

VALORACIÓN TÉCNICA DEL EXPEDIENTE

S-05850-2023 - RTVEPLAY-i Implantación solución transcripción y catalogación de audio

El objeto del presente documento es realizar la valoración técnica del expediente S/05850/2023. Para realizar esta valoración se va a tener en cuenta lo establecido en el pliego de condiciones técnicas respecto de los requisitos técnicos, así como lo establecido en el pliego de condiciones generales respecto de la valoración del expediente.

El expediente presenta un único lote y, tal y como establece el pliego de condiciones generales (PCG, en adelante) y la valoración técnica se hará según lo establecido en el Anexo II páginas 19 y 20 del PCG, de tal manera que *las propuestas de los licitadores se van a valorar en base a criterios sujetos a juicio de valor con hasta 45 puntos, distribuidos de la siguiente forma:*

- Arquitectura Global de la solución: 30 puntos.
- Plan de proyecto: 7 puntos.
- Equipo de servicios profesionales y modelo de soporte: 8 puntos.

Para la valoración de cada uno de los epígrafes se ha de tener en cuenta:

- **Arquitectura Global de la solución:** se valorará *la mejor solución arquitectural, valorándose, entre otros: la arquitectura general de la solución, el mecanismo de captación de la señal, el decalaje entre audio e información (ver PCT), el mecanismo (algoritmos, servicios, etc.) de extracción y transcripción de audio, los mecanismos de redundancia, la propuesta de monitorización y de gestión de la calidad, el proceso de mejora continua, la riqueza del API de consulta y su dimensionamiento y escalabilidad. Además, también se tendrá en cuenta la arquitectura de seguridad que se propone y el detalle de la arquitectura técnica y de cómputo presentada. Finalmente, también se tendrá en cuenta el riesgo de la solución planteada, y, especialmente, el **alcance** de la solución.*
- **Plan de proyecto:** *Se valorará el plan de proyecto, otorgándose la mayor puntuación al más coherente, robusto, viable y la reducción del tiempo que requiere para su implantación. Además, se valorará la idoneidad del equipo de trabajo de implantación, especialmente la capacitación de cada recurso a las tareas que han de ejecutar. Es por ello que ha de incluirse en los cronogramas los recursos que van a participar, junto a sus currículums con información suficiente para que permita la valoración de la capacitación. Será penalizada la falta de información específica.*
- **Equipo servicios profesionales y modelo de soporte.** *Se valorará la idoneidad del equipo de trabajo propuesto para la ejecución de los servicios profesionales, especialmente la capacitación de cada recurso a las tareas que han de ejecutar. Es por ello que ha de incluirse en los cronogramas los recursos que van a participar, junto a sus currículums con información suficiente para que permita la valoración de la capacitación. Será penalizada la falta de información específica. Se valorará el modelo de soporte, los tiempos de respuesta, así como los ANS que el licitador indique en su propuesta.*

Finalmente, y como se indica en el PCG, página 20, *para cada epígrafe, se otorgará el total de puntos a la mejor propuesta, mientras que el resto serán valoradas en función de lo ofertado por el mejor. Se*

valorará el detalle, pertinencia, la coherencia, la aplicabilidad al caso de RTVE, el cumplimiento a lo establecido al Pliego de condiciones técnicas, y la precisión de la información aportada.

Se penalizará la falta de concreción, de información específica y de detalle. Además, se penalizará especialmente la falta de coherencia de la propuesta, incluso, la incoherencia entre los epígrafes anteriores, así como la presentación de documentación genérica no concretada para el caso de RTVE. En caso que la propuesta técnica contravenga cualquiera de las especificaciones del PCT será valorada con cero puntos. También será valorada negativamente las propuestas no creíbles.

Finalmente, se descartará toda propuesta que utilice tecnologías de terceros (Microsoft, Google, Amazon, etc.) que impliquen que el audio podrá ser utilizado por esos terceros para otras finalidades diferentes (entrenamientos de sus herramientas a posteriori, etc.). Será descartada toda propuesta que emplea servicios de tercero y no se indique de forma expresa en su oferta técnica que esos servicios respetan lo anteriormente indicado. La no inclusión de esta información pone en riesgo los activos digitales de RTVE y supone la substracción de la propiedad intelectual de RTVE.

VALORACIÓN TÉCNICA

Para el expediente se han recibido un total de cinco propuestas técnicas:

- Aeternal Mentis
- Cinfo
- Cires 21
- Crosspoint
- Rebold

Se procede a analizar cada una de ellas a continuación.

AETERNAL MENTIS

Tras la lectura de la propuesta, ésta no es completa ni muestra información suficiente para comprender la solución que propone o los modelos de procesamiento de la media. Por ejemplo, en el caso de los directos, el licitador únicamente indica que:

Transcripciones en directo. En este punto Aeternal Mentis, a través de la URL de cada una de las emisiones transcribirá las emisiones que en ese momento se estén produciendo en directo.

Aeternal Mentis, a través de cada una de las URLs indicadas mediante su correspondiente API, sabrá qué emisiones hay en directo, así como las que se emitirán próximamente.

Para el caso de los ficheros indica que los ficheros subirán por SFTP o similar, que no se almacenarán, solo transcribirán, para a continuación indicar los siguiente:

A modo de ejemplo y teniendo en cuenta algunos de los enlaces donde se muestran los diferentes programas emitidos:

<https://www.rtve.es/play/audios/gente-despierta/tercera-hora-22-11-23/7016442/>

Los ficheros se podrían catalogar de la siguiente forma:

RTVE_RNE_gentedespierta_tercerahora221123_7016442_metadato1_metadato2.txt

Donde <metadato1>, <metadato2>, etc. podrían ser aquellas variables que se hayan acordado para incluir dentro de cada uno de los ficheros transcritos y que permitan una búsqueda más ampliada. Por ejemplo: DEPORTES / INVESTIGACIÓN / LIGA / etc.

En la propuesta, el licitador hace una referencia a variables (que resultan ser metadatos) que no define, si no que indica que hay que “acordar”. En la licitación objeto de valoración, se solicita una solución completa, definida, determinada y concreta. Dicha solución va a ser evaluada mediante criterios sujetos a juicio de valor. En la evaluación de la solución propuesta por AETERNAL MENTIS, no se da respuesta a las siguientes cuestiones:

1. ¿Cómo se procesan los streams?
2. ¿Cómo se metadatan?
3. ¿Qué algoritmos, tecnologías, bases técnicas, etc., va a utilizar el licitador para generar la información?

Respecto del API de consulta, el PCT, página 5, indica que *la transcripción, así como toda la metadata asociada/generada ha de estar disponible a través de un api enriquecida para ser empleada en tiempo de distribución y que permita obtener información a partir del canal y de dos parámetros para indicar el start y end time, o bien a partir del código SGCE/código de emisión (dato que viene dentro del api de parrilla) /idasset (identidad dentro del api rest o del cms de RTVE.es) según el caso.*

Sin embargo, en la propuesta técnica, el licitador indica que el tipo de llamadas será de esta forma:



A modo de ejemplo:

```
'Content-Type': 'application/x-www-form-urlencoded'
```

```
{  
  "canal": " RNE",  
  "programa": "GENTE_DESPIERTA",  
  "tipo": "INFORMATIVO",  
  "fecha_1": "2023_11_22"  
  "fecha_2": "null"  
  "metadato1": "deportes"  
  "metadato2": "liga"  
}
```

Respuesta

```
{  
  "archivo1": "RTVE_RNE_gentedespierta_tercerahora_221123_7016442.txt"  
  "archivo2": "RTVE_RNE_gentedespierta_tercerahora_221123_7016444.txt"  
  "validationResponse": true,  
  "responseItems": [...]  
}
```

La propuesta del licitador para el API es insuficiente por dos motivos: para los directos RTVE quiere indicar un canal y un starttime y un endtime para obtener como respuesta toda la metadata generada en ese canal para ese lapso de tiempo. Sin embargo, el licitador emplea el día (2023_11_22) y el canal. No permite hacer extracciones específicas en rangos de tiempos muy concretos (lo habitual para generar transcripción en tiempo real). Tampoco incorpora en su definición de API endpoints y filtros el idasset o el código SGCE, que es la forma de acceder a un elemento concreto.

Además, de lo anterior, la solución del licitador tiene un error importante: todo lo que ha planteado gira a partir de la generación de un fichero de texto con la transcripción y con otra metadata (que menciona, pero no concreta). Sin embargo, desde RTVE se ha solicitado algo diferente: así en la página 6 del PCT RTVE indica que el API de distribución ha de permitir la extracción de transcripciones parciales. Sin embargo, el licitador distribuye archivos. Esto no solo se puede ver en el ejemplo anterior, sino también en distintas referencias, tales como en la página 12 donde el licitador indica *que puede soportar al menos 100 peticiones/seg para la consulta de los diferentes **archivos***.

Dentro de la solución técnica, el licitador no incluye:

- La arquitectura global de la solución
- El mecanismo de captación de la señal
- El decalaje entre audio e información
- El mecanismo de extracción y transcripción de audio
- Los mecanismos concretos de redundancia
- Detalle de la arquitectura técnica y de cómputo
- Identificación de productos y desarrollos
- Definición detallada de las APIs de consulta y de ingesta.
- Currículums detallados de los recursos.
- Modelo de soporte y ANS

Finalmente, respecto de la prueba técnica establecida en el PCT, el licitador no ha dado respuesta a la comunicación de RTVE para realizarla.

*Debido a que la solución propuesta por el licitador Aeternal Mentis **no está detallada ni completa** como se ha expuesto y, además, el licitador no ha contestado al requerimiento de prueba técnica necesaria para una correcta evaluación de las propuestas, se declara la propuesta como **NO APTA**.*

CINFO

CINFO ha presentado una propuesta compleja que analizamos siguiendo el esquema de propuesta técnica que ha presentado.

De forma general, la arquitectura técnica que presenta el licitador tiene un problema estructural: no diferencia entornos (solo tiene un entorno productivo y no contempla entornos de integración) y tampoco diferencia entre procesos internos o de gestión y procesos externos o de distribución. Cuando se conciben soluciones que van a ser empleadas en procesos de tiempo real, las buenas prácticas indican que hay que separar los elementos que intervienen en la gestión (CMS, APIS de ingesta, etc.) de los elementos que se emplean en distribución (principalmente APIS). Esto se hace para evitar pérdidas de información por efectos bola de nieve, por seguridad en el acceso, diferenciando la DMZ del entorno de distribución, y, finalmente, para evitar que un problema en una capa afecte a las otras.

La arquitectura que plantea el licitador contiene al menos dos elementos compartidos por todas las tipologías de acceso: el NGINX que se emplea como elemento dinamizador y gestor de las peticiones, y, dentro del API, y tal y como indica el licitador (PT, página 38), “... existe un subsistema alojado en la solución para recibir las peticiones a las APIs, **tanto externas como internas** que redirige dichas peticiones a los nodos correspondientes”. A juicio de RTVE, esta estrategia arquitectural no es correcta por lo motivos antes indicados.

Respecto de los modelos de ingesta, el licitador indica que el único modelo para la ingesta de podcast proviene el API de publicación de RTVE (página 12), lo que, siendo una opción válida, resulta que establece una limitación en las capacidades. Otras opciones plantean la ingesta también vía carpetas calientes o FTPs.

Respecto del alcance, la propuesta tiene algunas inconsistencias. Del epígrafe 2.3.1.2 se concluye que la solución genera:

- Transcripción basada en Whisper.
- Extracción de huella de audio.
- Extracción de forma de onda.

En ese epígrafe se indica, además:

Traducción de transcripción a lenguas cooficiales

Una vez se ha producido la transcripción de contenidos, el sistema será capaz de generar una traducción de dicha transcripción automática a las diferentes lenguas cooficiales del Estado (Catalan, Gallego y Euskera) con la posibilidad de añadir otros idiomas al pool. Las traducciones de estos contenidos pueden realizarse on premise si así se requiere.

Esa redacción no se entiende ya que la referencia a hacer on premise no tiene aplicación alguna en el contexto en el que se pretende implantar la solución (esto es, todo es responsabilidad del licitador, y RTVE no provee nada de cómputo ni recursos). Además, esa referencia a “... será capaz ...”, así como con el término “... con la posibilidad ...”, inducen que es algo que podría ofertar.

En la página 30 indica esta capacidad como idea para desarrollos evolutivos con IA.

Generación de subtítulos en varios idiomas

Tras las tareas de transcripción de subtítulos, y teniendo los mismos catalogados y sincronizados con los programas de los que proceden, se puede proceder a realizar una traducción a las lenguas cooficiales del Estado. De esta manera es factible integrar en los contenidos pistas secundarias de subtítulos sincronizadas en los diferentes idiomas. En este caso se plantea dentro del proyecto la colaboración con sistemas de terceros especializados en las lenguas cooficiales del estado español. **cinfo** puede aportar dentro del coste del proyecto las licencias completas para gallego, euskera y catalán.

Si leemos la página 89, el licitador indica:

Propuestas definidas de evolutivo usando IA	no se indican	Aunque las necesidades y evolutivos a aplicar al sistema dependan enteramente de la decisión de RTVE, cinfo propone ya de partida posibles aplicaciones de la IA a los metadatos generados por el sistema, incluyendo traducción automática a otras lenguas, generación de podcasts en otros idiomas diferentes del original, etc.
Aportamos dentro del precio del proyecto las licencias de traducción a lenguas cooficiales	no se pide traducción	Se incluye un sistema adicional para la incorporar la traducción de los textos (transcripciones) a idiomas cooficiales del territorio español especializado en este servicio.

Como se puede ver en la imagen adjunta, el licitador está incluyendo las licencias de traducción (opción de abajo), pero la inclusión de la traducción automática a otras lenguas las sitúa (opción de arriba) como evolutivos que dependerán de decisiones de RTVE. Esto nos lleva a concluir que el licitador no oferta la solución en su instalación, solo las licencias para poder implementarlo a futuro.

Respecto a consideraciones de ciber seguridad e instalación de líneas y ubicación física de la solución, el licitador habla de características técnicas que su solución va a tener, pero no indica dónde, es decir, en qué CPD o ubicación física lo va a instalar. Esto impide determinar si la ubicación es un lugar robusto, seguro y/o homologado según los estándares de mercado.

Respecto de las redundancias, el licitador indica que todo está redundado con una redundancia N+1 o 1+1. Sin embargo, en el epígrafe 2.3.1.3 indica que

Cada nodo tiene la capacidad de procesamiento necesaria para satisfacer las necesidades de extracción y transcripción de audio de 65 canales, por lo que estarán disponibles 3 nodos. Dos de ellos serán los nodos principales que realizarán los mismos trabajos de forma duplicada, mientras que el tercero se destinará a tareas de procesamiento en diferido y tendrá la capacidad de absorber carga de los trabajos en directo en caso de fallo de los nodos principales.

De esa afirmación del licitador deducimos que no está redundado el procesamiento de los ficheros en diferido. Esto puede suponer un problema de robustez de la solución.

Dentro de la propuesta técnica hemos encontrado dos grandes errores de concepto del licitador. El primero de ellos es sobre la comprensión de qué es una emisión. Según indica el licitador, página 22, se entiende un programa como un bloque temporal en la emisión de un canal. Esta afirmación no es correcta ya que, como se puede ver en el JSON que se proveyó en el PCT, un programa puede tener más de un bloque temporal dentro de la parrilla, siendo el conjunto de todos los bloques temporales lo que constituye la emisión completa.

El segundo error de concepto tiene que ver con el concepto de determinación de entidades. La detección de entidades fue clarificada a raíz de una pregunta que se formuló por parte de un licitador. En esa respuesta se indicó que *“la determinación de entidades es un concepto semántico que tiene que ver con el análisis de entidades que aparecen en un texto.”*. Sin embargo, el licitador indica que las entidades están relacionadas con la extracción de artistas. Como ejemplo podemos ver un pantallazo extraído de la prueba técnica donde se puede ver:

Entidades y música detectadas:

Hora (Madrid)	Título	Autor	Álbum
20/12/2023, 21:32:47	Spirit in the Sky	Keiino	Spirit in the Sky
20/12/2023, 21:08:42	greedy	Tate McRae	greedy
20/12/2023, 21:00:53	Rauw Alejandro: Bzrp Music Sessions, Vol. 56	Bizarrap, Rauw Alejandro	Rauw Alejandro: Bzrp Music Sessions, Vol. 56
20/12/2023, 20:59:22	Energía Bacana	Sebastian Yatra	Energía Bacana
20/12/2023, 20:48:58	Reasoning	Oliver Spencer	Coping Skills Vol. 2
20/12/2023, 20:43:19	Shallow Reef	James Copperthwaite	CHILLED BEDS

Así, por ejemplo, el licitador no contempla que el presentador del programa, los invitados, cualquier persona que opine, las ciudades que se mencionan, los organismos que se mencionan, etc., no se extraen como entidades, lo que implica que la solución es insuficiente.

Al hilo del tema semántico, el licitador no es concreto respecto de la metadata que va a generar. No hay listado de qué datos va a generar. Se echa en falta, por ejemplo, que cada emisión la catalogue contra una categorización (IPTC, por ejemplo) para determinar la temática del programa o incluso de cada trozo del programa. Esto daría mucho valor diferencial al procesamiento permitiendo la construcción de nuevas soluciones a partir de estos procesamientos.

Respecto de la herramienta de edición, el licitador indica, página 30:

El editor muestra la forma de onda y permite al operador reproducir el fichero, fijar un punto de inicio y de final y crear como subclip el audio entre ambos puntos pulsado el botón de “descarga”.

Este modelo de subclipping basado es ineficiente y muy antiguo requiriendo de demasiados pasos para el usuario. Los sistemas actuales de subclipping permite la multi selección de trozos dentro de un asset, la fusión dentro de la misma interfaz con otros ficheros, y, finalmente, la creación del clip final uniendo todos los cortes que se hayan solicitado utilizando la ordenación que el redactor decide. Sin embargo, la solución del licitador implica crear como clip cada uno de los trozos, luego irse a la herramienta, seleccionar los trozos y hacer la fusión. Demasiado complicado.

Respecto a la monitorización y gestión de la calidad, indica el licitador que va a realizar catas diarias y mensuales, pero, no indica cuántas.

Monitorización y de gestión de la calidad

Desde el punto de vista de las transcripciones y metadata generadas, se propone la creación de una tarea periódica de monitorización manual de calidad de las transcripciones. Para ello se realizarán diariamente catas manuales sobre los contenidos de las últimas 24 horas, sean estos de tipo directo, o archivo de audio, actuando sobre el sistema usando los módulos descritos previamente ante la presencia de cualquier fallo.

En el proceso de mejora continua, el licitador indica que tendrá un equipo de operación con expertos en metadatos que corregirá y mejorará los resultados, así como añadirá metadatos que servirán de retroalimentación de la IA. Este punto, a juicio de RTVE, es incompleto. Si el licitador toma como fuente de datos whisper large.v2, es decir, un sistema entrenado por la gente de open AI, es decir, un tercero, ¿cómo va a mejorar la IA? ¿Va a hacer reentrenamiento? No se comprende el modelo propuesto por el licitador porque no ha sido desarrollado convenientemente. Se echa en falta la estrategia de mejora de la base tecnológica, e incluso, si plantea algún modelo de sustitución de esta base ante lo cambiante del mercado.

En el punto 4 se desarrolla el API de ingesta y de consulta. Ya se ha comentado los problemas arquitecturales que presenta, por lo que no incidiremos en este aspecto. Sin embargo, desde el punto de vista de la riqueza del API, el licitador no ha incluido la definición de los endpoints. Así, no hay ningún ejemplo de llamadas y de respuestas. Esto impide valorar la calidad de la definición que presenta. Además, en la descripción hace menciones genéricas, tales como “... toda la metadata asociada a un programa ...” lo que dificulta el trabajo de valoración. Como ejemplo añadimos una definición de endpoint:

- **GET**

api-metadata/v1/metadata/by_id/{id}/media_metadata

Devuelve un JSON formado por los diferentes metadatos del contenido identificado por el ID. Se emplea para obtener la información completa de un recurso. En caso de que el metadato contenga información binaria (por ejemplo un PNG) se proporcionará una url para descarga del mismo.

Se echa en falta saber, por ejemplo, a qué se refiere con “... información completa de un recurso ...”, o, en qué caso devuelve un PNG.

El punto 5 elabora el plan de implantación. La solución se implementa como un proyecto donde se han de desarrollar la mayoría de los módulos. Esto es positivo porque de esa forma RTVE poseerá casi todo el código de la solución, pero desde un punto de vista de riesgos, prácticamente se está desarrollando una solución desde cero. Los recursos que asigna el licitador son adecuados para las tareas que ha de desarrollar y el proyecto se planta en 15 semanas.

En el punto 6 se elabora la metodología de trabajo. Según indica el licitador, página 77:

- **Elaboración, presentación y validación del plan de trabajo inicial**

Una vez analizada la situación inicial se procederá a la definición, refinado de los objetivos y desarrollo de la planificación y metodologías necesarias para lograr los objetivos y el alcance del contrato.

El plan de trabajo inicial se presentará para validación de RTVE y se distribuirá a todos los agentes implicados en la ejecución del contrato, de forma que todos sean conocedores del marco de actuación general del servicio que prestará **cinfo**.

Desde RTVE no nos parece razonable que el licitador plantee estas cuestiones para empezar con los trabajos. RTVE contrata un proyecto *llave en mano* que ha de ser desarrollado e implementado asumiendo la ventura y riesgo el licitador. Sin embargo, el licitador plantea que RTVE valide el plan de trabajo del licitador en tiempo de ejecución del contrato, plan de trabajo que es objeto de valoración técnica.

Lo anterior ocurre también con respecto de la gestión de la calidad. El licitador indica que el marco de calidad se acordará entre RTVE y cinfo (página 81 de la PT). Sin embargo, el PCT establece en su sección 4.3, que “*El licitador ha de establecer una estrategia de mejora continua y de gestión de la calidad, que será objeto de valoración*”. Es incongruente que el licitador pretenda que lleguemos a un acuerdo en tiempo de ejecución sobre algo que es objeto de valoración.

Finalmente, el punto 8 describe el modelo de soporte y los ANS. Son correctos.

La prueba técnica solo permite validar algunos cortes desde la emisión, y la generación de subtítulos y de canciones. El canal que se nos ha añadido era R1 del cual sólo se muestra una selección (de audios repetidos).



Nombre	Fecha de inicio de la emisión	Hora de inicio ...	ID del programa	Sgce Programa	Nombre Canal
<input type="checkbox"/> Boletines RNE	20231211	19:00	1750	MKA00079	Radio Nacional
<input type="checkbox"/> Por tres razones	20231211	19:06	119313	MLA10660	Radio Nacional
<input type="checkbox"/> Por tres razones	20231211	19:06	119313	MLA10660	Radio Nacional
<input type="checkbox"/> 24 horas	20231211	20:00	1871	MKA00005	Radio Nacional
<input type="checkbox"/> 24 horas	20231211	20:00	1871	MKA00005	Radio Nacional
<input type="checkbox"/> Radiogaceta de los dep...	20231211	21:00	2071	MKA00078	Radio Nacional
<input type="checkbox"/> Radiogaceta de los dep...	20231211	21:00	2071	MKA00078	Radio Nacional
<input type="checkbox"/> 24 horas	20231211	22:00	1871	MKA00005	Radio Nacional
<input type="checkbox"/> 24 horas	20231211	22:00	1871	MKA00005	Radio Nacional
<input type="checkbox"/> Las mañanas de RNE c...	20231212	08:00	119312	MKA00075	Radio Nacional
<input type="checkbox"/> Las mañanas de RNE c...	20231212	09:00	119312	MKA00075	Radio Nacional
<input type="checkbox"/> Boletines RNE	20231220	15:00	1750	MKA00079	Radio Nacional
<input type="checkbox"/> Boletines RNE	20231220	15:00	1750	MKA00079	Radio Nacional
<input type="checkbox"/> Boletines RNE	20231220	15:00	1750	MKA00079	Radio Nacional
<input type="checkbox"/> Boletines RNE	20231220	15:00	1750	MKA00079	Radio Nacional

Para cada uno de esos audios validamos que se genera una transcripción correcta. No incluye subclipping y el entorno propuesto es muy limitado. La prueba siendo muy limitada no permite clarificar qué otra metadata genera la solución. Funciona lo esencial (transcripciones y subtítulo) pero es muy limitada.

Visto todo lo ya indicado, RTVE valora la propuesta de la siguiente forma:

- **Arquitectura Global de la solución:** 5 puntos.
- **Plan de proyecto:** 4 puntos.
- **Equipo servicios profesionales y modelo de soporte.** 8 puntos.

CIRES 21

La empresa Cires 21 ha presentado una propuesta técnica muy completa y con un amplio desarrollo de los elementos que se piden en la licitación.

Si empezamos por la arquitectura técnica, el licitador plantea la instalación de sus sistemas en un Tier-3 con una conectividad de 1ms a Interxion. A lo largo de la propuesta se van descomponiendo cada una de las capas que presenta la arquitectura, diferenciando los entornos de DMZ de los del Api de distribución. El modelo de seguridad se cimenta en el uso de firewalls Fortinet y VPNs de Fortinet, y se crea una estructura redundada N+1 para los elementos críticos. La instalación cuenta con 16 unidades de rack para cómputo y 5u para elementos de red y seguridad.

Desde un punto de vista lógico, la propuesta desarrolla los modelos de captación, los algoritmos de procesamiento para cada una de los procesamientos que va a realizar, las herramientas, productos y servicios que se emplean o se van a desarrollar. Es decir, la documentación desarrolla de forma exhaustiva el qué (alcance) y el cómo (licenciamiento, plan de proyecto, productos y herramientas, desarrollos, arquitectura y procesos).

Respecto de las tecnologías que integra el licitador destaca:

- WhisperX
- OpenChat-3.5
- Langchain
- Faiss
- Yamnet
- Spacy
- Google API Translate

El alcance de la solución incluye:

- Transcripción de hasta 130 canales de directos.
- Al menos 400 horas de AOD.
- Transcripción y subtitulado.
- Extracción de entidades a partir del texto.
- Catalogación del audio basado en IPTC.
- Extracción del sentimiento.
- Traducción a las lenguas cooficiales utilizando Google API Translate.
- Traducción al inglés y francés utilizando WhisperX.
- Detección de músicas y carátulas.
- Resumen del audio.
- Highlights del audio.
- Forma de onda.
- Subclipping.
- Fusión de clips.
- Guardado de las grabaciones hasta un año.

Comparativamente hablando, esta propuesta supera a todas las demás en cuanto alcance y prestaciones.

El plan de proyecto es exhaustivo y coherente con la propuesta, planteando que la solución estará a los 3 meses desde el inicio del contrato. Los perfiles se adaptan a las tareas y se trata de personal muy experimentado.

El modelo de soporte es adecuado y los SLAS son correctos.

Respecto del modelo de calidad, el licitador indica que:

- Hará catas aleatorias que estarán disponibles para que RTVE pueda acceder y validar si así lo considera.
- Se desarrollarán sistemas de monitorización y vigilancia continua para la detección de casos esquina.
- Contratará a dos personas para revisar y corregir durante toda la vida del contrato. Estos perfiles dedicaran 4 días semanales a validar/corregir, y, el quinto día a generar informes de calidad.

El API que define el licitador es correcto y cumple con lo especificado. Además, utiliza sistemas de tokenización para hacerlo más robusto y seguro.

En el proceso de mejora continua, el licitador indica que ha creado una estructura técnica modular que permite el cambio rápido entre tecnologías vía el encapsulamiento utilizando APIS. Indica que sus ingenieros van a hacer seguimiento de la evolución del mercado y que, en función de esta evolución, actualizarán las versiones. Además, incluyen mecanismos basados en vocabularios para mejorar pequeños errores. Dentro de estos procesos, incluyen una declaración de Google sobre el trabajo en la mejora de sus APIS.

A juicio de RTVE, el principal riesgo de esta propuesta está en que el licitador no es autónomo en la evolución de la base tecnológica, sino que dependerá de la evolución del sector al apoyarse en tecnologías Open Source.

Respecto de la prueba técnica, el sistema propuesto está procesando los seis canales principales de RNE. Tras la lectura de la escaleta, entre otras, fusiona las emisiones de un programa para simplificar el acceso a los elementos de la parrilla (siempre que estén adyacentes).

Seleccione programa a modificar

Buscar

PROGRAMA	DETALLES	CANAL	DATETIME INICIO	TIME FIN	ID PROGRAMA	SGCE
24 horas	Detalles del Programa	RNE Radio 1	2023-12-21 22:00:00	00:00:00	1871	MKA00005
Radiogaceta de los deportes	Detalles del Programa	RNE Radio 1	2023-12-21 21:00:00	22:00:00	2071	MKA00078
24 horas	Detalles del Programa	RNE Radio 1	2023-12-21 20:00:00	21:00:00	1871	MKA00005
Informativo Territorial	Detalles del Programa	RNE Radio 1	2023-12-21 19:50:00	20:00:00	0	MTA50126
Por tres razones	Detalles del Programa	RNE Radio 1	2023-12-21 19:06:00	19:50:00	119313	MLA10660
Informativo Territorial	Detalles del Programa	RNE Radio 1	2023-12-21 19:06:00	19:06:00	0	MTA50126

FILTRO

Por Canal

Todo

RNE Radio 1

Hora Inicio

Desde

Hasta

Buscar Reinciar

Por Emitidos

Iniciados

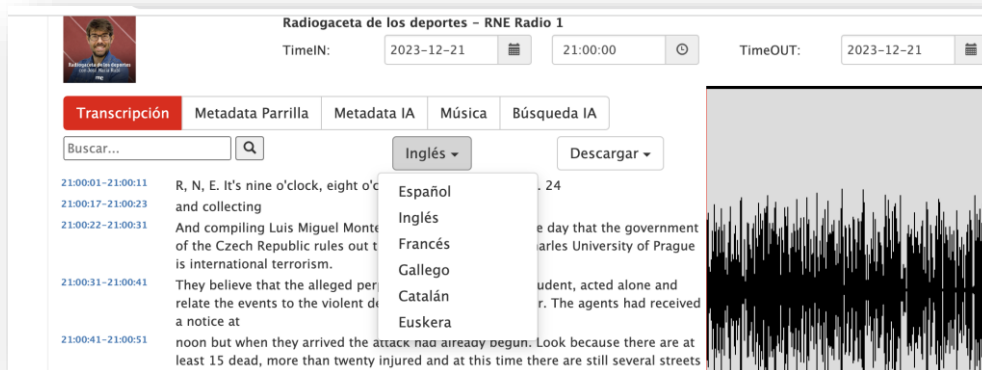
Todo

Cuando entras en un elemento presenta la siguiente pantalla:



En la pantalla de emisión se puede ver los subtítulos generados (con tiempo absoluto) así como el player para subclipping o consulta de la grabación.

Dentro de esta misma pantalla se puede consultar la traducción en los idiomas propuestos:



También permite visualizar la metadata adicional que se genera:

Transcripción	Metadatos Parrilla	Metadatos IA	Música	Búsqueda IA
resumen	El programa de radio "4 en la medular con Antonio Sánchez, Samu Costa, Manu Morlanes y Nadir Rodríguez. Y arriba Kyle Lari en el 11. De Osasuna, de Yago Barrasate sale Sergio Herrera" se enfoca en el fútbol, en particular el partido entre Real Madrid y Deportivo Alavés. El equipo de Javier Aguirre saldrá con Raikovic en portería, Maffeo, Gio González, Raillo, y el equipo de Deportivo Alavés cuenta con Abkar y Abde en la portería, Aurelio Calvente en la técnica. El programa también aborda temas de actualidad como el fallo del Tribunal Europeo de Justicia sobre la posible creación de un nuevo torneo llamado Superliga. También se menciona el sorteo extraordinario de la Lotería de Navidad y las opiniones de la audiencia sobre si este evento les provoca emoción o indiferencia. Además, se discuten temas políticos como la reunión entre el presidente del Supremo y el ministro de Justicia, Félix Bolaños, y la decisión del presidente del Gobierno, Pedro Sánchez, de adelantar elecciones autonómicas en Galicia. También se menciona la reconciliación entre el Gobierno español y la Generalitat y las cinco nuevas medidas acordadas. El programa incluye entrevistas con jugadores de fútbol y análisis deportivos, así como debates sobre los partidos y las noticias del día. La atmósfera es lúdica y entretenida, con un enfoque en el fútbol y la actualidad deportiva.			
entidades	entrevistas, 4 en la medular, Generalitat, Abde, Félix Bolaños, Maffeo, Manu Morlanes, debates, Abkar, Raikovic, Javier Aguirre, Gobierno español, Kyle Lari, Analistas deportivos, Sergio Herrera, Antonio Sánchez, Nadir Rodríguez, Aurelio Calvente, Tribunal Europeo de Justicia, Ministerio de Justicia, partidos, Pedro Sánchez, Samu Costa, Superliga, Galicia, Lotería de Navidad, Raillo, Supremo, Gio González, noticias del día			
compañías	Entrevistas a deportistas, Fútbol y otros deportes populares			
highlights	El equipo de Javier Aguirre saldrá con Raikovic en portería, Deportivo Alavés cuenta con Abkar y Abde en la portería, El fallo del Tribunal Europeo de Justicia sobre la posible creación de un nuevo torneo llamado Superliga, La reunión entre el presidente del Supremo y el ministro de Justicia, Félix Bolaños, La decisión del presidente del Gobierno, Pedro Sánchez, de adelantar elecciones autonómicas en Galicia.			

Músicas:

Transcripción	Metadatos Parrilla	Metadatos IA	Música	Búsqueda IA
Detección	21:04:05-21:04:13			
Canción	Dinero			
Álbum	Dinero			
Artista	Jennifer Lopez, DJ Khaled, Cardi B			
Género	Pop			
Duración	03:33			
Detección	21:05:35-21:05:43			
Canción	Rauw Alejandro: Bzrp Music Sessions, Vol. 56			
Álbum	Rauw Alejandro: Bzrp Music Sessions, Vol. 56			
Artista	Bizarrap, Rauw Alejandro			
Género	Latin Music			
Duración	03:11			
Detección	21:14:35-21:14:36			
Canción	greedy			
Álbum	greedy			
Artista	Tate McRae			
Género	Pop, Alternative			
Duración	02:09			
Detección	21:14:35-21:14:36			
Canción	Meero Baal Xaina			
Álbum	Na Rishau Hai Mero Maya			
Artista	Ranjita Kumari			
Género	Asian Music			
Duración	05:34			
Detección	21:26:06-21:26:08			
Canción	Houdini			
Álbum	Houdini			
Artista	Dua Lipa			

Toda esa información (salvo resúmenes y highlights) se tienen disponible incluso en la emisión que está en vivo.

Por todo ello se da por válida la prueba técnica, certificando que las capacidades ofertadas están todas disponibles en esta plataforma de pruebas.

En base a lo indicado, RTVE valora la propuesta de la siguiente forma:

- **Arquitectura Global de la solución:** 30 puntos.
- **Plan de proyecto:** 7 puntos.
- **Equipo servicios profesionales y modelo de soporte.** 8 puntos.

CROSSPOINT

CROSSPOINT ha presentado una propuesta compleja que analizamos siguiendo el esquema de propuesta técnica que ha presentado.

No se añadirá ningún tipo de imagen ni referencia literal de la propuesta técnica, ya que el licitador ha declarado confidencial la oferta y no puede ser objeto de reproducción total o parcial.

La arquitectura técnica presentada por el licitador se basa en una solución de mercado. Sobre esta corren todos los procesos de la solución. El licitador ha montado una infraestructura donde identifica redundancias de tipo software para mantener una escalabilidad. Sin embargo, desde un punto de vista de arquitectura, el licitador no concibe la separación de entornos. En las imágenes que acompañan a la arquitectura no hay una independencia entre la DMZ y las apis de distribución. En un entorno de gestión, el cual se va a emplear para captar información, gestionarla y finalmente tiene que distribuir la metadata, la no separación entre los procesos de gestión interna y las apis de distribución supone una mala decisión arquitectural que puede suponer un problema para los sistemas de producción de RTVE. Esto supone un riesgo y error conceptual.

El licitador indica en su página 3 que la solución que plantea se va a instalar en el CPD de Interxion Madrid, el mismo CPD que emplea RTVE. Esto hace que no haya problemas de latencias, y, nos da fiabilidad sobre la ubicación.

La solución que plantea el licitador, página 11, plantea tener un sistema de backup basado en la TDT para los canales lineales (los 6 principales canales broadcast). En la página 16 indica que va a adquirir hasta 130 señales de entrada de audio.

En la página 18 y en la página 19 de la PT el licitador indica que, para los directos, no se va a poder disponer de la transcripción hasta dentro de al menos 30 minutos desde el comienzo de un directo. Esta limitación que presenta el licitador es inasumible para RTVE. Recordar que en el PCT RTVE indicaba en la página 5:

*Respecto al API de distribución, éste será consultado en tiempo real por distintos servicios, tanto para la inclusión en los directos digitales (generando SRT o VTT parciales) como en los procesos de generación del encapsulado de la media VOD. El licitador ha de asumir una capacidad de al menos **100 peticiones por segundo**.*

El decalaje indicado por el licitador impide la inclusión en los directos digitales de la transcripción o del subtítulo, ya que estos serán provistos más de media hora después del directo. El modelo propuesto puede tener algún sentido si lo que se hace es únicamente el metadato y catalogación de podcasts, pero para directos no es razonable.

Lo anterior también choca con lo indicado en el PCT sección 3.3:

El formato de la transcripción tiene que ser tal que pueda ser utilizada out of the box para que un player pueda incluirla como subtítulo.

Para cualquier directo menor de 30 minutos, nunca podremos tener su subtítulo.

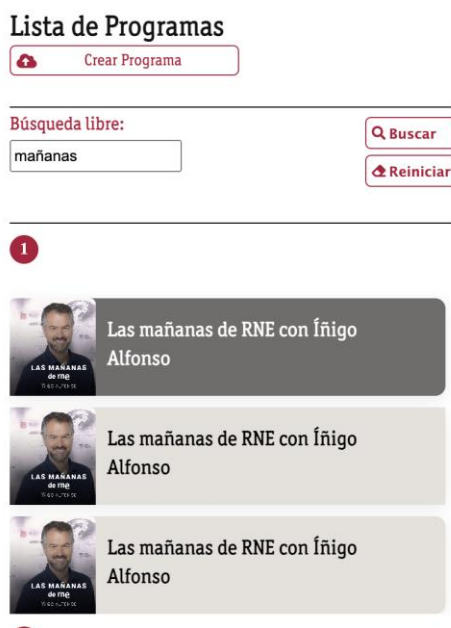
En la página 18 nos ha sorprendido encontrar que el sistema notifica a través de correo electrónico. Este mecanismo de notificación es inútil, máxime cuando existen múltiples opciones automatizables para realizar notificaciones a sistemas que puedan implementar soluciones reactivas a eventos tales como la disponibilidad de un asset.

Respecto a la monitorización, el licitador indica que se harán catas de 200 minutos al mes.

El documento desarrolla los elementos de seguridad de la solución. Estos están basados principalmente en que el CPD es el mismo que utiliza RTVE. Sin embargo, el licitador ha centrado sus explicaciones en la parte hardware (edificios, máquinas, etc.) pero no ha descrito la solución de seguridad: no aparecen en el documento los firewalls, estrategias de DDOS, separación entre DMZ y servicios de distribución, etc., lo que se penalizará.

El plan de proyecto es razonable, y el equipo es adecuado. El modelo de soporte y el plan de calidad parecen correctos.

Respecto de la prueba técnica, la solución puesta a disposición de RTVE tiene problemas graves de usabilidad tales como que el scroller de parrilla (un cuadro rojo) no es intuitivo o el buscador de programas muestra resultados repetidos.



Respecto de funcionalidades, la solución no permite el subclipping múltiple, sino tan solo la extracción de fragmentos.

La parrilla se ingesta tal cual viene desde RTVE y no se hace un asset con todas las emisiones, sino un asset por franja horaria (aunque compartan código SGCE y código de programa):



Para ver la transcripción y determinar la calidad de la extracción, hemos accedido a ver el subtítulo. El análisis del día 20 de diciembre de 2023 encontramos un error en la generación de la transcripción/subtítulo. Básicamente está mal el offset respecto del directo de los textos (ver imagen de ejemplo) entre otras, en las siguientes emisiones:

- Emisión programa 24horas de 23:00 a 00:00.
- Emisión programa 24horas de 22:00 a 23:00.
- Radio Gaceta de los deportes, de 21:00 a 22:00.
- Emisión programa 24horas de 20:00 a 21:00.
- Boletines RNE (19:50)

LISTADO DE SUBTITULACIÓN	
1	00:00:00:00 28 inglés.
	00:00:01:00
2	00:00:01:00 Qué tal. Muy buenas noches.
	00:00:02:00
3	00:00:02:00 Bienvenidos al capítulo 2 de esta trilogía Inter se al de
	00:00:03:00
4	00:00:03:00 y en 2ª División antes de irnos de vacaciones antes
	00:00:04:00
5	00:00:04:00 las uvas antes de la Navidad hay todavía mucho
	00:00:05:00

Además, las transcripciones de muchas de las emisiones están mal (ya no los códigos de tiempo, sino el texto), como por ejemplo en:

- **Mañana Mas (de 11:06 a 12)**



LISTADO DE SUBTÍTULACIÓN	
1	00:00:00:00 1 00:-1-6,-40g -->
	00:00:00:00 00:-1-5,-40g
2	00:00:00:00 2 00:-1-5,-40g -->
	00:00:00:00 00:-1-4,-40g
3	00:00:00:00 3 00:-1-4,-40g -->
	00:00:00:00 00:-1-3,-40g
4	00:00:00:00 4 00:-1-3,-40g -->
	00:00:00:00 00:-1-2,-40g
5	00:00:00:00 5 00:-1-2,-40g -->
	00:00:00:00 00:-1-1,-40g

- **Mañana Mas (de 12:06 a 13)**

LISTADO DE SUBTÍTULACIÓN	
1	00:00:00:05 00:00:00:06

- **El ojo crítico (de 15:04-16:00)**

LISTADO DE SUBTÍTULACIÓN	
1	00:00:00:00 1 00:00:-4,-200 -->
	00:00:00:00 00:00:-3,-200
2	00:00:00:00 2 00:00:-3,-200 -->
	00:00:00:00 00:00:-2,-200
3	00:00:00:00 3 00:00:-2,-200 -->
	00:00:00:00 00:00:-1,-199
4	00:00:00:00 4 00:00:-1,-199 -->
	00:00:00:00 00:00:00,-200
5	00:00:00:00 5 00:00:00,-200 -->
	00:00:00:00 00:00:00,800

Hemos detectado algunas transcripciones que parecen estar bien, salvo que no todo el audio está transcrito. Como ejemplo Boletín RNE del 21 de diciembre a las 00:00 (falta más de 30 segundos de transcripción) o Gente despierta (de 00:05 a 01:00) que le falta más de un minuto de transcripción.

Respecto a la puesta a disposición de la metadata, podemos comprobar que para los directos es necesario más de media hora para disponer de la transcripción.

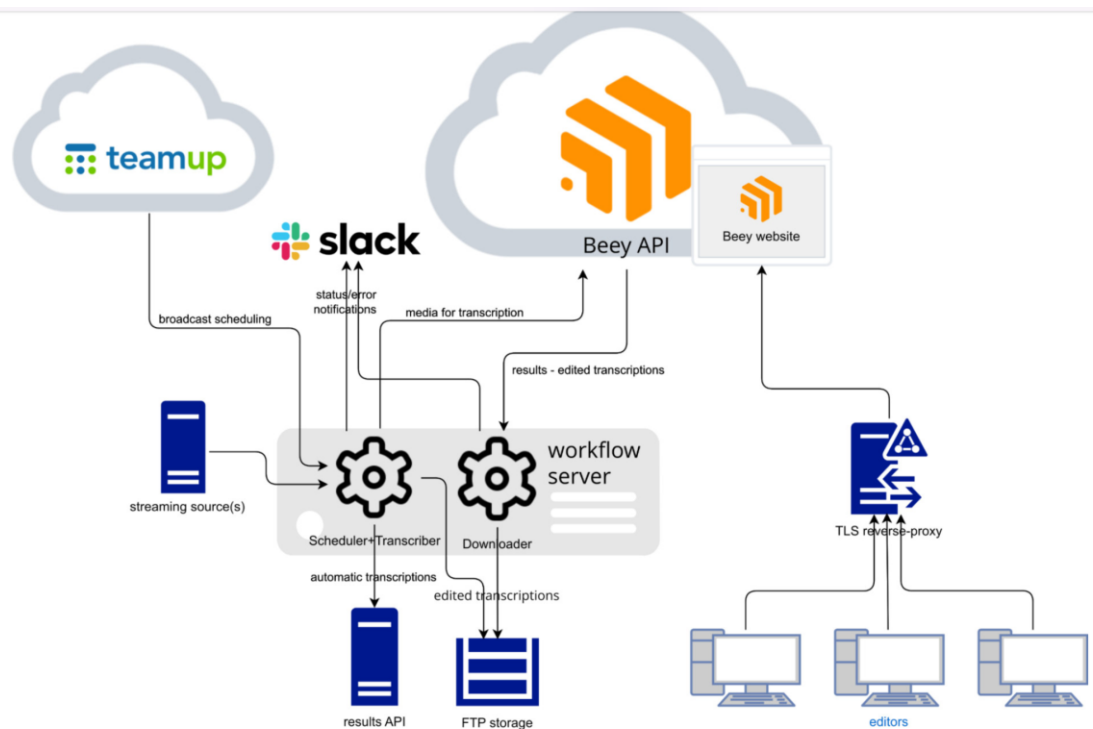
Finalmente, resaltar que, aunque la solución que aporta el licitador está basada en el uso de RapidAPI para la obtención de las canciones y carátulas, la prueba técnica utiliza Shazam, lo que se penaliza en esta valoración.

Debido al nivel de errores en la prueba técnica y debido a que la solución técnica no cumple los requisitos mínimos establecidos en el PCT, la propuesta del licitador CROSSPOINT se declara como **NO APTA**.

REBOLD

El licitador ha presentado una propuesta basada en su infraestructura en nube (Azure) llamada Beey. El hecho de estar en Azure o que sea una infraestructura que se puede licenciar en la actualidad no exime al licitador de presentar toda la información que se exige en el pliego de condiciones técnicas. Así, el licitador omite casi toda la información que se requiere. Enumeramos las faltas de información que no aparecen en la propuesta técnica:

- Detalle de la arquitectura de la solución. El licitador indica como arquitectura técnica la siguiente imagen:



La arquitectura se basa en un conjunto de elementos que ingestan o se integran con Beey Website/Api. De esa pieza lo único que indica el licitador son sus funcionalidades, pero no el detalle técnico que permita a RTVE valorarla. Desconocemos número de máquinas, sistemas de seguridad, sistemas de escalabilidad, etc. También desconocemos los algoritmos, mecanismos internos de gestión del audio, etc. Toda esta ausencia de información no permite valorar la arquitectura de la solución.

- Modelo de gestión: no se detalla. Se habla de procesos en general, pero no detalla el flujo de los procesos.
- Mecanismos de extracción de datos: el licitador no provee esa información mencionando únicamente que utiliza tecnologías de IA.
- Alcance de la solución:
 - Únicamente transcripción.
 - La solución no tiene capacidad de subclipping.
 - La solución no genera forma de onda o entidades.
- Productos empleados y licenciamiento

El licitador no indica qué licencias son las que usa para dar la solución. Únicamente nos indica que:

Nuestra solución utiliza la robusta infraestructura proporcionada por los servicios incloud Azure, evitando la necesidad de ser propietarios de hardware físico.

Este modelo basado en la nube asegura el acceso a recursos computacionales de alta calidad y las últimas tecnologías. Operamos completamente dentro del ecosistema de Azure, aprovechando su infraestructura escalable validada a través de su implementación en numerosos proyectos punteros. Todas las licencias necesarias son gestionadas a través de nuestra suscripción de Azure, garantizando el cumplimiento y el acceso a una amplia gama de servicios y herramientas.

Es decir, que utiliza servicios de Azzure pero no los enumera. Esto contraviene la exigencia de concreción indicada en el PCG, por lo que se penalizará en la valoración.

- Definición detallada de las APIs de consulta y de ingesta.

El API entregada no contienen los endpoints mínimos exigidos por RTVE (página 4 del PCT).

- Documentación sobre modelo de aseguramiento de la seguridad.

El licitador no da detalle de cómo asegura la seguridad, no da información sobre firewalls, sistemas de protección perimetral, separación de redes, separación de DMZ respecto a APIS de distribución, etc. Se limita a indicar generalidades como:

Nuestra propuesta de modelo de aseguramiento de la seguridad para RTVE se alinea con los estándares más rigurosos, garantizando la protección de los recursos relacionados con la gestión de la información. Cumplimos con la legislación vigente, preservando la confidencialidad y privacidad de la información, asegurando la disponibilidad, acceso, integridad, calidad, trazabilidad, autenticidad y conservación.

Política general de seguridad de la información: Implementamos una política general de seguridad de la información que abarca todos los recursos, servicios y procesos de RTVE, así como a todos los sistemas de información utilizados en la prestación de servicios y en el soporte de funciones y responsabilidades.

Notificación y gestión de incidentes: Establecemos un procedimiento eficiente para la notificación y gestión de incidentes de seguridad de la información. Utilizamos una herramienta de ticketing para el registro y seguimiento de incidencias. El proceso incluye la evaluación de daños, resolución y cierre, con especial atención a la clasificación de incidentes y su impacto potencial.

Adicionalmente indica:

Seguridad Avanzada:

- Integración de cortafuegos y sistemas de prevención de intrusiones.
- Medidas de cifrado robustas en la transmisión y almacenamiento de dato

Objetivo	Métricas de Evaluación	KPI Objetivo	Detalle
Calidad del Código	Análisis estático del código, revisión de pares, métricas de complejidad ciclomática	> 90% de código con alta calidad	Prácticas de desarrollo limpio y herramientas automáticas de análisis de código.
Seguridad	Pruebas de seguridad, análisis de vulnerabilidades, cumplimiento de estándares de seguridad	0 vulnerabilidades críticas	Prácticas de desarrollo seguro y pruebas exhaustivas de seguridad.

La información aportada por el licitador es insuficiente.

- Plan de implantación de la solución y currículums de las personas que forman parte de la implantación.

El plan de proyecto presentado no está detallado dentro de la propuesta técnica. En su lugar el licitador indica lo siguiente:

Fase 2: Diseño y Planificación (Semana 3-4)

- Diseño detallado de la arquitectura de la solución, incluyendo la integración y desarrollo de nuevos módulos.
- Definición de roles y responsabilidades dentro del equipo de desarrollo.
- Planificación detallada del proyecto, identificando hitos clave y asignando plazos específicos.
- Confirmación de la aceptación por parte de RTVE del diseño propuesto.

Es decir, el licitador pretende entregar un plan de proyecto y una planificación detallada en la fase 2 en ejecución. Esto no es admisible ya que el plan de proyecto y la planificación detallada forman parte del objeto de valoración.

Otro elemento que no es baladí, es que el licitador pretende delegar el mantenimiento en personal de RTVE. Así, en el plan de proyectos el licitador indica:

Fase 5: Implementación y Entrega (Semana 11-12)

- Implementación de la solución en el entorno de producción de RTVE.
- Entrenamiento del personal de RTVE en el uso y mantenimiento de la nueva funcionalidad.

Esta pretensión no es aceptable, ya que el soporte y mantenimiento lo ha de realizar el prestador con sus propios medios.

- Plan de aseguramiento de la calidad y de informes de calidad.

No está detallado suficientemente.

El licitador ha presentado una implementación para la prueba técnica. El resultado del análisis es que, aunque la solución no falla, es decir, no tiene errores, y ha transcrito un trozo de R5, verificamos que

la herramienta es muy limitada respecto de lo que aportan otras soluciones. La solución está centrada únicamente en la transcripción y en el subtitulado, pero no en la gestión de la media.

Tras el análisis de la propuesta técnica, y dadas las faltas de información que se han reseñado, la propuesta técnica de la empresa Rebold se puntúa con **cero puntos**.

RESUMEN

Tras el análisis de la documentación aportada, se valora cada propuesta de la siguiente forma:

<i>Licitador</i>	Arquitectura Global	Plan proyectos	Equipo servicios profesionales	Total
<i>Aeternal Metis</i>	NO APTA			
<i>Cinfo</i>	5	4	8	17
<i>Cires 21</i>	30	7	8	45
<i>Crosspoint</i>	NO APTA			
<i>Rebold</i>	0	0	0	0

Madrid, diciembre de 2023