

La muerte y el regreso de *Bibframe*

The death and return of *Bibframe*

Daniel Martínez-Ávila

Martínez-Ávila, Daniel (2023). "La muerte y el regreso de *Bibframe*". *Anuario ThinkEPI*, v. 17, e17a41.

<https://doi.org/10.3145/thinkepi.2023.e17a41>

Publicado en *IweTel* el 4 de diciembre de 2023

Daniel Martínez-Ávila

<https://orcid.org/0000-0003-2236-553X>

<https://www.directorioexit.info/ficha1675>

Universidad de León

Área de Biblioteconomía y Documentación

Campus de Vegazana

24071 León, España

dmarta@unileon.es



Resumen: Se presenta una breve revisión de las críticas a MARC y sus propuestas de reemplazo basadas en tecnologías semánticas y *linked (open) data*, centrándose principalmente en la reciente crítica de Jeff Edmunds *Bibframe must die*. Se enumeran y analizan los argumentos de Edmunds y se presentan, de forma sintética y comentada, algunas de las respuestas publicadas en la lista de distribución *Bibframe*, incluyendo aportaciones de Karen Coyle, Shawne Miksa y otros. Se concluye la necesidad de profundizar en el uso y divulgación de tecnologías semánticas basadas en *linked (open) data*, como *Bibframe*, para poder alcanzar una mejor representación del conocimiento en la era de *big data* y las inferencias de inteligencia artificial, ya que estos son aspectos que son indicados por Edmunds como el futuro de la representación en la catalogación.

Palabras clave: *Bibframe*; MARC; *Linked (open) data*; Catalogación.

Abstract: In this text I present a brief review of the critiques of MARC and its replacement proposals based on semantic technologies and linked (open) data, focusing mainly on Jeff Edmunds' recent critique titled *Bibframe must die*. Edmunds' arguments are presented and analyzed in addition to some of the responses published on the *Bibframe* mailing list, including contributions by Karen Coyle, Shawne Miksa, and others. I conclude with the need to deepen the use and dissemination of semantic technologies on the basis of linked (open) data such as *Bibframe* to achieve a better representation of knowledge in the era of big data and artificial intelligence inferences, as these are aspects that are designated by Edmunds as the future of representation in cataloguing.

Keywords: *Bibframe*; MARC; *Linked (open) data*; Cataloguing.

1. Introducción

Pese a que hace ya más de 20 años que Roy Tennant declaró que "MARC debería morir" ("*MARC must die*") (Tennant, 2002), lo cierto es que los nuevos estándares de metadatos adaptados a las tecnologías y posibilidades semánticas, tanto de contenido, como RDA, como de estructura e intercambio, como *Bibframe*, se resisten a despegar definitivamente en cuanto a implementación. RDA se publicó en 2010 y ha sido ampliamente adoptado internacionalmente, aunque la tendencia es que usando perfiles de aplicación creados por grandes bibliotecas (como los que también hace disponibles la BNE).

<https://www.bne.es/es/servicios/servicios-para-bibliotecarios/rda>

Sin embargo, el hecho de que el *RDA Toolkit* sea de pago y pocas bibliotecas públicas (incluso en Estados Unidos) tengan acceso a él provoca ciertas reticencias. Por ejemplo, en febrero de 2023 se generó

un debate en la lista de distribución *Educat* sobre el sentido de enseñar esta herramienta específica en lugar de tan solo los principios teóricos a una comunidad de futuros profesionales cuya gran mayoría jamás tendrá acceso al *Toolkit* o posibilidad de crear perfiles.

<https://listserv.loc.gov/cgi-bin/wa?A2=ind2302&L=EDUCAT&P=1448>

El anuncio también este año de que las *Original RDA* (que contemplan el formato MARC) no serían continuadas a favor de las *Official RDA* (que ya no están vinculadas a la estructura de MARC) fue también cuestionado por sus implicaciones profesionales y prácticas (parece ser que MARC continúa muy vivo en el subconsciente de los profesionales). Por otra parte, también a un nivel profesional, el desencanto con RDA y su filosofía de pago ha provocado el desarrollo de iniciativas quizás anecdóticas como las *Open Rules for Cataloging*

<https://openrulesforcatalogingproject.github.io>

En lo que respecta a la crítica a MARC de Tennant, los argumentos originales eran claros y directos. No obstante, esta declaración ha sido objeto de continuos debates y críticas, muchas veces infundadas y rozando lo personal (Tennant, 2017), que siguen coleando hasta nuestros días. Estas críticas, en sus vertientes más razonables, han estado provocadas principalmente por la posibilidad de una viabilidad práctica de su reemplazo (debido al amplio número de registros existentes en MARC) y, de una forma más visceral, por la resistencia irracional al cambio de algunos profesionales. Kroeger (2013), revisó algunas de las principales iniciativas de reemplazo de MARC durante el período 2002-2012 discutiendo desde la declaración de Tennant y otros autores, el *Informe Calhoun* de la *Library of Congress* (Calhoun, 2006); el informe *On the Record* (*Library of Congress*, 2008); el código de catalogación de Martha Yee en RDF y las iniciativas de desarrollo de *Bibframe* de aquel momento.

<http://myee.bol.ucla.edu/ucatrul&rdflntro.htm>

Recordemos que *Bibframe* es un modelo entidad-relación basado en RDF, adaptado al entorno *linked data* y diseñado como la mejor alternativa de reemplazo de MARC. Kroeger (2013), desde una posición no exenta también de crítica, concluyó que *Bibframe* representa un progreso en la dirección adecuada, aspecto que por otra parte no parece ser compartido por todos los colegas. Como ejemplo, en la fecha equivalente al día de los inocentes en Estados Unidos de 2022 (el 1 de abril), se publicó en la lista de *Bibframe* el anuncio de la *Library of Congress* de que en tres meses no habría registros MARC creados o disponibles desde esta institución. La seriedad de la información era dudosa por la fecha en la que se publicó, pero los argumentos eran convincentes. Xavier Agenjo (quien por cierto ha publicado junto a Francisca Hernández-Carrascal numerosos trabajos seminales en español sobre *Bibframe*, como por ejemplo Agenjo-Bullón; Hernández-Carrascal, 2015; 2016; 2017; 2018; 2020) se hizo eco del mensaje en *IweTel*, confiando sin muchas esperanzas que no fuera una broma, dejando entrever la razón que tenía el mensaje en su fondo.

2. ¿*Bibframe* debe morir?

El día que justamente se cumplieron 21 años del artículo de Tennant, el 15 de octubre de 2023, Jeff Edmunds –quien ya había discutido en 2017 el futuro fracaso del sistema en la lista de distribución de *Bibframe* y diversos trabajos (Edmunds, 2017)– publicó un texto con el provocativo título *Bibframe must die* (*Bibframe debe morir*) (Edmunds, 2023). El anuncio de este texto fue distribuido en la lista de *Bibframe* y provocó un interesante debate.

<https://listserv.loc.gov/cgi-bin/wa?A2=ind1702&L=BIBFRAME&P=61>

El texto comienza con una crítica o comentario sarcástico sobre el chascarrillo original de Tennant sobre la edad de MARC y los Beatles. Continúa comentando que las tecnologías sobre las que se basa *Bibframe*, (*Official*) *RDA* y *LRM*, surgidas en la mitad de la década de los 90 (supuestamente *linked data* y la web semántica), son anteriores a las posteriores revoluciones de *Google*, smartphones, internet de las cosas, *ChatGPT* y otras tecnologías que han cambiado la forma en la que buscamos, recuperamos y pensamos la información. Las tecnologías sobre las que se basa *Bibframe* estarían basadas en concepciones anticuadas del descubrimiento, recuperación y descripción de recursos e información en un mundo conectado. El problema, según Edmunds, es que *Bibframe* finge ser moderno cuando está anticuado, fundamentalmente defectuoso y es “un peligro para el futuro de la catalogación”.

“¿Por qué dejar a la inteligencia artificial la posibilidad de hacer inferencias (suponemos que cada vez más) probabilísticamente correctas cuando se puede proporcionar esa base de conocimiento y razonamiento de forma explícita por parte de creadores y profesionales?”

El objeto de la crítica de Edmunds también recae sobre FRBR, citando que las tareas del usuario de su modelo y de RDA son insuficientes para satisfacer sus necesidades, y afirmando que la idea de que usuarios y catalogadores piensan o deberían pensar en términos de obras, expresiones, manifestaciones e ítems (las entidades de FRBR) es absurda, debido a las características actuales del universo bibliográfico. La solución para Edmunds no se encontraría en la costosa y forzada integración de relaciones entre entidades que RDA y *Bibframe* proponen, sino en la indización del texto completo, *big data* y sistemas de inteligencia artificial que puedan realizar inferencias precisas a partir de datos poco limpios e inconsistentes. La idea es que la misma crítica que se posicionó y se mantiene sobre MARC respecto a las posibilidades tecnológicas de los sistemas bibliográficos es igualmente aplicable a *Bibframe* en la actualidad. Tras citar las dificultades y observaciones de algunas bibliotecas que han completado o trabajado en la transición a *Bibframe* (como la *Biblioteca Nacional de Suecia*), **Edmunds** (2023) lista las que son a su entender las cinco razones por las que *Bibframe* debe morir:

1) *Bibframe* costaría caro a las bibliotecas. Aunque se reconoce que con el tiempo se “podría” ganar en eficiencia, se resalta un dramático aumento de costes por algunos años, incluso después de terminada su transición, y se citan casos de subvenciones millonarias terminadas y deudas y recortes en universidades norteamericanas (cuyas bibliotecas supuestamente se encuentran en incluso mejor situación que las públicas).

2) *Bibframe* es inviable. Aquí se citan numerosos problemas:

- bibliotecas, vendedores y editoriales tendrían que abandonar MARC;
- se deberían resolver las inconsistencias de las implementaciones locales de *Bibframe*;
- se debería acordar y adoptar un único tipo de intercambio de *Bibframe*;
- todos los editores que cataloguen nativamente con *Bibframe* deberían generar un resultado que sea por lo menos compatible;
- sistemas de gestión de entidades deberían ser diseñados para una gestión compartida de *linked data*, obviando los silos;
- sistemas de gestión bibliotecaria deberían adaptarse a *Bibframe* y *linked data*;
- proveedores de registros MARC deberían migrar a *Bibframe*;
- OCLC debería transformar su base de datos de registros bibliográficos y de autoridades a tripletes de entidades y relaciones basados en *Bibframe*;
- OCLC debería disponer de editores y APIs para la creación nativa de datos *Bibframe* en *WorldCat*, para todo tipo de formatos y para su intercambio entre miembros;
- *Official RDA* debería ser adoptada universalmente y mapeada a *Bibframe*;
- catalogadores de cientos de miles de bibliotecas de todo el mundo deberían aprender los principios de *Bibframe*;
- catálogos y herramientas de descubrimiento deberían ser rediseñados para seguir los principios de *Bibframe* y *linked data*;
- se debería diseñar, crear y mantener una base de tripletes para evitar copiar datos locales y almacenarlos centralmente.

3) *Bibframe* no añade valor al ecosistema de descubrimiento. Según Edmunds la supuesta superioridad de *Bibframe* en cuanto a experiencia de usuario y descubrimiento no está demostrada. Según el autor, la tecnología de *linked data* se trataría de una solución a problemas que tienen un origen técnico, mientras que la verdadera solución consistiría en mejorar la calidad de los metadatos y disminuir la cantidad de materiales que no están o están infra catalogados.

4) *Bibframe* no es amigable para el usuario, quienquiera que este sea. Aquí el argumento se conecta con el retraso de la implementación del *Official RDA Toolkit* por parte de la *Library of Congress* hasta que todos los catalogadores acepten *Bibframe*. En ambos casos, Edmunds apunta a su “jerga rocambolesca”, su diseño poco amigable para el usuario y su complejidad como problemas.

“Lo contrario de ontología no es no-ontología, sino mala ontología, y eso es lo que MARC supondría para las inferencias de la inteligencia artificial”

“Gran parte de los problemas son debidos a la insuficiente difusión que se le da a *Bibframe* y otros estándares y normas semánticas por parte de los organismos que los crean”

5) *Bibframe* nace y está impregnado de inequidad. Se indica de nuevo que la implementación de *Bibframe* al igual que el *Official RDA Toolkit* son caras, sólo al alcance de unas pocas instituciones norteamericanas y predominantemente blancas.

Edmunds concluye que mientras Tennant proclamó que MARC debería morir por razones teóricas, él lo hace por razones prácticas. La crítica de Edmunds parece efectiva por varias razones:

- primero, porque, aunque de forma selectiva, aporta datos y conocimiento práctico desde el área;
- segundo, porque, en principio, no lo hace para defender MARC (pese a que, aunque no lo coloque de forma explícita en el texto, en una de las respuestas del debate de la lista de distribución sí afirma que *Bibframe* no supone ningún beneficio sobre MARC, ya que según él todas las ventajas de la nueva alternativa son hipotéticas y no comprobadas);
<https://listserv.loc.gov/cgi-bin/wa?A2=ind2311&L=BIBFRAME&P=8403>
- tercero, porque deja entrever que existe una alternativa incluso mejor que *Bibframe* basada en inteligencia artificial, *big data* y nuevas tecnologías (poco se puede argumentar contra un argumento tan abstracto pero a la vez tan efectivo). En esta línea argumental, una defensa de *Bibframe* no sería vista como una defensa del progreso sino del pasado, pese a que la alternativa del futuro nunca queda realmente clara y lo que se muestra como realidad (MARC) está mucho más atrasado incluso para su procesamiento automático.

Sin embargo, el debate sí logró otro objetivo inesperado: el cuestionamiento de las características de *Bibframe* entre aquellas personas que por estar inmersas en una realidad de MARC no lo conocían en profundidad, incluyendo algunas de esas personas que no habían dado el paso a su implementación haciendo realidad esa crítica de que "*Bibframe* es inviable". Las ventajas teóricas de *Bibframe* fueron una vez más discutidas por profesionales y académicos y, aunque este no es el espacio para enumerarlas todas, sí me gustaría resaltar algunas cuestiones colocadas como respuesta al texto de Edmunds junto a algunas reflexiones propias.

En lo que respecta a inclusividad, *Bibframe* también pasa de una codificación de caracteres MARC-8 a UTF-8, que permite una correcta representación de caracteres no latinos en lenguajes bibliográficos, incluyendo por ejemplo multitud de lenguajes asiáticos. Aquí me gustaría destacar que el problema de la romanización de caracteres ha tenido un amplio recorrido de discusión en MARC (Jeong; Oh, 1991) y también como ventaja de RDA (Seikel, 2009).

Karen Coyle, quien también es citada como autoridad en uno de los puntos del artículo de Edmunds, volvió a recordar el problema práctico de la falta de escalabilidad de MARC (algunos campos han agotado las posibilidades de subcampos a-z, 0-9), límites insuficientes de registros (99999 bytes) y campos (9999 bytes), incapacidad de manejar tablas de contenido, imágenes y otros formatos asociados y, sobre todo, que las búsquedas a texto libre no pueden sustituir las prácticas semánticas de la organización del conocimiento. Las posibilidades de poder navegar por una ontología de relaciones (jerárquicas, no solo de autoridad) en el catálogo es un aspecto relevante para los usuarios y el descubrimiento (aunque cuestionado por el propio Edmunds en la discusión posterior). Mi propia reflexión en este punto es la siguiente: ¿por qué dejar a la inteligencia artificial la posibilidad de hacer inferencias (suponemos que cada vez más) probabilísticamente correctas cuando se puede proporcionar esa base de conocimiento y razonamiento de forma explícita por parte de creadores y profesionales? Las ventajas y posibilidades de la agencia en organización del conocimiento y *linked (open) data* en la era de *big data*, aunque sin citar explícitamente *Bibframe*, han sido también discutidas en algunos trabajos (Ibekwe-SanJuan; Bowker, 2017; Martínez-Ávila et al., 2018). Tim Thompson, bibliotecario de la Universidad de Yale, también citó a Guizzardi (2010) a este respecto indicando que lo contrario de ontología no es no-ontología, sino mala ontología, y eso es lo que MARC supondría para las inferencias de la inteligencia artificial.

3. Conclusión

Shawne Miksa, de la *Universidad de North Texas*, indicaba en uno de los últimos mensajes del debate que gran parte de los problemas son debidos a la insuficiente difusión que se le da a *Bibframe* y otros estándares y normas semánticas por parte de los organismos que los crean (y en entornos LOD eso no debería restringirse necesariamente a las gran-

“Las búsquedas a texto libre no pueden sustituir las prácticas semánticas de la organización del conocimiento”

“*Bibframe* está más vivo que nunca, no quizás tanto en su estado de implantación sino por la necesidad de profundizar en su divulgación en todos los foros posibles”

des instituciones). En ese sentido, *Bibframe* está más vivo que nunca, no quizás tanto en su estado de implantación sino por la necesidad de profundizar en su divulgación en todos los foros posibles. Todas estas iniciativas y publicaciones sobre *Bibframe* son parte de esa solución que Miksa apunta y, en el caso del español, los trabajos de Xavier Agenjo y Francisca Hernández-Carrascal (algunos de ellos en el *Anuario ThinkEPI*) también han sido un buen ejemplo. Por mi parte, espero que esta nota también contribuya a generar el debate y la divulgación que el modelo requiere. Para más información sobre *Bibframe* remito a la página del modelo de la *Library of Congress*:

<https://www.loc.gov/bibframe>

4. Referencias

Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2015). "Avances de *Bibframe* en 2014". *Anuario ThinkEPI*, v. 9, pp. 269-277.

<https://doi.org/10.3145/thinkepi.2015.64>

Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2016). "*Bibframe* en 2015: avances y perspectivas del nuevo modelo bibliográfico". *Anuario ThinkEPI*, v. 10, pp. 158-164.

<https://doi.org/10.3145/thinkepi.2016.32>

Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2017). "Avances de *Bibframe* en 2016: perspectivas del nuevo modelo bibliográfico". *Anuario ThinkEPI*, v. 11, pp. 310-318.

<https://doi.org/10.3145/thinkepi.2017.58>

Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2018). "*Bibframe* 2017, entre la internacionalización y el desembarco en Europa". *Anuario ThinkEPI*, v. 12, pp. 373-377.

<https://doi.org/10.3145/thinkepi.2018.62>

Agénjo-Bullón, Xavier; Hernández-Carrascal, Francisca (2020). "2018-2019, *Bibframe* en producción". *Anuario ThinkEPI*, v. 14, e14f03.

<https://doi.org/10.3145/thinkepi.2020.e14f03>

Calhoun, Karen (2006). *The changing nature of the catalog and its integration with other discovery tools. Final report*. March 17. Washington, DC: Library of Congress, 2006.

<https://www.loc.gov/catdir/calhoun-report-final.pdf>

Edmunds, Jeff (2017). *Bibframe as empty vessel*. ScholarSphere. PennState University Libraries.

<https://doi.org/10.26207/zs35-2546>

Edmunds, Jeff (2023). *Bibframe must die*. ScholarSphere. PennState University Libraries.

<https://doi.org/10.26207/v18m-0g05>

Guizzardi, Giancarlo (2020). "Ontology, ontologies and the 'I' of FAIR". *Data Intelligence*, v. 2, n. 1-2, pp. 181-191.

https://doi.org/10.1162/dint_a_00040

Ibékwe-SanJuan, Fidelia; Bowker, Geoffrey C. (2017). "Implications of big data for knowledge organization". *Knowledge organization*, v. 44, n. 3, pp. 187-198.

<https://doi.org/10.5771/0943-7444-2017-3-187>

Kroeger, Angela (2013). "The road to *Bibframe*: The evolution of the idea of bibliographic transition into a post-MARC future". *Cataloging & classification quarterly*, v. 51, n. 8, pp. 873-890.

<https://doi.org/10.1080/01639374.2013.823584>

Library of Congress (2008). *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Washington, DC: Library of Congress.

<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

Jeong, Pil-Mo; Oh, Dong-Geun (1991). "The influence of pronunciation and scripts on the East-Asian MARC formats". *International cataloguing & bibliographic control*, v. 20, n. 4, pp. 56-60.

Martínez-Ávila, Daniel; Smiraglia, Richard P.; Szostak, Rick; Scharnhorst, Andrea; Beek, Wouter; Siebes, Ronald; Ridenour, Laura; Schlais, Vanessa (2018). "Classifying the LOD cloud: digging into the knowledge graph". *Brazilian journal of information sciences: Research trends*, v. 12, n. 4, pp. 6-10.

<https://doi.org/10.36311/1981-1640.2018.v12n4.02.p6>

Seikel, Michele (2009). "No more romanizing: The attempt to be less anglocentric in RDA". *Cataloging & classification quarterly*, v. 47, n. 8, pp. 741-748.

<https://doi.org/10.1080/01639370903203192>

Tennant, Roy (2002). "MARC must die". *Library Journal*, October 15, pp. 26-28.

<http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf>

Tennant, Roy (2017). "'MARC must die' 15 years on". *Hanging together: the OCLC research blog*, October 15.

<https://hangingtogether.org/marc-must-die-15-years-on/>

* * *

Un salto al vacío Hugo Agud-Andreu

Mi aportación es en referencia a por qué dejarlo en manos de la inteligencia artificial. No hablo tanto de inteligencia artificial, sino de la capacidad de cálculo de las máquinas. Soy consciente que mi punto de vista también tiene fallos, pero es que la información en sí misma, llámese MARC21 o *Bibframe* es solo un eslabón más de la cadena de difusión, como un átomo, pero la ingente cantidad de información hace que no baste con que esa información exista, sino que debe ser procesada, indexada y puesta a disposición del usuario.

¿Por qué *Google* ha tenido tanto éxito? ¿Por qué *ChatGPT* lo está “petando”? porque indexan esa información y la procesan mediante una serie de algoritmos que la enriquecen y la validan. Digamos que tienen en cuenta el impacto de ese recurso, las visitas, los enlaces, y mil variables más que soy incapaz de comprender y que desconozco. Pensar que el secreto está en los ingredientes, para mí es un error...; el secreto está sin duda en el “cocinero”, en cómo trata y gestiona y presenta la ingente cantidad de información de millones de fuentes.

No debemos olvidar quién domina la “difusión” de la información. Empresas como *Google* o como *OpenAI* son expertos en gestionar y difundir información, y si *Bibframe* fuera relevante para el descubrimiento, estoy absolutamente seguro de que hubieran apostado e invertido en *Bibframe*. Yo aún recuerdo *Google Books* cuando les interesó invertir.

Por otro lado, siempre he pensado que menos es más, y para que algo triunfe debe ser “sencillo” y a poder ser económico... y *Bibframe*, ni barato, ni sencillo.

Lo que realmente daría alas a *Bibframe* sería un informe serio de alguna institución que lo haya implementado, que explique detalladamente los costes de implementación y el impacto que ha tenido en los usuarios y en su instrucción.

Sin esa experiencia real y positiva, un salto al *Bibframe* para las pequeñas bibliotecas sería un salto al vacío.

Hugo Agud-Andreu
Orex
hagud@orex.es

Dialnet | métricas

Nueva edición 2020

Fundación Dialnet

dialnet.unirioja.es/metricas