# Audio-Visual Semantics: propuesta de una ontología para la descripción de secuencias audiovisuales

Juan-Antonio Pastor-Sánchez[1], Tomás Saorín[2],
Virginia Bazán[3], Manuel Escribano[4] y María-José Baños-Moreno[5]

[1] ORCID 0000-0002-1677-1059. Departamento de Información y Documentación, Universidad de Murcia, España
pastor@um.es
[2] ORCID 0000-0001-9448-0866. Departamento de Información y Documentación, Universidad de Murcia, España
tsp@um.es
[3] ORCID 0000-0003-4920-2212. Radio Televisión Española, España
virginia.bazan@rtve.es
[4] ORCID 0000-0003-2521-7030. VSN Video Stream Networks, España
mescribano@vsn.es
[5] ORCID 0000-0001-9137-1330. ODILO, España
mbm41963@um.es

**Resumen.** El presente trabajo aborda la descripción de los aspectos conceptuales de contenidos audiovisuales mediante la ontología Audio-Visual Semantics (AVS) que permite representar acciones, características e interacciones entre entidades y/o elementos con una granularidad multinivel. Para ello se ha partido de una aproximación en el que las piezas audiovisuales están compuestas por secuencias. En dichas secuencias es posible identificar diferentes sucesos sobre los que pueden elaborarse descripciones. Los sucesos están compuestos por una serie de elementos tales como agentes, acciones, el objeto de dichas acciones u otros sucesos. Se contempla el uso de cualificadores (especificando su alcance y su valor) para definir cualidades o atributos de los diferentes elementos que intervienen en un suceso. Los agentes, acciones, objetos y cualificadores no se definen de forma extensiva en la propia ontología, sino que son referenciados como conceptos de vocabularios SKOS, lo que hace innecesario alterar la ontología AVS y permite definir relaciones semánticas entre conceptos y el etiquetado multilingÜe. La ontología AVS se encuentra en proceso de desarrollo y en la actualidad se está procediendo a su validación mediante la descripción de piezas audiovisuales y la verificación de los resultados obtenidos.

**Palabras clave**: Contenidos Audiovisuales, Ontologías, Descripción semántica, SKOS

Pastor-Sánchez, J.A., Saorín, T., Bazán, V., Escribano, M., Baños-Moreno, M.J. (2020). Audio-Visual Semantics: propuesta de una ontología para la descripción de secuencias audiovisuales. En J. Tramullas, P. Garrido-Picazo y G. Marco-Cuenca (eds.) *Actas del IV Congreso ISKO España y Portugal 2019* (pp. 337-347). https://doi.org/10.5281/zenodo.3738620

**Abstract.** This paper shows the description of the conceptual aspects of audiovisual content through the Audio-Visual Semantics ontology (AVS) that allows to represent actions, characteristics and interactions between entities and / or elements with a multilevel granularity. For this, it is considered that the audiovisual pieces are composed of sequences. In these sequences it is possible to identify different happenings that can be described. Happenings are composed of a series of elements such as agents, actions, targets or other events. The use of qualifiers (specifying their scope and value) is contemplated to define qualities or attributes of the different elements that intervene in an event. The agents, actions, objects and qualifiers are not defined extensively in the ontology itself, but are referenced as concepts of SKOS vocabularies, which avoids modifying the AVS ontology and allows to define semantic relations between concepts and multilingual labeling. Currently, the AVS ontology is in development and in the validation phase through the description of audiovisual pieces and the verification of the results obtained.

**Keywords:** Audiovisual Contents, Ontologies, Semantic Description, SKOS

# 1 Introducción

La tecnología digital actual ha facilitado en gran medida la creación y difusión de material audiviosual en múltiples ámbitos: entretenimiento, noticias, publicidad, educación, investigación, etc. Además, la mayor disponibilidad de estos contenidos en Internet ha aumentado su reutilización y redistribución. La producción y consumo de vídeo y podcast se ha generalizado, sin que sea necesario el uso de servicios o elementos tecnológicos complejos. Cualquier usuario puede grabar, editar y distribuir un documento audiovisual en Internet (YouTube, iVoox, redes sociales) utilizando únicamente un smartphone.

La descripción de escenas es un aspecto básico del trabajo diario de los documentalistas. Periodistas, editores y personal de archivo utilizan esta herramienta para una descripción detallada de todo el material audiovisual para preparar noticias, documentales, describir eventos deportivos, etc. Estas tareas hasta el momento suelen realizarse mediante la descripción de escenas utilizando texto libre. Igualmente es posible acceder a todo tipo de contenidos audiovisuales desde cualquier dispositivo, así como utilizar plataformas comerciales de *streaming* para su consumo.

Todo ello ha tenido un impacto significativo en la gestión y distribución de documentos audiovisuales por parte de los productores y operadores (Evain, Matton y Vaervagen, 2017), así como en los procesos de preservación digital (Plank, 2018; Corrado y Moulaison, 2016; Evens y Hauttekeete, 2011). Igualmente, esta situación ha provocado que los sistemas de recomendación en las plataformas de contenidos audiovisuales en streaming estén cobrando cada día mayor importancia (Hallinan y Striphas, 2016).

La interoperabilidad en los procesos de gestión y acceso de documentos audiovisuales ha sido abordada por otros trabajos (Höffernig y Bailer, 2009; Dimoulas, Veglis y Kalliris, 2015). En este sentido, para acceder a un contenido audiovisual de manera eficiente se precisa un procesamiento previo de su semántica (Höffernig et al., 2011). Esta tarea no es sencilla, debido a la propia la naturaleza de este tipo de documentos. Mientras que para las personas resulta relativamente sencillo describir y categorizar objetos, imágenes o sonidos en función de su significado, para las máquinas es algo ciertamente complejo. De hecho, los contenidos audiovisuales no son fácilmente procesables debido a los diferentes elementos que aparecen y la dinámica entre los mismos.

Son abundantes los trabajos que han abordado el uso de tecnologías semánticas en este ámbito, especialmente en aquellos aspectos relacionados con la edición, producción, distribución, programación y emisión de contenidos audiovisuales integrándose con MPEG-7 que permite una inclusión de aspectos semánticos básicos (Fourati, Jedidi y Gargouri, 2014; Hunter y Nack, 2001; Isaac y Troncy, 2004; Raymond et al., 2010). Por su parte, la European Broadcasting Union (EBU) han elaborado la ontología EBUCore que ofrece un conjunto de elementos de metadatos para la descripción de aspectos técnicos y estructurales[1] así como una serie de vocabularios SKOS que pueden utilizarse para la descripción de recursos[2].

El presente trabajo aborda la descripción de los aspectos conceptuales del contenido que trascienden la mera identificación de entidades o elementos contextuales tales como personas, organizaciones, lugares, etc. Esto permitiría plantear consultas y procesos que vayan más allá de la búsqueda y reutilización de contenidos audiovisuales únicamente a partir de un análisis y explotación de simples descripciones textuales. Para ello se propone la ontología Audio-Visual Semantics (AVS) que permite representar acciones, características e interacciones entre dichas entidades y/o elementos con una granularidad multinivel. Dicha propuesta es el resultado de la primera fase de un proyecto de Investigación y Desarrollo financiado Centro de Desarrollo Tecnológico Industrial, E.P.E., una entidad pública empresarial, dependiente del Ministerio de Ciencia, Innovación y Universidades. En dicho proyecto participan Video Stream Network (VSN), Radio Televisión Española (RTVE) y la Universidad de Murcia

1 Puede consultarse la ontología EBUCore en: https://www.ebu.ch/metadata/ontologies/ebucore/

2 Dichos vocabularios están disponibles en: https://www.ebu.ch/metadata/ontologies/skos/

## 2 Metodología

El objetivo general del proyecto de investigación en el que se enmarca el desarrollo de la ontología AVS es el desarrollo de un software para que el proceso de descripción de contenidos audiovisuales pueda realizarse de forma rápida, en un contexto multilingüe, centrado en la semántica de los contenidos y con un incremento en la precisión de las tareas de búsqueda y recuperación de contenidos audiovisuales.

El sistema debe permitir la definición de declaraciones completas y semánticamente significativas con un alto grado de formalización y no vinculado a un idioma específico. Un requisito indispensable es el uso de estándares abiertos y ampliamente extendidos. La descripción de escenas debe ser flexible e incluso cercana al estructuras sintácticas sencillas y generales. En este sentido, sería posible identificar sujetos, acciones y objetos de dichas acciones con la suficiente expresividad semántica para modelar una amplia variedad de declaraciones.

El objetivo último del proyecto es dotar al Gestor de Contenidos Audiovisuales de VSN de la capacidad para catalogar y buscar contenido multimedia usando descripciones semánticas. Para ello se apoyará en AVS que define con precisión el dominio de las descripciones que se pueden construir para representar cualquier escena.

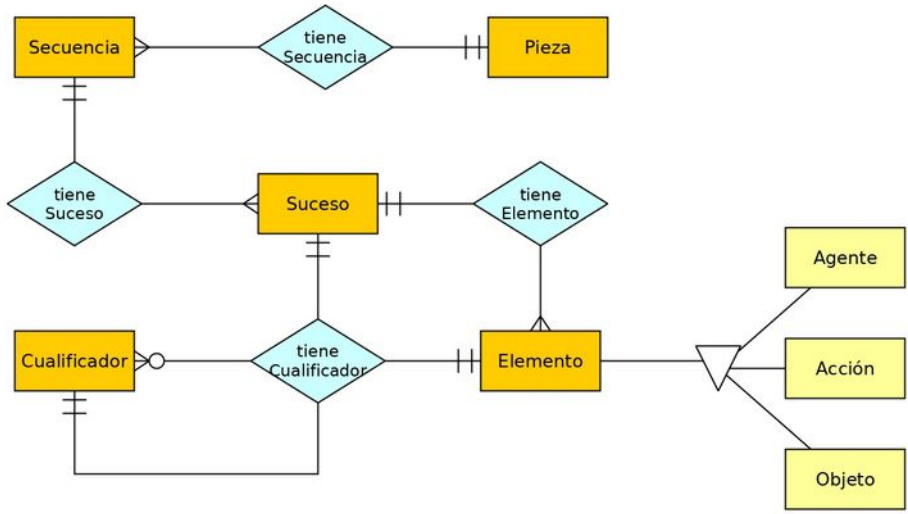


**Figura 1.** Modelo conceptual. Fuente: elaboración propia.

La metodología para la elaboración de la ontología AVS parte de un modelo conceptual en el que las piezas audiovisuales están compuestas por secuencias. En dichas secuencias es posible identificar y describir diferentes sucesos. Los sucesos están compuestos por una serie de elementos que pueden vincularse con 1) agentes que realizan una acción, 2) las acciones realizadas y 3) el objeto u objetos de dichas acciones.

También es posible que dichas vinculaciones sean con otros sucesos, por ejemplo, cuando una persona hace declaraciones sobre un hecho determinado, o cuando un suceso es causa de otro, etc. Para dotar de una mayor flexibilidad al modelo, es posible definir cualificadores sobre los sucesos, elementos y sobre los propios cualificadores. Este mecanismo de cualificación permite adecuar el nivel de detalle de las descripciones a las necesidades específicas de los contenidos o de los sistemas en los que se realizan.

Existe una gran diversidad de propuestas metodológícas para la elaboración de ontologías. Para la construcción de AVS se han seguido los principios metodológicos de Mendonça y Soares (2017) que se resumen en los siguientes pasos realizados:

1. Especificación de los objetivos y requisitos de la ontología.
2. Obtención del conocimiento mediante el análisis de las descripciones de minutados de piezas de contenidos audiovisuales y visualización de las mismas.
3. Conceptualización de la ontología a través de la identificación de los conceptos, relaciones y atributos.
4. Formalización e implementación de la ontología OWL especificando sus clases, propiedades y axiomas.
5. Evaluación y validación de la ontología mediante su aplicación en la descripción semántica de secuencias de piezas audiovisuales específicas.
6. Documentación de la ontología.

Cabe resaltar dos de los requisitos que se tuvieron en cuenta para el desarrollo de la ontología. El primero de ellos se relaciona con la necesidad de que la ontología debía ser de fácil uso en un entorno de uso a través de una herramienta de descripción minutada de los documentos audiovisuales. Es decir, el operador o documentalista debería poder utilizar esta ontología de un modo totalmente transparente y a través de una interfaz lo más sencilla posible. Lo anterior implicaba ofrecer una estructura de clases y propiedades lo más sencilla posible.

El segundo requisito tiene en cuenta los aspectos relacionados con el mantenimiento de la ontología. Considerando lo anterior se debía tener en cuenta que la identificación de nuevas entidades (personas, organizaciones, lugares) acciones, características, etc, podría darse durante el proceso de descripción. Por lo tanto, la incorporación de las mismas no debía modificar la ontología, sino que debía realizarse en una serie de vocabularios controlados. La gestión de los mismos se realizaría de forma dinámica, añadiendo nuevos elementos durante el proceso de descripción y con mecanismos de autocompletado para la búsqueda y selección de los mismos.

## 3 Resultados

AVS[3] se basa en la definición de un conjunto relativamente reducido de elementos. Un punto a tener en cuenta en la definición de clases es la jerarquía "Clip → Secuencia → Suceso", que establece la dinámica de descripción de un contenido audiovisual. Según este principio las piezas audiovisuales (avs:Clip) se dividen en secuencias o escenas (avs:Sequence) en las que se identifican eventos o sucesos que tienen lugar durante las mismas (avs:Happening). La ontología también incluye una serie de propiedades para la definición de intervalos que permitan localizar temporalmente las secuencias y los sucesos dentro de la piezas audiovisuales.

Por su parte, los sucesos se conciben como la combinación de elementos (avs:Element) que pueden ser agentes, acciones y objetos de las acciones. En vez de crear una jerarquía de clases a partir de los elementos, se han utilizado una serie de propiedades (avs:hasAgent, avs:hasAction y avs:hasTarget) que definen la naturaleza del vínculo entre un elemento y un concepto de un vocabulario SKOS, determinando de esta forma el tipo de elemento que interviene en el suceso.

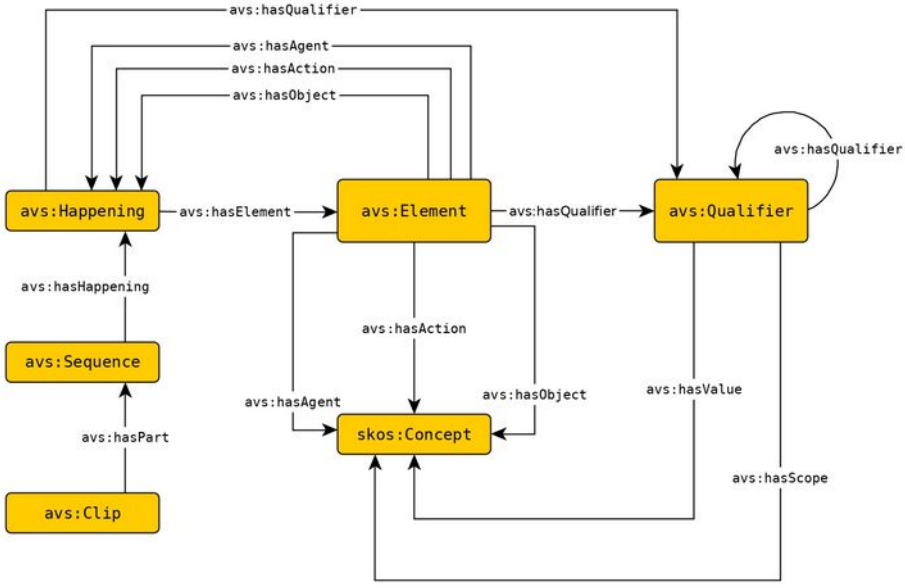


**Figura 2.** Elementos de la ontología AVS. Fuente: elaboración propia.

El uso de cualificadores (avs:Qualifier) permite definir cualidades o atributos de los diferentes elementos que intervienen en un suceso. Para ello los cualificadores tienen dos componentes que se definen mediante sus correspondientes propiedades:

3 La ontología estará disponible en http://purl.org/umu/avs y actualmente se encuentra en la fase de validación final, tras la cual será documentada.

- El alcance del cualificador (avs:hasScope) que permite identificar la característica que se está cualificando (lugar del suceso, vestimenta de una persona, estado de ánimo, situaciones meteorológicas, momento del suceso, etc).
- El valor del cualificador (avs:hasValue) que permite seleccionar un valor concreto para el alcance del cualificador (amarillo, vestido, enfadado, tormenta eléctrica, amanecer, etc).

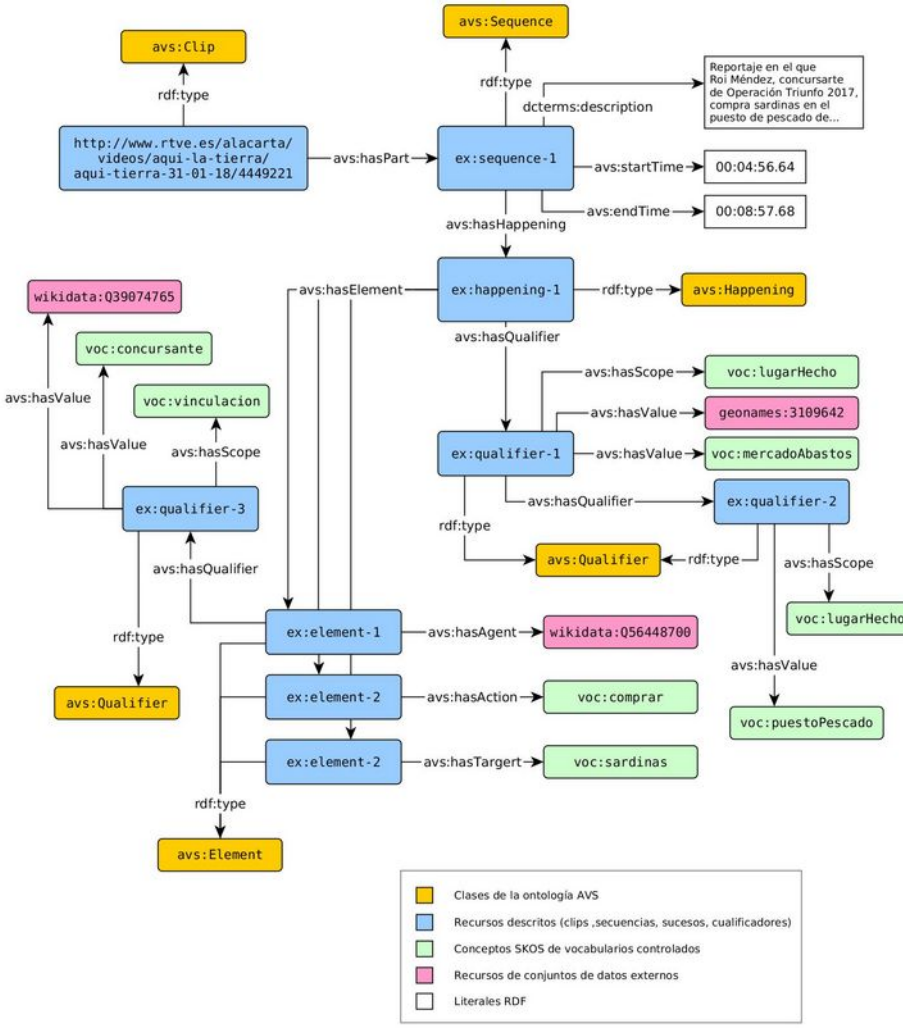


**Figura 3.** Elementos de la ontología AVS. Fuente: elaboración propia.

El ejemplo de la figura 3 delimita una secuencia dentro de un clip. Al mismo tiempo se identifica un suceso específico, que se cualifica mediante el recurso avs:qualifier-1 para indicar la ubicación en la que se desarrolla el suceso mediante la intersección del concepto "Mercado de abastos" con el elemento de Geonames correspondiente a Santiago de Compostela. Dicho cualificador es a su vez cualificado para establecer dicha ubicación con mayor detalle. El suceso tiene asociados tres elementos: agente que realiza la acción (el elemento de wikidata correspondiente a Roi Méndez), acción realizada (comprar) y el objeto de la acción (sardinas). El elemento correspondiente al agente se cualifica para definir su vinculación como concursante de Operación Triunfo 2017 (elemento de Wikidata). Este ejemplo también muestra como es posible reutilizar conjuntos de datos externos en la descripción de los sucesos de una secuencia (en este caso Wikidata y Geonames).

El uso de SKOS es un aspecto relevante de AVS. Los agentes, acciones, objetos y cualificadores no se definen de forma extensiva en la propia ontología, sino que son referenciados como conceptos de vocabularios SKOS. Los vocabularios pueden organizarse en torno a esquemas de conceptos o colecciones SKOS para representar grupos de conceptos para lugares, acciones, alimentos, organizaciones, personas, programas de televisión, roles, etc. Incluso, la cualificación permitiría definir cuando el elemento objeto de un suceso es receptor directo o indirecto de la acción. De este modo se consigue una mayor flexibilidad durante la aplicación de la ontología en la representación de diferentes atributos, autoridades, lugares, etc. La definición de nuevos valores únicamente requerirán la creación de un nuevos conceptos SKOS. De este modo, se evita alterar la ontología AVS y se reutilizan las diversas relaciones que ofrece SKOS para la organización jerárquica y asociativa de conceptos, así como las propiedades para la desambiguación terminológica y el etiquetado multilingüe. AVS también permite reutilizar recursos de conjuntos de datos externos como Wikidata, Geonames, VIAF, etc.

## 4 Conclusiones y líneas de trabajo futuras

La ontología AVS tiene una estructura sencilla y adecuada para la descripción rápida de la semántica de los aspectos conceptuales de secuencias de documentos audiovisuales. En la actualidad se está procediendo a la evaluación y validación de la ontología mediante su aplicación en la descripción de piezas audiovisuales facilitadas por RTVE y la verificación de los resultados obtenidos. También se están realizando consultas SPARQL sobre los conjuntos de datos generados para evaluar la viabilidad de la ontología en los procesos de búsqueda. Por el momento no se disponen de datos suficientes para asegurar si la metodología utilizada es válida tanto para un entorno de contenidos audiovisuales de carácter general (ficción, informativos, documentales) o si es más adecuada para colecciones más homogéneas desde el punto de vista de su contenido.

En la actualidad, la recuperación de objetos audiovisuales digitales se realiza principalmente a partir de descripciones textuales de los mismos y la aplicación de técnicas de recuperación de información donde la precisión y la exhaustividad son afectados por los clásicos problemas relacionados con el lenguaje natural: sinonimia, polisemia, multilingüismo, ambigüedad semántica, etc. La formalización de la descripción del contenido mediante ontologías es una solución de gran interés para evitar dichos problemas. En este sentido, AVS aporta un alto grado de formalización de dichas descripciones que además pueden aplicar técnicas de inferencia y descubrimiento de datos para mejorar los procesos de búsqueda de contenidos.

Por otro lado, la aplicación de vocabularios SKOS supone una solución válida para incorporar nuevos elementos de descripción sin que sea preciso definir una compleja estructura de clases y subclases sujeta a las nuevas necesidades identificadas durante el proceso de descripción. Por su parte, el uso adecuado de etiquetas preferentes y alternativas permite agilizar los procesos de descripción y recuperación y aporta una mayor coherencia terminológica. También es posible utilizar relaciones jerárquicas y asociativas para organizar adecuadamente los elementos de los vocabularios. Otra ventaja que aporta el uso de SKOS es la posibilidad de compartir y reutilizar los vocabularios entre diferentes operadores y de integrar su aplicación en herramientas de búsqueda.

El paso más inmediato será analizar la operatividad de la ontología en una herramienta de descripción de documentos audiovisuales. La interacción entre datos, interfaz y usuario debe realizarse del modo más ágil posible. Dentro de este punto también se contemplará la generación y gestión de los vocabularios SKOS durante el proceso de descripción. La búsqueda y recuperación secuencias concretas también deberá ser diseñada cuidadosamente para que el planteamiento y ejecución de consultas resulte una tarea sencilla.

Son interesantes las propuestas de trabajos que apuntan la integración de las tecnologías semánticas en los sistemas de recomendación (Sotelo, Juayek y Scuoteguazza, 2013; Sotelo y Juayek, 2015). Algunas posibles líneas de trabajo podrían ser la integración con dichas propuestas, así como con otras para la representación de narrativas transmedia (Pastor y Saorín, 2018) y la reutilización/mapeado de otros vocabularios y ontologías (p.e. Dublin Core, EBUCore) para la anotación o enriquecimiento de las descripciones.

## Referencias

Corrado, E.M. & Moulaison Sandy, H.L. (2016). *Archiving Conference, Archiving 2016 Final Program and Proceedings*, pp. 161-166. https://doi.org/10.2352/issn.2168-3204.2016.1.0.161

Dimoulas, C., Veglis, A., & Kalliris, G. (2015). *Audiovisual hypermedia in the semantic Web. En: Encyclopedia of Information Science and Technology* (3ª edición), pp. 7594-7604. IGI Global.

Evain, J.P., Matton M. & Vaervagen, T. (2017) Wikipedia and DBpedia for Media - Managing Audiovisual Resources in Their Semantic Context. En: van Erp M. et al. (eds) *Knowledge Graphs and Language Technology ISWC 2016*, pp. 41-56. Lecture Notes in Computer Science, vol 10579. Springer, Cham. https://doi.org/10.1007/978-3-319-68723-0_4

Evens, T., & Hauttekeete, L. (2011). Challenges of digital preservation for cultural heritage institutions. *Journal of Librarianship and Information Science, 4*3(3), 157–165. https://doi.org/10.1177/0961000611410585

Fourati, M., Jedidi, A. & Gargouri, F. (2014), Towards a Semantic Multi-modalities Description of Audiovisual Documents. En *Proceeding ISM '14 Proceedings of the 2014 IEEE International Symposium on Multimedia*, 259-262. https://doi.org/10.1109/ISM.2014.28

Hallinan, B., & Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society, 18*(1), 117–137. https://doi.org/10.1177/1461444814538646

Höffernig, M., & Bailer, W. (2009). Formal metadata semantics for interoperability in the audiovisual media production process. En *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe 2009)*. http://ceur-ws.org/Vol-539/paper_6.pdf

Höffernig M., Bailer W., Nagler G. & Mülner H. (2011) Mapping Audiovisual Metadata Formats Using Formal Semantics. In Declerck et al. (eds) *Semantic Multimedia. SAMT 2010*. Lecture Notes in Computer Science, vol 6725. Springer, Berlin, Heidelberg. https://pdfs.semanticscholar.org/5c30/8ceaa43bdd9c061522d9245ae0e510ddeea5.pdf

Hunter, J. & Nack, F. (2001). An overview of the MPEG-7 description definition language (DDL). *IEEE Trans. Circuits Syst. Video Techn*, 11, 765-772. https://pdfs.semanticscholar.org/635a/6745f57ba892f28d854c84f8f54c39fa746f.pdf

Isaac, A.; Troncy, R. (2004). Designing and Using an Audio-Visual Description Core Ontology. *Workshop on Core Ontologies in Ontology Engineering at EKAW'0*4, Whittlebury. https://pdfs.semanticscholar.org/bca6/00fb958d5b97709fd383c14f25b14ea2ddd9.pdf

Mendonça, F.M., & Soares, A.L. (2017). Construindo ontologias com a metodologia ontoforinfoscience: uma abordagem detalhada das atividades do desenvolvimento ontológico. *Ciência da Informação, 46*(1), p. 43-59. http://revista.ibict.br/ciinf/article/view/4013/3713

Pastor-Sánchez, J.A., Saorín, Tomás (2018). A Conceptual model for an OWL ontology to represent the knowledge of transmedia storytelling. En *Proceedings of the Fifteenth International ISKO Conference*, pp. 511-520. http://hdl.handle.net/10760/33908

Plank, M. (2018). Managing Born-Digital Audiovisual Media. *International Association of Sound and Audiovisual Archives (IASA) Journal*, (47), 61–67. http://journal.iasa-web.org/pubs/article/view/56

Raimond Y., Scott T., Oliver S., Sinclair P. & Smethurst M. (2010) Use of Semantic Web technologies on the BBC Web Sites. En: Wood D. (eds) *Linking Enterprise Data*, pp. 263-283. Springer, Boston, MA.

Sotelo, R., Juayek, M. & A. Scuoteguazza (2013). A comparison of audiovisual content recommender systems performance: Collaborative vs. semantic approaches. En *2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1-5. DOI: 10.1109/BMSB.2013.6621791

Sotelo, R. & Juayek, M. (2015). Incidence of specific semantic characteristics on the performance of recommender systems of audiovisual content. En *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1-4. DOI: 10.1109/BMSB.2015.7177277