

# **Especificaciones técnicas**

---

## **Metadato automático de contenidos del fondo documental de RTVE**

---



## Contenido

<b>1</b>	<b><i>Introducción</i></b> .....	<b>3</b>
<b>2</b>	<b><i>Objeto del contrato</i></b> .....	<b>3</b>
<b>3</b>	<b><i>Descripción de los requisitos técnicos y de calidad</i></b> .....	<b>4</b>
3.1	Alcance del proyecto. ....	4
3.2	Herramienta de control de calidad del metadato .....6	6
3.3	Fases y tiempos. ....6	6
3.4	Flujo de intercambio de información.....7	7
3.5	Tipo de servicio. ....8	8
<b>4</b>	<b><i>Control de calidad</i></b> .....	<b>8</b>
4.1	Calidad del procedimiento de intercambio de información. ....8	8
4.2	Calidad del metadato automático.....9	9
4.2.1	Control Metadato en Fase 1: Configuración. .... 9	9
4.2.2	Control Metadato en Fase 2: servicio..... 11	11
<b>5</b>	<b><i>Formación</i></b> .....	<b>11</b>
<b>6</b>	<b><i>Otros</i></b> .....	<b>12</b>
6.1	Validación de niveles de calidad mínimos.....12	12
6.2	Soporte.....12	12
6.3	Horas de servicios profesionales. ....12	12
<b>7</b>	<b><i>Gestión del proyecto</i></b> . ....	<b>12</b>
7.1	Composición del equipo de trabajo.....12	12
7.2	Reuniones de seguimiento. ....13	13
7.3	Entregables. ....13	13
<b>8</b>	<b><i>Formato de la Oferta</i></b> . ....	<b>13</b>
	<b><i>Anexo 1- Pruebas técnicas</i></b> .....	<b>15</b>
	<b><i>Anexo 2- Niveles de calidad en el metadato: métricas</i></b> .....	<b>17</b>
	Audio.....17	17
	Video. ....19	19

## 1 Introducción.

El Fondo Documental de RTVE gestiona millones de documentos audiovisuales, sonoros y textuales que abarcan toda la historia de la Radio Televisión Pública, muchos de los cuales no han sido catalogados por un documentalista o han sido tratados de forma que ahora resulta insuficiente para su recuperación y reutilización.

Dado que las herramientas basadas en el uso intensivo de algoritmos de Inteligencia Artificial, fundamentalmente las desarrolladas sobre redes neuronales, han experimentado una espectacular mejora en los últimos años, desde RTVE se plantea su uso para el metadado automático de los contenidos del archivo que actualmente no están suficientemente descritos.

## 2 Objeto del contrato

El objeto del contrato es la implementación de un servicio de metadado automático, aplicable a las colecciones de video gestionados por el Fondo Documental de RTVE.

La empresa adjudicataria deberá poner a disposición de RTVE un sistema que se integre en los actuales procesos de trabajo en el Archivo y, que permita generar de forma automática metadatos para mejorar la recuperación y por tanto la reutilización del material de archivo por parte de las distintas áreas de la Corporación.

El tipo de metadatos y los requerimientos técnicos y de calidad que se habrán de cumplir se explicitan en puntos posteriores de este mismo documento.

Con estas premisas, el objeto de este expediente es el metadado automático de aproximadamente 11.000 horas de contenidos de video, en dos fases:

### 1. Fase 1:

- Durante un periodo de tiempo de 4 meses se procederá a la configuración, aprendizaje del sistema y puesta en marcha del servicio.
- En este periodo, todos los resultados obtenidos automáticamente por el servicio contratado serán supervisados, controlados y auditados por personal de RTVE antes de su incorporación a los respectivos campos de las bases audiovisuales del gestor documental de RTVE ARCA.
- La empresa adjudicataria recibirá la información resultante de la revisión realizada por personal de RTVE para que pueda afinar su sistema, de modo que se cumplan los requerimientos de calidad exigidos por RTVE para el servicio.
- En este periodo se tratarán del orden de 160 horas en las condiciones expuestas más adelante.
- La empresa adjudicataria facilitará una interfaz para realizar el control de calidad antes mencionado. Dicha interfaz quedará lista para su uso de forma fiable y estable.
- Todos los desarrollos necesarios para que el flujo de trabajo pase a una fase estacionaria estarán concluidos antes del final de esta fase 1.
- Si una vez alcanzado el final de esta fase 1 (mes 4), el sistema no ha alcanzado los niveles de calidad establecidos, RTVE valorará la ampliación de este periodo, siempre que la Dirección del Proyecto considere que esta ampliación puede contribuir a alcanzar dichos niveles de calidad.

### 2. Fase 2:

Se trata del servicio propiamente dicho. Esta fase comenzará una vez finalizada la fase 1 y tendrá una duración de 12 meses, durante los cuales se procederá al metadado automático de las aproximadamente 11.000 horas de contenidos, a razón de unas 280 horas semanales, en las condiciones que más adelante se explicitan.

Durante el periodo de vigencia del contrato, la empresa adjudicataria hará entrega a RTVE del metadado del contenido audiovisual facilitado, en el formato definido, así como los archivos originales procedentes del tratamiento de dichos recursos por parte de terceros proveedores si los hubiera.

### **3 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 metadado de archivos audiovisuales, utilizando una herramienta que permita a RTVE situarse en la vanguardia en el uso de estas tecnologías.

El servicio consistirá en la recogida de contenidos para catalogación previamente seleccionados y su tratamiento mediante un sistema adaptado a los procedimientos definidos por RTVE al adjudicatario.

Por lo tanto, la empresa adjudicataria, para la consecución de los objetivos de este pliego, puede utilizar herramientas propias de metadado o integrar y adaptar herramientas de proveedores de servicios externos, sean éstos generalistas (Amazon, Google, Microsoft, etc.) o específicos para el sector audiovisual.

De cualquier modo, se ha de permitir la posibilidad de la coexistencia en la herramienta de soluciones de metadado de diversos proveedores, de tal manera que, en la medida de lo posible, cada uno de los elementos del metadado que enunciaremos después sea tratado por aquel proveedor de servicios que tenga la superioridad tecnológica en cada campo concreto.

La empresa adjudicataria ha de poder actuar como un integrador de estos servicios y, ofertará una herramienta que ejecutará sobre dichos contenidos los procesos necesarios y suficientes, para que los resultados obtenidos cumplan con las especificaciones planteadas, aplicando sus propios algoritmos de machine learning para adaptar los resultados obtenidos a partir de las herramientas de metadado automático de material audiovisual, a las necesidades de RTVE.

El adjudicatario incluirá una capa de Inteligencia Artificial, utilizando redes neuronales, o cualquier otra tecnología que será entrenada con material de RTVE, que permita que los resultados obtenidos durante la vigencia del contrato mejoren en lo que se refiere a la adecuación a las necesidades del Archivo de RTVE. Con el fin de ayudar a la empresa adjudicataria a alcanzar dicho objetivo RTVE pondrá a su disposición la Base de Datos RTVE2020 Database, (<http://catedrartve.unizar.es/rtvedatabase.html>).

En todas las fases del proyecto, los resultados del metadado deberán ser enviados a RTVE ajustándose a los requisitos de formato necesarios para su incorporación al Fondo Documental de la Corporación.

Los archivos que contengan el metadado generado automáticamente incluirán un identificador de que la media a la que se refiere ha sido tratada automáticamente.

#### **3.1 Alcance del proyecto.**

El contenido sobre el que se aplicará el proceso de metadado objeto de este pliego, consta de aproximadamente 11.000 horas de contenidos de vídeo.

Dado que el coste de los servicios es diferente para el tratamiento del audio y del video, se diferencian dos bloques para el tratamiento de estos contenidos.

Sobre cada uno de los contenidos se aplicarán procesos de transcripción de voz a texto, reconocimiento de entidades y clasificación automática. De forma adicional, y sobre aproximadamente un 20% del total de los contenidos, se aplicarán procesos de análisis de la imagen tal y como se especifica más adelante.

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

#### 1. Tratamiento de Audio (sobre todos los contenidos de video)

En este ámbito, los procesos mínimos requeridos serán:

- Transcripción de audios 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.

#### 2. Tratamiento de imagen (sobre el 20% de los contenidos de video)

Los procesos mínimos requeridos serán:

- Reconocimiento facial entendido como el reconocimiento de rostros y sus atributos.
- Reconocimiento de identidad o de personalidades.
- Extracción de caracteres alfanuméricos (OCR) en imagen en forma de rótulos y/o subtítulos.
- Reconocimiento de logos y marcas.
- Identificación de escenas.
- Reconocimiento de objetos.

El suministrador deberá facilitar un mecanismo por el que RTVE podrá diferenciar los contenidos a los que se aplicará el metadato del vídeo.

Los procesos arriba explicitados son los procesos mínimos exigidos. No obstante, si el proveedor de servicios ofreciera otro tipo de datos, todos ellos deberán de ser facilitados también a RTVE en las condiciones que se especifiquen. Además, si en el tiempo de vigencia del contrato hubiera nuevos procesos que alcanzaran una madurez suficiente, RTVE podrá solicitar la inclusión de los mismos en el alcance del proyecto. El formato de archivo de intercambio con RTVE incluirá todos los metadatos obtenidos.

Todos los archivos serán devueltos a RTVE asociados al ítem al que se refieran y en el formato definido por RTVE.

Además, en todos los casos, tanto para los elementos de metadato considerados imprescindibles, como para el resto de elementos obtenidos, se proporcionará el código de tiempo correspondiente al momento en que aparece cada elemento detectado en el archivo original.

Los resultados obtenidos han de darse en castellano. Para la clasificación temática de los contenidos audiovisuales se utilizará un léxico propio de RTVE que será facilitado al adjudicatario del proyecto, 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.

Así, el adjudicatario ha de devolver, para cada contenido audiovisual, uno o varios archivos con el formato y las especificaciones definidas por RTVE y, además, los ficheros originales generados por parte de los proveedores de servicios a los que haya acudido.

### 3.2 Herramienta de control de calidad del metadato

---

Además de los metadatos obtenidos para cada contenido, y como parte de los trabajos de la Fase 1 del proyecto, 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.

Esta herramienta será proporcionada como SaaS y se accederá a ella mediante un cliente web que permitirá como mínimo 4 accesos simultáneos. Contará al menos con las siguientes características:

- Una interfaz que ha de permitir el acceso a cualquiera de los archivos de video sobre los que se ha realizado el metadato a elección por parte del personal de RTVE.
- Para el contenido audiovisual elegido se han de poder 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 como retroalimentación para la mejora del servicio.
- Incluirá una funcionalidad que permita calcular el WER de un fragmento, seleccionado de la transcripción mediante la comparación de ésta con una transcripción exacta.
- Se ha de incluir la posibilidad de corregir 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.
- Dispondrá de un botón de aceptar/rechazar. En caso de aceptación se desencadenará el proceso para el envío de los metadatos a RTVE. Se ha de permitir la aceptación de los archivos analizados de forma individual o total para el periodo analizado.
- 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.

### 3.3 Fases y tiempos.

---

Una vez firmado el contrato, como se ha mencionado más arriba, el proceso de metadato automático se aplicará sobre aproximadamente 11.000 horas de contenidos de video, en dos fases:

1. Fase 1: durante un periodo de tiempo de 4 meses se procederá a la configuración, aprendizaje y puesta en marcha del servicio. Al final de este período se habrán tenido que alcanzar los niveles de calidad que luego se explicitan para cada tipo de programa y servicio y todo el flujo de información deberá funcionar adecuadamente. Durante este periodo se tratarán 160 horas de contenidos.
2. Fase 2: Se trata del servicio propiamente dicho. Esta fase tendrá una duración de 12 meses.

### 3.4 Flujo de intercambio de información.

El procedimiento de intercambio de archivos será definido en la Fase 1, pero en términos generales se caracterizará por:

- Desde el momento en que el material está disponible en el repositorio de salida de RTVE hasta que esté listo para validar han de pasar, como máximo, 24 horas.
- El flujo del proceso empezará en un repositorio temporal de salida, desde donde la empresa adjudicataria recogerá los ficheros de video para su tratamiento.
- El formato de los ficheros de salida desde RTVE será en forma de archivos MP4 a 1,5 Mbps, en baja resolución en calidad H264 (generalmente 352x288) y 8 canales de audio en castellano (AAC estéreo 44kbps cada canal), de los que habitualmente, tienen información únicamente los 4 primeros.
- El adjudicatario será el encargado de garantizar el espacio de almacenamiento en la nube, como mínimo de 1TB y de proporcionar la línea con RTVE con el suficiente ancho de banda (1 Gbps de subida como mínimo). Se ha de garantizar el flujo adecuado tanto en la fase de salida de los archivos desde RTVE, como en la vuelta de los ficheros resultantes.
- La empresa enviará los ficheros de video para su tratamiento por parte del proveedor de servicios de Inteligencia Artificial con el que trabaje y, tratará el resultado obtenido con sus propias herramientas para adaptarlo a las necesidades de RTVE ya explicitadas en distintos apartados de este documento.
- Mediante la herramienta de control proporcionada por el adjudicatario en el marco del contrato (definida en el punto 3.2), desde RTVE se procederá a controlar diariamente la calidad del metadato obtenido. Durante la Fase 1 del proyecto se procederá a analizar todo el contenido tratado. En la Fase 2, el control se realizará de forma cualitativa sobre fragmentos aleatoriamente elegidos de la media tratada cada día.
- Una vez haya entrado el servicio en fase estacionaria (Fase 2), en caso de darse por bueno el metadato de las catas realizadas, RTVE validará la totalidad de la media tratada ese día, lo que dará lugar al envío a RTVE de los metadatos validados. De no cumplir con los criterios de calidad mínimos, que serán los que deben haberse alcanzado al final del período de prueba, RTVE no validará el contenido.
- El adjudicatario recibirá las indicaciones pertinentes y 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.
- El flujo ha de ser continuo y el adjudicatario ha de contemplar todos los mecanismos necesarios para evitar los cuellos de botella. El material no validado por RTVE ha de ser recatalogado según lo explicitado en el apartado 4.2.2.
- El archivo de vuelta tendrá el formato y las características que se expliciten por parte de RTVE en la fase de configuración y ajustes. En él constarán, asociados al identificador de la media de origen, los metadatos obtenidos, siempre con código de tiempo.
- Además, se incluirán debidamente identificados y clasificados los archivos originales proporcionados por los proveedores de Inteligencia Artificial.
- Todo intercambio de información digital entre RTVE y el adjudicatario será mediante protocolo estándar de Internet, no dependiente de los suministradores de servicios.
- Para el flujo de entrada no será necesario el retorno de la media, que deberá ser destruida después de su tratamiento por parte de la empresa adjudicataria y de la validación por parte de RTVE del resultado obtenido mediante la herramienta de control descrita en el punto 3.2. El adjudicatario certificará con una periodicidad mensual 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.

Todo el flujo de trabajo descrito, deberá ser comprobado y validado por los responsables de RTVE designados. No se dará por finalizada la fase de configuración y puesta en marcha hasta que todo el sistema funcione, de acuerdo a especificaciones, de una forma fluida y armónica.

No se autoriza el uso del material audiovisual facilitado por RTVE para ningún uso distinto a este proyecto.

### **3.5 Tipo de Servicio.**

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

## **4 Control de 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.

A la finalización del periodo de configuración y puesta en marcha, se realizará una prueba de aceptación del sistema en la que se demostrará que el mismo cumple con los estándares definidos. La superación de esta prueba dará lugar al paso a la fase de servicio. Esta prueba se recogerá en un entregable que aporte información acerca del tipo de test realizado y los resultados obtenidos.

### **4.1 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 3.4 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, estarán de vuelta en el plazo mencionado anteriormente en el punto 3.4.
3. RTVE recibe, para cada archivo audiovisual proporcionado para su tratamiento, un archivo/archivos con la estructura definida por RTVE en la Fase 1, identificando la media de la que provienen. También se entregarán todos los archivos originales proporcionados por terceros, provenientes del tratamiento del referido archivo audiovisual, debidamente identificados, de tal manera que permita su relación unívoca con el archivo de origen.
4. El funcionamiento de la herramienta para la validación de los metadatos y control de calidad se ajusta a las especificaciones expuestas en el punto 3.2.
5. No existen cuellos de botella debido a malos funcionamientos en la red que interfieran en el flujo arriba mencionado.

## 4.2 Calidad del metadato automático

En lo referente a la calidad del metadato propiamente dicho, existirán dos niveles de control diferentes en las dos fases en las que se divide el proyecto, según se explica en los apartados siguientes.

Del mismo modo los parámetros de calidad exigibles serán distintos para los procesos sobre el audio y sobre el video y en función de la clase de contenidos audiovisuales objeto de análisis. Las 11.000 horas de contenidos audiovisuales están formadas por brutos o material original de rodaje, programas editados y material telecinado. A continuación, se definen las características de estos grupos.

- **Material original:** este grupo está formado material original de rodaje preservados tal y como fue grabado, sin manipular, editar o posproducir. Los materiales son en su mayoría entrevistas en estudio o en interior que responden a un modelo pregunta respuesta o están articulados en forma de diálogo entre dos o más personas y, en las que predomina el habla espontánea. La temática de las entrevistas es variada y la calidad del audio es buena.
- **Material emitido:** Este grupo está compuesto por programas emitidos, generalmente magazines de actualidad informativa, pero también especializados en deportes, tauromaquia, cine, política, cultura o crónica social. Pueden contener entrevistas en el estudio, conexiones en directo y debates. Se incluye en esta categoría los premontajes, materiales resultantes de la producción de noticias y reportajes incluidos en los programas con pistas de audio separadas y sin posproducción.
- **Material telecinado:** Este grupo lo forman tanto programas emitidos como material original rodado en soporte fílmico y transferido luego a soporte video en la década de los 2000. Se trata de contenidos generalmente en blanco y negro con una calidad de audio desigual.

### 4.2.1 Control Metadato en Fase 1: Configuración.

Durante la fase de configuración se auditará toda la media metadada. El control consistirá en la revisión por parte de profesionales de RTVE de la media tratada y el retorno continuo al adjudicatario para el entrenamiento del sistema.

Al final de la Fase 1, se han de haber alcanzado de forma mantenida en el tiempo los niveles de calidad, explicitados en la Tabla 1 para cada funcionalidad y tipo de media. Estos niveles serán comprobados en una prueba de aceptación. 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 2.

	Material emitido	Material original	Material telecinado
<b>Tratamiento del audio</b>			
Transcripción de audio a texto	≤20% (WER)	≤10% (WER)	≤25% (WER)
Puntuación y Capitalización	≤25% (WER)	≤15% (WER)	≤30% (WER)
Segmentación por hablantes	≤25% (DER)	≤15% (DER)	≤25% (DER)
Reconocimiento de entidades	≤35% (SER)	≤35% (SER)	≤35% (SER)
Extracción palabras clave	Requerido	Requerido	Requerido
Clasificación contenido	≥75% (Precisión)	≥75% (Precisión)	≥75% (Precisión)
<b>Tratamiento del vídeo</b>			
Reconocimiento facial	<45% (DER)	<35% (DER)	<40% (DER)
Reconocimiento de identidad	Requerido	Requerido	Requerido
OCR	≤7% (WER)	No aplica	≤7% (WER)
Marcas y logos	Requerido	Requerido	Requerido
Identificación escenas	Requerido	Requerido	Requerido
Reconocimiento objetos	Requerido	Requerido	Requerido

*Tabla 1. Niveles de calidad exigidos*

“Requerido” implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo.

#### 4.2.2 Control Metadatado en Fase 2: Servicio.

En la fase de servicio, los controles de calidad se harán, por parte de personal de RTVE, sobre el material ya tratado por el integrador a través de la herramienta definida en el punto 4 y con las condiciones explicitadas en el punto 4.2.

Las mediciones se realizarán con carácter diario, auditándose una muestra de la media tratada durante el día anterior. Si el material no es validado por RTVE, de acuerdo a los criterios establecidos en la Tabla 1, ha de ser recatalogado correctamente en un máximo de 48 horas.

En caso de no validarse el tratamiento realizado durante 2 días seguidos, 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 de la Fase 2. Las horas que se hubieran dejado de procesar durante el periodo de parada, tendrán que poder procesarse antes de 30 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.

No obstante, y ante la apreciación de un deterioro de la calidad del metadatado suministrado a lo largo de la Fase 2, en cualquier momento RTVE podrá exigir la realización de una nueva medición de la calidad en los términos especificados para la Fase 1 (apartado 4.2.1) en un plazo no superior a 48h desde la comunicación del requerimiento.

## 5 Formación.

Como se ha mencionado, además de los metadatos obtenidos para cada archivo, y como parte de los trabajos de la Fase 1 del proyecto, el adjudicatario proporcionará la herramienta necesaria para que personal de RTVE asignado a tal propósito, valide desde su puesto de trabajo los resultados y acepte o rechace los mismos.

La empresa adjudicataria proporcionará la documentación necesaria para que el personal de RTVE pueda manejar dicha herramienta y proporcionará formación en su uso para las personas que RTVE considere oportunas.

La formación deberá impartirse en español, será de manera presencial en las instalaciones de RTVE en Madrid y deberá contar al menos con:

- Cursos de operación.

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 puedan realizar su trabajo sin dificultad alguna.

La formación se referirá tanto al uso de la herramienta de control de calidad y metadatado automático como las bases teóricas sobre las que se sustenta.

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

- Curso técnico.

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 y su funcionamiento. Se deberá proporcionar documentación de apoyo en formato electrónico o acceso web.

## 6 Otros.

### 6.1 Validación de niveles de calidad mínimos.

Todas las ofertas que sean admitidas para su evaluación técnica, 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 el Anexo 1, cumpliendo los requisitos de calidad del metadato allí expuestos. El no cumplimiento implicará la desestimación de la oferta.

La convocatoria llegará a cada uno de los ofertantes a través del departamento de compras y deberán responder a ella en el plazo de 3 días laborables. Una vez cumplido este plazo si el oferente no ha respondido o no se ha presentado, la oferta quedará desestimada.

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

### 6.3 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.

## 7 Gestión del proyecto.

### 7.1 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 cuantas personas considere oportunas.

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.

## 7.2 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.

## 7.3 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.

Se presentará otro informe al final de la primera fase del proyecto y cuando éste se dé por terminado.

## 8 Formato de la Oferta técnica

La oferta técnica presentada 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.
- 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 y video si no se tratara del mismo.

Se especificará, asimismo, para la primera parte del proyecto las fases y tiempos estimados para la entrada en servicio de las diferentes funcionalidades, especificando, entre otros, los tiempos estimados de:

- Formalización del equipo de trabajo.
- Envío a proveedores externos de los primeros archivos de trabajo.
- Procesos de movimiento de archivos.

- 
- Habilitación de la herramienta de control de calidad mencionada en el apartado 3.2.

## Anexo 1- Pruebas técnicas

RTVE realizará junto con cada licitador una batería de pruebas en la que se verificará el cumplimiento de todas las funcionalidades y requerimientos especificados más adelante en este apartado.

Con la finalidad de ejecutar estas pruebas, RTVE lanzará un requerimiento a todos los licitadores que hayan presentado oferta, solicitando que ponga a disposición los recursos técnicos y humanos que resulten necesarios para completar esta fase de validación.

Los licitadores tendrán la obligación de acudir a las instalaciones de RTVE en Prado del Rey para realizar una prueba sobre ficheros de video en la fecha que RTVE se les solicite, tal y como se establece en el apartado 6.1. Para ello, deben indicar en sus ofertas técnicas la información de contacto más conveniente a tal efecto.

Se han de cumplir los siguientes criterios de calidad mínimos:

	Material emitido	Material original	Material telecinado
<b>Tratamiento del audio</b>			
Transcripción de audio a texto,	≤25% (WER)	≤13% (WER)	≤30% (WER)
Puntuación y Capitalización	Requerido	Requerido	Requerido
Segmentación por hablantes	≤30% (DER)	≤20% (DER)	≤30% (DER)
Reconocimiento de entidades	≤40% (SER)	≤40% (SER)	≤40% (SER)
Extracción palabras clave	Requerido	Requerido	Requerido
Clasificación contenido	Requerido	Requerido	Requerido
<b>Tratamiento del vídeo</b>			
Reconocimiento facial	≤50% (DER)	≤40% (DER)	≤45% (DER)
Reconocimiento de identidad	Requerido	Requerido	Requerido
OCR	≤10% (WER)	No aplica	≤10% (WER)
Marcas y logos	Requerido	Requerido	Requerido
Identificación escenas	Requerido	Requerido	Requerido
Reconocimiento objetos	Requerido	Requerido	Requerido

Tabla 2. Niveles de calidad para la prueba técnica

---

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 video, que tendrán una duración no superior a 60 minutos y que estarán convenientemente identificados y clasificados, de acuerdo a lo establecido en el apartado 4.2.

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

## Anexo 2- 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 tanto en las pruebas de los sistemas ofertados que se solicitarán a todos los oferentes, como el de la solución adjudicada al final de la Fase 1, están recogidos en el Anexo 1 y en el apartado 4.2.1 respectivamente.

### 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).

- **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, comparecencias, 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.

Para medir el rendimiento del sistema se tendrá en cuenta que éste sea capaz de transcribir, reconocer y clasificar correctamente las entidades nombradas y para ello se comparará el resultado del reconocimiento de entidades a partir del texto obtenido por el reconocedor automático con la transcripción real y convenientemente anotada.

La fórmula utilizada para medir la calidad será el Slot Error Rate (SER en su acrónimo en inglés) que responde a la siguiente fórmula matemática:

$$SER = \frac{S + D + I}{C + D + S} = \frac{S + D + I}{R}$$

Donde:

- S es el número de sustituciones o entidades que se identificaron correctamente, pero fueron erróneamente clasificadas.
- D es el número total de entidades no detectadas.
- I es el número de inserciones o entidades reconocidas de forma incorrecta.
- C es el número total de anotaciones correctas proporcionadas por el sistema.
- R es el número total de entidades en la referencia.

Finalizada la Fase 1 del proyecto el cálculo del SER se realizará asignando pesos específicos a los errores de borrado, inserción y sustitución de acuerdo con la siguiente fórmula.

$$SER = \frac{S + 2D + 1,5I}{R}$$

Donde:

- S es el número de sustituciones entendidas como entidades que se identificaron correctamente, pero fueron erróneamente clasificadas o entidades cuyos límites no se definieron correctamente.
- D es el número total de entidades no detectadas.
- I es el número total de inserciones e entidades reconocidas de forma incorrecta.
- **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**: en este caso, la exactitud se entiende como la proporción entre los contenidos correctamente clasificados y el número total de contenidos a clasificar.

$$Exactitud = \frac{C}{R}$$

Donde:

- C es el número de contenidos clasificados correctamente.
- R es el total de contenidos clasificados.

Al finalizar la Fase 1 del proyecto se espera que la exactitud en la clasificación del sistema sea igual o superior al 75%.

## 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**, será requerido, lo que implica demostrar que la funcionalidad está disponible pero no se exigirá nivel de calidad mínimo.

- 
- **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.
  - **Identificación de escenas**, 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.