



**“SERVICIO DE TRANSCRIPCIÓN E INDEXACIÓN
PROBABILÍSTICA AUTOMÁTICA DE CUATRO
SERIES DOCUMENTALES CONSERVADAS EN
LOS ARCHIVOS ESTATALES” (4 LOTES)**

Lote 1: 71.015,92 € (IVA excluido) 85.929,26 (IVA incluido)

Lote 2: 71.015,92 € (IVA excluido) 85.929,26 (IVA incluido)

Lote 3: 77.762,27 € (IVA excluido) 94.092,35 (IVA incluido)

Lote 4: 64.287,99 € (IVA excluido) 77.788,47 (IVA incluido)

PRESUPUESTO BASE: 284.082,10.- € (IVA EXCLUIDO)

343.739,34.- € (IVA INCLUIDO)

*Financiado con fondos de los Planes Next Generation
UE: Plan de Recuperación, Transformación y Resiliencia
(PRTR) (Instrucción quinta del artículo 7 Ley 11/2020, de
Presupuestos Generales del Estado para el año 2021*



MINISTERIO
DE CULTURA
Y DEPORTE



Financiado por la
Unión Europea
NextGenerationEU

DIRECCIÓN GENERAL
DE BELLAS ARTES

SUBDIRECCIÓN GENERAL
DE LOS ARCHIVOS ESTATALES

PLIEGO DE PRESCRIPCIONES TÉCNICAS A REGIR EN LA CONTRATACIÓN MEDIANTE PROCEDIMIENTO ABIERTO DE UN SERVICIO DE TRANSCRIPCIÓN E INDEXACIÓN PROBABILÍSTICA AUTOMÁTICA DE CUATRO SERIES DOCUMENTALES CONSERVADAS EN LOS ARCHIVOS ESTATALES.

Financiado con fondos de los Planes Next Generation UE: Plan de Recuperación, Transformación y Resiliencia (PRTR) (Instrucción quinta del artículo 7 Ley 11/2020, de Presupuestos Generales del Estado para el año 2021)

FIRMADO

JUDIT DIEZ MANZANEDO - 2021-12-30 14:05:32 CET, Presidenta de la Junta de Contratación
SEVERIANO HERNANDEZ VICENTE - 2021-12-29 12:58:37 CET

La autenticidad del documento puede ser comprobada mediante el CSV: OIP_HP32KQFP7A3JUD6XW9D6NNE5BSFO en <https://www.pap.hacienda.gob.es>

Contenido

1. Organismo Interesado	3
2. Antecedentes.....	3
3. Objeto del contrato	5
4. Especificaciones técnicas comunes aplicables a los cuatro lotes	6
5. Objeto del contrato del lote 1	9
6. Objeto del contrato del lote 2	10
7. Objeto del contrato del lote 3	12
8. Objeto del contrato del lote 4	13
9. Propiedad intelectual	15
10. Control de calidad.....	15
11. Entregables	15
12. Confidencialidad y Protección de Datos de carácter personal	16
13. Cláusula de daño no significativo.....	16

1. Organismo Interesado

El Organismo interesado en la presente contratación es la Subdirección General de los Archivos Estatales (en adelante SGAE) del Ministerio de Cultura y Deporte.

2. Antecedentes

El Real Decreto 509/2020, de 5 de mayo, por el que se desarrolla la estructura orgánica básica del Ministerio de Cultura y Deporte y se modifica el Real Decreto 817/2018 de 6 de julio, establece que corresponde a la Secretaría General de Cultura y Deporte la promoción, la protección y difusión de los archivos estatales a través de Dirección General de Bellas Artes, dependiente de la misma, centro directivo al que pertenece la SGAE, a quien le corresponde la gestión de los archivos de titularidad y gestión estatal adscritos al Departamento ministerial, así como la función del fomento de la conservación del patrimonio documental y su promoción y difusión nacional e internacional.

Desde hace décadas, ejerciendo las responsabilidades que la normativa encomienda a la SGAE y siguiendo las recomendaciones, normas y directrices nacionales e internacionales, los Archivos Estatales han hecho un gran esfuerzo por describir el patrimonio documental conservado en sus centros para ofrecerlo a la ciudadanía a través del Portal de Archivos Españoles (en adelante PARES), señalado en el Real Decreto 1708/2011, de 18 de noviembre, por el que se establece el Sistema Español de Archivos y se regula el Sistema de Archivos de la Administración General del Estado y de sus Organismos Públicos y su régimen de acceso; como la herramienta de difusión de los documentos de los centros de archivo del Ministerio de Cultura y Deporte, con el fin de fomentar el acceso libre y gratuito de los usuarios a los contenidos culturales e informativos que se custodian.

En la actualidad, los Archivos Estatales gestionados por el Ministerio de Cultura y Deporte custodian 259.825 metros lineales (casi 260 kilómetros) de documentación muy diversa, no solo documentación textual en papel, sino también a modo de ejemplo, más de 300.000 pergaminos y alrededor de 1.700.000 documentos fotográficos.

Los trabajos de ordenación, identificación, descripción y digitalización de tan ingente volumen documental, así como el posterior volcado de los resultados de todo ello en PARES, son procesos lentos que requieren de tiempo y recursos humanos especializados. Únicamente la labor de descripción documental implica un conocimiento profundo de las instituciones históricas productoras de los documentos, además del conocimiento de la documentación, de disciplinas como la paleografía y diplomática para poder leer los documentos en otros tipos de escritura diferentes a la actual, es decir, requiere del trabajo especializado de los archiveros de los departamentos de descripción, que tras la lectura y comprensión de los documentos, extraen su información para trasladarla a estándares estructurados que son los que

finalmente se publican como resultado final en PARES. Este proceso, por tanto, es lento y requiere de efectivos humanos numerosos, teniendo en cuenta el gran volumen de documentación mencionado.

Por todo ello, solo una pequeña parte del Patrimonio Documental conservado en nuestros archivos está disponible para su acceso por parte de los ciudadanos en PARES que, a día de hoy, difunde:

- 5.373.761 de descripciones de documentos
- 35.578.068 de imágenes vinculadas a las descripciones anteriores.

Además, hay que tener en cuenta que no toda la documentación descrita en PARES tiene imágenes asociadas, tan solo 2.026.785 descripciones de documentos están digitalizadas, lo que supone aproximadamente un 36% del total de registros descriptivos difundido en el Portal.

Por otro lado, la demanda social para acceder a los documentos de archivo, bien sea por interés científico de investigación, ocio cultural, búsqueda de antecedentes familiares o de historia local o para ejercer derechos civiles, ha crecido exponencialmente en las últimas décadas, siendo insuficiente la cantidad de descripciones que ofrecemos en PARES comparadas con la demanda en aumento, no sólo de la ciudadanía del estado español. Debemos tener en cuenta la fuerte demanda de otros países fruto de la importancia histórica de la corona española. Los documentos de los Archivos Estatales tienen una gran demanda internacional, ya que se conservan en estos centros los documentos que permiten reconstruir gran parte de la historia de América, Europa, África, Asia y Oceanía.

Para hacer frente a esta insuficiencia de efectivos, han resultado parcialmente efectivas las aplicaciones como el reconocimiento óptico de caracteres (en inglés, *“optical character recognition”*, OCR). Esta tecnología ha sido de utilidad para la transcripción e indexación de documentos de texto impreso o mecanografiado. Sin embargo, estas soluciones tecnológicas no pueden aplicarse al ámbito de textos manuscritos, por la variabilidad de las letras, tintas, formatos, tipologías, entre otros. Sólo recientemente se están encontrando nuevas soluciones tecnológicas con resultados aceptables para la transcripción automatizada de texto manuscrito e indexación probabilística en el ámbito de los archivos, denominadas por sus siglas en inglés, HTR (*“handwritten text recognition”*) que permite la transcripción automática del texto contenido en las imágenes y su indexación Probabilística o “PrIx” (del inglés *Probabilistic Indexing*), que se refieren al servicio que se quiere contratar.

Recientemente se ha mostrado la utilidad de estas tecnologías emergentes para los Archivos Estatales, mediante la conclusión en noviembre de 2021 de una experiencia piloto desarrollada en el marco del proyecto europeo aprobado por el Programa Europa Creativa *“European Digital Treasures”*, que cuenta con la participación de los Archivos Estatales/Nacionales de varios países (España, Hungría, Noruega, Portugal y Malta), el *International Centre for Archival Research (ICARUS, Austria)* y la *Munster*

Technological University (MTU, Irlanda). Con este proyecto piloto se probó la eficacia del sistema de reconocimiento de texto manuscrito (HTR) en una serie documental de cada uno de los Archivos Nacionales socios de proyecto y, en el caso de los Archivos Estatales de España, se obtuvo satisfactoriamente la indexación probabilística automática de la serie documental conservada en el Archivo General de la Administración *Registro de pasaportes del Consulado de España en Buenos Aires (1933-1938)*.

Otra iniciativa similar con resultados satisfactorios que se está desarrollando mediante un proyecto de la convocatoria de ayudas a “*Proyectos de I+D+i*», en el marco del *Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema de I+D+i y del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020*” es el proyecto “*Searching in the Simancas Archive*” (SimancasSearch), proyecto que está en estos momentos trabajando en la indexación automática de la serie “*Libros registro de cédulas*”, conservadas en el Archivo General de Simancas.

Ambos constituyen los antecedentes por los que la SGAE quiere continuar con esta línea de trabajo mediante la contratación del servicio que se explica en el siguiente punto.

3. Objeto del contrato

El presente pliego de prescripciones técnicas tiene por objeto la contratación de cuatro servicios de transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial de cuatro series documentales conservadas en los Archivos Estatales.

Se ha dividido el objeto del contrato en cuatro lotes, teniendo en cuenta la especificidad de las cuatro series documentales y la custodia en diferentes centros dependientes de la SGAE.

LOTE 1: Serie documental conservada en el Archivo General de la Administración *Registro-Índice de la Población Reclusa (1938-1955)* con un volumen aproximado de 380.000 fichas digitalizadas.

LOTE 2: Serie documental conservada en el Archivo General de la Administración *Propuestas de Libertad Condicional (1940-1965)* con un volumen aproximado de 400.000 fichas digitalizadas.

LOTE 3: Serie documental conservada en el Archivo de la Corona de Aragón *Registros de Cancillería de los reyes Fernando II de Aragón (el Católico) y Carlos I de Aragón (1479 y 1556)* con un volumen aproximado de 244.000 imágenes digitalizadas de documentos.

LOTE 4: Serie documental conservada en el Archivo General de Simancas *Registro General del Sello de Corte (1502 y 1518)* con un volumen aproximado de 312.000 imágenes digitalizadas de documentos.

Las empresas licitadoras podrán presentar ofertas para uno de estos lotes o para los cuatro.

4. Especificaciones técnicas comunes aplicables a los cuatro lotes

La especificidad de las cuatro series documentales, las necesidades propias del proceso de trabajo individualizado a realizar mediante modelos informáticos específicos para cada serie y los diferentes archivos donde se conservan han permitido dividir el objeto del contrato en cuatro lotes. Sin embargo, existen una serie de especificaciones técnicas de carácter procedimental para la aplicación de técnicas de Reconocimiento de textos manuscritos (HTR) que son comunes a las cuatro series, teniendo en cuenta que los procesos técnicos se encaminan a obtener en todos los casos una propuesta de transcripción automática del texto contenido en las imágenes y la Indexación Probabilística (PrIx).

Para la correcta prestación del servicio será preciso realizar en cada uno de los lotes los siguientes procesos y tareas generales:

1. Proceso de entrega de los objetos digitales

Las series documentales que forman parte de los cuatro lotes se encuentran previamente digitalizadas y se encuentran almacenadas en las instalaciones de los Archivos Estatales y en los servidores del Ministerio de Cultura y Deporte. La SGAE coordinará con la empresa adjudicataria realizar el traspaso de las copias digitales en el plazo máximo de 15 días desde la firma del contrato, según el volumen y detalle de las especificaciones técnicas de los lotes. La empresa adjudicataria guardará los objetos digitales en servidores propios con todas las medidas de seguridad que garanticen su integridad e impidan el acceso a estas imágenes a personas no autorizadas. La empresa adjudicataria se compromete a garantizar la confidencialidad de los datos y a no hacer ningún uso de los mismos fuera de lo establecido en este contrato.

2. Proceso de análisis documental de la serie

Para el entrenamiento y enseñanza del sistema de cara al reconocimiento automático de los textos manuscritos de las cuatro series, se precisará entrenar para enseñar al sistema a reconocer el texto. Se analizarán por parte de la empresa adjudicataria las siguientes variables, que serán entrenadas por separado para establecer un modelo específico para cada lote. El análisis de cada una de las siguientes variables constituirá una tarea derivada del proceso de análisis documental:

a. Análisis de la maquetación (layout) o forma en la que el texto se reparte formalmente en la página.

- b. Análisis de los idiomas presentes en la serie.
- c. Análisis de la variación de las letras que se encuentran en la serie, fundamentalmente gótica cursiva castellana cortesana aragonesa y humanística contemporánea. En este caso, no sólo se tendrán en cuenta las diferencias de letra en términos paleográficos, es decir, de evolución de las letras a lo largo del tiempo, sino también las diferentes letras en función de los amanuenses, escribanos o autores de los textos manuscritos.
- d. Análisis del estado de conservación de los documentos, teniendo en cuenta la frecuencia de la presencia de manchas, roturas, faltas de soporte, mala calidad de las tintas o transferencia de estas de una cara a la otra de los documentos, pues el modelo debe ser entrenado según estas complejidades añadidas para el proceso de reconocimiento.
- e. Análisis de la calidad de reproducción de la imagen.

Una vez concluido el proceso del análisis documental para cada uno de los lotes, la empresa adjudicataria emitirá un informe que refleje el análisis de las variables y tareas, en el que debe aparecer una propuesta detallada de las actuaciones con el número inicial de objetos digitales para el entrenamiento del sistema relacionados con el proceso 3 (véase siguiente punto) y la previsión de reentrenamientos para alcanzar el mayor grado de precisión en la transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial. La entrega de este informe se hará en un plazo máximo de un mes y medio desde la recepción de los objetos digitales del proceso 1.

3. Proceso de entrenamiento o preparación de *Ground Truth* (en adelante GT)

Para permitir la posterior indexación automática y búsqueda por contenido textual directamente sobre los objetos digitales, la empresa adjudicataria implementará un sistema basado en modelos de *Machine Learning / Deep Learning* para poder reconocer el texto manuscrito y ofrecer capacidades de búsqueda sobre este. El entrenamiento del sistema, se desglosará en las siguientes tareas:

- a. Selección de imágenes representativas de cada serie. Como resultado del proceso anterior de análisis documental, la empresa adjudicataria en coordinación con la persona designada como coordinador de cada uno de los lotes por parte de la SGAE, elegirá los objetos digitales representativos de las variables de entrenamiento señaladas anteriormente. La muestra debe ser suficientemente representativa de la serie, ya que sobre esta muestra se realizará el primer entrenamiento del sistema. El número de imágenes será

propuesta en el informe de finalización del proceso 2 y deberá contar con el visto bueno del coordinador del lote por parte de Archivos Estatales.

- b. Creación de una colección de entrenamiento (GT). La empresa adjudicataria pondrá a disposición del servicio a un especialista en gestión documental o paleografía, que se detallará en el objeto del contrato de los lotes, que será el encargado de entrenar al sistema mediante la producción de las anotaciones de referencia que van desde la estructura (layout) de la página hasta la transcripción de cada palabra, incluyendo abreviaturas. La tarea de anotación consistirá en el marcaje de los bloques y líneas de texto de cada imagen seleccionada y la transcripción de todas las líneas marcadas.
- c. Entrenamiento de modelos y evaluación de calidad de resultados. La empresa adjudicataria realizará al menos durante el periodo especificado para cada uno de los lotes, al menos tres entregas de resultados mostradas en un demostrador creado al efecto, según el entrenamiento de los modelos. Tras cada entrega, se emitirá un informe a la SGAE en el que se precise la cantidad de imágenes que deben ser entrenadas de nuevo para obtener una optimización de las prestaciones en el rendimiento de la *PrIx* y el porcentaje de mejora esperado.

4. Proceso de indexación probabilística de toda la serie

La empresa adjudicataria generará un fichero *PrIx* por cada imagen de la serie. Igualmente, para cada línea de texto automáticamente detectada, se proporcionará la mejor transcripción automática derivada del contenido del *PrIx* asociado a dicha línea según los requisitos de la tarea 3c. La generación de los ficheros de toda la serie y la entrega de estos ficheros a la SGAE será acompañada por parte del adjudicatario de un informe de resultados, al que la SGAE dará su visto bueno por escrito. Este hito marcará el segundo pago del contrato.

La duración total de los procesos 3 y 4 será especificada en el objeto del contrato de cada uno de los lotes.

5. Proceso de integración en PARES. La empresa adjudicataria deberá proporcionar un software que combine la transferencia de conocimiento, la utilización de la tecnología desarrollada, el mantenimiento y reposición de la base, el motor de búsqueda y la atención al cliente durante el proceso de integración. Asimismo, deberá integrar la búsqueda de los índices probabilísticos automáticos en el entorno de PARES. Las soluciones de la empresa adjudicataria deberán integrarse en el entorno tecnológico en el que se encuentran desarrollados los diferentes sistemas sobre los que corre PARES, cuyas características tecnológicas son las siguientes:

- Aplicaciones Web
 - Arquitectura de aplicaciones Web: MVC
 - Tecnología: Java J2EE
 - Frameworks: Struts e hibernate
 - Sistemas Operativos: Linux sobre plataformas virtualizadas en VMware.
 - Servidor Web: Apache
 - Servidores de aplicaciones: Jboss
 - Herramientas de desarrollo: Eclipse, (con Subversión)
- Base de Datos Relacional Oracle, con extensiones de Oracle Text.

Los requisitos para los nuevos elementos a integrar en PARES deben ajustarse a este entorno tecnológico. La duración estimada de este proceso es de un mes.

5. Objeto del contrato del lote 1

El lote 1 tiene por objeto la contratación de un servicio de transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial de 380.000 objetos digitales procedentes de la reproducción digital del mismo número de fichas que integran la serie documental denominada **Registro-Índice de la Población Reclusa (1938-1955)**. Esta serie fue generada por la Sección de Tratamiento Penitenciario del Servicio de Prisiones del gobierno de Burgos, posteriormente transformado éste en la Dirección General de Prisiones, documentos fechados entre los años 1938 y 1955, que se conservan desde 1983 en el Archivo General de la Administración. El volumen de la colección a indexar es de 160 cajones o gavetas con 380.000 fichas, aproximadamente, con el historial carcelario registrado de cada preso. La media de campos a indexar y transcribir es de 26 campos por documento o ficha.

Dada la singularidad de los procesos número 3 y 4, se estima que por la homogeneidad de la información contenida en las fichas se realice en un plazo máximo de cinco meses.

PERFILES PROFESIONALES

Para la ejecución del servicio, se considera imprescindible que la empresa adjudicataria disponga de recursos humanos adecuados tanto en número como en cualificación técnica. Mediante la firma del contrato, el adjudicatario se comprometerá a disponer de los medios personales suficientes para la realización de los servicios objeto del presente contrato.

Teniendo en cuenta la experiencia de los proyectos piloto explicados en el punto de antecedentes, se consideran adecuados los perfiles que se especifican a continuación, dejando no obstante a la empresa la determinación del personal que ponga a disposición de este contrato. Se estima que serán necesarios los siguientes perfiles para ejecutar este servicio:

- 1 Jefe de proyecto a tiempo completo. Tareas: planificación, logística y coordinación por parte de la empresa, tramitación, generación de informes y documentación, dirección técnica y supervisión de los trabajos.
- 1 Informático a tiempo completo con experiencia en tecnologías de indexación automática de texto manuscrito. Tareas: programación informática.
- 1 especialista en archivística y a media jornada con experiencia en lectura de letra humanística cursiva del siglo XX. Tareas: entrenamiento del sistema verificación de los resultados de los trabajos de indexación y para la verificación y corrección de las transcripciones utilizadas en la indexación automática de los documentos integrantes de la serie mediante muestreo sistemático.

En todo caso, la empresa adjudicataria establecerá el número de personal que considere adecuado para la realización de este servicio con los perfiles expuestos o similares. El Jefe de Proyecto será el interlocutor con el personal a cargo del proyecto de la SGAE para todas las tareas de planificación, dirección y seguimiento de las actuaciones contempladas en el presente pliego.

Corresponde al adjudicatario la ejecución, la dirección y la coordinación directa de los medios personales que realicen las actuaciones objeto del contrato. Las ofertas de las empresas licitadoras deberán incluir indicación del número de personas que asignarían al servicio junto con información detallada de sus perfiles profesionales y experiencia.

MEDIOS MATERIALES

Los trabajos objeto de este contrato no precisan de realizarse en las instalaciones del Archivo General de la Administración, pudiéndose realizar en remoto o en las instalaciones de la empresa adjudicataria.

La empresa adjudicataria pondrá a disposición del equipo de trabajo todos los medios materiales necesarios para llevar a cabo las tareas objeto del contrato, equipamiento informático, tecnológico y de procesamiento de imágenes, material de oficina y los elementos ofimáticos y de comunicaciones que sean precisos.

6. Objeto del contrato del lote 2

El lote 2 tiene por objeto la contratación de un servicio de transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial de 400.000 objetos digitales procedentes de la reproducción digital del mismo número de fichas que integran la serie documental denominada ***Propuestas de Libertad Condicional***

(1940-1965). Esta serie fue generada por la Sección de Libertad Condicional y Redención de Penas por el trabajo de la Dirección General de Prisiones, documentos fechados entre los años 1940 y 1965, que se conservan desde 1983 en el Archivo General de la Administración. El volumen de la colección a indexar y transcribir asciende a unas 400.000 fichas, aproximadamente, en la que se consignan, entre otros datos la prisión en la que se encuentra recluido cada preso, la condena que sufre, la fecha en la que comenzó a cumplir la condena, la de ratificación de la misma y la fecha en la que comenzó a disfrutar la libertad condicional. La media de campos a indexar y transcribir es de 18 campos por documento o ficha.

Dada la singularidad del proceso número 3 y 4, se estima que por la homogeneidad de la información contenida en las fichas se realice en un plazo máximo de cinco meses.

PERFILES PROFESIONALES

Para la ejecución del servicio, se considera imprescindible que la empresa adjudicataria disponga de recursos humanos adecuados tanto en número como en cualificación técnica.

Teniendo en cuenta la experiencia de los proyectos piloto explicados en el punto de antecedentes, se consideran adecuados los perfiles que se especifican a continuación, dejando no obstante a la empresa la determinación del personal que ponga a disposición de este contrato. Se estima que serán necesarios los siguientes perfiles para ejecutar este servicio:

- 1 Jefe de proyecto a tiempo completo. Tareas: planificación, logística y coordinación por parte de la empresa, tramitación, generación de informes y documentación, dirección técnica y supervisión de los trabajos.
- 1 Informático a tiempo completo con experiencia en tecnologías de indexación automática de texto manuscrito. Tareas: programación informática.
- 1 especialista en archivística y gestión documental a media jornada con experiencia en lectura de letra humanística cursiva del siglo XX. Tareas: entrenamiento del sistema verificación de los resultados de los trabajos de indexación y para la verificación y corrección de las transcripciones utilizadas en la indexación automática de los documentos integrantes de la serie mediante muestreo sistemático.

En todo caso, la empresa adjudicataria establecerá el número de personal que considere adecuado para la realización de este servicio con los perfiles expuestos o similares. El Jefe de Proyecto será el interlocutor con el personal a cargo del proyecto de la SGAE para todas las tareas de planificación, dirección y seguimiento de las actuaciones contempladas en el presente pliego.

Corresponde al adjudicatario la ejecución, la dirección y la coordinación directa de los medios personales que realicen las actuaciones objeto del contrato. Las ofertas de las empresas licitadoras deberán incluir indicación del número de personas que asignarían al servicio junto con información detallada de sus perfiles profesionales.

MEDIOS MATERIALES

Los trabajos objeto de este contrato no precisan de realizarse en las instalaciones del Archivo General de la Administración, pudiéndose realizar en remoto o en las instalaciones de la empresa adjudicataria.

La empresa adjudicataria pondrá a disposición del equipo de trabajo todos los medios materiales necesarios para llevar a cabo las tareas objeto del contrato, equipamiento informático, tecnológico y de procesamiento de imágenes, material de oficina y los elementos ofimáticos y de comunicaciones que sean precisos.

7. Objeto del contrato del lote 3

El lote 3 tiene por objeto la contratación de un servicio de transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial de 244.000 objetos digitales procedentes de la reproducción digital de la serie documental denominada **Registros de Cancillería de los reyes de Aragón. Reinados de Fernando II (el Católico) y Carlos I (1479-1556)**. Esta serie fue generada por la Cancillería Real de los reyes de Aragón entre los años 1479 y 1556, cuyos documentos originales se conservan en el Archivo de la Corona de Aragón. El volumen de la colección a indexar y transcribir es de 556 registros, con un total aproximado de 244.000 imágenes digitalizadas. La estructura formal de los registros es similar, habiéndose establecido una clasificación artificial de los correspondientes a cada uno de los reinados para facilitar su identificación. Sin embargo, los asientos de los Registros de la Cancillería Real de la Corona de Aragón de estas dos fracciones de serie tienen la complejidad de estar redactados en cinco lenguas diferentes: catalán, castellano, altoaragonés, latín e italiano.

Dada la singularidad del proceso número 3 y 4, se estima que por variabilidad de lenguas en esta serie documental se realice en un plazo máximo de seis meses.

PERFILES PROFESIONALES

Para la ejecución del servicio, se considera imprescindible que la empresa adjudicataria disponga de recursos humanos adecuados tanto en número como en cualificación técnica.

Teniendo en cuenta la experiencia de los proyectos piloto explicados en el punto de antecedentes, se consideran adecuados los perfiles que se especifican a continuación, dejando no obstante a la empresa la determinación del personal que ponga a

disposición de este contrato. Se estima que serán necesarios los siguientes perfiles para ejecutar este servicio:

- 1 Jefe de proyecto a tiempo completo. Tareas: planificación, logística y coordinación por parte de la empresa, tramitación, generación de informes y documentación, dirección técnica y supervisión de los trabajos.
- 1 Informático a tiempo completo con experiencia en tecnologías de indexación automática de texto manuscrito. Tareas: programación informática.
- 1 especialista en paleografía y diplomática aragonesa a media jornada con experiencia en lectura de letra cortesana aragonesa de los siglos XV y XVI. Tareas: entrenamiento del sistema verificación de los resultados de los trabajos de indexación y para la verificación y corrección de las transcripciones utilizadas en la indexación automática de los documentos integrantes de la serie mediante muestreo sistemático.

En todo caso, la empresa adjudicataria establecerá el número de personal que considere adecuado para la realización de este servicio con los perfiles expuestos o similares. El Jefe de Proyecto será el interlocutor con el personal a cargo del proyecto de la SGAE para todas las tareas de planificación, dirección y seguimiento de las actuaciones contempladas en el presente pliego.

Corresponde al adjudicatario la ejecución, la dirección y la coordinación directa de los medios personales que realicen las actuaciones objeto del contrato. Las ofertas de las empresas licitadoras deberán incluir indicación del número de personas que asignarían al servicio junto con información detallada de sus perfiles profesionales.

MEDIOS MATERIALES

Los trabajos objeto de este contrato no precisan de realizarse en las instalaciones del Archivo de la Corona de Aragón, pudiéndose realizar en remoto o en las instalaciones de la empresa adjudicataria.

La empresa adjudicataria pondrá a disposición del equipo de trabajo todos los medios materiales necesarios para llevar a cabo las tareas objeto del contrato, equipamiento informático, tecnológico y de procesamiento de imágenes, material de oficina y los elementos ofimáticos y de comunicaciones que sean precisos.

8. Objeto del contrato del lote 4

El lote 3 tiene por objeto la contratación de un servicio de transcripción e indexación probabilística automática mediante el empleo de inteligencia artificial de 312.000 objetos digitales procedentes de la reproducción digital de la serie documental denominada **Registro General del Sello de Corte (1502-1518)**. Esta serie fue generada por la Real Cancillería de los Reyes de Castilla entre los años 1502 y 1518, cuyos documentos originales se conservan en el Archivo General de Simancas. La estructura

formal de los registros es homogénea existiendo una división por años y los documentos están redactados en castellano en letra cortesana.

Dada la singularidad del proceso número 3 y 4, se estima que por homogeneidad de letra y lengua de esta serie documental se realice en un plazo máximo de cuatro meses.

PERFILES PROFESIONALES

Para la ejecución del servicio, se considera imprescindible que la empresa adjudicataria disponga de recursos humanos adecuados tanto en número como en cualificación técnica.

Teniendo en cuenta la experiencia de los proyectos piloto explicados en el punto de antecedentes, se consideran adecuados los perfiles que se especifican a continuación, dejando no obstante a la empresa la determinación del personal que ponga a disposición de este contrato. Se estima que serán necesarios los siguientes perfiles para ejecutar este servicio:

- 1 Jefe de proyecto a tiempo completo. Tareas: planificación, logística y coordinación por parte de la empresa, tramitación, generación de informes y documentación, dirección técnica y supervisión de los trabajos.
- 1 Informático a tiempo completo con experiencia en tecnologías de indexación automática de texto manuscrito. Tareas: programación informática.
- 1 especialista en paleografía y diplomática castellana a media jornada con experiencia en lectura de letra cortesana castellana del siglo XVI. Tareas: entrenamiento del sistema verificación de los resultados de los trabajos de indexación y para la verificación y corrección de las transcripciones utilizadas en la indexación automática de los documentos integrantes de la serie mediante muestreo sistemático.

En todo caso, la empresa adjudicataria establecerá el número de personal que considere adecuado para la realización de este servicio con los perfiles expuestos o similares. El Jefe de Proyecto será el interlocutor con el personal a cargo del proyecto de la SGAE para todas las tareas de planificación, dirección y seguimiento de las actuaciones contempladas en el presente pliego.

Corresponde al adjudicatario la ejecución, la dirección y la coordinación directa de los medios personales que realicen las actuaciones objeto del contrato. Las ofertas de las empresas licitadoras deberán incluir indicación del número de personas que asignarían al servicio junto con información detallada de sus perfiles profesionales.

MEDIOS MATERIALES

Los trabajos objeto de este contrato no precisan de realizarse en las instalaciones del Archivo General de Simancas, pudiéndose realizar en remoto o en las instalaciones de la empresa adjudicataria.

La empresa adjudicataria pondrá a disposición del equipo de trabajo todos los medios materiales necesarios para llevar a cabo las tareas objeto del contrato, equipamiento informático, tecnológico y de procesamiento de imágenes, material de oficina y los elementos ofimáticos y de comunicaciones que sean precisos.

9. Propiedad intelectual

El Ministerio de Cultura y Deporte es el único titular de los derechos de propiedad intelectual, de la explotación de los registros y demás resultados del servicio realizado.

No podrá realizarse comercialización ni distribución alguna de los contenidos generados mediante este contrato, por parte del adjudicatario ni de terceros, sin el permiso expreso de los titulares de derechos o propietarios mencionados.

10. Control de calidad

La empresa adjudicataria deberá aplicar controles de calidad sobre los distintos procesos y tareas mencionados en este pliego, a lo largo de todo el período de desarrollo del servicio.

Al finalizar el entrenamiento del proceso 3, la empresa adjudicataria deberá realizar pruebas en el demostrador mediante una evaluación objetiva. Los resultados del demostrador deberán tener al menos una tasa de error de palabra ("*Word error rate*", WER) no superior al 50 %.

11. Entregables

Tras la conclusión del proceso 2, la empresa adjudicataria deberá entregar un informe exhaustivo sobre los procesos que deban ejecutarse con posterioridad. Una vez que el coordinador del contrato por parte de la SGAE otorgue su visto bueno por escrito al citado informe, podrá realizarse el primer pago.

Con la entrega de los resultados del informe que engloba los procesos 3 y 4 y el visto bueno por escrito de la SGAE se podrá realizar el segundo pago.

La entrega de los resultados mediante la integración del servicio en PARES será requerimiento obligatorio para la recepción final y abono del último pago por los trabajos desarrollados.

12. Confidencialidad y Protección de Datos de carácter personal

La empresa adjudicataria del servicio queda expresamente obligada a mantener la absoluta confidencialidad y reserva, por tiempo ilimitado, sobre cualquier dato, información y /o documentos que pudiera conocer con ocasión del cumplimiento del contrato, por lo que habrá de establecer las medidas necesarias para la protección de los mismos y la imposibilidad de utilizarlos, publicarlos o manipularlos de ninguna manera distinta a la establecida en este pliego. El acceso por parte de la empresa adjudicataria a la información de titularidad del Ministerio de Cultura y Deporte se realizará única y exclusivamente con la finalidad de la prestación de este servicio. En estas condiciones y siempre que los datos sean tratados conforme a las instrucciones dadas y con la finalidad prevista, dicho acceso no se considerará comunicación de datos.

13. Cláusula de daño no significativo

Esta medida no afecta a las emisiones de gases efecto invernadero, ni provoca un aumento de los efectos adversos de las condiciones climáticas actuales y/o futuras, no utiliza recursos hídricos o marinos, no genera residuos, no genera emisiones contaminantes a la atmósfera y por tanto no causa un perjuicio o tiene influencia en el medio ambiente, ni afecta a la biodiversidad o a los ecosistemas pues pretende transponer las directivas 2019/789 SatCAb y 2019/790 de derechos de autor así como completar otros procedimientos normativos para adaptar las normas de propiedad intelectual a la nueva realidad digital. La medida tiene un impacto previsible nulo o insignificante sobre el objetivo medioambiental relacionado dada su naturaleza. No procede realizar una evaluación sustantiva al resultar suficiente la evaluación simplificada.

EL SUBDIRECTOR GENERAL DE LOS ARCHIVOS ESTATALES

SEVERIANO HERNÁNDEZ VICENTE