

Grafos de conocimiento y bases de datos en grafo: conceptos fundamentales a partir de una "obra maestra" del *Museo del Prado*

Knowledge graphs and graph databases: Essential concepts based on a *Museo del Prado's "masterpiece"*

Tomás Saorín

Saorín, Tomás (2019). "Grafos de conocimiento y bases de datos en grafo: conceptos fundamentales a partir de una 'obra maestra' del *Museo del Prado*". *Anuario ThinkEPI*, v. 13, e13f05.

<https://doi.org/10.3145/thinkepi.2019.e13f05>

Publicado en *IweTel* el 1 de diciembre de 2019

Tomás Saorín

<https://orcid.org/0000-0001-9448-0866>

Universidad de Murcia

Departamento de Información y Documentación

Campus de Espinardo, 30071 Murcia, España

tsp@um.es



Resumen: Se presentan los conceptos básicos sobre bases de datos en grafo y grafos de conocimiento, a partir del estudio de caso del modelo semántico digital del *Museo del Prado*, y el impacto en el rediseño de su sitio web, centrado en el acceso enriquecido a su colección y recursos vinculados, y metadatos para el descubrimiento de contenidos.

Palabras clave: Grafos de conocimiento; Bases de datos en grafo; Colecciones digitales; Metadatos; Descubrimiento de contenidos; *Museo del Prado*.

Abstract: The basic concepts of graph databases and knowledge graphs are presented, based on the case study of the digital semantic model of the *Museo del Prado*, and the impact on the redesign of its website, focused on the enriched access to its collection and linked resources, and metadata for content discovery.

Keywords: Knowledge graphs; Graph databases; Digital collections; Metadata; Content discovery; *Museo del Prado*.

1. Introducción

En esta nota vamos a tratar de hacer una aproximación a un concepto cada vez más presente como forma de organizar información, datos o metadatos, denominado grafos de conocimiento, bases de datos en grafo o *Knowledge Base*, y que encontramos habitualmente como uno de los componentes de nuevos sistemas de información, entornos para el descubrimiento de información, sistemas de recomendación e inteligencia artificial. Las tecnologías y sistemas implicados en todos estos sistemas de nueva generación son muchos, pero la idea de grafo, y sus implementaciones, son un elemento esencial que

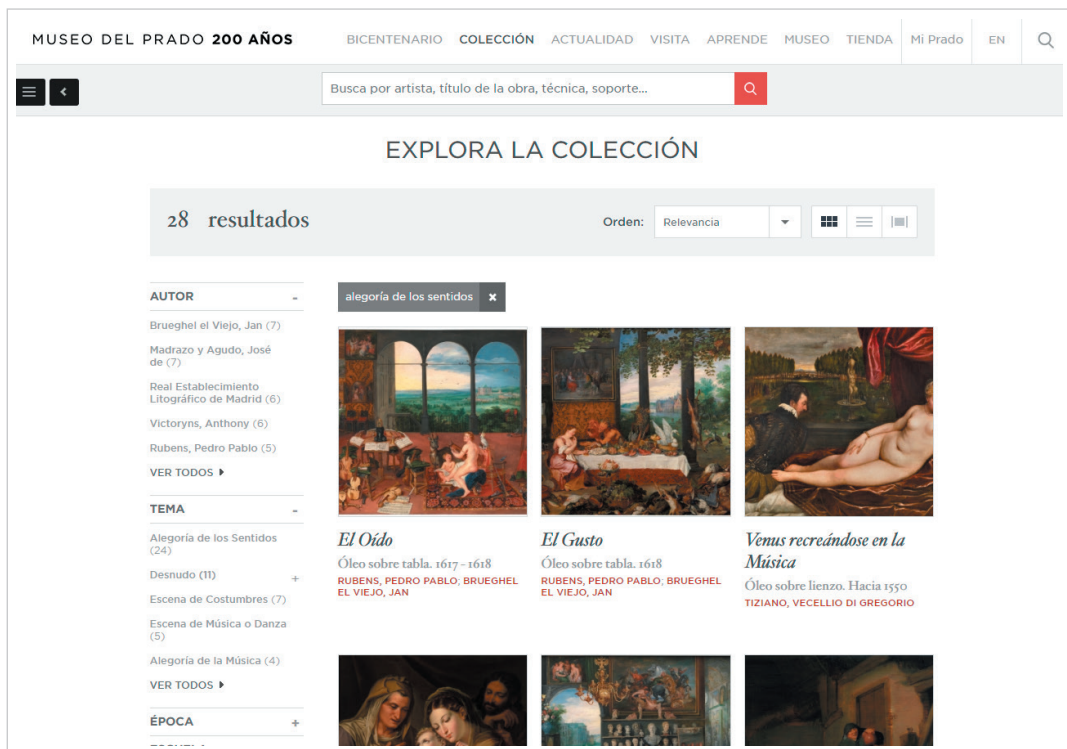


Figura 1. Exploración de la colección en la web del *Museo del Prado*.

conviene comprender si trabajamos en el campo de la web de datos, los metadatos, la reutilización de información o el posicionamiento web.

La idea de una base de datos, con una estructura de campos y relaciones entre tablas, basada en el modelo relacional, forma parte de los conocimientos básicos sobre tecnologías para manejar información de forma estructurada. El concepto de ficha es intuitivo y su conversión en un conjunto de tablas relacionadas, a través del análisis E-R de entidades y relaciones, forma parte de las destrezas básicas en diseño de bases de datos. La idea de conjuntos de campos normalizados en los que se concreta la información que gestionamos coincide también con el concepto de esquema de metadatos. Desde estas bases de datos relacionales, óptimas para gestionar datos y procesos transaccionales, las bases de datos documentales, que almacenan contenidos semiestructurados (documentos, contenidos intensivos en lenguaje natural, con profundidad de significados) ofrecen posibilidades de búsqueda textual avanzada, mediante índices y procedimientos para recuperar información en búsquedas inexactas, aproximativas, exploratorias. Este es el marco estándar cuando hablamos de bases de datos.

Para hablar de grafos, empezemos con un ejemplo basado en un caso de éxito: el *Museo del Prado* ha transformado sus sistemas de información de archivos, biblioteca y colecciones mediante un grafo de conocimiento. Se trata de un proyecto de largo recorrido, que desde 2012 ha ido creciendo y dando resultados, inicialmente como biblioteca digital, luego a nivel interno como "Integración de servicios documentales" (Martín-Bravo; Pantoja-Ferrari, 2013), dejándose ver ya en *microsites* de gran relevancia como el de *Goya en el Prado*, y llegando finalmente a transformar la web del museo (Pantoja-Ferrari, 2016). Fruto de un proyecto tecnológico con financiación de *Telefónica* y con la participación de la empresa *GNOSS* (Alonso-Maturana, 2015), tenemos una integración de información que demuestra todo su potencial convirtiéndose en el núcleo de la propia web del museo. El rediseño del concepto de web llevado a cabo por el museo pone en el centro de la experiencia de usuario el descubrimiento de sus colecciones, pero aplicando un modelo basado en un grafo de conocimiento, frente al modelo convencional que podemos encontrar, por ejemplo, en *Ceres*, la *Red Digital de Colecciones de Museos de España*. La transformación digital del *Museo del Prado* merece más atención que la que le dedicamos en esta nota, y es la "obra maestra" a la que nos referimos en el título, y sobre la cual hemos incluido en la bibliografía el enlace a varias exce-

"Los metadatos de un grafo pueden ser el núcleo del diseño de experiencias digitales enriquecidas"

lentes conferencias del responsable del *Área de Desarrollo Digital*, Javier Pantoja-Ferrari y otros técnicos implicados, que recomendamos vivamente ver.

<https://www.museodelprado.es/coleccion>

<http://ceres.mcu.es/>

<https://www.goyaenelprado.es>

A cierta distancia podemos pensar que ambos, *Ceres* y *El Prado*, son proyectos análogos, pero entre ellos hay una diferencia importante en la capacidad de navegar por las colecciones y encontrar relaciones no evidentes, o encontrar propuestas de otras lecturas de las obras. Los recursos canalizados por ambos tienen un alto valor por sí mismos (contenido), y además por el tratamiento sistemático que han recibido (metadatos). Sin embargo, en la solución de *El Prado* no solo encontramos un diseño elegante, sino algo que podríamos denominar como "profundidad y riqueza de contenido descubrible". Mientras *Ceres* pone a la ficha en el centro del proyecto y permite consultar o navegar por términos controlados planos (Material/Soporte; Clasificación: Objeto; Estilo; Iconografía) en *El Prado* encontramos, además, elementos de contexto y descubrimiento, en la forma de refinado de la navegación por etiquetas (búsqueda por facetas) que nos sugieren rutas de recorrido de la colección y conecta la densidad de los metadatos con la experiencia de uso del sitio web. Pantoja sugiere explorar la colección a través de los pájaros o instrumentos musicales que aparecen en los cuadros, en una

"experiencia de búsqueda y descubrimiento constante en el que de una obra vas saltando a otra en un recorrido que es propio y no delimitado a ningún criterio prefijado" (*My art diary*, 2016).

Además, en el mismo acceso a la colección podemos encontrar la bibliografía básica vinculada a cada obra, contenidos multimedia de conferencias y eventos, un acceso al autor como un contenido de primer orden: obras, autores coetáneos o línea de tiempo. Y, si fuéramos un investigador dentro de la intranet del museo, acceso a los expedientes de archivo relacionados con el ciclo de vida de la obra (*Comunidad Baratz*, 2017), aunque la integración de sistemas documentales la apreciamos ya con bastante claridad al usar el sitio web *Goya en el Prado*. En el acceso a la colección, se aplica una estrategia editorial, a lo *Europeana Collections*, creando selecciones temáticas que se presentan con textos y elementos muy cuidados (Dibujos y estampas, Pintura del siglo XIX, etc.). El resultado es brillante, en muchos sentidos¹, pero por dentro siempre sentimos el latido del "ingenio" que consulta a la base de datos integrada de la colección, que es su motor, su *powered by*. Parece un sitio web, pero es en realidad una capa sobre una potentísima base de datos en grafo, y ese es el concepto sobre el que queremos llamar la atención.

2. Teoría de grafos

El término grafo tiene una vida propia fuera de las bases de datos, a través de la teoría matemática de grafos y sus aplicaciones. Los grafos matemáticos crean modelos simplificados de la realidad a través de redes conectadas, permiten hacer cálculos de centralidad, encontrar caminos óptimos para ir de un punto a otro, establecer relaciones indirectas entre elementos y forman parte, en el mundo académico, de las medidas para los estudios de citas y patrones de producción bibliográfica. También alimentan la capacidad de análisis de la interacción en redes sociales y fenómenos en los que hay un alto grado de interacciones entre elementos conectados o conectables. Euler, vértices, aristas ... son el sustrato matemático para conceptos como nodos y relaciones, que tienen una comprensión intuitiva cuando hacemos mapas mentales y diagramas, pero que pueden pasar del papel de dibujo a una base de datos procesable.

3. Bases de datos en grafo

El universo SQL ha sido desbordado por la necesidad de modelar mejor los datos conectados: datos que se enriquecen con otros datos, con potencial de generar, mediante su análisis y explotación, nuevos *insights* más allá del *reporting* tradicional. La conexión de datos permite sugerir nuevas ideas, encontrar patrones, inferir comportamientos. Los grafos están por todas partes, en gigantes globales como *Google*, *Twitter*, *Facebook* o *Amazon*, en gigantes cercanos como *BBVA* o *Telefónica*, pero también hay ya tecnologías asequibles para cualquier organización o empresa con algo de sangre en las venas y ganas de explorar el *graph thinking*.

Las bases de datos en grafo operan sobre datos conectados (*connected data*), entendidos como "datos cuya interpretación y valor requieren que primero comprendamos la forma en la que se relacionan sus elementos constitutivos", datos que suelen requerir, para su comprensión "nombrar y calificar las conexiones entre cosas"².

"Los grafos están por todas partes, en gigantes globales como *Google*, *Twitter*, *Facebook* o *Amazon*, en gigantes cercanos como *BBVA* o *Telefónica*"

Existen varios modelos de datos para estas bases de datos, pero uno de los que tienen más tracción es el denominado *Labeled Property Graph Model*, en el que las relaciones que se establecen entre nodos están "tipadas" o etiquetadas: hay diferentes clases de relaciones (tripleas, si se lee esto desde el ámbito de la web semántica). En este modelo los nodos contienen propiedades (pares de clave-valor), las relaciones tienen una dirección y también pueden tener propiedades. *Neo4j* es una de las tecnologías que tomamos como referencia en este texto.

Un aspecto importante al hablar de estas bases de datos es

que son entornos de gestión de datos, y no solo herramientas de análisis. En el vibrante universo del *big data*, muchas de las herramientas procesan e integran datos de diferentes fuentes, pero no son bases de datos donde se almacenan de forma primaria los datos. Las primeras las podemos denominar OLAP o *graph compute engines*"; las segundas, las que nos interesan aquí, OLTP o *graph databases*, y son el equivalente a las bases de datos relacionales transaccionales, y, como ellas, incorporan las funciones CRUD (*Create, Read, Update, y Delete*). Podemos diferenciar estas bases de datos según incorporen de forma nativa el almacenamiento en forma de grafo y/o el procesamiento. Pero, saltando estos detalles técnicos, nos interesa cualquier base de datos que se comporte como un gestor de grafos.

Las relaciones son "ciudadanos de primera clase" en este modelo, haciendo más intuitivo modelarlas al no usar artificios, como son en SQL las claves foráneas entre tablas. Las relaciones no se infieren, se establecen entre los datos. Por eso estas bases de datos son, desde este punto de vista, más expresivas, coherentes y simples que otras implementaciones en SGBD SQL o NOSQL. Una misma necesidad de gestionar información puede moldearse usando tecnologías SQL u otras, pero los grafos proporcionan las siguientes ventajas:

- Rendimiento: para realizar consultas y cálculos que implican gran cantidad de relaciones, frente a la necesidad de complejas consultas SQL.
- Flexibilidad: permiten evolucionar el diseño conforme se manifiesta la necesidad en un proyecto, de forma que posibilitan que "structure and schema to emerge in tandem with our growing understanding". La incompletitud de comprensión de un problema en sus inicios, y su transformación conforme se aborda, son asumidos como elemento esencial en los grafos que son "aditivos" de forma natural. Empezamos sin saber adónde.
- Agilidad: los datos del grafo pueden evolucionar al mismo compás que el resto de aplicaciones de nuestro ecosistema o de otros actores con los que nos relacionamos.

Las bases de datos en grafo son *schema-free*, y requieren otra aproximación al control y gobernanza de los modelos de datos: menos rígida, más transparente por necesidad. Las metodologías ágiles de desarrollo cuadran bien con los grafos. En los grafos es fácil añadir nodos y nuevas relaciones sin poner en riesgo al conjunto de datos previamente almacenado, ni necesidad de hacer migraciones: los datos originales permanecen intactos mientras el *dataset* crece y evoluciona.

En resumen, en estas bases de datos las relaciones se explicitan, forman parte del núcleo del modelado, y los datos se almacenan de forma nativa como "datos conectados". Así la implementación tiene una fuerte correspondencia con el análisis de dominio que se haya realizado.

Una vez que tenemos relaciones almacenadas, la red formada por ellas puede recorrerse: existen caminos o *paths* que podemos atravesar. Estas bases de datos disponen de lenguajes de consulta que ejecutan las *traversing queries*, aprovechando de forma eficiente las extensas redes de datos conectados a muchos niveles o grados de distancia (digamos, buscar algo de un amigo de un amigo de un amigo). Esta es una razón para que demuestren validez en entornos con datos en red muy dinámicos: redes sociales, sensores, ecosistemas, análisis deportivo, procesamiento del lenguaje natural y, por supuesto, los sistemas de recomendación en sentido amplio.

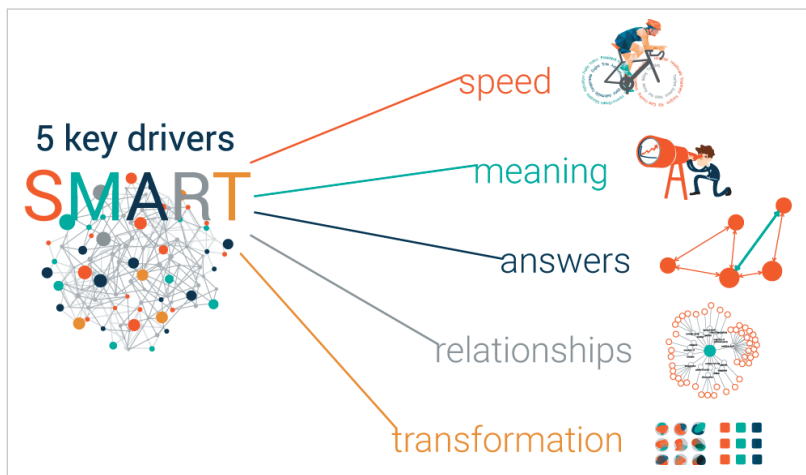


Figura 2. The 5 Key Drivers of Why Graph Databases are Gaining Popularity
Fuente: *OntoText Blog*, 2017.
<https://www.ontotext.com/blog/5-key-drivers-graph-databases-gaining-popularity>

Los grafos son, por naturaleza, multidimensionales, por lo que permiten plantear consultas complejas, que amplían en muchos casos las capacidades de cualquier conocimiento experto, obteniendo análisis muy afinados del tipo "qué sabores de helado han consumido los usuarios a los que les gusta el café *latte* pero aborrecen la coliflor y que compran habitualmente en un barrio determinado". En estas consultas por patrones (*patterns*) está comprimida la potencia de estos entornos para que extraigamos, con relativa sencillez, información nueva derivada de los datos.

4. Grafos de conocimiento

Entonces ¿es lo mismo un grafo de conocimiento que una base de datos en grafo? Veamos cómo define *El Prado* su propio grafo:

"un sistema de representación del conjunto de sus contenidos y recursos digitales que entiende hechos relacionados con los autores, las obras de arte, sus contenidos, temas, épocas y estilos, así como cualquier objeto potencialmente enlazado con ellos y, especialmente, entiende el modo en que este conjunto de entidades está todo él conectado"³.

El elemento clave es "entender", es decir, la capacidad de que la formalización de los datos y sus relaciones permita a una aplicación tratar correctamente un conjunto de entidades, teniendo en cuenta el sentido con el que han sido definidas y descritas para "colaborar con las personas en sus procesos de interrogación, recuperación de la información y descubrimiento de conocimiento". En el grafo de conocimiento hay una dosis de inteligencia establecida que enriquece las meras relaciones entre elementos y sus etiquetas. Hay clases, dependencias, herencia, y por lo tanto se materializa un entorno de información con una organización que modela el conocimiento de un dominio. Tomando la conocida fórmula de Elaine Svenonius que establece que

"la efectividad de un sistema de acceso a la información depende directamente de la inteligencia puesta en organizarla" (**Svenonius**, 2000),

los conjuntos de datos que forman un grafo de conocimiento poseen, en su estructura o modelo de datos, una inteligencia declarada, activable y observable.

El grafo de conocimiento del *Museo del Prado* se utiliza para "anotar, organizar y presentar la información del museo de modo significativo" y permite la consolidación de información preexistente en diferentes "silos" (sistemas de información especializados), dentro de un sistema unificado, extensible, expresivo e interrogable, que vincula el conjunto de entidades que conforman la colección como nodo central con el resto de objetos de conocimiento que componen la actividad museográfica. Supone más de 20.000 recursos, 310.000 entidades y unas 2.830.000 relaciones.

Podemos tomar como definición de grafo de conocimiento (*knowledge graph*) la de aquel sistema que "acquires and integrates information into an ontology and applies a reasoner to derive new knowledge" (**Ehrlinger; Wöß**, 2016), donde vemos que aparecen los conceptos de integración, ontología y razonamiento.

Hemos de entender la integración como la capacidad de incorporar información de diferentes fuentes de datos –estructurados y semiestructurados–, por lo que se entiende que el grafo está construido *on top* de otras bases de datos, que al integrarse se enriquecen e incrementan su significatividad. Este incremento de significado procede de la ontología en la que se definen entidades y propiedades, se tipifican, se clasifican y se modela el conocimiento de dominio. De esta forma un sistema entiende que una entidad del tipo futbolista es un ser humano, y que está vinculado con el deporte fútbol, que se juega con una pelota y donde hay una serie de competiciones. Y, una vez modelado un dominio, otro grafo tendrá la posibilidad de conectarlo con otro modelo de dominio, combinando, por ejemplo, el universo del fútbol con el universo del cine. La formalización en una ontología permite establecer además una serie de inferencias; siguiendo con el ejemplo, un futbolista, dado que es un ser humano, tiene que tener una fecha de nacimiento, tener un idioma nativo y sería muy raro que pudiera vivir más de 130 años. *GraphDB* es una tecnología de *Ontotext* para gestionar grafos semánticos⁴.

El conocimiento deriva no de la forma de grafo –que facilita la flexibilidad y explorar relaciones– sino de la ontología. El modelado de datos incluye el contexto, porque la semántica está definida explícitamente: es smart, es adaptable, es flexible. En el caso del *Museo del Prado* su modelo ontológico deriva del modelo de referencia *CIDOC Conceptual Reference Model (CRM)* de los museos y de los FRBR del universo bibliográfico, y se denomina *Modelo Semántico Digital del Prado*⁵. Las ontologías son formalizaciones que definen de forma estricta y sin ambigüedades los conceptos que pretenden manejar. Dado que, siguiendo a David Weinberger, la realidad es miscelánea (**Weinberger**, 2007) –cuando no, directamente, confusa y conflictiva– las ontologías funcionan habitualmente mediante implementaciones híbridas, en la que se combinan elementos de varias. Este es el caso de *El Museo del Prado*, o de otros

modelos de datos o perfiles de aplicación, como *Europeana Data Model* y todo aquel que se enfrente, de un modo u otro, a un contexto de recursos compartidos por varios agentes con intereses variables. Unir dominios es un aspecto esencial de las ontologías y los grafos: las mismas entidades pueden ser descritas conforme a modelos de dominio diferentes, o niveles de agregación en un mismo dominio, y podemos trabajar con "particiones" de un grafo ajustadas al contexto semántico que necesitemos, donde solo tendremos las relaciones y propiedades adecuadas al contexto. Y, en el caso que nos traemos entre manos, la formalización de la ontología se realiza mediante OWL, el estándar del W3C para definir ontologías en la web de datos, aunque existan también otras formas alternativas de hacer más semántico un grafo (Barrasa, 2017).

5. Entonces, ¿esto es web semántica?

Pues lo es y no lo es. El proyecto del *Prado Digital* hace uso de los estándares de la web semántica y *linked open data*, pero aún no es un conjunto de datos publicado en abierto. No puede descargarse ni consultarse mediante una API pública. Es un buen ejemplo de que los estándares de interoperabilidad también son útiles dentro de los complejos y diversos sistemas de información de una organización, donde conviven sistemas con diferente origen y propósito, que solo un milagro puede integrarlos –temporalmente– en un único sistema, pero el cual puede conectarse o consolidarse mediante las tecnologías de datos abiertos y enlazados.

Los *RDF triplestores* son otro miembro más en la familia de las bases de datos en grafo, basado en datos abiertos, ontologías ligeras o completas, datos enlazados y sobre los que existe un lenguaje de interrogación normalizado denominado *Sparql*. Para no alargar esta nota, solo señalar que es la orientación hacia la publicación, y no hacia la gestión, lo que caracteriza a las bases de datos de tripletas RDF frente a otras bases de datos en grafo (Anadiotis, 2017). Como referencia podemos tomar el consorcio LDDB (*The Graph & RDF benchmark reference*) que aborda ambas tecnologías para facilitar un mercado más eficiente.

<http://lddbouncil.org/public/why-graph>

Los grafos de conocimiento están también en uno de los muchos cerebros conectados que hacen posible la búsqueda de *Google*. El *Google Knowledge Graph* se alimenta de información recolectada de la Web, sirviendo como base para responder preguntas, afinar la relevancia, validar contenidos, combinar conceptos, interpretar la realidad y alimentar a las máquinas. El de *Google* es una caja negra, propiedad corporativa, pero *Wikidata* está ahí como *knowledge base* abierta, colaborativa, multidominio y con un ritmo de crecimiento intenso. Y no solo eso, sino que ya aparecen señales de alerta relevantes de que el software con el que está construida, *Wikibase*, puede convertirse en una plataforma accesible para gestionar y publicar datos semánticos abiertos para los LAM (bibliotecas, archivos y museos) (Godby et al., 2019). Si queremos ser expertos en SEO, tenemos que movernos en el ámbito de los grafos de conocimiento (Aljal, 2019). Si nos interesan los sistemas de recomendación de filtrado colaborativo, pensemos en grafos; si nos interesa la recomendación centrada en el contenido, pensemos en grafos de conocimiento (Ricci et al., 2011). Y lo mismo podemos decir, y ya lo dice Gartner (Paneta, 2019), si nos movemos entre el *big data*, la analítica de negocio, el marketing digital o la predicción de tendencias.

"Si queremos ser expertos en SEO, tenemos que movernos en el ámbito de los grafos de conocimiento"

Para finalizar, quería reflexionar sobre el viaje del *Museo del Prado*: si en 2012 se anuncia como *Biblioteca Digital* y el microsite *Goya en el Prado* para 2015, es la digitalización de los metadatos de las colecciones la que alimenta la transformación de la web del museo, de forma que el valor central del tratamiento documental se pone al servicio del núcleo de la experiencia digital. Los metadatos son arquitectura de la información de sitios web con enfoque *bottom-up*⁶ (partir de los contenidos y sus metadatos para desplegar el sitio web). Si tienes buenos metadatos, si tienes un grafo de conocimiento, puedes transformar tu web. *Metada matters*, también es diseño web, es gestión de contenidos.

6. Notas

1. El proyecto recibió el reconocimiento a la excelencia técnica y la cantidad y calidad de sus contenidos con dos premios *Webby* (jurado y público) en 2016.
2. Una detallada presentación de las bases de datos en grafo, vinculada a la plataforma *Neo4j*, de donde tomamos las citas de este apartado, es:

Robinson, Ian; Webber, Jim; Eifrem, Emil (2015). *Graph databases: New opportunities for connected data* (2nd Edition). O'Reilly Media. ISBN: 978 1491930892

3. *Museo del Prado*: "El grafo de conocimiento del Museo del Prado".
<https://www.museodelprado.es/grafos-de-conocimiento/el-grafos-de-conocimiento-del-museo-del-prado>
4. *Ontotext*. "What is a Knowledge Graph?".
<https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph>
5. *Museo del Prado*: "Modelo ontológico del Museo del Prado".
<https://www.museodelprado.es/grafos-de-conocimiento/modelo-ontologico>
6. Poco frecuentado concepto aportado por Rosenfeld, derivado de su tan gastado manual sobre arquitectura de la información para sitios web.
Rosenfeld, Louis (2005). *Enterprise information architecture Roadmap* (v2).
http://www.louisrosenfeld.com/home/bloug_archive/images/EIARoadmap2.pdf

7. Referencias

- Aijal, Julian** (2019). "What is a knowledge graph and how does one work? It will do wonders for your SEO strategy". *The next Web*, 11 julio.
<https://thenextweb.com/podium/2019/06/11/what-is-a-knowledge-graph-and-how-does-one-work>
- Alonso-Maturana, Ricardo** (2015) "Las web semánticas de la web: el negocio de construir y explotar grandes grafos de conocimiento especializados". *Red Gnos, Watermelon*, 8 junio.
<http://red.gnos.com/comunidad/watermelon/recursos/las-web-semanticas-de-la-web-el-negocio-de/50e5810a-2cd2-4fc4-a327-653cc0e95236>
- Anadiotis, George** (2017). "Graph databases and RDF: It's a family affair". *ZDNET*, 19 mayo.
<https://www.zdnet.com/article/graph-databases-and-rdf-its-a-family-affair>
- Barrasa, Jesús** (2017). "RDF triple stores vs. labeled property graphs: What's the difference?". *Neo4J blog*, 18 agosto.
<https://neo4j.com/blog/rdf-triple-store-vs-labeled-property-graph-difference>
- Comunidad Baratz* (2017). "El Museo del Prado es el primer museo español que hace público y accesible su archivo histórico". *Comunidad Baratz*, 29 noviembre.
<https://www.comunidadbaratz.com/blog/el-museo-del-prado-es-el-primer-museo-espanol-que-hace-publico-y-accesible-su-archivo-historico>
- Ehrlinger, Lisa; Wöß, Wolfram** (2016). "Towards a definition of knowledge graphs". *Semantics: Posters and demos track*.
<http://ceur-ws.org/Vol-1695/paper4.pdf>
- Godby, Jean; Smith-Yoshimura, Karen; Washburn, Bruce; Davis, Kalan-Knudson; Detling, Karen; Fernsebner, Christine; Eslao, Steven-Folsom; Li, Xiaoli; McGee, Marc; Miller, Karen; Moody, Honor; Tomren, Holly; Thomas, Craig** (2019). *Creating library linked data with Wikibase: Lessons learned from project passage*. Dublin, OH: OCLC Research.
<https://doi.org/10.25333/faq3-ax08>
- Martín-Bravo, Ana; Pantoja-Ferrari, Javier** (2013). "Integración de servicios documentales: la nueva web del Museo del Prado" (Video). En: *II Jornadas sobre de Bibliotecas de Museos*. Ministerio de Educación, Cultura y Deportes (Madrid, noviembre 2013).
<https://www.museodelprado.es/actualidad/multimedia/conferencia-integracion-de-servicios-documentales/56f2e76f-6750-4cbc-83ab-e92f6225cecc>
- My art diary* (2016). "La nueva web del Museo del Prado. Entrevista a Javier Pantoja". *My art diary*, 9 mayo.
<http://myartdiary.com/la-nueva-web-del-museo-del-prado-entrevista-con-javier-pantoja-ferrari/>
- Panetta, Kasey** (2019). "5 trends appear on the Gartner hype cycle for emerging technologies, 2019". *Smarter with Gartner*, 29 agosto.
<https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019>
- Pantoja-Ferrari, Javier** (2016). *Una nueva experiencia: el website del Museo del Prado* (Video), Conferencia, Museo Nacional del Prado, 13 enero.
<https://www.museodelprado.es/actualidad/multimedia/conferencia-una-nueva-experiencia-el-website-del/2ebd0f17-02df-49d1-9003-4f5743b2874b>
- Ricci, Francesco; Rokach, Lior; Shapira, Bracha; Kantor, Paul B.** (eds) (2011). *Recommender systems handbook*. Springer. ISBN: 978 0 387 85820 3
- Svenonius, Elaine** (2000). *The intellectual Foundation of Information Organization*, MIT, p. ix. ISBN: 978 0262512619
- Weinberger, David** (2007). *Everything Is miscellaneous: The power of the new digital disorder*. Times Books. ISBN: 978 0805088113