



## Aplicación de los principios de la Web Semántica y los Datos Vinculados a los archivos audiovisuales de televisión. Antecedentes y perspectivas

Concha Soler Monreal<sup>1</sup>; Isidoro Gil Leiva<sup>2</sup>

Recibido: 13 de noviembre de 2022/ Aceptado: 10 de marzo de 2023

**Resumen.** El artículo aborda la aplicación de las tecnologías de la web semántica en los archivos de televisión. Desde una perspectiva general, se plantea la implantación de los datos vinculados en los archivos audiovisuales de emisoras de televisión con el fin de optimizar y reutilizar recursos tanto para la gestión de la información como para la difusión de los productos multimedia del medio de comunicación. La propuesta se proyecta en todas las etapas de la cadena documental en dos direcciones: hacia el interior del archivo, para mejorar los metadatos, y hacia el exterior, a través de las publicaciones del medio, con el fin de evitar la dispersión de los espectadores mediante el control del acceso a segundas pantallas. Se relatan proyectos de datos vinculados desarrollados por medios de comunicación, así como por archivos audiovisuales y se proponen nuevas formas de reutilización de recursos.

**Palabras clave:** Web semántica; Datos vinculados; Archivos audiovisuales; Archivos de televisión; Cadena documental; Segundas pantallas; Metadatos.

### [en] Application of the principles of Semantic Web and Linked Data to TV broadcasting archives. Background and perspectives

**Abstract.** This article addresses the application of Semantic Web technologies to TV broadcasting archives. From a general perspective, an approach is presented for the implementation of Linked Data to TV broadcasting archives with the aim of optimizing and reusing resources both for the management of information and for the broadcast of the multimedia products of the media. The process is carried out at every stage of the documentary chain in two directions: inside the archive, to improve the metadata, and outside, through the media's publications, to control access to second screens and, thus, prevent the viewer from getting distracted. Linked Data projects developed by the media as well as by broadcasting archives are listed, and new methods for reusing resources are proposed.

**Keywords:** Semantic Web; Linked Data; Broadcasting archives; Archives; Documentary chain; Second screens; Metadata.

---

<sup>1</sup> Institut Valencià de Cultura  
E-mail: solermonreal@telefonica.net

<sup>2</sup> Universidad de Murcia  
E-mail: isgil@um.es

**Sumario.** 1. Introducción. 2. Antecedentes. 3. Precedentes en la documentación audiovisual. 4. Propuesta de aplicación de datos vinculados a la gestión de archivos audiovisuales. 5. Oportunidades y proyectos futuros. 6. Conclusiones. 7. Referencias bibliográficas

**Cómo citar:** Soler Monreal, C.; Gil Leiva, I. (2023), Aplicación de los principios de la Web Semántica y los Datos Vinculados a los archivos audiovisuales de televisión. Antecedentes y perspectivas, en *Revista General de Información y Documentación* 33 (1), 231-256

## 1. Introducción

Los profesionales de la información, debido al aumento incesante de carga de trabajo sin que los medios humanos lo hagan en la misma medida, se ven obligados a buscar nuevas fórmulas de optimización de la gestión tanto desde el punto de vista tecnológico como para integrarse en un mundo interconectado. La información estructurada procedente del análisis de los documentos audiovisuales precisa de esfuerzos que cada vez cuesta más asumir. Las piezas informativas y los programas de televisión en formato digital siguen requiriendo descripciones minuciosas para su recuperación posterior. Al mismo tiempo, las peticiones que realizan los usuarios de los servicios tienden a una mayor especificidad, o incluso son demandas de frases de audio concretas.

En este sentido, autores como López de Quintana (2019) señalan que la unidad documental no es ya la escena o la pieza informativa, sino el plano. López de Quintana y León-Carpio (2021b) destacan el hecho de que, si los recursos humanos no crecen en la misma proporción en que lo hace la producción audiovisual y la especificidad en las demandas, es complicado pensar en abarcar toda la información generada por el medio. Es necesario, pues, aprovechar los recursos existentes, así como permitir que la información fluya también hacia el exterior. Estos autores (López de Quintana y León-Carpio, 2021a) insisten en la necesidad de transformar la labor del documentalista para que dedique menos tiempo a la gestión de metadatos y resolución de demandas, cambiando el perfil hacia la gestión del conocimiento y la generación de contenido.

Por consiguiente, la labor del documentalista debe enfocarse de otra manera. Ya no debe ser un mero transcriptor y descriptor de imágenes y audio, sino que debe evolucionar y ampliar expectativas hacia la automatización, y la gestión del conocimiento.

Las experiencias de medios periodísticos que vinculan sus productos informativos a fuentes externas con la finalidad de ampliar y descubrir contenidos relacionados, de forma análoga a los formatos bibliotecarios adaptados a la interconexión de fuentes y datos, conducen a plantear si los recursos que ofrece la web semántica pueden aportar mejoras a la información periodística y más específicamente a la gestión de la información audiovisual.

También los profesionales de la información trabajan con datos vinculados. Son conocidas varias iniciativas de publicación de datos abiertos vinculados en el sector del patrimonio cultural, también conocido como GLAM (Galerías, Bibliotecas, Archivos y Museos) en los que la Biblioteca Nacional participa con varios

proyectos. La Secretaría de Estado de Digitalización e Inteligencia Artificial mantiene un portal ([datos.gob.es](https://datos.gob.es)) donde se explica la importancia de la publicación de los datos abiertos y su reutilización, además de ofrecer un catálogo de proyectos, acceso a conjuntos de datos y soporte para llevar a cabo los procesos. Allí se citan trabajos desarrollados en los sectores de turismo, medio ambiente, educación, salud, transporte, y el sector cultural. Para este último se señalan una serie de ventajas: alcanzar mayor audiencia, facilitar acceso único a colecciones dispersas geográficamente, limitar daños en la copia original, garantizar su sostenibilidad y, por último, dar posibilidad de crear nuevos productos y servicios de valor ([datos.gob.es](https://datos.gob.es), 2020).

Para Jean Gabriel Minel, jefe de Investigación y Desarrollo del INA (*Institut National de l'Audiovisuel*): "La Inteligencia Artificial (IA) no trata solo de catalogar y preservar, sino de conectar su contenido a las oportunidades" (Minel, 2019). Según este autor, los desafíos a los que se enfrentan las empresas televisivas son dos: el constante crecimiento de la cantidad de contenido que archivar, así como el mayor uso de contenido de archivo en la producción o distribución. Para hacer frente a ambos desafíos plantea la Inteligencia Artificial como determinante en el tratamiento de la información audiovisual y de los archivos, con herramientas cada vez más desarrolladas, entre las que se puede citar: el *speech to text*, la clasificación automática, el reconocimiento de voz y facial, la diarización de hablantes, o incluso el análisis de emociones. Así pues, se propone sumar a todo ello la integración de los archivos audiovisuales con la tecnología de la web semántica y los datos vinculados, tal y como hacen los formatos bibliográficos desde hace tiempo. En su aplicación se valorarán las dificultades que se pueden presentar en el proceso, así como sus ventajas.

Por lo tanto, el objetivo principal de este artículo es el desarrollo de una propuesta para aplicar la tecnología de datos vinculados y web semántica a los archivos audiovisuales de los centros de documentación de televisión. Mediante la implantación integral de las tecnologías semánticas en todas las etapas de la cadena documental se persiguen las siguientes finalidades: optimizar los procesos de catalogación, mejorar la recuperación de información e imágenes, y a la vez obtener un beneficio para el medio con recursos relacionados a través de sus páginas web y otros productos derivados.

El método utilizado parte de un enfoque descriptivo, estudiando distintos usos de los datos vinculados en los medios de comunicación. Más concretamente, se revisan antecedentes de implementación en archivos audiovisuales y su relación con las iniciativas, más asentadas, del mundo bibliotecario. Se ha examinado el proceso de generación de datos vinculados para compararlo e integrarlo con la cadena documental en sus operaciones de entrada, tratamiento y salida. A partir de la agregación de ambos procedimientos, se traza un marco teórico de aplicación valorando las opciones disponibles en los distintos aspectos del tratamiento documental de los materiales audiovisuales, prestando atención especial a las experiencias previas estudiadas y a las posibilidades de nuevos desarrollos.

## 2. Antecedentes

La aplicación de las tecnologías conocidas como *Linked Data*, de Datos Enlazados o Datos Vinculados surgió en 2002 con el objetivo de crear una Web Semántica que permita la conexión entre máquinas, o lo que es lo mismo, pasar de la web del hipertexto, de los documentos, a la web de los datos, los cuales se conectan e interrelacionan para crear algo mucho más grande cuyo rendimiento se verá a largo plazo (Berners-Lee, 2009).

El mundo de la web semántica abre oportunidades impensables hace unos años, entre las que se encuentra la posibilidad de enriquecer cualquier contenido con sus relaciones, las cuales pueden ser infinitas. En la definición que Hendler, Berners-Lee y Miller (2002) hacen de la Web Semántica ya apuntan a una red basada en datos vinculados: “La Web Semántica es una extensión de la Web actual en la que la información tiene un significado bien definido, lo que permite que las computadoras y las personas trabajen mejor en cooperación. Se basa en la idea de tener datos en la Web definidos y vinculados de modo que puedan usarse para un descubrimiento, automatización, integración y reutilización más efectivos en varias aplicaciones. Para que la Web alcance su máximo potencial, debe evolucionar hacia esta Web Semántica, proporcionando una plataforma de acceso universal que permita que los datos sean compartidos y procesados por herramientas automatizadas y por personas”. Y más adelante: “Para los documentos en nuestras vidas, vincular contenido es mucho más fácil que nunca”.

En resumen, la finalidad de la web semántica es convertir la *World Wide Web* en contenido estructurado como si fuera una gran base de datos global, con la que, a partir de los datos vinculados, se pueden encontrar otros datos relacionados. Tim Berners-Lee (2006) sentó las bases para poder publicar datos en la web en forma de datos vinculados con estas cuatro normas:

1. Utilizar URIs para nombrar cosas
2. Que esos URI sigan el protocolo HTTP para que las personas puedan encontrar esos nombres.
3. Que cuando alguien busca un URI, proporcione información útil, utilizando los estándares (RDF, SPARQL).
4. Proporcionar vínculos a otras URI, para que puedan descubrir más cosas de forma como se hace en la web clásica.

La expresión *Linked Data* (o Datos Vinculados) en la mayor parte de las ocasiones se asocia a *Open Data* (o Datos Abiertos), formando la expresión *Linked Open Data* (o Datos Abiertos Vinculados), así como su abreviatura LOD, aunque no necesariamente ha de ser así: el concepto de *Linked Data* se define como datos que “tratan de crear enlaces para que una persona o máquina pueda explorar la red de datos” (Berners-Lee, 2006), mientras que la expresión *Open Data* hace referencia a la “filosofía que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo en forma de datos accesibles y reutilizables” (*Open Knowledge Foundation*, s.f.). Es decir, para crear la Web

Semántica se necesitan enlaces, pero no se puede acceder a ellos si los propietarios no proporcionan acceso a los mismos. La tecnología *Linked Data* aplicada a los medios de comunicación podría denominarse *Linked Media Data*, y en el caso de incorporar una vocación de apertura, *Linked Open Media Data* (Ríos Hilario, 2014). O si se prefiere, Datos de Media Vinculados o Datos de Media Abiertos Vinculados.

Las tecnologías de la web semántica se están aplicando en las televisiones europeas desde la década de 2000. Las apuestas más significativas son las de la *BBC (British Broadcasting Corporation)* y la *EBU (European Broadcasting Unión)*. El germen de la implantación de la web semántica en la BBC comenzó en 2004, cuando se creó una web para cada programa de radio, estableciendo una url para cada uno. En 2007, se comenzaron a vincular los programas musicales con los identificadores de artistas de MusicBrainz (Kobilarov y otros, 2009). Ese mismo año, se creó una ontología para programas y se empezaron a publicar también los datos propios. Pero la aplicación de datos vinculados más conocida de la BBC es la creación de una página web para el mundial de fútbol de 2010 con información sobre equipos y jugadores, que culminó en 2012 con el sitio web de los juegos olímpicos (Rayfield, 2010). La plataforma de datos vinculados de la BBC, empleados en sus sistemas internos desde 2012, proporciona un amplio conjunto de ontologías y vocabularios controlados que se pueden utilizar para categorizar el contenido en una amplia gama de dominios.

The screenshot shows the BBC Sport website for the 2010 World Cup, specifically the page for the Spain team. The page features a navigation bar with 'SPORT', 'FOOTBALL', 'WORLD CUP 2010', 'GROUPS & TEAMS', and 'FIXTURES & RESULTS'. The main content area is titled 'Spain' and includes a list of 'Latest matches' with dates and opponents, a 'Latest stories' section with headlines, and a 'Latest audio and video' section. A table titled 'Group H Teams' shows the standings for Spain, Chile, Switzerland, and Honduras. A red circle highlights the 'Around the web' section, which contains links to 'Spain Fifa Profile', 'Spain Wikipedia profile', 'Spanish FA site', 'BBC Search - country page', and 'BBC Mundo (In Spanish)'. A red arrow points to the 'BBC Search - country page' link.

A	B	C	D	E	F	G	H
Group H Teams							
	W	D	L	GD	PTS		
Spain	2	0	1	2	6		
Chile	2	0	1	1	6		
Switzerland	1	1	1	0	4		
Honduras	0	1	2	-3	1		

Figura 1. Extracto de la página de la Selección Española de Fútbol del sitio web de la BBC del mundial de 2010 con enlaces a Wikipedia, FIFA, o la Federación Española de Fútbol. Fuente: [http://news.bbc.co.uk/sport2/hi/football/world\\_cup\\_2010/groups\\_and\\_teams/team/spain](http://news.bbc.co.uk/sport2/hi/football/world_cup_2010/groups_and_teams/team/spain) a través de Archive.org

Por su parte, EBU trabaja con web semántica desde 2008. En el año 2009 se presentó una implementación combinada de *EPG (Electronic Program Guide)* con ontologías de noticias. La ontología EBU Core, basada en Dublin Core se utiliza como referencia en el proyecto EUScreen y por *Europeana*.

Diarios como *The New York Times*, *The Guardian* o agencias de noticias como Thomson Reuters han sido de las primeras empresas periodísticas en publicar datos abiertos vinculados. *The New York Times* publicó por primera vez en 2009 su lista de encabezamientos de materia como datos abiertos vinculados. Desde ese año ha ido publicando ampliaciones del vocabulario periódicamente, así como varias API (*Application Programming Interface*) para permitir el enlace a sus datos. Aunque más cercano a la gestión del big data, *The New York Times* obtuvo el premio *Pulitzer* en 2021 por su proyecto de seguimiento del coronavirus a partir de la recopilación de datos de casi diez millones de fuentes de todo el mundo, proyectados diariamente en gráficos y mapas desde el inicio de la pandemia y que se encuentran disponibles en GitHub. Por su parte, *Thomson Reuters*, hace uso de tecnologías semánticas desde el año 2008, a través de un servicio llamado Calais. Una de las herramientas desarrolladas por Calais es la API abierta *SemanticProxy*, capaz de generar metadatos semánticos en RDF a partir de cualquier url, reconociendo personas, entidades, lugares, y eventos. En la actualidad, el servicio se aloja en <https://www.refinitiv.com/en> donde se puede ver una demo de etiquetado inteligente, simplemente introduciendo un texto en inglés, chino, japonés, alemán, español o francés. El sistema devuelve el texto con la extracción de etiquetas. Por ejemplo, copiando el texto de una noticia del diario El País titulada “Boeing vs Space X: la carrera hacia la órbita baja”, publicada el 19 de mayo de 2022 extrae las etiquetas que se pueden ver en la figura 2.

The screenshot displays the SemanticProxy interface. On the left, under 'FOUND IN DOCUMENT', there is a list of categories and their associated entities with relevance bars:

- ENTITIES** (Relevance: ?)
- Company**
  - Elon Musk (20%)
  - THE BOEING COMPANY (40%)
- Country**
  - Estados Unidos (20%)
- Currency**
- Natural Feature**
  - Cabo Caňaverar (20%)
- Organization**
  - National Aeronautics and Space ... (20%)
- Person**
  - Jacques Cousteau (20%)

On the right, under 'DOCUMENT VIEW', the text of the article is shown with highlighted entities:

**Boeing vs Space X: la carrera hacia la órbita baja**

El lanzamiento de la nueva cápsula tripulable Starliner, realizado a las 18.54, hora de la **costa este en EE UU** (00.54, hora peninsular española) de este jueves, es el segundo intento de certificar el esfuerzo de **Boeing** por competir con Space X en el transporte de astronautas a órbita. El primero fue a finales de 2019, cuando ambas empresas iban codo con codo en esa carrera. Pero, entonces, la misión no fue bien. La Starliner (bautizada Calypso, como el buque oceanográfico de **Jacques Cousteau**) entró en órbita, pero no consiguió acoplarse a la Estación Espacial Internacional. Problemas de software.

Hizo falta año y medio para resolver todos los inconvenientes que presentaba la primera versión de la cápsula. El segundo vuelo se planificó para el verano de 2021, pero, cuando todo parecía a punto y el vehículo ya estaba en la plataforma de lanzamiento, se descubrió un nuevo inconveniente: 13 de las válvulas del sistema de propulsión estaban atascadas.

La **NASA** aplaza el viaje a la Luna previsto para 2024

La causa del problema se atribuyó al clima casi tropical de **Cabo Caňaverar**. No por la humedad en sí, sino porque el **mar de su costa** es un componente del ambiente. Debido a él, se prevé que se produzcan fenómenos de este tipo.

Figura 2. Etiquetado automático de una noticia del diario El País.

Fuente: <https://www.refinitiv.com/en>

También hacen uso de los datos vinculados servicios populares entre los usuarios de la red. Uno de ellos es *Spotify*, que ofrece una API para desarrolladores que permite conectar con metadatos musicales avanzados, crear reproductores específicos o recomendaciones basadas en las características de los audios, como por ejemplo tempo, vivacidad, o “bailabilidad”.

```
const endpoint = "https://api.spotify.com/v1/recommendations";
const artists = '6sFIWsNpZYqfjUpaCgueju';
const danceability = encodeURIComponent('0.9');

fetch(`${endpoint}?seed_artists=${artists}&target_danceability=${danceability}`, {
  method: "GET",
  headers: {
    Authorization: `Bearer ${userAccessToken}`
  }
})
.then(response => response.json())
.then(({tracks}) => {
  tracks.forEach(item => {
    console.log(`${item.name} by ${item.artists[0].name}`);
  })
})
})
```

Figura 3. Aplicación para personalizar las recomendaciones.

Fuente: [https://developer.spotify.com/?\\_ga=2.128826316.177463250.1652891843-261397917.1652891843](https://developer.spotify.com/?_ga=2.128826316.177463250.1652891843-261397917.1652891843)

En el ámbito de la gestión de la información, la *Library of Congress* puso en marcha en 2012 BIBFRAME, un modelo de descripción bibliográfica basado en la tecnología de datos abiertos vinculados que aspira a sustituir al formato MARC. A su vez, el estándar de MARC y sus versiones a lo largo del tiempo se crearon con la finalidad de compartir e intercambiar información superando barreras idiomáticas.

El reciente informe publicado por *OCLC Research Library Partners Metadata Managers Focus Group* (Smith-Yoshimura, 2020), dedicado al cambio del tratamiento de los metadatos en los catálogos bibliográficos, señala el cambio fundamental en los metadatos que nutren las bases de datos. Ya no serán “cadenas de caracteres” sino identificadores permanentes que enlazarán los contenidos a otros datos, vocabularios e informaciones, enriqueciendo el contexto, superando barreras idiomáticas y disciplinarias.

Salvando las distancias, la reutilización de datos, aunque no se trate de un contexto científico sino mediático, entronca con la filosofía FAIR (*Findable, Accesible, Interoperable and Reusable*), es decir, hacer que los datos sean encontrables, accesibles, interoperables y reutilizables. Para ello se necesita que datos y metadatos puedan ser encontrados a través de identificadores únicos y

persistentes, mediante protocolos estandarizados, abiertos, universales, y gratuitos. Además, deben ser interoperables para permitir el intercambio y la reutilización:

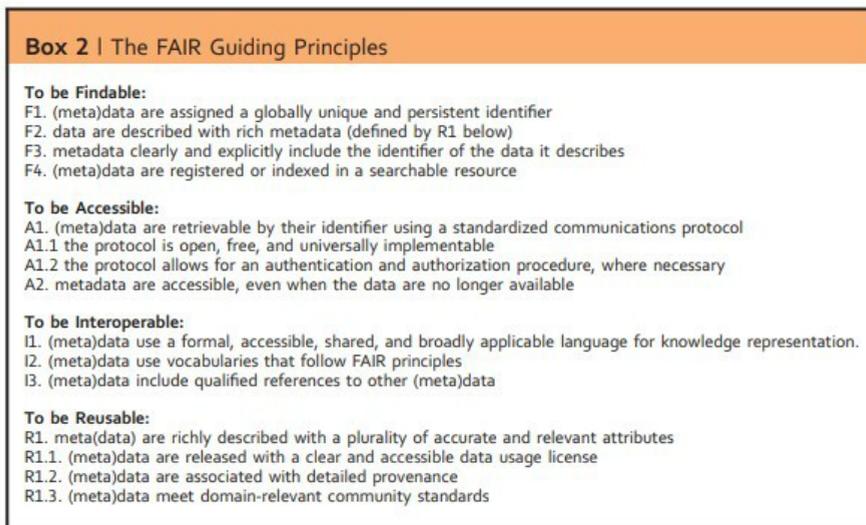


Figura 4. Principios FAIR. Fuente: Wilkinson y otros, 2016.

Evidentemente, se trata de contextos diferentes, pero coincidentes en cuanto que ambos tratan de reutilizar los datos. La filosofía de FAIR, desarrollada en pro de la ciencia abierta, comparte aspectos de los LOD, en este caso, aplicados a los medios de comunicación, y enlaza con el programa Horizonte 2020 de Integración e Innovación de la Unión Europea (Datos.gob.es, 2017), exceptuando aquellas cuestiones afectadas por derechos de autor, que también reconoce el lema fundamental de los principios FAIR: “tan abierto como sea posible, tan cerrado como sea necesario”. Los principios FAIR, además de los datos, abarcan a los algoritmos, herramientas y flujos de trabajo que llevan a esos datos, lo cual sintoniza perfectamente con la tecnología de los Datos Abiertos Vinculados<sup>3</sup>.

### 3. Precedentes en la documentación audiovisual

Hasta ahora se ha expuesto cómo se empezaron a utilizar los datos vinculados en medios periodísticos y televisivos. A continuación se exponen aplicaciones relacionadas específicamente con la documentación audiovisual y televisiva: los proyectos EUscreen y LinkedTV, junto la adaptación del modelo BIBFRAME al esquema de metadatos PBCore el portal TIB/AV y la adaptación en el ámbito cinematográfico.

El repositorio *EUscreen* (<https://euscreen.eu/about.html>), llamado *EUscreenXL* a partir de 2013, que proporciona acceso a una selección de documentos del

<sup>3</sup> En esta línea, Poveda-Villalón, Espinoza Arias, Garijo y Corcho 2020 defienden tratar las ontologías como objetos de investigación.





Figura 6. Aplicación de contenido relacionado con LinkedTV en un programa de noticias de la emisora alemana RBB. Fuente: Nixon, s.f.

Por tanto, la información ofrecida por los medios a través de Internet, bien de actualidad, bien sobre programas en emisión, se puede adaptar a este modelo sugiriendo al usuario la consulta de datos o enlaces con información adicional. Por ejemplo, ofreciendo biografías de los actores de una serie, o datos sobre su temática, introduciendo servicios de valor añadido sin que los espectadores accedan a fuentes de información externas a las del propietario del contenido.

También se sitúa en la línea de la documentación audiovisual el trabajo de Strobel y Marín-Arraiza (2015) sobre el portal TIB/AV, en donde se establece un DOI (Digital Object Identifier) para cada vídeo de contenido científico alojado en el repositorio. El OCR de vídeo recupera textos de las imágenes y también se transcribe el audio a texto. Los metadatos formales y técnicos son descritos manualmente. El metadato automático se realiza sobre imagen, voz y segmentación de vídeo. El OCR de vídeo transcribe texto escrito y diapositivas. El sistema es capaz de identificar conceptos visuales como conferencia, imagen de microscopio, detalle de fachada, animación por ordenador, etc. Con los metadatos textuales extraídos del texto se etiqueta según una lista de encabezamientos de materia. La recuperación se realiza mediante una ontología bilingüe alemán-inglés.

Inspirado en la idea de adaptar el modelo de datos de BIBFRAME al ámbito de la descripción de imágenes en movimiento se llevó a cabo un proyecto en el seno de la Biblioteca del Congreso de los Estados Unidos (Pierce, 2020). Se intentó adaptar el esquema de metadatos PBCore (estándar diseñado para facilitar el intercambio de registros descriptivos entre emisoras públicas de los Estados Unidos) al de BIBFRAME en la descripción de un conjunto de datos perteneciente a la colección de la National Educational Television. En concreto, se seleccionaron registros de una serie de ficción y otra serie documental de no ficción. Las lagunas de datos encontradas se solventaron con propiedades específicas de EBU Core. Además, el proyecto intentó la conciliación con la Lista de Encabezamientos de Materia de la Biblioteca del Congreso (LCSH), con el Archivo de Autoridad de Nombres de la Biblioteca del Congreso (LCNAF) y con los identificadores de EIDR (*The Entertainment Identifier Registry*), todos ellos publicados como datos vinculados. El estudio está en la línea de otros trabajos de adaptación de formatos, como los de AMIA (The Association of Moving Image Archivists) para

transformar PBCore a MARC o de la propia Biblioteca del Congreso (Library of Congress) para adaptar MARC 21 a BIBFRAME 2.0.

En el entorno cinematográfico, *Linked Movie Database*, LMDb, (<https://www.cs.toronto.edu/~oktie/linkedmdb/>) relaciona filmes con información sobre actores, directores, bandas sonoras, etc., con datos estructurados en RDF con una estructura de tripletas: sujeto, verbo y predicado y con enlaces a fuentes externas: *DBPedia*, *Freebase*, *RottenTomatoes*, *Geonames*, *IMDb* y *MusicBrainz* entre otras.

En 2015 la comisión de catalogación y documentación de la FIAF (International Federation of Film Archives) planteó por primera vez la aplicación de los datos vinculados a los archivos filmicos. Ante el cambio de paradigma que supone la adopción de los datos vinculados, se creó un grupo de trabajo formado por archiveros e informáticos (Heftberger, 2019). Sus conclusiones resultan muy útiles para emprender una tarea análoga en los archivos audiovisuales de televisión.

#### **4. Propuesta de aplicación de datos vinculados a la gestión de archivos audiovisuales**

Cuando los documentos audiovisuales eran contenidos en soportes analógicos, la única manera de acceder a contenidos concretos era a través de instrumentos creados al efecto: índices, catálogos, y, sobre todo, bases de datos referenciales. Desde el momento en que los soportes físicos dejan paso a material accesible por ordenador, su descripción empieza a formar parte indisoluble de los mismos a través de los metadatos. Los metadatos siempre se han definido como “datos sobre los datos” (Senso y Rosa-Piñero, 2003), es decir, son elementos descriptivos de objetos digitales, que proporcionan información semántica sobre recursos con la finalidad de poder ser recuperados. Permiten interactuar la inteligencia humana con la artificial, categorizando esos recursos mediante estándares de descripción y lenguajes de etiquetado. Lo que contiene el MAM (*Media Asset Management*) son los objetos digitales junto a los modelos de datos que permiten su descripción.

La complejidad de la estructura de los esquemas de metadatos asociados a los documentos audiovisuales a la hora de plantear una estrategia de integración de la web semántica junto con la tecnología de datos abiertos vinculados lleva a planificar una implementación en todos los procesos de la cadena documental. Por tanto, se propone crear un modelo conceptual para el análisis y la catalogación de material basado en LOD. Se trata de enriquecer los documentos audiovisuales que se conservan en los archivos de televisión, así como de todos los productos derivados del mismo, de manera análoga al formato BIBFRAME para bibliotecas. Es decir, conectar los documentos con sus autores, con su temática, relacionarlos con información complementaria, e incluso plantear su publicación con información añadida que puede ser ofrecida a través de la página web del medio al que sirve, aplicable tanto a la información de actualidad como a grabaciones de ficción. De modo que, por un lado, los elementos relacionados con la ficción formen parte de la narrativa transmedia y, por otro, se contribuya a lograr un periodismo semántico, un tipo de periodismo tecnológicamente avanzado basado

en las tecnologías de la web semántica que permitirá que los metadatos legibles por máquina de una publicación se puedan reutilizar y vincular a artículos semánticamente conectados (Panagioditis y Veglis, 2020).

Así pues, se plantea una estrategia en dos direcciones: por un lado, hacia el archivo en sí mismo, la descripción documental y los metadatos, y, por otro lado, hacia el medio en el que el archivo desarrolla su acción.

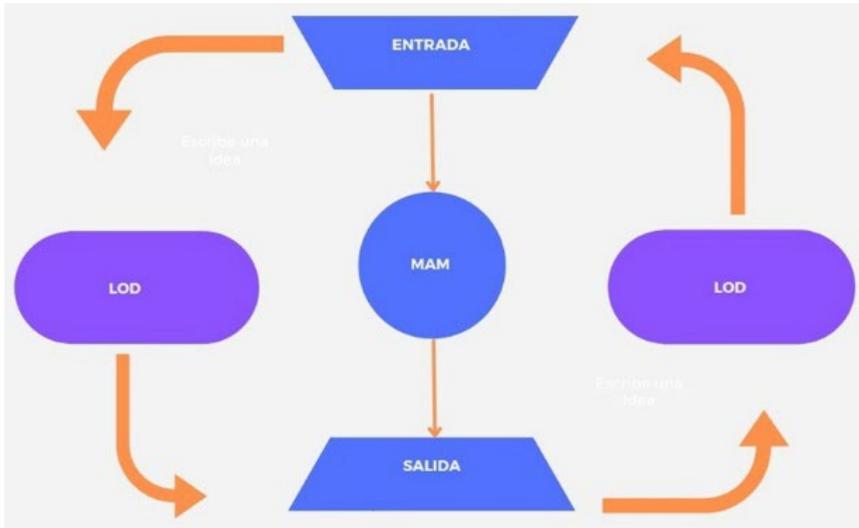


Figura 7. Integración de la cadena documental con datos vinculados.

Fuente: Elaboración propia.

#### 4.1. Aplicaciones de las tecnologías de la web semántica hacia el interior del archivo

El proceso de generación de datos vinculados, al igual que el documental, consta de una serie de tareas, según exponen Radulovic y otros 2015. Cada una de ellas contiene elementos de entrada o inputs, y elementos de salida u outputs. El procedimiento se compone de ocho tareas: selección de la fuente de datos que se transformará en datos vinculados, obtención de acceso a la misma, análisis de la licencia de uso, especificación de una estrategia de nomenclatura de recursos, desarrollo de una ontología para describir los datos y su transformación en formato RDF. Finalmente, los datos generados se vinculan con otros conjuntos de datos vinculados. La figura 8, muestra el proceso de generación de datos vinculados, de modo que las tareas consecutivas se representan con líneas completas, y las no consecutivas mediante líneas discontinuas. Los productos finales son representados con líneas dobles.

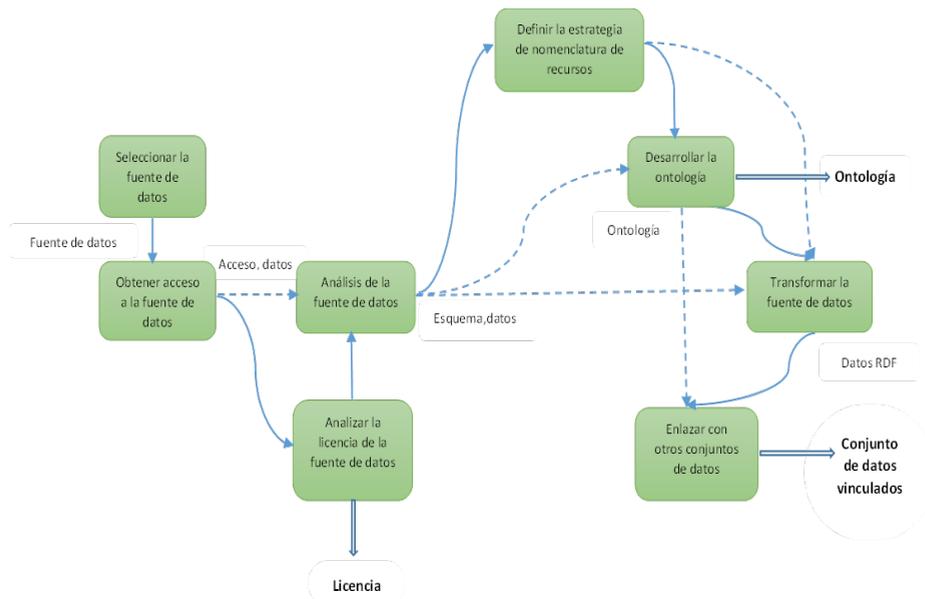


Figura 8. Proceso de generación de datos vinculaos.  
Fuente: Radulovic y otros, 2015.

El modelo conceptual que se propone para transformar los datos de un archivo audiovisual de televisión a datos vinculados quedaría integrado en las operaciones habituales de entrada y salida de información centrados en el MAM (*Media Asset Management*), el sistema gestor de los *assets* u objetos de media junto con su descripción.

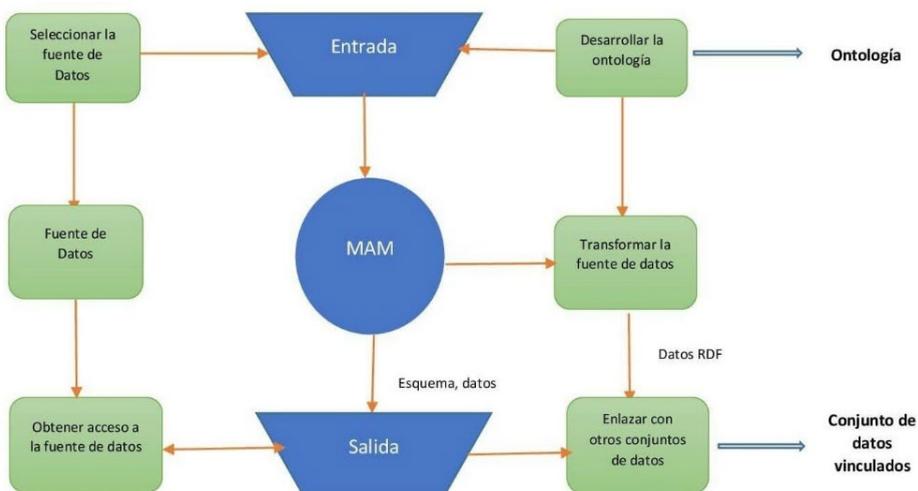


Figura 8. Proceso de elaboración de datos vinculados en un archivo audiovisual.  
Elaboración propia a partir de Radulovic y otros 2015

Los MAM (*Media Asset Management*) de los archivos de televisión contienen dos tipos de información: la información de registro de materiales, y la información del contenido.

Antes de la digitalización estos datos estaban separados en bases de datos referenciales con contenido diferenciado. En la primera se registraban datos de descripción física, mención de responsabilidad, y a veces incluso un breve resumen, junto con datos de gestión, de emisión y de reemisión, propiedad de los derechos y distribución. En la segunda, en cambio, además de la información de registro de materiales se describe el contenido, añadiendo descriptores referentes a personas o entidades, a materias y a lugares.

En el primer caso se incluyen todos los materiales emitidos por la cadena y en algunos casos también los preparados para su difusión en todas sus versiones (rotulada, limpia, doblada, versión original), informativos completos, documentales, retransmisiones, series y películas. Forman parte del segundo grupo los programas susceptibles de ser reutilizados (en todo o en parte) o explotados comercialmente, para lo que es necesario describir su contenido mediante identificadores (onomásticos, temáticos y geográficos) y, cuando es preciso, también un minutado más o menos exhaustivo. En este grupo entrarían las piezas de los informativos, los brutos o grabaciones de rodaje, el material procedente de agencias de noticias o de intercambio, algunas retransmisiones y series de producción propia con contenido documental o de interés.

### **Aplicación de los datos vinculados enfocada al registro de materiales: identificadores persistentes y ontologías para la gestión de los derechos de autor.**

El objetivo es aplicar la tecnología de datos vinculados para ambos grupos, pero, como concluyeron los expertos de *FIAF* (Heftberger, 2019), con una implantación paulatina, estableciendo procesos técnicos automatizados, comenzando por los registros filmográficos y los registros de autoridad y en una fase posterior abordar la indización de materias, los vocabularios controlados y la información sobre las colecciones. Por esa razón, se propone comenzar en una primera fase por el primer grupo de documentos audiovisuales: aquellos de los que hasta ahora solo se introducían datos de registro (metadatos técnicos, autoría, responsabilidad, propiedad de derechos y datos de emisión).

Del mismo modo, el primer paso para poner en práctica la web semántica y conseguir la interoperabilidad es contar con un sistema de identificadores persistentes o PID (*Persistent Identifier*), representados en forma de URI, con una institución que los sustente y seguir los principios FAIR, es decir, enlaces estables, fiables, permanentes, unívocos y persistentes. Mediante los PID se pretende “proporcionar una referencia duradera a una entidad (física, digital o abstracta), por ejemplo, un documento digital, un sitio web, una persona o una institución” (Marín-Arraiza, 2022). Generalmente los PID se utilizan para identificar publicaciones, conjuntos de datos e investigadores individuales, como por ejemplo los DOI que identifican objetos de investigación, o los números ORCID, que

identifican investigadores. Se pueden definir 5 tipos básicos de identificadores persistentes (Dappert, Farquhar y Kotarski, 2017): a) Agentes: personas individuales, organizaciones y entidades, b) Recursos: publicaciones, datos y productos de investigación, c) Declaraciones de derechos: acuerdos, licencias, patentes, d) Eventos: procesos relevantes para la procedencia de los recursos, como creación, curación, acceso, reclamación, actualizaciones, citas, revisión, e) Entidades derivadas: como proyectos, que pueden verse como una agregación de la entidad legal involucrada, la organización que la financia, los productos producidos y las declaraciones de derechos que la rigen.

Estos cinco tipos se pueden extrapolar a los identificadores de material de televisión. En la industria de los contenidos audiovisuales existe una variedad de números normalizados, como ISLI, ISAN, ISNI, ISWC, ISRC o EIDR que sirven para desambiguar nombres de autores y títulos y versiones de la misma obra. Muchos de ellos, al estar normalizados como DOI, URI, o IRI (según se muestra en la tabla 1) se pueden adoptar como PID.

Tabla 1. Identificadores estandarizados.

NOMBRE	NORMA	CARACTERÍSTICAS
VIAF (Virtual International Authority File) <a href="http://viaf.org">http://viaf.org</a>		Base de datos de autoridades creada por OCLC y varias bibliotecas nacionales. Se enlaza mediante IRI.
ISNI (International Standard Name Identifier) <a href="https://isni.org">https://isni.org</a>	ISO 27729:2012	Identifica unívocamente identidades relacionadas con la cadena de creación: persona natural, persona legal o personaje de ficción. Los códigos ORCID son un bloque de ISNI reservados a investigadores.
ISLI (International Standard Link Identifier) <a href="http://www.isli.org">www.isli.org</a>	ISO 17316:2014	Identifica enlaces entre entidades que pueden ser documentos, recursos multimedia, personas o elementos más abstractos como tiempos o lugares.
ISAN (International Standard Audiovisual Number) <a href="http://www.isan.org">www.isan.org</a>	ISO 17316:2015	Código internacional que identifica registros audiovisuales y todas sus variantes
EIDR (The Entertainment Identifier Registry) <a href="https://www.eidr.org/">https://www.eidr.org/</a>	No es en sí mismo un estándar ISO, pero al ser miembro de DOI, se rige por la norma ISO 26324:2012	Sistema de identificación único global para una amplia gama de objetos audiovisuales, incluyendo películas y programas de televisión y radio. Identifica cada versión de una obra y sus derivados. No persigue sustituir a los sistemas de identificación existentes sino incluirlos. Es compatible con múltiples sistemas como EBUCore, MARC, ISAN... Los registros están referenciados por DOI, que puede ser representado por URI.
MRL (Media Resource Locator)		Es un URI usado para identificar y localizar de manera unívoca recursos multimedia. Utilizado por reproductores multimedia y la aplicación de Java Media Framework.

Fuente: elaboración propia

Al igual que el grupo de trabajo de FIAF (Heftberger, 2019), se recomienda utilizar EIDR dada su flexibilidad e interoperabilidad con múltiples sistemas, en particular con EBU Core, en la línea del esquema conceptual planteado.

Para finalizar la primera fase, relativa a los datos registro y descripción, es necesario hacer referencia a la gestión de los derechos de autor y distribución, tarea esencial en los archivos audiovisuales de televisión. En este sentido, se propone la adopción de la ontología Media Contract Ontology (MCO). Creada por la RAI y otras televisiones europeas a partir de un glosario de términos, fue concebida para relacionar los derechos de la propiedad intelectual con las condiciones para su explotación y forma parte del estándar de MPEG-21 (FIAT-IFTA, 2019).

### **Aplicación de los datos vinculados enfocada a la descripción semántica del contenido.**

En la segunda fase se abordaría la descripción del contenido. El uso de vocabularios estandarizados, junto con los identificadores permanentes y las descripciones RDF son los pilares básicos de las tecnologías LOD (Marcondes, 2018). En las operaciones de salida es el momento de definir una ontología, que sirve para describir conceptos del dominio y sus relaciones y, transformar la fuente de datos en formato RDF, para, posteriormente, permitir las conexiones con otros conjuntos de datos. Aplicado al esquema propuesto, normalizaría y homogeneizaría conceptos y relaciones de los metadatos descriptivos (temáticos, onomásticos y geográficos) para el enlazado de recursos y la reutilización. La incorporación de las ontologías a la descripción semántica del contenido, además de necesarias para la tecnología LOD y el desarrollo de la web semántica, son fundamentales para la superación de los problemas generados por el lenguaje natural y son de utilidad para la gestión. Las ventajas del uso de ontologías son múltiples:

Tabla 2. Ventajas de uso de ontologías.

Problemas del lenguaje natural	<ul style="list-style-type: none"> <li>- Superan problemas de sinonimia y polisemia</li> <li>- Salvan barreras idiomáticas</li> <li>- Son necesarias para el control de autoridades</li> <li>- Crean relaciones semánticas más ricas y relaciones de inferencia</li> </ul>
Gestión	<ul style="list-style-type: none"> <li>- Vence ambigüedades en cuanto a identificación de obras, versiones, autoría, etc.</li> <li>- Ayudan a superar la escasez de recursos</li> <li>- Evitan la duplicación de esfuerzos</li> <li>- Favorecen la interoperabilidad</li> <li>- Favorecen el procesado automático</li> <li>- No requieren de gran inversión en software o infraestructura</li> </ul>

Fuente: elaboración propia

Para lograrlo, se presentan dos escenarios posibles: o bien adaptar un tesoro o vocabulario controlado al lenguaje ontológico, o bien adoptar una ontología existente. Debido a la naturaleza enciclopédica de la información de actualidad y a

la complejidad de crear una ontología o adaptar un tesoro, se considera conveniente comenzar con una taxonomía ya existente. A continuación, se recogen distintas opciones de taxonomías y ontologías disponibles:

- Media Topics (<http://cv.iptc.org/newscodes/mediatopic>) de IPTC (*The International Press Telecommunications Council*). Esta taxonomía puede ser un buen punto de partida por su clasificación jerárquica de hasta 5 niveles y 17 categorías semánticas, junto con su conexión con Wikidata. Publicada por primera vez en 2010, con actualización anual, está disponible en 11 idiomas. Está construido con SKOS ampliado y puede convertirse en RDF/XML, RDF/Turtle y JASON-LD.
- Audiovisual Semantics, AVS (Pastor-Sánchez, Bazán-Gil y Escribano, 2019), es una ontología elaborada específicamente para la descripción de contenido audiovisual, utilizando vocabularios OWL y SKOS, partiendo de un modelo conceptual en el que las piezas audiovisuales están compuestas por escenas, dentro de las cuales se pueden identificar acontecimientos, la entidad mínima que puede ser objeto de descripción. Además de objetos y acciones, se incluyen descriptores temáticos, onomásticos y geográficos.
- Dublin Core Metadata Elements: conocido también como The Dublin Core, es un vocabulario para la descripción de recursos del DCMI (Dublin Core Metadata Initiative) basado en quince elementos y expresado en RDF. Última versión disponible en: [www.dublincore.org/specifications/dublin-core/dcmi-terms/](http://www.dublincore.org/specifications/dublin-core/dcmi-terms/)
- Schema.org (<https://schema.org/>): Proyecto conjunto de Google, Microsoft, Yahoo y Yandex para promover el uso de datos estructurados en la web. Usado por 10 millones de sitios. Sus descripciones incluyen trabajos creativos como series de televisión o películas.
- TVRadioSchema: iniciativa conjunta de la BBC, EBU y Google para extender Schema.org a programas de televisión y radio.
- *Dbpedia* (<https://es.dbpedia.org/>): Proyecto inicialmente creado a partir de la Wikipedia en inglés, que persigue transformar la conocida enciclopedia web en datos semánticos expresados en RDF. A partir de 2011 el proyecto se abrió a otros idiomas, en la actualidad en 15. Auspiciado por la Universidad Politécnica de Madrid en su versión en español se han extraído doscientos millones de tripletas.
- *Wikidata* (<https://www.wikidata.org/>): Base de datos secundaria de Wikipedia, libre, colaborativa y multilingüe. Creado en 2012 por la fundación Wikimedia, identifica elementos de Wikipedia a los que se identifica con un número y se les agrega valor mediante declaraciones en forma de pares clave-valor. Centraliza los datos de la Wikipedia en todos los idiomas, unificando los enlaces.



(taxonomías, sistemas de clasificación, tesauros, folksonomías y similares) en formatos RDF, lo que permite la creación y publicación de conceptos en la web y vincularlos en otros esquemas (WC3 Working Group Note, 2009). En cuanto a software, la Universidad de Stanford mantiene el editor Protegé, gratuito y de código abierto, para la gestión de ontologías (<https://protege.stanford.edu/>).

## 4.2. Aplicaciones de las tecnologías de la web semántica hacia el exterior del archivo

Hacia el exterior del archivo y en relación con el medio en el que se inserta el archivo, los datos vinculados pueden gestionarse en tres vertientes:

- Añadiendo valor a las producciones audiovisuales (series o películas) para enriquecer la información de las emisiones en las EPG o guías electrónicas de programación, lo cual también es importante porque contribuye a que los buscadores tipo *Movistar* o *Netflix* "aprendan" de nuestros gustos en base a las búsquedas previas.
- Implementando servicios de valor añadido en la web del propio medio, ofreciendo información fiable y enriquecida, o lo que es lo mismo, dotar a la información textual y visual (noticias y vídeos) de elementos relacionados.
- Controlando el acceso a las segundas pantallas.

En el primer caso, la ontología OntoTV (Redondo, Botón Fernández y Lozano-Tello, 2012) recopila información de distintas fuentes y emisoras (Internet, televisión convencional, o plataformas digitales) para publicar información más amplia que la de una *EPG* en forma de datos vinculados basándose en la ontología *BBC Programmes* de la *BBC*.

En cuanto a la segunda de las aplicaciones propuestas, sería necesario dotar de identificadores persistentes a los vídeos de programas y noticias alojados en los servidores web del medio con el fin de poder establecer conexiones con otros conjuntos de datos y que a su vez puedan ser vinculados. Con la tecnología LOD pueden realizarse representaciones visuales de contenido y datos que ayuden tanto a redactores como documentalistas. Por ejemplo, se podría dibujar un gráfico temático que muestre de qué tratan las noticias de los informativos, trazar un mapa de dónde se originan más noticias (o en qué comarcas, si se trata de una televisión autonómica o local), de forma parecida a cómo la *BBC* tomó datos de las ubicaciones de las grabaciones de archivo y las cruzó con una aplicación de mapas, aunque también se puede realizar la operación inversa y tomar assets relacionados con un punto del mapa para permitir a los usuarios navegar por los datos sin un plan de recuperación preestablecido (Alexander, 2012).

En tercer lugar, íntimamente relacionado con lo anterior, es el control de las segundas pantallas, aludido con anterioridad al mencionar el proyecto *LinkedTV*, que proporciona un método para establecer vínculos hacia el exterior del archivo y de los contenidos de la web del medio televisivo. En un mundo en el que los medios se desdibujan y se hacen cada día más híbridos y en el que los contenidos

se piensan para ser emitidos por distintas vías (televisión convencional, redes sociales, web, etc.), los espectadores suelen interactuar con varios dispositivos a la vez, participando con comentarios en redes sociales, buscando información complementaria sobre el programa o ficción en emisión. Para evitar que la dispersión de la atención del espectador acabe en el abandono del medio propio, es fundamental el control de estas segundas pantallas. Se consigue ofreciendo a los consumidores un producto de valor añadido relacionado con los programas sin tener que recurrir a fuentes externas, mediante enlaces a contenidos web relacionados con el producto en emisión, ya se trate de noticias o ficción, ofreciendo al espectador información extra, tanto en modo push como en modo pull. Esto es, de forma pasiva, sin que lo solicite el usuario, bien de forma activa, cuando sea el televidente quien desee acceder a contenido relacionado.

Una de las formas de control de segundas pantallas muy utilizada en los últimos tiempos son los códigos QR. Aunque se inventaron hace años, han resurgido como elementos gráficos en programas, concursos, publicidad y sobre todo en informativos. Sobreimpresionados en la imagen para ser escaneados por dispositivos móviles, conducen al espectador a páginas web con información adicional. En España empezó a utilizar este recurso el grupo Atresmedia en los especiales de la noche electoral norteamericana en noviembre de 2020. Como apunta Borja Terán (2020), el uso del código QR se reintrodujo en la vida cotidiana a través de las cartas de los restaurantes, para evitar tocarlas debido a la pandemia de Covid-19. En los grafismos de televisión se utiliza como símbolo de valor añadido y modernidad, sin ser nuevo, y aunque, según este autor, no importa mucho si el espectador realmente hace uso de él, pues es más importante el aspecto visual.

Sin embargo, tanto los códigos QR como las segundas pantallas son cada vez más utilizados por los telespectadores, según se deduce de los datos obtenidos de la encuesta a usuarios de Internet “Navegantes en la Red” realizada por AIMC (Asociación para la Investigación de los Medios de Comunicación, 2022). De 15623 encuestados entre octubre y diciembre de 2021, casi el 80 por ciento afirma haber hecho alguna consulta de información a través de los códigos QR/Bidi en el último mes. En cuanto al tipo de consulta, aunque mayoritariamente se han utilizado para acceder a cartas de bares o restaurantes, es significativo que un 24,1 por ciento declare que los ha utilizado para ampliar información sobre algo y un 7,7 por ciento para acceder a contenidos multimedia. La tendencia se ve más claramente al comparar los resultados con los de años anteriores:



Figura 11. Frecuencia de uso de los códigos QR entre 2013 y 2021. Fuente: Encuesta Navegantes en la Red 2021. Infografías. <https://www.aimc.es/otros-estudios-trabajos/navegantes-la-red/infografia-resumen-24o-navegantes-la-red/>

El informe de la AIMC destaca la tendencia al alza del uso de códigos QR: “Un crecimiento que, a raíz de la pandemia, se vio claramente amplificado en 2020, y que, en 2021, su uso ha sido incluso mayor. Así, casi uno de cada dos encuestados (47,8%) ha escaneado códigos de este tipo en la semana anterior a la entrevista, es decir, 6 puntos más que en 2020 y 21 más respecto a 2019. Ampliando el foco al mes, la cifra se eleva al 77,7% (+4 puntos en un año y +24 en dos). Y sólo el 7% (la mitad que en 2019) no los ha escaneado nunca”.

La misma encuesta también proporciona datos significativos sobre el acceso a las segundas pantallas: un 36,7 por ciento contesta que ha buscado información sobre el contenido de un programa mientras lo veía y solo un 9,5 por ciento utiliza la web o la app del programa. No se da una cifra global, pero es evidente que una gran proporción de espectadores realiza actividades relacionadas con el contenido que consume en televisión a través de Internet, bien en la web, bien en redes sociales.

Esto lleva a sugerir la posibilidad de combinar la información enriquecida que se puede ofrecer mediante la tecnología de datos vinculados con los códigos QR sobreimpresionados en imagen como otra forma de controlar las segundas pantallas, aunque también puedan suponer una distracción para el espectador respecto a la emisión televisiva convencional.

## 5. Oportunidades y proyectos futuros

Tanto los antecedentes estudiados como la propuesta de aplicación expuesta anteriormente de los datos vinculados a los archivos de televisión conduce a dos escenarios de actuación factibles a corto plazo: por un lado, en la puesta en marcha del Archivo Audiovisual de la Comunidad Valenciana y por otro, en el seno de la FORTA.

Así, pues, el proyecto *EUscreen*, donde los programas de televisión seleccionados para ser publicados como datos vinculados procedían de distintos

medios (en sus inicios constaba de 30.000 archivos procedentes de 20 fuentes, y hoy son más de 60.000 archivos procedentes de 25 empresas televisivas de 19 países), conduce a pensar en la oportunidad de crear catálogos colectivos de información audiovisual procedentes de diversos orígenes mediante procesos de armonización de metadatos.

Puede ser especialmente adecuado en el caso de la puesta en marcha del futuro Archivo Audiovisual de la Comunidad Valenciana, al tener que armonizar esquemas de metadatos procedentes de fondos distintos (según se expone en el artículo 7 punto 1 de la Ley de creación del Consell Audiovisual de la Comunidad Valenciana (CACV): “Se crea el Archivo Audiovisual de la Comunitat Valenciana para coordinar los archivos y fondos audiovisuales de la Generalitat y sus instituciones, en especial los archivos del Institut Valencià de Cultura y los de la Radiotelevisió Valenciana, SAU, en liquidación, así como de aquellas otras instituciones públicas o privadas que quieran adherirse”.

Sobre todo, en los apartados c y e del mismo artículo referidos a los objetivos de creación del CACV:

“c) Apoyar y asesorar a aquellas instituciones públicas o privadas que quieran digitalizar sus archivos audiovisuales e integrarlos o conectarlos con el Archivo Audiovisual de la Comunitat Valenciana.

e) Fomentar la difusión e internacionalización del audiovisual y los activos audiovisuales de la Comunitat Valenciana”.

Es decir, el futuro Archivo Audiovisual de la Comunidad Valenciana tendrá su germen en el de la extinta Radiotelevisión Valenciana, pero a él se sumarán archivos audiovisuales de distinto contenido y procedencia con variados esquemas de metadatos que será necesario armonizar. Habrá que adaptar los modelos actuales, basados en el *Minimum Data List* y compararlos con otros basados en Dublin Core, como por ejemplo EBU Core, disponible como ontología RDF. La homogeneización de los esquemas de metadatos no son un fin en sí mismos, sino un requisito previo en la publicación de los catálogos (o una parte de ellos) como datos abiertos vinculados. Sin embargo, este paso previo proporciona unas posibilidades interesantes de cara al futuro, puesto que, además de lograr la integración de assets de media de varias procedencias, se están estableciendo las bases para la publicación del archivo como datos abiertos vinculados en una fase posterior. Siguiendo la metodología empleada en *EUScreen*, el proceso consta de tres pasos principales: la ingesta y armonización de los metadatos, la transformación de esos metadatos en un esquema de referencia común y su publicación en forma de datos vinculados (LOD, *Linked Open Data*).

En relación a la propuesta de compartir recursos en el seno de la FORTA (Federación de Organismos de Radio y Televisión Autonómicos), y siempre vigilando la gestión de derechos, se propone la creación de un modelo de datos mínimo armonizado (que podría incluir una ontología multilingüe) entre las emisoras integrantes del servicio de intercambio de imágenes con el fin de optimizar esfuerzos. Si el intercambio de archivos que se realiza entre las

televisiones autonómicas lo es también de los metadatos, cuando una de las cadenas incorpore material procedente de otra no será necesario describir de nuevo el material audiovisual, además de superar las barreras idiomáticas. El mismo modelo podría ser aplicable a servicios de compraventa de imágenes o incluso agencias de noticias. Los metadatos asociados se incorporarían automáticamente al sistema, facilitando la tarea de los catalogadores.

Por último, otra de las posibles aplicaciones sería crear URI de los programas disponibles en la web, por ejemplo, a partir de los números de producción, ya que, aunque de uso interno, se trata de identificadores únicos y normalizados para cada programa y cada una de sus versiones. A través de los URI de los programas, se pueden establecer mecanismos para vincularlos a sus contenidos, de modo que se permita buscar y descubrir contenido de archivo a través de los datos vinculados (EBU, 2015).

## 6. Conclusiones

Los servicios de documentación de las televisiones se esfuerzan por aplicar nuevas tecnologías que liberen a los documentalistas de tareas precisas pero laboriosas y que a la vez faciliten la reutilización de las imágenes para uso del medio propio o con fines comerciales. Los métodos basados en la inteligencia artificial, cada vez más perfeccionados, todavía necesitan de supervisión humana. Mientras tanto, la especificidad requerida en el momento de la recuperación de la información es cada vez mayor, al tiempo que los departamentos de documentación se involucran más en el medio en el que se circunscriben. Estos factores impelen a buscar fórmulas para racionalizar esfuerzos, aprovechar los recursos existentes, y a la vez crear nuevos cauces para enriquecer el producto informativo y su difusión mediante la conexión con fuentes externas.

De todo ello, tras estudiar la aplicación integral de las tecnologías de la web semántica en la gestión de los archivos audiovisuales se pueden extraer las siguientes conclusiones:

Queda pendiente explorar el camino del periodismo semántico mencionado anteriormente (Panagioditis y Veglis, 2020): si la web semántica y los datos vinculados (con identificadores persistentes, fiables, estables, accesibles y unívocos) pueden contribuir a la verificación de noticias.

La tecnología basada en la web semántica y de datos vinculados no es reciente. En el campo de los archivos audiovisuales ha habido algunas iniciativas, pero no se ha aplicado en la totalidad de los procesos de la cadena documental. Conscientes de que se trata de una empresa de envergadura que implica muchos cambios en el tratamiento “clásico” de la documentación, proporciona expectativas inexploradas en la apertura de los archivos de televisión (tradicionalmente cerrados) al exterior. Así como supone una oportunidad para el enriquecimiento del propio archivo y del medio al que pertenece mediante los enlaces desde la página web. La evolución natural de las emisiones de televisión a través de web es transformar los metadatos y publicar los contenidos como datos enlazados para vincularlos a otras fuentes de contenido (Simou, Evain, Drsopoulos y Tzouvaras, 2013).

Todavía falta mucho camino para que las fuentes de datos sean fácilmente accesibles, así como mejorar los procedimientos y aplicaciones que permitirán obtener un mejor resultado de la web de los datos. Se trata de explorar las posibilidades y los beneficios que los archivos audiovisuales, y en concreto, los de televisión pueden obtener de la aplicación de la tecnología de los datos vinculados. Posiblemente, y con los recursos actuales no sea posible hacerlo desde un departamento de una empresa de comunicación, pero, sin duda, sí es un campo abierto a la colaboración con entidades públicas o privadas.

Incluso sin terminar por completo el proceso de aplicación de la metodología de la web semántica a los archivos audiovisuales, se obtienen mejoras significativas como la integración de distintos modelos de datos, su transformación a formatos legibles por máquina, el control de las segundas pantallas, la gestión de derechos o la conexión con las guías de programación, además de las tareas tradicionales de la gestión de la información.

Las posibilidades que puede ofrecer en un futuro la tecnología de los datos abiertos vinculados son prometedoras. Conduce a abrir horizontes tanto a la posibilidad de intercambiar información y datos con otros archivos, como a crear catálogos colectivos, teniendo en cuenta el lema de los principios FAIR: “tan abierto como sea posible, tan cerrado como sea necesario”. Pero para todo ello, lo determinante para los responsables de los medios o instituciones dependientes de los repositorios audiovisuales es saber hasta qué punto están dispuestos a abrir sus colecciones y si están dispuestos a compartir los datos como datos abiertos vinculados.

## 7. Referencias bibliográficas

- AIMC (Asociación para la Investigación de los Medios de Comunicación). (2022). *24ª Encuesta a usuarios de Internet – Navegantes en la Red*. <<http://www.aimc.es/otros-estudios-trabajos/navegantes-la-red>>. [consulta: 27/03/2022].
- Alexander, F. (2012). Simplifying Semantics: Linked data rules ok!. *Iasa Journal*, nº38. <<https://www.iasa-web.org/sites/default/files/iasa-journal-38-part7.pdf>>.
- Berners-Lee, T. (2006). *Linked Data*. <<https://www.w3.org/DesignIssues/LinkedData.html>> [consulta: 29/04/2022].
- Berners-Lee, T. (2009). *The next web*. <[www.ted.com/talks/tim\\_berners\\_lee\\_the\\_next\\_web?utm\\_campaign=tedspread&utm\\_medium=referral&utm\\_source=tedcomshare](http://www.ted.com/talks/tim_berners_lee_the_next_web?utm_campaign=tedspread&utm_medium=referral&utm_source=tedcomshare)>. [consulta: 09/04/2022].
- Dappert, A.; FARQUHAR, A.; KOTARSKI, R.y HEWLETT, K. (2017). “Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research”. *Data Science Journal*, vol. 16. ISSN 1683-1470. <https://doi.org/10.5334/dsj-2017-028>.
- Datos.gob.es. (2017). *Principios FAIR: Buenas prácticas para la gestión y administración de datos científicos*. <<https://datos.gob.es/es/blog/la-importancia-de-abrir-los-datos-culturales#interact>><https://datos.gob.es/es/noticia/principios-fair-buenas-practicas-para-la-gestion-y-administracion-de-datos-cientificos>>. [consulta: 08/05/2022].
- Datos.gob.es. (2020). *La importancia de abrir los datos culturales*. <<https://datos.gob.es/es/blog/la-importancia-de-abrir-los-datos-culturales#interact>>. [consulta: 08/05/2022].

- EBU European Broadcasting Union. (2015), EBU-MIM Semantic Web Activity Report. Geneva, EBU. <<https://tech.ebu.ch/docs/techreports/tr019.pdf>>.
- FIAT-IFTA. (2019). *MPEG-21 Media Contract Ontology in RAI: a new way to manage audiovisual rights*. <[https://fiatifta.org/wp-content/uploads/2019/09/Introduction-MCO\\_last-version.pdf](https://fiatifta.org/wp-content/uploads/2019/09/Introduction-MCO_last-version.pdf)>.
- Generalitat Valenciana. Ley 10/2018 de creación del Consell del Audiovisual de la Comunitat Valenciana (CACV). <<https://www.boe.es/eli/es-vc/l/2018/05/18/10>>.
- Heftberger, A (2019). Building Resources Together—Linked Open Data for Film Archives. *Journal of Film Preservation*, nº 101, 65–73.
- Hendler, J.; Berners-Lee, T. y Miller, E. (2002). “Agent Technology on the Internet. 3. Integrating Applications on the Semantic Web”. *The Journal of the Institute of Electrical Engineers of Japan*, vol.122 (2002), 10, 676–680. ISSN 1881-4190. <https://doi.org/10.1541/ieejjournal.122.676>.
- Kobilarov, G.; Scott, T.; Raimond, Y.; Oliver, S.; Sizemore, C. et al. (2020). Media meets semantic web—how the bbc uses dbpedia and linked data to make connections. ARROYO, Lora et.al. (eds.). *The Semantic Web: Research and Applications. ESWC 2009. Lecture Notes in Computer Science*, 723–737. Berlin, Heidelberg, Springer. ISBN 978-3-642-02121-3. [https://doi.org/10.1007/978-3-642-02121-3\\_53](https://doi.org/10.1007/978-3-642-02121-3_53).
- López de Quintana, E. (2019). Navegando entre la tecnología, los flujos de información y el lenguaje. Mesa redonda: Organización y recuperación de información en archivos audiovisuales. IV Congreso ISKO España-Portugal. Barcelona, ISKO.
- López de Quintana, E. y León-Carpio, A. (2021a). Algoritmos e IA en la Documentación de Atresmedia: una realidad en explotación. *XXII Jornada de Gestión de la Información de SEDIC: Algoritmo & Compás: El profesional de la información al son de la inteligencia artificial ¿o no?* <[www.youtube.com/watch?v=bhDd6sbFog](http://www.youtube.com/watch?v=bhDd6sbFog)>. [consulta: 20/04/2022].
- López de Quintana, E. y León-Carpio, A. (2021b). Artificial intelligence for a role change in television archives: The Atresmedia—Etiqmedia experience”. *Journal of Digital Media Management*, 10 (2), 177–187.
- Marcondes, C. H. (2018). Datos abiertos enlazados de archivos, bibliotecas y museos en la web. Barcelona, UOC.
- Marín-Ariza, P. (2022). Madurez de sistemas de identificadores persistentes: oportunidades en el contexto español. *Anuario ThinkEPI*. ISSN 2564-8837. <https://doi.org/10.3145/thinkepi.2022.e16a06>.
- Minel, J. G. (2019). “Cómo dar vida a los archivos e integrarlos en la producción”. *TM Broadcast*, marzo, 14–15. <<https://issuu.com/daromedia/docs/tmbroadcast121>>.
- Nixon, L. [s.f.]. *Linked television. Seamlessly linking television to the Web*. <<https://lydonnixon.wixsite.com/linkedtvs/showcase/home>>. [consulta: 16/06/2022].
- Nixon, L (2015). Delivering related web content synchronized to online television: The LinkedTV solution. *2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*. Berlin, IEEE, 323-327. ISBN 9781479987481. <https://doi.org/10.1109/icce-berlin.2015.7391270>.
- Nixon, L.; Troncy, R.; Mezaris, V. (2013). “TV’s future is linked: Web and television across screens”. *EuroITV '13: Proceedings of the 11th European Conference on Interactive TV and Video*, pp. 177-178. <https://doi.org/10.1145/2465958.2465987>.
- Open Knowledge Foundation. [s. f.]. *Definición de conocimiento Abierto*. <<https://opendefinition.org/od/1.0/es>>. [consulta: 13/07/2022].
- Panagiotidis, K. y Veglis, A. (2020). Transitions in Journalism—Toward a Semantic-Oriented Technological Framework. *Journalism and Media*, 1 (1), 1–17. ISSN 2673-5172. <https://doi.org/10.3390/journalmedia1010001>.

- Pastor-Sánchez, J.-A.; Bazán-Gil, V. y Escribano, M. (2019). Audiovisual content description using semantic scene representation. *Journal of Digital Media Management*, 8 (1), 47-57.
- Pierce, C. (2020). Unique challenges facing Linked Data implementation for National Educational Television. *International Journal of Metadata, Semantics and Ontologies*, 14 (2), pp. 98. ISSN 1744-263X. <https://doi.org/10.1504/ijmso.2020.108323>.
- Poveda-Villalón, M., Espinoza-Arias, P., Garijo, D. y Corcho, O (2020). Coming to Terms with FAIR Ontologies. Keet, C.M y Dumontier, M. (eds.), *Knowledge Engineering and Knowledge Management. EKAW 2020. Lecture Notes in Computer Science*, Springer, Cham, v12387, 255-270. [https://doi.org/10.1007/978-3-030-61244-3\\_18](https://doi.org/10.1007/978-3-030-61244-3_18).
- Radulovic, F.; Poveda-Villalón, M.; Vila-Suero, D.; Rodríguez-Doncel, V.; García-Castro, R. et al. (2015). Guidelines for Linked Data generation and publication: An example in building energy consumption. *Automation in Construction*, 57 178–187. ISSN 0926-5805. <https://doi.org/10.1016/j.autcon.2015.04.002>.
- Rayfield, J (2010). “BBC World Cup 2010 dynamic semantic publishing”. *BBC Internet Blog*. <[https://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc\\_world\\_cup\\_2010\\_dynamic\\_sem.html](https://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html)> [consulta: 29/04/2022]
- Redondo-García, J. L.; Botón-Fernández, V. y Lozano-Tello, A. (2012). Linked Data methodologies for managing information about television content. *International Journal of Interactive Multimedia and Artificial Intelligence*, 1(6), 36-43. <https://doi:10.9781/ijimai.2012.165>.
- Ríos-Hilario, A. B. (2014). La tecnología de datos vinculados al servicio de los medios de comunicación. *Cuadernos de Documentación Multimedia*, 2, 118–140. ISSN 1575-9733. [https://doi.org/10.5209/rev\\_cdmu.2014.v25.47470](https://doi.org/10.5209/rev_cdmu.2014.v25.47470)
- Senso, J. A. y Rosa-Piñero, A. de la. (2003). El concepto de metadato: algo más que descripción de recursos electrónicos. *Ciência da Informação*, vol. 32, pp. 95-106.
- Simou, N.; Jean-Pierre E.; Drosopoulos, N. y Tzouvaras, V. (2013). “Linked European television heritage”. *Semantic Web*, 4 (3), 323–329. ISSN 1570-0844. <https://doi.org/10.3233/sw-130104>.
- Smith-Yoshimura, K. (2021). *La transición a la siguiente generación de metadatos: informe OCL Research*. MARINAS, Iván Pérez (traducción). Biblioteca Nacional de España. <[www.bne.es/export/sites/BNWEB1/webdocs/Inicio/Perfiles/Bibliotecarios/La-transicion-siguiente-generacion-metadatos.pdf](http://www.bne.es/export/sites/BNWEB1/webdocs/Inicio/Perfiles/Bibliotecarios/La-transicion-siguiente-generacion-metadatos.pdf)>.
- Strobel, S. y Marín-Arraiza, P. (2015). Metadata for Scientific Audiovisual Media: Current Practices and Perspectives of the TIB|AV-Portal. *Communications in Computer and Information Science*, pp. 159–170. Cham, Springer International Publishing. ISBN 9783319241289. [https://doi.org/10.1007/978-3-319-24129-6\\_14](https://doi.org/10.1007/978-3-319-24129-6_14).
- Terán, B. (2020). La expansión de código QR en televisión. *La información*. 02/12/2020. <<https://www.lainformacion.com/opinion/borja-teran/la-expansion-del-codigo-qr-en-television/2822464/>>. [consulta: 10/03/2022].
- W3C Working Group Note. (2009). *SKOS Simple Knowledge Organization System Primer*. <<https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>>. [consulta: 21/04/2022].
- Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, vol. 3 (1), 1–9. ISSN 2052-4463. <https://doi.org/10.1038/sdata.2016.18>