 16.3 de la Ley 41/2002).

En efecto, tal y como dispone el Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP para el ámbito del sector público estatal, las Administraciones públicas deben proceder a la disociación de los datos personales, siempre y cuando los medios técnicos y económicos lo permitan, para permitir la reutilización de los datos por otras personas. De este modo, como señala el Reglamento General de Protección de Datos “los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo”.

La disociación de los datos personales es aquel tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable, es decir, que permita obtener un dato disociado.

Por lo que respecta a la disociación, como se desprende de la lectura de las Orientaciones y garantías en los procedimientos de anonimización de datos personales de la Agencia Española de Protección de Datos de 2016, el fin del proceso de anonimización es eliminar o reducir al mínimo los riesgos de reidentificación de los datos anonimizados manteniendo la veracidad de los resultados del tratamiento de éstas, es decir, además de evitar la identificación de las personas, los datos anonimizados deben garantizar que cualquier operación o tratamiento que pueda ser realizado con posterioridad a la anonimización no conlleva una distorsión de los datos reales. 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 [6].

La anonimización se puede llevar a cabo utilizando diferentes técnicas. Según el Dictamen 05/2014 sobre técnicas de anonimización del Grupo de Trabajo sobre protección de datos del artículo 29, se pueden utilizar técnicas de anonimización como la aleatorización (adición de ruido, la permutación, la privacidad diferencial), la generalización (agregación, anonimato k, diversidad de proximidad) y la pseudonimización (cifrado de clave secreta, función hash). Sin embargo, estas técnicas pueden generar riesgos de singularización (posibilidad de extraer de los datos registros que identifican a una persona); vinculabilidad (capacidad de vincular dos o más registros de una persona de una o varias bases de datos) o inferencia (posibilidad de deducir el valor de un atributo a partir de los valores del resto de atributos). Para reducir los riesgos de identificación se pueden seguir unas buenas prácticas tales como evaluar periódicamente los riesgos, no olvidar la posibilidad de que haya atacantes, utilizar de manera combinada diversas técnicas o revelar la técnica utilizada [7].

Sin embargo, es importante ser conscientes de que cuando se faciliten datos para que puedan ser reutilizados como recursos para el procesamiento del lenguaje natural, la Administración pública pierde el control sobre los quien los reutiliza y los usos que hace con ellos. En última instancia, aumentan considerablemente las posibilidades que un tercero tenga en su poder la tecnología que permita reidentificar a los interesados lo que constituye una amenaza importante [4].

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. Así lo advierte el Grupo de Trabajo del Artículo 29 en su Dictamen 5/2014, en el que observar que “(...) los responsables del tratamiento deben ser conscientes de que un conjunto de datos anonimizado puede entrañar todavía riesgos residuales para los interesados. Efectivamente, por una parte, la anonimización y la reidentificación son campos de investigación activos en los que se publican con regularidad nuevos descubrimientos y, por otra, incluso los datos anonimizados, como las estadísticas, pueden usarse para enriquecer los perfiles existentes de personas, con la consiguiente creación de nuevos problemas de protección de datos. En suma, la anonimización no debe contemplarse como un procedimiento esporádico, y los responsables del tratamiento de datos han de evaluar regularmente los riesgos existentes” [7].

De este modo, como recuerda la Autoridad Catalana de Protección de Datos en su dictamen sobre el proyecto VISC+, en el marco del principio de minimización en el tratamiento de los datos personales, “sería necesario limitar los datos comunicados a los mínimos indispensables para conseguir la finalidad perseguida por el cliente final”.



La anonimización tiene un impacto importante cuando las tecnologías del lenguaje se apliquen en el ámbito de la salud. Así, por ejemplo, en el marco de un proyecto que se base en el análisis de los datos contenidos en las historias clínicas como era el caso del proyecto VISC+ que se propuso impulsar el Departamento de Salud de la Generalitat de Catalunya. De acuerdo con la memoria, “el proyecto VISC+ tiene por objetivo relacionar la información de salud que se genera en Cataluña de una manera totalmente anonimizada y segura para impulsar y facilitar la investigación y la innovación en medicina y las ciencias de la salud”.⁵ Al respecto, no se puede desconocer que según la Ley 41/2002, de 14 de noviembre, reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica dispone que para tratar los datos que constan en la historia clínica será necesario el consentimiento de su titular excepto que concurra alguna de las causas previstas en la ley o se anonimice la información (artículo 16.3 de la Ley 41/2002).

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*). Así lo señaló en su día la Autoridad Catalana de Protección de Datos en el Dictamen en relación con la consulta planteada por una entidad de derecho público en relación con el modelo de gestión y de servicios para dar valor a la información del sistema sanitario catalán en el marco de las políticas públicas (en adelante VISC+).

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. En efecto, el Reglamento General de Protección de Datos 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.

En particular, se dispone que “cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del

⁵ Memoria del proyecto VISC+ (enero 2015) http://aguas.gencat.cat/web/.content/minisite/aguas/projectes/antic_visc/memoria_visc_aguas2015.pdf (última consulta: marzo de 2017).



tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales. Una única evaluación podrá abordar una serie de operaciones de tratamiento similares que entrañen altos riesgos similares” (artículo 35). Esta evaluación de impacto será necesaria, entre otros casos, cuando se quiera llevar a cabo un tratamiento a gran escala de categorías especiales de datos como datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física o datos relativos a la salud. En cualquier caso, cuando entre en vigor el Reglamento General de Protección de Datos, será la Agencia Española de Protección de Datos la que deberá establecer y publicar una lista de los tipos de operaciones de tratamiento que requieran una evaluación de impacto relativa a la protección de datos.

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.

En esta dirección, resulta de interés la consulta de la *Guía para una evaluación de impacto en la protección de datos* elaborada por la Agencia Española de Protección de Datos [8].

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. En particular, pueden incluir una cláusula de protección de datos. Siguiendo al Grupo de Trabajo del Artículo 29, estas cláusulas pueden ser de dos tipos:

- cláusulas de protección de datos relativas a datos anonimizados. Estas cláusulas pueden hacer referencia al carácter anonimizado de los datos; a la prohibición de reidentificar a las personas; a la prohibición de tomar cualquier medida o decisión respecto a las personas de que se trate; o la obligación de notificar la detección de que las personas pueden ser reidentificadas.
- cláusulas relativas a las finalidades compatibles con la finalidad para la que los datos fueron inicialmente recabados.

Además de la inclusión de estas cláusulas 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.

En primer lugar, porque la cesión de datos personales sin contar con legitimación para ello es un hecho constitutivo de infracción grave o muy grave por parte del responsable del tratamiento que puede ser sancionada con multas desde 40.001 a 600.000 euros.

En segundo lugar, porque el incumplimiento de las condiciones impuestas en la correspondiente licencia o en la normativa reguladora aplicable que puede conllevar la imposición de una sanción al reutilizador de multa de 10.001 a 50.000 euros por la comisión de infracciones graves y que al graduarse podrá tenerse en cuenta los daños y perjuicios causados por lo que se refiere a la protección de datos de carácter personal.

2.3.3 La confidencialidad de la información

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

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:

- a) La seguridad nacional.
- b) La defensa.
- c) Las relaciones exteriores.
- d) La seguridad pública.
- e) La prevención, investigación y sanción de los ilícitos penales, administrativos o disciplinarios.
- f) La igualdad de las partes en los procesos judiciales y la tutela judicial efectiva.
- g) Las funciones administrativas de vigilancia, inspección y control.



- h) Los intereses económicos y comerciales.
- i) La política económica y monetaria.
- j) El secreto profesional y la propiedad intelectual e industrial.
- k) La garantía de la confidencialidad o el secreto requerido en procesos de toma de decisión.
- l) La protección del medio ambiente.

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”.

En la actualidad, las administraciones públicas disponen de ingentes cantidades de información relativas a las empresas. Una parte de esta información se refiere a aspectos organizativos, de funcionamiento, sobre los productos o los servicios que presta la empresa o sobre los procesos que se siguen cuyo conocimiento por terceros puede tener una incidencia negativa en el desarrollo de la actividad de la empresa y en su posición en el mercado.

Algunos de los límites del derecho de acceso a los que se ha hecho referencia persiguen proteger determinados datos empresariales para evitar las negativas consecuencias a que se ha hecho referencia. En particular, se debe traer a colación la confidencialidad vinculada a los límites relativos a los derechos de propiedad intelectual e industrial y, eventualmente, la protección de los datos personales. Asimismo, debe tomarse en consideración el límite previsto en la legislación básica relativo a los intereses económicos y comerciales. Eventualmente también pueden ser de aplicación otros límites como el relativo a la investigación y la sanción de las infracciones administrativas.



En cualquier caso, como ya afirmado el Tribunal de Justicia de la Unión Europea, “no puede considerarse que cualquier informativa relativa a una sociedad y a sus relaciones comerciales pueda acogerse a la protección que debe garantizarse a los intereses comerciales” (Sentencia del TJUE de 15 de diciembre de 2011, asunto *CDC Hydrogene Peroxide/Comisión*, T-437/08, apartado 44). En esta dirección, la Sentencia del Tribunal Supremo de 30 de abril de 2012 pone como ejemplo de materias a incluir en el secreto comercial la documentación relativa a las características técnicas específicas de un nuevo producto, las líneas generales de una campaña publicitaria estratégica, una fórmula, un compuesto químico, el modelo para una máquina o el nombre de una empresa que se pretende absorber, pero en cambio no se incluiría de acuerdo con dicha sentencia en la noción de secreto comercial la relación de trabajos, trabajadores, maquinaria, facturación o cuenta de resultados de una empresa.

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 natural.

En primer lugar, se puede hacer referencia a la legislación en materia de defensa de la competencia, la Ley 15/2007, de 3 de julio, de Defensa de la Competencia que prevé que en cualquier momento del procedimiento se podrá ordenar, de oficio o a instancia de parte, que se mantengan secretos los datos o documentos que consideren confidenciales (art. 43). La divulgación o explotación de secretos industriales o empresariales a los que se haya tenido acceso legítimamente pero con deber de reserva o ilegítimamente sin autorización de su titular podrá ser considerado como una violación desleal de un secreto (art. 13 Ley 3/1991, de 10 de enero, de Competencia Desleal).

En segundo lugar, en materia de contratación pública, 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 atribuye a los empresarios la determinación de qué información incluida en la propuesta de licitación debe ser considerada como confidencial siendo esta declaración un límite al derecho de acceso. Al respecto, la Sentencia del Tribunal de Justicia de la Unión Europea de 14 de febrero de 2008 recuerda que para alcanzar el objetivo relativo a la apertura a la competencia no falseada en materia de contratos públicos “es necesario que las entidades adjudicadoras no divulguen información relativa a procedimientos de adjudicación de contratos públicos cuyo contenido pueda ser utilizado para falsear la competencia, ya sea en un procedimiento de adjudicación en curso o en



procedimientos de adjudicación ulteriores” (STJUE VAREC, de 14 de febrero de 2008, asunto C-450/06, apartado 35).

Pero los licitadores no podrán declarar que toda la información aportada tenga este carácter. Al respecto, podemos traer a colación la Resolución 62/2012 del Tribunal Administrativo Central de Recursos Contractuales en la que se afirma que “esta obligación de confidencialidad no puede afectar a la totalidad de la oferta realizada por el adjudicatario, habida cuenta de que el propio artículo 124.1 de la Ley de Contratos del Sector Público garantiza que este deber de confidencialidad no debe perjudicar el cumplimiento de las obligaciones en materia de publicidad e información que debe darse a candidatos y licitadores, obligaciones entre las que se encuentran incluidas las enumeradas en el artículo 135.4 de la propia LCSP (art. 151.4 TRLCSP)” a pesar de que la adjudicataria del contrato había señalado como confidencial la totalidad de la documentación que componía su proposición. Señala esta misma Resolución que “en el conflicto entre el derecho de defensa del licitador descartado y el derecho a la protección de los intereses comerciales del licitador adjudicatario se ha de buscar el necesario equilibrio de forma que ninguno de ellos se vea perjudicado más allá de lo estrictamente necesario”. Al respecto, el Auto del Tribunal Supremo de 15 de febrero de 2007 recuerda que “[s]e impone, pues, una valoración circunstanciada de cada caso concretamente examinado a fin de cohonestar de forma singularizada el derecho a la defensa y la protección de los intereses públicos y privados que conducen a las limitaciones de acceso al expediente administrativo”, “sin que sea dable una declaración formulada globalmente sobre la pertinencia o impertinencia de que conste en el expediente toda la documentación solicitada”.

En esta dirección, los órganos de contratación deben analizar si esta declaración de reserva es tan amplia que desnaturaliza el principio de transparencia y, por ende, la posibilidad de reutilizar la información de las licitaciones como recurso de las tecnologías del lenguaje natural.

Además, los órganos de contratación también pueden decidir no comunicar determinados datos relativos a la adjudicación cuando considere, justificándolo debidamente en el expediente, que la divulgación de esa información puede obstaculizar la aplicación de una norma, resultar contraria al interés público o perjudicar intereses comerciales legítimos de empresas públicas o privadas o la competencia leal entre ellas o se haya declarado el contrato como secreto o reservado (art.153 TRLCSP).

3 El proceso

Las tecnologías de procesamiento del lenguaje natural permiten realizar multitud de procesos como el análisis, la traducción, el resumen, la relación o la explotación de textos.

El procesamiento de lenguaje natural se basa en el uso de procesadores lingüísticos [2, 21].

En las próximas páginas se analizan los requisitos y las condiciones que deben concurrir en los procesadores lingüísticos y, en general, en el procesamiento del lenguaje natural.

3.1 El papel de las administraciones públicas en el desarrollo de procesadores del lenguaje natural

Las Administraciones públicas utilizan procesadores de lenguaje natural 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, encargan su desarrollo a empresas privadas. Más allá de sus propias necesidades, las Administraciones públicas pueden fomentar el desarrollo de procesadores del lenguaje natural 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 natural y la traducción automática en España. Asimismo, se fija como objetivo la creación de las Plataformas comunes de procesamiento de lenguaje natural y de traducción automática de las Administraciones Públicas. En particular, este objetivo se traduce en los siguientes objetivos específicos:

- dotar de herramientas comunes para el procesamiento del lenguaje natural y la traducción automática a las Administraciones Públicas;
- disponer de herramientas comunes para la generación, evaluación y explotación de los recursos lingüísticos;
- compartir infraestructuras y sistemas entre los proyectos innovadores que sean apoyados desde las instituciones públicas.

Así pues, el Plan de Impulso de las Tecnologías del Lenguaje atribuye a las Administraciones públicas diversas funciones relacionadas con el desarrollo de procesadores del lenguaje natural. En el marco del Plan de Impulso de las Tecnologías de la Información, las Administraciones públicas deben fomentar el desarrollo de estas infraestructuras por parte de la industria. Asimismo, deben



desarrollar infraestructuras para el procesamiento del lenguaje natural 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 natural 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.

Por lo que se refiere a la actividad de fomento de la industria de las tecnologías del lenguaje por parte de las Administraciones públicas, cabe tener presente no sólo lo previsto en la Ley 38/2003, de 17 de noviembre, General de Subvenciones sino, en particular por lo que ahora se analiza, la Ley 15/2007, de 3 de julio, de Defensa de la Competencia (LDC) y la normativa de desarrollo. En particular, 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.

En esta dirección pueden resultar de utilidad los criterios identificados por la Comisión Nacional de los Mercados y la Competencia para desarrollar la prueba de sopesamiento prevista en el Real Decreto 261/2008, de 22 de febrero, por el que se aprueba el Reglamento de Defensa de la Competencia, entre los que se incluye que el régimen de ayudas debe responder a un objetivo de interés general claramente definido o a la corrección de un fallo de mercado previamente identificado; que el régimen de ayudas debe ser el instrumento apropiado para alcanzar dicho objetivo, lograr un efecto incentivador (de forma que el beneficiario varíe su conducta) y resultar proporcional al problema afrontado (de manera que no se puedan lograr los mismos objetivos con un menor importe de ayuda o con un esquema menos distorsionante) o que la medida debe arrojar un saldo positivo en su impacto sobre el bienestar, compensando las eventuales distorsiones sobre la competencia y los flujos de comercio [9].

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 sus funciones.



Cuando las Administraciones públicas decidan ofrecer servicios en el mercado basados en las aplicaciones desarrolladas por ellas debe garantizarse la libre competencia respecto a desarrolladores del lenguaje natural 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. 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 [10]. 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.

La delimitación entre ambas funciones puede tener nuevamente un impacto en las normas de competencia si, por ejemplo, la Administración pública ofrece sus servicios bajo unas condiciones más favorables que las que pueda tener la industria para desarrollar esas mismas actividades u ofrecer al mercado los mismos servicios.

Para dar una primera respuesta a esta cuestión, se pueden traer a colación las reflexiones realizadas por la Comisión Europea al fijar los límites de la noción de tarea pública que recoge la Directiva RISP donde se plantea un problema que presenta ciertas analogías respecto a qué servicios o qué información de valor añadido pueden facilitar las Administraciones públicas basada en la información pública que poseen y que, eventualmente, ofrecen para que también pueda ser reutilizada. En particular, la Comisión Europea reconoce que la tarea debe de ser desarrollada por una entidad pública y debe ser definida por una ley o por otra norma o de acuerdo con la práctica común de un Estado miembro a pesar de que dicha entidad no tenga un encargo claro en esta dirección. Este es un aspecto importante en la medida en que cuanto más se amplíen estas nociones más se restringen las oportunidades de los desarrolladores privados.

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.

3.2 La contratación del desarrollo de los procesadores

Buena parte de los procesadores de lenguaje natural 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 natural.

En primer lugar, la relativa a los derechos de propiedad intelectual o industrial vinculados a dichas aplicaciones (que ya vimos en el primer capítulo y que tendremos ocasión de examinar más en detalle en el último subcapítulo). 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 (artículo 301). Sin embargo, los pliegos de cláusulas administrativas o en el documento contractual pueden disponer otra cosa. En cualquier caso, se hayan cedido o no los derechos de propiedad intelectual, se podrá autorizar el uso del procesador a todos los entes, organismos y entidades que conforman el sector público.

En segundo lugar, la relativa a la innovación que puede acompañar el desarrollo de estas aplicaciones. En estos casos, el TRLCSP dispone diferentes cuestiones que se sintetizan a continuación con la salvaguarda de que la futura transposición de la Directiva 2014/24/UE del Parlamento Europeo y del Consejo, de 26 de febrero de 2014, sobre contratación pública y por la que se deroga la Directiva 2004/18/CE que supondrá un cambio significativo de esta cuestión al haber incorporado la innovación como una política pública.

Asimismo, debe tenerse presente que la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación establece que “se impulsará la contratación pública de actividades innovadoras, con el fin de alinear la oferta tecnológica privada y la demanda pública, a través de actuaciones en cooperación con las Comunidades Autónomas y con las Entidades Locales, de acuerdo con lo señalado por la Ley 2/2011, de 4 de marzo, de Economía Sostenible”.

Además, no puede desconocerse que las PYMES tienen un papel significativo en el impulso de la innovación tal y como se desprende, entre otros, de la lectura del artículo 22.2 TRLCSP “los entes,



organismos y entidades del sector público velarán por la eficiencia y el mantenimiento de los términos acordados en la ejecución de los procesos de contratación pública, favorecerán la agilización de trámites, valorarán la innovación y la incorporación de alta tecnología como aspectos positivos en los procedimientos de contratación pública y promoverán la participación de la pequeña y mediana empresa y el acceso sin coste a la información, en los términos previstos en la presente Ley”.

En la actualidad, la innovación se puede canalizar a través de la compra pública pre-comercial o la compra pública de tecnología innovadora [11]. 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 (artículos 175 y siguientes):

- La compra pública pre-comercial persigue buscar soluciones a posibles demandas futuras respecto a las que no existe una solución en el mercado. Esta modalidad contractual busca impulsar actividades de investigación y desarrollo como la exploración y el diseño de soluciones, la creación de prototipos o el desarrollo de un volumen limitado de servicios en forma de prueba. La compra pública pre-comercial también se caracteriza porque los órganos de contratación y las entidades adjudicatarias comparten los riesgos y los beneficios de la investigación realizada. El artículo 4.1.r TRLCSP excluye la compra pública pre-comercial de su ámbito de aplicación por lo que serán los pliegos los que, de acuerdo con los principios generales de la contratación, los que concreten el régimen de cada contrato. Los derechos de propiedad intelectual o industrial estarán compartidos de acuerdo con lo que señale el pliego. Las soluciones alcanzadas podrán ofrecerse al mercado a través de una nueva licitación que se impulse.
- La compra pública de tecnología innovadora consiste en la compra de un bien o servicios que no existe en el momento de la compra pero que puede desarrollarse en un período de tiempo razonable a través del desarrollo de una tecnología nueva o mejorada [11]. A diferencia del supuesto anterior, la compra pública de tecnología innovadora está sometida al TRLCSP con la particularidad que con frecuencia los órganos de contratación utilizarán como procedimiento de adjudicación la negociación o el diálogo competitivo.

3.3 La normativa lingüística

El Plan de Impulso de las Tecnologías del Lenguaje define como objetivo general fomentar el desarrollo del procesamiento del lenguaje natural 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.

Únicamente en caso de la Ley 1/1998, de 7 de enero, de Política Lingüística de Cataluña se prevé explícitamente que el gobierno de la Generalidad debe favorecer, estimular y fomentar la investigación, producción y comercialización de todo tipo de productos en catalán relacionados con las industrias de la lengua, como son los sistemas de reconocimiento de voz, de traducción automática y similares u otros posibles de acuerdo con los avances tecnológicos (artículo 29).

Ni la Ley 3/1983, de 15 de junio, de normalización lingüística del Gallego, ni la Ley 3/1986, de 29 de abril, de normalización lingüística en las Illes Balears ni la Ley 10/1982, de 24 de noviembre, básica de normalización del uso del euskera recogen ninguna mención al respecto lo que no es obstáculo para que en el marco de la competencia genérica de la administración autonómica para el fomento del uso de la lengua. Por todas, se puede traer a colación el artículo 26 Ley 10/1982, de 24 de noviembre, “los poderes públicos vascos tomarán las medidas oportunas y los medios necesarios tendentes a fomentar el uso del euskera en todos los ámbitos de la vida social, a fin de posibilitar a los ciudadanos el desenvolvimiento en dicha lengua en las diversas actividades mercantiles, culturales, asociativas, deportivas, religiosas y cualesquiera otras”.

3.4 La interoperabilidad de las aplicaciones

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 natural 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 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”.

La Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público persigue promover la reutilización de sistemas y aplicaciones de propiedad de la Administración.

En particular, se dispone que las Administraciones públicas deben poner a disposición de cualquier Administración pública que lo solicite las aplicaciones que bien hayan desarrollado sus propios servicios técnicos, bien hayan sido objeto de contratación y de cuyos derechos de propiedad intelectual sean titulares (artículo 157).

La LRJSP establece que Administraciones públicas cedentes y cesionarias pueden acordar la repercusión del coste de adquisición o fabricación de las aplicaciones cedidas.

Para facilitar la reutilización y la interoperabilidad, la LRJSP también prevé que las aplicaciones pueden ser declaradas de fuentes abiertas. De este modo, se puede fomentar su reutilización además de dar mayor transparencia y, como reconoce la propia norma, fomentar la incorporación de los ciudadanos a la Sociedad de la información.

Todo esto está previsto y desarrollado en el Esquema Nacional de Interoperabilidad y en la Norma Técnica de Interoperabilidad. Estas normas dan cobijo a los objetivos definidos por el Plan de Impulso de las Tecnologías del Lenguaje relativos a la adopción de normas técnicas de interoperabilidad.

Desde el punto de vista de la eficiencia y la economía, la LRJSP prevé que antes de desarrollar una aplicación propia, las Administraciones públicas deben consultar el directorio general de aplicaciones de la Administración General del Estado si existe alguna aplicación que pueda ser reutilizada y que satisfaga total o parcialmente sus necesidades. En este caso, es necesario también que los requisitos de interoperabilidad y de seguridad lo permitan.

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.



La Administración General del Estado debe crear un directorio de aplicaciones, que debe integrar los directorios de aplicaciones del resto de Administraciones públicas. En él, deberían incluirse específicamente los procesadores desarrollados por las distintas Administraciones públicas.

Asimismo, la Administración General del Estado debe prestar apoyo para la libre reutilización de aplicaciones así como impulsar el desarrollo de aplicaciones, formatos y estándares comunes en el marco de los esquemas nacionales de interoperabilidad y seguridad.

En relación a la reutilización de aplicaciones, se debe poner de relieve que el Plan de Impulso de las Tecnologías del Lenguaje prevé como medida específica, la creación de las Plataformas comunes de procesamiento de lenguaje natural y de traducción automática de las Administraciones Públicas. Asimismo, se prevé la creación de sendas plataformas comunes de procesamiento de lenguaje natural y de traducción automática, basadas en componentes reutilizables de procesamiento del lenguaje natural y traducción automática para las Administraciones públicas.

3.5 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 (podéis ver la definición que propone el artículo 96.1 TRLPI). Para la protección de los programas, se consideraron tres posibilidades: la de las **patentes** (el programa como invención), la del **derecho de autor** (el programa como obra literaria) y el establecimiento de una **tutela ad hoc** (el programa como creación de un tipo nuevo). Esta última habría sido probablemente la ideal, pero la urgencia en asegurar la máxima cobertura internacional y la inexistencia de requisitos (formalidades) de registro fue determinante para optar por la opción del derecho de autor. Ésta fue la opción de la TRLPI española de 1987, la Directiva 91/250/CE, el Acuerdo sobre Derechos de Propiedad Intelectual relacionados con el Comercio de 1994 (art. 10.1) y el Tratado OMPI sobre derecho de autor de 1996 (art. 4): que consideran los programas como "obras literarias", a pesar de tener un régimen específico dentro de la ley (art. 96 y siguientes TRLPI).



La posibilidad de patentar programas es uno de los caballos de batalla actuales para la industria del software. En España no es posible (al menos formalmente) patentar un programa de ordenador, si bien es posible que éste forme parte de una invención patentable. También a nivel de la UE, el Convenio de la Patente Europea (CPE) excluye las patentes de programas de ordenador. Aun así, ello no significa que una funcionalidad concreta de un programa de ordenador – que pueda ser vinculado con un proceso o aplicación industrial- pueda ser objeto de patente.

Para la protección del programa de ordenador bajo el derecho de autor, serán de aplicación las normas generales vistas hasta ahora para el resto de obras -salvo para lo específicamente regulado en los artículos 95 y ss TRLPI. El régimen especial para la protección del software establece reglas concretas para determinar qué está protegido (**el programa, la documentación preparatoria, documentación técnica y manuales de uso**), quién es su titular (el coordinador del programa de ordenador creado como obra colectiva es autor) y qué límites son de aplicación.

El programa de ordenador puede quedar protegido por el TRLPI siempre y 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 (incluyendo la interfaz, definida como “la parte del programa que establece la conexión entre el código y los componentes físicos del ordenador” (cons. 10 y 11 Directiva 91/250/CE). 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 (art. 96.1 TRLPI).

En principio, para determinar la **autoría y la titularidad** del programa de ordenador, aplicaremos las mismas reglas generales previstas para el resto de obras. Sin embargo, el art.97 prevé algunas especificidades.

Artículo 97. Titularidad de los derechos.

1. Será considerado autor del programa de ordenador la persona o grupo de personas naturales que lo hayan creado, o la persona jurídica que sea contemplada como titular de los derechos de autor en los casos expresamente previstos por esta Ley.
2. Cuando se trate de una obra colectiva tendrá la consideración de autor, salvo pacto en contrario, la persona natural o jurídica que la edite y divulgue bajo su nombre.
3. Los derechos de autor sobre un programa de ordenador que sea resultado unitario de la colaboración entre varios autores serán propiedad común y corresponderán a todos éstos en la proporción que determinen.
4. Cuando un trabajador asalariado cree un programa de ordenador, en el ejercicio de las funciones que le han sido confiadas o siguiendo las instrucciones de su empresario, la titularidad de los derechos de explotación correspondientes al programa de ordenador así creado, tanto el

programa fuente como el programa objeto, corresponderán, exclusivamente, al empresario, salvo pacto en contrario.

5. La protección se concederá a todas las personas naturales y jurídicas que cumplan los requisitos establecidos en esta Ley para la protección de los derechos de autor.

Así pues, dejando a salvo la autoría individual (autor es quien crea -art.5 TRLPI) y la coautoría (obra en colaboración – art.7 TRLPI), 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 no sólo de titular (como hacía el art.8 TRLPI) sino 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 más amplio, ya que no se limita a lo "necesario para el ejercicio de la actividad habitual del empresario en el momento de la entrega de la obra" ni exige que "en ningún caso podrá el empresario utilizar la obra o disponer de ella para un sentido o fines diferentes" (como hacía el art.51 TRLPI).

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:

Art.301.2 TRLCSP (RDL 3/2011):

Salvo que se disponga otra cosa en los pliegos de cláusulas administrativas o en el documento contractual, 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 contratante.

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 [12] (p.18-19):

“[Entidad] adquirirá la propiedad de todo los productos (tanto software en cualquier forma o fijación, así como datos y/o información, incluida la documentación preparatoria, especificaciones, presentaciones, dll, scripts, etc.) que sean elaborados por el adjudicatario, [incluidos sus empleados y en su caso cualquier empresa subcontratada], en ejecución del Contrato (en adelante, “Resultados”), y, en particular, todos los derechos de propiedad intelectual y/o industrial que deriven de los mismos. Se incluye en esta cesión cualquier derecho de explotación de dichos Resultados, en cualquier modalidad y bajo cualquier formato, para todo el mundo, reservándose [Entidad] cualquiera otras facultades anejas a los derechos de propiedad intelectual e industrial.

[Entidad] será titular de todos los derechos referidos en el párrafo anterior por la duración máxima de los derechos, para todo el mundo y la única entidad que, por tal concepto, podrá explotar los Resultados, antes o después de la terminación del Contrato, correspondiendo a los autores y/o inventores de los mismos únicamente los derechos que les reconoce el artículo 14 de la Ley de propiedad intelectual y/o artículo 14 Ley de Patentes en su caso. A los efectos previstos en el párrafo anterior, el adjudicatario se compromete a la entrega a [Entidad] de toda la documentación técnica y materiales generados en el proceso de análisis, diseño, desarrollo, implantación y realización de pruebas, en cuyo poder quedarán a la finalización del Contrato, no pudiendo el adjudicatario utilizarla para otras personas o entidades salvo autorización expresa por parte de [Entidad].

[Entidad] puede decidir poner a disposición de terceros los programas de ordenador, datos e información resultantes de este contrato, usando para ello la Licencia Pública de la Unión Europea (EUPL) (<http://joinup.ec.europa.eu/software/page/eupl/licence-eupl>), u otras licencias que garanticen lo dispuesto en el artículo 16 del Real decreto 4/2010.

En el caso de que el adjudicatario incorpore a los Resultados productos, módulos o elementos propiedad de terceros sujetos a una licencia de Aplicación de Fuentes Abiertas (según lo definido en el Anexo de la Ley 11/2007), deberá comprobar y garantizar que los mismos podrán ser distribuidos a terceros como parte de los Resultados bajo la correspondiente licencia EUPL (u otra licencia que garantice lo dispuesto en el artículo 16 del Real decreto 4/2010). Las excepciones sobre este requerimiento sólo serán posibles previa autorización de [Entidad]. En el caso de que el adjudicatario incorpore a los Resultados productos, módulos o elementos propiedad de terceros sujetos a otro tipo de licencia, deberá obtener previamente de sus titulares los derechos suficientes para ceder la propiedad de los Resultados a [Entidad] de acuerdo con las presentes condiciones, y para que esta entidad pueda realizar los licenciamientos referidos en el párrafo anterior.”

El **plazo de protección** de un programa de ordenador se rige por las normas generales (art. 98 TRLPI): 70 años *post mortem auctoris* o, cuando el autor es persona jurídica, 70 años desde la divulgación.

En cuanto a los derechos, a pesar de que el art.99 sólo se refiere a los **derechos de explotación** (no se mencionan los derechos simple remuneración, ni los derechos morales), normalmente se



considera que el autor (persona física o jurídica) del programa también podrá gozar de los derechos morales que el art.14 TRLPI da al autor.

La explotación de un programa de ordenador puede hacerse a través de una gran variedad de licencias. Existe la posibilidad de utilizar licencias (**End User License Agreements**) con los "usuarios legítimos" (usuarios licenciados); En el 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 – sin que sea posible su reutilización.

La explotación de un programa de ordenador puede efectuarse también 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*. 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. El objetivo último del software libre es, pues, asegurar que la protección jurídica de los **programas de ordenador** bajo el régimen del *copyright* no sea un obstáculo para la subsiguiente reutilización del software. Y para ello, se utiliza el instrumento básico que ofrece el régimen del *copyright* (y que también utilizan los titulares de software propietario): la licencia.

Las licencias "libres" permiten a los usuarios del software hacer cualquier uso de éste (para cualquier finalidad), copiarlo y modificarlo (para adaptarlo o mejorarlo) y redistribuirlo (tanto en formato original, como en versión modificada). Todo ello, con **una única condición**: que se dé crédito al autor y se identifique la autoría de las subsiguientes modificaciones que se efectúen en el programa.

Más allá de este **mínimo común a todas las licencias de software libre**, existen variedades importantes. Por ejemplo, la licencia GNU-GPL (*General Public License*) obliga a que cualquier nueva versión del programa se distribuya **bajo la misma licencia GNU-GPL**, de manera que se asegure que los nuevos programas o versiones (software derivado) continuarán siendo "libres" y de que nadie se apropiará del esfuerzo ajeno en beneficio propio. El *copyleft* es, pues, una obligación contractual que obliga a comercializar el software derivado bajo la misma licencia GPL e impide que se privatice, que



se convierta en software "propietario". No todas las licencias software libre son, propiamente, *copyleft*.

Licencia **GNU-GPL** es la **licencia *copyleft*** por excelencia. La GNU-GPL se distingue claramente de las licencias de software "propietario" que permiten al usuario legítimo del software instalarlo y utilizarlo en una unidad concreta, pero no permiten ningún acto de explotación del mismo. 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.

La licencia se adjunta directamente al código fuente (licencia y obra son inseparables) y debe reproducirse siempre, en todo el software redistribuido.

Open Source

No hay que confundir "libre" con "gratis". Aunque inicialmente, software "libre" y *copyleft* eran sinónimos (ya que la casi totalidad del software "libre" se distribuía bajo la licencia *copyleft* GNU-GPL), a finales de los noventa, el movimiento del software "libre" vivió un momento de crisis con la formulación de los postulados de la Open Source Initiative (OSI) y la definición del código fuente abierto (*open source*). De esta manera, se quiere facilitar a usuarios, programadores y especialmente a empresas, una mejor comprensión del software libre y, en última instancia, promover nuevos modelos de negocio basados en el mismo y atraer la inversión empresarial. Con este objetivo, 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).

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"). Pero también deberemos tenerlas en cuenta a la hora de decidir como sujetar el procesador creado a la reutilización por parte del público.



Salvo que venga vinculado por una obligación contractual de *copyleft* por haber utilizado software de GPL, el titular del derecho de autor sobre el programa tiene libertad para decidir cómo proceder a la explotación / reutilización del mismo: ya sea mediante licencias tradicionales con usuarios o mediante licencias públicas (como puede ser la GPL). Es decir, el procesador -y no solo los resultados- también podrá quedar sujeto a la reutilización y para ello, la GPL es una opción más (que – recordemos- siempre obliga al *copyleft*).

De hecho, el RD 4/2010, de 8 enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica, establece las condiciones de licencia para la reutilización del software en los siguientes términos:

CAPÍTULO VIII

Reutilización y transferencia de tecnología

Artículo 16. Condiciones de licenciamiento aplicables.

1. Las condiciones de licenciamiento de las aplicaciones y de la documentaci de Interoperabilidad en el ámbito de la Administración Electrónica, establece las condiciones el procesador -y no solo los resultados- también y que éstas puedan poner a disposición de otras Administraciones públicas y de los ciudadanos, sin contraprestación y sin necesidad de convenio, tendrán en cuenta que el fin perseguido es el aprovechamiento y la reutilización, así como la protección contra su apropiación en exclusiva por parte de terceros, en condiciones tales que eximan de responsabilidad al cedente por el posible mal uso por parte del cesionario, as de licenciamiento de las aplicaciones y de la documentaci de Interoperabilidad en el ám, ni de compensación alguna en caso de errores en la aplicación.

2. Las administraciones utilizarento de las aplicaciones y de la documentaci de Interoperabilidad en el ám, ni de compensación alguna en caso de errores en la aplicación. el procesad

a) Pueden ejecutarse para cualquier propósito.

b) Permiten conocer su ca cualquier p

c) Pueden modificarse o mejorarse.

d) Pueden redistribuirse a otros usuarios con o sin cambios siempre que la obra derivada mantenga estas mismas cuatro garantón a

3. Para este fin se procurará la aplicación de la Licencia Pública de la Unión Europea, sin perjuicio de otras licencias que garanticen los mismos derechos expuestos en los apartados 1 y 2.

La EUPL (Licencia Pública de la Unión Europea) fue redactada en el 2008 a instancias de la Comisión Europea para la distribución de software de los organismos de la UE. Esta licencia es una licencia “open source” que no obliga propiamente al *copyleft* (como hace la GPL) sino que simplemente obliga a proporcionar el código fuente (o indicar donde se puede conseguir de forma rápida y fácil). Para más información, nos remitimos a la [12] (p.23-25).

A modo de resumen: a los efectos de favorecer la reutilización del programa de ordenador, son las licencias con *copyleft* “suave” (como la EUPL o la Lesser GPL que se utiliza para los manuales y



documentos que acompañan al programa de ordenador) y no las de *copyleft* “fuerte” (como la GPL) las que mejor se acomodan a los objetivos de reutilización.

4 El resultado

La última etapa del procesamiento del lenguaje natural 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.

En función de quien haya llevado a cabo el procesamiento, así como del origen de los recursos utilizados, las posibilidades serán muy diversas, con lo cual es imposible un estudio pormenorizado de ellas. Por ello, en adelante se establecen los criterios generales que deberían ser considerados y aplicados -según corresponda- a cada escenario concreto.

4.1 La reutilización de los resultados

Más allá de las aplicaciones (programas de ordenador), los resultados obtenidos por el procesamiento del lenguaje natural 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.

Así pues, en el segundo y tercer casos, además de obtener la autorización administrativa para reutilizar la información del sector público (ver siguiente subcapítulo), debe haber una licencia por la que se concreten las condiciones a las que se someta esta reutilización

En relación con estas condiciones, se tiene que partir de la idea de que la LRISP trata de favorecer la reutilización de la información del sector público, y el mercado comunitario de contenidos basado en información del sector público. Por este motivo, apuesta porque las condiciones incorporadas en las licencias respeten los criterios siguientes:

- Sean claras, justas y transparentes.
- No restrinjan las posibilidades de reutilización ni limiten la competencia.
- No sean discriminatorias para categorías comparables de reutilización.
- No contengan acuerdos exclusivos. De todos modos, se prevé que, en caso de que sea necesario para la prestación de un servicio de interés público, es posible otorgar derechos exclusivos, que tienen que ser transparentes y públicos. En este caso, se tiene que hacer una revisión periódica, y en todo caso cada tres años, de la permanencia del motivo que justificó la concesión de esta exclusividad.

Las condiciones que se establezcan, que tienen que respetar los criterios a los cuales se acaba de hacer mención, 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 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.

En relación a las condiciones, la Norma Técnica de Interoperabilidad señala que las que haya definido la Administración General del Estado podrán ser utilizadas como referencia por otras Administraciones públicas.

Así mismo se dispone que “dichas condiciones de reutilización globales a un organismo, disponibles en formato digital y procesables electrónicamente, podrán ser complementadas por condiciones específicas aplicadas a categorías de documentos o recursos de información concretos mediante licencias-tipo, disponibles en las mismas condiciones que las globales”.

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 -véase Capítulo 3 “Proceso”) pueden ponerse a disposición del público para su reutilización. En la medida en que estos resultados pueden ser objeto de protección por la propiedad intelectual, es conveniente establecer una licencia de PI que cubra tales aspectos.

Pero incluso antes de decidir la licencia que mejor se ajusta a los objetivos de reutilización, es fundamental asegurar que la Administración está legitimada para sujetar tales contenidos a la reutilización. Primero, debemos distinguir qué resultados pueden quedar sujetos a una licencia de reutilización (ya sea porque son titularidad de la Administración o porque no son contenidos protegidos por la propiedad intelectual – contenido en el dominio público o los contenidos excluidos de protección por el art.13 TRLPI) y cuales no (por ejemplo, porque se trata de contenidos protegidos de titularidad ajena para los cuales no se cuenta con la debida autorización para sujetarlos a la reutilización).



No se trata ahora de examinar el uso de recursos (privados o públicos) para fines de tratamiento lingüístico (que fue objeto de estudio en el capítulo 2 – Recursos), sino de examinar la posible reutilización (y las condiciones) de los resultados de este tratamiento. A los efectos de la propiedad intelectual, estos resultados pueden ser obras “nuevas” (compuestas o derivadas) o simplemente consistir en reproducciones modificadas de los recursos utilizados para el tratamiento. Tanto en un caso como en el otro, para su reutilización la Administración deberá ser titular de los derechos de PI o haber obtenido la correspondiente autorización del titular. Por supuesto, si el resultado del tratamiento simplemente consiste en contenidos que se hallan en el dominio público, no será necesaria licencia de PI alguna para favorecer su reutilización.

Tal como ya apuntábamos en el capítulo 2 – Recursos, es posible que el tratamiento lingüístico que se haga de algunos recursos puede conllevar la creación de resultados que sean, a su vez, objeto protegido por la propiedad intelectual, y ello con independencia de que los recursos procesados fueran una “obra o prestación” protegida por la propiedad intelectual o que simplemente se tratara de contenidos en el dominio público. Cuando el tratamiento lingüístico realizado sea suficientemente original, estaremos ante una nueva obra, protegida. En concreto, es resultado del tratamiento podrá ser una obra compuesta o derivada, creada sin la intervención del autor de la obra utilizada o adaptada. La autoría y titularidad de esta obra compuesta o derivada (resultante del tratamiento), se regirá por las normas generales (autoría, coautoría, obra colectiva, creación asalariada, creación por encargo, etc).

Recordémoslos: quien haya realizado el tratamiento será el autor / titular de los resultados obtenidos, siempre y cuando estos puedan ser considerados “obra” a los efectos del TRLPI. Si el tratamiento ha sido realizado por encargo fuera de la Administración, ésta deberá haber obtenido a su favor una cesión de derechos de PI para poder sujetar estos resultados a la reutilización pública.














Si el contenido pre-existente (sujeto al tratamiento) no estaba protegido por el derecho de autor (porque ya se halla en el dominio público o porque se trataba de contenidos que nunca estuvieron protegidos), el titular de los resultados gozará de plenos derechos de propiedad intelectual sobre los mismos y podrá decidir cómo autorizar su reutilización.

Si el contenido pre-existente (recurso sujeto al tratamiento) estaba protegido por el derecho de autor o conexos (obras o prestaciones protegidas), la reutilización de estos resultados requerirá la previa autorización del titular de los mismos. Al tratarse de una obra compuesta o derivada, sin esta autorización, no será posible sujetar los resultados del tratamiento a la reutilización por parte del público.

Para la reutilización de los resultados del tratamiento lingüístico, las **licencias Creative Commons** pueden ser de utilidad. Recordemos que se trata de licencias estandarizadas (sus cláusulas 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 (NC)
- excluyendo la modificación de su obra (ND); o,
- permitiendo la modificación, a condición de que la nueva obra quede sujeta a la misma licencia (esta es la cláusula "copyleft") (SA).

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

		Reconocimiento (by): Autoriza cualquier uso o acto de explotación que se haga de la obra (incluidos los usos comerciales y las transformaciones).	
		Reconocimiento (by) - Sin obra derivada (nd): Excluye la transformación de la obra para crear otra nueva (obra derivada).	
		Reconocimiento (by) - No comercial (nc): Excluye la posibilidad de utilizar la obra con fines comerciales.	
			Reconocimiento (by) - Sin obra derivada (nd) - No comercial (nc): No se permite ni hacer obras derivadas, ni los usos comerciales.
		Reconocimiento (by) - Compartir igual (sa): 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.	
			Reconocimiento (by) - No comercial (nc) - Compartir igual (sa): 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

A la hora de determinar qué licencia de propiedad intelectual es la que mejor se ajusta al objetivo de la reutilización, la Administración deberá tener en cuenta que las más permisivas son las que mejor facilitan la reutilización (BY), mientras que las que no permiten los fines comerciales (NC) o la adaptación y transformación (ND) son las que menos contribuyen a la reutilización.

La licencia BY-NC podría restringirla al excluir el procesamiento que se realizara con fines directa o indirectamente comerciales (por ejemplo, obteniendo subvenciones o algún otro tipo de ayuda económica). También es necesario valorar el alcance de las licencias que no permiten la



transformación (=) a los efectos de autorizar el procesamiento lingüístico (respuesta que dependerá, como vimos anteriormente, de la calificación que se dé -como simple reproducción o como transformación- a las anotaciones y demás tratamiento que se realizan en los procesos de HLT). También es importante tener presente que al utilizar recursos sujetos a las licencias CC con la obligación del “copyleft” (*share-alike*), como la BY-SA o la BY-NC-SA obligarían a sujetar los resultados del procesamiento a la misma licencia con lo cual se puede estar restringiendo las oportunidades de reutilización de tales contenidos (al imponer una condición contractual de explotación a quien reutilice los resultados), especialmente, en el caso de la BY-NC-SA que no permite los fines comerciales. En este punto, no coincidimos del todo con la recomendación realizada en la Guía del ENI, # 78-79 (pag.25) que recomienda las licencias CC-BY-SA y CC-BY-ND como recomendadas para la reutilización de activos.

Además de estas 6 licencias tradicionales, existe una licencia CC0 (Public Domain Dedication) <https://creativecommons.org/publicdomain/zero/1.0/deed.es> **ES** que “suspende” los derechos de autor y *sui generis* sobre bases de datos y, como su nombre indica, pone los contenidos protegidos en el dominio público. En las *Directrices sobre las licencias normalizadas recomendadas, los conjuntos de datos y el cobro por la reutilización de los documentos*, la Comisión Europea recomienda el uso de esta licencia CC0 para fomentar la reutilización de contenidos por parte del sector público. Ello es en cierta medida, comprensible en tanto que estas Directrices sólo se refieren a la reutilización de contenidos que son titularidad de la Administración o que no están protegidos por el régimen de propiedad intelectual (dominio público) [13]. Sin embargo, para la reutilización de los resultados del tratamiento, posiblemente no sea esta licencia CC0 la que mejor se ajuste a la necesidad de salvaguardar la propiedad intelectual en todos los casos. La licencia CC0 no acaba de encajar con los sistemas de derecho de autor de la mayoría de países europeos (en concreto, el de España) donde la decisión de qué obras están protegidas y cuales no es una cuestión de derecho imperativo, y no es dispositivo: la ley (y no el autor) decide qué está protegido y cuando dejará de estarlo. Así pues, la licencia CC0 es válida para favorecer la reutilización de contenidos del sector público que no están protegidos por el derecho de autor (dominio público) o contenidos excluidos de protección (Art.13 TRLPI: normativa, resoluciones judiciales, actos de organismos públicos, etc); pero no parece la más recomendable para la reutilización de otros contenidos protegidos de titularidad pública y, menos aún, para la reutilización de contenidos protegidos de titularidad ajena para cuya licencia la Administración deberá contar con la previa autorización de los titulares.

En cualquier caso, es el mismo documento de *Directrices* la Comisión recomienda que cada Estado miembro examine “la posibilidad de desarrollar **una licencia abierta nacional adecuada**”.



Ciertamente, 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] J.-P. Triaille, J. d. M. d'Argenteuil, and A. de Francquen, "Study on the legal framework of text and data mining (TDM)," 2014.
- [2] Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información, "Plan de Impulso de las Tecnologías del Lenguaje," 2015.
- [3] Comisión Europea, "Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones sobre Datos abiertos. Un motor para la innovación, el crecimiento y la gobernanza transparente - COM(2011) 882 final," 2011.
- [4] Grupo de trabajo sobre protección de datos del artículo 29, "Dictamen 06/2013 sobre datos abiertos y reutilización de la información del sector público (ISP)," 2003.
- [5] Grupo de trabajo sobre protección de datos del artículo 29, "Dictamen de 3/2013 sobre limitación de la finalidad," 2003.
- [6] Agencia Española de Protección de Datos, "Orientaciones y garantías en los procedimientos de anonimización de datos personales," 2016.
- [7] Grupo de Trabajo sobre protección de datos del artículo 29, "Dictamen 05/2014 sobre técnicas de anonimización," 2014.
- [8] Agencia Española de Protección de Datos, "Guía para una evaluación de impacto en la protección de datos," 2014.
- [9] Comisión Nacional de los Mercados y la Competencia, "Informe de la CNMC sobre las ayudas públicas en España -2016 IAP/CNMC/001/16," 2016.
- [10] B. Lundqvist, Y. Forsberg, M. de Vries, and M. Maggiolino, "Open Data and Competition Law. Some issues regarding access and pricing of raw data," *Masaryk University Journal of Law and Technology*, vol. 9, 2015.
- [11] Ministerio de Economía y Competitividad, "Guía 2.0 para la compra pública de innovación," 2015.
- [12] Dirección de Tecnologías de la Información y las Comunicaciones, "Reutilización de activos. Guía de publicación y licenciamiento de activos," Ministerio de Hacienda y Administraciones Públicas, Madrid, 2015.
- [13] Comisión Europea, "Directrices sobre las licencias normalizadas recomendadas, los conjuntos de datos y el cobro por la reutilización de los documentos (2014/C 240/01)," 2014.

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