

## VALORACIÓN TÉCNICA DEL EXPEDIENTE

### S-01185-2023 - RTVEPLAY-i Implantación solución subtítulo

El objeto del presente documento es realizar la valoración técnica del expediente S/01185/2023. 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, 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: 25 puntos.
- Plan de proyecto: 10 puntos.
- Equipo de servicios profesionales y modelo de soporte: 10 puntos.

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

- **Arquitectura Global de la solución:** *se valorará la arquitectura general de la solución, el mecanismo de alineamiento propuesto, 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 las herramientas que se proponen y el detalle de la arquitectura técnica y de cómputo presentada. Finalmente, también se tendrá en cuenta el riesgo de la solución planteada, así como 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.*

## **VALORACIÓN TÉCNICA**

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

- Cires 21
- Cross Point
- Euskaltel
- Five Flames
- Unuware

Se procede a analizar cada una de ellas a continuación.

### **CIRES 21**

Tras el análisis de la propuesta, hay que decir que es una propuesta robusta y muy completa. Se valora especialmente que la solución sea robusta y redundada, y que el licitador haya puesto énfasis en la claridad de los dimensionamientos, así como en detalles como las instalaciones y cableados. En la oferta el licitador ofrece hasta 8 canales, incluyendo la capacidad de llegar hasta 32 tarjetas SDIs con su correspondiente cableado. Sin embargo, dentro de la propuesta se echa en falta algunos elementos que no aparecen, como, por ejemplo, la presencia de switches o mapas de red.

Para el tema del alineamiento, el licitador ha planteado soluciones muy interesantes. Se valora especialmente las distintas tipologías de alineamiento (enlatado versus directo), así como la utilización de los recursos tanto de señales como JASON. También se valora positivamente el modelo de estimación del decalaje, así como su claridad de cálculo, aunque no tanto el valor que da.

Una cuestión que no es baladí es la solución speech to text seleccionada por el licitador. Éste utiliza <https://github.com/openai/whisper>, una solución open source para la traducción. El problema que plantea el uso de whisper son los idiomas que puede gestionar. Según se desprende del enlace que añade el licitador (<https://raw.githubusercontent.com/openai/whisper/main/language-breakdown.svg>) el euskera no está soportado (ver imagen 1). Esto es un primer problema para la detección del idioma que se está empleando, y supone un limitante a la hora de abordar la inclusión de nuevos canales.

Además, y respecto de la introducción de este tipo de tecnologías, no está nada claro cómo hace el entrenamiento el licitador (hay cinco modelos de entrenamiento), pero, lo que es más importante, no introduce en el modelo de mejora continua la actualización de los modelos. De forma general, los sistemas basados en IA necesitan de cierta actualización para mejorar su rendimiento y eficacia. Cada cierto tiempo se añaden nuevos entrenamientos que mejoran los modelos y, que finalmente, acaban mejorando el producto. Es por ello que, cuando se introduce una tecnología basada en IA, salvo que se consuma en modelo SaaS, hay que contemplar la actualización del modelo. En otro caso, no hay una

mejora continua del producto y de la solución. Es por ello que se valora negativamente el modelo de mejora continua.

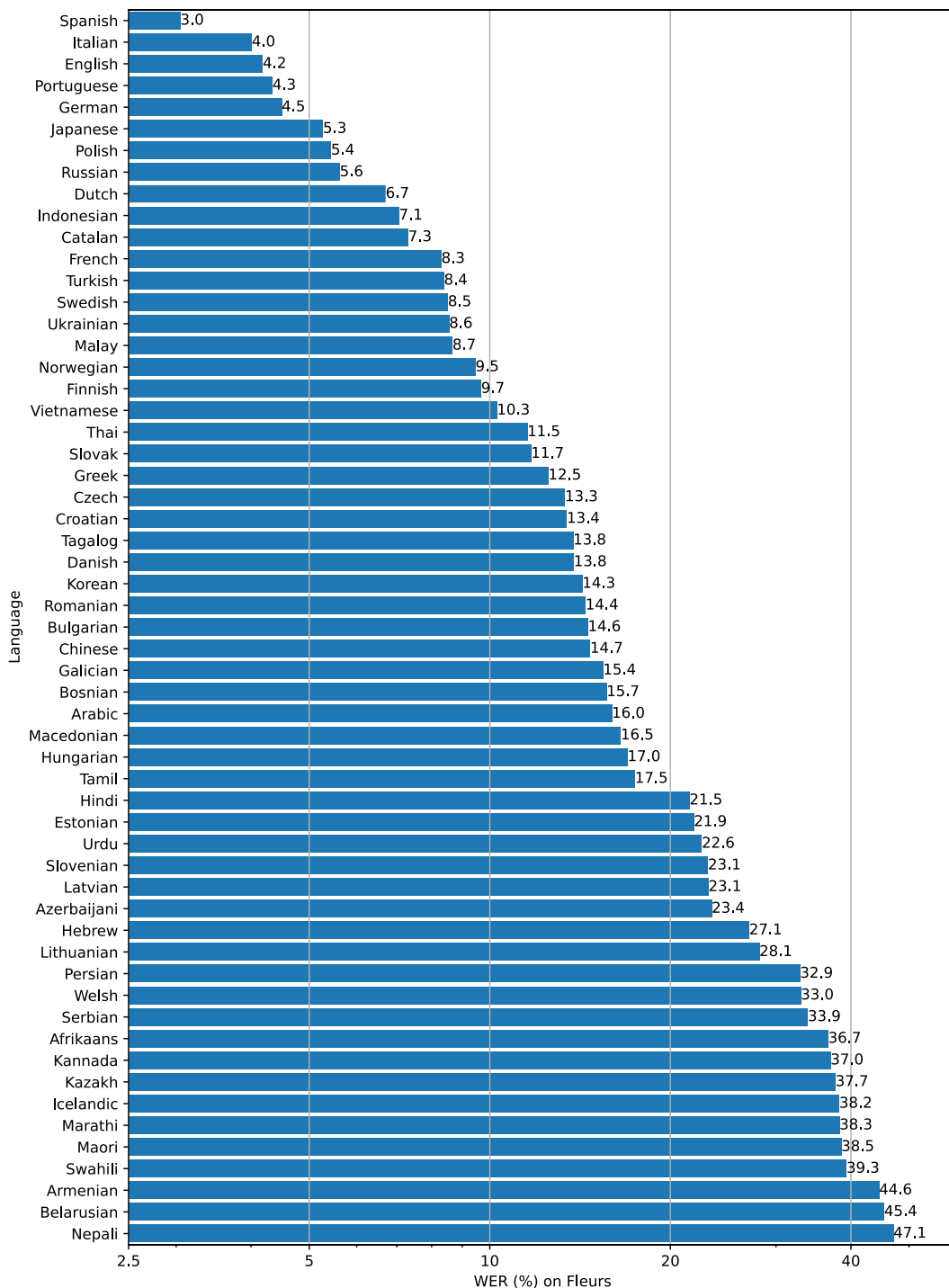


Imagen 1: idiomas soportados por whisper

Dentro de la gestión de JASON, el licitador ha pasado muy por encima de cómo se filtran los elementos publicitarios, y cómo se unen programas (futbol) que tienen un elemento en medio (telediario) para hacer la composición del VOD. El único punto donde se trata esto es la página 25 de la propuesta.

Respecto de la traducción, su propuesta sugiere la utilización de la tecnología de Google, por lo que debe gestionar la salida a Internet para las comunicaciones. Sin embargo, la propuesta no aborda cómo se garantiza esta conectividad, switches, etc. Propone el uso de otra tecnología de Meta como alternativa para abaratar costes, pero no desarrolla el modelo de entrenamiento, ni tampoco, el modelo de cómo se actualizará el modelo de dicha tecnología.

Respecto del API es completa y detallada.

Respecto del modelo de gestión de la calidad, desde un punto de vista técnico está muy bien, pero no, desde una perspectiva funcional. Se echa en falta la realización de catas aleatorias para revisar la calidad de las traducciones diarias, ya que los informes son mensuales y trimestrales.

Respecto del plan de proyectos, aunque ha presentado un desglose de tareas de alto nivel, se echa en falta una mayor precisión de las tareas a ejecutar, así como los recursos concretos que va a destinar para cada tarea. El plan de proyectos es hacerlo en 3 meses, aunque nos parece un poco justo.

Finalmente, el equipo profesional, y el modelo de soporte nos parece correcto. ANS 99,99%.

Tiempos de respuesta/ resolución:

- Crítica: 15/60.
- Alta: 60/240.
- Normal: 480/480.

## **CROSSPOINT**

En líneas generales, la solución nos parece menos clara que la de Cires 21, comprende más elementos, intervienen más actores y la responsabilidad está mas diluida. El sistema que plantean tiene un vicio: tan solo plantean 7 u 8 entradas en sus codificadores, dependiendo de SDI o multicast. Es cierto que, en la parte de redundancia incidan que hay el doble de equipamiento, pero también es cierto que indica en su propuesta que la configuración es activo-pasivo, por lo que en la práctica hay un problema con las entradas.

El alineamiento lo está realizando sobre la señal de la TDT (ya que la señal limpia no la puede utilizar ni lo indica), y no se entiende en qué momento emplea los códigos de tiempo para el ajuste fino. Valoramos de forma muy negativa la falta de concreción al respecto del uso del elemento de sincronización. Es básicamente un documento comercial más que una especificación técnica.

Tampoco se entiende el rol del listener FAB. Siendo un elemento opcional, se le da un valor que no debería tener.

Respecto del equipamiento y la redundancia es exhaustivo, salvo que no indica en qué hardware estarán alojados los desarrollos de Crosspoint. La solución está suficientemente redundada, y está

indicado para el resto de elementos. De igual forma, indica el uso de hardware de comunicaciones, pero no especifica cuál es, incluye circuitos, pero no da detalles de proveedor ni caudal.

Valoramos positivamente el modelo udp de buffering, el modelo para corregir términos de la traducción (aunque se puede volver farragoso), y, en general, el modelo de API.

Respecto de la monitorización y mejora de la traducción, el licitador propone que personal nativo haga catas de 200 minutos al mes por idioma.

Para la traducción utiliza los servicios de Google.

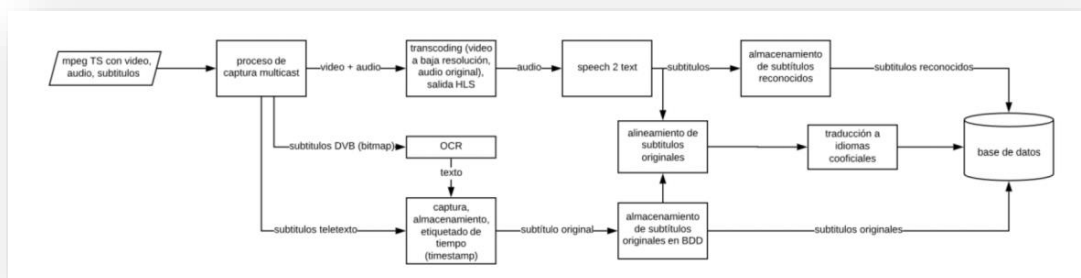
El plan de proyectos es correcto, pero no identifica los recursos con las tareas como se exige en el PCT. El personal que dedica está cualificado.

Finalmente, el equipo profesional nos parece correcto el modelo de soporte nos parece correcto. Sin embargo, los tiempos de respuesta no están definidos en todos los casos.

## EUSKALTEL

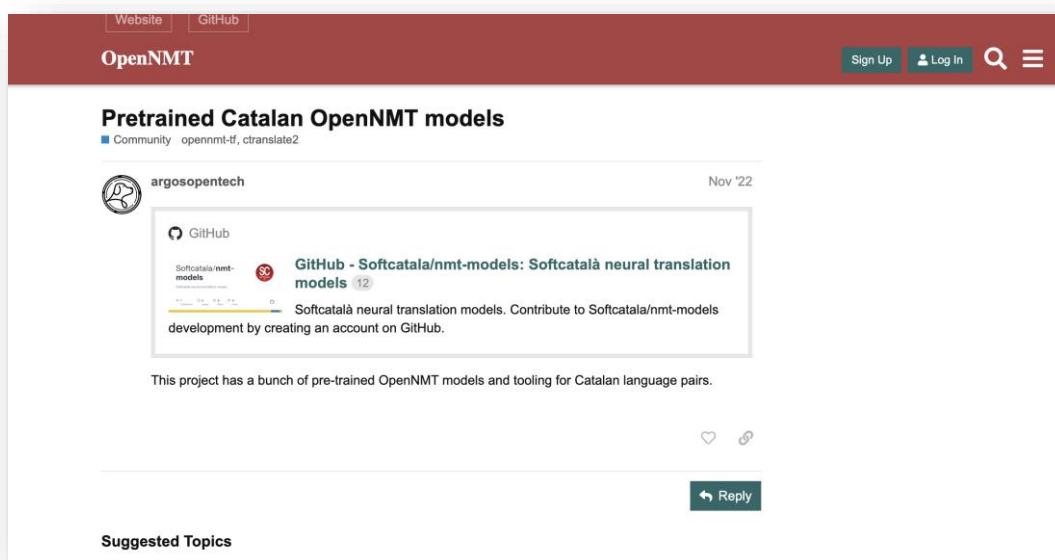
La solución que aporta el licitador se basa en la contribución únicamente de la señal mpeg TS que incluye el audio, el video y los subtítulos. Esto es, en la señal que luego se distribuye a la TDT. Además, la captura de las señales se está haciendo únicamente por multicast. Esto no es correcto porque las señales podrán venir por multicast o SDI. Sin embargo, el licitador no incluye ninguna tarjeta ni opción para la entrada de señales SDI.

Para el alineamiento de los subtítulos (ver pág. 10 de la propuesta técnica), *“con los subtítulos originales (de la tdt) y los obtenidos por la IA (de la tdt) se realiza el alineamiento de los originales a los tiempos reales obtenidos mediante algoritmos de matching de textos ...”*. Es decir, el alineamiento lo está haciendo sobre la señal de la tdt y no sobre la señal limpia. Si atendemos a lo que se indica en el PCT (pág. 4) *“RTVE produce directos con un decalaje de cuatro segmentos de cinco segundos respecto de la señal limpia”*, encontramos que la solución del licitador no es válida para los directos por la sencilla razón que no está teniendo en cuenta que los directos los emitimos en función de lo indicado en la señal limpia, o, de lo que se indica en el feed de JASON. El gráfico de la página 12 no deja lugar a dudas de la no utilización de estas dos fuentes.



Respecto del sistema de traducción hay dos elementos que hay que considerar. Lo primero es el modelo de gestión de la cola de traducciones pendientes. Dice el licitador (página 16 de la propuesta técnica) que “... para ello, además de una gestión basada en una cola LIFO que respeta el orden de entrada para el procesado ...”. Si tenemos en cuenta que una cola LIFO (Last In, First Out) lo que hace es extraer primero los últimos trabajos pendientes, la solución del licitador va a provocar en situaciones de alta carga que las traducciones más antiguas sean postergadas, lo que va a llevar al encolamiento de traducciones y a la pérdida de traducciones para el directo. Claramente el esquema que tendría que haber definido es una FIFO (First In, First Out).

El segundo aspecto sobre el sistema de traducción tiene que ver con la tecnología que propone emplear. Según indica en la propuesta técnica, plantea la instalación de una tecnología open source, que luego es entrenada con un modelo de idioma que publica una comunidad con licencia open source (OpenNMT). Indagando en OpenNMT, por ejemplo, para el caso de catalán podemos ver que el modelo de catalán es provisto por alguien de la comunidad (en este caso una asociación sin ánimo de lucro):



Es decir, que para buscar los modelos del catalán tenemos que irnos a un tercero que es un usuario de github (Softcatala) que publicó un modelo para la traducción catalán-español en 2022. Desde la perspectiva de RTVE, y dado lo relevante del proyecto que se está llevando a cabo, y desde una perspectiva de riesgos del proyecto, no nos parece que este enfoque sea apropiado. El licitador con el pretexto de abaratar costes se desprende de la responsabilidad de la traducción, no teniendo ningún control sobre los modelos.

Esta solución contrasta con la de otros licitadores que se han apoyado en soluciones profesionales de alto rendimiento y alta eficacia (Google, Microsoft), y carece de un elemento fundamental: tener un modelo entrenado actualizado. De hecho, otro detalle que destaca dentro de la propuesta es que el modelo de mejora continua se basa en “una revisión semestral de los modelos machine learning utilizados” (página 30). Es decir, en el mejor de los casos, el licitador tardará seis meses en actualizar su tecnología de traducción (si es que alguien de la comunidad ha actualizado los modelos).

Finalmente, el licitador no ha entendido el uso de JASON (de hecho, en la propuesta pone de forma insistente JSON), y no ha montado lógica para eliminar los bloques promocionales dentro del subtítulo. En otras palabras, para preparar el VOD, el licitador no hace una curación de los subtítulos. En el anexo I se explicaba cómo JASON indica no sólo el comienzo real de una emisión (que resulta estar alineada con la señal limpia) sino también los emplazamientos publicitarios dentro de una emisión. Esto no ha sido considerado por el licitador.

Plan de proyectos 4 meses.

Tiempos de respuesta/ resolución:

- Crítica: 15/60.
- Alta: 30/240.
- Normal: 240/480.

## **FIVE FLAMES**

La solución de Fiveflames incluye la lectura de siete fuentes SDI por cada servidor (hay 4), así como entradas multicast. Estas son procesadas extrayendo el OCR y el teletexto, primando el teletexto frente al OCR.

Para el alineamiento utiliza tecnología de Google para el speech to text (en principio cada 10 minutos), así como los eventos de JASON para determinar a qué programa pertenece el subtítulo, así como para hacer algunas inferencias:

- Ya está subtítulo.
- Estimar desalineamiento a priori.
- Determina el cambio de cadencia del subtítulo.

Con esa información almacena la información en la base de datos y la pone a disposición del api. Con estos datos monta un mecanismo diferenciado entre el directo y el vod. La solución es útil para el sistema que va a inyectar los subtítulos en los directos, ya que en el caso de los directos utiliza mecanismos de ajuste para mejorar la legibilidad. El api presenta parámetros específicos que permite mejorar la sincronización y la entrega de los subtítulos para el directo. Es quizá el api más pensado para la distribución, incluyendo parámetros como desfase máximo, gestión del contador de subtítulos para directos o epoch para la ingesta en directos. El único problema que presenta la solución es que el cálculo del decalaje del alineamiento de una emisión lo hace cada 10 minutos. Es por ello que hace una fase de calentamiento para estimar el decalaje promedio. La solución está bien para emisiones enlatadas, pero puede presentar problemas en el caso de directos que tenga delay diferenciado.

Dentro de la propuesta nos ha parecido muy interesante lo referente a la curva de ajuste. Básicamente, para evitar efectos de bola de nieve en los subtítulos del directo, el licitador utiliza un mecanismo para hacer el cambio de forma gradual de un modelo de delay a otro, con el objeto de garantizar la legibilidad de los subtítulos en el player o en el móvil.

La solución, además, incluye la capacidad de subir subtítulos y hacer el alineamiento de forma manual para corregir errores de subtitulación. Además, la solución ofrece hasta 10.000 horas de Speech to Text para poder generar subtítulo de programas que no trajesen subtítulos y se quisiesen incorporar.

Respecto del sistema, presenta una redundancia completa e independiente. Se valora positivamente que se haya añadido toda la especificación de todo el hardware, inclusive el firewall para la salida a internet. En contraparte, se echa en falta algo más de detalle en la arquitectura técnica.

Respecto del modelo de desarrollo, todos los elementos (salvo Google) son entregables para RTVE. Además, se valora positivamente que es el único licitador que monta entornos de desarrollo y de integración en sus oficinas para poder realizar los trabajos de implantación o de evolución.

Respecto del plan de proyectos es completo, indica los recursos a las tareas y son perfiles adecuados a las tareas que van a realizar. Incluye el cableado Ethernet, señales de video SDI y energía. Estima algo más de 3 meses.

El modelo de calidad incluye catas manuales diarias sobre contenidos de las últimas 24 horas, actuando sobre el sistema ante fallo. No incluye entrenamiento sobre traducción al delegar en Google.

Tiempos de respuesta/ resolución:

- Crítica: 10/50.
- Alta: 50/210.
- Normal: 420/480.

## UNUWARE

La propuesta técnica presentada por Unuware presenta tres aspectos que no se ajustan a lo solicitado en el pliego de condiciones técnicas.

1. El primero tiene que ver con el plan de proyecto. Según consta en la propuesta técnica, página 50 (ver imagen adjunta), la primera semana de la ejecución del contrato el licitador plantea el análisis y la planificación.

### 1. Análisis y planificación (1 semana)

- Reunión inicial con el equipo de RTVE para definir y confirmar los requerimientos y objetivos del proyecto.
- Elaboración del plan de proyecto detallado, incluyendo la asignación de recursos y responsabilidades.
- Compra del hardware.

Como resultado de esta fase, el licitador indica que va a elaborar el plan de proyecto detallado, incluyendo la asignación de recursos y responsabilidades. Hay que tener en cuenta que el plan de proyectos es uno de los elementos que forman parte de la valoración técnica y por tanto, tenía que ser entregado por el licitador para su valoración.

La pretensión del licitador, es decir, elaborar el plan de proyectos tras la adjudicación del contrato, pone al licitador en una posición de ventaja competitiva respecto de los demás

licitadores, incumple lo exigido en el PCT, restringe la concurrencia que debe presidir la contratación pública y la igualdad de trato y libre competencia de todos los licitadores.

2. El segundo elemento que se detecta en la propuesta tiene que ver con la falta de decisión del licitador respecto de los elementos que propone. Así, en la propuesta técnica, página 42, indica que:

En el módulo de traducción es posible realizar la traducción de subtítulos utilizando tanto modelos de lenguaje a gran escala como métodos de traducción tradicionales. Sin embargo, recomendamos el uso de modelos de lenguaje a gran escala, ya que emplean técnicas avanzadas de aprendizaje profundo y procesamiento del lenguaje natural (NLP) para generar traducciones más precisas y coherentes. Estos modelos capturan mejor el contexto y las relaciones entre palabras y frases en los idiomas de origen y destino, lo que resulta en traducciones más fluidas y naturales en comparación con los enfoques tradicionales basados en reglas o estadísticas.

Es decir, el licitador **recomienda** *el uso de modelos de lenguaje a gran escala*. Además, en la página 46 indica:

#### 6.9.5 Módulo de traducción

Para el proyecto propuesto disponemos de un módulo de traducción, el cual es parte de la plataforma ShowMeText Off, y es usado por varias empresas que se encargan de crear accesibilidad para cadenas de televisión, productoras, etc.

Actualmente, disponemos de dos enfoques para llevar a cabo traducciones: mediante modelos de lenguaje a gran escala y mediante métodos tradicionales. La elección entre estos dos enfoques dependerá de los resultados obtenidos durante la implementación del proyecto.

El licitador no indica qué va a implementar, haciendo una recomendación que no determina ni concreta con la pretensión de emplazar la solución técnica a otro momento del proyecto, justo después de la adjudicación, impidiendo valorar la solución propuesta en el momento de valoración técnica con carácter previo a la adjudicación. El resto de licitadores sí que indican qué solución van a emplear, lo que permite hacer una determinación de la calidad de la solución en tiempo de valoración. Hay que recordar que el objeto del contrato es la implantación de una solución de subtítulo que se valora mediante criterios sujetos a juicio de valor, lo que nos lleva a un plus de exigencia en cuanto a las pretensiones de los licitadores. Permitir esta propuesta pondría a este licitador en ventaja respecto de la competencia, vulnera las exigencias del pliego, así como la igualdad de trato con respecto del resto de licitadores.

3. El tercer elemento que hay que destacar versa sobre un elemento que introduce en la propuesta técnica, la plataforma SIGAS. El licitador indica en la página 47:

#### 6.9.10 Plataforma

Para este proyecto se adaptará la plataforma de SIGAS agregando las interfaces adicionales necesarias para el funcionamiento del sistema propuesto.

La mayoría de las interfaces propuestas aún no se han desarrollado. Hasta el momento, solo se encuentra disponible la interfaz "Cuadro de mando" que se usa en la plataforma SIGAS.

Para, a continuación, indicar que (página 48) *"se adaptará la plataforma SIGAS incorporando únicamente las nuevas interfaces. Todas las interfaces previamente definidas en la presente serán desarrolladas, con excepción del "Cuadro de Mandos" y las interfaces comunes de la plataforma SIGAS, como la interfaz de inicio de sesión, menús, cabecera, pies de página, etc"*.

Contar con una especificación técnica de qué interfaces incluye SIGAS es fundamental para poder evaluar el alcance de los trabajos del proyecto. Dicha definición técnica no está incluida en la oferta y el licitador añade lo siguiente en su propuesta:

S-07673-2022-Sistema Inteligente para la Gestión Automática del Subtitulado (SIGAS) en TDP, América y Asia. (2022, November 16). Portal de Licitación Electrónica de Radio y Televisión Española.  
<https://licitaciones.rtve.es/licitacion/licitaciones/detalle?id=1984138>

Es decir, referencia a un pliego público a través de una url. Esto contraviene lo establecido en el PCG que establece que toda la documentación del proyecto ha de ser entregada en los sobres electrónicos. No es admisible que haya que acceder a documentos externos a la licitación para hacer una valoración técnica.

En conclusión, la propuesta presentada por Unuware no ofrece una solución técnica concreta y determinada. Además, el licitador pretende que aspectos clave de la misma como el plan de proyecto o la solución definitiva a implantar, sean configuradas posteriormente conjuntamente con RTVE.

Por los motivos indicados, la propuesta del licitador UNUWARE no puede ser valorada y se declara como **no apta**.

## RESUMEN

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

<i>Licitador</i>	<b>Arquitectura Global</b>	<b>Plan proyectos</b>	<b>Equipo servicios profesionales</b>	<b>Total</b>
<i>Cires 21</i>	25	6	8	39
<i>Crosspoint</i>	10	8	6	24
<i>Euskaltel</i>	0	9	9	18
<i>Five Flames</i>	20	10	10	40
<i>Unuware</i>	NO APTA			

Madrid, abril de 2023