



Financiado por
la Unión Europea
NextGenerationEU



Plan de
Recuperación,
Transformación
y Resiliencia

rtve

RTVEPLAY-i Implantación grafo conocimiento semántico

PLIEGO DE CONDICIONES TÉCNICAS

rtve



Índice	Página
1 Introducción.....	3
2 Objeto	3
3 Descripción de la solución requerida.....	3
3.1 Unificación de tesauros y vocabularios.....	4
3.2 Implantar la base tecnológica del grafo de conocimiento	5
3.3 Poblado del grafo de conocimiento	6
3.4 Publicación de buscador facetado y portal navegable de consulta de la ontología	7
3.5 Implementación de interfaces para la actualización desde FD y RTVE Play	7
3.6 Licenciamiento, conectividad	8
3.7 Gestión de la calidad y acompañamiento	9
3.8 Servicios profesionales	9
3.9 Fases del servicio	9
3.10 Coordinación de los servicios	9
3.11 Otros aspectos del servicio.....	10
4 Estructura solicitada para la propuesta técnica	10
4.1 Resumen Ejecutivo	10
4.2 Oferta técnica	10

1 Introducción

La plataforma técnica de RTVE Play (en adelante, la plataforma) tiene múltiples relaciones con diferentes sistemas, tanto internos de RTVE como de terceros. Para RTVE los intercambios de información son siempre difíciles y requieren de mapeos entre los distintos sistemas para una adecuada integración. De entre los sistemas con los que se relaciona la plataforma destaca el fondo documental, archivo de la producción audiovisual de RTVE con millones de horas de contenido producido por RTVE en sus más de 50 años de vida.

RTVE, como miembro de la UER, participa de grupos de estandarización como, por ejemplo, el grupo EBU-AIM (<https://tech.ebu.ch/groups/aim>), grupo de la UER que gestiona y fomenta el uso de sistemas estándares de metadatos como el ebucoresplus. La adopción de estándares dentro de una organización en el ámbito semántico permite la mejora de la interoperabilidad interna y externa, facilita el link de datos y el vínculo con otras ontologías de representación del contenido.

Teniendo en cuenta los dos elementos anteriores, y en cumplimiento del encargo que hace la Secretaría de Estado de Telecomunicaciones y Plataformas Digitales (SETELECO), se lanza el presente expediente que tiene como objetivo la mejora de la interoperabilidad semántica de la nueva plataforma RTVE Play con distintos sistemas, actores del mercado y con el Archivo de RTVE. En este contexto, son también objetivos del proyecto:

- Implementar un estándar europeo de referencia dentro del sector media español;
- Crear un grafo semántico de conocimiento como repositorio de términos y vocabularios que se pondrá a disposición del sector para, entre otros, favorecer la evolución tecnológica de los sistemas de Inteligencia Artificial y la interoperabilidad semántica;
- Poblar el grafo de conocimiento con los contenidos de RTVE Play y mostrar modelos de uso, como son los buscadores facetados.

Este proyecto se lanza con la participación del Fondo Documental de RTVE y la Dirección de Sistemas de RTVE.

2 Objeto

El objeto de este pliego es definir los requisitos técnicos necesarios para la implantación de un grafo de conocimiento a partir de los tesauros del Fondo Documental, así como el poblado del grafo con los contenidos de la plataforma RTVE Play, en los términos que se describen en la sección 3.

3 Descripción de la solución requerida

Para la ejecución del presente proyecto el licitador ha de realizar un conjunto de acciones interrelacionadas que tienen como objeto final la implantación de un grafo de

conocimiento basado en una ontología de referencia encadenada con otras ontologías (Geonames, IMDB, DBPedia o Wikipedia) en los términos que se describen en este expediente. Para ello el proyecto presenta las fases indicadas en la sección 3.9. De entre las fases indicadas tan solo la primera será objeto de financiación por los Fondos Next Generation en virtud del encargo que hace SETELECO a RTVE en su línea de acción número 6, “Evolución en metadato y búsqueda semántica”.

La primera fase, tiene el siguiente alcance:

1. Unificación de los tesauros del Fondo Documental de RTVE y vocabularios de la plataforma en los términos que se indican en la sección 3.1.
2. Implantar la base tecnológica del grafo de conocimiento en los términos que se indican en la sección 3.2.
3. Poblar el grafo de conocimiento con el tesoro unificado y con las entidades de la plataforma, en los términos que se indican en la sección 3.3.
4. Publicar un buscador facetado y un portal navegable que permita consultar la ontología, desarrollado en la sección 3.4.
5. Implementar las interfaces que permitan la actualización de los tesauros y las entidades de Play, descrito en la sección 3.5.

Como segunda fase, y ya fuera de la financiación de los fondos europeos, el adjudicatario tendrá que incorporar las noticias en formato texto así como sus relaciones con categorías, vídeos y audios de todo tipo de verticales (informativos, radio, televisión, deportes, playz) dentro del grafo de conocimiento, procesando las noticias que RTVE tiene almacenadas dentro del CMS, y actualizando los elementos y relaciones de la fase 1 en consecuencia. En su oferta técnica el licitador tendrá que documentar la metodología, herramientas, procesos, técnicas y cualquier otra información relevante para llevar a cabo este proceso. Será objeto de valoración.

3.1 Unificación de tesauros y vocabularios

RTVE cuenta con dos conjuntos de vocabularios que provienen del CMS de Digital y del Fondo Documental (FD, en adelante). Para poder crear una sinergia entre ambos vocabularios, es necesaria su unificación.

Dado el volumen de términos, el licitador tendrá que unificar los cincuenta y un (51) tesauros con los que cuenta el Fondo Documental (con un total de casi un millón y medio de términos) teniendo en cuenta que los tesauros no sólo mantienen términos, sino también sinónimos, términos equivalentes, términos relacionados, notas de alcance, usado por, término general/específico, etc.

Durante el proceso de unificación es necesario que el licitador mantenga la correspondencia entre los términos unificados y los términos originales provenientes del FD. Asimismo, ha de mantener la información del término origen (incluidas las relaciones), así como la tipología de origen (ver anexo I) desde la que parte. Esta información es necesaria para los procesos de sincronización y exportación de los términos hacia el fondo documental.

Además de los tesauros mencionados, la parte digital cuenta con un volumen de más de seis mil (6.000) categorías y temáticas que también hay que unificar con los términos

anteriores (ya que existen redundancias). Los términos del tesoro unificado que existan o provengan del CMS han de incorporar la correspondencia con este origen, así como con la identificación de la categoría o temáticas del CMS de RTVE (idcategoria). Estos términos, así como sus respectivos mapeos, forman parte de los entregables de esta fase.

Finalmente, los términos unificados han de ser clasificados en la tipología de entidad ontológica de la ontología de referencia (o pivote) a la que corresponden, manteniendo la tipología de tesoro de la que procede (fechas, personas, lugares, temas, formas y series de programas).

El tesoro unificado ha de estar completamente desambiguado, no siendo admisible ningún porcentaje de ambigüedad. Además, el adjudicatario ha de entregarlo en un formato que será especificado por RTVE (preferentemente SKOS). Aunque los términos de los tesauros están en mayúsculas sin tildes, signos diacríticos, etc., el adjudicatario ha de devolverlo en texto enriquecido.

Será objeto de valoración la metodología propuesta por el licitador para la realización de este proceso, incluyendo la solución que plantea para solventar los mapeos con el FD y el CMS, qué información y atributos almacena, modelo de desambiguación de términos, la determinación y el enlazado con fuentes externas, el modelo de gestión de la calidad propuesto y los mecanismos de validación de la calidad de la clasificación.

Finalmente, será objeto de valoración la aportación de información ampliada, como por ejemplo la aportación de la url de foto/imagen para las personas, los lugares, etc.

3.2 Implantar la base tecnológica del grafo de conocimiento

Es objeto del proyecto es implantar una solución técnica para alojar, gestionar y explotar un grafo de conocimiento semántico que ha de disponer de las siguientes características mínimas:

- Ha de contar con herramientas de gestión de la ontología, definición de clases, atributos, equivalencias y el encadenamiento con otras ontologías.
- Ha de contar con capacidades de gestión de las entidades y sus relaciones tanto para equipos documentalistas a través de interfaces web, así como por medio de apis.
- Ha de contar con distintos niveles de tipología de usuarios, diferenciando el documentalista, el desarrollador o el administrador.
- Ha de contar con mecanismos de consulta de la ontología y de las entidades vía la URI que se le asigna, así como con mecanismos que faciliten la navegación a través de las relaciones, de las ontologías vinculadas, así como el descubrimiento de nuevas relaciones. A estos efectos el licitador ha de proponer un modelo de nombrado semántico acorde al modelo adoptado.
- Ha de contar con un sistema avanzado de consultas semánticas al estilo SPARQL que permita la ejecución de queries complejas y la extracción de información tanto individual como múltiple, así como las capacidades de hacer extracciones masivas de datos para enriquecer APIS. Será objeto de valoración tanto la riqueza de la extracción de datos, como las capacidades de inferencia semántica

que provea la solución.

El licitador ha de asumir que la arquitectura propuesta ha de tener una capacidad de al menos **100 peticiones concurrentes por segundo**.

El adjudicatario tendrá que incorporar a los productos online los distintos censales que RTVE utiliza para la medición analítica.

El licitador ha de entregar la arquitectura global de la solución identificando los mapas de red, mecanismos de aislamiento de las herramientas de backend (herramientas de authoring o de gestión), distribución y arquitectura de capas de front end y de API, sistemas de seguridad, etc. Si alguna pieza estará instalada fuera de los CPDs de RTVE, el licitador debe incorporar las líneas de comunicación para conectar con la DMZ de RTVE, y alternativamente las VPNs para la intercomunicación. El licitador en su propuesta técnica ha de describir toda la información indicada. En caso que alguna pieza se haya planteado en remoto (fuera de RTVE) y pueda suponer un riesgo de seguridad, RTVE podrá exigir sin coste que se haga en local. Además, si las comunicaciones antes mencionadas no tuviesen caudal suficiente, las latencias sean muy altas o dificulten la operativa, RTVE podrá exigir la instalación de líneas dedicadas que tendrá que instalar el adjudicatario sin coste, o la instalación en local. En cualquiera de los dos casos, el plazo máximo del que dispone es de 30 días naturales.

Como ya se ha comentado en la introducción, el licitador ha de adoptar un estándar de representación ontológico. RTVE prefiere que dicho estándar sea ebucoreplus, aunque el licitador puede plantear otro modelo, que será objeto de valoración. En cualquier caso, el licitador está obligado a clasificar las entidades según ebucoreplus, y al enlazado, según el caso, con Geonames, IMDb, Wikipedia y DBPedia.

Para la valoración de la propuesta técnica, el licitador ha de entregar en su propuesta la arquitectura técnica que propone, incluyendo el detalle de los entornos propuestos, dimensionamiento técnico que propone para el caso de RTVE, modelos de escalabilidad, elementos de seguridad, modelo de servicio, identificación de elementos técnicos esenciales, separando los elementos de front end y de back end, herramientas de soporte, la utilización de software libre, y cualquier detalle técnico relevante. Toda esta información será objeto de valoración.

Dentro de las capacidades que serán valoradas, el licitador debe documentar los mecanismos técnicos que incluye su solución para el modelo de similitud de entidades o instancias con la finalidad de montar recomendadores semánticos, valorándose especialmente los mecanismos que emplea la solución para este fin, la utilización de los conceptos de la ontología de referencia en modelos de recomendación, el rol de vocabularios, atributos y relaciones, la capacidad de adaptabilidad y set up de la solución y los algoritmos que se ponen a disposición.

3.3 Poblado del grafo de conocimiento

El adjudicatario, tras la unificación del tesoro, ha de poblar el grafo de conocimiento con el tesoro enriquecido, clasificado y linkado.

Además, de lo anterior, el adjudicatario ha de transformar las entidades de la plataforma RTVE Play (al menos audio, video, temporada y programa) a sus correspondientes

entidades dentro del modelo ontológico e instanciar sus atributos y relaciones. Además, el licitador ha de proponer cómo va a realizar la asociación de las entidades y atributos de RTVE Play con los elementos del tesoro unificado. Hay que tener en cuenta que las entidades del CMS cuentan con elementos textuales que hay que identificar con los correspondientes items del tesoro.

Será objeto de valoración la metodología, las técnicas y herramientas, así como el modelo de gestión de la calidad.

3.4 Publicación de buscador facetado y portal navegable de consulta de la ontología

Además del poblado de la ontología, será responsabilidad del licitador la creación de dos elementos de explotación de la información semántica.

Por un lado, RTVE quiere publicar la ontología para que cualquier persona o actor del sector pueda de forma pública consultar la información de la que disponemos. Para ello el licitador ha de proponer un modelo de publicación de la ontología que permita verla (vista de clases, atributos y relaciones), consultarla (vista instancias) así como navegar a través de ella. En su propuesta ha de incluir la definición del **portal web** navegable indicando la tecnología que propone, el modelo de servicio, la arquitectura, el modelo de seguridad y wireframes explicativos (o, en su defecto y si es que dispone, URLs explicativas).

La vista anterior muestra la información para expertos, pero el licitador también ha de implementar un buscador semántico facetado que permita la búsqueda textual, la búsqueda contextual, la búsqueda avanzada y generar páginas enriquecidas con la información encontrada. En su propuesta ha de incluir la solución técnica que propone, indicando los elementos tecnológicos, el modelo de servicio, la arquitectura, el modelo de seguridad y wireframes explicativos (o, en su defecto y si es que dispone, URLs explicativas).

Toda la información será objeto de valoración, teniéndose en cuenta, entre otras, la usabilidad, accesibilidad, profundidad y capacidad de búsqueda, el diseño de las páginas, la capacidad multidispositivo y la adaptabilidad al dispositivo, y, finalmente, la escalabilidad de la solución.

3.5 Implementación de interfaces para la actualización desde FD y RTVE Play

El licitador ha de implementar e implantar, de acuerdo con la Dirección de Sistemas y con la Dirección de Explotación y Desarrollo de Interactivos, los procesos de actualización y volcado de los tesoros, categorías del CMS y entidades de Play, y luego de noticias. Estas interfaces tienen como objetivo mantener actualizado el grafo de conocimiento a partir de las actualizaciones a nivel de tesoros, categorías y contenidos que se produzcan en los sistemas del Fondo Documental y del CMS una vez ya se haya poblado el grafo de conocimiento (sección 3.3).

Las distintas direcciones técnicas darán instrucciones precisas para poder realizar estos trabajos. En cualquier caso, el licitador tiene que tener en cuenta que se tratará bien del desarrollo de interfaces que reciban eventos desde cualquiera de los sistemas antes mencionados con la información para crear el/los elemento/s en el grafo de

conocimiento, bien de volcados semiestructurados con la información a añadir en el grafo de conocimiento que se ejecuten como crons. Los artefactos a implantar podrán ser instalados en servicios o sistemas de RTVE.

Con independencia de lo anterior, el licitador debe documentar en su propuesta las interfaces que propone, indicando las distintas capacidades de integración que su solución tiene. Si así lo requiere RTVE, las interfaces podrán ser ajustadas sin coste para cumplir con los estándares de seguridad o para cumplir los requisitos específicos que desde RTVE se determine. El licitador ha de tener en cuenta que la ubicación de los artefactos y procesos que reciban los eventos desde FD o CMS, al acceder a datos potencialmente sensibles, será en ubicaciones provistas por RTVE.

Adicionalmente, el licitador ha de tener en cuenta que la adición de nuevos elementos dentro del grafo, especialmente la entrada de nuevos términos, pueden estar sujetos a aprobación por parte de un gestor. Es por ello que la solución ha de proveer la capacidad de implementar workflows de aprobación que acordará con RTVE. El licitador en su propuesta técnica ha de documentar las capacidades de definición de workflows que su solución tiene.

Finalmente, y para facilitar las interfaces hacia los sistemas de RTVE, la solución ha de disponer de la capacidad de hacer volcados bajo demanda y volcados diarios y periódicos (bulks) de todos los términos, sus relaciones y atributos adicionales de las categorías ingestadas del CMS y de los términos que el grafo almacena, tanto totales como segmentados para cada uno de los tipos de tesoro (fechas, personas, lugares, temas, formas y series de programas) en el formato que especifique el FD.

Todos los elementos indicados en esta sección serán objeto de valoración.

3.6 Licenciamiento, conectividad

Todas las herramientas, productos y licencias que requiera la ejecución del proyecto ha de ser provista por el licitador. Todos los elementos que forman parte de la solución han de estar libres de costes para RTVE durante, al menos, cuatro años, y, el licitador estará obligado a dar soporte durante ese período de tiempo, y garantía de al menos un año tras la aceptación por parte de RTVE.

El licitador está obligado a indicar qué elementos de la solución existen, y cuáles va a desarrollar. De estos últimos, el adjudicatario está obligado a entregar todo el código fuente, así como los medios de desarrollo que desde RTVE se determinen. RTVE adquiere la propiedad industrial de esos desarrollos (incluso en el caso que sea sobre un producto del adjudicatario) y podrá tener plena disposición de los mismos. RTVE, si así lo decide, incluso puede poner a disposición del sector el código completo de los desarrollos que adquiere, así como hacer cuantas comunicaciones estime oportunas. En su caso, el adjudicatario no podrá hacer ningún tipo de comunicación ni comercialización salvo que RTVE lo apruebe de forma expresa.

La solución del grafo de conocimiento no requiere que todos los elementos de la solución tengan que estar instalados en la infraestructura de RTVE. Es admisible elementos de la solución que puedan prestarse en remoto/cloud de terceros sin coste alguno adicional. RTVE podrá determinar qué elementos de las propuestas técnicas

tengan que ser obligatoriamente instalados en su cloud híbrida, especialmente aquellos que puedan disponer de datos sensibles o supongan un riesgo de seguridad.

En el caso de disponer de elementos en remoto, el licitador tendrá que entregar un plan de conectividad para el acceso a esos servicios. RTVE podrá cambiar sin coste dicho plan de conectividad para garantizar el acceso controlado o para garantizar los caudales necesarios, e incluso, si es necesario, añadir líneas punto a punto con un caudal de al menos de 1Gbps redundado.

3.7 Gestión de la calidad y acompañamiento

La creación de un grafo de conocimiento es un trabajo complejo que requiere de un ajuste fino, y continuado. Es por ello que el licitador ha de incluir el plan de mejora continua, de actualización de elementos de la solución, y de gestión de la calidad del metadato y la clasificación, que será objeto de valoración

Además de lo anterior, RTVE requiere de formación específica, sin coste, tanto técnica como para documentalistas. El licitador ha de presentar un plan de formación que será objeto de valoración.

Finalmente, RTVE valorará el modelo de acompañamiento, supervisión y seguimiento. El licitador ha de entregar dicho modelo en la propuesta técnica.

3.8 Servicios profesionales

Durante el período de prestación, RTVE podría requerir de la ejecución de evolutivos para ajustar las capacidades de la solución a las necesidades del momento. Es por ello que el licitador ha de contar con un conjunto de profesionales expertos en las tecnologías que incluye la solución propuesta. Estos servicios profesionales se activan a petición de RTVE y solo se ejecutan en caso que RTVE apruebe el evolutivo. Como evolutivo que son, RTVE adquiere la propiedad industrial de los desarrollos que se realicen.

Los servicios profesionales se ejecutarán en remoto. En caso que la instalación de los evolutivos requiera de trasladar al equipo del adjudicatario a Torrespaña, los costes los tendrá que asumir el adjudicatario.

3.9 Fases del servicio

Los licitadores han de tener en cuenta de que disponen de un **máximo de cuatro meses** para tener la solución 100% operativa. RTVE valorará positivamente que el tiempo de implantación sea menor. A partir del cuarto mes, RTVE dispondrá del soporte en caso que sea necesario, así como la posibilidad de hacer peticiones de servicios profesionales en los términos antes indicados por un período de hasta cuatro años.

La inclusión de noticias se ejecutará durante 2024, y, el adjudicatario tendrá que terminar los trabajos antes del 1 de septiembre de 2024.

3.10 Coordinación de los servicios

El adjudicatario nombrará un coordinador de los servicios que será el contacto con RTVE durante la instalación del servicio y para los servicios profesionales.

Para el soporte, el adjudicatario ha de poner a disposición de RTVE un teléfono de atención en español en horario 24x7, así como una herramienta de ticketing para poder reportar incidentes.

3.11 Otros aspectos del servicio

El adjudicatario deberá ejecutar el proyecto con su propio material y equipo informático (tanto el hardware como todas las licencias necesarias de software de escritorio).

El servicio tendrá que ser prestado íntegramente en castellano.

CRTVE no admitirá la ejecución del servicio en oficinas que estén fuera de la Unión Europea: el adjudicatario deberá notificar el lugar de ejecución de las tareas y CRTVE tendrá que dar por válidas las ubicaciones.

Bajo ningún concepto CRTVE permitirá que la ubicación del servicio pueda suponer un problema para los activos digitales de CRTVE, ni que puedan suponer una potencial vulneración de la ley de protección de datos.

RTVE proporcionará conectividad a los entornos de trabajo a través de una VPN, que se configurará al principio de la prestación del servicio. RTVE, en función de la calidad de la conectividad, u otras incidencias de seguridad o que afecten a la prestación, podrá exigir la instalación de una línea dedicada punto a punto a las instalaciones del adjudicatario, sin coste para RTVE. En caso de ser requerido, el adjudicatario tendrá que proveer de esta línea en un plazo máximo de un mes.

4 Estructura solicitada para la propuesta técnica

La propuesta debe tener, como mínimo, la siguiente estructura:

4.1 Resumen Ejecutivo

Deben explicarse los aspectos más importantes de la propuesta: antecedentes, objetivos, beneficios, descripción. Su extensión no deberá ser superior a 5 páginas.

4.2 Oferta técnica

Para la evaluación técnica será necesaria la presentación de la oferta técnica incluyendo toda la documentación solicitada en este pliego de condiciones técnicas.

La oferta debe contener al menos los siguientes apartados:

- Descripción detallada de la solución propuesta, fases, cronograma, incluyendo detalle de la arquitectura de la solución, redundancias, APIS, modelo de alineamiento, mecanismos de traducción, productos empleados, licenciamiento, identificación de productos y desarrollos.
- Definición detallada de las capacidades de consulta de la solución, tanto en la



versión directa como la avanzada.

- Detalle de la información indicada en la sección 3.4.
- Plan de implantación de la solución y currículums de las personas que forman parte de la implantación.
- Plan de aseguramiento de la calidad, modelo de acompañamiento y plan de formación.
- Modelo de soporte y herramientas propuestas.
- Modelo de similitud semántica.

Madrid, abril de 2023