

## **Acuerdo Marco para la compartición de datos de las memorias de traducción generadas por la traducción de documentos de las Administraciones Públicas.**

### **1. INTRODUCCIÓN.**

Desde la adopción en 2003 del primer conjunto de normas sobre reutilización de la información del sector público, el volumen de datos generado ha aumentado exponencialmente en todo el mundo, al tiempo que se están generando y recopilando nuevos tipos de datos.

Paralelamente, estamos asistiendo a una evolución permanente de las tecnologías del procesamiento del lenguaje para el análisis, explotación y tratamiento de datos.

Esta rápida evolución tecnológica permite la creación de nuevos servicios y aplicaciones basados en el uso, la agregación o la combinación de datos.

Los datos masivos (Big Data) recopilados a través de distintas herramientas tecnológicas o extraídos de grandes colecciones de información en distintos formatos son capaces de generar nuevo conocimiento en los sectores más diversos, pero, a la vez, plantean problemas tanto de titularidad como de su uso posterior.

Las Administraciones públicas y los organismos y entidades que conforman el sector público son grandes productoras de datos de gran utilidad para el procesamiento del lenguaje natural. En el desarrollo de sus actividades generan ingentes cantidades de datos cuyo uso y explotación puede resultar de gran interés tanto para la propia Administración como para la industria.

El uso de estos datos por las Administraciones públicas combinado con las técnicas de minería de datos y de Big Data y las tecnologías de procesamiento del lenguaje natural puede facilitar la toma de decisiones públicas y la eficacia de las políticas públicas.

El presente Acuerdo Marco, se permite ser una recomendación a los estados miembro para la aplicación de licencias abiertas que remuevan los posibles obstáculos legales con el fin de incorporar estos tipos de datos para la creación de nuevo conocimiento.

En la actualidad ya viene siendo práctica que, distintos gobiernos, corporaciones y organismos internacionales productores de datos públicos se apoyen en estas licencias —en particular, las modalidades xxx xxx xxx (PPDL) y Creative Commons (CC) en las que se renuncia a todos los derechos sobre las bases de datos— como mecanismo adecuado para favorecer el uso y reutilización de estos grandes volúmenes de datos.

## **2. Porqué un Acuerdo Marco Europeo para la gestión de la compartición de datos.**

### ***Razones del acuerdo marco para la compartición de datos provenientes de memorias de traducción.***

#### **2.1. Planteamiento general.**

El AMSCODA TMX (**A**cuerdo **M**arco **S**obre **C**ompartición de **D**atos TMX) no es casual ni puramente episódico o puntual. Al contrario, responde a una necesidad que encuentra diversas motivaciones. Las administraciones públicas europeas generan miles de traducciones a diario , y esas traducciones generan bases de datos cuya liberación como datos ya es un mandato legal (Directiva Europa sobre Datos Abiertos).

En este momento, existen dos puntos de más relevancia y que reflejan la carencia de un consenso básico para una adecuada regulación y gestión de los problemas ligados a LA COMPARTICIÓN DE DATOS GENERADOS POR LAS TRADUCCIONES DE LOS TEXTOS GENERADOS POR LAS ADMINISTRACIONES PÚBLICAS. Es este **uno de los sectores que mayor cantidad de datos genera de esta clase** en el desempeño de su actividad habitual y que más invierte en infraestructuras de almacenamiento de datos a nivel mundial y en particular los datos generados por las memorias de traducción de los documentos propiedad de la Administración Pública, dadas las necesidades del

administrado, y la pluralidad de idiomas ya no sólo dentro de un territorio o país, sino en todos los países que conforman la UE.

Pues bien, tan importante como el cumplimiento de las recomendaciones para que los datos sean reutilizables e interoperables, es el establecimiento de sus condiciones legales de uso. Sin duda alguna, lo idóneo es que los datos sean totalmente accesibles y que su reutilización no requiera permisos concretos y así lo establece la Directiva Europea sobre Datos Abiertos.

Sin embargo, la realidad exige ciertas condiciones, y es por ello que se hacen necesarias las licencias, que en este caso de las memorias de traducción deben ser lo más abiertas posibles. Es nuestro objetivo prioritario establecer un marco de actuación que favorezca su mayor difusión y capacidad de reutilización, siendo más que recomendable que no se generen más tipos de licencia, sino todo lo contrario, simplificando los tipos como recomienda la *Agenda Digital Europea* (Comisión Europea, 2010).

En la actualidad, dos son las principales alternativas internacionales que impulsan el uso de licencias en el contexto de datos e información, y que regulan y promueven el libre acceso y uso de la información. Se trata de Creative Commons (CC), (xxx xxx xxx) ODC. Las primeras son aplicables tanto a datos como a documentos de otro tipo, mientras que las segundas únicamente se utilizan en la esfera de los datos. La tendencia internacional nos indica que para regular las condiciones de acceso y uso de los datos se están utilizando modelos de licencias abiertas, basadas principalmente en CC, aunque con variaciones adaptadas a las características propias de cada portal de datos y de cada país donde se apliquen.

Todo ello, en beneficio de una Administración cada día más electrónica que sea capaz de **medir el impacto de sus acciones en términos de retorno social de la inversión**, que es lo que se persigue conseguir con acciones como la que propone este Acuerdo

Marco Sobre Compartición de Datos generados por las memorias de traducción de textos traducidos por las Administraciones públicas.

Por tanto, no existiendo un consenso en la actualidad en torno a cómo se ha de regular el tratamiento, uso y compartición de los datos de las memorias de traducción generadas a partir de publicaciones de las administraciones públicas, se propone bien una opción abstencionista, remitiendo a las leyes particulares de cada Estado Miembro de carácter general ya existentes, o bien se produce una regulación particular para este hecho, o se propone un modelo regulador basado en "normas blandas", de carácter unilateral, no vinculantes y enfoque práctico tal como es un ACUERDO MARCO.

AMSCODA toma partido en todos los aspectos: En el primero dejando bien claro que el problema existe, que es real para las organizaciones y que hay que intervenir, si bien rechazando los excesos y los abusos. En el segundo, proponiendo un marco complementario y específico de regulación, así como de gestión sistemática, pero flexible, adaptable para todas las Administraciones, entes privados y Estados Miembro.

Para ello, es altamente recomendable y necesario en aras de una sociedad cada día más preocupada por el valor de los datos generados que los datos se pongan a disposición del ente responsable contratista o ente central de cada Estado Miembro, con cada uno de los contratos de servicio de traducción que se realicen, poniendo a disposición no sólo el texto o documentos de la traducción en cuestión sino la memoria de dicha traducción.

Asimismo, gracias a la implementación de la base de datos NEC TM, se establecen por ende unas condiciones que facilitan la reutilización de dichas memorias de traducción gracias a su accesibilidad, y sobre todo a que no serán sometidas a restricciones técnicas o jurídicas que limiten o dificulten dicha reutilización.

**3. Multidimensionalidad del problema de la compartición de datos de las memorias de traducción generadas por las Administraciones Públicas: fundamentos de la tutela ofrecida por el AMSCODA**

En el plano estrictamente de su naturaleza jurídica, este Acuerdo Marco pretende ser un Acuerdo cuya aplicación se desarrolle conforme a las prácticas y procedimientos de cada experiencia estatal y teniendo en cuenta prácticas ya comunes y no desconocidas para los actores de la industria de la traducción.

Atendiendo a las consideraciones anteriores, parece claro que no es posible reconocer a los Acuerdos Marco una eficacia jurídica directa como si de una norma comunitaria más se tratara, de fuente autónoma en este caso, pero tampoco cabe infravalorar su alcance como reglas de cumplimiento puramente voluntario, estrictamente dependientes del poder de los sujetos firmantes para vincular a las organizaciones.

Así pues, AMSCODA TMX entendido como Acuerdo Autónomo pretende ser más bien un instrumento práctico, esto es, un conjunto de pautas orientadoras que concretan en reglas efectivas para la gestión, centralización y compartición de datos, siendo el Acuerdo Marco estrictamente complementario a la legislación, comunitaria y nacional.

De este modo, un Acuerdo no normativo, como es AMSCODA TMX, puede constituir, y aquí constituye, un útil instrumento de concreción e interpretación clarificadora de la compartición de los datos generados por las memorias de traducción de los trabajos realizados por los licitadores. La función de AMSCODA es, pues, clarificar y especificar el derecho de las Administraciones Públicas a solicitar la totalidad de los datos generados en los contratos de traducción, su propiedad sobre los textos originales y la traducción como servicio contratado y derivado de un original, reflejando el consenso de los actores implicados, entendiendo como los implicados de un lado a las empresas licitadoras y de otro el ente administrativo garante de cada Estado Miembro y finalmente el organismo central de la UE que a tal efecto se determine.

En este sentido, debe recordarse que la configuración más acreditada de este modelo o paradigma de juridicidad ilustra la existencia de instrumentos que, sin ajustarse a tipos tradicionales de norma jurídica, no pueden tampoco excluirse del mundo del Derecho, y por tanto de la obligatoriedad, para expresar un firme compromiso de los sujetos

Acuerdo Marco para la compartición de datos de las memorias de traducción generadas por la traducción de documentos de las Administraciones Públicas. AMSCODA TMX. PROYECTO NEC TM.

firmantes que llevan a cabo el contrato de servicios en el ejercicio de sus competencias bajo un procedimiento legalmente previsto, poniendo en práctica un instrumento dotado de una mínima o relativa juridicidad.

El contenido fundamental del AMSCODA se centra más en la fijación de modelos y pautas conformando un instrumento típico del Derecho Reflexivo.

En última instancia, este paradigma de regulación concretado por AMSCODA TMX se enmarca de plano dentro de los nuevos enfoques y nuevos métodos de regulación y de acción propuestos en la actualidad, además de por la UE, la cual apuesta más por instrumentos que transformen la norma legal en práctica que por más normas, siendo además el acceso, uso y posible reutilización de la información contenida en las bases de datos regulado por las leyes, así como mediante contratos y/o licencias.

Desde la CEE, la publicación de la comunicación sobre datos abiertos y la propuesta de modificación de la Comisión Europea, conforma una política de apertura de datos y de impulso de un mercado de información que tiene su punto central en la aprobación de la Directiva 2003/98/CE del Parlamento Europeo y del Consejo de 17 de noviembre de 2003 relativa a la reutilización de la información del sector público.

A ello deberemos añadir la revisión de la Directiva de Reutilización de la Información del Sector Público, presentada a finales de 2011, y que forma parte de la Agenda Digital Europea, iniciativa emprendida por la Comisión Europea (2010), estando dicha Agenda, orientada a favorecer los servicios en línea dentro de la Unión, y ocupando en ella un objetivo prioritario la apertura de datos públicos para su reutilización, la simplificación del sistema de licencias para el intercambio de contenidos y la implantación de normas de interoperabilidad.

## **5. La opción reguladora de AMSCODA**

Los datos públicos existentes en organismos públicos de la Unión Europea tienen un tratamiento específico en virtud de las Directivas de Reutilización de Información del Sector Público de 2003 y 2013, en las que se prevé la puesta a disposición del público para la reutilización de sus documentos “cualquier contenido sea cual sea el soporte (escrito en papel o almacenado en forma electrónica o como grabación sonora, visual o audiovisual” conservados por organismos del sector público para fines comerciales o no comerciales. Por tanto, el principio general es la libre puesta a disposición de tales contenidos, en caso de aplicar alguna tarifa, ésta “se limitará a los costes marginales en que se incurra para su reproducción, puesta a disposición y difusión” (art. 6.1).

En suma, AMSCODA se revela básicamente como una “Recomendación” basada en una mínima o relativa juridicidad, con base a contratos de servicios ya existentes y datos (memorias de traducción) generadas ex profeso. Esta “Recomendación” es la especificación práctica del deber de compartición de los datos generados en forma de memorias de traducción originadas por los servicios de traducción contratados por Administraciones Públicas, sea cual sea su soporte o formato.

## **6. Contenidos del Acuerdo Marco: el sistema de gestión de la cesión de datos de las MEMORIAS DE TRADUCCIÓN promovido por el acuerdo.**

Una memoria de traducción es una base de datos lingüística que va almacenando de forma continua las traducciones que generan profesionales para poder utilizarlas en el futuro en vistas a obtener una mayor consistencia terminológica y de estilo, además de ahorros por coincidencias totales o parciales entre nuevos textos y antiguas traducciones ya realizadas.

Las **memorias de traducción** son almacenes digitales compuestos por líneas de texto proveniente del contenido en la lengua original alineadas con su traducción en otras lenguas. Estos textos además pueden alinearse eficientemente por unidades de traducción. Las unidades de traducción que se almacenan junto con sus equivalentes se definen de forma variable (por frase, párrafo, palabra o grupo de palabras, etc.) siendo

típicamente la segmentación tras un signo de puntuación marcando el final de la frase o un salto de párrafo las más frecuentes ofrecidas por defecto en el entorno de los sistemas de traducción asistida por ordenador.

La principal función de las memorias de traducción (en adelante, MT) es extraer sugerencias totales o parciales de una frase y concordancias para términos. Durante la traducción, se buscan segmentos del idioma de origen en la base de datos de la MT. Si la MT posee un segmento en el idioma de origen que coincide exacta o parcialmente, este segmento se sugerirá al traductor, junto con la traducción recuperada de esa base de datos, y cualquier otra información adicional que haya sido guardada con el segmento en la base de datos. Las herramientas de traducción asistida por ordenador muestran grados de similitud (conocidos en la industria como *fuzzy match*). En la industria de la traducción, estas coincidencias resultan útiles a partir de un 50%-65% para los lingüistas, estableciéndose también baremos de precios relativos al esfuerzo que supone traducir frases completamente nuevas u otras para las que existen sugerencias y frases porcentualmente similares, dentro de un mismo contexto. Se considera un grado de similitud 100 % una coincidencia completa entre una frase solicitada y otra idéntica en la base de datos.

El formato libre de intercambio de memorias de traducción es el TMX (Translation Memory eXchange), generalmente en versión 1.4b. Se trata de un estándar XML del tipo DTD (*document type definition*). Fue creado por el comité OSCAR (Open Standards for Container/Content Allowing Re-use).

Mediante la aplicación del formato TMX es más viable la colaboración en proyectos de traducción de personas o empresas. El formato TMX también hace más fácil la migración de un sistema de traducción asistida a otro, lo que favorece la competitividad entre las tecnologías ofertadas y el desarrollo constante de las mismas para marcar diferencias con respecto a sus competidores. Como otros estándares abiertos, este formato se desarrolló con vistas a reducir los problemas de compatibilidad, impulsar la reutilización de los recursos lingüísticos, simplificar el intercambio de datos y estimular, de esta manera, la innovación tecnológica.



Este Acuerdo Marco, pretende establecer, como bien hemos venido definiendo, un “marco de buena conducta” para la obtención de los datos generados en los contratos de servicios de traducción por las Administraciones Públicas europeas, su organización mediante la implantación del sistema de software NEC TM y las iniciativas de centralización de datos que designe la Comisión Europea con vistas a la adopción de un Protocolo en materia de centralización de datos bilingües por parte de las Administraciones Públicas, su uso y beneficio para la sociedad en general y los propios Estados Miembro, ayuda a la creación de un corpus de Big Data nacional y la compartición de las partes que la Administración considere pertinentes a nivel europeo.

Para ello, el organismo competente del Estado en cuestión -por ejemplo la Secretaría de Estado específica para el Avance Digital- o el Ministerio competente en el Estado Miembro en cuestión será quien gestione la centralización de los datos alojados en las infraestructuras informáticas nacionales. Estos datos podrán compartirse, en caso de adhesión al presente Acuerdo Marco y si así lo determina la autoridad competente, a un nivel superior con el organismo pertinente de la Comisión Europea (ELRC-Share, por ejemplo).

Dicho lo cual, el primero de los pasos será la centralización de los datos provenientes de las memorias de traducción generadas por las Administraciones Públicas mediante la adopción del presente Acuerdo, y tras ello, y de conformidad con el mismo, su compartición con un organismo paneuropeo para beneficio de las propias AAPP de los demás Estados Miembro de la UE. El uso del software “National and European Central Translation Memory” (NEC TM) proporcionará la conexión de entrada y salida (E/S o input/output en inglés) para las administraciones, las cuales podrán guardar de forma privada sus memorias de traducción, alojadas en un servidor central y podrán compartirlas con quienes decidan para efficientar y ahorrar costes en sus trabajos a nivel nacional, crear Big Data nacional y seleccionar los datos que deseen con el servidor central donde se aloje la memoria de traducción central europea.

Relación con los entes privados y cesión de los TMX

Acuerdo Marco para la compartición de datos de las memorias de traducción generadas por la traducción de documentos de las Administraciones Públicas. AMSCODA TMX. PROYECTO NEC TM.

Los trabajos de traducción realizados por aquellas empresas licitadoras conllevarán la aceptación de la cesión de las memorias de traducción generadas por las mismas como resultado de la prestación de su servicio. A tal fin, en el anuncio que aparece en el Boletín o Gaceta Oficial del Estado Miembro, región o municipalidad, es decir en el portal del contratista, deberá incluirse una cláusula que hará referencia al AMSCODA de NEC TM, incluyendo el CÓDIGO CPV, que determinará QUE LOS DATOS DE LA MEMORIA DE TRADUCCIÓN, QUEDARÁN ALOJADOS EN EL SERVIDOR de la AUTORIDAD PÚBLICA COMPETENTE a los efectos de que en un futuro, y si así se estimase por dicha autoridad nacional pudieren compartirse con el organismo de la Comisión Europea que se determine (ELRC-Share o iniciativa similar, por ejemplo).

**A los efectos, las empresas licitadoras entregarán junto con sus trabajos las memorias de traducción generadas, aplicándose el principio de irretroactividad, significando éste que no debe tener efectos hacia atrás en el tiempo, y que sólo operará esa obligatoriedad desde la adhesión y la firma de la cláusula por la que se comprometen a entregar junto con sus trabajos realizados, los datos generados por sus traducciones y que conforman las memorias de traducción.**

## **7. PROPIEDAD INTELECTUAL y CONDICIONES DE USO DE LAS BASES DE DATOS**

En materia de propiedad Intelectual, no cabe discusión alguna de que los datos que se generan en la realización del trabajo en cumplimiento del contrato, por el cual se ha adjudicado la traducción en cuestión, es total y exclusivamente de la Administración Pública autora o gestora del texto original objeto de la traducción.

No se puede desconocer el hecho de que en la medida en que 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.

Así pues, la adjudicataria cederá de forma exclusiva y sin que operen en esa cesión límites temporales o territoriales algunos, los derechos de cualquier tipo de documentación o datos generados independientemente de su soporte o formato,

entendiendo en esa cesión de derechos los de reproducción, distribución y transformación.

Por ende, los datos generados por dichas memorias de traducción no podrán ser utilizados por el adjudicatario con fines lucrativos, y solo podrán utilizarse para informar, si fuera el caso de los trabajos realizados.

#### Propiedad intelectual y condiciones de uso de las bases de datos.

La base de datos DGT-TM es propiedad exclusiva de la Comisión Europea. La Comisión cede gratuitamente y a escala mundial, durante todo el período de protección de estos derechos, sus derechos no exclusivos a los reutilizadores para todo tipo de usos que cumplan las condiciones establecidas en la Decisión de la Comisión, de 12 de diciembre de 2011, relativa a la reutilización de los documentos de la Comisión, publicada en el Diario Oficial de la Unión Europea L 330 de 14 de diciembre de 2011, páginas 39 a 42.

Toda reutilización de la base de datos o de los elementos estructurados contenidos en la misma debe ser identificada por el reutilizador, que tiene la obligación de indicar la fuente de los documentos utilizados: la dirección del sitio web, la fecha de la última actualización y el hecho de que la Comisión Europea conserva la propiedad de los datos.

Esta base de datos es, pues, óptima para poblar inicialmente la versión de NEC TM de un Estado Miembro.

## **8. PROTECCIÓN DE DATOS.**

### **APLICACIÓN DEL REGLAMENTO UE 2016/679 y de REGLAMENTO 2018/1725**

En relación con los datos personales que tengan que ser objeto de tratamiento en base al cumplimiento del contrato por parte de la Administración Pública y el adjudicatario, ambas partes quedarán obligadas al cumplimiento de los siguientes Reglamentos Generales en materia de Protección de Datos.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, así como al cumplimiento de la normativa estatal vigente en materia de protección

Acuerdo Marco para la compartición de datos de las memorias de traducción generadas por la traducción de documentos de las Administraciones Públicas. AMSCODA TMX. PROYECTO NEC TM.

de datos de cada uno de los Estados Miembro que adopten este Acuerdo Marco, así como el Reglamento del Parlamento Europeo y del Consejo de 23 de Octubre de 2018, sobre la protección de las personas físicas con respecto al tratamiento de datos personales por parte de las instituciones, organismos, oficinas y agencias de la Unión y sobre la circulación de dichos datos.

## **REFLEXIÓN FINAL.**

La Administración Pública como impulsora de la industria de las tecnologías del lenguaje, con la creación de plataformas comunes de procesamiento del lenguaje natural y la traducción automática y el desarrollo de recursos para la Reutilización de Información del Sector Público (RISP), está obligada a desarrollar políticas de compartición de datos y a sentar las bases para que esa compartición sea real y efectiva entre todos los entes que participen en el proceso, velando además porque los participantes del proceso ajenos a la Administración adquieran el compromiso a la firma de la cláusula de aceptación de sus licitaciones, de obligarse a entregar junto con sus trabajos realizados, los datos generados por sus traducciones (memorias de traducción) conformando todos ellos un conjunto de recursos lingüísticos de valor inestimable.

**REFERENCIAS.** Ferrer-Sapena y otros (2011) en su artículo sobre el acceso a los datos públicos; Ferrer-Sapena y Peset (2012) acerca de la reutilización de datos culturales o Ramos Simón y otros (2012) en su estudio sobre los portales de datos europeos.

*Licensing Open Data: A Practical Guide* para el Higher Education Funding Council for England (Korn y Oppenheim, 2011). También son reseñables la *Guide to Open Data Licensing* (Open Knowledge Foundation, s.f.) y las directrices emanadas de la *Agenda Digital Europea* (Comisión Europea, 2010).