

Especificaciones técnicas

Metadato automático de contenidos del fondo documental de RTVE



Contenido

1	<i>Introducción.....</i>	3
2	<i>Objeto del contrato</i>	3
3	<i>Lote 1. Servicio de metadatado automático.</i>	3
3.1	<i>Fases y tiempos.</i>	3
3.1.1	Fase de configuración y puesta en marcha.	3
3.1.2	Fase de servicio.	4
3.2	<i>Descripción de los requisitos técnicos y de calidad.....</i>	4
3.2.1	Alcance del proyecto	4
3.2.2	Flujo de intercambio de información.	5
3.2.3	Interfaz de usuario	7
3.2.4	Tipo de Servicio.....	9
3.3	<i>Calidad.....</i>	9
3.3.1	Calidad del procedimiento de intercambio de información.	9
3.3.2	Calidad del metadatado automático	9
3.4	<i>Formación.</i>	10
3.5	<i>Otros.....</i>	11
3.5.1	Validación de niveles de calidad mínimos.	11
3.5.2	Soporte	12
3.5.3	Horas de servicios profesionales.	12
3.5.4	Gestión del proyecto.	12
3.6	<i>Formato de la Oferta técnica</i>	13
4	<i>Lote 2. Control de calidad.</i>	13
	<i>Anexo 1- Niveles de calidad en el metadatado: métricas.</i>	17
	Audio.	17
	Video.....	19

1 Introducción.

El Fondo Documental de RTVE es responsable de la preservación de los contenidos audiovisuales y sonoros producidos por la Corporación Radio Televisión Española con el fin de garantizar su disponibilidad para la emisión, la producción de nuevos contenidos y la comercialización, sin olvidar, que muchos de estos recursos tienen un valor patrimonial esencial para la historia de España de los siglos XX y XXI. Para que esto sea posible es necesario un adecuado nivel de metadato para su recuperación.

Los actuales modelos de trabajo, basados en la catalogación e indización manual de los contenidos, han demostrado ser insuficientes para asegurar la explotación de un archivo en el que todavía queda un porcentaje significativo de material sin identificar y que, al mismo tiempo, sigue creciendo exponencialmente.

Por este motivo, RTVE ha aplicado en los dos últimos años herramientas basadas en el uso intensivo de algoritmos de Inteligencia Artificial para el metadato automático de los contenidos del archivo, mejorando su accesibilidad y favoreciendo la reutilización de material hasta ahora inédito. Al mismo tiempo, las soluciones aplicadas han demostrado un rendimiento excelente sobre el material de producción más reciente por lo que es intención de RTVE aumentar el número de contenidos procesados mediante técnicas de IA.

2 Objeto del contrato

El objeto de este expediente es la contratación de un servicio de metadato automático de archivos de vídeo, cuyas características y especificaciones se exponen en este PCT, que utilice herramientas de Inteligencia Artificial de última generación, y que se integre en los actuales procesos de trabajo del Fondo Documental de RTVE.

Además, se contratará un servicio de control de calidad con el objetivo de mantener la calidad de los resultados.

3 Lote 1. Servicio de metadato automático.

Sobre los contenidos seleccionados por RTVE y durante la vigencia del contrato se realizará el análisis automático de aproximadamente:

- 15.000 horas de análisis básico de audio en español.
- 3.000 horas de análisis básico de la imagen de los contenidos.
- 1.600 horas de análisis de audio bilingüe en catalán y español.
- 500 horas de otros procesados.

En los contenidos a analizar puede pedirse el análisis de una o varias de las modalidades antes mencionadas.

El tipo de metadatos que habrán de obtenerse para cada modalidad de análisis, los procedimientos de integración con los sistemas RTVE y los requerimientos técnicos y de calidad que se habrán de cumplir se explicitan en puntos posteriores de este mismo documento.

3.1 Fases y tiempos.

3.1.1 Fase de configuración y puesta en marcha.

Una vez firmado el contrato el adjudicatario dispondrá de un mes para la configuración y puesta en marcha del servicio de forma que, a la finalización de esta fase, se tengan operativas todas las funcionalidades descritas en este PCT. Esta fase de configuración y puesta en marcha será sin coste para RTVE.

Los oferentes incluirán en su oferta una interfaz de usuario cuyos requisitos se mencionan más adelante, que permitirá a personal designado por RTVE auditar los resultados obtenidos por la/s herramienta/s de Inteligencia Artificial ofertadas en el servicio.

Al final de este período se ha de constatar por parte de RTVE que la integración del servicio con los procedimientos de RTVE es satisfactoria, que el flujo de intercambio de información funciona adecuadamente y que se alcanzan los niveles de calidad que luego se explicitan para los metadatos obtenidos. Para ello, se tratarán unas 60 horas de contenidos.

El servicio comenzará el 1 de septiembre de 2023 y durante un mes se solapará en el tiempo con el sistema aún en uso en RTVE. Como se han de realizar todas las pruebas necesarias para garantizar que el flujo de información entre RTVE y el adjudicatario de este expediente funciona correctamente, el adjudicatario garantizará que de ser necesario dichas pruebas se realizarán en los horarios definidos por RTVE, en cualquier intervalo horario, incluyendo horario nocturno y en fin de semana.

Si una vez alcanzado el final de esta fase de configuración y puesta en marcha el sistema no ha alcanzado los niveles de calidad establecidos, el contrato quedará anulado y RTVE no se hará cargo de ningún coste en que pueda haber incurrido el adjudicatario.

3.1.2 Fase de servicio.

Se trata del servicio propiamente dicho. Esta fase comenzará una vez finalizada la fase de configuración y puesta en marcha y validada la calidad del servicio por parte de RTVE, y tendrá una duración de 12 meses.

3.2 Descripción de los requisitos técnicos y de calidad

La empresa adjudicataria del lote 1 de este pliego, actuará como un integrador de soluciones de Inteligencia Artificial aplicada al metadato de archivos audiovisuales, utilizando herramientas que permitan a RTVE seguir en la vanguardia en el uso de estas tecnologías.

Para la consecución de los objetivos de este pliego, la empresa adjudicataria, puede utilizar herramientas propias de metadato o integrar y adaptar herramientas de proveedores de servicios externos, sean éstos generalistas (Amazon, Google, Microsoft, etc.) o específicos para el sector audiovisual, siempre y cuando la solución ofertada permita el cambio de una herramienta a otra que ofrezca mejores resultados en las funcionalidades solicitadas.

El adjudicatario incluirá una capa de Inteligencia Artificial, utilizando redes neuronales, o cualquier otra tecnología, que ejecutará, los procesos necesarios y suficientes para que los resultados finales cumplan con las especificaciones planteadas, adaptando los resultados obtenidos a partir de las herramientas de metadato automático de material audiovisual a las necesidades de RTVE.

Con el fin de ayudar a la empresa adjudicataria a alcanzar dicho objetivo RTVE pondrá a su disposición la Base de Datos RTVE2022 Database, (<http://catedrartve.unizar.es/rvedatabase.html>).

El sistema permitirá la coexistencia de soluciones de metadato de diversos proveedores, de tal manera que, en la medida de lo posible, cada uno de los elementos del metadato que enunciaremos después sea tratado por aquel proveedor de servicios que tenga la superioridad tecnológica en cada campo concreto.

En este sentido, dada la vertiginosa evolución del conocimiento en estos ámbitos, RTVE se reserva la posibilidad de poder exigir la realización de test de diferentes herramientas y el cambio de proveedor de herramientas de Inteligencia Artificial al menos dos veces durante la vigencia del contrato cuando los resultados obtenidos en los test anteriormente mencionados resulten, a criterio de RTVE, de una clara superioridad de las herramientas testadas frente a las facilitadas originalmente por el adjudicatario.

3.2.1 Alcance del proyecto

La elección de los contenidos a los que se aplique el tratamiento del audio, video o ambos será potestad de RTVE, en función de sus necesidades.

Procesos mínimos requeridos en cada caso:

a) Análisis de básico de audio en español.

- Transcripción de audio a texto.
- Puntuación y capitalización del texto transcrito.
- Segmentación por hablantes (diarización).
- Reconocimiento de entidades nombradas (NERC) o extracción de entidades. Entendiendo como tal la tarea que supone identificar y clasificar en categorías predefinidas como personas, organizaciones, lugares, fechas, etc., las entidades nombradas en el texto resultante de la transcripción automática.
- Extracción de palabras clave.
- Clasificación automática de contenido, entendida como la asignación de una o varias categorías en función de los conceptos expresados en el texto resultante de la transcripción automática.

b) Análisis básico de la imagen.

- Reconocimiento facial entendido como el reconocimiento de rostros y sus atributos
- Reconocimiento de identidad o de personalidades. Posibilidad de entrenar el modelo para reconocer identidades de interés para RTVE.
- Extracción de caracteres alfanuméricos (OCR) en imagen.
- Reconocimiento de logos y marcas.
- Reconocimiento de objetos.

c) Análisis de audio bilingüe en catalán y español.

Previo al análisis básico del audio el adjudicatario deberá transcribir el audio en el idioma en el que se hable y la parte que esté en catalán, traducirla al español.

d) Otros procesados

En este apartado se incluyen aquellos análisis que solo se aplicarán a los casos de uso que determine RTVE pudiendo referirse tanto a análisis específicos de audio como de vídeo como pueden ser:

- Identificación de escenas.
- Resúmenes automáticos.
- Segmentación por noticia.
- Descripción de imagen

3.2.2 Flujo de intercambio de información.

Para cada fichero de media que se quiera analizar, RTVE proporcionará:

- Un archivo MP4 con video en baja resolución a un mínimo de 1,5 Mbps, en formato H264 (generalmente 352x288) y 8 canales de audio en español (AAC estéreo 44kbps cada canal), de los que habitualmente, tienen información únicamente los 4 primeros.
- Un archivo XML con la información ya disponible en RTVE relativa a dicha media para facilitar su procesamiento e identificación. Contendrá información explícita (si ha de procesarse audio/vídeo o ambos), pero también campos que el adjudicatario podrá utilizar para establecer reglas de análisis o validación llegado el caso.

Y deberá recibir, para cada documento analizado, los siguientes archivos, todos ellos con la nomenclatura adecuada que permita asociarlos al fichero de media al que correspondan y con el código de tiempo asociado al momento en que aparece cada elemento detectado en el archivo original:

- Un archivo XML con el resultado del análisis de la media, en el formato y las características definidas por RTVE para su ingesta en el Gestor Documental, aceptándose que este formato puede variar en función de nuevos requerimientos mientras dure el contrato.
- Todos los archivos originales proporcionados por los proveedores de Inteligencia Artificial asociados al documento objeto de análisis, debidamente identificados y clasificados.

Los resultados obtenidos han de darse en español. Para la clasificación temática de los contenidos audiovisuales se utilizará un léxico determinado de común acuerdo entre el adjudicatario y RTVE, siendo necesario que el adjudicatario adapte a este léxico los resultados obtenidos de los motores de IA.

Las labores de RTVE serán exclusivamente las de selección de materiales, comprobación de resultados e incorporación de resultados en su destino final, cualquier otra labor, instalación o proceso deberá ser realizado por el adjudicatario.

A continuación, se describe el flujo de información y las características técnicas:

- El flujo del proceso empezará en un repositorio temporal, proporcionado por RTVE, donde la empresa adjudicataria recogerá cada hora los ficheros de video que allí hayan sido depositados para su tratamiento.
- Atendiendo a la información contenida en el XML anexo a la media y con las instrucciones definidas por RTVE, el adjudicatario deberá:
 - determinar el tipo y el modelo de análisis a aplicar a cada fichero, caso de ser necesario.
 - determinar el sistema de validación, de los definidos más adelante, a aplicar a cada fichero
 - mejorar la coherencia de los datos incluidos
- El XML no incluye información de la pista de audio que debe ser analizada. El sistema ofertado aportará la tecnología necesaria para la selección de la mejor pista a analizar de forma automática
- El adjudicatario enviará los ficheros de media para su procesado al proveedor/es de servicios de Inteligencia Artificial con el/los que trabaje y, tratará el resultado obtenido con sus propias herramientas para adaptarlo a las necesidades de RTVE.
- Desde el momento en que el material salga de RTVE hasta que esté listo para la revisión/validación han de pasar, como máximo, 24 horas.
- Mediante la interfaz de usuario proporcionada por el adjudicatario (definida en el punto 3.2.3), RTVE podrá chequear la calidad del análisis obtenido y validar los resultados.
- El sistema ha de implementar estos tipos de validación de contenidos de forma simultánea:
 - Automática: sin intervención humana.
 - Manual, con revisión por parte de un documentalista designado por RTVE previo análisis con la interfaz de usuario.

RTVE definirá la media a la que se aplica cada forma de validación según serie de programas, fecha de producción, cadena o cualquier otra casuística que se irá definiendo y variando a lo largo de la duración del proyecto en función de los resultados obtenidos. El adjudicatario será el responsable de desarrollar las reglas o algoritmos necesarios para tal fin.

- Según se vaya validando el contenido, se desencadenará el proceso para el envío de los metadatos a RTVE, para ello, el adjudicatario depositará en una carpeta facilitada por RTVE para cada fichero de media analizado los archivos definidos con anterioridad.

- Para el flujo de entrada en RTVE no será necesario el retorno de la media, que deberá ser destruida por la empresa adjudicataria con el siguiente criterio, cada mes se destruirán aquellos contenidos que hayan sido validados y enviados los metadatos a RTVE dos meses antes y además, que no hayan sido consultados o editados con posterioridad a dicho envío. El adjudicatario certificará mensualmente la destrucción de la media tratada.
- No obstante, el adjudicatario conservará los ficheros con el metadato (tanto los generados por los proveedores externos como los generados por él) durante el tiempo de vigencia del contrato, por si fuera necesario usarlos como backup.
- El adjudicatario será el encargado de garantizar el espacio de almacenamiento en la nube, como mínimo de 1TB.

Todo intercambio de información digital entre RTVE y el adjudicatario será mediante protocolo estándar de Internet, no dependiente de los suministradores de servicios.

El adjudicatario será el responsable de garantizar que el flujo descrito funcione correctamente tanto en la fase de recogida de los archivos desde RTVE, como en la vuelta de los ficheros resultantes del análisis de los documentos para que no se produzcan cuellos de botella que impidan el normal funcionamiento del sistema.

3.2.3 Interfaz de usuario

Además de los metadatos obtenidos para cada contenido, el adjudicatario proporcionará la herramienta para que personal de RTVE asignado a tal propósito valide, desde su puesto de trabajo, los resultados y acepte o rechace los mismos.

La interfaz será proporcionada como SaaS y se accederá a ella mediante un cliente web que permitirá como mínimo 20 accesos simultáneos.

Ha de permitir la habilitación de espacios personales diferenciados para cada usuario y garantizar la posibilidad de definir diferentes perfiles de usuarios con distintos niveles de permisos. En términos generales se han de posibilitar al menos tres niveles de usuarios:

- “Básico”, que será el utilizado por el adjudicatario del Lote 2. Nivel bajo de permisos. Podrá realizar las siguientes tareas:
 - Consulta de toda la media tratada.
 - Búsqueda y acceso a la media.
 - Selección de ficheros (transcripciones y metadatos) y descarga de los mismos.
 - No podrá editar resultados ni modificar los archivos.
- “Documentalista”. Nivel medio de permisos. Podrá realizar todas las funcionalidades de la interfaz.
- “Administrador”. Será el que cuente con el nivel más alto de permisos y además de las anteriores podrá dar de baja/alta nuevos usuarios y definir sus perfiles y permisos.

La interfaz contará al menos con las siguientes características:

- Ha de permitir búsquedas para acceder a todos los archivos aplicando distintos filtros (temporales, textuales, según el estado de procesamiento, según el tipo de análisis, etc.)
- Ha de permitir el acceso a cualquiera de los archivos de video sobre los que se ha realizado el metadato.
- Para el contenido audiovisual elegido se podrán visualizar, en un formato amigable y en una línea de tiempo los resultados obtenidos en el análisis del audio y/o de la imagen, para permitir por parte de RTVE la revisión cualitativa del metadato proporcionado.
- La interfaz ha de permitir acceder a cada punto de la media tratada:

- Situándose sobre la línea de tiempo.
 - Situándose sobre uno de los frames característicos de cada escena.
 - Situándose sobre una palabra en la transcripción del audio.
 - Seleccionando las etiquetas de los metadatos obtenidos.
- Se han de poder marcar los resultados que no se consideren correctos e incluir la facilidad de devolverlos al adjudicatario o al proveedor de Inteligencia Artificial como retroalimentación para la mejora del servicio.
 - Se ha de incluir la posibilidad de editar y modificar lo que no se considere correcto y que los cambios sean incluidos en el archivo que se devolverá a RTVE.
 - El contenido revisado ha de quedar marcado como tal.
 - El contenido enviado a RTVE ha de quedar marcado como tal.
 - El documentalista dispondrá de un botón de validar/rechazar.
 - Se ha de poder reprocesar un contenido si el análisis de audio o de vídeo ha fallado o se considera que cambiando algún parámetro del tratamiento los resultados pueden mejorar. En los documentos a reprocesar:
 - Si se incorporan en la plataforma diferentes proveedores de IA, se ha de dar la posibilidad de reprocesar el contenido que no haya cumplido con los criterios de calidad esperados cambiando de proveedor.
 - Para el tratamiento de audio se dará al personal de RTVE la opción de reprocesar los contenidos cambiando la pista de audio a analizar o cualquier otro parámetro de análisis.
 - La arquitectura del sistema se dimensionará de tal manera que se puedan reprocesar contenidos sin que afecte al normal funcionamiento del servicio.
 - Dispondrá de un cuadro de mando que recogerá, para cada día, los resultados obtenidos de la revisión realizada por el personal de RTVE, el cual servirá para hacer un seguimiento del servicio.
 - Además, se incluirá la posibilidad de exportar a una hoja de cálculo todos los datos derivados del procedimiento de análisis mediante Inteligencia Artificial:
 - los facilitados por RTVE para cada contenido: identificadores, nombre de programa, serie, duración, tipo de contenido, fecha de producción, etc.
 - los propios de proceso: fecha de recepción del fichero por parte de RTVE, fecha de análisis del audio, fecha de análisis del video, fecha de revisión por parte de RTVE, fecha de envío a RTVE de los metadatos, solicitud de reprocesado si lo hubiera, el motivo, el tipo de modelo solicitado para el reprocesado, el cambio de herramienta solicitado, la pista de audio para la que se solicita reprocesado, etc.
 - los procedentes del análisis de los documentalistas dentro del proceso de control de calidad: categorías, palabras clave, objetos, OCR, caras, identidades etc.... añadidas, cambiadas o eliminadas, así como los comentarios del documentalista.
 - los de estado del documento dentro del proceso (herramienta de IA empleada, modelo de análisis empleado (si los hubiera), pista de audio analizada, tipo de tratamiento del fichero (audio/video/audio+video) y estado del contenido dentro del proceso (revisado/no revisado, y dentro de los revisados si ha sido validado o no por un documentalista). También se ha de guardar información de si los archivos resultantes del análisis ya han sido enviados al gestor documental de RTVE o aún permanecen en local en el sistema del proveedor.

3.2.4 Tipo de Servicio.

Todo el servicio funcionará en nube o SaaS, externo a RTVE.

3.3 Calidad.

En este apartado se enumeran los criterios mínimos imprescindibles para dar por buenos los resultados obtenidos en el tratamiento de la media por el adjudicatario.

La calidad se entenderá en dos aspectos:

1. Calidad del procedimiento de intercambio de información (que los archivos y campos se devuelven de forma correcta, que el flujo de trabajo sea adecuado, etc.)
2. Calidad del metadato automático.

Si a la finalización del periodo de configuración y puesta en marcha, el sistema no ha alcanzado los niveles de calidad establecidos en ambos ámbitos, el contrato quedará anulado y RTVE no se hará cargo de ningún coste en que pueda haber incurrido el adjudicatario. Para garantizar que se cumplen los criterios exigidos, al principio de la cuarta semana de la fase de configuración se hará una prueba de aceptación que garantice que el servicio funciona de forma adecuada. Dicha prueba consistirá en testear todas las funcionalidades del sistema en condiciones de servicio en cuanto a número de horas enviadas en un día, movimiento de ficheros, calidad y cantidad de los metadatos obtenidos y su incorporación al Gestor Documental de RTVE.

3.3.1 Calidad del procedimiento de intercambio de información.

En lo referente al flujo de trabajo, el control de calidad abarcará todos los procesos definidos en el punto 3.2.2 de este documento.

En este ámbito, el adjudicatario ha de garantizar que:

1. Son procesados todos los archivos dejados por RTVE en el repositorio de salida y que el resultado obtenido es aceptado por RTVE antes de la destrucción de la media por el adjudicatario.
2. Los ficheros con los metadatos obtenidos y en el formato adecuado, están preparados para su revisión mediante la interfaz de usuario en un plazo máximo de 24 horas desde su recepción por parte del adjudicatario.
3. Tras la validación de los datos en cualquiera de sus modalidades (automática/manual) RTVE recibe, para cada archivo audiovisual proporcionado para su tratamiento los archivos con los metadatos que se indican en el punto 3.2.2 con los formatos adecuados y debidamente identificados.
4. El funcionamiento de la interfaz para la validación de los metadatos y control de calidad se ajusta a las especificaciones expuestas en el punto 3.2.3.
5. No existen cuellos de botella debido a mal funcionamiento en la red que interfieran en el flujo arriba mencionado.

3.3.2 Calidad del metadato automático

Una vez el servicio entre en funcionamiento los controles de calidad se harán, por parte de personal de RTVE, y por el adjudicatario del lote 2 de este expediente sobre el material ya tratado por el integrador a través de la interfaz de usuario definida en el punto 3.2.3.

Si el material no es validado, de acuerdo con los criterios establecidos en la Tabla 1, ha de ser reprocesado correctamente en un máximo de 48 horas.

En caso de no validarse por parte de RTVE el tratamiento realizado durante dos días consecutivos o que el informe semanal del adjudicatario del lote2 fuera desfavorable, se parará el procesado de datos durante un periodo máximo de 5 días. Durante este periodo el adjudicatario realizará los cambios oportunos para que se garantice de nuevo una calidad de servicio igual o superior a la alcanzada al principio del servicio. Las horas que se hubieran dejado de procesar durante el periodo de parada, tendrán que poder procesarse antes de 7 días a contar

desde que dicha parada se produjo. Por cada uno de los días en que el servicio esté parado, el adjudicatario será penalizado en las cantidades que se enuncian en el Anexo II del Pliego de Condiciones Generales.

El adjudicatario se comprometerá a resolver las incidencias que hayan podido dar lugar a la caída de la calidad sin que ello represente ningún coste adicional para RTVE; si el material ha de ser enviado de nuevo a un proveedor externo será a costa del adjudicatario.

Además, el sistema ofertado proporcionará una mejora continua a lo largo de todo el contrato.

Las informaciones acerca de las métricas de calidad se definen en el Anexo 1.

Tabla 1. Requisitos de calidad del metadato automático

	Material en español	Material en catalán
Transcripción de audio a texto	≤15% (WER)	≤20% (WER)
Puntuación y Capitalización	≤25% (WER)	≤25% (WER)
Segmentación por hablantes	≤25% (DER)	≤25% (DER)
Reconocimiento de entidades	Requerido	-
Extracción palabras clave	Requerido	-
Clasificación contenido	Requerido	-
Reconocimiento facial	<10% (DER)	-
Reconocimiento de identidad	<10%	-
OCR	≤7% (WER)	-
Marcas y logos	Requerido	-
Reconocimiento objetos	Requerido	-

3.4 Formación.

La formación deberá ser proporcionada por la empresa adjudicataria, sin coste para RTVE. Deberá impartirse en español, será de manera presencial en las instalaciones de RTVE en Madrid o telemáticamente si así se acuerda una vez adjudicado el proyecto, y deberá contar al menos con:

- Cursos de operación para personal de RTVE y de la empresa adjudicataria del Lote 2.

Se deberán impartir dos convocatorias en turnos de mañana y tarde y con capacidad cada una de ellas para un mínimo 10 personas, con la duración necesaria y suficiente para que los operadores queden capacitados para realizar su trabajo sin dificultad alguna.

La formación se referirá al uso de la interfaz de usuario.

De igual manera se deberá proporcionar manual de consulta en español, en formato electrónico o acceso web.

- Curso técnico para personal de RTVE.

Se deberá impartir un curso técnico, en una única convocatoria, a acordar con la Dirección del Proyecto si de mañana o tarde, donde se facilite de manera detallada la información acerca de la solución técnica adoptada para la ejecución del proyecto, sus bases teóricas y su funcionamiento. Se deberá proporcionar documentación de apoyo en formato electrónico o acceso web.

3.5 Otros.

3.5.1 Validación de niveles de calidad mínimos.

Todas las ofertas deberán realizar una prueba piloto, en las instalaciones de RTVE en Prado del Rey, donde deberá quedar demostrada la idoneidad de la solución presentada. El piloto deberá responder a una batería de pruebas que verifique el cumplimiento de las condiciones explicitadas en la tabla 1 del punto 3.3.2, cumpliendo los requisitos de calidad del metadato allí expuestos. El no cumplimiento implicará la desestimación de la oferta.

Además, han de demostrar que disponen de una interfaz de usuario con las características definidas en el punto 3.2.3 a excepción de:

- Se han de poder marcar los resultados que no se consideren correctos e incluir la facilidad de devolverlos al adjudicatario o al proveedor de Inteligencia Artificial como retroalimentación para la mejora del servicio.
- Que los cambios realizados en los metadatos sean incluidos en el archivo que se devolverá a RTVE.
- El contenido enviado a RTVE ha de quedar marcado como tal.
- El contenido revisado ha de quedar marcado como tal.
- Se ha de poder reprocesar un contenido si el análisis de audio o de vídeo ha fallado o se considera que cambiando algún parámetro del tratamiento los resultados pueden mejorar. En los documentos a reprocesar:
- La arquitectura del sistema se dimensionará de tal manera que se puedan reprocesar contenidos sin que afecte al normal funcionamiento del servicio.
- Dispondrá de un cuadro de mando que recogerá, para cada día, los resultados obtenidos de la revisión realizada por el personal de RTVE, el cual servirá para hacer un seguimiento del servicio.
- Incluirá la posibilidad de exportar a una hoja de cálculo todos los datos derivados del procedimiento de análisis mediante Inteligencia Artificial:

Los licitadores tendrán que superar todas y cada una de las pruebas técnicas en un plazo inferior a 3 horas, a contar a partir del momento en que RTVE les facilita el acceso a los ficheros de vídeo, que tendrán una duración no superior a 60 minutos y que estarán convenientemente identificados y clasificados.

El licitador correrá con todos los costes que sean necesarios para realizar las pruebas.

3.5.2 Soporte

El adjudicatario garantizará que en todo momento se dan las condiciones necesarias para que el flujo de trabajo definido en puntos anteriores de este documento se mantenga durante toda la vigencia del contrato, comprometiéndose a responder a posibles incidencias mediante un servicio de atención al usuario de 12 horas/día de lunes a viernes, de 08 a 20 horas, y con un tiempo máximo de resolución de 48 horas.

Además, el adjudicatario incluirá un sistema automático de alertas que detecte cualquier incidencia que implique paros en los procesos lo comunique al personal que RTVE designe mediante un sistema de notificación consensuado.

El no cumplimiento de estas especificaciones será penalizado en las condiciones expuestas en el Anexo II del Pliego de Condiciones Generales.

3.5.3 Horas de servicios profesionales.

Ante la posible eventualidad de que sean necesarios durante el periodo de vigencia del contrato desarrollos o actividades imposibles de prever en esta fase del proyecto, se solicita a los oferentes que incluyan, en su propuesta económica, el coste de 200 horas de servicios profesionales.

Los desarrollos se definirán en su momento y el adjudicatario presentará una valoración de horas que el responsable de RTVE deberá evaluar y aprobar. Se certificarán las horas una vez terminado el desarrollo.

3.5.4 Gestión del proyecto.

a) Composición del equipo de trabajo

Durante la duración del contrato se establecerá un equipo de trabajo permanente en el que estarán representados, por parte de RTVE, los departamentos de Archivo, Innovación y Sistemas, en el que la empresa adjudicataria estará representada por cuantos recursos considere oportunos, al menos uno de ellos deberá ser castellanoparlante.

El adjudicatario nombrará un responsable del servicio dentro de su equipo, que será el interlocutor único con el personal de RTVE para la coordinación y priorización de tareas. Así mismo, desde RTVE también habrá otro coordinador, que será su único contacto para cualquier duda o problema que pueda surgir.

Algunas de las actividades que realizará el equipo de trabajo serán:

- Planificación y asignación de tareas.
- Definición del flujo de trabajo en la recogida de media y la entrega de metadatos.
- Supervisión de actividades.
- Supervisión del cumplimiento de los trabajos.
- Gestión de prioridades.
- Optimización de procesos.
- Asistencia a las reuniones de coordinación.
- Proponer y aceptar nuevas funcionalidades a ser incluidas en los procesos de metadato.
- Elaboración de informes en base a métricas que permitan medir el rendimiento, la efectividad y la calidad del servicio.

Por su parte, la Dirección de Proyecto nombrada por RTVE tendrá las atribuciones necesarias para definir, en cada momento, las prioridades dentro del mismo, así como para convocar cuantas veces considere oportunas reuniones de seguimiento con el grupo de trabajo.

La Dirección de Proyecto de RTVE será la responsable de aprobar los retrasos justificados.

En general y salvo causas de fuerza mayor, el cauce de comunicación en el seno del proyecto será entre los coordinadores de Proyecto de ambas partes.

b) Reuniones de seguimiento.

Sin perjuicio de las reuniones que la Dirección del Proyecto pueda convocar fuera de agenda, el grupo de trabajo se citará en reuniones de seguimiento con periodicidad semanal durante los cuatro meses de la primera fase del proyecto y con periodicidad mensual durante la fase de servicio. En todos los casos, la empresa adjudicataria estará encomendada de levantar acta de reunión y hacérsela llegar a todo el grupo de trabajo.

c) Entregables.

La empresa adjudicataria preparará de cara a su discusión y análisis en las reuniones de seguimiento arriba mencionadas un informe periódico, en base a métricas, que permitan medir el rendimiento, la efectividad y la calidad del servicio.

La definición del contenido de este informe tendrá lugar en la reunión de inicio del proyecto a propuesta de la empresa adjudicataria y validado por RTVE.

3.6 Formato de la Oferta técnica

Será obligatoria la entrega de una oferta técnica que incluirá información suficiente que permita hacer la evaluación de la solución ofertada y tendrá la estructura que se indica a continuación:

- Número y perfiles del personal involucrado en el proyecto con aportación del CV de cada uno. Si bien no se requiere ninguna titulación o categoría concreta, si es imprescindible que el equipo propuesto tenga formación y experiencia en el campo de la inteligencia artificial aplicada al procesamiento del lenguaje natural, machine learning, big data y cualquier otra tecnología relacionada con los procesos descritos en este Pliego.
- Descripción detallada de la infraestructura técnica que se usaría para dar el servicio.
- Descripción del flujo de la información durante el servicio y mecanismos de interacción adjudicatario-RTVE.
- Identificación del/de los proveedores de servicios de Inteligencia Artificial con el/los que se procedería al metadato automático de los contenidos, diferenciando para audio/video/otros procesados si no se tratara del mismo.
- Equipo de trabajo.
- Relación de proveedores de los servicios de IA con los que el oferente pueda trabajar en su plataforma.
- Planificación de los trabajos a realizar en la primera fase con diagrama temporal de los mismos especificando las tareas a realizar y explicitando el momento de la realización de los siguientes hitos:
 - Plataforma lista para el envío de los primeros archivos para su metadato.
 - Plataforma lista para procesos de movimiento de archivos.
- Persona de contacto.
- Todos los puntos explicitados en la oferta económica deberán estar incluidos y desarrollados en la oferta técnica para su valoración.

4 Lote 2. Control de calidad.

El servicio descrito en este lote entrará en funcionamiento cuando se dé por buena la entrada en modo estacionario del Lote 1, previsiblemente el 1 de octubre de 2023.

4.1 Descripción del servicio.

El adjudicatario del Lote 2 deberá realizar un control de calidad sobre los metadatos obtenidos de la media procesada por el sistema del adjudicatario del Lote 1.

Para ello, se le habilitará un espacio de trabajo en la interfaz de usuario donde podrá localizar y descargar aquellos archivos elegidos para revisión y al que podrá acceder mediante las credenciales asignadas al perfil básico. El adjudicatario de este Lote 2 se compromete a asistir a los cursos de formación que imparta sobre el funcionamiento de su interfaz de usuario el adjudicatario del Lote 1.

La evaluación consistirá en un análisis de los resultados presentados por el automatismo del sistema del Lote 1 en las categorías y métricas especificadas en la tabla 1 del punto 3.3.2.

Este análisis se realizará siguiendo las métricas descritas en el Anexo I sobre una o varias catas con un total de aproximadamente 5 minutos diarios de contenido tratado por el servicio de metadatado, de los cuales una muestra cada semana corresponderá a programas en catalán. A efectos del cálculo presupuestario, se estima que se necesitan 45 minutos para el análisis de 5 minutos incluidas las tareas especificadas a continuación.

- Cálculo del WER de la transcripción automática, segmentación por hablantes y capitalización. En el caso de las muestras de programas en catalán, solo se pedirá el WER en este idioma
- Cálculo del WER del OCR y del DER del reconocimiento facial y reconocimiento de identidad.
- Chequeo de que el sistema ejecuta el resto de las funciones requeridas.

Las reglas para la selección de las catas serán comunicadas al adjudicatario de este Lote por parte de RTVE al comienzo del proyecto. No obstante, y en función de la marcha del proyecto, RTVE podrá variar los criterios de elección de los contenidos a chequear, manteniendo siempre la media de minutos diarios.

El adjudicatario de este lote deberá realizar las siguientes acciones:

- Seleccionar y descargar desde la plataforma de gestión el texto transcrito y los archivos con los metadatos obtenidos correspondiente a las catas de media que han de ser analizadas.
- Corregir cada parámetro a validar sobre una copia del texto original y realizar los cálculos arriba descritos.
- Entregar el archivo original y el archivo corregido a RTVE y al adjudicatario del Lote 1 una vez finalizado el proceso de corrección.
- El adjudicatario del Lote 2 deberá generar un informe semanal donde se recojan los resultados del análisis, así como los comentarios que considere necesarios sobre el funcionamiento de la herramienta, tanto en lo referido a los resultados ofrecidos como a las incidencias que puedan surgir durante su uso.
- En un plazo de 4 días laborables al finalizar cada semana analizada se adjuntará la siguiente documentación:
 - **Excel de datos semanales.** En un formato a indicar por RTVE una vez adjudicado el expediente, se presentará de forma incremental los resultados de la última semana analizada.
 - Un informe por cada una de las muestras analizadas con la siguiente estructura:
 - **Resumen de datos.** Un cuadro en que se refleje de forma individual el recuento numérico de cada uno de los parámetros descritos.
 - **Análisis de la transcripción.** Un documento de texto con el contenido de la transcripción sobre el cual se marcarán de forma distinguible con un código de colores, los distintos errores especificados en lo puntos anteriores.

- **Comentarios / consideraciones.** Un análisis subjetivo acerca de las impresiones derivadas del análisis realizado.

Por último, al término del proyecto se incluirá un informe final donde se incluya un resumen de lo analizado y se explique la evolución de la calidad.

- Aunque las entregas de los informes se solicitan una vez por semana, cualquier incidencia ya sea en el servicio o incumplimiento de los márgenes de calidad establecidos deberá reportarse a RTVE en el mismo día del análisis. RTVE podrá requerir el envío de los resultados con periodicidad inferior a la semana si se dan unos niveles de calidad críticos y así lo considera.

Evidentemente la actividad de este Lote está supeditada a la del Lote 1, por lo tanto, en caso de que por alguna circunstancia se produzca una interrupción en la actividad del Lote 1, los trabajos recogidos en este Lote se verán afectados de la siguiente manera:

- Paro menor que una semana: conllevará una redistribución de las catas a analizar.
- Paro superior a una semana: la actividad de este Lote quedará interrumpida. En función de las circunstancias que originaran este paro, podría dar lugar a un cambio en la fecha de finalización del contrato o a la cancelación del mismo.
- La cancelación del contrato al adjudicatario del Lote 1 si no alcanzara los niveles de calidad exigidos en la prueba de aceptación, daría lugar a la cancelación del contrato de este Lote.

Por si fuera necesidad de RTVE evaluar de forma extraordinaria algunos contenidos de su interés, los oferentes a este Lote deberán incluir en su oferta 20 horas/persona de análisis de media, tanto de cálculo de WER, como del resto de parámetros de medida descritos anteriormente.

4.2 Equipo de trabajo

Durante la duración del contrato se establecerá un equipo de trabajo permanente en el que estarán representados, por parte de RTVE, los departamentos de Archivo, Innovación, en el que la empresa adjudicataria del lote 2 estará representada por cuantos recursos considere oportunos y al menos uno de ellos deberá ser castellanoparlante.

El adjudicatario nombrará un responsable del servicio dentro de su equipo, que será el interlocutor único con el personal de RTVE para la coordinación y priorización de tareas. Así mismo, desde RTVE también habrá otro coordinador, que será su único contacto para cualquier duda o problema que pueda surgir.

4.3 Reuniones de seguimiento.

Sin perjuicio de las reuniones que la Dirección del Proyecto pueda convocar fuera de agenda, el grupo de trabajo se citará en reuniones de seguimiento cuya periodicidad se ajustará al devenir del proyecto pudiendo éstas celebrarse simultáneamente con el adjudicatario del Lote 1 o sólo con RTVE.

4.4 Formato de la Oferta técnica

Será obligatoria la entrega de una oferta técnica que incluirá información suficiente que permita hacer la evaluación de la solución ofertada y tendrá la estructura que se indica a continuación:

- Número y perfiles del personal involucrado en el proyecto. Con aportación del CV de cada uno. Si bien no se requiere ninguna titulación o categoría concreta, si es imprescindible que el equipo propuesto tenga formación y experiencia en el control de calidad de los resultados obtenidos mediante la aplicación de automatismos al procesamiento del lenguaje natural.
- Descripción del servicio.
- Relación de instalaciones, medios técnicos y humanos.
- Compromiso de calidad del servicio.

- Organización del trabajo y sistema de aseguramiento de la calidad del servicio.
- Cumplimiento y mejoras de las especificaciones técnicas requeridas.
- Persona de contacto.
- Todos los puntos explicitados en la oferta económica deberán estar incluidos y desarrollados en la oferta técnica para su valoración.

Anexo 1- Niveles de calidad en el metadato: métricas.

A continuación, se describen las métricas que se utilizarán para la evaluación de cada una de las funcionalidades solicitadas en el presente expediente.

Los niveles de calidad solicitados están recogidos en la Tabla 1.

Audio.

- **Transcripción de audios a texto:** el control de calidad de la transcripción de audio a texto se calculará mediante la tasa de error por palabras (WER en su acrónimo en inglés) que se obtiene comparando la transcripción automática con la transcripción real. Responde la siguiente fórmula matemática:

$$WER = \frac{N_{INS} + N_{BOR} + N_{SUS}}{N}$$

donde:

- N_{INS} es el número de inserciones o palabras reconocidas que no se han pronunciado.
- N_{BOR} es el número de borrados o palabras no reconocidas pero que se han pronunciado.
- N_{SUS} es el número de sustituciones o palabras reconocidas erróneamente.
- N el número total de palabras a reconocer.

Para calcular esta tasa se utilizará la herramienta SCLITE incluida en el conjunto de herramientas SCTK (Speech Recognition Scoring Toolkit) del National Institute of Standards and Technology (NIST en su acrónimo inglés) o una herramienta con características similares.

- **Signos de puntuación y capitalización:** el control de calidad de la puntuación y capitalización, se realizará mediante la cuantificación de la tasa WER, donde N_{INS} , N_{BOR} N_{SUS} incluyen puntos y mayúsculas. Se requerirán comas, pero no se tendrán en cuenta para el cálculo del WER.
- **Segmentación por hablantes:** el rendimiento de los sistemas se medirá en función del Error de Diarización (DER en su acrónimo en inglés) comparando la segmentación obtenida por los sistemas con una segmentación real.

El DER contabiliza la fracción de tiempo que no se atribuye correctamente a un hablante específico, incluyendo las regiones en las que hay más de un hablante y responde a la siguiente fórmula matemática:

$$DER = \frac{T_{PER} + T_{FA} + T_{HAB}}{T_{VOZ}}$$

Donde:

- T_{HAB} : es el tiempo de error del hablante o la cantidad de tiempo que se ha asignado a un hablante incorrecto.
- T_{PER} : es el tiempo de hablante perdido o la cantidad de tiempo que el hablante está presente pero no está etiquetada por el sistema de diarización.
- T_{FA} : es el tiempo de alarma falsa o la cantidad de tiempo que un sistema de diarización ha etiquetado con un hablante, pero no está presente.

Segmentos consecutivos del mismo hablante, con un silencio de menos de 2 segundos, se juntarán y se considerarán como un solo segmento. Se permitirá un margen de error de 0.25 segundos, antes y después de cada límite de referencia, para tener en cuenta tanto las anotaciones inconsistentes como la incertidumbre acerca de cuándo comienza o termina un hablante.

- **Reconocimiento de entidades**

El sistema deberá ser capaz de identificar al menos, los siguientes tipos de entidades:

- **Personas (PER):** nombres de personas individuales y nombres de familias, nombres que hacen referencia a grupos de personas que se integran en una organización formal (madridistas, atléticos), personajes de ficción, gentilicios (europeos, italianos, madrileños), iniciales y pseudónimos, títulos, tratamientos o cargos cuando no están asociados a un nombre propio (Presidente del gobierno, Conde de Ahumada, Primer ministro).
 - **Organizaciones (ORG):** partidos políticos, instituciones públicas, administraciones públicas, Organizaciones no gubernamentales (ONG's), organizaciones internacionales, organizaciones militares, organizaciones religiosas, compañías, equipos deportivos, instituciones educativas y de investigación, grupos de música, empresas de entretenimiento y de radio y televisión.
 - **Localizaciones (LOC):** topónimos de ciudades, países, provincias, regiones, continentes, masas de agua y formaciones geológicas.
 - **Eventos (EVT):** conferencias, ruedas de prensa, comparencias, conciertos, festivales, periodos de tiempo (Semana Santa, Navidad, Carnaval, etc.), guerras, batallas, desastres provocados por la acción del hombre, desastres naturales, fenómenos atmosféricos, juicios, etc.
 - **Productos (PRO):** nombres de productos electrónicos, vehículos, armamento, servicios web, libros, software, películas, programas de televisión, periódicos y periódicos digitales, obras de arte, nombres de documentos legales, tratados, iniciativas o programas, etc.
 - **Fechas (DAT):** menciones a años o periodos de años (1977, 1936-1939), décadas (los ochenta, los 70), fechas concretas (25 de abril, 14 de julio, 25/12/1992) menciones a un siglo o periodo de siglos (expresados con números romanos o con referencias del tipo "Siglo de oro", Renacimiento, etc.) o referencias a fechas emblemáticas como 11-S, 23F.
 - **Instalaciones (FAC):** edificios y estructuras desarrolladas por el ser humano: estadios, parques, museos, teatros, hoteles, aeropuertos, puertos, estaciones de tren, iglesias, calles, autopistas, puentes, túneles, aparcamientos y otros edificios, equipamientos o servicios urbanos o no.
- **Extracción palabras clave,** será requerido, lo que implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo.
 - **Clasificación del contenido:** será requerido, lo que implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo
 - **Transcripción en catalán:** Se utilizarán las mismas métricas que para la transcripción en español.

Video.

- **Reconocimiento facial** Se utilizará la tasa de error de diarización (DER) tal y como se ha definido en la funcionalidad de “segmentación por hablantes”, sustituyendo en este caso la voz (hablantes) por la imagen (rostros).
- **Reconocimiento de identidad**, Se utilizará la tasa de error de diarización (DER) tal y como se ha definido en la funcionalidad de “segmentación por hablantes”, sustituyendo en este caso la voz (hablantes) por la cara de la persona a reconocer).
- **OCR** (en rótulos y títulos de crédito no en imágenes dentro de la imagen). En este caso se utiliza la misma métrica, WER, explicada en la funcionalidad de “transcripción de audios a texto con signos de puntuación y capitalización”.
- **Marcas y logos**, será requerido, lo que implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo.
- **Reconocimiento de objetos**, será requerido, lo que implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo.