

Temario de Ayudante de Biblioteca de la Administración General del Estado

Este temario ha sido elaborado por un opositor, para presentarse al proceso selectivo de Ayudante de Bibliotecas de la Administración General del Estado en la [convocatoria de 2021](#).

Incluye todos los temas, de legislación y específicos de bibliotecas, del programa correspondiente a la convocatoria de la Administración General del Estado para cubrir plazas de Ayudante de Bibliotecas en el Ministerios de Cultura y Deporte, Ministerio de Defensa, Ministerio de Asuntos Exteriores, Unión Europea y Cooperación y Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática. «BOE» núm. 149, de 23 de junio de 2021.

Temario completo disponible en:

<https://www.bibliopos.es/>



Temario de Ayudante de Biblioteca de la Administración General del Estado, cedido por su autor a Bibliopos.es para su publicación bajo licencia [Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License](#).

Bajo esta licencia puedes utilizar libremente el temario para uso personal y compartirlo siempre que [cites la fuente](#) y proporciones un enlace a la [licencia](#). No puedes hacer uso comercial del documento.

B13 Datos enlazados

Vocabularios y ontologías de uso en bibliotecas. BIBFRAME.

Los seres humanos piensan asignando significado a signos (palabras, imágenes, situaciones, etc.) y organizándolos en categorías. La Web evoluciona para incorporar estos significados y categorías, acercándose más a la forma de pensar humana mediante una indización y clasificación previas. Los “documentalistas robotizados” no tiene por qué comprender la información, sólo precisan de metadatos y ontologías para indizar y clasificar los documentos o recursos de la Web. Según la W3C (World Wide Web Consortium, consorcio internacional encargado de definir los estándares para la Web), la **Web semántica**, impulsada por Tim Berners-Lee, es una Web extendida dotada de mayor significado, en la que cualquier usuario de Internet podrá encontrar soluciones a sus preguntas en un menor tiempo y de forma sencilla gracias a una información mejor definida.

Datos enlazados

Junto con el desarrollo de la Web semántica se popularizó el concepto de *linked data* (**datos enlazados o vinculados**), que permite la publicación de datos estructurados para que puedan ser interconectados y más útiles.

La mayoría de los sitios web están construidos en lenguaje HTML (HyperText Markup Language, desarrollado por W3C), con marcas o etiquetas, que describen el formato del documento, en base a unos estándares que permiten que todos los ordenadores puedan reproducir correctamente esos documentos. Sin embargo, el problema de este lenguaje es que se encuentra destinado principalmente a la visualización y no a la estructura de la información. Se puede decir que la World Wide Web, basada en documentos y enlaces de hipertexto, fue diseñada para la lectura humana y no para que la información que contiene pudiera tratarse de forma automática.

Además, la Web actual es una bastísima red de documentos, que funciona más como una base de datos que como una red de documentos, de modo que si se desea buscar información (por ejemplo escribiendo una palabra clave en un buscador) es necesario procesar los documentos que la contienen para llegar a ella. En la Web no se realizan preguntas, sino que los buscadores exploran los documentos de la Red para encontrar las palabras claves de una búsqueda, de ahí que muchas búsquedas devuelvan resultados redundantes. La diferencia entre la Web del hipertexto y la Web semántica es que mientras una vincula páginas o documentos en HTML, la segunda enlaza datos estructurados. Los datos enlazados es la forma que tiene la Web semántica de vincular los distintos datos que están distribuidos en la Web, de manera que se referencian de la misma forma que lo hacen los enlaces de las páginas web. La idea básica es que se puede dar un localizador a cada concepto que se quiera representar y enlazar los conceptos directamente, en vez de enlazar a los documentos que contienen esos conceptos. Así, se obtendría una red de conceptos en vez de documentos, obteniendo respuestas precisas al hacer preguntas.

La primera presentación pública de la propuesta la realizó Tim Berners-Lee en 2009. Los datos enlazados, como parte de la Web semántica, se basa en la aplicación de ciertos principios básicos y necesarios. Los **cuatro principios** definidos que caracterizan los datos enlazados son:

- a) Utilizar URI (del inglés Uniform Resource Identifier, en español identificador uniforme de recursos) para identificar unívocamente los recursos publicados en la Web.
- b) Usar URI compatibles con el protocolo HTTP de manera que las personas puedan localizar y consultar (es decir, desreferenciar) estos recursos en la Web.
- c) Proporcionar información útil acerca del recurso en esas direcciones, usando estándares como el caso de RDF (Resource Description Framework, Marco de Descripción de Recursos, estándar para describir los recursos web, especificar metadatos y representar

información, es decir, que proporciona información descriptiva simple sobre los recursos que se encuentran en la Web).

- d) Incluir enlaces a otras URI relacionadas con los datos contenidos en el recurso, de forma que se potencie el descubrimiento de información en la Web.

Por tanto, los datos enlazados permiten publicar, compartir y conectar datos a través de la Web semántica mediante el uso de URI y RDF. Todo este proceso se desarrolla a través de la tecnología Web pero con un cambio fundamental: los datos publicados están pensados para su lectura, interpretación y procesado automático por parte de programas informáticos. El uso de datos enlazados permite publicar información migrada desde diferentes fuentes en las que se encuentra almacenada, siempre usando el lenguaje RDF, para permitir enlazarla con otros datos, de forma que aumente la conectividad entre la información y se facilite así la recuperación y la agregación de información relacionada. La información que se provee es de especial usabilidad, ya que se puede combinar e integrar fácilmente con otra información, como por ejemplo para obtener datos estadísticos o cruzar datos.

Los datos enlazados permiten crear vínculos entre elementos de un mismo recurso o de recursos diversos, de modo que sean legibles por las máquinas. Tales datos se enlazan desde tres perspectivas: sintáctica (se establecen estructuras con RDF sujeto-predicado-objeto), semántica (se instauran relaciones significativas de diversa naturaleza: equivalencia, jerarquía, asociación, creación, datos de edición, etc.) e informática (se posibilita la comprensión de tales relaciones en un sistema). Los datos enlazados en la Web posibilitarán un espacio interconectado en el que los recursos tendrán una descripción unitaria alta e integrarán aspectos de contenido con asociaciones significativas al mismo recurso o a otros.

El término *linked data* no implica que los datos deban ser gratuitos, de utilización libre o abiertos como postula el concepto *open data* (datos abiertos, son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen). Sólo cuando ambos conceptos se unen surge el término **linked open data (LOD) o datos abiertos enlazados**, que indica que son datos abiertos en RDF. Esto significa que el usuario puede enlazar datos provenientes de diversas fuentes, instituciones u organizaciones, explorar y combinar estos datos de manera libre y sin restricciones de copyright para nuevos desarrollos web.

A escala internacional, en 2004 el W3C recomendó a las bibliotecas publicar sus almacenes de datos (registros bibliográficos y de autoridades) utilizando tecnologías de la Web semántica, asegurándoles que así aumentaría su impacto digital y su utilidad social. Actualmente existen distintas iniciativas, entre los que se puede mencionar:

- La Digital Public Library of America (DPLA, iniciativa de la Harvard University para brindar acceso digital a colecciones de bibliotecas, archivos y museos de Estados Unidos) y Europeana (biblioteca digital de las colecciones que se encuentran en las bibliotecas nacionales e instituciones culturales de los Estados miembros de la Unión Europea), lanzaron en el año 2013 una aplicación de datos abiertos enlazados que permitía la consulta conjunta de las dos bibliotecas digitales: **DPLA/Europeana Query**.
- datos.bne.es (Datos enlazados en la BNE), desde 2011, es un proyecto conjunto del Ontology Engineering Group (OEG, Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid) y la Biblioteca Nacional de España destinado al enriquecimiento de la Web semántica con datos bibliográficos provenientes de los catálogos bibliográficos y de autoridades en formato RDF. Así, los datos de la BNE se encuentran enlazados o conectados con otros datos procedentes de fuentes de datos internacionales como VIAF (Virtual International Authority File) o DBpedia (proyecto para la extracción de información estructurada de Wikipedia, que permita hacer preguntas sofisticadas contra Wikipedia). La versión 2.0 (2014) incluye los enlaces a ejemplares digitalizados en la Biblioteca Digital Hispánica y presenta los datos mediante una ontología propia de BNE.

Vocabularios y ontologías de uso en bibliotecas

En cuanto a la organización de la información se proponen estructuras que facilitan el ordenamiento de conceptos y temáticas. Estas proveen el acceso a los contenidos de una manera ágil, rápida y con lenguajes comprendidos por el usuario. Los **sistemas de organización del conocimiento** (en inglés, *knowledge organization systems*, KOS) son instrumentos que posibilitan categorizar recursos de acuerdo con un sistema de organización determinado para facilitar su posterior recuperación y permiten al usuario interpretar estructuras de conocimiento de una manera organizada. Cualquier sistema de organización del conocimiento se basa en una serie de unidades de conocimiento llamadas conceptos. Estos conceptos se traducen en palabras del lengua natural y luego se transforman en términos que permiten describir el contenido de los documentos. Para llevar a cabo estas tareas se utilizan herramientas que permiten, por un lado clasificar o categorizar los conceptos (los sistemas de clasificación y las taxonomías como herramientas de categorización) y por el otro realizar un control de ese vocabulario (los tesauros y ontologías).

Tanto los tesauros como las ontologías son herramientas que sirven para estructurar conceptualmente determinados ámbitos del conocimiento por medio de vocabularios controlados. La diferencia entre los tesauros y las ontologías radica en la complejidad ya que estas últimas introducen un mayor nivel de profundización semántica y proporcionan una descripción lógica y formal que puede ser interpretada tanto por las personas, como por las máquinas, mientras que los tesauros sólo pueden ser interpretados por humanos.

En este contexto, las **ontologías** figuran como un nuevo abordaje computacional de representación y organización de contenidos informativos. Reformulando la definición de Tom Gruber, una ontología es una especificación formal, explícita, de una conceptualización compartida. Una conceptualización consiste en identificar los conceptos y las relaciones conceptuales que se asumen que existen y son relevantes para crear la abstracción de algún fenómeno, llamado dominio.

En el ámbito bibliotecario, el vocablo *ontología* fue introducido en la literatura especializada en 1997 por Brian C. Vickery y en 2000 Hodge las instaure como un sistema para la organización del conocimiento para ambientes digitales. Una ontología puede ser definida como un sistema de representación del conocimiento que posibilita describir formalmente las propiedades y relaciones de un determinado modelo conceptual, favoreciendo la realización de inferencias automáticas en los procesos de organización y recuperación de recursos de información. Las ontologías potencian la representación de las relaciones existentes entre los conceptos que traducen los contenidos informativos, instituyendo nuevas formas de acceso y favoreciendo mayor flexibilidad y calidad en la recuperación de informaciones, por medio de la realización de inferencias. Por tanto, las ontologías añaden capacidad de deducción de conocimiento no expresado de manera explícita.

Las ontologías suponen un salto cualitativo respecto de los sistemas de organización del conocimiento tradicionales. La semejanza las ontologías y los tesauros radica en que son herramientas de representación y recuperación de información y estructuran conceptualmente determinadas porciones de la realidad. En cambio, la diferencia se basa en el nivel de abstracción con el que se construye la organización conceptual.

Aunque se componen principalmente de un vocabulario (listas de términos o conceptos que suelen ir acompañadas de definiciones), si solamente fuera un vocabulario tendría un uso muy limitado, ya que no se podría expresar el significado pretendido de los términos contenidos en la ontología. Como todos los conceptos de una ontología pueden ser utilizados como puntos de acceso para consultas, posibilitan expresiones de búsqueda más adecuadas y precisas. Definen los conceptos de forma específica describiendo las propiedades y relaciones de cada concepto y facilitan la comunicación entre los seres humanos y las máquinas. Las ontologías favorecen la representación conceptual de un vocabulario común dentro de un dominio específico, hacen que se incluyan definiciones de los conceptos del dominio, y todas aquellas relaciones que tienen que ser interpretadas por la máquina. Trabajan con conceptos relacionados (como tesauros y redes semánticas) pero permiten definir relaciones semánticas complejas, reglas y axiomas, que no están

presentes en el resto de sistemas de organización del conocimiento. Según Gruber, para conseguirlo, recurre a ciertos **elementos**: conceptos, relaciones entre esos conceptos, funciones (un tipo concreto de relación, como asignar-fecha), instancias (elementos concretos de conceptos) y axiomas o reglas de restricción (normas válidas en el dominio de aplicación de la ontología).

Para el desarrollo de ontologías se hace uso de la Web semántica a través de lenguajes de representación, los cuales generan un significado coherente y estructurado de la información procesada. La tecnología **RDF** (Resource Description Framework) facilita un esquema de datos básico, el cual se apoya en XML (eXtensible Markup Language, traducido como lenguaje de marcado extensible), lenguaje destinado a precisar los metadatos. **OWL (Web Ontology Language)** es un lenguaje de etiquetado semántico para publicar y compartir ontologías en la Web. Se trata de una recomendación del W3C, y puede ser usado para representar explícitamente el significado de términos en vocabularios y las relaciones que existen entre ellos, esto configura una ontología. En realidad, OWL es una extensión del lenguaje RDF: emplea las tripletas de RDF, aunque es un lenguaje con más poder expresivo. La versión actual es el OWL 2, publicado en 2012.

Las ontologías pueden utilizarse para dos **funciones** principales: favorecer una mejor clasificación y mejorar la renovación y tasa de precisión de recuperación por medio de búsquedas semánticas. Para Adriana Suárez Sánchez, las ontologías presentan diferentes aplicaciones en el ámbito bibliotecario como:

- representación estructurada de dominios (o ámbito del conocimiento, que busca esquematizar el saber humano, permitiendo agregar atributos, definiciones y relaciones asociadas con conceptos o desplegadas para la exploración temática de los dominios),
- indización de recursos de información digitales (especialmente en bibliotecas digitales y repositorios),
- generación de aprendizaje entre estudiantes y usuarios (actúan como mapas disciplinares en los que se explora y conoce el dominio, los temas, los subtemas, los conceptos, las instancias específicas, etc.),
- construcción de la Web semántica (mediante sentencias RDF formadas por tripletas sujeto-predicado-objeto, que implica que un sistema estará en posibilidades de dar respuesta a búsquedas más complejas), y
- estructuración de redes de datos enlazados (posibilitan la creación de redes, ya sea a nivel de un recurso, al establecer asociaciones entre los atributos de una entidad bibliográfica, o entre varios recursos, al crear vinculaciones a partir de elementos seleccionados).

SKOS (Simple Knowledge Organization System) es una iniciativa del W3C (se publicó como recomendación W3C en 2009) que ofrece una manera de representar esquemas de clasificación, tales como listas de encabezamientos de materia, taxonomías, folksonomías, esquemas de clasificación, tesauros y cualquier tipo de vocabulario controlado. SKOS constituye una de las especificaciones más relevantes en el área de la Web semántica para la actividad de bibliotecas y centros de documentación. Por esta razón, la Subdirección General de Coordinación Bibliotecaria decidió en 2011 utilizar la *Lista de encabezamientos de materia para bibliotecas públicas* (LEMBP, segunda edición revisada en 1993) como base para construir una lista de materias adaptada a los principios de los datos abiertos vinculados. Así, la LEMBP ha sido transformada al formato SKOS, actualizada, ampliada con encabezamientos de otros registros de autoridad, como el de la Biblioteca Nacional de España, y vinculada con otras listas internacionales como las bases de datos RAMEAU (de la Bibliothèque Nationale de France), Gemeinsame Normdatei (de la Deutsche Nationalbibliothek) y LCSH (Library of Congress Subject Headings). Los conceptos pueden ser identificados mediante URIs, con rótulos en una o más lenguas, y ser documentados con diferentes tipos de notas. Los diversos conceptos pueden relacionarse semánticamente entre sí, en jerarquías y redes de asociación informales, así como también se los puede agrupar en esquemas conceptuales.

BIBFRAME

Mientras que en los motores de búsqueda de la Web se puede consultar rápida y fácilmente reseñas, portadas, citas, comentarios y valoraciones de otros usuarios y obras relacionadas, obteniéndose de forma integrada diversidad de información heterogénea relativa al criterio de búsqueda documental, los catálogos han encontrado grandes dificultades para formar parte de este contexto, y el contenido que albergan fundamentalmente ha permanecido circunscrito al ámbito informacional bibliotecario. Uno de los principales responsables de esta situación es MARC, estándar concebido por y para la comunidad bibliotecaria, inteligible y procesable sólo dentro de este ámbito, creado muchos años antes del advenimiento de la Web, y por tanto ajeno a los aspectos esenciales de los datos enlazados. La forma en que plasma su contenido, por medio de tablas organizadas en campos, etiquetas, indicadores y códigos de subcampo, lo hace inflexible y no apropiado para expresar relaciones asociativas, de pertenencia o jerárquicas, vital para el despliegue de información en el seno de Internet. Las arañas o indexadores web son incapaces de descifrar esos datos de forma nativa, y en todo caso su recolección estaría a expensas de la manera en que éstos son interpretados y eventualmente publicados en Internet por la heterogénea diversidad de sistemas de gestión o portales bibliotecarios capacitados para procesarlos. Miles de millones de registros en formato MARC, resultado de muchos años de prácticas de catalogación haciendo uso de ese formato, yacen en los catálogos de un ingente número de bibliotecas y permanecen aislados de Internet, y que deben ser migrados a un contenedor que permita la publicación de su contenido como datos enlazados.

La Library of Congress y especialistas de distintas partes del mundo identificaron la necesidad de una adaptación de los formatos MARC21 al modelo de la Web semántica y datos enlazados. Como consecuencia, en mayo de 2011 la Library of Congress anunció una iniciativa llamada **BIBFRAME** (anagrama de Bibliographic Framework) para desarrollar los estándares de descripción bibliográfica hacia un modelo de datos enlazados. La iniciativa tiene como objetivo adaptarse mejor a las futuras necesidades para representar la información bibliográfica en la Web. En marzo de 2013 se lanzó BIBFRAME 1.0 y en el 2016 se publicó la versión 2.0.

El formato de codificación de metadatos BIBFRAME es un estándar de descripción bibliográfica concebido por la comunidad bibliotecaria, que emplea tecnologías de datos enlazados para sacar el máximo partido a las enormes ventajas ofrecidas por la Web semántica y que permite así exponer los datos bibliográficos en la Web. Existe una importante sintonía entre las directrices de RDA (Resource Description and Access) y las entidades, atributos y relaciones de su modelo conceptual LRM (Library Reference Model) y las clases y propiedades (correspondientes a características y a relaciones) de BIBFRAME. El vocabulario se compone de clases y propiedades de RDF. Dicho marco extiende la estructura de enlaces de la Web mediante el uso de tripletas semánticas o sentencias RDF, cada uno de los cuales se identifica a través de una URI (Uniform Resource Identifier). La consulta y manipulación de datos almacenados en formato RDF, se realiza a través del lenguaje SPARQL (acrónimo de Simple Protocol and RDF Query Language). Por otra parte, para hacer accesible la información dispuesta en la Web a través de los buscadores de Internet, ésta debería ofrecerse en un formato de datos estructurados marcado por *Schema.org*, conjunto de vocabularios creados para el intercambio datos estructurados en la Web. Mientras que las tripletas RDF, respondiendo al modelo de BIBFRAME, permiten especificar el contenido de los recursos bibliográficos y de autoridad con gran nivel de detalle y son apropiados para su interacción con los futuros SIGB (Sistemas Integrados de Gestión Bibliotecaria) de última generación, *Schema.org* define vocabularios de carácter más general y de diversa tipología, con el fin de incrustar datos estructurados en páginas web para su uso en motores de búsqueda y otras aplicaciones de la Web.