Media in the Cloud:

Ontology and Semantic Web Technology Navigation Guide

November 2022

Contents

Summary

This is a primer on the use of ontologies — which provide the framework for the data held in knowledge graphs — and other semantic web technologies within the media domain.

The current media landscape is characterized by movement of workflows into the cloud. This has brought with it expectations around automation, agility, and scalability; and challenges in areas such as interoperability, portability, discovery, and orchestration. It is an ever more data-driven eco-system with an increased need for consistent, interoperable, and actionable metadata and semantics to drive and manage distributed and cloud-based processes and workflows.

Fortunately, semantic technologies have great potential to address some of these issues, supporting the combination of standards-based machine-readability with the representation of human subject matter expertise. An array of mechanisms and specifications have emerged that make the utilization of semantic technologies in media workflows both more practical and more desirable, including microservices and application programming interfaces; artificial intelligence and machine learning; and media-specific ontology specifications from the European Broadcasting Union, MovieLabs, and the dwerft Linked Metadata for Media project.

What is an Ontology?

An ontology is a formal model that represents a given "knowledge domain," meaning the entities that are meaningful within that space and the relationships between them, using a set of vendor agnostic specifications developed by the World Wide Web Consortium (W3C).

The core of semantic technology is the "triple," a subject-predicate-object assertion such as "Johanna-knows-Paula." Each component of the triple is regarded as a resource and is uniquely identifiable. Extremely complex models can be built out using this simple, flexible basic structure. They can be stored in databases known as triple-stores and can be combined into graphs that provide extremely flexible and extensible representation of the meaningful objects within an enterprise or a broader field.

Business Uses of Semantic Web Technology:

Semantic web technology is enabling, and when smartly stacked with other tools it can support the following areas:

Knowledge Management: data integration through the connection of disparate data sets; data contextualization, because data is situated within a frame of reference; data governance that utilizes shared vocabularies; and knowledge extraction, inference and artificial intelligence that utilizes semantic vocabularies with axiomatic statements and logical constraints.

Application and Service Integration: semantic web technologies can significantly reduce the difficulty of managing data interchange between applications, providing system inputs and annotations in machine-readable language that supports automatic service discovery and composition.

Asset and Content Management: vocabularies can be integrated with machine learning for automatic metadata extraction from transcripts or video or audio analysis for accurate content description. They can also support automation of content curation and management workflows, particularly in decentralized environments.

Search and Discovery: improved discovery of smartly tagged and contextualized content, and improved recommendations based on semantic relations to other content, such as a shared topic or actor.

Introduction

This is a primer on the use of ontologies and other semantic web technologies within a media domain characterized by the movement of workflows into the cloud.

Tim Berners-Lee *et al* proposed the semantic web in 2001,[1] meaning moving from the mere presentation of content on the web to actionable human and machine-readable data. Movement towards such a web has been slow, but may now be regarded as accelerating as an array of mechanisms and specifications have emerged that make the utilization of semantic web technologies in media workflows both more practical and more desirable. These include the cloud itself, which is not a requirement for the use of semantic web technologies but is certainly a catalyst; microservices and application programming interfaces (APIs); artificial intelligence (AI) and machine learning (ML); and media-specific ontology specifications.

Problem Statement

The movement into the cloud comes with attendant expectations around automation, agility, and scalability and challenges in areas such as interoperability, portability, discovery, and orchestration. It is an ever more data-driven eco-system with an increased need for consistent, interoperable metadata and semantics to drive and manage distributed processes and workflows.

Fortunately, semantic web technologies — which may be regarded as internet-native — have great potential to address some of these challenges, supporting the combination of standards-based machine-readability with the representation of human subject matter expertise. Even more fortunately, there are media-specific ontologies already in existence.

Unfortunately, there is a lack of awareness about what semantic technologies are, what ontologies are, where they are applicable, and how to deploy them.

What is an Ontology?

For the purposes of this paper, which does not delve into philosophy, an ontology is a formal model that represents a given "knowledge domain" — meaning the entities that are meaningful within that space and the relationships between them — using a set of specifications developed by the World Wide Web Consortium (W3C): primarily RDF; RDFS; OWL; SKOS and SPARQL. In practical terms an ontology represents explicit business knowledge and provides the scaffolding for the core data infrastructure of an enterprise or a broader field.

One can imagine that building out such a model would be challenging and that would be correct. However, it is unnecessary to start the modelling process from scratch because others have kindly developed models that can be shared and re-used, and it is relatively simple to extend them and evolve where necessary. This shortens the time required to create a model and supports standardization both within organizations and across domains.

The utilization of RDF and related specifications provides a mechanism for information exchange between applications without a loss of meaning and for creating linked data, meaning data that is interlinked with, and enriched by, data from heterogenous and distributed sources.

---

[1] Berners-Lee, Tim; James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284(5), May 2001.

Current Media Ontology Landscape

The following ontologies and related resources have been published by bodies with a deep understanding of media workflow requirements — namely the European Broadcasting Union (EBU), MovieLabs, and the dwerft Linked Metadata for Media project — and are all available in some form of accessible or downloadable form. (For more information on each ontology see Appendix 1.)

- **Digital Media Ontology** (DMO), Version 1.0, 2020: the DMO models pre-distribution processes and provides mappings to other ontologies and metadata schemas. It also proposes a technical stack — the Linked Media Data Cloud (LMDC) — as a mechanism for how data could be stored and shared between applications and processes.
- **EBU Class Conceptual Model** (EBU CCDM / EBU Tech 3351), Version 2.2, 2020: provides a high-level framework of basic classes and properties to describe the core business objects of the end-to-end value chain of media enterprises. Intended to be used with EBUCore.
- **EBUCore** (EBU Core Metadata Set / Tech 3293), Version 1.10, 2020: available in both XML and RDF, and intended to provide more detailed descriptive and metadata properties to be used with CCDM.
- **EBU Reference Data Sets:** vocabularies, for example “media types” or “editorial control codes,” presented in SKOS format for use with EBUCore and CCDM classes.
- **MovieLabs Ontology for Media Creation**, (OMC) 2021: an ontology intended to support the production process, providing a conceptual framework and a set of defined terms for use in software-defined workflows to reduce ambiguity and support data interoperability.

Also worth noting is the W3C Ontology for Media Resources**,** released in 2012 and intended to provide a core vocabulary and set of descriptive properties for media resources; and the BBC Ontologies, which are not currently publicly maintained but included various models (programs, creative works, sport, wildlife, etc.) released over time and originally intended to support publishing content on the BBC websites.[2]

Related Metadata Schemas

The difference between a metadata schema and an ontology is not absolute and to some extent is dictated by the technology used to publish or manifest a specification and/or the date of publication. Clearly defined data models with defined vocabularies and relationships between entities will be worthwhile whether they are manifested in RDF or other technologies, and several specifications are published or have been implemented in various formats over time. [3]

---

[2] BBC Ontologies may be accessed by entering https://www.bbc.co.uk/ontologies into the Wayback Machine of the Internet Archive.

[3] For those interested in the development of broadcasting metadata standards over time TV-Anytime (now ETSI TS 102 822-3-1, Version 1.11.2, 2019-06) is also worth exploring. It is a metadata specification that first emerged in 2000 from a broader effort looking at the requirements of on-demand consumption, including content referencing

Which standard, or combination of standards, is relevant or useful in any particular scenario will depend on the unique requirements of the given use case. Generally, metadata schemas are more precise and less flexible than ontologies, but several of the ontology standards listed above build upon, incorporate, or provide mappings to the metadata schemas listed below, and others not discussed here.

- Dublin Core: a set of fifteen broad metadata elements for use in digital resource description first released in 1995 and since extended and represented as RDF properties. A seminal standard that has formed the basis of multiple subsequent specifications.
- Entertainment ID Registry (EIDR): known for providing unique identifiers for entertainment content, EIDR also provides a related metadata schema and element vocabularies, including links to other data sources for the content being identified.
- International Standard Audiovisual Number (ISAN): a numbering system and metadata schema for audiovisual works and versions designed by the International Standards Organization (ISO).
- MovieLabs Common Metadata (TR-META-CM, Version 2.10, 2021): part of a broader Digital Distribution Framework[4] providing metadata used in the business-to-business transfer of media resources, and references multiple published technical standards.
- PBCore: a cataloging standard that first emerged from the U.S. public broadcasting domain in 2001 and now provides an element set along with vocabularies, mappings and extensive user tools.
- Schema.org: a consortium that provides multiple metadata schemas potentially relevant in the broadcasting and audiovisual space.

Primary Business Uses of Semantic Technology

Semantic web technologies are enabling: when incorporated into an eco-system they can support a wide range of functions that would otherwise be prohibitively burdensome, or achieved only incompletely, or and perhaps also achieved multiple times (slightly differently each time). Below are some of the primary areas where semantic technologies could be useful when combined with other tools and human subject matter expertise.

Knowledge Management

- Data Integration: this remains the core business use case for semantic technology, which can connect disparate data sets with greater agility than other approaches and be especially useful for activities that span a diverse range of business contexts, as with content moving through the production and supply chains.
- Data Association: semantic technology is inherently good at linking data and can facilitate actions such as pulling in relevant data from external sources — for instance from DBPedia or Wikidata, both associated with Wikipedia[5] — or publishing data for either public or private consumption.

---

and rights management. The EBU CCDM distribution section was originally designed to map to TV-Anytime as well as the BBC Programmes Ontology.

[4] MovieLabs Digital Distribution Framework (MDDF) represents a suite of compatible standards and specifications for the automation of digital workflows and supply chain efficiency.

[5] See www.dbpedia.org and www.wikidata.org for more information on the two projects.

- Data quality and governance: by providing shared vocabularies, semantic technology can improve the consistency and exchangeability of data, and in so doing diminish data silos and redundancy.
- Contextualization: because terms and concepts within ontologies are not orphaned, but always presented in relation to other terms, associating content with any term will also locate it within a broader knowledge structure.
- Knowledge extraction or inference: because ontologies rely on axiomatic statements and logical constraints it is possible to use them to infer certain information. The classic example of this is the inference that an instance belongs to one class based on its membership of another: for instance, because it has been stated that all humans are mammals, if it is further stated that person A is a human, it can be concluded that Person A is also a mammal. That is, semantic technology may be used to make implicit knowledge explicit.
- Artificial Intelligence (AI): vocabularies and knowledge graphs can create virtuous circles where they improve the quality of data provided to artificial intelligence and machine learning (ML), which conversely can enrich and improve vocabularies. Semantic technology can also help reduce “black box” issues where it is hard to understand what an algorithm is doing and generally support a “human-in-the-loop” approach to cognitive computing and AI-assisted decision-making.

Application and Service Integration

- Data integration is not merely an end in itself: shared vocabularies used across multiple applications supports data exchange between applications and services.
- Semantic technology can significantly reduce the effort associated with managing data interchange between applications. It can provide or annotate system inputs and outputs and characterize functionality in machine-readable language, supporting automatic service discovery and composition and potentially significantly reducing the cost associated with system evolution and maintenance.
- This in turn supports workflows, or the interaction of decoupled systems, and greater automation of processes that may have previously required human mediation — supporting a move to human intervention being required only for exception handling.

Asset and Content Management.

- Vocabularies have long been used in the tagging and description of content and may now be combined with machine learning applications for automatic metadata extraction from transcripts and video and audio analysis.
- Well-described content is easier to manage, and semantic vocabularies and ontologies support content curation and management workflows, especially in non-centralized eco-systems, which might face greater issues around silos and fragmentation.

Search and Discovery

- Content recommendations based on semantic relationships to other content, such as a shared topic, actor, or producer.
- Improved search and discovery of smartly tagged and contextualized content — providing a mechanism to support the navigation of the overwhelming amounts of data that most enterprises are encountering and storing.

Technical Specifications

The primary W3C semantic technology specifications for vocabulary building are RDF (Resource Description Framework), RDFS (RDF Schema), OWL (Web Ontology Language), and SKOS (Simple Knowledge Organization System), which may be used along with the query language SPARQL (SPARQL Protocol and RDF Query Language). Which of these is the most appropriate in any given case depends on the complexity and rigor of the intended application and the available expertise and tools, and they are mutually compatible and may be used in combination. Moreover, the specifications are vendor agnostic and offer technology independence. (This paper does not discuss property graphs, a form of graph that does not utilize the W3C semantic specifications and is less standardized and interoperable. However, there have been discussions of the two forms of graph merging, as well as discussions of extensions to RDF to improve its functionality.[6] See the glossary section for more information on property graphs and additions to RDF.)

Semantic technology utilizes a simple yet powerful mechanism for representing data, that of subject-predicate-object statements called triples, which is the basis of the RDF data model:

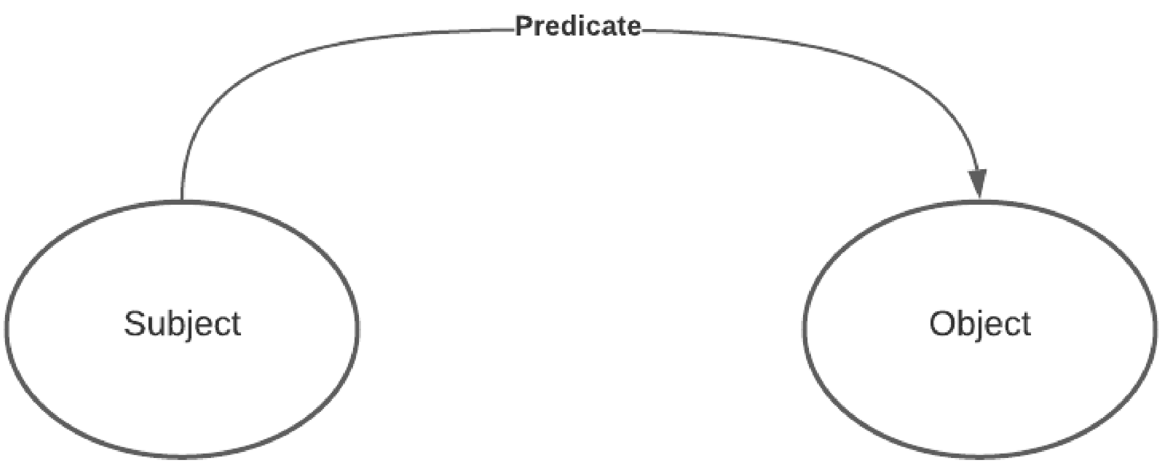

Figure 1: An RDF Triple.

[6] For instance, see the *W3C Workshop on Web Standardization for Graph Data*, "Creating Bridges: RDF, Property Graphs and SQL," 6-9 March 2019, and "RDF-star and SPARQL-star Draft Community Group Report, 21 January 2022."

This could represent a real-world relationship as below:

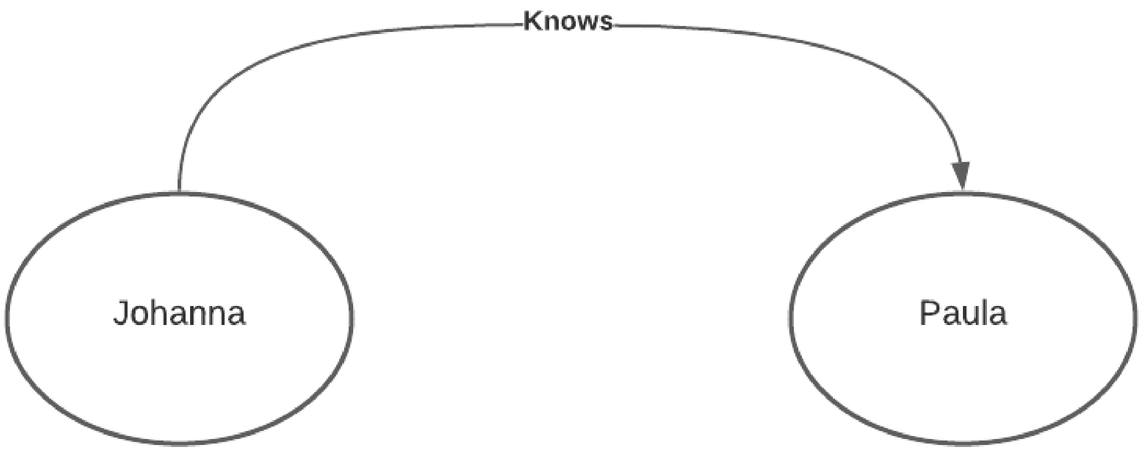

Figure 2: RDF Triple with Values.

Each component of the triple is regarded as a resource, and each is associated with a unique identifier (a uniform resource identifier or URI, of which a uniform resource locator or URL is an example).

Extremely complex models can be built out using this simple, flexible basic structure. They can be stored in databases known as triple-stores and can be combined into graphs. The very granular nature of semantic technology makes it simple to add new relations, classes, and instances as new triples, meaning a schema can constantly evolve in response to new requirements. This agility is a strength of semantic graphs compared to relational or property graph data models. Semantic graphs are neither schema-less, which can be useful for prototyping, but less helpful in the long-term, nor do they require a schema to be completely defined upfront, which is then cumbersome to change.

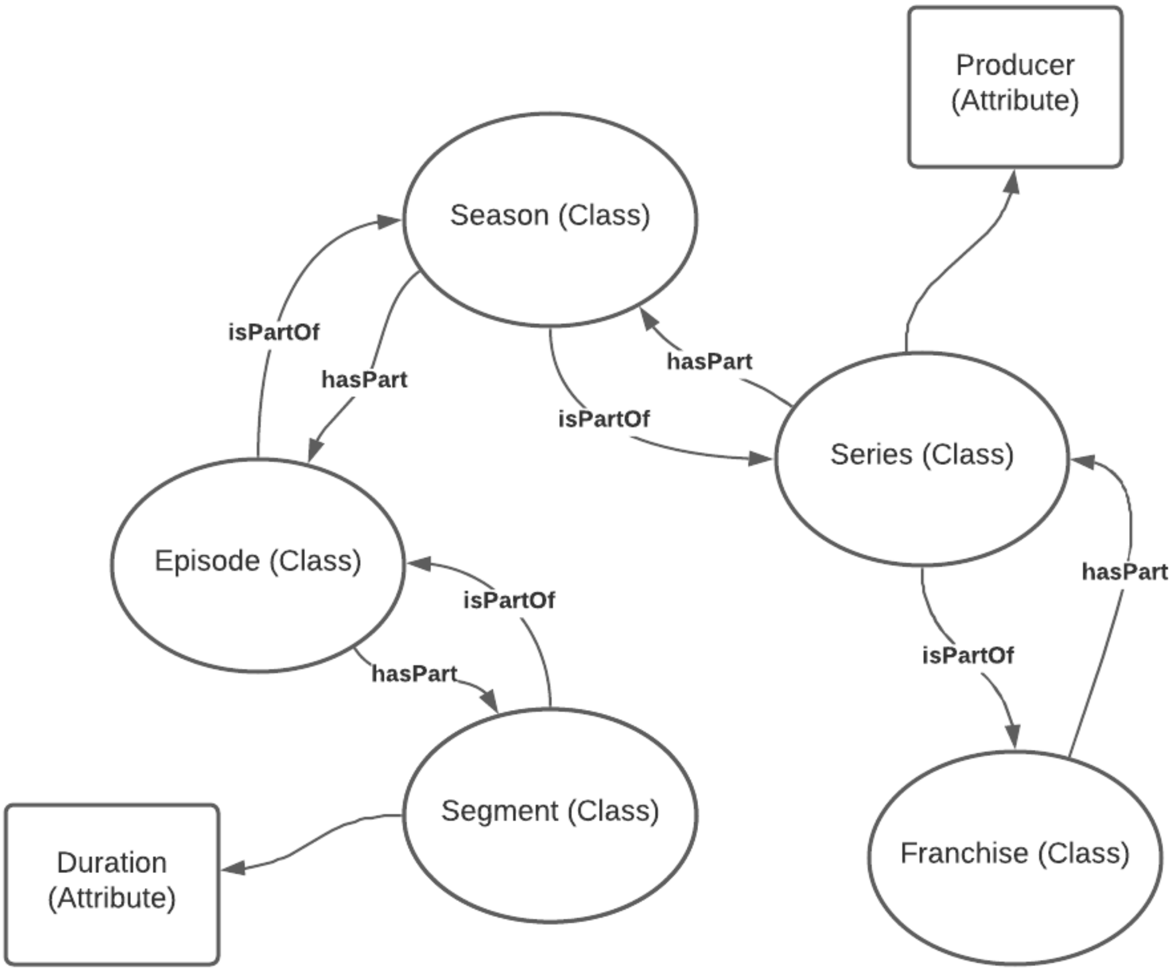

Figure 3: Simple Graph with Classes, Relationships, and Attributes.

RDFS allows resources to be defined with classes, properties (attributes or relationships defined by predicates), and constraints upon their use, allowing the construction of a simple ontology. RDF and RDFS together provide the base syntax of the semantic web, while the OWL specification layers on the capability for ontologies with greater expressivity and more sophisticated logical constraints, and SKOS is intended for the construction of hierarchical taxonomies, including capturing alternative and primary terms for given concepts, which may be used to provide the instances of classes.

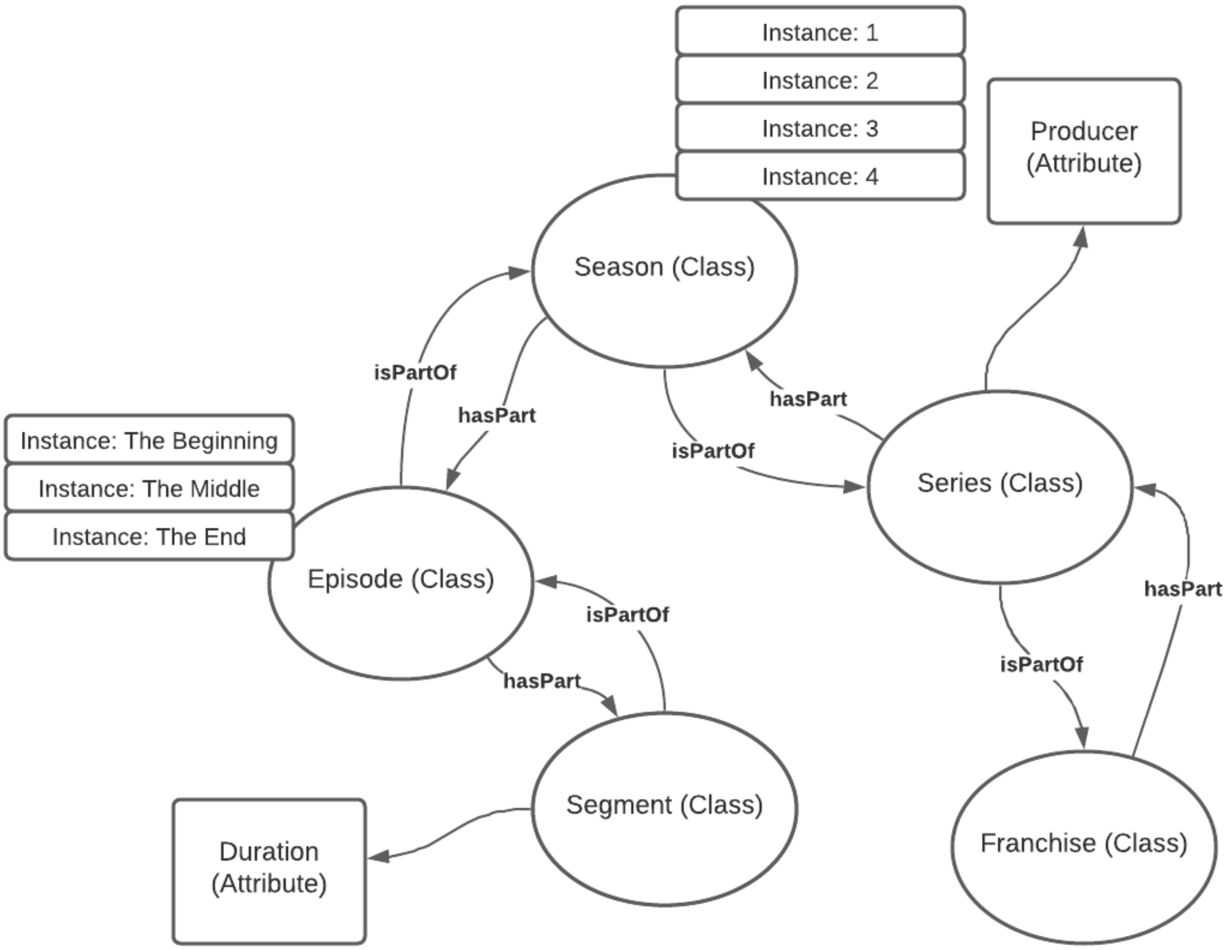

Figure 4: Simple Graph with Classes, Relationships, Attributes, and Instances.

Using these tools (in combination with any others that are appropriate) it is possible to model out given domains in a meaningful way, for instance one may define a class of "Series" or "Season" in RDFS and apply that to a taxonomy created in SKOS, which lists the names of all series broadcast on a given network, along with the name or number of any seasons. The model could be extended to include associated cast and crew, with information on each, and extended again to include the characters featured in each series, or any other relevant information, and relate it back to the series.

Mapping and Merging

Mapping — the matching of an element in one information system more or less exactly to an element in another, or the disambiguation of two elements that superficially seem similar — is a crucial and often painful component of data integration exercises and interoperability. We are again fortunate that extensive mapping between ontologies and metadata schemas has already been done by others, and that semantic technology has in-built mechanisms supporting matching. For example, SKOS specifies several properties for mapping between different concept schemes: broadMatch; narrowMatch are used to state a hierarchical mapping; relatedMatch is used to indicate a more general or looser association; and closeMatch and exactMatch may be used to indicate a higher degree of similarity between concepts. OWL provides for establishing that concepts are either equivalent — that is having the same data or set of instances — by the use of equivalentClass and equivalentProperty, or equal, using sameAs to allow the merging of classes. (For a simple example of a new SKOS mapping exercise, see Appendix 2.)

Semantic graphs are also inherently composable, meaning that if two resources have the same URI they will be automatically merged, supporting the aggregation of graphs from different sources and the disambiguation of resources.

Existing Mappings

There has been a great deal of cross fertilization between the metadata standards and ontologies listed here, and others not discussed in detail. This has included multiple mapping exercises that may be utilized in data integration, though admittedly specific mappings are sometimes hard to locate.

- PBCore provides multiple mappings to other standards, including EBUCore, the preservation metadata standard PREMIS; Dublin Core; and IPTC (International Press Telecommunications Council) concepts.[7]
- The EBU CCDM Distribution Domain was originally designed to seek maximum mapping to the then current BBC Programmes ontology and TV-Anytime.
- EBUCore RDF has mappings to schema-org, Dublin Core and the widely used Friend of a Friend (FOAF) ontology.
- MovieLabs Creative Works Ontology has mappings to the EIDR[8] and the Internet Movie Database (IMDB) data models.
- The Digital Media Ontology has mappings to EBUCore, FOAF, Dublin Core, DBPedia and other specifications.
- The W3C Ontology for Media Resources has extensive mappings to metadata formats current when it was published, including CableLabs 1.1; EBUCore, Dublin Core, and TV-Anytime.

Case Studies

Semantic technology is still an emerging area in media and entertainment, but there are already several compelling production and proof of concept implementations, and other industries have also successfully utilized semantic technology. Below are examples of the use of semantic and graph technology in various production and broadcasting scenarios:

- **Semantic Content Annotation:** The Finnish broadcasting company Yle employs multiple data sources and vocabularies, including linked open data from Wikidata and the broad-based KOKO Ontology (the Finnish Collaborative Holistic Ontology) from Finto, the Finnish Thesaurus and Ontology Service, in semi-automatic tagging of articles, images and programs. Their work provides extensive insights on both the practical aspects and the potential of managing internal and external vocabulary sources and utilizing them for machine learning, personalization, and

[7] See https://pbcore.org/mappings for more details on the mapping and a link to a simple mapping from PBCore to existing RDF predicates.
[8] EIDR itself provides metadata mappings to several other standards. See https://www.eidr.org/standards-and-interoperability/

recommendations. For instance, utilizing existing ontologies makes creating new concepts less onerous, as they are, to some extent, crowdsourced, and it is possible to enrich tags as they flow from system to system.[9]

- **Semantic Digital Asset Management:** France TV (FTV) utilized semantic technology to support a move to a more data-centric eco-system, implementing a knowledge graph-oriented digital asset management system over a data link. This allows mapping of heterogeneous data to a purpose-built ontology (that draws on existing standards), providing data interoperability, access, and contextualization, and in turn supporting artificial intelligence and microservice-based workflows, such as thumbnail generation and video segmentation, with the overall effect of shortening production cycles and supporting content use and re-use.[10]

- **Enterprise-Wide Linking of Data from Arbitrary Systems at Any Time During Processing**: Data siloes along a value chain are a barrier to process-oriented integration, as the information within data objects often cannot be reused simply because they are not connected to a broader structure. German broadcaster SWR's "metagraph" project is intended to create an enterprise-wide, cross-system network of data. Within this knowledge graph, data objects are registered using their existing references and then linked in a common graph, allowing the derivation of many additional links between objects. For a system-agnostic view of the information, this graph is mapped onto a canonical model, based on CCDM, and canonical objects are assigned with desired identifiers for enterprise-wide linking, allowing the referencing of canonical concepts — such as production jobs, editorial objects, essences, publication events — in arbitrary systems.[11]

- **Preservation and Management of Data through the Production Lifecycle:** The Entertainment Technology Center (ETC) at the University of Southern California proposed a mechanism for the persistence of data through production using the "C4 framework"— the Cinema Content Creation Cloud (C4), semantics, and NoSQL technologies. Production of the short film, "The Suitcase" was used by the ETC Production in the Cloud Working Group as a technical test for the development of a semantic data model utilizing linked data to curate authoritative metadata on-set and retain it for downstream use (removing the need for later recreation or replication); of automated processes for creating additional or enhanced metadata when necessary; and for the exploration of rapid software development utilizing RDF data and SPARQL queries.[12]

- **Use of Semantic Technology in Time-Based Tagging**: A proof-of-concept project at the U.S.-based Public Broadcasting Service (PBS) combined machine learning and semantic technology to create a scalable process for assigning high-quality time-based metadata to moving image content. This involved training machine learning algorithms to recognize puppet characters and

---

[9] Described in presentations "Using Wikidata for Tagging Content at Yle," Pia Virtanen, EBU MIM-AI Monthly Call, 2021-09-22 and "Tags at Yle," Mickael Hindsberg, Pia Virtanen, BBC Data, 2021-03-01.
[10] See presentation "Data Governance at France TV for Better Usage and Exploitation of Programmes," EBU MDN Workshop 2021, by Matthieu Parmentier (FTV).
[11] See presentation "Metagraph — A Graph Based Repository to Connect Arbitrary Systems," EBU MDN Workshop 2021, by Jürgen Grupp (Südwestrundfunk).
[12] "The Suitcase: Using C4, Semantics, and NoSQL for Managing Motion Picture Data," *SMPTE Motion Imaging Journal*, May/June (2017), by Nancy Silver, Kaki Ettinger, Joshua Kolden, Michael Malgeri, and Erik Weaver.

provide textual input — the character's names and other concepts extracted using both image analysis and from episode transcripts — that could be matched with an in-house taxonomy. The content chosen for the project was a selection of episodes from "Sesame Street," a show aimed at preschoolers produced by the Sesame Workshop, while the taxonomy was developed by the PBS Education division. [13]

Functional Stack

Below is an illustrative, rather than prescriptive, representation of the components of a semantic layer within an architecture. The key characteristics of an eco-system employing a semantic layer or knowledge graph is the separation of the logical and physical layers and the use of predictable and standardized technologies in the logical representation.

The separation of "meaning" or a machine-readable semantic layer from content files and application code supports broad accessibility away from the technical constraints of any given system. This makes the knowledge graph available to multiple systems through standard interfaces so that, for instance, rights management; production; post-production; media asset management; distribution processes; business intelligence and analytics; and ML and AI can all utilize not only shared vocabularies but shared meaning. The integration of semantic technology into an eco-system can support the contextualization and enrichment of distributed and heterogenous content with a standardized business vocabulary, and support data-driven workflows.

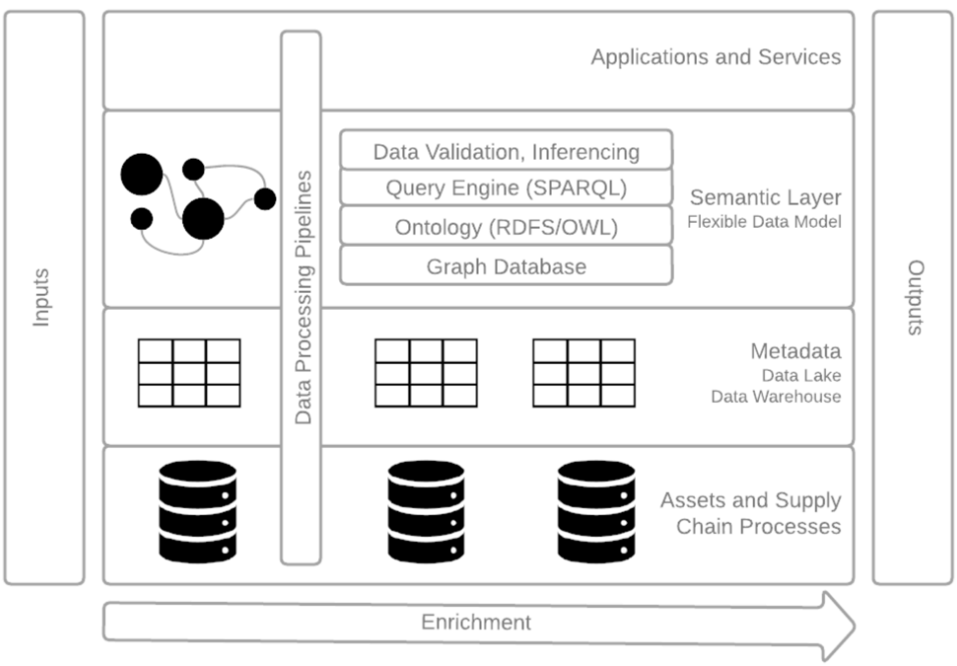

Figure 5: Semantic Layer Components within an Architecture.

[13] "Smart Stacking of Machine Learning and Semantic Technologies in Media Management," *Journal of Digital Asset Management,* Volume 9, Number 4 (2021), by Sally Hubbard, Maureen Harlow, and Athina Livanos-Propst.

Where Next?

Semantic technology is not a magic bullet. It does not necessarily replace other technologies and is not a good fit for every use case. It is not as mature as the relational database/SQL eco-system, which may be more appropriate for predictable and consistent data that is not characterized by many or complex relations, and for which there is a greater pool of expert human resources. However, semantic technology can be smartly stacked with other tools and is in broad terms a good fit for cloud-based eco-systems, having its genesis as a web-based approach to knowledge management.

Many organizations are already implementing enterprise knowledge graphs that utilize semantic technologies, or are exploring doing so, because nearly all organizations are challenged by the management and analysis of their data. Most are looking to become more data-driven or data-centric and to utilize machine learning and analytics, both of which are supported by consistent ontologies and taxonomies.

Where there is likely room for development in the near future is in the media factory — the production and distribution supply chain and in media asset management and media processing workflows. As these migrate to the cloud or hybrid cloud/on-premises and microservice-based deployments, it will become easier to integrate semantic technologies into day-to-day operations, and more vendors will integrate them into their offerings.

References and Resources:

Case Studies and Reports

- EBU Technology and Innovation, particularly the Metadata Developer Network (MDN) and Metadata and Artificial Intelligence (MIM) groups, provides a wealth of resources on current implementations of semantic and related technologies.

W3C Semantic Technology Primers

- *OWL 2 Web Ontology Language Primer (Second Edition),* W3C Recommendation, 2012.
- *RDF 1.1 Primer*, W3C Working Group Note 24, 2014.
- *SKOS Simple Knowledge Organization System Primer*, W3C Working Group Note, 2009.

Papers and Reports

- *A Creative Works Ontology for the Film and Television Industry*, MovieLabs White Paper, 2018.
- *EBU-MIM Semantic Web Activity Report*, EBU Tech Report 019, 2015.
- Evain, Jean-Pierre, Tobias Burger, *Semantic Web, Linked Data and Broadcasting — More in Common Than You Think!*, EBU Technical Review, Q1, 2011.
- *The Evolution of Production Workflows*, MovieLabs White Paper, 2020.

Books and Monographs

- Barassa, Jesus; Amy E. Hodler; Jim Webber, *Knowledge Graphs: Data in Context for Responsive Businesses,* 2021. (Note this assumes an eventual convergence of semantic and property graphs and discusses both.)
- DuCharne, Bob, *Learning SPARQL: Querying and Updating with SPARQL 1.1*, 2nd Edition, 2013.
- Hendler, James, Dean Allemang, Fabien Gandon, *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*, 3rd Edition, 2020.
- Martin, Sean; Ben Szekely; Dean Allemang; *The Rise of the Knowledge Graph: Towards Modern Data Integration and the Data Fabric Architecture*, 2021.
- Wood, David; Marsha Zaidman; Luke Ruth; Michael Hausenblas, et al, *Linked Data: Structured Data on the Web,* 2014.

Further Reading

These articles provide broader insights on the use of semantic technologies in various scenarios: within supply chains; with microservices; and for recommendations systems.

- Chatzimichailm, Angelos; Evangelos A. Stathopoulos; Dimos Ntioudis; Athina Tsanousa; Maria Rousi; Athanasios Mavropoulos; Georgios Meditskos; Stefanos Vrochidis; Ioannis Kompatsiaris, "IoT and Semantic Web Technologies," *Semantic IoT: Theory and Applications*, 2020.
- Diepenbrock, Andreas; Florian Rademacher; Sabine Sachweh, *An Ontology-Based Approach for Domain-Driven Design of Microservices Architecture*, Lecture Notes for Informatics (LNI), Informatik, 2017.
- Hilal, Samia, "A Survey of Ontology-Based Frameworks for Sustainable Supply Chain Interoperability and Collaboration," *Handbook of Research on Interdisciplinary Approaches to Decision Making for Sustainable Supply Chains,* 2020.

- *Information, Special Issue: Information Retrieval, Recommender Systems and Adaptive Systems,* 12, no. 8, 2021.
    - Kopsachilis, Vasilis, et al, "Semantically-Aware Retrieval of Oceanographic Phenomena Annotated on Satellite Images."
    - Esheiba L, Elgammal, et al, "A Hybrid Knowledge-Based Recommender for Product-Service Systems Mass Customization."
- Petersen, Niklas, et al, "Realizing an RDF-Based Information Model for a Manufacturing Company — A Case Study," *SEMWEB*, 2017.
- Steindl, Gernot, and Wolfgang Kastner, "Semantic Microservice Framework for Digital Twins," *Applied Sciences* 11, no. 12, 2021.
- Taherizadeh, Salman; Dimitris Apostolou; Yiannis Verginadis; Marko Grobelnik and Gregoris Mentzas, "A Semantic Model for Interchangeable Microservices in Cloud Continuum Computing," *Information* 2021, 12, 40.

Glossary

The meaning of terms used in the broad area of knowledge management is often fuzzy. The following definitions are deliberately brief and intended to give a thumbnail view of the various technologies discussed in this paper. More detailed definitions are available from the published glossaries listed below, and other sources may provide slightly different definitions. Terms from the broader metadata management and utilization domain are included, as are terms describing property graphs in acknowledgement of the similarities between the two forms of graphs and of recent work that has explored merging them or developing common standards and mechanisms.

Terms

- **AI (Artificial Intelligence):** a broad field within computer science that encompasses **machine learning** and knowledge-based systems that don't rely directly on data and refers to the ability of a computational system to autonomously perform tasks commonly associated with intelligent beings.
- **API (Application Programming Interface):** an intermediary mechanism that specifies the rules for the exchange of structured data between software applications, including **microservices**, at a given location or **endpoint**.
- **Arc:** a directed or ordered relationship between two **nodes** within a **property graph**. Analogous to a **Directed Relation** in a **semantic graph**. See also **Edge**.
- **Attribute**: a characteristic or value that is associated with a **class** within an **ontology** (or an **instance** of that class) via a **property**.
- **Axiom:** a premise assumed to be true and used in reasoning. In an **ontology**, each **triple statement** is an axiom.
- **Big Data:** a term that refers to very large heterogeneous data sets, characterized by volume, variety and velocity, that present challenges in management, storage, analysis, and exchange. See **Data Lake** and **Data Fabric**.
- **Class:** a general classification or grouping of resources within an **ontology**, the members of which are known as **instances. RDF properties** will use **classes** as their **domain** and (optionally) their **range.**
- **Cloud:** remote networked server infrastructure used for storage, processing and/or services and accessed via internet protocols.
- **Concept:** a "unit of thought" within **SKOS**, generally meaning a term used within a given knowledge organization system or **vocabulary**.
- **Controlled Vocabulary (CV)**: a broad term encompassing various ways of providing lists of allowed terms for use in particular circumstances to keep data consistent, from simple pick lists to hierarchical **taxonomies** to **ontologies.**
- **Data Fabric**: a term indicating a data management and integration framework that supports access to and sharing of data in a distributed data environment and typically includes a **knowledge graph** or **semantic** component to improve contextualization.
- **Data Lake:** a repository of data from heterogeneous sources stored in its raw and/or transformed state, be that structured, semi-structured or unstructured, for analytics or other purposes. See also **Big Data** and **Data Swamp**.
- **Data Mesh:** may be used interchangeably with **data fabric** to refer to data analytics in a distributed architecture with an emphasis on domain-specific teams; or refer to the organizational or procedural aspects of integrated data management rather than to a technological data architecture.

- **Data Warehouse:** a repository of structured data from various sources, generally utilizing some combination of extraction, transformation, and loading (ETL) mechanisms for aggregation and management.
- **Data Swamp:** a term used to refer to a badly maintained **data lake**, or alternatively to characterize the challenges of extracting meaningful information from large and heterogeneous data sets.
- **DBpedia:** a **linked open data** project that extracts structured data from infoboxes in Wikipedia articles, utilizes URIs for entities based on Wikipedia, and publishes **RDF** data about entities. See **Wikidata**.
- **Disjoint:** having no common elements, used within an **ontology** to specify where two classes cannot share an **instance** and utilized in logical inference.
- **Domain:** ontologies model knowledge within a given knowledge domain. In RDF specifically a **property** has a defined domain as its starting point, stating that a given **class** is the **subject** of the property. See also **Range**.
- **Directed Graph**: a **property graph** within which the links (**arcs**) between **nodes** are directed, meaning they run only in one specified direction.
- **Directed Relation:** an RDF / OWL relationship (**property**) that runs in one direction only, from **subject** to **object.** Analogous to an **arc** in a **property graph.** See **Inverse** and **Symmetric Relation.**
- **Edge**: a relationship, particularly a non-directed relationship, between two **nodes** within a **Graph**. See also **Arc**. Edges within **property graphs** may have associated attributes.
- **Endpoint**: a component, generally a URL, of an **API** that provides a location for the exchange of data.
- **Graph:** a data structure in which relationships are persistent and integral components. Organized either as a collection of **nodes** linked by **edges** or **arcs** in a **property graph,** or as **subjects** and **objects** linked by **predicate**-defined **properties** in a **semantic graph**. Note that various sources predict that the two types of graph will merge.
- **Graph Database (GDB):** a type of **NoSQL** database that stores graphs. There are various kinds of graph database, with the primary distinction being between those that store **semantic graphs** and those which store **property graphs**.
- **Graph Theory:** the study of graphs as mathematical structures, initiated by a 1736 paper by Leonhard Euler on the Seven Bridges of Königsberg that examined the possible paths between a given set of locations (**nodes**) within a city using the bridges (**edges**) and also suggested an abstract mechanism for examining any graph (any collection of nodes and edges).
- **Hybrid Deployment Model:** the utilization of some combination of **cloud** hosted, on-premises, and/or virtualized computing resources and infrastructure for workflows and/or storage.
- **Inferencing:** the process of deriving new logical conclusions from existing premises. The logic encapsulated within the **statements** of an **ontology** form the basis for deriving new conclusions from or about data.
- **Instance:** an individual member of a **class**.
- **Inverse Relation:** an RDF/OWL relationship that has two components, with one relation running from **subject** to **object** and the another running in the opposite direction**,** swapping the subject and object roles**.** See **Directed** and **Symmetric Relation.**
- **Knowledge Base (KB):** a broad term that may refer to any kind of repository of information on a given topic or area. In **semantic** terms a knowledge base implies a store of interlinked entities that are machine-readable and traversable.

- **Knowledge Graph (KG):** may be used interchangeably with **knowledge base**, or more specifically refer to an instantiation of a knowledge or data base using **graph** or **semantic technology**, when it may also be used interchangeably with **ontology.** The term was popularized in 2012 with the launch of the Google Knowledge Graph.
- **Labeled Property Graph (LPG):** see **Property Graph.**
- **Linked Data (LD)**: a set of machine-readable mechanisms for inter-relating and contextualizing **instance** and **class** data, using the **RDF** data model and uniform resource identifiers (URIs) to name and disambiguate data **resources**.
- **Linked Open Data (LOD)**: information published using **linked data** principles where the represented knowledge is freely accessible and re-useable. Often used by public institutions and libraries, and famously in the Wikipedia-derived DBpedia.
- **Machine Learning (ML):** various processes that broadly involve "teaching" computer algorithms to recognize data patterns that can be leveraged to in various ways, for instance in classification or to make predictions. "Supervised" machine learning requires labeled data sets for training, while "unsupervised" sub-fields such as deep learning and neural networks can also ingest unstructured and non-labeled data and are more scalable. See **AI**.
- **Metadata**: a broad term that generally indicates supplementary or contextual data describing content or data entities. **Vocabularies** are a source of consistent metadata.
- **Microservice:** a modular service-based capability provided by a granular software component that handles a narrow task and can be managed in isolation, but needs to exchange messages and metadata with other microservices through the supply chain.
- **Named Graph:** an extension to the **RDF** model that adds a "context" resource to triples (making them quads) and allowing them to be grouped by context or provenance and making it easier to partition **graphs** into sub-graphs. See **Semantic Graph**, **Quad-Store**, **Reification**.
- **Natural Language Processing** (**NLP):** the extraction of meaningful information from natural language or unstructured text resources. See **Information Extraction.**
- **Node:** also known as a vertex, the basic unit of a **property graph,** and the thing to which **edges** or relationships are attached and where they intersect. Analogous to **subject** and **object** resources in RDF.
- **Information extraction (IE):** the automatic extraction of structured information from unstructured and/or semi-structured machine-readable documents, likely to utilize **machine learning** and/or **semantic web technologies**. See **Ontology-Based Information Extraction**.
- **Object:** part of the **subject, predicate**, object structure used within **RDF triples**. May be either a **class** or **instance**, or an **attribute** value.
- **Ontology-Based Information Extraction (OBIE):** the use of an **ontology** (or other controlled vocabulary) as the source for appropriate metadata with which to tag resources during **information extraction**.
- **Ontology:** the most expressive and powerful form of **controlled vocabulary,** providing a formal representation of knowledge within a given **domain** that defines concepts, the term(s) used to name them, their properties, and connections between them.
- **Predicate:** part of the **subject**, predicate, **object** structure of **triples**. Predicates define the nature of the relationship between the **subject** and the **object** or assign **attributes** to subjects. See **Property**.
- **Property:** in RDF a property is defined by **predicate**-and may be either a relationship between **classes,** or an **attribute** assigned to a **class.** Properties specify a **domain**, a class that represents the

subject of a **triple**; and a **range** representing the **object** (either a class or an attribute value). In property graphs properties are key/value pairs that be assigned to **edges** as well as **nodes**.

- **Property Graph**: any of various **edge** and **node**-based **graphs** that are associated with various proprietary query languages.[14] Both nodes and edges are assigned (local) identifiers and properties (as key/value pairs). Nodes may be labelled or unlabeled, while edges may be directed or undirected, labeled or unlabeled, and may be associated with attributes. Often optimized for data storage and analytics such as "shortest path" or "most relationships." Currently to be distinguished from a **semantic graph**.
- **Property Graph Database**: a database that stores a property graph.
- **Quad-Store:** a database that stores one or more **named graphs.** See **Semantic Graph.**
- **Quoted Triple:** a triple treated as a single entity and used as the subject or object of another **triple** within the proposed **RDF-star** extension to **RDF**. May also be called an embedded triple.
- **Range**: within RDF the range of a **property** specifies the end point or destination of a directed relationship, which may be a **class** or a metadata value, and is the **object** within a **triple**. See **Domain**.
- **RDF Graph:** See **Semantic Graph.**
- **Relationship**: the connections between entities, which can be fully specified within an ontology: named; given directionality; and have allowed **domains** (starting points) and **ranges** (end points). See **Predicate.**
- **Relational Database (RDB):** a data structure wherein data is stored according to a pre-defined **schema** as rows within tables, with each row having a "primary key" element used to make it unique, and in which data between tables is associated via join operations during queries.
- **Reasoner:** software providing semantic reasoning or a rules engine able to infer logical consequences from a given set of **axioms**.
- **Reification:** various mechanisms by which an **RDF statement** is modeled as a **resource** referenced by another statement. See **RDF-star**, **Named Graph**.
- **Resource:** an abstraction that represents a physical or a conceptual object, or any entity that can be identified by an URI. Resources are what is described by statements within an **RDF** representation.
- **Schema:** a model that describes or determines how data is organized within a database. Semantic and **graph** schemas are considerably more flexible and extensible than those required for **relational databases**.
- **Semantic AI:** the combination of a semantic graph with **AI** and **ML** mechanisms to address issues such as low data quality, a lack of data, or "black box" issues of a lack of insight into how algorithms are arriving at conclusions.
- **Semantic Web Technology:** broadly refers to the various formats and specifications, centered around the **RDF** standard, developed by the W3C to promote common, interchangeable data and knowledge representation formats for use on the web. See **OWL**, **SKOS, SPARQL.**
- **Semantic Graph:** a directed **graph** based on **RDF** that uses the **SPARQL** query language. Data is represented as **triples** (or as quads when an additional context resource is included), and each resource has an assigned URI. Currently to be distinguished from a **property graph.**
- **Semantic Graph Database**: a database that stores a **semantic graph** and can export and import all standard RDF serializations, making data interchange straightforward.

[14] There is a current ISO project to create a standard Graph Query Language (GQL).

- **Statement**: a **triple** within an **ontology**, consisting of a s**ubject**, **predicate**, **object** assertion. Each statement is regarded as an **axiom**.
- **Subject:** part of the subject, **predicate**, **object** structure used within **RDF triples**.
- **Symmetric Relation:** an RDF/OWL relationship property that is the same in both directions, from **subject** to **object** or from object to subject, as the **domain** and **range** are the same **class**. See **Directed** and **Inverse Relation.**
- **Taxonomy:** a form of **controlled vocabulary** that provides a hierarchical organization of things or entities of interest within a given organization or area. May be used together with ontologies, with the taxonomy providing the **instances** of the **classes** defined in the **ontology**.
- **Triple:** a base unit of **RDF** representation, comprised of a **subject**, **predicate**, **object statement** with the subject and property (and optionally the object) referenced by URIs.
- **Triple-Store (triplestore):** a database that stores **RDF** triples. See **Semantic Graph**.
- **Upper Ontology:** an ontology that describes general concepts that are common across all knowledge domains, intended to support cross-domain semantic interoperability.
- **Vertex (Vertice)**: see **Node**
- **Vocabulary**: See **Controlled Vocabulary.**
- **Virtual Graph:** used to provide scalability to **semantic graphs** and allow data held in relational or **NoSQL** storage to be queried via **SPARQL**.
- **Wikidata**: a **linked open data** project that supplements Wikipedia articles, utilizes URIs for entities based on Wikipedia, and publishes **RDF** data about entities. See **DBpedia**.

Additional Glossaries

- W3C Glossary
- ISO Online Browsing Platform Terms and Definitions.

Data Standards and Specifications

The following represents a non-exhaustive list of relevant data and metadata specifications.

- **Dublin Core:** a metadata schema originally comprised of fifteen elements designed to describe resources on the web.
- **FOAF (Friend of a Friend):** an RDF schema for machine-readable modeling of social networks.
- **GraphQL:** a query language designed for use with APIs.
- **Graph Query Language (GQL):** a proposed standard query language for property graphs, and which it has been proposed to extend to support the querying of RDF data. Current property graph query languages are proprietary.
- **JSON (JavaScript Object Notation):** a data format for representing structured data often used for data exchange via API.
- **JSON-LD (JavaScript Object Notation for Linked Data):** an **RDF** serialization syntax based on **JSON**.
- **NoSQL (No SQL** or **Not Only SQL):** used to refer to various forms of non-relational databases, including **graph databases**.
- **N-Triples:** a serialization syntax for **RDF**.
- **N-Quads**: a serialization syntax for **RDF**.
- **OWL (Web Ontology Language):** designed for designing ontologies. Built on **RDF** with greater expressivity, more logical constraints, and annotation support.

- **PBCore:** a metadata schema designed to describe audiovisual assets.
- **RDF (Resource Description Framework):** a structure for representing information that uses the syntax of **triples**, each consisting of a **subject**, a **predicate** and an **object**, to make **statements** about **resources**. The core specification of **semantic technologies**.
- **RDFS (RDF Schema):** a structure for representing a schema or simple ontology in RDF.
- **RDF-star (RDF*):** a proposed extension of RDF that would allow a triple to be treated as a single entity, or as a "quoted" or "embedded" triple. See **Reification**.
- **RIF (Rule Interchange Language)**: a standard for exchanging rules between systems.
- **SHACL (Shapes Constraint Language):** a language for validating RDF graphs.
- **SKOS (Simple Knowledge Organization System**): a technology for structuring **taxonomies** and other **controlled vocabularies,** supporting both dictionary-like definitions and relations between terms.
- **SPARQL (SPARQL Protocol and RDF Query Language):** the standard semantic query language.
- **SPARQL-star (SPARQL***): a proposed extension to SPARQL that would support querying RDF-star content.
- **SQL (Structured Query Language):** the language used to query relational databases.
- **Trig:** a serialization syntax for **RDF**.
- **Turtle**: a serialization syntax for **RDF**.
- **XML (eXtensible Markup Language):** a markup language for representing structured data. often used for data storage.

Appendix 1: Current Ontologies and Standards Organizations

Further information of the principal media ontologies, including on the mission and intent of the producing organizations, and the formats in which the ontologies are available, the license under which it may be used, and any known implementations.

dwerft Digital Media Ontology (DMO)

The DMO is a product of the dwerft: Linked Metadata for Media project, which brought together industry partners to develop innovative media technology solutions and was sponsored by the German Federal Ministry of Education and Research and ran from 2014 to 2021.

The project was intended to address a metadata functionality gap wherein most production tools are focused on content rather than metadata, and their metadata functionality is not interoperable. This makes metadata management and utilization downstream cumbersome. The proposed solution is the Linked Media Data Cloud (LMDC), comprised of the DMO, adapters, a CRUD (Create, Read, Update, Delete) REST API, and a triple store. The API allows tools to be connected via adapters to the LMDC; the adapters map proprietary data to the DMO; and the triple store holds the data in a standardized format ready for optimized and automated re-use downstream, with mapping to required distribution metadata formats.

The DMO has been utilized by dwerft partner transfermedia for a proof-of-concept implementation of an “X-Ray-like” feature wherein scene-based information on cast and locations was extracted from the script, shooting schedule, script-continuity and post-production processes, transformed into the DMO format, exported into a JSON-timeline, and displayed by an HbbTV-application running on a TV/VoD Player. More information on use cases along the value chain are provided on the dwerft scenarios page.

The Digital Media Ontology 1.0 is published and available for download from github in OWL format, without any licensing restrictions.

European Broadcasting Union (EBU) Family of Standards

The European Broadcasting Union is an international alliance of public service media organizations founded in 1950. EBU Technology and Innovation has published a family of semantic web standards over several years which can be used in combination with each other, and with other relevant specifications metadata specifications from the EBU or other sources. Together they provide “a framework for descriptive and technical metadata for use in service-oriented architectures and audiovisual ontologies for semantic web and linked data development.”

The EBU semantic specifications are the EBU Class Conceptual Model ontology; the EBUCore metadata specification, and Reference Data Sets vocabularies for use with EBUCore and CCDM classes. CCDM and EBUCore are currently maintained and downloadable as human-readable PDF format documents and machine-readable RDF schemas. EBUCore is also available in XML, and CCDM online in a graphical RDF representation, while the reference vocabularies are presented online in SKOS format.

The specifications have been available for several years and have been utilized in various ways by multiple organizations, including YLE, TV2 Norway, ABC Australia, the BBC, France TV, NRK, SVT and the ERT Archives.

The specifications are published under a Creative Commons “Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)” license. Users and implementers have the freedom to change them to address their respective needs; must give attribution without suggesting an EBU endorsement; and are encouraged to share the results alike. The standards can be implemented into products for commercial purposes but cannot be traded as a specific feature.

MovieLabs Ontology for Media Creation (OMC)

Motion Picture Laboratories, Inc., known as MovieLabs, is a non-profit joint venture founded by the major Hollywood (i.e., U.S.-based) studios to find solutions to industry challenges. It is dedicated to enabling more efficient production workflows utilizing the latest technologies and modern software systems; improving backend infrastructure for distribution and production; enabling new and better digital delivery for media content; and improving media experiences for consumers.

MovieLabs have identified that improved interoperability and communication between both people and machines is required to realize the potential of new workflow systems, and that this can be supported by a common ontology. The first MovieLabs ontology released was the Creative Works Ontology (CWO), 2018, which described creative works in distribution for purposes of analytics. The primary current MovieLabs ontology is the Ontology for Media Creation (OMC), 2021. The OMC focuses mainly on content creation from concept through production and post. But it also extends into distribution, incorporating and superseding the previous CWO with connected ontologies for creative works and media distribution. It remains under active development.

The OMC is available for download in multiple formats: the ontology in RDF and Turtle (TTL); the vocabularies in both SKOS and comma separated values (CSV); and human-readable documentation in PDF. It is also available online in a graphical RDF representation. The OMC is available for use by companies developing or implementing products, solutions, or services for media creation and distribution. The OMC specifications and resources are distributed under the Creative Commons “Attribution 4.0 International License (CC BY 4.0).”

Society of Motion Picture and Television Engineers (SMPTE)

SMPTE was founded in 1916 and is a U.S.-based, global professional association of engineers, technologists, and executives working to drive the media and entertainment industry forward. It is an internationally recognized standards organization; publishes the *SMPTE Motion Imaging Journal*; and provides extensive membership and education services. The SMPTE Media in the Cloud initiative is intended to share and promote best practices for those navigating the migration to the cloud and to foster a multivendor, global media cloud eco-system that supports content interchange.

Appendix 2: Mapping Exercises

Example SKOS Matches between "Person" Class of EBUCore, dwerft DMO, and MovieLabs CWO				
Mapping between specifications is not intended to be prescriptive or make them "agree" or identical — as each may have specific reasons for modelling the entity as they do, for instance CWO assumes that adopters will build out "Person" according to their own requirements and is therefore comparatively minimalist — but to show where there is useful commonality or overlap, or conversely where there is no match.				
EBUCore	*Ontology*	dwerft Digital Media Ontology (DMO)	*Ontology*	MovieLabs Creative Works Ontology (CWO)[15]
Agent	*SuperClass*	Agent	*SuperClass*	Anything with an ID
Person	*Class*	Person	*Class*	Person
Properties	SKOS Match	*Properties*	SKOS Match	*Properties*
Affiliation [ebucore:hasAffiliation]	has exact match	affiliation		
Career event	has broader match	biography		
Birthplace	has exact match	birthPlace	has exact match	hasBirthplace
Country of birth	has exact match	countryOfBirth		
		nationality	has close match	hasCountryofCitizenship
Country of death	has exact match	countryOfDeath		
Date of birth [ebucore:dateOfBirth]	has exact match	dateOfBirth		
		birthDate		
Date of retirement	has broader match	biography		
Death place				
Date of death	has exact match	dateOfDeath		
Age	has exact match	age		
Description [ebucore:personDescription]	has broader match	biography		
Education	has broader match	biography		
Family information				
Given name	has exact match	givenName		
Local given name				
		birthName		
Middle name	has exact match	middleName		
Family name	has exact match	familyName		
Local family name				
Person name	has close match	name	has close match	hasPersonName
		hasName		
Nickname [ebucore:nickName]	has exact match	nickname		
Gender [ebucore:gender]	has exact match	gender	has close match	hasGender
				isTransgender
Hobbies	has broader match	biography		
Homepage (private)				
Identification picture	has close match	relatedPicture		
Identification picture locator				
Person identifier			has close match	hasPersonIdentifier
Marital Status	has exact match	maritalStatus		
Occupation	has exact match	occupation	has close match	isPersonContributing
Person height				
Person weight	has exact match	weight		
Person type				
Personal event	has broader match	biography		
Salutation title	has exact match	salutationTitle		
Suffix				
Telephone (private) [ebucore:officeMobileTelephoneNumber])	has broader match	telephoneNumber		
Telephone (private) [ebucore:officeTelephoneNumber]	has broader match	telephoneNumber		
Telephone (private) [ebucore:privateMobileTelephoneNumber]	has broader match	telephoneNumber		
Telephone (private) [ebucore:privateTelephoneNumber]	has broader match	telephoneNumber		
Private email	has broader match	emailAddress		
Office email	has broader match	emailAddress		
Username	has exact match	username		
				contributesTo

[15] Note that The MovieLabs CWO is being superseded by additions to the MovieLabs Ontology for Media Creation.

		job	has close match	isPersonContributing
				isSeenAsStar

Example SKOS Matches: Portraying Characters

Connecting characters to the actor or thing (e.g., a computer graphics model) that represents them is important information and can be quite complex: the same actor can be used for multiple characters, the same character can be played by two actors simultaneously (voice and physical presence), and the same character can be played by multiple people or things (young and old versions, or a real version and an animated version.) DMO and OMC treat this connection as an entity in its own right (a mechanism called reification), whereas Wikidata attaches it to the Character entity. The reified version is compatible with the non-reified one, and just needs an “extra hop” when traversing the graph.

It is also important to connect the portrayal of a Character to a Scene or other place it is used. DMO and OMC support this through hasRolePlay (DMO) and includesPortrayal (OMC) relationships; these two relationships are a close match, and their inverses (i.e., the connections running in the opposite direction in an inverse or bidirectional relationship) get you from the RolePlay/Portrayal back to where it is used.

Wikidata	*Ontology*	dwerft Digital Media Ontology (DMO)	*Ontology*	MovieLabs Ontology for Media Creation (OMC)
			SuperClass	Depiction
	Class	RolePlay	*Class*	Portrayal
Properties	SKOS Match	*Properties*	SKOS Match	*Properties*
		hasRolePlayCharacter	has exact match	portrays
		RolePlayedBy	has close match	hasPortrayer
		hasProps	has close match	uses
		^hasRolePlay	has close match	^includesPortrayal.
fictional character/film character	*Class*	Character	*Class*	Character
Properties	SKOS Match	*Properties*	SKOS Match	*Properties*
performer	has close match	playedBy	has exact match	isPortrayedBy/hasPortrayer
name	has exact match	birthName	has exact match	hasCharacterName
family name	has exact match	familyName	has close match	has last name
given name	has exact match	givenName	has close match	has first name
				has species
narrative age	Has close match	age		

Credit and Acknowledgements

This guide was written by Sally Hubbard with generous input, feedback and references provided by many others, particularly: Jeff Davidson (Viacom CBS); Raymond Drewry (MovieLabs); Mark Guelbahar (transfermedia); Jim Helman (MovieLabs); Daniel Lucas (MovieLabs); Louis Matignon (Perfect Memory); Guillaume Rachez (Perfect Memory); Alexandre Rouxel (EBU); Jeremy Tarling (BBC); Tormod Værvågen (EBU); Lucille Verbaere (EBU).