

III. OTRAS DISPOSICIONES

MINISTERIO DE LA PRESIDENCIA, JUSTICIA Y RELACIONES CON LAS CORTES

16648 *Resolución de 2 de agosto de 2024, de la Subsecretaría, por la que se publica el Convenio entre la Agencia Estatal Consejo Superior de Investigaciones Científicas, M.P., el Instituto Cervantes, la Asociación Española de Terminología, el Barcelona Supercomputing Center, la Universidad Politécnica de Madrid y la Real Academia de Ingeniería, para la realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA).*

La Presidenta de la Agencia Estatal Consejo Superior de Investigaciones Científicas, M.P., el Director del Instituto Cervantes, el Presidente de la Asociación Española de Terminología, el Director de Barcelona Supercomputing Center - Centro Nacional de Supercomputación, el Rector de la Universidad Politécnica de Madrid y el Presidente de la Real Academia de Ingeniería han suscrito un convenio para la realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA).

Para general conocimiento, y en cumplimiento de lo establecido en el artículo 48.8 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, dispongo la publicación en el «Boletín Oficial del Estado» del referido convenio como anejo a la presente resolución.

Madrid, 2 de agosto de 2024.—El Subsecretario de la Presidencia, Justicia y Relaciones con las Cortes, Alberto Herrera Rodríguez.

ANEJO

Convenio entre la Agencia Estatal Consejo Superior de Investigaciones Científicas, M.P., el Instituto Cervantes, la Asociación Española de Terminología, el Barcelona Supercomputing Center, la Universidad Politécnica de Madrid y la Real Academia de Ingeniería, para la realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA)

16 de julio de 2024.

INTERVIENEN

De una parte, doña Eloísa del Pino Matute, que actúa en su calidad de Presidenta de la Agencia Estatal Consejo Superior de Investigaciones Científicas, M.P. (en adelante, CSIC), con NIF Q2818002D y domicilio institucional en calle de Serrano, 117 Madrid 28006, que interviene en nombre y representación de este organismo en virtud de nombramiento efectuado mediante Real Decreto 498/2022, de 21 de junio de 2022, y de las competencias atribuidas por el artículo 11.2.i) del Estatuto del CSIC, contenido en el Real Decreto 1730/2007, de 21 de diciembre, por el que se crea la Agencia Estatal Consejo Superior de Investigaciones Científicas y se aprueba su Estatuto (BOE de 14 de enero de 2008).

De otra parte, don Luis Manuel García Montero, Director del Instituto Cervantes (en adelante, IC), cargo para el que fue nombrado por Real Decreto 933/2018, de 20 de julio, en nombre y representación del Instituto Cervantes, de acuerdo con lo previsto en el artículo 9 de la Ley 7/1991, de 21 de marzo, por la que se crea el Instituto Cervantes;

con domicilio a los efectos del presente convenio en la calle Alcalá, 49, 28014 Madrid, España, y NIF Q2812007I.

De otra parte, don Joaquín García Palacios, con DNI núm. ***2612**, en calidad de presidente de la Asociación Española de Terminología (AETER) según acuerdo de la Junta directiva de 3 de noviembre de 2023, y en virtud del poder de representación otorgado al presidente en el artículo 22 de los Estatutos de la Asociación, de fecha 18 de septiembre de 2015.

De otra parte, don Mateo Valero Cortés, actuando en nombre y representación del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (en adelante BSC-CNS), provisto del CIF S0800099D y con domicilio social sito en calle Jordi Girona 31, 08034, Barcelona, con el cargo de Director que ostenta en virtud de la renovación de su nombramiento por el Consejo Rector del BSC-CNS en la reunión celebrada el día 17 de diciembre de 2019 y en virtud de las competencias atribuidas al director en el artículo 15.i) de los Estatutos de este Consorcio.

De otra parte, la Universidad Politécnica de Madrid (en adelante UPM), con CIF Q2818015F, domiciliada en calle Ramiro de Maeztu, 7, 28040 Madrid y en nombre y representación don Guillermo Cisneros Pérez, mayor de edad, con DNI núm. ***9392**, en calidad de Rector, nombrado por Decreto 109/2020, de 25 de noviembre (BOCM núm. 289), del Consejo de Gobierno de la Comunidad de Madrid.

De otra parte, don Jaime Domínguez Abascal, que actúa en su calidad de Presidente de la Real Academia de Ingeniería (en adelante RAI), con NIF Q2862005B, y domicilio institucional en calle Don Pedro, 10, Madrid 28005, que interviene en nombre y representación de esta institución, en virtud de las facultades que le confiere el artículo 25.1.a de los estatutos de la Real Academia de Ingeniería aprobados por Real Decreto 397/2013 de 7 de junio (BOE de 10 de julio de 2013).

CSIC, IC, AETER, BSC-CNS, UPM y RAI podrán ser denominadas conjuntamente como «las partes» e individualmente como «la parte».

Las partes se reconocen capacidad jurídica y competencia suficientes para formalizar el presente convenio, y a tal efecto

EXPONEN

Primero.

Que el Plan de Recuperación, Transformación y Resiliencia traza la hoja de ruta para la modernización de la economía española, la recuperación del crecimiento económico y la creación de empleo, para la reconstrucción económica sólida, inclusiva y resiliente tras la crisis de la COVID-19, y para responder a los retos de la próxima década. Este Plan recibe la financiación de los fondos Next Generation EU, entre ellos el Mecanismo de Recuperación y Resiliencia.

Que el Plan de recuperación, estructurado en treinta y un componentes tras su modificación por Adenda, aprobada por la Comisión Europea el 2 de octubre de 2023 cuya decisión fue ratificada el 17 de octubre del mismo año por el ECOFIN (Consejo de Ministros de Finanzas de la UE), incluye un conjunto de reformas e inversiones que lo convierten en una herramienta para transformar el modelo productivo de España a través de la transición energética, la digitalización, la cohesión territorial y social, y la igualdad.

Segundo.

Que el Componente 16, Reforma 1, del Plan de recuperación, se refiere a la Estrategia Nacional de Inteligencia Artificial (ENIA), una estrategia lanzada el 23 de julio de 2020 por el Gobierno de España dentro de la Agenda España Digital y concebida para encauzar el proceso de transformación digital del país. En el eje estratégico número 3 de la ENIA, «Desarrollar plataformas de datos e infraestructuras tecnológicas

que den soporte a la IA», se encuadra la medida 14: «Impulso al Plan Nacional de Tecnologías del Lenguaje».

El objetivo general del Plan Nacional de Tecnologías del Lenguaje es desarrollar la investigación y desarrollo industrial del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en España, tanto en lengua española como en lenguas cooficiales.

Tercero.

Que el proyecto de creación del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) está directamente vinculado al desarrollo de una plataforma pública de terminología para dar acceso unificado a toda la terminología del español mediante un potente metabuscador, en el marco PERTE (Proyecto Estratégico para la Recuperación y Transformación Económica) Nueva Economía de la Lengua. El proyecto está basado en modelos de lenguaje y aprendizaje profundo y en el trabajo cualitativo de especialistas en distintos ámbitos y de lingüistas, los cuales incluyen herramientas para identificar neologismos y generar definiciones precisas de términos. Las terminologías se han convertido en activos esenciales para poder organizar los grandes volúmenes de texto existentes en las bases de datos documentales de las diferentes instituciones y organizaciones, y juegan un rol clave en el desarrollo y entrenamiento de modelos de lenguaje natural basados en inteligencia artificial. El desarrollo de este portal contribuye, por tanto, a la apertura de terminologías y a su uso compartido por la comunidad científica en primera instancia, y adicionalmente del tejido productivo español.

Cuarto.

Que la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) ejerce, entre otras funciones, la elaboración, formulación, coordinación y evaluación de la Estrategia Nacional de Inteligencia Artificial (ENIA), dentro del Componente 16, Reforma 1, del Plan de Recuperación, Transformación y Resiliencia, a través de la Dirección General de Digitalización e Inteligencia Artificial y dentro de ella, de la Subdirección General de Inteligencia Artificial y Tecnologías Habilitadoras Digitales, de acuerdo con lo dispuesto en el artículo 9.bis del Real Decreto 403/2020, de 25 de febrero, por el que se desarrolla la estructura orgánica básica del Ministerio de Asuntos Económicos y Transformación Digital. No obstante, a fecha de firma del documento, esta función recae en el Ministerio para la Transformación Digital y de la Función Pública (Real Decreto 829/2023, de 20 de noviembre, por el que se reestructuran los departamentos ministeriales), al cual pasa la SEDIA como Secretaría de Estado.

Quinto. Que el CSIC, de conformidad con el artículo 47 de la Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación («Ley de la Ciencia»), es un organismo público de investigación (OPI) constituido actualmente como Agencia Estatal y adscrito al Ministerio de Ciencia, Innovación y Universidades a través de la Secretaría General de Investigación, que tiene como objeto el fomento, la coordinación, el desarrollo y la difusión de la investigación científica y tecnológica, de carácter multidisciplinar, con el fin de contribuir al avance del conocimiento y al desarrollo económico, social y cultural, así como a la formación de personal y al asesoramiento a entidades públicas y privadas en esta materia. Dentro de las líneas de investigación, se encuentran las actuaciones de desarrollo y coordinación de líneas investigación en el ámbito del procesamiento del lenguaje natural.

Que el Instituto de Lengua, Literatura y Antropología (ILLA-CSIC) es un instituto del CSIC cuyo objetivo primordial es la investigación del patrimonio cultural hispánico en su triple dimensión antropológica, lingüística y literaria. Dentro de la línea de análisis lingüístico, los investigadores del Departamento de Lengua y Análisis del Discurso abordan el estudio de la lengua española tanto en su aspecto fónico (estudio del que se

ocupa principalmente el Laboratorio de Fonética) como en el plano de las relaciones gramaticales y léxicas. Presta atención a las propiedades compartidas entre el español y las demás lenguas naturales, así como a la variación lingüística de toda índole (histórica, dialectal, sociolingüística y de registro, textual) propia de una lengua tan ampliamente extendida en el espacio y el tiempo. Una subárea emergente de investigación atiende al procesamiento del lenguaje natural y las tecnologías lingüísticas, atendiendo a las necesidades del español en el desarrollo de la inteligencia artificial.

Sexto.

Que en virtud de Resolución de la Secretaría de Estado de Digitalización e Inteligencia Artificial por la que se regula la aportación de la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), al Consejo Superior de Investigaciones Científicas (CSIC), para la realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) en el marco de la Estrategia Nacional de Inteligencia Artificial, el PERTE Nueva Economía de la Lengua y el Plan de Recuperación, Transformación y Resiliencia –financiado por la Unión Europea– Next Generation EU (en adelante, la resolución), dicho órgano realizó al CSIC una aportación económica de un millón cuatrocientos cincuenta mil euros (1.450.000,00 €) para la realización del mencionado Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA).

Séptimo.

Que de acuerdo con la propuesta remitida a la SEDIA para la realización del mencionado Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) el CSIC contará con la colaboración de las entidades que se detallan en los siguientes exponendos.

Octavo.

Que el Instituto Cervantes (IC) es un organismo público creado al amparo de la Ley 7/1991, de 21 de marzo, que cuenta con una red de centros en el exterior y cuyos fines y objetivos primordiales son la promoción universal de la lengua española, la difusión de la cultura en el exterior y atender al patrimonio lingüístico y cultural que es común a los países y pueblos de la comunidad hispanohablante.

Noveno.

Que la Asociación Española de Terminología (AETER) es una asociación de ámbito nacional, sin ánimo de lucro, creada en 1992, cuyos fines se centran en el desarrollo y difusión de la terminología tanto en su vertiente teórica como en la aplicada, persiguiendo la armonización de la terminología del español con la terminología de las otras lenguas de España y con la del español de los distintos países hispanoamericanos. En su intento por contribuir al mantenimiento de una terminología actualizada en lengua española, integrada en los sistemas más actuales de comunicación y de gestión de la lengua, realizó en 2005 la primera propuesta para la realización de un sistema unificado para el acceso a la terminología del español, así como para conseguir que la terminología tenga los marchamos de calidad que pueden otorgarle la validación de los expertos en los distintos ámbitos de conocimiento y la sanción lingüística para la terminología nueva. Esta iniciativa, que ahora se concreta en TeresIA, figura como uno de sus proyectos fundamentales. Las instituciones y los profesionales integrados en esta asociación certifican la calidad de las aportaciones que se pueden hacer desde ella para la configuración y el desarrollo del proyecto en cada una de sus fases.

Décimo.

Que el Consorcio BSC-CNS se constituye como una entidad de derecho público integrada por la Administración General del Estado, por la Administración de la Generalitat de Catalunya y por la Universitat Politècnica de Catalunya y creada mediante convenio de colaboración de fecha 9 de marzo de 2015 para la gestión y promoción de la colaboración científica, técnica, económica y administrativa de las instituciones que la integran para el equipamiento y gestión del BSC-CNS, centro español de supercomputación y de I+D+I y agente de ejecución del Sistema Español de Ciencia, Tecnología e Innovación (de acuerdo con lo establecido en el artículo 3.4 de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación) cuyas infraestructuras forman parte del mapa nacional de Infraestructuras Científicas y Técnicas Singulares (ICTS).

Undécimo.

Que la Universidad Politécnica de Madrid (UPM) tiene entre otros fines la creación, desarrollo, transmisión y crítica de la ciencia, de la técnica y de la cultura, y el apoyo científico y técnico al desarrollo cultural, social y económico de la sociedad, así como la preparación para el ejercicio de actividades profesionales que exijan la aplicación de conocimientos y métodos científicos y técnicos o de creación artística y el apoyo y estímulo a la empresa pública y privada en el proceso de actualización e innovación tecnológica, según se establece en los Estatutos de la UPM aprobados por Decreto 74/2010, de 21 de octubre (BOCM de 15 de noviembre) y su Modificación parcial: Decreto 26/2018, de 3 de abril (BOCM de 9 de abril).

Duodécimo.

Que la Real Academia de Ingeniería (RAI) es una corporación de derecho público con personalidad jurídica propia, que tiene como fines promover la calidad y competencia de la Ingeniería española e impulsar el desarrollo de todas aquellas actividades que contribuyan a fomentar el estudio, la discusión y la difusión de las técnicas y de sus fundamentos científicos y sociales.

Que la RAI, concedora de que la lengua española es la segunda lengua del mundo por el número de hablantes que la tienen como lengua materna y que, además, se ha convertido en un importante vehículo de comunicación y desarrollo económico, desde su fundación consideró entre sus fines estatutarios la elaboración y mantenimiento de un Diccionario Español de Ingeniería (en adelante DEING). Tras la primera edición, que contaba con más de 50.000 entradas y 1.500 glosarios, la RAI ha llevado a cabo un proyecto de mantenimiento y actualización encaminado a incorporar nuevas acepciones de uso habitual en el español técnico.

Decimotercero.

Que el Departamento de Lengua Española de la Dirección General de Traducción de la Comisión Europea reitera su voluntad de contribuir, con sus terminólogos y con los recursos terminológicos abiertos de la Comisión Europea, a las actividades necesarias para la creación del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA).

Decimocuarto.

Que, teniendo en cuenta la experiencia del CSIC en la gestión de este tipo de proyectos, se considera oportuno que sea este organismo el encargado de coordinar la ejecución de los fondos asignados y la participación de las distintas instituciones españolas en el proyecto.

Decimoquinto.

Que el artículo 34.1 de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación establece que los agentes públicos de financiación o ejecución del Sistema Español de Ciencia, Tecnología e Innovación, incluidas las Administraciones Públicas, las universidades públicas, los organismos públicos de investigación de la Administración General del Estado, los consorcios y fundaciones participadas por las administraciones públicas, los organismos de investigación de otras administraciones públicas, y los centros e instituciones del Sistema Nacional de Salud, podrán suscribir convenios sujetos al derecho administrativo. Podrán celebrar estos convenios los propios agentes públicos entre sí, o con agentes privados que realicen actividades de investigación científica y técnica, nacionales, supranacionales o extranjeros, para, entre otras actividades, la realización conjunta de proyectos y actuaciones de investigación científica, desarrollo e innovación, así como la creación o financiación de centros, institutos, consorcios o unidades de investigación, e infraestructuras científicas.

Decimosexto.

Que según lo dispuesto en el artículo 59.1 del Real Decreto-ley 36/2020, de 30 de diciembre, por el que se aprueban medidas urgentes para la modernización de la Administración Pública y para la ejecución del Plan de Recuperación, Transformación y Resiliencia, la tramitación de los convenios que celebre la Administración General del Estado, sus organismos públicos y entidades de derecho público, vinculados o dependientes, para la ejecución de los proyectos con cargo a fondos europeos previstos en el Plan de Recuperación, Transformación y Resiliencia, se regirá por lo previsto en el capítulo VI del título preliminar de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, con las especialidades recogidas en dicho artículo 59.

En consecuencia y de conformidad con lo dispuesto en la Ley 40/2015, de 1 de octubre, en el Real Decreto-ley 36/2020, de 30 de diciembre, y en la Ley 14/2011, de 1 de junio, y con lo anteriormente expuesto, las partes Acuerdan suscribir el presente convenio que se regirá por las siguientes

CLÁUSULAS

Primera. *Objeto del convenio.*

El objeto de este convenio es recoger las condiciones de la colaboración de las distintas instituciones que participan en el proyecto de creación del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA), en el marco del Componente 16 del Plan de Recuperación, Transformación y Resiliencia.

Estas instituciones son el CSIC, el IC, la AETER, el BSC-CNS, la UPM y la RAI.

En el anexo I se recoge la distribución de fondos entre las partes firmantes de este convenio.

En el anexo II se recoge la memoria técnica de la participación de estas instituciones en las distintas tareas, en calidad de involucradas y coordinadoras según cada caso.

Segunda. *Obligaciones de las partes.*

Las partes firmantes de este convenio se comprometen a:

a) Desarrollar las actividades necesarias para la creación del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) de acuerdo con lo establecido en la memoria técnica recogida en el anexo II. Estas actividades se realizarán con cargo a los fondos transferidos por la resolución por la que se regula la aportación de la SEDIA al CSIC y del CSIC, en su función de coordinador, al resto de partes de este convenio, según el reparto establecido en el anexo I. De acuerdo a lo

establecido en la citada resolución y siguiendo lo dispuesto en el apartado cuarto de las instrucciones publicadas de la Resolución 1/2022, de 12 de abril, de la Secretaría General de Fondos Europeos, el CSIC actúa como entidad instrumental, el Ministerio de Asuntos Económicos y Transformación Digital como entidad ejecutora y la Dirección General de Digitalización e Inteligencia Artificial como órgano gestor del proyecto.

Las actuaciones derivadas de la creación del Portal TeresIA deberán llevarse a cabo por las partes en el periodo determinado para la ejecución de los fondos según la citada resolución, esto es, antes del 31 de diciembre de 2025.

Las partes se comprometen, con cargo a la financiación recibida a tal fin, a realizar las acciones que se les asignen con la finalidad de ayudar a la entidad ejecutora a desarrollar las medidas contenidas en el Plan de Recuperación y, en concreto, a cumplir con los hitos y objetivos asignados como parte de la Decisión de Ejecución del Consejo de acuerdo con sus indicadores de seguimiento, en la forma y los plazos previstos y bajo las directrices de la entidad ejecutora. Asimismo, facilitarán a la entidad ejecutora toda la información de seguimiento que ésta precise, y en particular proporcionarán la información que ha de suministrarse al sistema de información de gestión y seguimiento del Plan de Recuperación, Transformación y Resiliencia, conforme a lo dispuesto en la Orden HFP/1031/2021, de 29 de septiembre, y según se establezca en el manual de uso y funcionamiento del sistema de información, Manual de usuario CoFFEE-MRR.

b) Las partes se comprometen a cumplir con todos los principios transversales del Mecanismo de Recuperación y Resiliencia de acuerdo con el Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021, y con los principios de gestión específicos del Plan de Recuperación, Transformación y Resiliencia, previstos en el artículo 2 de la Orden HFP/1030/2021, de 29 de septiembre, con el mismo nivel de exigencia que el requerido por dicha Orden para las entidades ejecutoras, incluidas, entre otras, las obligaciones de:

a. Disponer de un Plan de medidas antifraude.

b. Cumplir con las normas de etiquetado verde y etiquetado digital de acuerdo con lo previsto en el Plan de Recuperación, Transformación y Resiliencia y el Mecanismo de Recuperación y Resiliencia, establecido por el Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo, de 12 de febrero de 2021. En este caso, al tratarse de una actuación de la medida C16.R1 contribuye en un 100 % a la consecución del objetivo digital (etiqueta 09.bis) y en un 0 % al objetivo de transición verde.

c. Realizar un análisis de riesgos en relación con posibles impactos negativos significativos en el medioambiente (do no significant harm, DNSH), seguimiento y verificación de resultado sobre la evaluación inicial.

En este sentido, se deberá garantizar el cumplimiento con las siguientes condiciones:

i. Mitigación del cambio climático: Los equipos que se utilicen cumplirán con los requisitos relacionados con el consumo energético establecidos de acuerdo con la Directiva 2009/125 / EC para servidores y almacenamiento de datos, o computadoras y servidores de computadoras o pantallas electrónicas. Para la instalación de las infraestructuras IT, se seguirá la versión más reciente del Código de conducta europeo sobre eficiencia energética de centros de datos, o en el documento CEN-CENELEC CLC TR50600-99-1 «Instalaciones e infraestructuras de centros de datos - Parte 99-1: Prácticas recomendadas para la gestión energética».

ii. Adaptación al cambio climático: Se realizará una evaluación del riesgo climático y la vulnerabilidad de las instalaciones de infraestructuras IT (tecnología de la información) y, si procede, se establecerán las soluciones de adaptación adecuadas para cada caso.

iii. Uso sostenible y la protección de los recursos hídricos y marinos: Los riesgos de degradación ambiental relacionados con la conservación de la calidad del agua y la prevención del estrés hídrico se identificarán y abordarán con el objetivo de lograr un buen estado del agua y un buen potencial ecológico, tal como se define en el artículo 2, puntos 22 y 23, del Reglamento (UE) 2020 / 852, de conformidad con la Directiva 2000/60 / CE del Parlamento Europeo y del Consejo, y un plan de gestión del

uso y la protección del agua, desarrollado en virtud de la misma para la masa o masas de agua potencialmente afectadas, en consulta con las partes interesadas pertinentes.

iv. Transición a una economía circular, incluidos la prevención y el reciclaje de residuos: Los equipos utilizados cumplirán con los requisitos de eficiencia de materiales establecidos de acuerdo con la Directiva 2009/125 / EC para servidores y almacenamiento de datos, u ordenadores y servidores de ordenadores o pantallas electrónicas. Los equipos no contendrán las sustancias restringidas enumeradas en el anexo II de la Directiva 2011/65 / UE, excepto cuando los valores de concentración en peso en materiales homogéneos no superen los enumerados en dicho anexo. Al final de su vida útil, el equipo se someterá a una preparación para operaciones de reutilización, recuperación o reciclaje, o un tratamiento adecuado, incluida la eliminación de todos los fluidos y un tratamiento selectivo de acuerdo con el anexo VII de la Directiva 2012/19/ UE.

v. Protección y restauración de la biodiversidad y los ecosistemas: Se asegurará que las instalaciones de infraestructuras IT no afectarán negativamente a las buenas condiciones y la resiliencia de los ecosistemas, tampoco al estado de conservación de los hábitats y las especies, en particular, los espacios de interés de la Unión. Por ello, cuando sea preceptivo, se realizará la Evaluación de Impacto medioambiental, de acuerdo con lo establecido en la Directiva 2011/92/EU.

d. Firmar los modelos del anexo IV de la Orden HFP/1030/2021, de 29 de septiembre:

- Modelo de declaración de ausencia de conflicto de intereses (DACI).
- Modelo de declaración de cesión y tratamiento de datos en relación con la ejecución de actuaciones del plan de recuperación, transformación y resiliencia (PRTR).
- Compromiso de cumplimiento de principios transversales mediante la firma del Modelo declaración de compromiso en relación con la ejecución de actuaciones del plan de recuperación, transformación y resiliencia (PRTR).

c) En caso de subcontratar parte o toda la actividad financiada con cargo al Mecanismo de Recuperación y Resiliencia, la entidad beneficiaria de la ayuda, ya sean contratistas o subcontratistas, tendrán que cumplir con la normativa europea y nacional que les resulte aplicable y, en particular, con las obligaciones reflejadas en la cláusula segunda b) de este convenio, así como con el resto de las obligaciones que se derivan del Reglamento del Mecanismo de Recuperación y Resiliencia, especialmente en materia de etiquetado digital y verde, principio de no causar daño significativo, evitar conflictos de intereses, fraude, corrupción, comunicación, transferencia de datos y publicación obligatoria, cuando corresponda, en la Plataforma de Contratos del Sector Público («PLACSP») o en la herramienta de contratación aplicable en cada caso.

d) Llevar un inventario de los bienes, materiales o inmateriales, adquiridos con cargo a las aportaciones previstas. El inventario deberá registrar, para cada elemento que se incluya en los apartados correspondientes, información sobre cada uno de los gastos imputados, concepto, importes abonados, fechas de los gastos y de los pagos, y denominación y NIF o equivalente de los proveedores o destinatarios, de acuerdo con lo recogido en el artículo 22, letra d) del Reglamento (UE) 2021/241.

e) Cumplir con las obligaciones en materia de información, comunicación y publicidad contempladas en el artículo 34 del Reglamento (UE) 2021/241, relativo al Mecanismo de Recuperación y Resiliencia, sobre la financiación europea en las medidas incluidas en el Plan de Recuperación, Transformación y Resiliencia.

Remitir con la periodicidad, procedimiento y formato, según se establece en las Órdenes HFP 1030/2021 y HFP 1031/2021, y con referencia a las fechas contables que establezca la Autoridad Responsable del Mecanismo de Recuperación y Resiliencia (MRR), información sobre el seguimiento de la ejecución contable de los gastos financiados con estos fondos. La información se emitirá de conformidad con lo dispuesto en el artículo 46.2 del Real Decreto-ley 36/2020, de 30 de diciembre. Las partes remitirán esta información al CSIC con tiempo suficiente para que, en su calidad de entidad

instrumental del subproyecto TeresIA, a efectos de la ejecución de los fondos MRR, el CSIC pueda tener la información exigida completa a efectos de que pueda enviársela a la entidad ejecutora en los plazos establecidos.

f) Cumplir los requisitos de pista de auditoría: obligación de guardar la trazabilidad de cada una de las actuaciones y cada uno de los perceptores finales de los fondos, en los términos definidos por el artículo 22.2.d) del Reglamento (UE) 2021/241 y las Órdenes Ministeriales HFP/1031/2021 y HFP/1030/2021, de 29 de septiembre. De conformidad con el artículo 8.1 de la Orden HFP/1030/2021, los requerimientos en relación con la identificación de los perceptores finales sean personas físicas o jurídicas son:

- NIF del perceptor final.
- Nombre de la persona física o razón social de la persona jurídica.
- Domicilio fiscal de la persona física o jurídica.
- Acreditación, en caso de perceptores finales que desarrollen actividades económicas, de la inscripción en el Censo de empresarios, profesionales y retenedores de la Agencia Estatal de Administración Tributaria o en el censo equivalente de la Administración Tributaria Foral, que debe reflejar la actividad económica efectivamente desarrollada a la fecha de solicitud de la ayuda.

Las partes deberán conservar la documentación vinculada a la actividad financiada por el MRR de conformidad con el artículo 132 del Reglamento Financiero, así como de disponer de un sistema que permita calcular y seguir el cumplimiento de objetivos y medición de indicadores.

g) Las partes quedan sometidas a los controles preceptivos a desarrollar por las autoridades de control que se establezcan, entre otras, el órgano gestor y responsable del seguimiento de las actuaciones financiadas, al control financiero de la Intervención General de la Administración del Estado y al control fiscalizador del Tribunal de Cuentas, así como a los controles señalados en el Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021. Las partes deberán también someterse a las actuaciones de control de las instituciones comunitarias, en virtud del artículo 22.2.e) del Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero, cuando sea preciso.

Los perceptores finales de los fondos están obligados a prestar colaboración y proporcionar cualquier documentación o información que se estime necesaria, así como a facilitar las inspecciones y comprobaciones que, en su caso, se efectúen. La oposición a la realización de estas comprobaciones e inspecciones podrá constituir causa de reintegro de la ayuda, sin perjuicio de la tramitación del correspondiente procedimiento sancionador. Cualquier irregularidad que afecte al control desarrollado por autoridades nacionales o de la Unión que impida el retorno de fondos al Tesoro, implicará el reintegro por las partes del importe del retorno no percibido en relación con las actuaciones responsabilidad de las partes.

h) Tomar las medidas necesarias para garantizar el cumplimiento de lo dispuesto en la Orden HFP/55/2023, de 24 de enero, relativa al análisis sistemático del riesgo de conflicto de interés en los procedimientos que ejecutan el Plan de Recuperación, Transformación y Resiliencia.

Dicho análisis se realizará en observancia de los requerimientos en el ámbito de la prevención, detección y corrección del conflicto de interés que la Comisión Europea exige a los Estados Miembros beneficiarios del MRR sobre la base del artículo 22 del Reglamento 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021, por el que se establece el Mecanismo de Recuperación y Resiliencia y de las normas financieras aplicables al presupuesto de la Unión Europea.

En este sentido, todas las entidades ejecutoras e instrumentales en ejecución de las obligaciones adquiridas en este convenio deberán realizar el análisis ex ante del riesgo de conflicto de interés sobre sus instrumentos jurídicos de ejecución (contratos, generalmente).

En concreto, y conforme a lo previsto en la disposición adicional centésima décima segunda de la Ley 31/2022, de 23 de diciembre, el análisis sistemático y automatizado del riesgo de conflicto de interés se llevará a cabo en cada procedimiento para las personas que realicen las siguientes funciones o asimilables, y aun cuando no se rija en su funcionamiento por la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público por la que se transponen al ordenamiento jurídico español las Directivas del Parlamento Europeo y del Consejo 2014/23/UE y 2014/24/UE, de 26 de febrero de 2014, o la Ley 38/2003, de 17 de noviembre, General de Subvenciones, que, a estos efectos, se denominarán decisores de la operación:

a) Contratos: órgano de contratación unipersonal y miembros del órgano de contratación colegiado, así como miembros del órgano colegiado de asistencia al órgano de contratación que participen en los procedimientos de contratación en las fases de valoración de ofertas, propuesta de adjudicación y adjudicación del contrato.

b) Subvenciones: órgano competente para la concesión y miembros de los órganos colegiados de valoración de solicitudes, en las fases de valoración de solicitudes y resolución de concesión.

El análisis del conflicto de interés se realizará sobre los empleados que participen en los procedimientos de adjudicación de contratos o de concesión de subvenciones, en el marco de cualquier actuación del PRTR, formen parte o no dichos empleados de una entidad ejecutora o de una entidad instrumental.

Las personas mencionadas en los apartados a) y b) anteriores serán las que deban firmar electrónicamente las Declaraciones de Ausencia de Conflicto de Interés (DACI) respecto de los participantes en los procedimientos de contratación o de concesión de subvenciones. Esta formulación se realizará una vez conocidos dichos participantes.

En el anexo I de la Orden HFP/55/2023, de 24 de enero, se recoge el modelo de DACI, con el contenido mínimo que ésta deberá contener.

A través de la herramienta informática referida en el artículo 4 de la citada Orden HFP/55/2023 se analizarán las posibles relaciones familiares o vinculaciones societarias, directas o indirectas, en las que pueda haber un interés personal o económico susceptible de provocar un conflicto de interés, entre las personas antes mencionadas y los participantes en cada procedimiento.

A estos efectos, se considera vinculación societaria susceptible de generar un conflicto de interés cuando ésta suponga titularidad real, o relación de vinculación de tipo administrador, directivo, o pertenencia al consejo de administración de la empresa en cuestión.

Así mismo, se prevé un sistema específico de obtención de información sobre titularidades reales para los casos en los que para el licitador o el solicitante de ayuda no existan datos de titularidad real en las bases de datos de la AEAT, entre otros supuestos, en el caso de determinadas entidades no residentes.

Además, el CSIC se compromete a:

De acuerdo con la resolución de la SEDIA, el CSIC, como entidad instrumental, bajo la responsabilidad y dirección del extinto Ministerio de Asuntos Económicos y Transformación Digital (ahora Ministerio para la Transformación Digital y de la Función Pública), se compromete a actuar como coordinador de la participación de los grupos de investigación de las instituciones españolas y europeas en el desarrollo del proyecto TeresIA cuyo alcance está dentro de este convenio. Además, como receptor de los fondos del Plan de Recuperación, Transformación y Resiliencia destinados por la SEDIA a financiar estos proyectos, una vez este convenio resulte eficaz y de acuerdo con lo recogido en el anexo I, se compromete a transferir a las otras partes firmantes de este convenio los fondos correspondientes a la financiación de las actividades a realizar por las mismas en un plazo máximo de cuarenta y cinco (45) días a partir de la inscripción del presente convenio en el Registro Estatal de Órganos e Instrumentos de Cooperación (REOICO).

Las cantidades a transferir por el CSIC al resto de partes firmantes de este convenio se imputarán a la partida presupuestaria 28.301.46PR.692 «Inversión nueva asociada al funcionamiento operativo de los servicios. Mecanismo de Recuperación y Resiliencia.» del presupuesto anual del CSIC.

Tercera. Compatibilidad del régimen de ayudas de Estado y prevención de la doble financiación.

La aportación dineraria mencionada en la cláusula sexta será compatible con otras subvenciones, ayudas, ingresos o recursos de una finalidad similar, siempre y cuando se garantice el cumplimiento con el artículo 9 del Reglamento (UE) 2021/241 del Parlamento Europeo y del Consejo de 12 de febrero de 2021.

En todo caso, las partes deberán comunicar al CSIC, y este a la SEDIA, la obtención de otra subvención, ayuda, ingreso o recurso que financie las actividades subvencionadas tan pronto esté en conocimiento de dicha obtención, y en todo caso, con anterioridad a la justificación de la aplicación dada a los fondos percibidos.

El receptor final de los fondos tiene la obligación de informar sobre cualesquiera otros fondos, no solo europeos, que hayan contribuido a la financiación de los mismos costes.

Cuarta. Uso de los fondos a cargo del Mecanismo de Recuperación y Resiliencia.

El uso de los fondos del MRR para la cofinanciación de actividades y proyectos estará sujeto a los siguientes principios: el marco temporal de las actuaciones financiadas por los fondos del MRR, de acuerdo con el artículo 17.2 y 18.4(i) del reglamento del MRR, y los principios de adicionalidad y complementariedad a los que se refiere el artículo 9 del citado reglamento y la sección 2 de la parte 3 de la guía de Estados Miembros publicada el 22 de enero de 2021(SWD(2021)12 part 1/2).

No podrán financiarse con cargo a este instrumento gastos ordinarios y recurrentes. Todos los gastos deberán responder de manera indubitada a la naturaleza de la actividad seleccionada para su financiación por el Mecanismo de Recuperación y Resiliencia que, en todo caso y circunstancias, deberán contribuir a los hitos y objetivos de la medida. Tampoco podrá financiarse el Impuesto sobre el Valor Añadido o impuesto equivalente.

En ningún caso, los fondos del MRR o cualquier otro fondo europeo podrán utilizarse para financiar dos veces el mismo coste.

Quinta. Verificación y justificación de los fondos.

Las actuaciones que se financien deben haber concluido materialmente antes del 31 de diciembre de 2025 y haberse verificado, como muy tarde, el 30 de junio de 2026. La verificación de la actividad realizada consistirá en la comprobación de que la financiación del MRR empleada se ha utilizado para los fines previstos y que se ha gestionado de conformidad con todas las normas aplicables y, en particular, si se han cumplido los hitos 251 y 253 recogidos en el anexo a la Decisión de Ejecución del Consejo (Council Implementing Decision o CID) vinculados al Componente 16, Reforma 1. Concretamente, los fondos recogidos en este convenio contribuyen al cumplimiento del hito 251 «Apoyo a proyectos en inteligencia artificial», que requiere haber comprometido, antes del cuarto trimestre de 2023, al menos quinientos (500.000.000,00 €) millones de euros de presupuesto en las medidas recogidas como parte de la Estrategia Nacional de Inteligencia Artificial (ENIA), y entre las que se encuentra el Plan Nacional de Tecnologías del Lenguaje, en el que se circunscribe esta actuación, y del hito 253 «Finalización de proyectos en inteligencia artificial», que establece el compromiso de finalización de los trabajos asociados a las medidas de la Estrategia Nacional de Inteligencia Artificial (ENIA) en el primer trimestre de 2026. Con objeto de justificar la ejecución presupuestaria, las partes deberán remitir al CSIC como máximo en los siguientes diez (10) días a partir del 31 de diciembre de 2025, un informe justificativo de la plena aplicación de fondos para la actuación regulada por este convenio. Por su parte,

el CSIC deberá remitir, el correspondiente informe justificativo a la SEDIA, como máximo en los siguientes veinte (20) días desde la fecha de finalización del plazo marcado en el apartado anterior. La justificación económica de gastos se realizará mediante cuenta justificativa, con aportación de justificantes de gasto, de acuerdo con el artículo 72 del Real Decreto 887/2006, de 21 de julio, por el que se aprueba el Reglamento de la Ley 38/2003, de 17 de noviembre, General de Subvenciones. Todas las partes deberán acreditar los costes reales incurridos, sin incluir el IVA u otros impuestos análogos.

Sexta. *Financiación.*

La financiación total, a cargo de fondos MRR, para las actividades recogidas en este convenio es de 1.450.000,00 euros.

Estos fondos proceden de:

a) La aportación realizada por el entonces Ministerio de Asuntos Económicos y Transformación Digital a la SEDIA con cargo a los Presupuestos Generales del Estado para el año 2023 para la realización del proyecto recogido en el Componente 16, Reforma 1: «Estrategia Nacional de Inteligencia Artificial».

b) La aportación realizada por la SEDIA al CSIC, con cargo a la aplicación presupuestaria 27.50.000X.730 de la SEDIA, por importe de un millón cuatrocientos cincuenta mil euros (1.450.000,00 €) en la anualidad 2023 y recogida en la resolución de la Secretaría de Estado de Digitalización e Inteligencia Artificial por la que se regula la aportación de la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), al Consejo Superior de Investigaciones Científicas (CSIC), para la realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) en el marco de la Estrategia Nacional de Inteligencia Artificial, el PERTE Nueva Economía de la Lengua y el Plan de Recuperación, Transformación y Resiliencia –financiado por la Unión Europea– Next Generation EU.

c) La aportación realizada por el CSIC al resto de instituciones participantes en este convenio, en cumplimiento de la resolución mencionada en el apartado anterior y siguiendo la distribución de fondos establecida en el anexo I. Esta aportación se realizará, con cargo a la aplicación presupuestaria 28.301.46PR.692 del CSIC, en un único pago y en un plazo máximo de cuarenta y cinco (45) días tras la inscripción del presente convenio en REOICO.

Séptima. *Comisión de Seguimiento.*

Para garantizar la correcta ejecución y el seguimiento de lo establecido en el presente convenio se constituirá una Comisión de Seguimiento que adecuará su funcionamiento a lo previsto en los artículos 15 y siguientes de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (en adelante LRJSP).

La Comisión de Seguimiento será mixta y paritaria estará integrada por un o una representante de cada una de las partes firmantes de este convenio.

Las partes designarán nominalmente a las personas titulares que les representen en la Comisión de Seguimiento y se lo comunicarán al CSIC en el plazo de UN MES tras la entrada en vigor del presente convenio y designarán igualmente a las personas que sustituyan a los titulares en caso de ausencia de estos.

a) Las funciones de la Comisión de Seguimiento serán las siguientes:

1. Realizar el seguimiento y la vigilancia de las actuaciones recogidas en este convenio.
2. Resolver los problemas de interpretación y cumplimiento que pudieran surgir en la aplicación e interpretación de las cláusulas y en el cumplimiento de este convenio.
3. Proponer modificaciones y la prórroga.
4. Establecer las definiciones, contenidos y entregables de las tareas recogidas en el anexo II, de acuerdo con el criterio conjunto de las instituciones involucradas en cada

una de ellas. En caso de discrepancia, se tendrá en cuenta el criterio de la institución coordinadora en cada caso.

b) La Comisión de Seguimiento se reunirá de forma ordinaria, al menos una vez al trimestre (ya sea de forma presencial o virtualmente) y, de forma extraordinaria, cuando lo solicite una de las partes.

c) Cada una de las partes dispondrá de un voto y los acuerdos serán adoptados por mayoría de votos.

d) En lo no previsto en el presente convenio, el régimen de organización y funcionamiento de la Comisión de Seguimiento será, de forma supletoria, el previsto para los órganos colegiados en la sección 3.^a del capítulo II del título preliminar de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

Octava. *Comunicación y publicidad de la financiación europea.*

a) En la documentación administrativa para la gestión de expedientes correspondientes a las actuaciones recogidas en este convenio se deberá incluir una mención a que el expediente está financiado por el Mecanismo de Recuperación y Resiliencia.

b) Las partes firmantes de este convenio deberán cumplir con las obligaciones de información y publicidad que las autoridades competentes establezcan relativas a la financiación del Mecanismo de Recuperación y Resiliencia.

c) En las medidas de información y comunicación del convenio así como en las actuaciones que desarrollen las inversiones recogidas en el mismo (carteles informativos, placas, publicaciones impresas y electrónicas, material audiovisual, páginas web, anuncios e inserciones en prensa, certificados, etc.), se deberá cumplir con lo establecido en la Orden HFP/1030/2021, de 29 de septiembre. En todo caso, se deberán incluir los siguientes logos:

- El emblema de la Unión;
- Junto con el emblema de la Unión, se incluirá el texto «Financiado por la Unión Europea – Next Generation EU»;
- Se usará también el logo del Plan de Recuperación, Transformación y Resiliencia disponible en el enlace: <https://planderecuperacion.gob.es/identidad-visual>.
- Se tendrán en cuenta las normas gráficas del emblema de la Unión y los colores normalizados establecidos en el anexo II del Reglamento de Ejecución 821/2014. También se puede consultar la siguiente página web: <http://publications.europa.eu/code/es/es-5000100.htm> y descargar distintos ejemplos del emblema en: https://europa.eu/european-union/about-eu/symbols/flag_es#download.

d) Los carteles informativos y placas deberán colocarse en un lugar bien visible y de acceso al público. Toda la publicidad deberá cumplir con la normativa vigente en materia de fondos europeos.

Asimismo, en las publicaciones, infraestructuras y otros resultados a los que puedan dar lugar las actividades financiadas, deberá también mencionarse que las mismas han sido financiadas por el Ministerio de Asuntos Económicos y Transformación Digital (hoy Ministerio para la Transformación Digital y de la Función Pública).

Novena. *Protección de datos.*

De conformidad con el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 (Reglamento General de Protección de Datos o RGPD), aplicable desde el 25 de mayo de 2018, las partes hacen constar de manera expresa que se abstendrán de cualquier tipo de tratamiento de los datos personales de que dispongan como consecuencia de este convenio, exceptuando aquel que sea estrictamente necesario para las finalidades del mismo. En este sentido se comprometen, especialmente, a no ceder a terceros los datos mencionados o los archivos que los contengan, así como a guardar estricta confidencialidad sobre los

mismos. Asimismo, las partes quedan sometidas a lo dispuesto en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, y demás normativa aplicable en materia de protección de datos.

El tratamiento de datos de los firmantes y personas de contacto de cada entidad se realizará en calidad de responsable por cada una de las partes, y de acuerdo a las políticas de privacidad accesible en la página web de cada entidad.

Sin perjuicio de lo anterior, en el caso que sea necesario para la finalidad y la ejecución del Proyecto transmitir o almacenar datos personales en el BSC-CNS –distintos a los relativos a las personas de contacto y de ejecución regulados en los párrafos anteriores–, éstos deberán venir anonimizados en origen por la entidad responsable de los datos. Si los datos no pudieran ser anonimizables, éstos no se transmitirán hasta que las partes correspondientes formalicen la herramienta jurídica aplicable al caso concreto. En todo caso, la entidad que vaya a ceder los datos ostentará la condición de responsable de tratamiento y será el encargado de redactar el acuerdo de tratamiento y el protocolo técnico de acceso, tratamiento, almacenamiento de los datos seudonimizados con el receptor de los datos. En el caso de datos personales anonimizados, el responsable de los datos deberá redactar el protocolo técnico de acceso, tratamiento, almacenamiento de datos anonimizados para garantizar que se han aplicado las medidas técnicas y organizativas necesarias para la anonimización irreversible de esos datos.

Décima. *Confidencialidad.*

Toda información intercambiada entre las partes en el marco de este convenio tendrá la consideración de reservada y confidencial, y no podrá ser divulgada a terceros sin el consentimiento previo por escrito de las partes afectadas ni utilizada para tareas que no estén específicamente acordadas.

La confidencialidad no será de aplicación cuando:

- La parte receptora pueda demostrar que conocía previamente la información recibida.
- La información recibida sea o pase a ser de dominio público.
- La parte receptora obtenga autorización previa y por escrito para su revelación, o la información sea requerida judicialmente o por orden de autoridad administrativa o gubernamental competente.
- Sea recibida de manera lícita de un tercero.
- Se haya generado de forma independiente y de buena fe por los miembros de su institución sin ninguna conexión con la información confidencial.

Las partes adoptarán cuantas medidas sean necesarias para que su personal cumpla con lo dispuesto en esta cláusula.

Las obligaciones en materia de confidencialidad subsistirán hasta los cinco años después de la pérdida de vigencia del presente convenio.

Undécima. *Publicación de resultados.*

En caso de que existan resultados publicables, las partes aplicarán las normas de publicación de resultados recogidas en el modelo DESCAs de Consortium Agreement de la Unión Europea para Horizonte Europa.

1. **Publicación conjunta.** Las partes podrán cooperar para publicar los resultados de la investigación en un artículo conjunto en una revista revisada por pares. En tal caso, la autoría de las publicaciones se basará en las normas recogidas en el modelo DESCAs. Las partes deberán reconocer adecuadamente todas las contribuciones científicas o de otro tipo en cualquier publicación o presentación relacionada con la Investigación realizada en virtud del presente convenio.

2. **Publicación por separado.** Durante el proyecto y por un período de 1 año tras la finalización del mismo, cada parte podrá hacer difusión de los propios resultados,

incluyendo, pero no limitándose a publicaciones y presentaciones, con sujeción a las siguientes disposiciones:

Cualquier publicación prevista se notificará previamente a las otras partes al menos 45 días naturales antes de la publicación. Cualquier objeción a la publicación deberá notificarse por escrito al coordinador y a la parte o partes que proponen la difusión en un plazo de 30 días naturales de la recepción de la notificación. Si no se presenta ninguna objeción en el plazo antes indicado, se autoriza la publicación.

Una objeción estará justificada si:

- a) afecta negativamente a la protección de los resultados o de los conocimientos previos de la parte que presenta la objeción, o
- b) perjudica significativamente los intereses legítimos de la parte que presenta la objeción en relación con sus resultados o conocimientos previos, o
- c) la publicación propuesta incluye Información confidencial de la Parte que presenta la objeción.

La objeción debe incluir una solicitud precisa de las modificaciones necesarias.

Si se presenta una objeción, las partes implicadas deberán discutir la forma adecuada de superar los motivos de la objeción (por ejemplo, modificando la publicación prevista y/o protegiendo la información). La Parte objetora no deberá continuar la oposición de forma injustificada si se han alcanzado medidas apropiadas tras la discusión.

La Parte objetora puede solicitar un aplazamiento de la publicación de no más de 90 días naturales a partir del momento en que plantee dicha objeción. Transcurridos los 90 días naturales, la publicación será permitida, siempre que se hayan tenido en cuenta las objeciones de la Parte objetora.

Una Parte no incluirá en ninguna actividad de difusión los Resultados o conocimientos previos de otra Parte sin obtener la previa aprobación por escrito de la Parte propietaria, a menos que ya estén ya publicados.

Duodécima. *Propiedad intelectual e industrial.*

Las partes respetarán los derechos de propiedad intelectual e industrial de terceros, e individualmente los de las otras partes. Cada parte seguirá siendo propietaria de los conocimientos previos aportados en el marco del presente convenio. Se entiende por «conocimientos previos», en particular: la información, el conocimiento, los métodos, las herramientas, el software y/o los derechos de propiedad industrial e intelectual aportados por cada una de las partes, anteriores a la firma del presente convenio. Los conocimientos previos tendrán la consideración de información confidencial.

Ninguna de las manifestaciones que se contienen en el presente convenio supone la cesión o transmisión de cualesquiera derechos de propiedad intelectual o industrial, titularidad de las partes firmantes.

Las partes se informarán mutuamente por escrito y sin demora de los resultados de sus proyectos de investigación concluidos en relación al desarrollo del proyecto TeresIA, y de la propiedad intelectual adquirida resultante, así como de cualquier mejora en todo o parte de los mismos. Las partes tratarán estos resultados como información confidencial, salvo acuerdo en contrario.

Los Resultados producidos únicamente por el personal de una de las partes será propiedad de dicha parte. Los resultados producidos por varias de las partes serán de propiedad conjunta de las partes que hayan participado en dicho resultado. Las condiciones de estos nuevos conocimientos adquiridos de propiedad conjunta serán acordadas por acuerdo escrito entre las partes.

En caso de Resultados Conjuntos, si cualquiera de las partes notifica por escrito su deseo de comercializar cualquiera de los resultados y alguna de las otras partes propietarias no puede/quiere hacerlo, las partes acuerdan negociar un acuerdo de

licencia adecuado y equilibrado donde se recoja el reparto de los beneficios de la comercialización de dichos resultados del proyecto o de la propiedad intelectual de los conocimientos adquiridos, o de cualquier parte de los mismos. En este proceso se deberán observar los trámites y autorizaciones que establece la normativa de carácter público para las entidades adscritas a la Administración General del Estado.

Decimotercera. *Vigencia, modificación y extinción del convenio.*

a) De conformidad con el artículo 48.8 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, el presente convenio se perfecciona con el consentimiento de las partes y resultará eficaz una vez inscrito en el Registro Estatal de Órganos e Instrumentos de Cooperación (REOICO) del sector público estatal, debiendo publicarse, a continuación, en el «Boletín Oficial del Estado».

b) Este convenio permanecerá en vigor desde la fecha de su firma y hasta el 31 de diciembre de 2026 y podrá ser prorrogado, de mutuo acuerdo entre las partes, por un período máximo de dos (2) años, mediante Adenda al mismo, adoptada antes de la finalización del periodo inicial de vigencia, todo ello con los límites que establece la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público. Entre las condiciones para su prórroga se encuentra la ampliación del plazo de ejecución de los fondos del Plan de Recuperación, Transformación y Resiliencia.

c) Las cláusulas de este convenio podrán ser modificadas por acuerdo unánime entre las partes.

d) De acuerdo con lo previsto en el artículo 51.1 de la LRJSP, este convenio se extinguirá por el cumplimiento de las actuaciones que constituyen su objeto o por incurrir en alguna de las siguientes causas de resolución:

1. El transcurso del plazo de vigencia del convenio sin haberse acordado la prórroga del mismo.
2. Por acuerdo unánime de las partes.
3. Por incumplimiento de las condiciones y compromisos asumidos por alguna de las partes.

En este caso, cualquiera de las partes podrá notificar a la parte incumplidora un requerimiento para que cumpla en un determinado plazo con las obligaciones o compromisos que se consideran incumplidos. Este requerimiento será comunicado a través la Comisión de Seguimiento.

Si transcurrido el plazo indicado en el requerimiento persistiera el incumplimiento, la parte que lo dirigió notificará a las partes firmantes la concurrencia de la causa de resolución y se entenderá resuelto el convenio.

- Por decisión judicial declaratoria de la nulidad del convenio.
- Si se produjesen circunstancias que hicieran imposible o innecesaria la colaboración.
- Por cualquier otra causa distinta de las anteriores prevista en el convenio o en otras leyes.

En caso de resolución del convenio, las partes quedan obligadas al cumplimiento de sus respectivos compromisos hasta la fecha en que ésta se produzca, y dará lugar a la liquidación del mismo con el objeto de determinar las obligaciones y compromisos de cada una de las partes en los términos establecidos en el artículo 52 de la Ley 40/2015.

Asimismo, de conformidad con el artículo 84 de la Ley 47/2003, la responsabilidad financiera derivados de la gestión de los fondos procedentes de la Unión Europea se someterá a lo previsto en la Ley 38/2003, de 17 de noviembre, y en la normativa comunitaria. Serán causas de reintegro total de la aportación dineraria:

a) El incumplimiento de los hitos y objetivos establecido para esta actuación, que afecten al cumplimiento de los hitos y objetivos establecidos para el Plan de

Recuperación, Transformación y Resiliencia. De este modo, la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) podrá exigir al Consejo Superior de Investigaciones Científicas (CSIC) el reintegro total por anticipado, según lo dispuesto en la Orden HFP/1030/2021, de 29 de septiembre, a efectos de garantizar el cumplimiento de los hitos y objetivos del Plan de Recuperación, Transformación y Resiliencia.

b) La falta de presentación de la documentación justificativa recogida en el apartado octavo y noveno.

c) La constatación, durante las actuaciones de comprobación, que los objetivos, productos, servicios o gastos justificados no se corresponden con lo establecido en esta resolución, o que hubieran sido desarrollados con anterioridad a la entrada en vigor de la presente norma, lo que evidenciaría la no aplicación de los fondos aportados a los fines que se persiguen con su aportación al Consejo Superior de Investigaciones Científicas (CSIC).

d) El incumplimiento de las obligaciones establecidas por la Orden HFP/1030/2021, de 29 de septiembre:

i. El incumplimiento del principio de «no causar un perjuicio significativo al medio ambiente» (Do No Significant Harm, -DNSH).

ii. El incumplimiento de las obligaciones de información y publicidad.

iii. El incumplimiento de las obligaciones en materia de etiquetado digital y verde.

iv. El incumplimiento de cualquiera de obligaciones establecidas en la citada Orden.

Será causa de reintegro parcial de la aportación dineraria la existencia de fondos transferidos que no se hayan destinado a financiar la actividad descrita en el anexo II o que no estén convenientemente justificados de acuerdo con lo establecido en esta resolución, y deberán ser reintegrados al Tesoro.

En caso de reintegro, la Parte afectada deberá reintegrar total o parcialmente los fondos recibidos por el CSIC, como institución coordinadora, para que ésta a su vez los devuelva al órgano competente.

Los eventuales reintegros a que hubiera lugar por no realizarse el gasto o, en su caso, por no poder cumplir total o parcialmente con los objetivos previstos, se tramitarán conforme a lo establecido en el artículo 7, séptima, de la Ley 31/2022, de 23 de diciembre, de Presupuestos Generales del Estado para el año 2023.

Decimocuarta. *Régimen jurídico y resolución de controversias.*

Este convenio tiene naturaleza interadministrativa de acuerdo con el artículo 47 de la Ley 40/2015, de 1 de octubre, y se regirá por sus propios términos y condiciones y por el ordenamiento jurídico administrativo, con las especialidades previstas en el artículo 59 del Real Decreto-ley 36/2020, de 30 de diciembre, por el que se aprueban medidas urgentes para la modernización de la Administración Pública y para la ejecución del Plan de Recuperación, Transformación y Resiliencia.

Las cuestiones litigiosas que puedan surgir en la interpretación y cumplimiento de este convenio que no cuenten con acuerdo en el seno de la Comisión de Seguimiento, serán de conocimiento y competencia del Orden Jurisdiccional de lo Contencioso-Administrativo.

Y en prueba de conformidad, las partes firman este convenio mediante firma electrónica, en la fecha indicada en la última de ellas.–La Presidenta de la Agencia Estatal Consejo Superior de Investigaciones Científicas, M.P., Eloísa del Pino Matute.–El Director del Instituto Cervantes, Luis Manuel García Montero.–El Presidente de la Asociación Española de Terminología, Joaquín García Palacios.–El Director de Barcelona Supercomputing Center-Centro Nacional de Supercomputación, Mateo Valero Cortés.–El Rector de la Universidad Politécnica de Madrid, Guillermo Cisneros Pérez.–El Presidente de la Real Academia de Ingeniería, Jaime Domínguez Abascal.

ANEXO I

Distribución de los fondos entre las instituciones participantes

Institución	Fondos en euros (según distribución acordada entre las partes)
CSIC (coordinador) a través del Instituto de Lengua, Literatura y Antropología del Centro de Ciencias Humanas y Sociales.	386076
Centro Nacional de Supercomputación (BSC).	235216
Ontology Engineering Group, Universidad Politécnica de Madrid (OEG-UPM).	213513
Asociación Española de Terminología (AETER).	113000
Instituto Cervantes (IC).	105000
Real Academia de Ingeniería.	397195
Total transferencia SEDIA.	1450000

En la cantidad asignada al CSIC se incluyen 128.576 euros previstos para la subcontratación del portal, según acuerdo de todas las partes.

La transferencia de estas cantidades será efectuada por el CSIC a cada una de las instituciones, en un único pago, que será realizado en un plazo máximo de cuarenta y cinco (45) días a partir de la inscripción del presente convenio en el Registro Estatal de Órganos e Instrumentos de Cooperación (REOICO).

ANEXO II

Detalle de las tareas dentro de la actuación «Realización del Portal de Acceso a Terminologías en España y Servicios de Inteligencia Artificial (TeresIA) en el marco de la Estrategia Nacional de Inteligencia Artificial, el PERTE de Nueva Economía de la Lengua y el Plan de Recuperación, Transformación y Resiliencia –financiado por la Unión Europea– Next Generation EU»

Conforme con la memoria técnica del proyecto, presentada por CSIC, el plan de actividades para el período 2024 -2025 es el siguiente:

Definición de tareas a desarrollar

Nota inicial: TeresIA y LexIA son dos proyectos independientes, aunque relacionados en algunos puntos. Hay procesos comunes que tratarán de realizarse conjuntamente, optimizando al máximo los recursos. El resultado del trabajo terminológico en Ingeniería proyecto LexIA) será incorporado al metabuscador TeresIA.

Paquete de trabajo 1. Coordinación de TeresIA

T1.1 Coordinación y gestión de la implementación del proyecto TeresIA. Instituciones involucradas: CSIC, IC, AETER, BSC-CNS, UPM y RAI. Institución coordinadora: CSIC.

Esta tarea incluye las tareas de coordinación y gestión del proyecto TeresIA, así como la organización de las reuniones del consorcio y las reuniones con el Ministerio. El equipo de gestión estará al cargo de revisar los avances del proyecto, el cumplimiento de los objetivos, y el envío de los entregables al Ministerio en tiempo y forma.

Esta tarea también incluye la revisión periódica de la calidad de los resultados y la creación de un plan de gestión de riesgos. En caso de que se produzca algún hecho que ponga en riesgo la realización de alguna tarea, se definirá un plan de contingencia y se informará al Ministerio para su valoración. Al mismo tiempo, el coordinador del proyecto será responsable de administrar el presupuesto asignado al proyecto y hacer los pagos pertinentes a las organizaciones participantes o subcontratadas.

T1.2 Gestión administrativa y financiera. Instituciones involucradas: CSIC, IC, AETER, BSC, UPM y RAI. Institución coordinadora: CSIC.

El coordinador será responsable de que se cumplan los objetivos, se envíen los informes técnicos al Ministerio en tiempo y forma, y se resuelva cualquier asunto administrativo o de carácter contractual. En cuanto a la gestión financiera, el coordinador será responsable de supervisar los costes incurridos durante el proyecto por las instituciones participantes y gestionar los pagos con el Ministerio. Cada una de las instituciones participantes en el proyecto se compromete a cumplir sus objetivos y plazos, como base de una adecuada coordinación general.

T1.3 Gestión de la protección de datos y de la propiedad intelectual. Instituciones involucradas: CSIC, IC, AETER, BSC-CNS, UPM. Institución coordinadora: CSIC.

El objetivo de esta tarea es la coordinación y supervisión de todos los aspectos relacionados con la gestión y explotación de la plataforma resultante, de todos los servicios asociados, y de los documentos y recursos involucrados. En esta tarea también se considerarán los aspectos relacionados con las licencias de los datos y recursos, y la propiedad intelectual de los mismos.

T1.4 Difusión del portal TeresIA. Instituciones involucradas: IC, CSIC, AETER, BSC-CNS, UPM. Institución coordinadora: IC.

Esta tarea se dedicará a la difusión del portal y del proyecto en general en foros académicos y profesionales, nacionales e internacionales (conferencias, talleres, etc.), para promocionar las actividades y resultados del mismo.

Se creará la identidad gráfica del portal, incluyendo el logo y un conjunto de materiales gráficos (diapositivas, pósters o infografías). En paralelo, se establecerán canales de comunicación con los potenciales usuarios de la plataforma, aprobando la implicación en el proyecto de instituciones relevantes en el ámbito terminológico y de promoción de las lenguas, como son AETER y el Instituto Cervantes.

Paquete de trabajo 2. Generación de repositorio de contenido e identificación de recursos existentes

T2.1 Creación de inventario de recursos terminológicos. Instituciones involucradas: IC, AETER, UPM, CSIC. Institución coordinadora: IC.

Creación de un inventario de recursos terminológicos (glosarios, tesauros, ontologías) para su uso en el despliegue del portal. Estos recursos se clasificarán de acuerdo con criterios como su licencia de uso o el formato en el que están representados. Se realizará también una evaluación de su conversión a los formatos estándares acordados para la publicación de datos en el portal.

T2.2 Modelo de representación (ontología) para la homogeneización del acceso y la recuperación de datos. Instituciones involucradas: UPM, BSC-CNS, CSIC. Institución coordinadora: UPM.

El sistema de almacenamiento de datos contará con numerosos recursos documentales y terminológicos de fuentes heterogéneas. Para poder acceder a dichos recursos mediante una única interfaz se propondrá una representación ontológica que

permita modelar principales características de los datos de forma homogénea, de forma que la información pueda ser consultada/consumida por distintos servicios y usuarios, preservando los datos en su forma original.

T2.3 Preparación del sistema de acceso a contenidos y corpus. Instituciones involucradas: CSIC, BSC, UPM, IC. Institución coordinadora: CSIC.

Junto con el resto de los socios, se analizarán los requerimientos del proyecto a nivel de datos para incorporarlos en un repositorio masivo y se caracterizarán para facilitar su posterior inventariado. Una vez definidos se implementarán herramientas de ingesta, preparación e inclusión de datos procedentes de los diferentes socios del proyecto. Los conjuntos de datos se almacenarán en la infraestructura ofrecida por el CSIC, aunque se tendrá en consideración en el desarrollo su posible migración a otro tipo de infraestructura común.

Paquete de trabajo 3. Desarrollo de herramientas y componentes avanzados de IA para extracción de términos y relaciones

T3.1 - Generación de conjuntos de datos de entrenamiento y evaluación de sistemas de Inteligencia Artificial. Instituciones involucradas: BSC-CNS, AETER, IC, CSIC, UPM. Institución coordinadora: BSC-CNS.

Para el desarrollo de sistemas automáticos de extracción de términos se requieren textos anotados con ejemplos de calidad que permitan a los algoritmos aprender de forma adecuada. Para ello se seleccionarán manualmente un mínimo de 500 párrafos de artículos de literatura científica y/o administrativa en castellano. Anotadores experimentados extraerán las palabras claves de estos párrafos siguiendo unas guías de anotación que se diseñarán y escribirán a partir de trabajos científicos previos como SemEval 2017 e Inspec.

T3.2 Implementación de sistemas de detección de términos relevantes en un documento científico. Instituciones involucradas: BSC-CNS, UPM. Institución coordinadora: BSC-CNS.

Los datos generados como parte de T3.1 se usarán para la adaptación, mejora y evaluación de arquitecturas de detección de términos relevantes en textos basadas en redes neuronales profundas. Estos sistemas estarán basados en el estado del arte descrito en la literatura, incluyendo herramientas que aprovechen los modelos de lenguaje, como hace, por ejemplo, la herramienta KeyBERT utilizando arquitectura de entrenamiento que utilicen la semántica contextual similares a AttentionRank y UCPhrase, aplicando estrategias para extracción de términos pertenecientes a dominios concretos. Estas técnicas más avanzadas se evaluarán respecto a las técnicas de aprendizaje automática tradicional para ver el impacto de los modelos de lenguaje en español en la extracción terminológica.

T3.3 Extracción de relaciones entre términos a partir de corpus. Instituciones involucradas: UPM, BSC-CNS. Institución coordinadora: UPM.

El objetivo de esta tarea es el diseño e implementación de algoritmos de extracción de relaciones entre términos inferidas a partir del propio corpus con el fin de enriquecer semánticamente el recurso. Las relaciones semánticas que se considerarán en un principio serán las propias de las taxonomías y tesauros, a saber, las jerárquicas (también conocidas como genérico-específicas) y las de relación entre términos. Se valorará la inclusión de relaciones específicas de dominio (ad-hoc) para los recursos generados en el marco de los casos de uso (PT7). En esta tarea también se reutilizarán trabajos previos tales como TermitUp. Las relaciones extraídas se compararán con las

que se hayan encontrado previamente en recursos existentes a los que se hayan enlazado los términos.

T.3.4 Evaluación de calidad, robustez, interoperabilidad y escalabilidad de las soluciones. Instituciones involucradas: BSC-CNS, UPM, CSIC, AETER, IC. Institución coordinadora: BSC-CNS.

Esta tarea estaría dedicada a la evaluación técnica de los algoritmos de extracción de terminologías y relaciones entre términos a partir de corpus anotados por experto. Para ellos se construirá un marco de evaluación de modelos de inteligencia artificial que permita obtener métricas comunes de evaluación como precisión, exhaustividad y métrica F1; pero también otras específicas del área de extracción de palabras clave como P@5K y P@50K.

Paquete de trabajo 4. Desarrollo de herramientas de conversión, desambiguación y enlazado de terminologías en TeresIA

T4.1 Herramienta de conversión de terminologías a formatos estándar. Instituciones involucradas: OEG, BSC, CSIC. Institución coordinadora: UPM OEG.

En esta tarea el objetivo es la creación de una herramienta de conversión de terminologías con dos objetivos: el primero, la armonización de las terminologías que se almacenarán y gestionarán en el portal para facilitar el acceso y recuperación de los datos en PT2; y el segundo, la integración o enlazado con otras terminologías similares preexistentes. Para ello se implementarán un conjunto de conversores entre los formatos más comunes de representación de glosarios terminológicos, tesauros, terminologías a los formatos más ampliamente adoptados por la comunidad de Web de datos lingüísticos.

T4.2 Herramienta de desambiguación de términos. Instituciones involucradas: BSC-CNS, UPM. Institución coordinadora: BSC-CNS.

El objetivo de esta tarea es la implementación de algoritmos de desambiguación de términos basados en modelos de lenguaje y/o su combinación con otros enfoques tradicionales. Estos algoritmos son necesarios para identificar el sentido con el que un término está siendo utilizado (o el concepto que éste representa) en casos de polisemia u homonimia. Se explorarán distintos escenarios según la disponibilidad de información semántica y contextual asociada a los términos en cuestión.

Por un lado, se analizarán los casos en los que se tenga información del contexto en el que aparezca el término utilizado (porque se disponga del corpus de extracción, por ejemplo), información de la semántica del término (porque se hayan establecido relaciones semánticas con otros términos), o cualquier otra información lingüística asociada (categoría gramatical, género, número, definición, etc.). Por otro lado, se tendrán en cuenta escenarios en los que la información contextual sea escasa o inexistente, o esté limitada a una lista de términos. Estos algoritmos van a permitir el posterior enlazado de los datos terminológicos con aquellos términos en terminologías existentes que representen el mismo concepto.

T4.3 Herramienta de enlazado o integración de terminologías. Instituciones involucradas: UPM, BSC-CNS. Institución coordinadora: UPM.

El propósito de esta tarea es la definición e implementación de los algoritmos de enlazado de las terminologías convertidas o creadas en TeresIA con términos definidos en terminologías o conjuntos de datos terminológicos ya existentes, según los principios de los datos enlazados (linked data), y su publicación para el consumo por otras aplicaciones en el contexto de los casos de uso. Esta tarea está intrínsecamente ligada a

la T4.2 e utilizará los resultados en su implementación. Se tomarán como punto de partida trabajos previos tales como TermitUp.

T4.4 Evaluación de los algoritmos de conversión, desambiguación y enlazado de términos. Instituciones involucradas: BSC-CNS, UPM, CSIC, AETER, IC. Institución coordinadora: BSC-CNS.

Esta tarea estará dedicada a la evaluación técnica de los algoritmos para la conversión, desambiguación y enlazado de términos. Para ello se tendrán en cuenta las métricas comunes de evaluación (precisión, recall o exhaustividad, métrica F1) y trabajos previos en el área como TermInteract.

Paquete de trabajo 5. Desarrollo del portal e interfaz de TeresIA

T5.1 Requisitos funcionales y no-funcionales. Instituciones involucradas: CSIC, IC, AETER. Institución coordinadora: CSIC.

Esta tarea tiene por objetivo extraer los requisitos funcionales y no funcionales de TeresIA. Los requisitos se extraerán de: (i) una colección de escenarios y casos de uso y (ii) el análisis de una selección de portales que se puedan considerar análogos. La arquitectura del software y todos los desarrollos posteriores se basarán en los requisitos especificados en el entregable correspondiente a esta tarea.

T5.2 Desarrollo del portal TeresIA. Instituciones involucradas: CSIC, BSC-CNS, UPM. Institución coordinadora: CSIC.

Esta tarea tiene por objetivo identificar y diseñar el conjunto de componentes de la arquitectura de TeresIA que den respuesta a los requisitos funcionales y no-funcionales. Dentro de esta tarea se definirán se implementarán los distintos componentes de la arquitectura, incluyendo aplicación web, almacenamiento de datos (PT2) y metadatos, y capa de servicios (PT3 y PT4). El resultado de esta tarea será una arquitectura software, una descripción funcional de los componentes, la descripción de una API HTTP REST pública, así como una recomendación de las tecnologías a emplear.

De acuerdo a las especificaciones y el diseño propuesto en la T4.1 y T4.2, se implementa la plataforma TeresIA, utilizando frameworks y tecnologías estándar en la medida de lo posible. Esta tarea aborda el desarrollo del motor de búsqueda propias y ajenas, la integración de los diferentes componentes de la plataforma TeresIA, en concreto, el acceso al repositorio de contenido, las herramientas desarrollados en los paquetes de trabajo 3 y 4, así como herramientas de acceso y mantenimiento.

T5.3 Acceso a datos externos a TeresIA. Instituciones involucradas: CSIC, BSC-CNS, UPM. Institución coordinadora: CSIC.

El objetivo de esta tarea es implementar las funcionalidades en backend de acceso a datos que residen en servidores externos a TeresIA. Se implementarán los clientes de API de terminologías multilingües de la EU, en concreto, IATE, como prueba de concepto. El resultado de esta tarea será el proyecto software con la implementación de los clientes accediendo a fuentes externas y el informe técnico con los detalles de implementación.

Paquete de trabajo 6. Desarrollo de herramientas de validación de terminologías en TeresIA por usuarios expertos

T6.1 Definición de los criterios de validación y sanción lingüística. Instituciones involucradas: AETER, CSIC, IC. Institución coordinadora: AETER.

El objetivo de esta tarea es la definición de los protocolos de validación y de sanción lingüística que implementará el módulo de validación y sanción (T6.2). Estos criterios

vendrán definidos por expertos en terminología y en neología terminológica del español. Una vez definidos, servirán de guía para la validación de los términos, en concreto para aquellos términos de nueva creación, desde un punto de vista lingüístico, con el fin de potenciar los usos adecuados de términos en español. Todo ello desde una perspectiva panhispánica que considere las distintas variantes geográficas de esta lengua y que afronte tanto la formación de términos nuevos utilizando las reglas del español, como la adaptación de términos procedentes de otras lenguas.

T6.2 Herramienta de validación y sanción por expertos. Instituciones involucradas: UPM, BSC-CNS, AETER, IC, CSIC. Institución coordinadora: UPM.

Esta tarea aborda el desarrollo e integración de una herramienta de validación y sanción en el portal TERESIA, teniendo en cuenta los criterios de validación y sanción lingüística previstos en la T6.1, y la gestión de usuarios definida en T1.2. La herramienta permitirá la validación de las terminologías resultantes de los procesos automáticos de extracción de terminologías y relaciones (PT3), y del enlazado entre terminologías (PT4), así como la sanción lingüística por parte de los expertos. Este módulo implementará un flujo de trabajo con varios niveles de interacción que permita la edición secuencial, en paralelo, de forma síncrona y asíncrona, así como funcionalidades de comunicación y debate por parte de los expertos, proporcionando un entorno de edición colaborativo amigable. Esta herramienta tomará como punto de partida experiencias previas llevadas a cabo por los participantes del proyecto, como es el caso de las herramientas TermInteract (UPM) y ASIT (BSC-CNS).

T6.3 Validación técnica y funcional de la solución. Instituciones involucradas: UPM, AETER, BSC-CNS, IC, CSIC. Institución coordinadora: UPM.

En esta tarea se llevará a cabo la evaluación técnica y no-técnica (usabilidad, aceptabilidad e impacto) de la herramienta de validación y sanción (T6.1 y T6.2). Para ello se involucrará a un conjunto de usuarios expertos en el ámbito de la terminología (pertenecientes a AETER y a la Dirección General de Traducción de la Comisión Europea) y servirá para evaluar la solución propuesta. Los resultados obtenidos en la prueba se utilizarán para identificar y solventar deficiencias en la implementación.

Paquete de trabajo 7. Escenarios de aplicación de generación de terminologías

T7.1 Escenario 1: Generación de terminologías en el dominio jurídico. Instituciones involucradas: IC, UPM, AETER, CSIC, BSC-CNS. Institución coordinadora: IC.

El objetivo es contribuir a la creación de recursos terminológicos en aquellas áreas de interés para la Justicia y la Administración pública españolas, así como el creciente ámbito de la mediación y arbitraje nacional e internacional, y su enlazado con recursos terminológicos a nivel nacional e internacional en los dominios de aplicación seleccionados.

T7.2 Escenario 2: Enriquecimiento de terminologías en el dominio biomédico. Instituciones involucradas: AETER, BSC-CNS, IC, CSIC, UPM. Institución coordinadora: AETER.

Las terminologías médicas tienen un papel esencial en el proceso de digitalización del sector de la sanidad. Este tipo de recurso lingüístico permite caracterizar y organizar los documentos biomédicos. Terminologías de alto impacto como SNOMED-CT han sido traducidas al español, sin embargo, adolecen de una falta de diversidad terminológica que limita la eficiencia de los sistemas de IA biomédico construidos sobre ésta. Por ejemplo, la versión internacional de SNOMED-CT en español dispone de un 18 % menos de entradas léxicas que la versión en inglés. Otros tesauros como CIE-10, que han sido diseñados para facilitar la comparación y análisis de datos sobre enfermedades, o de

naturaleza open-source como HPO, cuentan con una riqueza semántica mucho más limitada, con sólo uno o dos términos relacionados por concepto terminológico y con limitaciones para representar de forma adecuada algunas especialidades clínicas como la cardiología. El enriquecimiento de estas terminologías con unidades extraídas y enlazadas de forma semiautomática de conjuntos de datos mejorará la cobertura de éstas y la representatividad de los términos en español peninsular.

T7.3 Escenario 3: Caso de uso de terminologías en la gestión y recuperación de contenidos. Instituciones involucradas: CSIC, BSC-CNS, UPM, AETER, IC. Institución coordinadora: CSIC.

El trabajo con terminologías estructuradas permitirá la mejora de los sistemas de recuperación de información científica multilingüe de carácter internacional. Para que la producción científica en español y/o lenguas cooficiales pueda formar parte de los sistemas de recuperación multilingües es necesario contar con terminologías estructuradas, en el formato de la Web de Datos, que permitan llevar a cabo la indización sistemática de estos documentos o su descubribilidad. El volumen de producción científica en español permite pensar en corpus que serán una base importante para el posterior desarrollo de sistemas de indización automática que permitan implementar sistemas como el desarrollado por la National Library of Medicine en PubMed, capaces de asignar descriptores de terminologías controladas de forma automática a toda la documentación científica escrita en inglés. La futura integración de las terminologías y sistemas de indización permitirá un mayor impacto y reutilización de las publicaciones científicas en castellano.

Paquete de trabajo 8. (Conectado con el paquete de trabajo 1). Coordinación de LexIA, subproyecto de TeresIA

La coordinación del proyecto es necesaria para asegurar la colaboración y la comunicación efectiva entre los distintos participantes en orden a desarrollar los entregables establecidos en tiempo y forma, incluyendo la comunicación y organización de reuniones.

Será necesario el seguimiento y control del cumplimiento de los objetivos técnicos, asegurando la coherencia y consistencia de las contribuciones de cada parte y entre los diferentes paquetes de trabajo.

Asimismo, se seguirá la evolución del presupuesto del proyecto, monitorizando el avance de las actividades previstas y la evolución de los costes incurridos.

Se llevará a cabo la gestión de riesgos, estableciendo acciones de mitigación, así como la gestión asociada a la protección de datos y de la propiedad intelectual.

También, se desarrollarán actividades de difusión del proyecto y su implementación.

T8.1 Coordinación de paquetes de trabajo, socios, reuniones, etc. Instituciones involucradas: RAI y CSIC. Institución coordinadora: RAI.

T8.2 Seguimiento del proyecto y control presupuestario. Instituciones involucradas: RAI y CSIC. Institución coordinadora: RAI.

T8.3 Difusión del proyecto: avance, implementación y resultados. Instituciones involucradas: RAI y CSIC. Institución coordinadora: RAI.

Paquete de trabajo 9. Selección de corpus textuales para LexIA (Conectado con tareas 1.3 y 2.3.)

Establecimiento de los criterios para la selección de los documentos textuales a utilizar como referencia para conocer la evolución del uso de los términos y la detección de neologismos, que pueden tener origen en fuentes formales, redes sociales, etc.

Identificación de las fuentes del dominio de la ingeniería y selección de los documentos a considerar. Se recopilarán tesis doctorales, artículos, revistas especializadas en ingeniería.

Selección del corpus textual que alimentará el modelo de extracción. Los corpus textuales obtenidos, de fuentes académicas o especializadas, proporcionarán la diversidad terminológica y los conceptos técnicos, cuyo etiquetado y análisis constituirá el conjunto de datos enriquecido y actualizado, necesario para entrenar los modelos de aprendizaje automático.

Preparación de los corpus textuales para su análisis y etiquetado.

Análisis e implantación del marco contractual de cesión de los documentos.

Estudio de las posibilidades de automatización de la identificación de nuevos registros en los corpus seleccionados.

T9.1 Establecimiento de criterios para la selección de documentos textuales de interés (tesis doctorales, artículos, revistas, redes sociales, etc.). Instituciones involucradas: RAI y CSIC. Institución coordinadora: RAI.

T9.2 Identificación y selección de documentos de interés en el dominio de la ingeniería. Institución responsable: RAI.

T9.3 Selección de corpus textuales. Institución responsable: RAI.

T9.4 Preparación de los corpus textuales para su análisis. Institución responsable: RAI.

T9.5 Definición del marco legal de cesión de documentos y su incorporación automática al portal TeresiA. Institución responsable: RAI.

Paquete de trabajo 10. Anotación de corpus textuales seleccionados y extracción de terminología del corpus de Ingeniería (conectado parcialmente con los paquetes de trabajo 3 y 4)

Una vez recopilados los corpus textuales de tesis doctorales, artículos y revistas especializadas en diversas disciplinas de ingeniería, se procederá con el entrenamiento y ajuste de los algoritmos de detección y extracción de términos empleados en la Ingeniería en español.

Se realizará la anotación de un conjunto de párrafos extraídos de los corpus textuales seleccionados para identificar y etiquetar los términos relacionados con la Ingeniería presentes en los textos.

Se aplicarán técnicas de aprendizaje automático supervisado para que el modelo pueda aprender a reconocer términos relacionados con la Ingeniería en función de las características lingüísticas y contextuales presentes en los textos anotados.

T10.1 Anotación de corpus textuales seleccionados para esta etapa. Institución responsable: RAI.

T10.2 Apoyo en la adaptación del modelo a la identificación de terminología de Ingeniería. Institución responsable: RAI.

Paquete de trabajo 11. Validación de las definiciones de los neologismos e incorporación al DEI (conectado con el paquete de trabajo 6)

Una vez ajustado el modelo de extracción para el ámbito de la Ingeniería en español, se redactarán los borradores de las definiciones de los neologismos extraídos.

Los expertos de la RAI, tanto técnicos como lexicográficos, revisarán y enmendarán o validarán las citadas definiciones, alimentando el sistema con las definiciones validadas y contribuyendo al aprendizaje de la herramienta.

Se prestará atención a los requisitos lexicográficos recogidos en el documento «Planta del Diccionario Español de Ingeniería».

Este proceso continuará tras finalizar la fase de desarrollo y optimización de los modelos de extracción, una vez se inicie el proceso de alimentación recurrente de la herramienta a partir de los corpus seleccionados.

Una vez validadas las definiciones, se incorporarán al Diccionario Español de Ingeniería mediante una interfaz automatizada que actualice los contenidos del DEI.

T11.1 Revisión y validación de las definiciones (proceso continuo). Institución responsable: RAI.

T11.2 Elaboración de definiciones de los neologismos. Institución responsable: RAI.

T11.3 Diseño del proceso para la incorporación automática de las nuevas definiciones al DEI. Institución responsable: RAI.

T11.4 Diseño del trasvase de información del DEI al metabuscador TeresIA. Institución responsable: RAI.

Listado de entregables (TeresIA)

Entregable	Tipo
Informes del progreso de las actuaciones y cumplimiento de hitos.	Informe.
Desarrollo de la web e identidad gráfica del portal.	Web.
Inventario inicial de recursos terminológicos candidatos.	Informe.
Modelo de representación (ontología) de datos.	Informe - SW.
Despliegue del repositorio de datos.	SW.
Datos anotados.	Datos + Informe.
Algoritmos para la extracción de terminologías y su evaluación.	SW + Informe.
Algoritmos para la extracción de relaciones entre términos y evaluación.	SW + Informe.
Conversores para armonización de terminologías.	SW + Informe.
Herramienta de desambiguación de términos para enlazado.	SW + Informe.
Algoritmos para el enlazado de términos.	SW + Informe.
Requisitos funcionales y no-funcionales del portal.	Informe.
Desarrollo del portal web y casos de prueba.	SW + Informe.
Funcionalidades de acceso a datos externos.	SW + Informe.
Protocolos para la validación y sanción de términos por expertos.	Informe.
Herramienta de validación y sanción por expertos.	SW + Informe.
Evaluación de la herramienta de validación y sanción.	Informe.
Recursos terminológicos y aplicación al ámbito de la justicia.	Datos + Informe.
Recursos terminológicos y aplicación al ámbito de la biomedicina.	Datos + Informe.
Recursos terminológicos y aplicación al ámbito de la ingeniería.	Datos + Informe.

Listado de entregables (Lex_IA)

Entregable	Tipo
Informes semestrales de avance.	Informe.
Informe final.	Informe.
Informe de criterios.	Informe.

Entregable	Tipo
Informe de corpus textuales.	Informe.
Informe sobre el marco legal de cesión de documentos e incorporación automática a la herramienta.	Informe.
Informe de finalización del proceso de anotación del corpus seleccionado.	Informe.
Documento de requisitos de la solución de interfaz TERESIA-DEI.	Informe.
Herramienta interfaz entre «TERESIA» y el DEI.	SW.

Dentro de los entregables identificados la Secretaría de Estado de Digitalización e Inteligencia Artificial podrá definir hitos de entrega parcial de aquellos que por su complejidad y extensión se estime necesario de cara a verificar el cumplimiento de los hitos y objetivos ligados al Componente 16, Reforma 1.