Big data literario de raíz bibliotecaria: reflexiones sobre infraestructuras de anotación, catalogación, descubrimiento y recomendación de ficción narrativa

Literary big data powered by libraries: reflections on annotation, cataloging, discovery, and recommendation infrastructures for narrative fiction

Tomás Saorín

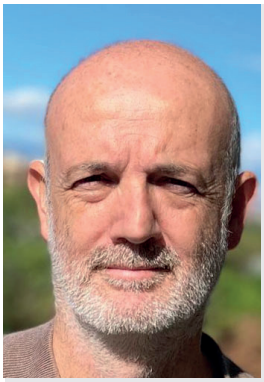
Saorín, Tomás (2020). "*Big data* literario de raíz bibliotecaria: reflexiones sobre infraestructuras de anotación, catalogación, descubrimiento y recomendación de ficción narrativa". *Anuario ThinkEPI*, v. 15, e15c01.

https://doi.org/10.3145/thinkepi.2021.e15c01

Publicado en *IweTel* el 11 de noviembre de 2021

Tomás Saorín
https://orcid.org/0000-0001-9448-0866

Universidad de Murcia
Facultad de Comunicación y Documentación
Campus de Espinardo, Edificio 3
30100 Murcia, España
tsp@um.es



Resumen: Se describe la relación entre el campo de los estudios literarios basados en datos de la corriente *distant reading* y las humanidades digitales, y la actividad de las bibliotecas y otras entidades del sector del libro en el ecosistema de la recomendación y el descubrimiento de lecturas. Se presentan proyectos de catalogación y descripción enriquecida de la ficción literaria, como *OCLC FictionFinder* y *Kirjasampo*, en el marco de los metadatos transmedia y abiertos, entendidos en relación con las prácticas de plataformas de consumo de contenidos digitales como *Netflix* o *Amazon Prime Video*, junto a otras prácticas de anotación y edición de textos literarios. Finalmente se plantea la oportunidad de desarrollo de laboratorios bibliotecarios digitales apoyados en infraestructuras de datos abiertas como *Wikidata* para la descripción enriquecida de ficciones narrativas de todas las épocas de forma colaborativa, para posibilitar proyectos y servicios de descubrimiento de lecturas relacionadas.

Palabras clave: Recomendación lectora; Generificación; Literatura; Metadatos; Universos de ficción; Descubrimiento; Transmedia; Catalogación; *Wikidata*; Ficción narrativa; *Library Labs*.

Abstract: This work explores the relationships between the field of literary studies based on data inspired by the "distant reading" school and the digital humanities and the activity of libraries and other agents of the book sector in the ecosystem of recommendation and discovery of readings. Projects for enriching catalogues and description resources about literary fiction are presented, such as *OCLC FictionFinder* and *Kirjasampo*, within the framework of transmedia and open metadata, understood in relation to the practices of digital content consumption platforms such as *Netflix* or *Amazon Prime Video*. Besides, other practices of annotation and editing of literary texts are outlined. Finally, I explain opportunities to devel-

op digital Library Laboratories supported by open data infrastructures such as *Wikidata* for the enriched description past and present of narrative fictions in a collaborative way, to enable projects and services for the discovery of related readings.

Keywords: Reader's advisory; Genrification; Literature; Metadata; Fictional universes; Discovery; Transmedia; Cataloguing; *Wikidata*; Fiction; Library labs.

Así de pronto, el término *big data literario* tiene gancho. Las grandes transformaciones que definen a los nuevos modelos de servicios de éxito siempre incluyen algo de explotación masiva de datos para producir valor mediante su remezcla e interpretación creativa. Pensar en la posibilidad de que las bibliotecas puedan jugar en un terreno que se llamase *big data literar*io podría sugerir un espacio de oportunidades. Sin embargo, es en el campo de la investigación filológica y literaria donde el uso de grandes volúmenes de datos sobre literatura es una corriente fuertemente asentada, y que se visibiliza con claridad en las investigaciones y proyectos sobre fuentes y análisis de literatura histórica o clásica, que se enmarcan en las comunidades de práctica que se reconocen como humanidades digitales, entendidas esencialmente como metodología de generación e interpretación de datos sobre corpus y textos digitalizados, anotados y marcados.

La actividad de edición crítica literaria está transformándose en el marco de la abundancia de colecciones de textos digitales. *Curation, analysis, editing, and modeling* son componentes esenciales de las humanidades digitales literarias (**Burdick** *et al.*, 2012). Proyectos digitales como los de textos del teatro clásico español, o un corpus anotado de sonetos renacentistas, o un buscador de concordancias, etc. son materia prima de datos para la corriente de estudios literarios de la escuela del *distant reading* (**Moretti**, 2013), que analizan la literatura como sistema a partir del análisis de datos y patrones en textos, paratextos o epitextos, o cualquier otro registro explotable sobre su producción, recepción y ciclo de vida (**Rodríguez**, 2017). Seguir la doble pista de la "lectura distante" y las "humanidades digitales literarias" es de interés para que las bibliotecas se orienten sobre su papel en la relación con la literatura y la lectura en la red.

Lo que nos preguntamos en este trabajo es si podríamos identificar un espacio de algo así como *big data* literario desde el dominio de las bibliotecas, editoriales, librerías, los lectores y la lectura (el mundo del libro y la lectura). Reconozco que he tomado el término "*big data* literario" de una conversación sobre un proyecto de descubrimiento y recomendación de lecturas en la *Biblioteca Regional de Murcia*. Surgió allí y suena bien para mover a reflexión.

Nos preguntamos por la participación de las bibliotecas en la revolución de los datos sobre ficción literaria, caso de que esta se esté produciendo o vaya a producirse. Con las revoluciones no siempre se sabe. Entendamos "literario" como lo relativo a la ficción narrativa, tomando la clásica taxonomía de librería entre ficción y no ficción. Hablaremos de literario y, por lo tanto, de lectura y placer, pero casi todo lo que aquí se diga se aplicaría a cualquier formato que proponga narraciones, tanto el clásico audiovisual y el cada vez más dominante constructor de universos de ficción, el videojuego.

El espacio en el que confluyen ahora mismo humanidades digitales literarias y bibliotecas es el de la digitalización de las colecciones históricas. Tradicionalmente el papel de las bibliotecas ha sido el de proveedoras básicas de bibliotecas o repositorios digitales en acceso abierto con sus fondos patrimoniales descritos y organizados sucintamente, en paralelo al rol de las humanidades digitales, lideradas desde la academia y los grupos o proyectos de investigación, produciendo colecciones y corpus coherentes, anotados y marcados para facilitar su procesamiento analítico. Ya se ha hablado de la necesidad de bibliotecas digitales con más potencia textual (**Rodríguez-Yunta**, 2014), y se busca que estas dos comunidades de práctica confluyan en el manifiesto de Santa Bárbara de 2017 titulado *Collections as data*, que pone el foco en una digitalización profunda de las obras y que facilite o provoque la explotación como datos, de forma que los repositorios sean mayores fuentes de datos sobre las colecciones que alojan (**Candela** *et al.*, 2020). Hay una evidente fuerza impulsora de un *big data* de la literatura histórica, en el que las bibliotecas juegan un papel y en donde encuentran nuevas formas de participar en la cadena de valor. Pero, y esa es la pregunta que provoca este trabajo: ¿puede existir un *big data* para la producción literaria actual?

"Usar la plataforma *Wikidata* para catalogar ficción de forma colaborativa rebaja al máximo las barreras de entrada de las bibliotecas para participar en la web de datos"

"Las colecciones disponibles en bibliotecas y repositorios digitales necesitan mayor tratamiento como textos anotados y marcados para facilitar el *big data* literario"

En realidad, no me gusta el término "*big data* literario", pero sí que creo que nos va a permitir hablar de esos datos que son afines al mundo bibliotecario, los metadatos literarios, desde una perspectiva diferente. Los metadatos son recursos descriptivos creados a partir de las obras originales, y que permiten manejar la producción cultural, acumulativa y en constante renovación, para hacerla accesible a través de plataformas de servicio digital, que podríamos llamar catálogos. Las estanterías de una biblioteca no son más que un catálogo con un formato singular: igual podrían ser fichas o una pantalla. Hablábamos de metadatos, de organización y descripción de recursos de información (**Glushko**, 2020), y en el mundo de la biblioteca y la lectura pública esos recursos los simplificamos en la etiqueta "libros". Y cuando en lugar de libros hablamos en términos de literatura, entendemos que hablamos de libros de poesía, teatro o novela, siendo esta última la dominante en el uso de las bibliotecas públicas. Lo que se presta en las bibliotecas públicas es casi siempre novelas. Novelas, cuentos, fábulas, relatos, ficción, esas palabras. ¿Podrían considerarse los metadatos sobre obras literarias una forma de *big data* literario? Antes parecía que la suma de registros de todos los catálogos combinados de todas las bibliotecas era una cantidad inmensa, pero, usando las unidades de medida *bigdatienses*, es una cantidad pequeña. Los catálogos colectivos no son *big data*, tanto por tamaño como por profundidad de sus datos.

Crear metadatos para libros ha sido una tarea central en el trabajo bibliotecario, aunque el vocabulario clásico nos habla de catalogación y registros bibliográficos. Pero es preferible hablar de "producir metadatos" (**Vukadin**, 2019, pp. 13-16). Metadatos que interpretan y seleccionan aspectos del contenido para facilitar su descubrimiento y conexión. Y Vukadin insiste en que usar metadatos es apostar por los aspectos de apertura y descripciones distribuidas. Los metadatos producen la posibilidad de catálogos, que si permiten acceder al contenido (darle al play o descargar el libro en préstamo, por ejemplo) llamamos plataformas o, ya puestos, bibliotecas digitales. No solo las bibliotecas operan sobre metadatos, sino que también lo hacen *Amazon*, *Netflix* o *Spotify*: no son meras tiendas *online* de zapatillas o mochilas, sino que se enfrentan al reto de organizar contenidos, es decir, objetos que transportan información, discurso, que requieren interpretación temática (*subject analysis*). Catálogo es cualquier reunión de metadatos interconectados y remezclados, enriquecidos por el proveedor de un servicio. ¿Existen los metadatos bibliotecarios específicamente literarios? Adelantaría como respuesta un no; existen metadatos sobre los libros y las ediciones de obras literarias, pero que penetran poco en su contenido y que, incluso a nivel de las obras y ediciones, aún juegan poco las posibilidades del modelo conceptual de referencia para bibliotecas (LRM) para describir obras como una red de relaciones consistente y navegable.

Creo, de nuevo siguiendo la argumentación de Ana Vukadin, que el enfoque de los metadatos literarios creativos será siempre el de *transmedia metadata*, alcanzado a cualquier ficción esencialmente narrativa. La creación de conexiones es la propuesta de valor de los metadatos, y mejorar los metadatos desde una perspectiva transmedia significa

> "contemplar los recursos de información como objetos con un potencial interés o interpretación metadisciplinar".

Todo el mundo puede catalogar cualquier cosa, que es uno de los lemas de la web semántica, expresado como "cualquiera puede decir cualquier cosa sobre cualquier cosa" (*anyone can say anything about anything*). *Netflix*, *Library of Congress*, *IMDb*, *Allmusic*, *Spotify* y el *WorldCat* de *OCLC* trabajan juntas, poniendo en circulación metadatos para que cada cual se monte, lo mejor que pueda, su plataforma de servicio. Hablaremos de bibliotecas, pero mirando hacia los lados en dirección a las plataformas de las industrias de los contenidos digitales, de las que se pueden obtener interesantes lecciones y buenas prácticas.

¿Qué hacen las bibliotecas con los metadatos de obras literarias de ficción?

Poco. El mecanismo de puesta a disposición de sus colecciones se basa en la disponibilidad en las estanterías del 82-3 –a la espera de que venga el usuario y localice la obra que quiere– o en las guías de lectura –una narrativa que reúne y agrupa obras con alguna excusa estacional o temática– o la exposición en punto de venta –en mostradores de novedades–. En el mejor de los casos, sobre todo en literatura infantil, se usan sencillas agrupaciones por géneros o centros de interés. De esta gestión de colecciones en el espacio, en los metadatos del catálogo quedan reflejados, y no siempre, y no con detalle, simplemente aspectos editoriales y de identificación (autor, traductor, editorial) y de género (novela

histórica, libros de ciencia ficción, novela romántica, etc.). La "genrification" (generificación) es un lento movimiento para organizar la literatura mediante esquemas más cercanos de las necesidades del público lector, y caracteriza a una bienintencionada comunidad de bibliotecarios e investigadores que desean repensar la práctica de la catalogación de la ficción, que no es un asunto menor (**Ward**; **Saarti**, 2018). Recomendaciones y servicios para transformar la organización de la colección hacia géneros son sin duda útiles, como las de la consultora *Follet Learning* (2019). La otra gran forma de recomendar lecturas desde las bibliotecas es social y experiencia: actividades, talleres, encuentros, clubes de lectura… de los que surge, mediante la socialización y de forma casi natural, el deseo de descubrir a ciertos autores, temas o estilos. Podríamos decir que el catálogo no es, ni mucho menos, uno de los canales preferentes para el descubrimiento de sugerencias de lecturas, sino tan solo el que permite localizar si está disponible la obra en la que ya estamos interesados.

"Las conexiones entre la ficción tienen una naturaleza transmedia y son datos que desea la industria de los contenidos digitales para mejorar su recomendación"

Estas prácticas son, en esencia, las mismas que hacen las librerías: estanterías, encuentros y géneros. La clasificación internacional IBIC para el mercado del libro llega hasta ese nivel de detalle de género para la ficción: novelas de aventuras y acción. Bien es cierto, que, en el caso de la industria editorial, se han puesto en marcha algunos proyectos de creación de catálogos de recomendaciones más detallados para facilitar el desarrollo de colecciones según público y temática. Pero, a mi juicio, su impacto es insignificante, tanto en los lectores como en la Red. Quizá estas fuentes de recomendación colectiva de editoriales y librerías, con respiración asistida de la financiación pública, puedan tener algún efecto en los prescriptores, esencialmente bibliotecarios responsables de compras y educadoras construyendo su biblioteca o plan lector escolar. Si hablamos de impacto en la red –visibilidad en buscadores o interconexión de datos bibliotecarios o editoriales sobre literatura actual en servicios y aplicaciones masivas– tampoco podríamos entusiasmarnos mucho, hay poca presencia perceptible de los catálogos colectivos de bibliotecas en la experiencia de búsqueda y uso de la web. Pese a las atractivas aventuras innovadoras digitales en la "recomendación editorial", muchas de ellas tienen una vida corta y no alcanzan a trascender (**Cordón-García**, 2018). El que busca literatura pronto es canibalizado por el catálogo de *Amazon* y otras grandes corporaciones de distribución del libro, por plataformas sociales construidas por lectores o por el contenido en estado semisalvaje disperso por infinidad de revistas de reseñas, magazines culturales o blogs amateurs. No tenemos aquí tiempo para hablar de las plataformas sociales de recomendación y etiquetado creadas por lectores, pero su atrevimiento, fluidez y concentración de contenidos y usuarios las hacen necesarias para renovar el discurso bibliotecario (**Cordón-García**; **Gómez-Díaz**, 2018). Podemos entender cualquier sistema de organización de lecturas literarias a partir de los metadatos en los que se concreta y qué se hace con ellos. En relación con la ficción hay dos casos bibliotecarios que nos pueden poner sobre la pista de las sendas a explorar:

- el experimento de *OCLC*: *FictionFinder*;
  https://www.oclc.org/research/areas/data-science/fictionfinder.html
- el sólido proyecto *Kirjasampo* de las bibliotecas finlandesas.
  https://www.kirjasampo.fi/

FictionFinder es un prototipo para manejar la ficción en *WorldCat*, aplicando el modelo conceptual FRBR a la organización de los elementos del catálogo, para agrupar las obras, expresiones, manifestaciones y ejemplares de forma más consistente. Además, profundiza en los metadatos descriptivos que tratan la ficción como contenido, incorporando además de géneros, el nivel de la audiencia y palabras clave para personajes, temas y lugares.

El proyecto *Kirjasampo,* llevado a cabo por el sistema finlandés de bibliotecas, supone, en sus propias palabras, *rethinking metadata* (**Hypén**, 2014). Este replantearse los metadatos se aborda desde los presupuestos de la web semántica y el uso de una ontología específica para la ficción conectada a una ontología genérica, que permite manejar la "descripción rica y diversa de la ficción": géneros, personajes, épocas de ambientación, lugares de ambientación, temas, motivos y eventos. Se trata de un esfuerzo colaborativo por describir con profundidad la ficción en lengua finesa, para poder sistematizar el conocimiento informal poseído por lectores y bibliotecarios y que permite responder, con cierto nivel de aproximación, a preguntas del tipo

> "¿En qué novela había una familia sobreviviendo en un escenario post-apocalíptico? ¿Me acuerdo de una novela de un detective que investiga el asesinato de un transexual?".

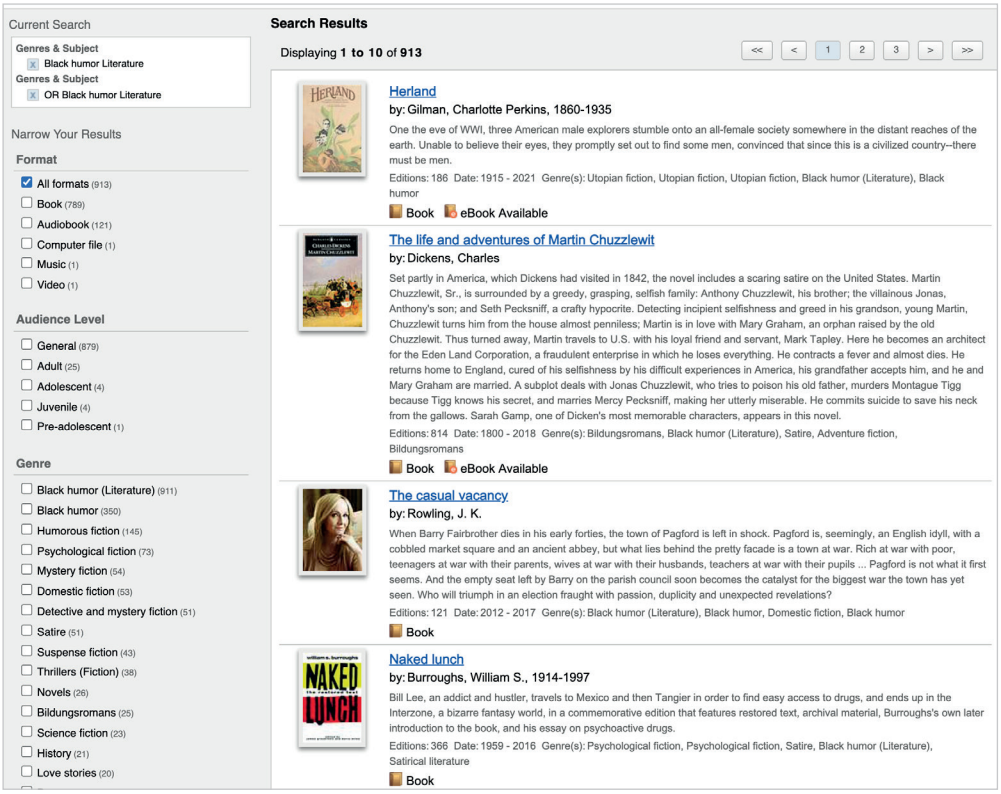


Figura 1. Facetas de filtrado por formato, nivel de la audiencia y género al buscar en el prototipo de *OCLC FictionFinder*.
http://experimental.worldcat.org/xfinder/fictionfinder.html

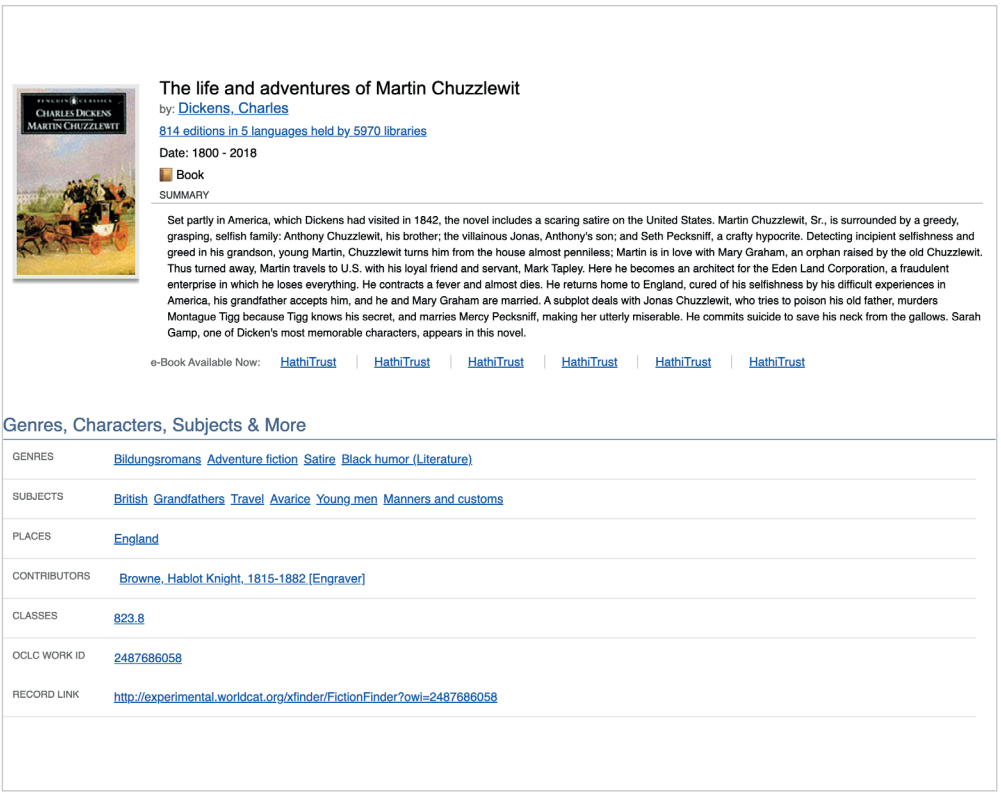


Figura 2. Registro de prototipo de *OCLC FictionFinder* que muestra metadatos sobre géneros, temas y lugares en una novela de humor negro.
http://experimental.worldcat.org/xfinder/fictionfinder.html

La confluencia entre los trabajos de Jarmo Saarti para definir un modelo para describir la ficción y la viabilidad del conjunto de tecnologías de *linked open data*, junto con la disponibilidad de infraestructuras nacionales para la publicación semántica de datos y una acción conjunta por parte de los sistemas de bibliotecas, permitieron poner en marcha un servicio pionero en la Web, que permite explorar y descubrir su literatura nacional –incluyendo las traducciones– de formas sugestivas.

¿Son, entonces, estos dos ejemplos *big data* literario? Desde luego, van en la buena dirección. Metadatos más profundos sobre la ficción, elaborados con una propuesta de valor clara para establecer conexiones entre contenidos de ficción y publicados como datos en la Web, alrededor de una URI para poder funcionar como datos enlazados. Sin embargo, no tienen suficiente volumen, diversidad y continuidad para llamarlos *big data*. Solo hay datos sobre las obras –contenido– y no tienen datos de uso –compras, visualizaciones– que son los que supervitaminan a las grandes plataformas como *Amazon* o *Netflix* y sus algoritmos de filtrado colaborativo. Yo, si tuviera que inventar una etiqueta, hablaría de que *OCLC* y *Kirjasampo* están produciendo datos muy valiosos que podríamos denominar *deeper data* o *Deep fiction metadata*. Pero, aunque no alcancen *per se* la categoría de *big data*, sí pueden formar parte de cadenas de valor de *big data* en la Red, al ser datos de naturaleza interoperable y conectiva (*linked open data*). ¿Por qué? Porque todos los complejos motores de recomendación de contenidos se alimentan de una endiablada combinación de fuentes de información variada, en la que también juegan un papel importante los metadatos sobre las obras. Los temas, los estilos, los lugares, las referencias cruzadas, las versiones, los autores, los grupos artísticos, las escuelas y tendencias, los premios o los personajes son combustible para los mecanismos de recomendación. Y se trata de metadatos descriptivos que requieren fuentes de calidad, sistemáticas y con credibilidad. Y aquí es donde entra en juego el campo bibliotecario convencional: si hay buenas fuentes pueden conectarse y remezclarse desde las industrias creativas. Tomemos el ejemplo del servicio *Prime Video* de *Amazon* y su funcionalidad *X-Ray* que nos ofrece una capa de metadatos integrada en la experiencia de navegación y uso: calificación, etiquetas, escenas, personajes, o música de fondo. Sus contenidos audiovisuales son catalogados mediante una cuádruple acción combinada de, por un lado, analítica automática, por otro, equipos de etiquetadores-catalogadores de la compañía, apoyados en los metadatos proporcionados en origen por los propios productores y, finalmente, por los datos de la gran base de datos web *IMDb* sobre audiovisual (propiedad de la compañía).

El *big data* de *Amazon* se apoya en un catálogo externo de formato tradicional, que podría ser un espejo en el que se mirasen las bibliotecas. Hay muchas lecciones a aprender de *IMDb*. La primera se infiere del escenario descrito anteriormente: la explotación de los metadatos será más habitual que la haga una plataforma externa. Las bibliotecas pueden producir metadatos, que al ser datos abiertos y enlazados podrán alimentar propuestas de valor y servicio de otros. Las bibliotecas pueden hacer una explotación básica, pero dejar abierta la puerta a complicidades para la innovación creativa y la colaboración de otros muchos actores en el ecosistema digital. En segundo lugar, podemos encontrar en *IMDb* otro ejemplo de catalogación profunda del contenido audiovisual de ficción: además de los esperables metadatos sobre intérpretes, directores, fechas y estudios, encontramos etiquetado de análisis de contenido muy arriesgado y sugerente, las *Plot keywords*. Una *folksonomía* sobre aspectos de la obra. Por ejemplo, 68 términos para la película Ana Karenina, entre ellos: *dance*, *imperial russia*, *character name as title*, *anna karenina character*, *saint petersburg russia*, *sex scene*, *russian empire*, *marriage*, *train station*, *train*, *year 1874*, *russian literature*, *married woman*, *unfaithful wife*, *etc.*

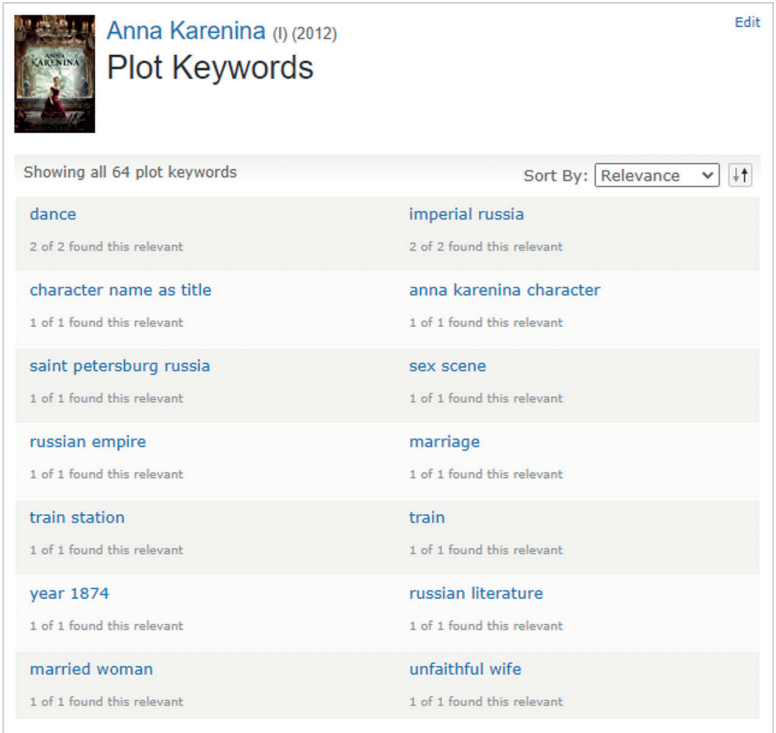


Figura 3. *Plot Keywords* de la película *Ana Karenina* de 2012.

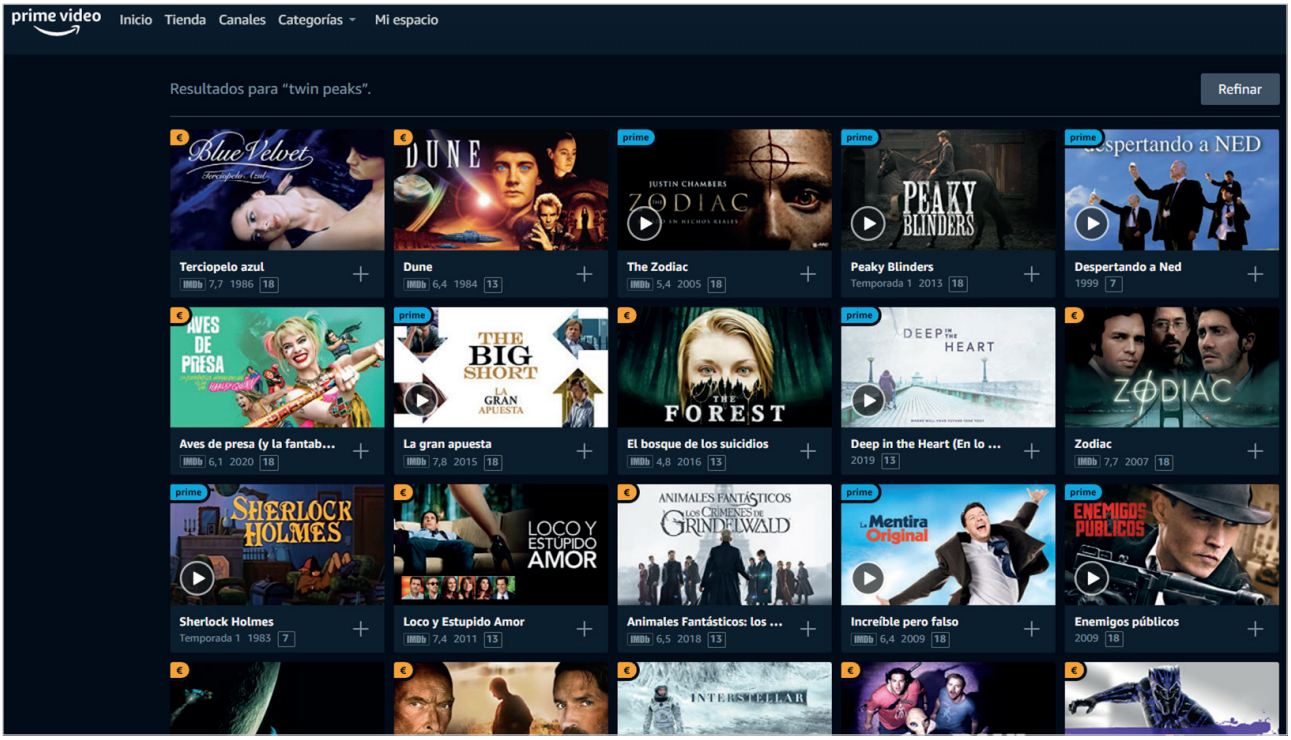


Figura 4. Resultados al buscar *Twin peaks*, película no disponible en *Prime Video*, pero de la que sí sabe que tiene que ver con David Lynch o algo de su argumento.

Otro uso que hace *Amazon Video* de tener a su servicio una base de datos con todo el cine del mundo es que al buscar en su plataforma una película, aunque no esté disponible en su colección, nos pueda ofrecer resultados relacionados. Dispone del índice de *IMDb* que tiene mayor cobertura que su limitado fondo, por lo que puede rastrear cualquier obra por la que se le pregunte sugiriendo otras con algún rasgo común. Si tuviera dentro también el catálogo de una biblioteca, podríamos preguntarle por un libro y nos recomendaría una serie. En la necesaria lectura del informe de *OCLC* sobre el futuro de los metadatos o la catalogación (**Smith-Yoshimura**, 2020), podemos incorporar este punto de vista sobre la combinación de catálogos, como oportunidad en el mundo de los servicios de lectura pública para producir un espacio común de información de mejor calidad y con mayor "curación de contenidos" y probabilidades de germinar. Aquí se aplica el lema de *metadata as a service*: metadatos como una fuente de datos de valor para otras aplicaciones de terceros.

Y existen inacabables conexiones entre todas las obras narrativas, de cualquier tipo, que forman una densa malla de referencias, que incluye además a la realidad. *Amazon* podría responder a búsquedas del tipo "series relacionadas con grupos terroristas en Europa", si pudiera identificar, desde una base de conocimiento externa, cuáles son los grupos terroristas que han existido en el continente y si tuviera etiquetadas las películas con el nombre de esos grupos. Para desentrañar todas las relaciones posibles, necesitaría conocer el mundo real o tener mapas de él. Necesitará, igual que los asistentes conversacionales *Alexa* o *Google Now*, conectar con grafos de conocimiento que ordenen y conecten cualquier faceta de la realidad.

El ejemplo más sugerente de un catálogo detallado de todo lo que contienen los libros lo encontramos en la empresa *Small Demons* (**Robertson**, 2013), que atrajo cierta atención en el sector alrededor del año 2013, y que posteriormente parece que su tecnología y *know-how* hayan sido absorbidas por alguna de las grandes compañías del contenido y recomendación. Se trataba de una base de datos generada mediante técnicas de minería de textos, para identificar cualquier referencia a objetos, personajes, lugares, música, obras, temas, acontecimientos, y establecer, a través de ellas, relaciones entre obras. Su concepto de un *storyverse*, sería el delirio de un lector/espectador curioso y memorioso.

También circularon reseñas sobre el servicio *X-Ray* para libros, mediante el cual *Amazon* podía usar el *big data* recogido de los lectores conectados a sus entornos de lectura en *Kindle*, para recopilar las frases más subrayadas, patrones de lectura y otros datos de *tracking* masivo. Se trata de capacidades que el gigante del comercio electrónico guarda en la recámara, y que aún no han explotado como servicio de verdadero valor añadido.

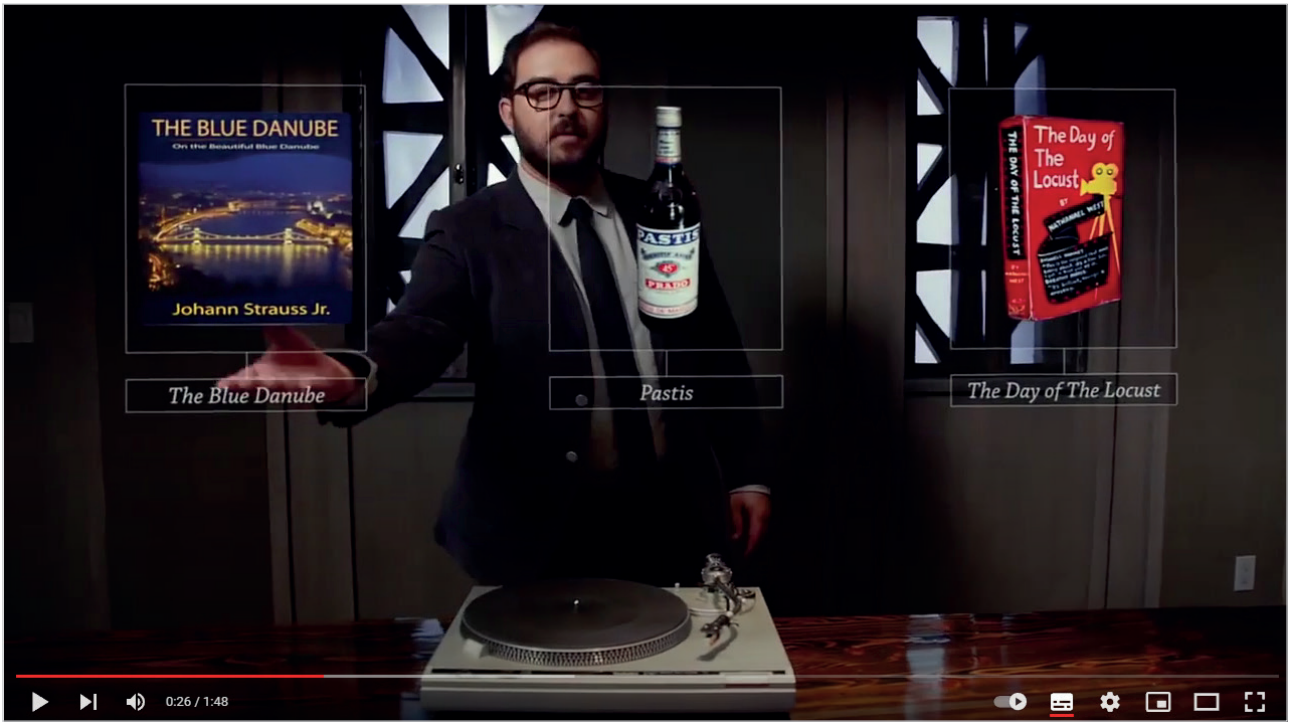


Ilustración: Fotograma del video promocional de la *startup Small Demons* (Junio 2014). Disponible en *https://www.youtube.com/watch?v=XOoOUBkdUMc*.

Entonces, ¿podríamos identificar un *big data* literario disponible en el ámbito bibliotecario, igual que lo hay en el ámbito audiovisual o de la música? ¿Es algo que afecta al sector de la lectura? La primera parte de la respuesta ha de ser prudente. Las plataformas gigantes de música y vídeo disponen de verdadero *big data*, resultado sobre todo de sus millones de usuarios que de forma masiva producen datos de uso, y que permiten el filtrado colaborativo que es, hasta cierto punto, autónomo de la necesidad de saber algo sobre el contenido. *Amazon*, *Spotify* o *Netflix* pueden recomendar sabiendo muy poco de lo que tratan sus contenidos, puesto que se apoyan en el comportamiento de una comunidad brutal de usuarios monitorizados y con una pauta de consumo continuo e inagotable. Tienen datos de uso concentrados en su plataforma, y departamentos de analítica e ingeniería para convertirlos en servicio. Las bibliotecas no tienen ni lo uno ni lo otro. Pocos y reticentes datos de uso y tracking. Pocos recursos de ingeniería e innovación. Los *Media Labs* (*ALA*, 2014), *Glam Labs* (**Mahey** *et al.*, 2019) o *Library Labs* (**Phetteplace**; **Brooks**; **Heller**, 2013) como los de la *British Library*, la *Biblioteca Virtual Cervantes* o la *BNE* son una excepción, y son necesarios para reconectar, desde las bibliotecas, con la información digital en la Red. Me gustaría explotar, en otro momento, la analogía entre los resultados y narrativas que conocemos del "periodismo de datos", con los que podrían surgir desde una "catalogación de literatura como datos" en las bibliotecas.

Hay un gran espacio pendiente de desarrollar relacionado con la descripción detallada y curación como datos del contenido literario y de los universos de ficción. Sobre *La Celestina, Hamlet*, *La regenta*, *Harry Potter*, *Grandes esperanzas* o *Los episodios nacionales* se ha escrito mucho desde los estudios literarios. Sus universos de ficción, tramas, personajes, referencias, lenguaje están densamente cartografiados, pero mediante procedimientos discursivos. Mediante publicaciones. Faltan datos conectados y activables, datos que representen esos universos de ficción y permitan incorporarlos a búsquedas y recomendaciones en entornos de formación del gusto lector y no meramente en el académico y de investigación. Lo mismo podríamos decir de la narrativa actual, de la cual tenemos reseñas y análisis, pero tampoco está disponible como datos. Y esos datos están ahí esperando que alguien los sistematice, y las bibliotecas pueden jugar ahí un papel directo e indirecto. Los datos sobre el interior de la ficción son de un alto interés para la industria de los contenidos digitales. Saber más sobre esto parece que interesa a *Amazon* o a *Netflix* (**Dye** *et al.*, 2020), y no solo de la ficción audiovisual, sino de una forma global: todas las ficciones están unidas, las obras se relacionan, versionan e influyen unas en otras.

Para conocer más sobre tramas y temas contamos ya con mucho contenido, pero poco *datificado*. Es decir, tenemos publicaciones/monografías, que proporcionan valiosísimos recursos para interpretar y descubrir el cine, la literatura, el cómic o los videojuegos. Estudios, guías, recopilaciones didácticas… narrativas para explicar, valorar, relacionar, organizar y entender la literatura. Pero carecemos de datos, de catálogos, de bases de conocimiento utilizables para incorporar ese conocimiento al medio digital. Una

búsqueda web sobre películas ambientadas en Salamanca nos dará seguro algunos resultados textuales, pero no un listado previsible y procesable de datos. La literatura, la literatura clásica, está anotada. Las ediciones críticas y anotadas son la capa de "realidad aumentada" que tradicionalmente se ha añadido a los textos. Son fuente de vínculos, conexiones e interpretación, necesarias para mantener el contexto y permitir una lectura informada a lo largo del tiempo. ¿No sería maravilloso poder recuperar de golpe todos esos miles de notas que se han hecho sobre las páginas de *El Quijote*? Actualmente estas búsquedas no son posibles, y sin embargo son tan importantes como las capas de datos sobre los mapas de *Google* y el geoposicionamiento, porque no todo está sobre el territorio, y nuestro suelo son las historias y las narraciones. Desde las humanidades digitales se producen anotaciones digitales sobre textos digitales que son realidad aumentada, porque los textos son realidad. La anotación de textos quizá ha sido abordada solo desde el ámbito erudito de la crítica textual, pero tiene una dimensión de práctica social muy relevante, sobre todo cuando puede ser compartida y en formato digital, y es un negocio académico, educativo y, ¡por qué no!, bibliotecario. Sobre este campo nos han brindado una brillante monografía **Kalir** y **Garcia** (2021) en la colección *MIT Essentials*, titulado simplemente *Annotation*. Un universo accesible y estructurado de todas las anotaciones sobre obras literarias sí serían, desde luego, un *big data* literario valiosísimo. Pero este no es el terreno de juego de las bibliotecas. En otro escenario, el de los gigantescos universos de fantasía de las grandes franquicias, también se encuentra la necesidad de elaborar catálogos para sostenerlos y garantizar su coherencia. Se trata de complejas bases de datos que recogen cada detalle de personajes, arcos narrativos, objetos, para ordenar la producción de contenido transmedia, como el *Holocron* del universo *Star Wars* (**Castillo**, 2017). Estos universos también motivan a su comunidad de fans a recopilar, de forma colaborativa y autónoma, datos sobre ellos y crear wikis o comunidades *online*. La *fan-fiction* produce especies singulares de *fan-catalogs* que no hay que perder de vista, porque trazan mapas de lectura y descubrimiento.

¿De qué podríamos hablar si aceptamos el término "*big data* literario bibliotecario"?

Pues, a mi modo de entender, de una competición y colaboración por movilizar metadatos descriptivos del contenido de la literatura, que para la narrativa tiene algunas singularidades. Y estamos hablando de metadatos, porque es algo diferente del continuo flujo de reseñas, promociones, análisis y contenido digital en forma discursiva que inunda la Web, a través de los medios generalistas, medios especializados, blogs colectivos y personales, páginas de las casas editoriales y redes sociales. Metadatos frente a discurso. Bibliotecas frente a revistas. Listados frente a artículos.

El modo convencional de recomendación y descubrimiento de lecturas desde las bibliotecas es el manual y cuerpo a cuerpo. Los sugerentes manuales de *ALA* de *Readers' advisory service in the public library* se centran en la creación de espacios en la biblioteca con selecciones de libros, en la interacción con usuarios y en la producción de guías de lectura temáticas, estacionales. Usan un enfoque discursivo y relacional, que bien podría valer para el entorno del aula, donde el profesor es un prescriptor humano y conversacional. De grandísimo valor, pero escaso alcance en la Red, donde los datos son los únicos que pueden participar en la posición dominante de la recomendación y la búsqueda como patrón masivo de comportamiento. El fenómeno asociado al auge del *big data* de "desaparición del experto" (**Mayer-Schönberger**; **Cukier**, 2013) también está hablando de la necesidad de competir en el terreno de los datos. No podemos competir solo cuerpo a cuerpo con la Red, no vale la guerrilla clásica, hace falta "guerrilla de datos".

¿Se puede catalogar la ficción, pueden hacerlo las bibliotecas?

Por supuesto que pueden, pero no solas, ni a la manera actual. Y esto es, además, uno de los retos que tenemos por delante para seguir aportando valor a la cadena de consumo, acceso y descubrimiento de nuevas lecturas. Para que las bibliotecas puedan participar con impacto masivo en el descubrimiento digital de lecturas de ficción tienen que catalogar los universos de ficción. Y hacerlo de determinada manera, conforme a estas pautas que me atrevo a recomendar:

- Catalogar en una plataforma no bibliotecaria abierta como *Wikidata*, la *knowledge-base* conectada a los artículos de *Wikipedia*.
- Describir elementos del mundo real presentes en la ficción, tales como lugares y épocas de ambientación, sucesos históricos que suceden dentro de la trama, personajes históricos protagonistas, obras mencionadas.
- Describir y conectar la literatura con temas y focos de interés, para lecturas sobre temas de interés: autismo, divorcio, duelo, terrorismo, etc., puesto que la ficción es una potente forma de conocimiento.
- Describir la ficción con una orientación transmedia, conectando cine, series, novelas, cuentos, teatro, cómic, videojuegos.

- Trabajar de forma de colaborativa y asimétrica entre ellas, desarrollando proyectos temáticos en redes de bibliotecas, y creando espacios para compartir resultados.
- Dinamizar la participación de otras entidades del ámbito de la cultura, la educación y la lectura, en esta catalogación de la ficción. Implicar a clubes de lectura, asociaciones culturales, grupos de interés, editoriales, librerías, etc.
- Incentivar el desarrollo de prototipos de explotación de estos datos para alimentar plataformas de recomendación y sugerencia basadas en datos. Y que estos prototipos vayan cogiendo forma en diversas plataformas digitales locales de visibilización de conexiones narrativas.
- Volver a disfrutar del placer de leer, conectar y anotar, creando actividades y experiencias que englobo bajo la etiqueta "LiteraDATA": actividades de socialización de la descripción y enriquecimiento de datos sobre obras literarias y de ficción.

"Los gigantes del contenido digital, como *Netflix* o *Amazon*, nos están dando lecciones de catalogación, enriquecimiento de metadatos y entornos de descubrimiento y recomendación"

En los detalles está el peligro. Aunque no es objeto de este trabajo detenerse en el cómo, sí hemos de plantear que la plataforma *Wikidata* se adivina como el lugar sobre el que aportar datos. *ARL* (2019) y *OCLC* ya han dedicado, tanto a *Wikidata* como a su software *Wikibase*, interesantes informes que la señalan como una infraestructura con capacidad transformadora. El uso de esta potente plataforma de datos enlazados rebaja las barreras tecnológicas de entrada, y puede funcionar como infraestructura de datos a todos los niveles (**Allison-Cassin**; **Scott**, 2018). Dan Scott lo expresa con agudeza en esta reflexión al hilo de la catalogación de grupos de música local:

> "rather than focusing on directly enhancing our own local data repository silos (for example, library catalogues, digital exhibits), libraries and archives should invest their limited resources in enriching Wikidata, a centralized data repository, to maximize the visibility of those entities and the reusability of that data in the world at large… and then pull that data back into our local repositories to enrich our displays and integration with the broader world of data" (Scott, 2017)
>
> "En lugar de centrarse en mejorar directamente nuestros propios silos de repositorios de datos locales (por ejemplo, catálogos de bibliotecas, exposiciones digitales), las bibliotecas y los archivos deberían invertir sus recursos limitados en enriquecer *Wikidata*, un repositorio de datos centralizado, para maximizar la visibilidad de esas entidades y la reutilización de esos datos en el mundo en general… y luego recuperar esos datos en nuestros repositorios locales para enriquecer nuestras pantallas y la integración con el mundo más amplio de los datos" (**Scott**, 2017).

Este es un tema estratégico para la supervivencia de las bibliotecas como actores relevantes en la batalla por la recomendación y descubrimiento de lecturas en la sociedad digital; las formas en las que se busca información en la Red han sido trastornadas, y los catálogos actuales tienden a la irrelevancia. Se requieren experiencias piloto, modelos de trabajo, capacidad de atraer la atención de otros agentes del mundo del libro y la lectura, dinámicas participativas y exploración de oportunidades.

¿Puede entenderse esto como un *big data* literario con denominación de origen bibliotecaria?

Un poco, pero sería más ajustado al escenario real pensar en *smart fiction metadata*, entendido como pieza de valor en el extenso universo de la recomendación de consumo de contenidos, literarios y no literarios. La lectura actual son muchas prácticas lectoras transmedia fluidas. Las bibliotecas pueden seguir catalogando la narrativa, pero con otros parámetros que produzcan mucha mayor capacidad de conectividad y que tengan al trabajo en colaboración abierta en su producción y explotación como eje posibilitador y, sobre todo, en plataformas abiertas que hagan que los datos puedan estar allí donde los usuarios están; en cualquier sitio de la Web, consumidas desde cualquier agente recomendador y prescriptor. Que puedan obtener beneficio de todo esto desde *eBiblio* hasta el temible *Amazon*, desde la *Red de Bibliotecas de La Coruña* a una app divertida pendiente aún de inventar.

Referencias

ALA (2014). *Digital media labs in libraries. Library Technology, Reports*, n. 6, August/September. American Library Association. *https://doi.org/10.5860/ltr.50n6*

Allison-Cassin, Stacy; **Scott, Dan** (2018). "Wikidata: a platform for your library's linked open data". *The code4Lib journal*, n. 40.
https://journal.code4lib.org/articles/13424

ARL (2019). *ARL White Paper on Wikidata: Opportunities and Recommendations*. Association of Research Libraries. *https://www.arl.org/resources/arl-whitepaper-on-wikidata*

Burdick, Anne; **Drucker, Johanna**; **Lunenfeld, Peter**; **Presner, Todd**; **Jeffrey, Schnapp** (2012). *Digital_humanities*. MIT. ISBN: 978 0262018470

Candela, Gustavo; **Sáez-Fernández, María-Dolores**; **Escobar-Esteban, María-Pilar**; **Marco-Such, Manuel** (2020). "Reusing digital collections from GLAM institutions". *Journal of information science*, first online, August. *https://doi.org/10.1177/0165551520950246*

Castillo, Diana (2017). "Be mindful of the future: information and knowledge management in Star Wars tie-in fiction". *Dalhousie journal of interdisciplinary management*, v. 13. *https://ojs.library.dal.ca/djim/article/view/6925*

Cordón-García, José-Antonio (2018). "Los sistemas de recomendación en el contexto editorial". En: Cordón-García, José-Antonio; Gómez-Díaz, Raquel (coords.). *Lectura, sociedad y redes: Colaboración, visibilidad y recomendación en el ecosistema del libro*. Madrid: Marcial Pons, pp. 187-234. ISBN: 978 84 9123 577 4

Cordón-García, José-Antonio; **Gómez-Díaz, Raquel** (coords.) (2018). *Lectura, sociedad y redes: Colaboración, visibilidad y recomendación en el ecosistema del libro*. Madrid: Marcial Pons, pp. 187-234. ISBN: 978 84 9123 577 4

Dye, Melody; **Ekanadham, Chaitanya**; **Saluja, Avneesh**; **Rastogi, Ashish** (2020). "Supporting content decision makers with machine learning". *Netflix technology blog*, 10 diciembre. *https://netflixtechblog.com/supporting-content-decision-makers-with-machine-learning-995b7b76006f*

Glushko, Robert J. (ed.) (2020). *Organización y descripción de recursos de información digital*. Madrid: Fesabid. *http://www.fesabid.org/wp-content/uploads/2021/05/organizacion_y_descripcion_de_recursos_informacion_digital_-_tdo_lite_esp_v_edit_-_fesabid_2020.pdf*

Kalir, Remi H.; **Garcia, Antero** (2021). *Annotation*. MIT. ISBN: 978 0262539920

Hypén, Kaisa (2014). "Kirjasampo: Rethinking metadata". *Cataloging & classification quarterly*, v. 52, n. 2, pp. 156-180. *https://doi.org/10.1080/01639374.2013.848389*

Mahey, Mahendra; **Al-Abdulla, Aisha**; **Ames, Sarah**; **Bray, Paula**; **Candela, Gustavo**; **Chambers, Sally**; **Derven, Caleb**; **Dobreva-McPherson, Milena**; **Gasser, Katrine**; **Karner, Stefan**; **Kokegei, Kristy**; **Laursen, Ditte**; **Potter, Abigail**; **Straube, Armin**; **Wagner, Sophie-Carolin**; **Wilms, Lotte** (2019). *Open a GLAM lab. Digital cultural heritage innovation labs*. Book Sprint, Doha, Qatar, 23-27 September. *https://glamlabs.pubpub.org*

Mayer-Schönberger, Viktor; **Cukier, Kenneth** (2013). *Big data: La revolución de los datos masivos*. Turner. ISBN: 978 8415832102

Moretti, Franco (2013). *Distant reading*. London: Verso. ISBN: 978 1781680841

Phetteplace, Eric; **Brooks, Mackenzie**; **Heller, Margaret** (2013). "Library labs". *Reference & user services quarterly*, v. 52, n. 3, pp. 186–190. *https://ecommons.luc.edu/lib_facpubs/24/*

Robertson, Adi (2013). "Building an atlas for books with small demons: A startup mines literary references for real-world connections". *The verge*, 1 March. *https://www.theverge.com/2013/3/1/4043298/building-an-atlas-for-books-with-small-demons*

Rodríguez, Jaime-Alejandro (2017). "Humanidades digitales: una oportunidad para allanar la brecha entre las dos culturas". En: Pereira G., José-Miguel (ed.). *Humanidades digitales, diálogo de saberes y prácticas colaborativas en red*. Bogotá: Editorial Pontificia Universidad Javeriana, pp. 137-146. ISBN: 978 9587811667

Rodríguez-Yunta, Luis (2014). "Ciberinfraestructura para las humanidades digitales: una oportunidad de desarrollo tecnológico para la biblioteca académica". *El profesional de la información*, v. 23, n. 5, pp. 453-462. *https://doi.org/10.3145/epi.2014.sep.01*

Scott, Dan (2017). "Wikidata, Canada 150, and music festival data". *Coffee code*, 2 junio. *https://coffeecode.net/wikidata-canada-150-and-music-festival-data.html*

Smith-Yoshimura, Karen (2020). *Transitioning to the next generation of metadata*. OCLC Research. https://doi.org/10.25333/rqgd-b343

Vukadin, Ana (2019). *Metadata for transmedia resources*. Chandos Publishing. ISBN: 978 0081013007

Ward, Mark; **Saarti, Jarmo** (2018). "Reviewing, rebutting, and reimagining fiction". *Cataloging & classification quarterly*, v. 56, n. 4, pp. 317-329. *https://doi.org/10.1080/01639374.2017.1411414*