
ARTÍCULO

Catalogación automatizada en archivos de medios de comunicación audiovisuales en base a la estandarización del evento. Estudio de caso

Automated cataloguing in audiovisual media archives
based on event standardization. Case study

Ferran Aragó Puigventós

Departamento de Periodismo y Ciencias de la Comunicación, Facultad de Ciencias de la Comunicación, UAB,
Cerdanyola del Vallès, Departament de Documentació, CCMA-3Cat, Sant Joan Despí
ferran.arago@autonoma.cat, <https://orcid.org/0009-0001-0388-2497>

Pilar Cid-Leal

Departamento de Filología Catalana, Facultad de Traducción y de Interpretación, UAB, Cerdanyola del Vallès
Pilar.Cid@uab.cat, <https://orcid.org/0000-0001-9820-7541>

María José Recoder Sellarès

Departamento de Periodismo y Ciencias de la Comunicación, Facultad de Ciencias de la Comunicación, UAB,
Cerdanyola del Vallès
MariaJosep.Recoder@uab.cat, <https://orcid.org/0000-0002-2329-6735>

Resumen: Este estudio, enmarcado en el ámbito de la documentación audiovisual, examina el caso de OPTA-Digiton, un sistema para la automatización de la catalogación de materiales audiovisuales de temática deportiva, desarrollado por el Departamento de Documentación de la CCMA-3Cat. El sistema, escalable y exportable, introduce un modelo novedoso para la estructuración y estandarización de metadatos basados en eventos, lo que genera resultados eficaces en la gestión documental. El estudio destaca el papel fundamental de los documentalistas en las primeras etapas de diseño y desarrollo. Durante estas fases, se emplean principios de Biblioteconomía y Documentación, así como conceptos de la Web Semántica, para construir modelos de datos de calidad, un recurso fundamental en la evolución de los sistemas de información impulsada por la IA.

Palabras clave: archivos de medios de comunicación, documentación audiovisual, catalogación automatizada, estandarización, gestión documental, eventos deportivos, CCMA-3Cat.

Abstract: *This study, framed within the field of audiovisual documentation, examines the case of OPTA-Digiton, a system for the automation of cataloguing of sports-themed audiovisual materials, developed by the Documentation Department of CCMA-3Cat. The system, which is both scalable and transferable, introduces a novel model for structuring and standardizing event-based metadata, yielding effective outcomes in document management. The study underscores the pivotal role of documentalists in the early design and development stages. During these phases, principles from Library and Information Science and Semantic Web concepts are employed to build high-quality data models, a critical asset in the AI-driven evolution of information systems.*

Keywords: *media archives, audiovisual documentation, automated cataloguing, standardization, document management, sports events, CCMA-3Cat.*

Recibido: 11-06-2025 / **Aceptado:** 24-07-2025 / **Publicado:** 22-06-2026

Cómo citar: Aragó Puigventós, F., Cid-Leal, P., Recoder Sellarès, M. J. (2026). Catalogación automatizada en archivos de medios de comunicación audiovisuales en base a la estandarización del evento. Estudio de caso. *Revista Española de Documentación Científica*, 49(1), 1853. <https://doi.org/10.3989/redc.2026.1.1853>

Copyright: © 2026 CSIC. Este es un contenido de acceso abierto diamante distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

Información complementaria ↓

1. INTRODUCCIÓN

Los servicios de documentación en medios de comunicación audiovisuales se hallan inmersos en una profunda transformación, iniciada a principio de los años 2000, como consecuencia de la digitalización (López de Quintana, 2014) y la irrupción de los medios sociales (Marcos y Edo, 2015). El paso de lo analógico a lo digital conllevó un aumento muy considerable de la gestión documental de una producción y emisión de contenidos en crecimiento exponencial (López de Quintana, 2014), acelerándose con la evolución de los medios de comunicación audiovisuales hacia entornos multiplataforma (Menke et al., 2018). Los cambios se agudizaron con la implementación de sistemas de automatización, especialmente a finales de la década de 2010, con la incipiente incorporación de herramientas de Inteligencia Artificial (a partir de ahora IA) (Caldera, 2025; Bazán, 2023).

La tendencia actual viene marcada por la creciente incorporación de herramientas IA en los procesos de automatización de la gestión documental y la recuperación de la información (Bazán, 2023); la finalización de la digitalización retrospectiva del fondo de archivo (European Broadcasting Union, 2025); la adaptación del tratamiento documental a las nuevas narrativas audiovisuales y su consiguiente diversidad de formatos (Marcos et al., 2021), así como la adecuación del servicio a nuevas normativas de protección de datos y gestión de derechos (European Broadcasting Union, 2025).

Dichas tendencias conllevan retos asociados, tales como la necesidad de modificar circuitos de trabajo para adaptar los sistemas documentales a la nueva realidad multiplataforma e integrar las nuevas herramientastecnológicas (Bazán, 2023; European Broadcasting Union, 2025), la búsqueda de estrategias para la asunción de la creciente gestión documental y la necesidad de mejorar la estructuración y estandarización de los metadatos (Soler y Gil, 2023); la adopción de procesos de verificación de la información, su trazabilidad (Caldera, 2025) y de la gestión de derechos de los materiales de origen (Grayson, 2023; European Broadcasting Union, 2025), y la formación de los documentalistas para ajustarse a las nuevas competencias y roles profesionales (Caldera, 2024; Codina y Lopezosa, 2024).

Todos estos retos, y en especial, el relativo a la creciente gestión documental, han supuesto un aumento muy considerable de la carga de trabajo, sin que esta se haya visto acompañada de un incremento proporcional de recursos humanos para hacerle frente (Soler y Gil, 2023). Ante la perspectiva de que la situación no cambie, la búsqueda de soluciones se ha convertido en un objetivo prioritario de buena parte del sector.

En este sentido y a modo de ejemplo, presentamos el departamento de Documentación de la Corporació Catalana de Mitjans Audiovisual, CCMA-3Cat, como un caso prototípico del escenario común descrito.

Según la propia definición de la organización, nacida en 1983 como Corporació Catalana de Ràdio i Televisió, CCRTV, rebautizada como Corporació Catalana de Mitjans Audiovisuals, CCMA en 2007 y, finalmente, como 3Cat en 2022, “es el ente público que gestiona los medios de comunicación audiovisual de la Generalitat de Catalunya (...) en cumplimiento de su misión de servicio público y con el compromiso de calidad, independencia y promoción de la cultura y lengua catalanas” (3Cat, 2024a). Actualmente, con una plantilla de 2.500 trabajadores y un presupuesto para el 2024 de 349 millones de euros, la CCMA-3Cat gestiona un conglomerado mediático, líder de audiencia en el ámbito catalán, dividido en cinco grandes áreas de producción y difusión: (1) Televisión lineal, con seis canales, capitaneados por TV3; (2) el grupo radiofónico de Catalunya Ràdio, compuesto por cuatro emisoras; (3) EVA, la

marca creada para distribuir contenidos en catalán para el público juvenil a través de las principales redes sociales; (4) la plataforma de distribución directa (OTT) 3Cat, que, desde 2024, unifica todos los contenidos radiofónicos, audiovisuales y digitales del grupo, y (5) los contenidos digitales generados por los medios del grupo, difundidos a través de productos propios (portales web, apps móviles, etc.) o mediante plataformas externas (redes sociales, plataformas de podcasting, etc.) (3Cat, 2024a).

El Departamento de Documentación fue creado en abril de 1983, meses antes de la primera emisión televisiva, con la misión de constituir y gestionar la videoteca, el archivo audiovisual de TV3. Desde su inicio, se estableció una política de preservación de su producción y emisión, siguiendo el modelo del Institut National de l'Audiovisuel (INA) francés y las recomendaciones de organismos internacionales tales como la UNESCO o la Federación Internacional de Archivos de Televisión (FIAT/IFTA), con una doble función: (1) patrimonial, preservando un fondo audiovisual único, testimonio de la producción televisiva y de la historia política, social y cultural del país, y (2) de producción, rentabilizando al máximo la explotación del fondo en la emisión y producción propia y de terceros (Conesa, 2012).

Durante los treinta primeros años, la actividad del servicio se estructuró en cuatro áreas: (1) videoteca, subdividida en las secciones de informativos, deportes, entretenimiento e históricos; (2) servicio de documentación escrita y archivo de prensa; (3) gestión de derechos, y (4) servicio de copias y venta de imagen de archivo.

En 2003 se implementó el proceso de producción y archivo digitalizado, mediante un sistema de gestión de contenidos audiovisuales, en inglés *Media Asset Management* (a partir de ahora, MAM) de diseño propio, apodado Digition, que se divide en tres entornos integrados: producción, emisión y archivo. Tres años después, se inició la digitalización del fondo de archivo, proceso prácticamente concluido en 2023, a falta de integrar conjuntos residuales de brutos de rodaje (Conesa, 2012; Bailac, 2023).

A mediados de la década de 2010, el departamento asumió progresivamente nuevos servicios, en el seno de un proceso de reestructuración de la corporación. En 2015, se integró el servicio de indización de noticias de agencia y señales de intercambio, hasta entonces gestionado por la redacción. En 2016, se incorporó el fondo y la gestión del archivo radiofónico. En 2017, se hizo cargo del archivo central administrativo y de gestión de la CCMA, creándose, en 2019, las bases de datos del archivo fotográfico. Finalmente, en 2024, se incorporó parte del tratamiento documental de medios digitales, redes sociales y la OTT, completándose así su evolución hacia un servicio de documentación transmedia (Bailac, 2023).

Estas nuevas responsabilidades coinciden en el tiempo con un crecimiento muy acusado del fondo documental a gestionar. Según datos obtenidos mediante consulta en el MAM, en la serie histórica del archivo (1984-2024) se aprecian tres grandes momentos de incremento de los materiales audiovisuales incorporados al fondo documental. El primero se produce en 1990, con la creación de un segundo canal de televisión (C33), con un incremento del 66% respecto a los datos del 1988. El segundo, en 2004, con la digitalización y la creación del canal de noticias de 24h, con un incremento del 95%, respecto al 2002 y, finalmente, a partir del 2016, con un crecimiento del 30% respecto al 2014, manteniéndose una progresión acelerada hasta la actualidad, con la excepción de un ralentí durante la pandemia de 2020. Si se comparan los datos de la última década, en 2024 la videoteca incorporó 196.587 documentos de audio y vídeo frente a los 103.697 de 2014, lo que supone un incremento del 90%.

Este aumento se explica por el crecimiento de la producción y emisión de contenidos en un entorno multiplataforma, que conlleva asociado, no solo el incremento de las tareas de selección y tratamiento documental de un material diversificado, sino también un incremento muy significativo de las tareas de recuperación de la información por parte de los documentalistas.

En lo referente a los recursos humanos, si se comparan datos obtenidos de documentos internos relativos a la planificación de febrero de 2015 con los de febrero de 2025, se observa que, mientras la carga de trabajo se ha duplicado, los recursos humanos apenas han crecido un 9% y los efectivos dedicados al análisis documental sólo un 4%. Las cifras expuestas demuestran que el servicio de

Documentación de 3Cat no escapa a la realidad descrita anteriormente. Según palabras de **Montse Català (2025)**, jefa del departamento de Documentación de 3Cat, estamos lidiando con un tsunami, con plantillas estancadas y contención presupuestaria.

Ante tal situación, el departamento lleva años implementando una serie de acciones que se resumen en: (1) optimización de la gestión documental; (2) externalización de servicios; (3) rediseño de circuitos y redistribución de efectivos según carga de trabajo; (4) automatización del tratamiento documental en base a herramientas de IA externas y (5) automatización en base a proyectos propios.

La optimización de la gestión documental se ha centrado, por un lado, en la revisión de las políticas de selección de materiales para el archivo, restringiendo la entrada de material propio, evitando duplicados o aquellos no útiles para su reutilización, y limitando la entrada de material cuyo origen de producción es ajeno. Según datos estadísticos internos, pese a que la producción y emisión de programas se ha duplicado en diez años, se ha reducido significativamente el paso al archivo de las piezas contenidas en ellos y las usadas como material original. En concreto, en 2024 se pasaron al archivo un 24% de editados menos respecto a su punto álgido en 2017, y las agencias archivadas disminuyeron un 32% respecto a su máximo en 2022.

Por otro lado, se ha simplificado el análisis documental, concentrando la descripción en la ficha documental y limitando la descripción de planos, donde solo se aporta información no contenida en la ficha, a fin de economizar al máximo la descripción necesaria para una posterior recuperación. Se asume, no sin cierto riesgo, la máxima de coste-beneficio en el binomio catalogación-recuperación, a sabiendas que un porcentaje pequeño de necesidades informativas no podrán ser resueltas mediante puntos de acceso textuales, sino mediante la comprobación visual de las imágenes.

En lo referente a la externalización de servicios, se ha licitado parte de la gestión documental del archivo de prensa, liberando así a los documentalistas de la selección y entrada de las noticias a la base de datos, para centrarse en el análisis documental y la salida de la información.

Asimismo, se han rediseñado los circuitos de trabajo para conseguir una mayor eficiencia de los efectivos disponibles, redistribuyéndolos según la carga de trabajo y las necesidades de producción, dando, por ejemplo, nuevos roles a los auxiliares de documentación en tareas documentales rutinarias que no requieren de análisis documental, pero que son de gran ayuda para los documentalistas, tales como la introducción manual de metadatos en procesos semiautomatizados.

Con respecto a la incorporación de herramientas IA, en 2020 se contrató los servicios de la empresa Etiquedia (<https://etiquedia.com/>), iniciándose un piloto para testar sus soluciones tecnológicas relativas al reconocimiento del habla, detección de idioma en el audio, segmentación de hablantes, transcripción de texto en la imagen, reconocimiento facial, segmentación del vídeo en planos, categorización de los planos segmentados y extracción de entidades semánticas. Actualmente, se ha incorporado parcialmente al sistema la herramienta de audio a texto, mientras que el resto de los módulos siguen su curso de evaluación (**Català, 2025**).

Finalmente, en relación a los proyectos de automatización propios, el archivo de 3Cat ha puesto en marcha un proceso de distribución automatizada de las noticias de agencia hacia las carpetas de las secciones de informativos en el área de producción del MAM, se han desarrollado sistemas de automatización del tratamiento de datos electorales, se ha semiautomatizado el uso de los subtítulos para la descripción de totales, y, finalmente, lo más complejo, se ha desarrollado un sistema para la automatización del análisis documental, apodado OPTA-Digition.

Gracias a la conjugación de las acciones descritas y del esfuerzo e implicación de los trabajadores, el departamento de documentación de 3Cat está sorteando el reto dando una respuesta efectiva a su misión, con reconocimiento explícito de su buen hacer por parte de los usuarios del servicio (**Català, 2025**).

En cifras obtenidas de la presentación anual de estadísticas internas del departamento, en 2024 se trataron

documentalmente 156.587 vídeos, en un sistema de producción audiovisual que generó un total de 44.198 horas de archivo. Documentación escrita trató 72.197 noticias de prensa, el equipo de radio gestionó 40.532 documentos sonoros, y, finalmente el archivo incluyó 14.504 nuevas fotografías. En total, estamos hablando de una gestión documental de casi 284.000 documentos anuales, a lo que debe sumarse 11.039 búsquedas realizadas por los documentalistas con una dedicación estimada de 5.900 horas.

No obstante, los cambios previstos en el área de Informativos (3Cat, 2024b) y la creciente oferta del entretenimiento hacen prever un futuro donde la producción seguirá aumentando. A falta de documentalistas, con una optimización de recursos exprimidos al máximo y la resistencia a dejar de catalogar colecciones, todo indica que estas soluciones pasan por ahondar en los procesos de automatización de la catalogación.

Así pues, con la premisa de la automatización como opción crítica para afrontar la gestión presente y futura de los archivos de medios de comunicación audiovisuales, el objetivo de este artículo es presentar el sistema OPTA-Digition, un proceso de automatización de la catalogación de los materiales audiovisuales de temática deportiva, sustentado en una nueva propuesta de estructuración y estandarización basada en el evento, entendido como la unidad semántica básica de la información periodística y documental (Caswell y Dörr, 2017).

La hipótesis es que el éxito de dicho proceso no lo determina la tecnología, sino el rol activo del documentalista y la aplicación de los fundamentos de la Ciencia de la Documentación en la fase inicial de su diseño y desarrollo.

2. METODOLOGÍA

La base metodológica de este trabajo es el estudio de caso (Yin, 2018; Codina, 2023). Para llevarlo a cabo, hemos desarrollado un análisis del sistema OPTA-Digition y como este incorpora los fundamentos de la Ciencia de la Documentación para la estructuración y estandarización de la catalogación automatizada de materiales audiovisuales.

Dicho análisis se ha realizado a partir de documentación interna del servicio elaborada por el autor -codesarrollador del proyecto- y de los datos obtenidos mediante consulta al MAM. También se ha revisitado las fuentes de información bibliográficas consultadas durante su diseño y desarrollo, referenciadas en el punto 3.3. (b) del artículo. En ese aspecto, y dado el lapso de cinco años transcurridos desde el diseño de su primera versión, se valoró si mantener o actualizar las referencias, en su mayoría modelos de datos publicados en sitios web. Se ha optado por lo segundo, al considerar que dicha decisión no altera los resultados, al tiempo que versiones actualizadas proporcionan mayor utilidad a las personas interesadas en su consulta.

Para establecer el marco teórico se ha realizado una revisión bibliográfica sobre el estado del arte de la documentación audiovisual y para la construcción del caso (Coller, 2005) se ha utilizado documentación interna del departamento de Documentación de 3Cat, el análisis de los datos obtenidos mediante consulta al MAM de 3Cat, además de comunicaciones personales con responsables del departamento.

3. ANÁLISIS DEL SISTEMA OPTA-DIGITION

OPTA-Digition es un sistema de automatización del análisis documental de materiales de temática deportiva desarrollado por el departamento de Documentación de 3Cat, en colaboración con el departamento de Ingeniería del ente. El sistema se basa en fundamentos de la Ciencia de la Documentación y en principios de la web semántica, enfocando la estructuración y estandarización de los metadatos para una posterior automatización, a partir de la descripción inicial del evento deportivo. La propuesta, sin partida presupuestaria y basada en tecnología asequible, integra las nuevas dinámicas de producción de contenido, respetando la coherencia del archivo, en base a las posibilidades y limitaciones, tanto del departamento de Documentación como de la organización de la cual depende.

3.1. Origen, objetivo y alcance

OPTA-Digition nace en 2020, como una evolución del proyecto TrAgEdi (acrónimo de Transmisiones, Agencias y Editados), un sistema de catalogación semiautomatizada basado en macros de Excel iniciado en 2016. Cuando, desde la dirección del departamento se hace el encargo, se marca como objetivo crear una nueva herramienta para automatizar la precatalogación de los materiales audiovisuales relativos a competiciones deportivas, a partir de datos externos proporcionados por la empresa OPTA Sports (<https://www.statsperform.com/opta/>), disponibles gracias a la suscripción existente del área de Deportes de 3Cat con dicha empresa para nutrir la sección de resultados deportivos del portal web.

A partir del objetivo inicial, y viendo la necesidad de reestructurar la información para poder automatizar, se aprovecha el desarrollo para replantear el tratamiento documental de los materiales de temática deportiva, y extender el cambio a todos los documentos, independientemente del tipo de material o de contenido, fueran o no candidatos a ser automatizados, garantizando la coherencia documental y adecuando los circuitos de trabajo a la nueva realidad de producción. Asimismo, se aprovecha para fijar nuevos criterios de análisis documental, con el fin de mejorar la calidad, evitar errores e incluir nuevos puntos de acceso a la información.

En el momento de plantear la propuesta, el departamento de Ingeniería estaba en pleno proceso (en el que continua) de desarrollo de la nueva versión del MAM. A su vez, y a petición del departamento de Documentación, se iniciaba el desarrollo de un nuevo Gestor de Metadatos (que también sigue en marcha), en la línea de las bases de conocimiento, para la gestión de vocabularios controlados. Así pues, y ante la urgencia de actualizar un TrAgEdi inoperativo, se optó por un desarrollo técnico menos ambicioso, centrado en herramientas Office y basado en reglas fijas, sin intervención de tecnologías IA.

El proyecto lo desarrolla un equipo formado por dos documentalistas y dos ingenieros trabajando a dedicación parcial, desde marzo de 2020. La primera versión se implementa en septiembre de 2020, revisada en julio de 2021. En julio de 2022 se actualiza con una segunda versión, operativa hasta la actualidad.

En lo referente a su alcance, el sistema desarrolla, por un lado, un corpus de estandarización para la catalogación del conjunto de materiales relativos a eventos de competición deportiva y de eventos no competición, aunque directamente asociados a estos, ya sea en la previa (entrenamientos, ruedas de prensa, desplazamientos, totales, etc.) o en el post evento (ruedas de prensa, totales, salidas, análisis técnico, etc.). Quedan excluidos los eventos no competición, tales como actos promocionales, presentaciones de fichajes, destituciones, etc.

Por otro lado, el sistema automatiza la precatalogación de los materiales relativos a piezas editadas (noticias, recursos de pantalla, montajes musicales, etc.), noticias de agencia, directos y diferidos de transmisiones deportivas, brutos de rodaje y brutos de transmisión, creando los títulos, descriptores, resúmenes y parte de la descripción de planos correspondientes a eventos tipo partido Equipo A vs Equipo B de las principales competiciones deportivas masculinas y femeninas con seguimiento informativo en 3Cat. Los deportes automatizados son: fútbol, baloncesto, hockey patines, balonmano, rugby, fútbol sala, hockey hierba y waterpolo.

Finalmente, el sistema replantea los flujos de trabajo en el tratamiento documental de los deportes, abarcando a todos los materiales en la entrada y selección, precatalogación y catalogación final.

3.2. Descripción del sistema

El sistema se configura en base a cinco componentes:

1. *Feeds* xml OPTA. El departamento de Deportes, mediante suscripción, recibe ficheros xml con datos de competiciones deportivas que se van actualizando constantemente. Dichos archivos se almacenan en dos carpetas compartidas: una, utilizada por Deportes y la otra, utilizada por el departamento de Documentación.

2. Digitition. El sistema actúa en las fichas documentales del área de producción y de archivo del MAM de 3Cat. El proceso incluye el uso de la herramienta del sistema para la exportación de listados de registros documentales, en formato Excel.
3. Base de Datos OPTA-Digitition de Microsoft Access. Es el corazón del sistema, donde se procesan los datos, haciendo de puente entre los *feeds* de OPTA y las fichas documentales del MAM. La base de datos se divide en tres áreas: (1) Metadatos, (2) OPTA y (3) Digitition.

El área de Metadatos gestiona los metadatos usados en el proceso. Un primer apartado administra su tipología, en un equivalente a lo que serían las clases de una ontología. En un segundo apartado, se organiza la biblioteca de instancias, similar a una base de conocimiento, donde se establecen las equivalencias de los datos importados de OPTA con la forma y ubicación (campo de destino de la ficha documental) del metadato de Digitition. Un tercer apartado controla las competiciones deportivas, indicándose los *feeds* de OPTA que se deben importar. Finalmente, en un cuarto apartado, se procesan los conversores-correctores de las cadenas de texto creadas automáticamente.

El área de OPTA ejecuta la acción de importación de los *feeds* xml, y se genera la catalogación automatizada de los campos a partir de los metadatos gestionados en la primera área. El área de OPTA incluye un apartado para la exportación de los títulos de los eventos deportivos creados automáticamente para su uso en la selección-identificación de los materiales de Digitition que se van a tratar.

Finalmente, el área de Digitition, contiene un apartado para la importación de los listados de materiales del MAM que se quieren precatalogar, otro para el emparejamiento de dichos materiales con la catalogación automatizada de los eventos realizada en el área de OPTA, y un tercero para la validación-concreción de la catalogación de las fichas documentales de los materiales de Digitition. En un último apartado, se ejecuta la exportación de los registros ya precatalogados hacia el MAM.

4. Macros y hojas de Microsoft Excel. Se utilizan como herramienta de exportación-importación de los datos entre los *feeds* xml de OPTA, la base de datos Access y Digitition. A su vez, las macros realizan tareas de estructuración de la información para un mejor procesamiento de los datos por parte de Access.
5. Digitition Importer. Es una aplicación creada con Java para ejecutar la migración de la información contenida en los ficheros Excel creados por la base de datos Access hacia los campos de la ficha documental de Digitition.

3.3. Fundamentos teóricos del sistema OPTA-Digitition

En concordancia con el objetivo del artículo, se evita entrar en los aspectos técnicos del sistema para centrarse en los fundamentos teóricos que lo sustentan, no sin antes elogiar la labor realizada por el equipo técnico en el traslado de los conceptos teorizados, explotando al máximo las posibilidades que ofrecen las herramientas ofimáticas usadas.

a) Análisis documental en base al evento

Los medios de comunicación basan la mayor parte de la creación de sus contenidos en eventos que suceden en el mundo (Dayan y Katz, 1992). La edición periodística se encarga de filtrar y jerarquizar dichos eventos, decidiendo qué es noticia y qué merece ser difundido (Shoemaker y Vos, 2009). Como se ha explicado, la multiplicidad de plataformas ha comportado un aumento muy destacado de la producción de contenidos en múltiples formatos. No obstante, el volumen de historias contadas se mantiene en una línea relativamente constante, si bien ligeramente al alza, a razón de las nuevas narrativas en entornos cross-media, en ningún caso comparable con el alud documental.

En este sentido, el concepto de evento se erige como elemento central para la organización, indización y recuperación del contenido multiplataforma, ya que permite aglutinar en un solo ítem y a modo de nodo semántico la información presentada en múltiples formatos (Tzelepis et al., 2016). En palabras de Caswell y Dörr (2017) catalogar documentos a partir del evento es clave para avanzar en sistemas automatizados de gestión documental.

Para desarrollar el concepto, partimos de un ejemplo real: el partido de vuelta de los cuartos de final de la Liga de Campeones de fútbol masculino 2024-2025 entre los equipos Borussia Dortmund y el Barça, jugado en el estadio Signal Iduna Park de Dortmund y que, pese a que terminó con victoria del Borussia por 3 a 1, el Barça se clasificó para la semifinal de la competición. Estamos ante la descripción textual de un evento deportivo noticiable, que ha generado gran cantidad de materiales, tanto en origen como de emisión, en diversidad de formatos y difundido en múltiples pantallas. Todos estos materiales giran, con sus variables y especificidades, alrededor de un mismo evento y a los acontecimientos circunscritos a él. Asimismo, dicho evento se desarrolla en un entorno relacional con otros eventos que también giran en torno a él.

Tenemos pues un ecosistema de decenas, incluso centenares de documentos que, con mayor o menor profundidad, pasan por un proceso documental que los describe con un conjunto de metadatos comunes relativos al evento en cuestión. Si partimos de un sistema basado en el análisis del documento, tendremos que repetir multitud de veces un tratamiento documental casi idéntico, donde se replicarán, una y otra vez, metadatos que referencian un único hecho. En cambio, si nos centramos en describir el evento, la reducción del esfuerzo resulta evidente.

En consecuencia, se estableció un cambio de paradigma en el tratamiento documental de los deportes en el archivo de 3Cat, consistente en invertir el foco de atención del análisis documental, pasando de describir documentos que relatan eventos, a describir eventos relatados en documentos, para luego replicar dicha descripción a los documentos que se refiere al evento descrito.

En el caso de la información periodística deportiva, esta inversión conceptual es ideal, dado que el evento deportivo es el centro de la producción de contenido, en base a tres etapas informativas muy marcadas: la previa, el evento en sí y el post evento.

b) Estructuración y estandarización en base principios de la web semántica, lenguajes y normas documentales

Una vez determinado que el objeto inicial de análisis es el evento, el siguiente paso ha sido encontrar la manera de representarlo.

El primer condicionante es la casuística del circuito de trabajo del análisis documental del archivo de 3Cat, donde el título del documento es el principal punto de acceso a la información, desde el primer al último estadio de la cadena documental. A pesar de seguir unas pautas normativas para su construcción, es un campo de texto libre, consecuentemente, no estructurado, una condición indispensable para su automatización.

Para atajar el problema, se planteó la solución a partir de los principios de la web semántica y los datos enlazados (Berners-Lee et al., 2001) con el propósito de establecer los eventos como conceptos unívocos, que pudiesen ser descritos en propiedades y relaciones que, por sí mismas, conformaran conceptos independientes y con valor semántico, gestionados a través de vocabularios controlados.

En este sentido, existen múltiples iniciativas para adaptar la catalogación a los preceptos de la web semántica. Entre ellas, destacan la ontología EBUCorePlus (<https://tech.ebu.ch/metadata/ebucoreplus>), elaborada por la Unión Europea de Radiodifusión, EBU, y el PBCore (<https://pbcore.org/>), creado por la comunidad de radiodifusión pública de los Estados Unidos de América. Sin embargo, como segundo condicionante, estos modelos no incorporan el concepto evento con la intencionalidad identificativa pretendida, sino que lo usan para representar el ciclo de vida del objeto documental (publicación, emisión, etc.).

Finalmente, el tercer obstáculo lo constituye el hecho de que el modelo de metadatos del MAM de 3Cat no está adaptado a los esquemas semánticos citados, sino que sigue una estructura de ficha documental con campos estructurados y no estructurados heredada de sistemas anteriores. Por consiguiente, a falta de un modelo que encajara y ante la necesidad de estructurar para automatizar, fue necesario idear una estructura que mantuviese el modelo de datos existente y que fuera aplicable al campo título.

Para ello, se analizaron múltiples esquemas semánticos para representar metadatos de eventos deportivos, así como bases de conocimiento y modelos de descripción bibliográfica. En concreto, los esquemas estudiados fueron: SportsML de IPTC (<https://iptc.org/standards/sportsml-g2/>), la estructura de datos de los feeds en xml de OPTA (<https://www.statsperform.com/opta-feeds/>), SportsEvent de Schema.org (<https://schema.org/SportsEvent>), Olympic Data Feed (<https://odf.olympictech.org/>), y la Sport Ontology de la BBC (<https://www.bbc.co.uk/ontologies/sport-ontology/>). En lo referente a las bases de conocimiento, el estudio se centró en Wikidata (<https://www.wikidata.org/>). Asimismo, se estudió el modelo de descripción bibliográfica BIBFRAME de la Library of Congress (<https://www.loc.gov/bibframe/>), basado en datos enlazados. Finalmente, para adecuar su forma y establecer las normas de concatenación, se trabajó en base a la norma **ISO 690 (2021)** y los encabezamientos de materia de la **Library of Congress (2019)**.

El resultado es la creación de una norma para referenciar eventos deportivos en formato cadena de texto, configurada en tres grupos semánticos: (1) identificación del evento, (2) resultado y fase del evento y (3), descripción del contenido del documento. La referencia se construye sobre una base de segmentos que serían el equivalente a las clases, que contienen enunciados, definidos como unidades básicas de información, de las cuales se puede extraer uno o más metadatos unívocos, el equivalente a las instancias. Mediante la agregación jerárquica de los segmentos, de más genérico a más específico, se va configurando un título descriptivo, preservando al máximo el lenguaje natural. La propiedad jerárquica de los bloques semánticos es relevante, ya que la suma de los segmentos añade significado al bloque anterior, ya de por sí, entendible y recuperable, permitiendo una ordenación alfabético-jerárquica de resultados y añadiéndole propiedades de lenguaje documental precoordinado a un campo de texto libre.

A todo lo expuesto, se sumó el desafío de encontrar una estructura de segmentos aplicables a todo tipo de eventos de competición deportiva, donde fuese posible encajar las múltiples casuísticas de los deportes, sus competiciones, y su representación en los documentos, según tipo de documento y contenido. Para establecer los patrones se analizaron dos años (2018 y 2019) de títulos y descriptores de editados catalogados por el archivo de 3Cat y se testó su idoneidad en más de un millar de competiciones de los deportes y disciplinas reconocidos por el Comité Olímpico Internacional (COI).

En aras de la preservación de la coherencia documental, la estandarización no puede aplicarse solo a los documentos susceptibles de ser precatalogados automáticamente, si no al conjunto de la colección, buscando a la vez un equilibrio en la incorporación del concepto evento sin romper del todo con las normas aplicadas hasta entonces. A grandes rasgos, la referencia se estructura de la siguiente manera:

Evento multideportivo: Deporte (disciplina, género, etc.): **Competición** (categoría, año/temporada): **Estructura de la competición** (fase, grupo, eliminatoria, jornada, etc.): **Concreción del evento** (participantes, recorrido, prueba, etc.): **Resultado: Fase del evento** (previa, post): **Descripción del contenido del documento.**

A continuación, se muestra una selección de ejemplos de títulos creados a partir de la norma, extraídos del MAM Digition (traducidos del catalán al castellano):

Estandarización automatizada:

- Baloncesto masculino: Liga NBA: 2023-2024: Play-off Conferencia Oeste: Final: Partido 3: Dallas Mavericks vs Minnesota Timberwolves: 116-107: Resumen.
- Balonmano masculino: Liga Europea: 2024-2025: Fase principal: Grupo 1: J4: GOG vs Granollers: Previa: Entrenamiento del Granollers en el Arena Svendborg.

- Fútbol femenino: Campeonato del Mundo: 2023: Semifinal: España vs Suecia: 2-1: Montaje musical postproducido “A la final” / Lead.

Estandarización no automatizada:

- JJOO 2024: Gimnasia artística femenina: Final por aparatos: Barra de equilibrios: Medalla de oro para Alice d’Amato (ITA).
- Tenis en silla de ruedas masculino: TRAM Barcelona Open: 2024: Final: Gustavo Fernández vs Martín de la Puente: 6-2, 4-6, 7-6(10). Partido entero.

c) Automatización basada en equivalencias y concatenación de texto en una estructura de metadatos fija

Resuelto el tema de la estructuración y estandarización, el siguiente desafío fue cómo transformar de forma automatizada una información codificada externa a una descripción de la ficha documental basada en estándares y normas de análisis propias.

Para ello, se utilizó el principio de equivalencia, una solución simple pero efectiva que establece el lugar y la forma de los datos de OPTA segmentados en clases en los metadatos, instancias que construyen la información contenida en los campos de la ficha documental título, resumen, descriptores, fecha del acontecimiento, ámbito geográfico y, parcialmente, la descripción de planos.

The screenshot shows a web form for entering metadata for a football team. The 'Esport' dropdown is set to 'Futbol masculí'. The 'Tipus metadada' dropdown is set to 'Equip'. The 'Arxiu origen' dropdown is set to 'IMP EQ EXCEL'. The 'Validat' checkbox is checked. The 'OPTA id:' field contains 't178'. The 'OPTA text:' field contains 'Fútbol Club Barcelona'. The 'Desambiguació:' field contains 'Equip de futbol masculí, Fútbol Club Barcelona, Espanya, Barcelona'. The 'Títol:' field contains 'Barça'. The 'Resum:' field contains 'el FC Barcelona'. The 'DeTema:' dropdown is set to 'BARÇA'. The 'Dpla:' field contains 'Barça'. The 'Text:' field is empty. The 'Observacions:' field is empty.

Figura 1. Equivalencias Equipo FC Barcelona

Fuente: captura de la sección metadatos de la BBDD OPTA-Digition de 3Cat, 16/4/2025

Para entender el proceso, observemos la **figura 1** y cómo se traduce el concepto equivalencia en el sistema OPTA-Digition. Partimos de un código OPTA, el *t178*, correspondiente a la clase *Tipo metadato: Equipo*, que identifica al FC Barcelona. En el formulario establecemos el lugar, es decir, el campo de la ficha documental, y la forma que toma el metadato. En este caso, al campo título, *t178* equivale a *Barça*, en el resumen, *el FC Barcelona*, en el descriptor *BARÇA*, y en la descripción de planos *Barça*.

The screenshot shows a web form for entering event details. The 'Esport' dropdown is set to 'Futbol masculí'. The 'Validat' checkbox is checked. The 'Títol:' field contains 'Futbol masculí: Lliga de Campions: 2024-2025: Quarts de final: Tornada: Borussia Dortmund vs Barça: 3-1'. The 'DeTema:' field contains 'FUTBOL * ESPORT MASCULI * LLIGA DE CAMPIONS * TEMPORADA 2024-2025 * BORUSSIA DORTMUND * BARÇA * DORTMUND * COMPETICIO'. The 'Resum:' field contains 'Partit de futbol masculí entre el Borussia Dortmund i el FC Barcelona corresponent a la tornada dels quarts de final de la Champions League 2024-2025, disputat al Signal Iduna Park de Dortmund, Alemanya i arbitrat per Maurizio Mariani, que ha acabat amb el resultat de 3 a 1. Victòria local del Borussia Dortmund. El guanyador de l'eliminàtoria ha estat el FC Barcelona'. The 'Dpla:' field contains 'LLOC 2025 0415 Signal Iduna Park, Dortmund, Alemanya'. The 'idCodTemp:' field contains 'c5-2024'. The 'IdJornada:' field contains 'c5-NR4-RQuarter-Finals-GV2'. The 'EstatLoc:' field contains 'FullTime'. The 'ID_Estat:' field contains 'v1369'. The 'Arbitre Nom i Cognoms:' field contains 'Maurizio Mariani'. The 'Id Equip Local:' field contains 't157'. The 'Id Equip Visitant:' field contains 't178'. The 'Guanyador Partit:' field contains 't157'. The 'Ganyador Eliminàtoria:' field contains 't170'. The 'Incidència 1:' field contains '1'. The 'Incidència 2:' field contains '1'. The 'Incidència 3:' field contains '1'. The 'Temps Afegit:' field contains '1'. The 'Resultat Local:' field contains '3'. The 'Resultat Visitant:' field contains '1'. The 'Resultat Penals Local:' field contains '1'. The 'Resultat Penals Visitant:' field contains '1'.

Figura 2. Catalogación automatizada de un evento deportivo OPTA-Digition.

Fuente: captura de la sección OPTA de la BBDD OPTA-Digition, 16/4/2025

En la parte inferior de la **figura 2** vemos una sucesión de los códigos OPTA que intervienen en la catalogación automatizada del evento, entre ellos el referente al *IdEquip Visitant*, con valor t178. A partir de dichos códigos, y la aplicación del principio de equivalencia descrito anteriormente, mediante consultas y en base a la estructura de metadatos diseñada para cada campo, se van construyendo las cadenas de texto que configurarán la descripción del evento. El proceso finaliza con la aplicación de reglas de sustitución-corrección del texto creado (apóstrofes, etc.), para que se adecue al lenguaje natural.

El sistema OPTA-Digition procesa los datos dinámicos de cada evento mediante una estructura fija, proporcionando una respuesta finalista sin posibilidad de alucinación o variaciones, por lo que el resultado es totalmente fiable. Pueden existir errores, mínimos, en la captación de los *feeds* de OPTA, pero su detección es muy rápida ya que, si la información de un campo es correcta, también lo será en los campos restantes.

d) Análisis en cascada en base al principio del mínimo común múltiple

A pesar de la inversión conceptual evento-documento, el sistema documental de 3Cat, sigue basándose en la gestión del documento en todas las fases de la cadena documental. Por este motivo, es necesario revertir el proceso.

Este cambio se realiza a partir del concepto de análisis en cascada, una adaptación un tanto libre de la catalogación por copia que rige, desde hace décadas, la catalogación bibliográfica. Este proceso se hace en base al principio de mínimo común múltiple, donde, mediante una acción de emparejamiento automatizada, se replican los metadatos comunes del evento a los documentos relacionados a partir de una búsqueda coincidente del título del evento con la raíz de los títulos de los documentos importados desde el MAM a la base OPTA-Digition. Una vez finalizado el emparejamiento, se valida la descripción y se exporta de vuelta al MAM integrándose a la ficha documental, donde el documentalista, finalmente, completará la catalogación añadiendo o eliminando, según el caso, información para singularizar el documento o aportar el valor añadido pertinente.

4. RESULTADOS

Para evaluar el impacto del sistema OPTA-Digition en la gestión documental de los documentos de temática deportiva del archivo de 3Cat, y en base a datos obtenidos mediante consulta al MAM, se ha realizado un análisis volumétrico de los materiales archivados que se emitieron durante una semana.

Entre el 10 y el 16 de marzo de 2025, y excluyendo las ruedas informativas de los servicios de 24h, 3Cat emitió 104 programas con información deportiva (82% en formato vídeo y 18% en audio), a través de sus múltiples ventanas (63% en televisión, 15% en radio y los 22% restantes en medios digitales). Estos programas crearon su contenido en base a 179 eventos: un 58% relativos a competiciones deportivas, un 37% a eventos de previa o post competición y un 5%, eventos no competición. Se informó sobre 40 competiciones de 13 deportes diferentes (61% fútbol, 11% motor, 7% baloncesto, 6% ciclismo en carretera, y el 15% restante, atletismo, fútbol sala, fútbol americano, balonmano, hockey patines, golf, tenis, deportes de nieve y waterpolo).

El 95% de los eventos fueron referenciados según el nuevo estándar. Dado que el sistema OPTA-Digition cubre el 52% de las competiciones informadas, un 65% de las 179 referencias se crearon de forma automatizada.

Estos eventos fueron relatados en 382 piezas informativas (94% editados en vídeo y 6% en secciones y cortes de audio de programas radiofónicos) y en 24 transmisiones (42% en radio y 58% en televisión) que, a su vez, fueron creadas en base a 326 materiales originales (72% noticias de agencias e intercambio, 28% brutos de rodaje y de transmisión). Sumados a los 118 materiales de programas (paralelos de antena, diferidos, etc.), generaron un total de 850 documentos.

Del total de 732 editados, originales y transmisiones, un 58% se referían a eventos competición, 34% a eventos previa o post competición y un 8% a no competición. De ellos, el 91% contenían un título con

la raíz estandarizada según las nuevas normas. El 64% de los títulos se crearon de forma automatizada y el 36% restante, de forma manual. Finalmente, el sistema OPTA-Digition permitió la precatalogación automatizada de un 40% de los materiales. Cabe subrayar que, con los *feeds* de competiciones disponibles en OPTA, si se completase el desarrollo del sistema y se incluyesen otros modelos de competición, se lograría automatizar más del 85% de los títulos y el 54% de la precatalogación.

En lo referente a los recursos humanos, y según los datos de planificación del ente, en febrero de 2015, 10 documentalistas realizaban el tratamiento documental de deportes. En febrero de 2025, la asignación de documentalistas a esa tarea se redujo a casi la mitad, en concreto a 5,5.

La constatación de la optimización de recursos humanos en la gestión documental es el estado actual de la catalogación. Una vez terminada la temporada de las principales competiciones deportivas, se ha realizado una consulta al MAM con el rango de fechas 15 de mayo de 2024 a 15 de mayo de 2025, para comprobar cuántos materiales están pendientes de catalogar. El resultado es que el 97% del fondo ha finalizado su tratamiento documental. Y del 3% del fondo pendiente, un 2% está precatalogado, consecuentemente, es recuperable.

A su vez, el sistema OPTA-Digition ha supuesto una remarcable mejora de calidad en la descripción documental.

Tabla 1. Comparativa descripción de editados 2015-2025.

Fuente: elaboración propia en base consulta al MAM Digition 3Cat, 25/4/2025. Traducido del catalán al castellano.

Campo	20215	2025
Título	Fútbol: Barça - Valencia (2-0). Resumen del partido	Fútbol masculino: 1a División: 2024-2025: J21: Barça vs Valencia: 7-1: Resumen
Resumen	Correspondiente a la 32ª jornada de 1ª división de la liga 2014-2015, que fue arbitrado por José Luis González González y que se jugó en el Camp Nou de Barcelona.	Partido de fútbol masculino entre el FC Barcelona y el Valencia CF correspondiente a la 21ª jornada de LaLiga, Liga de Primera División española 2024-2025, disputado en el estadio Olímpico Lluís Companys de Barcelona y arbitrado por César Soto Grado, que ha terminado con el resultado de 7 a 1. Victoria local del FC Barcelona.
Descriptor	FUTBOL * BARCELONA * LIGA (2014-2015) * BARÇA * VALENCIA (FC)	FUTBOL * DEPORTE MASCULINO * LIGA DE PRIMERA DIVISION * ESPAÑA * TEMPORADA 2024-2025 * BARÇA * VALENCIA (FC) * BARCELONA * COMPETICION
D. de planos	Partido de fútbol Barça - Valencia 2-0 Gol de Luis Suárez 1-0 min. 1 Claudio Bravo para un lanzamiento de penalti Gol de Messi 2-0 min. 93 (número 400 con el Barça)	LUGAR 2025 0126 Estadio Olímpico Lluís Companys, Barcelona Partido de fútbol masculino Barça vs Valencia 7-1 1-0 Barça GOL Frenkie de Jong ASISTENCIA Lamine Yamal min. 3 2-0 Barça GOL Ferran Torres ASISTENCIA Alejandro Balde min. 8 3-0 Barça GOL Raphinha ASISTENCIA Fermin López min. 14 4-0 Barça GOL Fermin López ASISTENCIA Pau Cubarsi min. 24 5-0 Barça GOL Fermin López min. 45+4 5-1 Valencia GOL Hugo Duro ASISTENCIA Diego López min. 59 6-1 Barça GOL Robert Lewandowski ASISTENCIA Fermin López min. 66 7-1 Barça GOL César Tárrega AUTOGOL Valencia min. 75

La **tabla 1** presenta un ejemplo de la descripción documental de dos editados relativos a resúmenes de partidos de fútbol masculino, que comparten los mismos protagonistas, lugar y competición, separados por diez años de diferencia. El primero, está descrito de forma totalmente manual y el segundo, está descrito de forma automatizada, exceptuando las acciones (goles) de la descripción de planos que, si bien están estructuradas según las nuevas normas y preparadas para ser automatizadas, aún falta el desarrollo técnico para hacerlo. Hay que señalar que, en cualquier caso, el análisis se realiza una sola vez en la base de datos Access, replicándose a todos los materiales referentes.

Si comparamos la descripción de ambos editados, la mejora resulta evidente. En el campo título, la descripción de 2015 es ambigua e incompleta, en el sentido que no indica la competición, la jornada o el género. Si sólo atendemos al título, este puede referirse a un partido de la Copa de la Reina de fútbol femenino o a un partido de Liga de Campeones de cualquier temporada, requiriéndose de otros campos documentales para que la información resulte completa. En cambio, en la descripción de 2025, se disipan todas las dudas, quedando representado unívocamente el evento. Asimismo, sus

características estructurales, relatadas en el apartado 3.3(b), mejoran la ejecución de búsquedas por texto. Por ejemplo, si se quiere obtener un listado de todos los partidos de la jornada 21 de la competición en cuestión, sólo es necesario hacer una búsqueda por frase exacta “Fútbol masculino: 1ª División: 2024-2025: J21” para obtener resultados pertinentes. Otro ejemplo de mejora es la posibilidad de satisfacer necesidades tales como saber todos los partidos del Valencia como visitante o cualquier otra temporada o competición: “Fútbol masculino & “vs Valencia”.

En lo referente al resumen, se ha automatizado buscando al máximo la apariencia de lenguaje natural, incluyendo información complementaria relevante (árbitro, estadio, victoria local o visitante, etc.) y aumentando los puntos de acceso con sinónimos (FC Barcelona, LaLiga, Primera División Española, etc.).

Si se compara el campo descriptor, observamos que, en el ejemplo de 2025, el número de términos ha aumentado significativamente. Eso se debe a la voluntad de adecuar la descripción a las herramientas de recuperación del MAM, que, además de la búsqueda por texto libre y por campos, incorpora un sistema de filtraje por campos controlados. La intención es que cada segmento del título disponga de su equivalente en un descriptor, pudiéndose hacer una búsqueda pertinente utilizando solo términos del tesauro.

Finalmente, el campo de descripción de plano de 2025 incorpora un encabezamiento de localización geo temporal, una descripción simplificada del evento (automatizados) y una sucesión de las acciones más destacadas, estructuradas en base a la información disponible en los feeds de OPTA. La información se presenta a partir de la concatenación de la acción (diferenciada en mayúscula) y sus protagonistas, en minúscula, creando binomios informativos recuperables por frase exacta. Pese a que la estructura no responde a la del lenguaje natural, se ha procurado que, a nivel visual, sea lo máximo entendible posible. Con los binomios y su estructura, se persigue resolver necesidades informativas tipo: goles de “x” equipo (“Barça GOL”) o goles de “x” jugador (“GOL Ferran Torres”).

Lo más relevante en la mejora de la calidad del análisis documental es la consecuencia derivada del propio proceso de automatización, que fija una estandarización común a todos los documentos donde se aplica, evitando omisiones, variaciones o errores en el análisis, y aportando la misma estructura de metadatos a todos los documentos, lo cual permite prever la información disponible para su recuperación.

Los resultados demuestran que se han alcanzado con creces los objetivos planteados en el inicio del desarrollo del proyecto. Podemos afirmar que OPTA-Digition es un caso de éxito de automatización del tratamiento documental de materiales de temática deportiva en medios de comunicación audiovisuales que ha contribuido a mitigar el reto de gestionar el aumento de la carga de trabajo con unos recursos humanos limitados.

5. DISCUSIÓN Y CONCLUSIONES

El sistema OPTA-Digition funciona a pleno rendimiento, pero no está terminado. Falta el desarrollo técnico de la descripción de planos y la incorporación de otros modelos de competición (ciclismo, motor, tenis, etc.). También se imagina la posibilidad de automatizar la descripción de las historias que no son eventos de competición, la mejora de automatismos, su posible integración en el MAM o su migración hacia otros sistemas tecnológicos. Después de cinco años, también se ha detectado la posibilidad de mejora y simplificación de la estructura de segmentos. Queda pendiente su implementación en medios digitales, por lo que no se puede hacer una evaluación global de su impacto en el entorno multiplataforma.

A nivel tecnológico, el sistema OPTA-Digition no supone ningún avance. OPTA-Digition es hijo de su tiempo, pero, sobre todo, de sus circunstancias. Pese a ello, actualmente sigue dando respuesta al objetivo para el que fue creado, optimizando muy eficazmente el análisis documental. Es un proyecto adaptable a otras tecnologías, que puede servir de inspiración para procesos similares en otros sistemas de información.

Lo novedoso del sistema es su estructuración y estandarización basada en el evento, lográndose un modelo para referenciar todo tipo de eventos deportivos. En este sentido, parte del éxito reside en la propia naturaleza del evento deportivo, muy estructurado y jerarquizado; su tratamiento informativo,

calendarizado y previsible, y la disponibilidad de un juego de datos externos. Dichas características no tienen paralelo en eventos que no sean de competición deportiva. Por consiguiente, trasladar el concepto de análisis por evento al resto de las secciones informativas o al entretenimiento es un reto mayúsculo, ya que los eventos culturales, sociales, políticos o económicos no son fáciles de automatizar mediante el método utilizado en el sistema OPTA-Digition. Su estructura y previsión son, por lo general, más variables, lo cual complica la tarea de extracción de patrones. Tampoco existen juegos de datos externos que permitan un modelo de automatización similar al desarrollado en deportes.

Llegados a este punto, cabe preguntarse ¿es posible escalar y exportar el modelo? La respuesta es sí. Sobre la escalabilidad, el hecho de tener un fondo documental amplio, bien descrito y coherente, augura una captura de patrones muy laboriosa pero factible. En cuanto a su posibilidad de exportación, aunque originalmente la propuesta desarrollada ha sido concebida como una solución individualizada, sus particularidades la hacen apta para ser implementada en otros sistemas de gestión documental, sean o no del ámbito deportivo. No obstante, en ambos casos, es necesario un esfuerzo importante, ajustando las expectativas de resultados según las posibilidades, y, sobre todo, disponer de tiempo para concebirlos. El trabajo puede aligerarse con la ayuda de la tecnología, ya sean herramientas de automatización tipo Power Automate de Microsoft, o las que se deriven de la IA, especialmente las relativas al Procesamiento del Lenguaje Natural, PLN.

En este sentido, y en una lógica de *work in progress*, las herramientas IA siguen avanzando en aplicaciones modulares, en previsión de que se acoplen y ofrezcan un servicio integral de catalogación automatizada para sustituir al máximo la catalogación humana, y, por qué no, incorporar los conceptos planteados en el Sistema OPTA-Digition. Pese a los logros de la IA en este ámbito, véanse las experiencias presentadas en la duodécima edición del FIAT/IFTA Media Management Seminar, celebrado en Cardiff, bajo el lema *IA and Human Collaboration: Partners in Archiving?* (FIAT/IFTA Media Management Commission, 2025), donde se intuye que esta realidad no será inmediata. Así las cosas, pendientes de una disminución de los costes, clarificación de aspectos legales o de gestión de metadatos (European Broadcasting Union, 2025), nos esperan años en que los servicios de documentación deberán seguir trabajando con estrategias complementarias a la IA.

En lo referente al rol del documentalista, se habla mucho del futuro de la profesión en un entorno dominado por la IA. No es una cuestión circunscrita a la Documentación audiovisual, sino que abarca a casi todos los ámbitos laborales donde, desde el Derecho al Periodismo, pasando por la Traducción y un largo etc., se plantea qué hacer cuando la máquina sustituya, aunque sea parcialmente, la labor del profesional. La incertidumbre de la respuesta es objeto de una apasionante discusión, con posicionamientos ludistas, escépticos o entusiastas, que trasciende el ámbito académico, convirtiéndose en un debate público general. Rehuyendo escenarios apocalípticos, y acotándolo al ámbito de la Documentación audiovisual, el colectivo asume lo inevitable y se presentan nuevos roles, tales como curadores de contenidos, validadores, verificadores o ‘aportadores’ de valor añadido como alternativa ocupacional. Todos estos roles encajan perfectamente con la profesión, teniendo en común el hecho de focalizarse en la fase final del análisis documental y en la recuperación de la información. Es más, en su mayoría, no dejan de ser nuevas terminologías para describir tareas que ya hacen los documentalistas, ajustadas, eso sí, a la nueva realidad tecnológica. Ante los avances tecnológicos, tal y como se demostró con la digitalización, no se duda de la capacidad de adaptación del colectivo ante el nuevo escenario.

Esta declaración de intenciones no está exenta de desafíos. Por ejemplo, si nos fijamos en el rol de validador en los procesos de catalogación automatizada con herramientas IA, el documentalista dispone del conocimiento acumulado de años de catalogación manual para resolver, en muy poco tiempo, su idoneidad. Pero, a la larga, ¿cómo se discernirá si el resultado que proporciona la máquina es pertinente si no se ha participado en la resolución de la duda o se ha perdido el hábito de hacerlo? En cualquier caso, esta observación se da de bruces con la realidad: una vez terminado el período de testeo, y ante la imposibilidad de destinar esfuerzos continuados a la validación de resultados, los servicios de documentación tienden a integrar en el circuito documental aquellas soluciones IA que proporcionan índices de fiabilidad altos, tales como la traducción automática o la transcripción del audio a texto, asumiendo el error sin corregirlo. ¿Pasará lo mismo con la indización automatizada?

Existe el riesgo de acomodarse a un resultado suficiente, pero no al más pertinente. Esta situación puede llevar a una mediocridad en el proceso de catalogación con la que los documentalistas no deberían sentirse cómodos, aunque los libere del trabajo de describir. Con ello, no se pretende obviar el hecho de que la catalogación manual contiene errores, y que procesos de automatización ayudan a mitigarlos, pero sí plantear un futuro donde aunque sea necesario, la clave no sea “pulir y abrillantar”, sino pensar más y mejor.

En este sentido, se alzan múltiples voces y experiencias donde el valor de la profesión no se relega al final del proceso, sino que el criterio documental es decisivo en la implementación de las nuevas tecnologías, reivindicando el papel central del documentalista en la arquitectura e implementación del proyecto tecnológico, decidiendo qué se quiere y cómo se quiere, como acto de empoderamiento hacia la IA, la cual no viene a sustituir a los documentalistas, sino a ampliar sus capacidades (Auserón, 2025).

Concluimos reafirmandonos en el concepto de la catalogación de la historia-evento como punto de partida para la optimización de la gestión documental multiplataforma, intuyendo soluciones en base a la conjugación de lenguajes documentales, ya sean antiguos (encabezamiento de materias, tesauros, etc.) como nuevos (ontologías, grafos de conocimiento), guiados por los preceptos de la web semántica, con la finalidad de conseguir un estándar común para todo tipo de eventos que permita el desarrollo de una norma de referencia general.

Sin desdeñar roles presentes y futuros, el ejemplo OPTA-Digition reafirma la hipótesis planteada. La óptica documental como parte activa en el diseño y desarrollo de los procesos de automatización se revela como determinante. Ante la indudable necesidad de automatizar la catalogación y de un futuro incuestionable marcado por la IA, el verdadero valor añadido del documentalista es su contribución, como experto en la estructuración y estandarización de la información, en la creación de modelos de calidad que sirvan para que la representación de la realidad sea lo más precisa y pertinente posible.

Ante la invisibilidad del algoritmo, hay que trabajar en lo visible, en estrecha colaboración con el área técnica para desarrollar herramientas de trabajo, según el punto de vista y necesidad de la Documentación. De ahí la relevancia de los fundamentos teóricos del sistema OPTA-Digition. Se trata de un ejercicio intelectual que demuestra el valor de relacionar conceptos de la Ciencia de la Documentación, uniéndolos y adaptándolos para resolver un caso práctico, donde la piedra angular no es la tecnología, sino el binomio documentalista-ingeniero en el proceso de creación de la arquitectura de los metadatos.

6. NOTAS

1. El presente trabajo es el primero de una serie de tres artículos previstos en el marco del programa de doctorado en Periodismo y Comunicación de la Universitat Autònoma de Barcelona, centrado en la estructuración, estandarización y automatización de la gestión documental en medios de comunicación audiovisuales.
2. A pesar de la vinculación laboral del autor con 3Cat, el artículo es un trabajo académico independiente. La discusión y conclusiones contenidas pueden no representar la línea de la dirección del departamento de Documentación de la empresa.

Información complementaria ↑

Fuentes de financiación

No aplica.

Material suplementario

No aplica.

Disponibilidad de datos

No aplica.

Agradecimientos

Al equipo de desarrollo del proyecto OPTA-Digition de 3Cat: Mònica Batlle Jordà, documentalista, codesarrolladora en el área de documentación; Montserrat Sort Serra, jefa de proyectos Business Intelligence y Carles Cano Fernández, ingeniero, de 3Cat, ambos codesarrolladores en el área técnica.

A Cristina Canales Martínez, responsable del área de deportes del departamento de Documentación de 3Cat, y a las documentalistas Agnès Lucena Santacreu y Anabel Martínez García, por su contribución en la adecuación de la nueva estructura, estandarización y circuito de trabajo a la gestión documental de los deportes en el archivo de la CCMA-3Cat.

A Montse Català Freixa, jefa del departamento de Documentación, por sus aportaciones mediante comunicación personal y documentos internos, y por facilitar el acceso a los datos necesarios para desempeñar la investigación.

Declaración de contribución de autoría

Ferran Aragó Puigventós: Conceptualización, Curación de datos, Análisis formal, Investigación, Metodología, Administración de proyecto, Redacción – borrador original, Redacción – revisión y edición.

Pilar Cid-Leal: Metodología, Redacción – revisión y edición.

María José Recoder Sellarès: Metodología, Redacción – revisión y edición.

Declaración de conflicto de intereses

Los/as autores/as de este artículo declaran no tener conflictos de intereses financieros, profesionales o personales que pudieran haber influido de manera inapropiada en este trabajo.

Declaración de uso de Inteligencia Artificial

No aplica.

BIBLIOGRAFÍA

- 3Cat. (2024a). Estat d'informació no financera 2024: Memòria de sostenibilitat. Disponible en: <https://statics.3cat.cat/multimedia/pdf/5/6/1744382905765.pdf>
- 3Cat. (2024b). 3Cat dona el tret de sortida a la transformació del seu model d'informatius [Nota de prensa]. Disponible en: <https://www.3cat.cat/premsa/3cat-dona-el-tret-de-sortida-a-la-transformacio-del-seu-model-dinformatius/nota-de-premsa/3317387/>
- Auserón, P. (2025). How the Next-Gen AV Archivist is Taking Off: ATRESMEDIA Keeps Evolving [Presentación]. *Media Management Seminar 2025: AI and Human Collaboration: Partners in Archiving?* FIAT/IFTA, Cardiff, Reino Unido. Disponible en: <https://fiatifta.org/seminar/media-management-seminar-2025/>
- Bailac, M. (2023). L'arxiu de TV3 a l'era digital, una mirada al seu valor, preservació i especificitats. *La Dada*. Disponible en: <http://arxivs.com/ladada/larxiu-de-tv3-a-lera-digital-una-mirada-al-seu-valor-preservacio-i-especificitats/>.
- Bazán-Gil, V. (2023). Artificial intelligence applications in media archives. *Profesional de la información*, 32(5), e320517. DOI: <https://doi.org/10.3145/epi.2023.sep.17>.
- Berners-Lee, T., Hendler, J., y Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43. DOI: <https://doi.org/10.1038/scientificamerican0501-34>.
- Caldera-Serrano, J. (2024). El documentalista audiovisual-periodista como perfil profesional para los servicios de información en televisión. *Revista Estudios de la Información*, 2(2), 5-15. DOI: <https://doi.org/10.54167/rei.v2i2.1552>.

- Caldera-Serrano, J. (2025). Aplicaciones de la inteligencia artificial para automatización de procesos documentales en los archivos audiovisuales televisivos. *Revista Interamericana de Bibliotecología*, 48(1), e356060. DOI: <https://doi.org/10.17533/udea.rib.v48n1e356060>.
- Caswell, D., y Dörr, K. (2017): Automated Journalism 2.0: Event-driven narratives. *Journalism Practice*, 11(6-7), 1-17. DOI: <http://dx.doi.org/10.1080/17512786.2017.1320773>.
- Català, M. (2025). *Estadístiques del 2016-2024 i IA*. [Diapositivas de PowerPoint]. Presentación interna del departamento de Documentación de 3Cat.
- Codina, L., y Lopezosa, C. (2024). Curación de contenidos en periodismo: características generales y uso de buscadores con inteligencia artificial. En F. Murcia-Verdú (ed.), *Periodismo y comunicación digital: retos y oportunidades*, 157-179. Comunicación Social.
- Codina, L. (2023). *Estudios de caso: características, tipología y bibliografía comentada*. Lluís Codina. Disponible en: <https://www.lluiscodina.com/estudios-de-caso/>.
- Coller, X. (2005). *Estudio de casos*. Centro de Investigaciones Sociológicas.
- Conesa, A. (2012). De la videoteca a l'arxiu digital: evolució del Departament de Documentació de Televisió de Catalunya. *Tripodos*, 31, 99-108. Disponible en: http://www.tripodos.com/index.php/Facultat_Comunicacio_Blanquerna/article/view/40.
- Dayan, D. y Katz, E. (1992). *Media events: The live broadcasting of history*. Harvard University Press.
- European Broadcasting Union (EBU) (2025). *Public Service Media Archives: PSM safeguarding a priceless resource (Public version)*. European Broadcasting Union. <https://www.ebu.ch/research/membersonly/report/public-service-media-archives>.
- FIAT/IFTA Media Management Commission. (2025). *Media Management Seminar 2025. AI and Human Collaboration: Partners in Archiving?* [Seminario]. University of South Wales, Cardiff, Gales (R.U.). Disponible en: <https://fiatifta.org/seminar/media-management-seminar-2025/>.
- Grayson, D. (coord.). (2023). *Unlocking the past: A comprehensive guide to managing rights and risks in audiovisual archives*. FIAT/IFTA, Value,
- International Organization for Standardization (ISO) (2021). *ISO 690:2021. Information and documentation — Guidelines for bibliographic references and citations to information resources* (4.^a ed.). ISO. Disponible en: <https://www.iso.org/standard/72642.html>
- Library of Congress. Policy and Standards Division. (2019). *Subject Headings Manual: H 1592 Events*. Library of Congress. Disponible en: <https://www.loc.gov/aba/publications/FreeSHM/H1592.pdf>.
- López-de-Quintana-Sáenz, E. (2014). Rasgos y trayectorias de la documentación audiovisual: logros, retos y quimeras. *El profesional de la información*, 23(1), 5-12. DOI: <https://doi.org/10.3145/epi.2014.ene.01>.
- Marcos, J. C., y Edo, C. (2015). Análisis de la nueva perspectiva de la documentación periodística en los medios de comunicación españoles. *Intracom: Revista de Información y Comunicación*, (14), 31-48 DOI: https://doi.org/10.5209/rev_RGID.2015.v25.n2.51235.
- Marcos-García, S., Alonso-Muñoz, L., y López-Meri, A. (2021). Periodismo y nuevas narrativas. Storytelling como formato de difusión informativa en redes sociales. *Estudios sobre el Mensaje Periodístico*, 27(2), 553-567. DOI: <https://doi.org/10.5209/esmp.71193>
- Menke, M., Kinnebrock, S., Kretschmar, S., Aichberger, I., Broersma, M., Hummel, R., Kirchhoff, S., Prandner, D., Ribeiro, N., y Salaverría, R. (2018). Convergence culture in European newsrooms: Comparing editorial strategies for cross-media news production in six countries. *Journalism Studies*, 19(6), 881-904. DOI: <https://doi.org/10.1080/1461670X.2016.1232175>.
- Shoemaker, P. J., y Vos, T. (2009). *Gatekeeping theory*. Routledge.
- Soler Monreal, C., y 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. *Revista General de Información y Documentación*, 33(1), 231-256. DOI <https://doi.org/10.5209/rgid.84670>.

Tzelepis, C., Ma, Z., Mezaris, V., Ionescu, B., Kompatsiaris, I., Boato, G., Sebe, N., y Yan, S. (2016). Event-based media processing and analysis: A survey of literature. *Image and Vision Computing*, 53, 3-19. DOI: <https://doi.org/10.1016/j.imavis.2016.05.005>

Yin, R. K. (2018). *Case Study: Research and Applications*. SAGE Publications.