

# Especificaciones técnicas

---

## IA aplicada al metadato de contenidos del Archivo RTVE

---



## 1 Introducción.

El Archivo de RTVE es responsable de la preservación de los contenidos audiovisuales y sonoros producidos por la Corporación Radio Televisión Española, garantizando su disponibilidad para su emisión, la producción de nuevos contenidos y su comercialización. Además, muchos de estos recursos poseen un valor patrimonial fundamental para la historia de España de los siglos XX y XXI.

Para hacer frente a este desafío, RTVE ha incorporado en los últimos años herramientas avanzadas basadas en Inteligencia Artificial para el metadado automático de los archivos audiovisuales, mejorando su accesibilidad y facilitando la explotación de material previamente inédito, demostrado ser eficaces en la gestión del material que por distintas razones no dispone de un tratamiento documental adecuado que facilite su recuperación.

Contar con un sistema de generación automática de metadatos en el Archivo sigue siendo esencial ante el crecimiento exponencial de los contenidos audiovisuales en el Archivo y la reducción de personal. Esta tecnología optimiza los recursos disponibles, agiliza la catalogación de grandes volúmenes de contenido y el acceso y puesta en valor del patrimonio audiovisual.

El objeto de este expediente es la contratación de un servicio de metadado automático de archivos audiovisuales, 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 Archivo RTVE.

## 2 Lote UNICO. Servicio de metadado automático.

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

- 18.000 horas de análisis de audio en español.
- 5.000 horas de análisis de la imagen de los contenidos.
- 5.000 horas de análisis de audio multilingüe.

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.

### 2.1 Fases y tiempos.

#### 2.1.1 Fase de configuración y puesta en marcha.

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.

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 contenidos hasta un máximo de 60 horas.

Esta fase comenzará el 1 de septiembre de 2025 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.

### *2.1.2 Fase de servicio.*

Se trata del servicio propiamente dicho. Esta fase comenzará el 1 de octubre, 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.

## **2.2 Descripción de los requisitos técnicos y de calidad**

La empresa adjudicataria 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 o específicos para el sector audiovisual, siempre y cuando la solución ofertada permita, en el periodo tiempo de vigencia del contrato, el cambio de una herramienta a otra que ofrezca mejores resultados en las funcionalidades solicitadas.

El adjudicatario incluirá una capa de Inteligencia Artificial adicional que ejecutará los procesos necesarios 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, teniendo en cuenta que estas necesidades de adaptación pueden ser diferentes según el tipo de serie o programa a analizar debiéndose atender todas ellas. En el momento que estamos viviendo en lo que a tecnología se refiere, resulta imprescindible que la solución ofertada sea ágil y adaptativa a un modelo de servicio iterativo y colaborativo, aplicando metodologías DevOps, de forma que permita introducir cambios de código frecuentemente y desplegarlos con rapidez aprovechando la retroalimentación de las revisiones realizadas en RTVE.

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.

### *2.2.1 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.

#### *2.2.1.1 Procesos mínimos requeridos.*

A continuación, se enumeran los procesos mínimos requeridos en cada caso:

#### **Análisis de audio en español:**

- Transcripción de audio a texto.
- Puntuación y capitalización del texto transcrito.

- Reconocimiento de entidades nombradas (NER) en el texto resultante de la transcripción automática.
- Extracción de palabras clave que sirvan para identificar los temas tratados. El sistema permitirá el establecimiento de un número máximo de palabras clave presentadas, atendiendo al criterio proporcionado por RTVE
- 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 y acorde con la clasificación de RTVE.
- Clasificación del tipo de contenido, identificando si es una entrevista, un documental, un informativo, un reportaje, una rueda de prensa...
- Realización de resúmenes automáticos. El sistema permitirá adaptar el formato de los resúmenes a los requerimientos de RTVE.
- Segmentación semántica, a modo de minutado automático de los contenidos.
- Segmentación por noticia para el análisis de programas informativos y magazines.
- Segmentación por hablantes (diarización).
- Identificación del género del hablante.
- Reconocimiento de identidad a partir de la voz exclusivamente de personajes públicos con la finalidad de poder realizar trabajos informativos relacionados con el servicio público esencial que RTVE tiene encomendado y con fines de archivo y preservación del patrimonio cultural y científico.

#### **Análisis de la imagen:**

- Reconocimiento de presencia de personas en la escena.
- Reconocimiento de identidad exclusivamente de personajes públicos con la finalidad de poder realizar trabajos informativos relacionados con el servicio público esencial que RTVE tiene encomendado y con fines de archivo y preservación del patrimonio cultural y científico.
- Extracción de caracteres alfanuméricos (OCR) en imagen.
- Reconocimiento de logos y marcas. Posibilidad de adaptar el modelo para reconocer las de interés para RTVE.
- Reconocimiento de objetos relevantes.
- Identificación de tipos de escenas.
- Descripción de imagen
- Análisis de programas informativos y magazines. Caso especial en los que el programa responde a estructuras marcadas: se deberá identificar y detectar con sus códigos de tiempo de inicio y final las siguientes secciones y elementos visuales:
  - Cabecera- Secuencia de introducción que aparece al inicio del programa y que generalmente incluye el título, logotipos, gráficos, imágenes y música distintiva.
  - Titulares- Enunciados breves que resumen las noticias, temas o segmentos más importantes que se abordarán durante la emisión.
  - Bloque deportes.
  - Bloque meteorológico.
  - Títulos de crédito.

- Señal en directo.
- Entradillas de corresponsales o enviados especiales.
- Copy- texto que aparece al final de la emisión, información legal, productoras responsables y otros datos relevantes sobre la propiedad intelectual del contenido, además de la fecha de emisión.

### **Análisis de audio multilingüe:**

El sistema ofertado deberá ser capaz de reconocer el idioma/s en que se hable y, para aquellos archivos en los que el audio no esté en español, y previo al análisis, deberá transcribir el audio en el idioma en el que se hable y traducirlo al español, al menos para el inglés y los idiomas hablados en Baleares, País Vasco, Comunidad Valenciana, Galicia y Cataluña.

Existirán dos tipos de casuísticas:

- Audios en un solo idioma distinto al español.
- Audios bilingües en otros idiomas y español.

En este último caso, se traducirá al español la parte que esté en un idioma distinto.

### **Interacciones entre los distintos procesados:**

- Entre los distintos procesados de audio:
  - Incluir en el resumen el tipo de contenido (entrevista, documental...).
- Entre los distintos procesados de vídeo
  - Incluir en “descripción de imagen” los datos obtenidos del reconocimiento de identidades, de OCR o de reconocimiento de objetos y en el caso de informativos, también los elementos estructurales o visuales detectados.
- Con los tesauros unificados de RTVE:
  - Dada la evolución continua de este tesoro, el adjudicatario deberá interactuar con el mismo al menos dos veces por semana para actualizarse con las nuevas versiones proporcionadas por RTVE.
- Con las bases de datos de RTVE VIRIA

En el momento de la adjudicación está previsto que RTVE cuente con un conjunto de bases de datos en un sistema de archivo de Linux de caras, voces y logos.

El adjudicatario ha de poder interactuar con la información contenida en dichas bases de datos al menos dos veces por semana y ser capaz de garantizar que la usará para mejorar sus sistemas de reconocimiento y adaptarse a los requerimientos de RTVE en este sentido. La interacción con dichas bases de datos se realiza vía API.

#### *2.2.1.2 Puntos de interés para RTVE*

El oferente podrá incluir en su oferta otras funcionalidades, las cuales servirán para obtener mejor puntuación en los “Criterios Técnicos Sujetos a Juicio de Valor” siempre que estén desarrolladas en el momento de la prueba técnica.

En este sentido, se enumeran también algunas funcionalidades de especial interés para RTVE.

### **Puntos de interés para RTVE en análisis de la imagen:**

- Reconocimiento de edificios y monumentos emblemáticos
- En análisis de programas informativos y magazines:

- Catch- conexión en directo con un reportero, corresponsal o una fuente externa. Se suele mostrar una pequeña ventana o ventana emergente en pantalla que muestra al periodista o el lugar de los hechos.
- Detección de imágenes de redes sociales,
- Detección de elementos de realidad aumentada.
- Detección piezas (combinan imagen con locución y sonido ambiente previamente producidas).
- Detección de ráfagas o cabeceras de secciones
- Detección de duplex, tríplex...
- Detección de colas (vídeos que sólo contienen imagen y sonido ambiente con locución del presentador).
- Reconocimiento de imagen tomada en plató, incluyendo la detección de videowall

#### **Puntos de interés para RTVE en análisis de audio multilingüe:**

- Reconocimiento y transcripción en otros idiomas. El oferente deberá enumerar en su oferta los idiomas en los que puede trabajar su sistema.

#### **Puntos de interés para RTVE en interacciones entre los distintos procesados:**

- Entre procesados de audio y vídeo, incluir en el resumen las identidades más relevantes reconocidas en el análisis de imagen y/o voz
- Sugerencias de indización para temas, personas, lugares y fechas basados en los tesauros de RTVE.

#### *2.2.1.3 PoC durante el contrato*

En el tiempo de duración del contrato, se realizarán algunas pruebas de concepto (PoC) con el fin de aplicar procesados especiales o diferentes a los habituales a colecciones concretas que así lo requieran, o de testear nuevas prestaciones y desarrollos que mejoren el tratamiento documental de nuestros archivos y que se aplicarán a los casos de uso que determine RTVE.

Estas pruebas pueden referirse tanto a análisis específicos de audio como de vídeo, por ejemplo:

- Detección de efectos como aplausos, ovaciones, gritos.
- Detección de contenido violento, discursos de odio.
- Búsqueda por imágenes.
- Detección y reconocimiento de obras musicales.

Para este fin, el oferente incluirá en su oferta económica un total de 200 horas de desarrollo.

No obstante, en caso de que el oferente ya disponga de estas funcionalidades en el momento de presentación de la oferta, podrá incluirlas en su oferta y se tendrá en cuenta en la valoración técnica.

Además, para atender al coste de horas de análisis generadas en las PoC y con el fin de facilitar la elaboración de la oferta económica se ha establecido como "Unidad de análisis PoC" el importe en euros/hora de un módulo que intervenga en la PoC, entendiendo como tal, la parte del procesado que realice una tarea concreta. El número de unidades que se consumirán para cada análisis dependerá de la complejidad de PoC.

Una vez definida por RTVE la PoC que se quiera hacer, el adjudicatario deberá presentar para su aprobación la estimación en horas de desarrollo y el importe de hora de análisis (que será la suma de las "unidades de análisis PoC" necesarias para analizar una hora de media).

### 2.2.2 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).
- 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 cuando corresponda, con el código de tiempo asociado al momento en que aparece cada elemento detectado en el archivo original:

- Un archivo XML por cada una de las funcionalidades con el resultado del análisis de la media, con la estructura 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, recogidos en un fichero con formato ZIP.

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 el proveedor, donde la empresa adjudicataria recogerá cada hora los ficheros de video y los XML asociados que allí hayan sido depositados para su tratamiento.
- 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.
- En circunstancias normales, el material estará listo para la revisión/validación en un tiempo que será igual o menor a la mitad del tiempo de duración del fichero a contar desde el momento en que el fichero haya llegado a la plataforma del adjudicatario. En momentos de alta demanda de ficheros a analizar, el material estará disponible para su validación como máximo 24 horas desde su salida de RTVE.
- El sistema permitirá a los usuarios “documentalistas” y “administradores” de RTVE la cancelación del análisis de aquellos documentos que aún no hayan sido analizados. Esta acción conllevará el borrado del documento en el sistema del adjudicatario.
- En principio el orden de análisis de ficheros será el de llegada de éstos al adjudicatario, no obstante, mediante la interfaz de usuario, el sistema permitirá que RTVE priorice aquellos ficheros que considere más urgentes de forma manual o automatizada utilizando los datos del XML facilitado por RTVE.
- El sistema será escalable en el sentido de que permitirá la incorporación de las evoluciones que ofrezcan las tecnologías emergentes, como es el caso de la IA Agéntica, que puedan aportar mejoras en los resultados o facilidades para los usuarios automatizando algunas tareas. Aunque la utilización de agentes deberá ir en aumento a lo largo del contrato, desde el inicio del mismo se dispondrá al menos

de uno, capaz de detectar problemas en la transcripción enviando una alerta por mail a los usuarios designados por RTVE.

- Mediante la interfaz de usuario proporcionada por el adjudicatario, definida en este pliego, RTVE podrá chequear la calidad del análisis obtenido y validar los resultados.
- El sistema incluirá estos tipos de validación de contenidos de forma simultánea:
  - Automática: sin intervención humana.
  - Manual, con revisión por parte de un documentalista designado por RTVE.

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 contrato 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 los metadatos obtenidos del análisis.
- 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 seis meses antes. El adjudicatario certificará la destrucción de la media tratada.
- En el caso de media analizada para una prueba de concepto, la media deberá permanecer en el sistema durante todo el periodo de tiempo que dure la PoC.
- 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 necesario para cumplir estos requisitos.

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.

A la finalización del contrato el adjudicatario hará entrega, en un soporte físico, de todos los metadatos generados en el transcurso del proyecto, organizados por carpetas denominadas con el mismo identificador que el utilizado para hacer llegar al Gestor documental de RTVE. Cada una de estas carpetas contendrá todos los metadatos generados para el ítem al que se refiere, en los formatos pedidos por RTVE.

### 2.2.3 Interfaz de usuario

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 soportará como mínimo 30 accesos simultáneos.

Ha de permitir definir diferentes perfiles de usuarios con distintos niveles de permisos. En términos generales se han de posibilitar al menos tres niveles de usuarios:

- “Básico”. 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 mediante palabras incluidas en el S2T, 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 metadatado.
- 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 metadatado proporcionado.
- En el caso de que el documento sea multilingüe, se visualizarán tanto el S2T del idioma/s original/es como la traducción al español
- La interfaz ha de permitir acceder a cada punto de la media tratada:
  - Situándose sobre la línea de tiempo.
  - Un botón que permita avanzar y retroceder 5 segundos desde el punto en que se esté posicionado en 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 ha de incluir la posibilidad de editar y modificar lo que no se considere correcto y que los cambios sean incluidos en los archivos que se devolverán a RTVE.
- Cuando se modifique una entidad en el campo de entidades, este cambio se aplicará también a todas las veces que aparezca la entidad en la transcripción y se conservará en la información que se envíe 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. Si la causa de este reprocesado se debe a errores en el análisis, no tendrá coste para RTVE. 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 configurable que servirá para hacer un seguimiento del servicio y que incluirá al menos la información descrita en el párrafo siguiente.
- 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, departamento o usuario que ha solicitado el análisis, etc.
  - los propios de proceso: fecha de recepción del fichero desde RTVE, fecha de análisis del audio, fecha de análisis del video, idioma/s procesado/s, fecha de revisión por parte de RTVE, fecha de envío a RTVE de los metadatos, solicitud de reprocesado si lo hubiera, el motivo, 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 y edición de los metadatos, incluyendo los cambios realizados y el usuario que los hizo.
  - El número de horas a analizar enviados por cada usuario o grupos de usuarios definidos por RTVE
  - los de estado del documento dentro del proceso. 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.

#### 2.2.4 Tipo de Servicio.

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

### 2.3 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.

#### 2.3.1 Calidad del procedimiento de intercambio de información.

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

En este ámbito, el adjudicatario ha de garantizar que:

1. Son procesados todos los archivos enviados por RTVE 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 los plazos contemplados en este documento.
3. Tras la validación de los datos RTVE recibe, para cada archivo audiovisual proporcionado para su tratamiento los metadatos 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 este documento.
5. No existen cuellos de botella debido a mal funcionamiento en la red que interfieran en el flujo arriba mencionado.

### 2.3.2 Calidad del metadato automático

Una vez el servicio entre en funcionamiento los controles de calidad se harán, por parte de personal de RTVE, sobre el material ya tratado por el integrador a través de la interfaz de usuario definida en este documento.

Si el material no es validado, de acuerdo con los criterios establecidos en la tabla siguiente, ha de ser reprocesado correctamente en un máximo de 24 horas.

**Tabla 1. Requisitos de calidad del metadato automático**

Transcripción de audio a texto	≤ 5% (WER)
Puntualización y Capitalización	≤7% (WER)
Segmentación por hablantes	≤10% (DER)
Traducción desde diversos idiomas	Requerido
Reconocimiento de entidades	Requerido
Extracción palabras clave	Requerido
Clasificación automática de contenido	Requerido
Tipo de contenido	Requerido
Realización de resúmenes automáticos.	Requerido
Segmentación semántica	Requerido
Segmentación por noticia	Requerido
Identificación del género del hablante	Requerido
Reconocimiento de identidad a partir de la voz.	Requerido
Reconocimiento personas en la escena	<10% (DER)
Reconocimiento de identidad	<10%
Extracción de caracteres alfanuméricos (OCR)	≤7% (WER)
Reconocimiento de logos y marcas	Requerido
Reconocimiento objetos	Requerido
Identificación de tipos de escenas.	Requerido
Descripción de imagen	Requerido
Análisis de informativos (puntos obligatorios)	Requerido

En caso de no validarse por parte de RTVE el tratamiento realizado durante dos días consecutivos 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 Pliego de Cláusulas Administrativas Particulares.

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.

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

## 2.4 Formación.

La empresa adjudicataria será la responsable de la formación. Deberá impartirse en español, será en modalidad on line, sin coste para RTVE y deberá contar al menos con:

- Cursos de operación para personal de RTVE.

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 al inicio del proyecto, 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. Durante la duración del contrato RTVE podrá solicitar un segundo curso que recoja información sobre los cambios tecnológicos introducidos.

## 2.5 Otros.

### 2.5.1 Validación de niveles de calidad mínimos.

Todos los oferentes 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:

- Que se realizan todos los procesos mínimos requeridos para cada tipo de análisis y las interacciones entre ellos, cumpliendo los requisitos de calidad del metadato expuestos en la Tabla 1, así como los descritos en las funcionalidades de interés para RTVE, que el oferente haya incluido en su oferta.
- Que la interfaz de usuario se ajusta a lo solicitado en este documento.

El no cumplimiento implicará la desestimación de la oferta.

El día de la prueba, RTVE pondrá a disposición de los oferentes varios archivos audiovisuales debidamente ordenados y clasificados y con una duración total no superior a 90 minutos. Los licitadores dispondrán de un máximo de 3 horas para la realización de todas las pruebas técnicas. Los resultados de estas pruebas deberán estar accesibles en la interface de usuario ofertada. El licitador habilitará un usuario y contraseña para el equipo de RTVE con permisos de edición y mantendrá los resultados de las pruebas y el acceso a RTVE por un periodo no inferior a 10 días. Además, entregará todos los metadatos obtenidos en ficheros .txt o XML.

A continuación, se enumeran las funcionalidades que **NO** serán requeridas en esta prueba:

- **Procesos de metadato:**
  - Análisis de audio: Reconocimiento de identidad a partir de la voz
  - Análisis de imagen: Reconocimiento de identidades
- **Interacciones entre distintos procesados:**
  - Entre los distintos procesados de vídeo: la información de reconocimiento de identidades en la “descripción de imagen”.
  - Con los tesauros unificados de RTVE
  - Con las bases de datos de RTVE VIRIA
- **Interfaz de usuario:**
  - El contenido enviado a RTVE ha de quedar marcado como tal.
  - Dimensionamiento del sistema en su arquitectura final, de tal manera que se puedan reprocesar contenidos sin que afecte al normal funcionamiento del servicio.
  - Disponer de un cuadro de mando configurable
  - Posibilidad de exportar datos a una hoja de cálculo

La convocatoria se enviará a los oferentes a través de correo electrónico con un plazo mínimo de cinco días naturales hasta la fecha de celebración de la prueba y deberán responder a ella en el plazo de tres días naturales. Una vez cumplido este plazo, si el oferente no ha respondido, la oferta quedará desestimada. Para ello, los oferentes deberán indicar en su oferta al menos **dos cuentas de correo electrónico de las personas de contacto**.

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

### *2.5.2 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 y 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 Pliego de Cláusulas Administrativas Particulares.

### *2.5.3 Gestión del proyecto.*

#### *a) Composición del equipo de trabajo*

Para el seguimiento del proyecto, se establecerá un equipo de trabajo permanente en el que estará personal del Archivo de RTVE, y 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 primeros meses de contrato que podrán pasar a periodicidad quincenal si la buena marcha del proyecto así lo permite. En todos los casos, la empresa adjudicataria estará encomendada a levantar acta de reunión y hacérsela llegar a todo el grupo de trabajo.

En principio, las reuniones de seguimiento serán en formato videoconferencia, no obstante, al menos una vez al trimestre y siempre que el sistema no esté funcionando acorde con lo solicitado en este documento, RTVE podrá solicitar reuniones presenciales.

#### *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.

## **2.6 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 de toda la información necesaria para evaluar la idoneidad 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 metadatado automático de los contenidos, diferenciando para audio y video/si no se tratara del mismo.
  - Relación de proveedores de los servicios de IA con los que el oferente pueda trabajar en su plataforma.
  - Explicación detallada de la forma en la que se hace el análisis de identidad y si se utilizan explícitamente marcadores biométricos o por el contrario se basa en vectores numéricos o algoritmos de análisis de elementos técnicos del contenido multimodal.
  - 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 metadatado.
    - Plataforma lista para procesos de movimiento de archivos.
  - Personas de contacto para la prueba técnica.

## 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.
- **Transcripción en otros idiomas:** Se utilizarán las mismas métricas que para la transcripción en español.
- **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.

#### **Video.**

Se emplearán procesos de reconocimiento de identidad/ voz no unívocos a través de los cuales, se pueda asociar cierta información con una persona, pero sin poder determinar con certeza e individualidad quién es. Es decir, que dentro de los distintos rasgos necesarios que se obtengan para su reconocimiento, no se obtengan todos los datos necesarios con el fin de reducir el conjunto de posibles identificados, pero no llegar a una única persona de forma concluyente, sin llegar a determinar de manera directa e inequívoca su identidad, aunque sí reduciendo el conjunto de posibles individuos a los que podría corresponder.

- **Reconocimiento de personas en la escena.** 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”.