

MetaMetrics: prototipo de visualización de la calidad de los metadatos en revistas científicas latinoamericanas publicadas en Open Journal System

MetaMetrics: prototype for visualizing the quality of metadata in Latin American scientific journals published in the Open Journal System

Biblioteca Universitaria, vol. 26, núm. 1, enero-junio, 2023. pp. 12-23.
DOI: <http://dx.doi.org/10.22201/dgbsdi.0187750xp.2023.1.1466>

MANUEL ALEJANDRO FLORES CHÁVEZ*

* Dirección General de Bibliotecas y Servicios Digitales de Información (DGBSDI) Universidad Nacional Autónoma de México (UNAM), Correo-e: mafloresc@dgb.unam.mx ORCID 0000-0002-2465-0724

Palabras Clave:

Calidad de metadatos, cosecha de metadatos, Open Journal System (OJS), visualización de datos, revistas científicas, América Latina.

Keywords:

Metadata quality, metadata harvesting, Open Journal System (OJS), data visualization, journals, Latin America.

RESUMEN

MetaMetrics es un tablero de datos interactivo que genera informes sobre la calidad de los metadatos en revistas científicas, a partir de tres características señaladas por indagaciones previas: suficiencia, precisión y consistencia. Para su desarrollo se llevó a cabo una investigación en la que se utilizaron 26 453 registros cosechados de los sistemas Open Journal System (OJS) de 716 revistas adscritas a instituciones de la Red de Macrouniversidades de América Latina y el Caribe. La evaluación de los datos se hizo mediante expresiones regulares en Google Sheets que siguieran el patrón de caracteres definido por las reglas ortográficas, sintácticas y semánticas de los *OpenAIRE Guidelines for Literature Repositories v3*. La visualización de datos aplicada al ámbito de los metadatos facilita la identificación de errores en la descripción bibliográfica que obstaculizan el proceso de indización de revistas en los índices internacionales.

ABSTRACT

MetaMetrics is an interactive data dashboard that generates reports on the quality of metadata in scientific journals, based on three characteristics identified by previous research: sufficiency, precision and consistency. For its development, 26,453 records harvested from the Open Journal System (OJS) of 716 journals belonging to institutions of the Network of Macrouniversities of Latin America and the Caribbean (Red de Macrouniversidades de América Latina y el Caribe) were used. The data was evaluated using regular expressions in Google Sheets following the character pattern defined by the orthographic, syntactic and semantic rules of the *OpenAIRE Guidelines for Literature Repositories v3*. Data visualization applied to the metadata domain facilitates the identification of errors in the bibliographic description that hinder the process of indexing journals in international indexes.

Introducción

El objetivo de este artículo es presentar, mediante una investigación llevada a cabo, el desarrollo y características de un prototipo de evaluador automático de calidad de metadatos, MetaMetrics, para revistas científicas alojadas en OJS. MetaMetrics está disponible para su descarga en formato de Excel y puede ser utilizado como una herramienta de diagnóstico por los equipos editoriales de las revistas, o como un filtro para la selección de publicaciones en índices, bases de datos y portales.

La evaluación de la calidad de los metadatos a nivel artículo debiera ser una tarea continua de cualquier equipo editorial y de los sistemas de información dedicados a la indización de revistas. El principal desafío que enfrentan estos actores es la cantidad de registros con los que deben trabajar y que, lógicamente, son imposibles de evaluar meramente a través del ser humano. Ante esta situación, es conveniente contar con herramientas informáticas que den al indizador humano la posibilidad de identificar rápidamente los problemas de codificación y asentamiento de los metadatos y, del mismo modo, corregirlos o brindar a los editores de revistas la retroalimentación apropiada. Como parte de la solución a esta problemática se plantea el diseño y funcionamiento de un tablero de datos en Excel creado exprofeso para evaluar la calidad de los metadatos a nivel documento en revistas científicas.

Los tableros de datos, también conocidos como *dashboards*, son representaciones visuales de las características y relaciones más importantes de un conjunto de datos; a través de gráficas, tablas, mapas u otros elementos intentan mostrar una situación derivada del procesamiento de datos, por ejemplo, las transacciones de una biblioteca o el uso de sus servicios (Ramos *et al.*, 2019). Como tal son parte de las técnicas de visualización de datos, un área que estudia la representación gráfica de los datos, especialmente en volúmenes tan grandes que serían difíciles de procesar para el ser humano, los cuales implican una sobrecarga cognitiva que dificulta la toma de decisiones sobre lo que se debe hacer con los metadatos (Nichols, 2008; Shankaranarayanan, 2021).

El uso de tableros en la evaluación de la calidad de los metadatos supone un escenario poco explorado o escasamente documentado. Si bien las bibliotecas siempre han sido partidarias de la calidad e interoperabilidad de los metadatos, no siempre han existido los recursos tecnológicos, humanos o económicos para aplicar estas herramientas (Harper, 2016).

Un caso en donde se haya utilizado la visualización de datos tuvo lugar en el año 2003. Luego de haber recolectado los metadatos de numerosos objetos de aprendizaje, basándose en los elementos *type* y *format* del estándar Dublin Core, Dushay y Hillmann (2003) se dieron cuenta de cuatro problemas: 1) había elementos vacíos, 2) se proporcionaban datos que debían ir en otros campos, 3) los vocabularios controlados para definir el tipo de documento y su formato eran diferentes en cada colección, y 4) aun cuando los elementos de Dublin Core son repetibles, se encontraron datos de un mismo tipo en un solo campo. El hallazgo de esas autoras se graficó con el software Spotfire, creado originalmente para el ámbito de los negocios. En Spotfire la representación de los datos se hace mediante cuadros de diferentes tamaños y colores, la proporción de la figura representa qué tanta presencia tiene una característica, por ejemplo, el uso de cierto vocabulario controlado para codificar el elemento *type*. Las autoras citadas también pudieron determinar en qué otros campos no se ingresó ningún dato (pp. 162-163).

Esa experiencia demostró que la calidad de los metadatos es relevante cuando son reutilizados para alimentar otros sistemas o al emplearse más allá de la recuperación de información (Király *et al.*, 2019). Con esto en mente, los estudiosos del tema comenzaron a preguntarse qué criterios de calidad deberían reunir los metadatos; tras revisar más de 40 mil registros provenientes de cuatro colecciones diferentes, Shreeves *et al.* (2005) se dieron cuenta de que de los 15 elementos proporcionados por Dublin Core solo 8 se utilizan con más frecuencia: *title*, *creator*, *subject*, *description*, *date*, *format*, *identifier* y *rights*, de manera que al cumplir con un mínimo de elementos se puede hablar de una **suficiencia** o **completitud** de los metadatos. Jackson *et al.* (2008) pudieron confirmar esta declaración al analizar una colección

distinta y en la que el 77% de sus registros satisficieron esos campos. Un año más tarde Park (2009) revisó diferentes estudios y concluyó que en la mayoría de ellos se enfatizaban tres criterios o dimensiones básicos de calidad: suficiencia, la cual determina qué tan completo está un registro bibliográfico según los lineamientos de descripción utilizados; precisión, que se refiere a la representación correcta del objeto catalogado a través de sus metadatos; y consistencia, también llamada coherencia, que es utilizada para asegurarse que los datos estructurados tengan siempre un formato específico, por ejemplo, DD-MM-AAAA (día-mes-año) para las fechas.

Cada uno de estos criterios evalúa aspectos de calidad específicos. De acuerdo con Elouataoui *et al.* (2022), la suficiencia se enfoca en la **usabilidad** de los registros, así, un registro es útil si está completo (suficiente); por otro lado, los metadatos que representan correctamente a un objeto tienen la característica de la **confianza** y cuando cumplen la estructura definida por estándar se dice que tienen el aspecto de la **regularidad** (p. 113).

Los criterios anteriores son un marco de referencia, al momento de evaluar los metadatos de una colección es importante considerar otros factores, como los propósitos de la institución en la que se generan los registros o el área de conocimiento que cubre. En las colecciones federadas y alimentadas con metadatos de fuentes externas es más probable encontrar variaciones en el cumplimiento de la calidad, debido a que el ingreso de datos depende de varias organizaciones (Park, 2009; Shankaranarayanan, 2021); por otra parte, la investigación en ciertas áreas puede tener métodos muy particulares que sean relevantes para la recuperación. La arqueología, por ejemplo, emplea materiales gráficos o tridimensionales; mientras que en el caso de la zoología la información taxonómica de los seres vivos estudiados aporta información útil para los investigadores, por lo que conviene asentarla en el registro (Kim, Yakel & Faniel, 2009).

El tipo de situaciones mencionadas en el párrafo anterior es lo que dificulta establecer un concepto de calidad único. Si bien los criterios a partir de los

cuales se medirá esta última pueden tener la misma denominación (suficiencia, puntualidad y consistencia), será el contexto lo que determine el contenido de estos (Tani, Candela & Castelli, 2013). En términos operativos la calidad es el grado de cumplimiento de un estándar o norma, el cual es convenido por una comunidad de expertos que tiene objetivos particulares (Hughes, 2004). En cualquier caso, el fin último de los metadatos es satisfacer las necesidades del usuario de estos (Palavitsinis, 2013) y garantizar la interoperabilidad.

Respecto al procedimiento para evaluar la calidad, el uso de porcentajes es uno de los métodos más comunes; se trata de calcular la proporción de registros que cumplen o no las especificaciones de un estándar (Tsiflidou & Manouselis, 2013), ya sea manualmente o por procedimientos informáticos (Ochoa & Duval, 2009). Otra alternativa es comparar los términos de indización con los de búsqueda, esta última opción es útil especialmente para los usuarios de los servicios de información.

Para las evaluaciones más complejas es necesario determinar qué aspectos o atributos son los que se espera encontrar en los datos, a dichas características Palavitsinis (2013) las llamó **dimensiones**; estas únicamente proporcionan una idea general de lo que se va a evaluar. Al describir cada una de las dimensiones se obtienen entonces las **métricas**, en ellas quedan definidas las consideraciones que se tienen en cuenta al momento de realizar la evaluación. Por último, para emitir un juicio hay que tener una referencia, esto es un valor o, dicho de otra manera, **indicadores**. En la tabla 1 se propone un ejemplo propio y sencillo de estos conceptos (p.25).

Tabla 1. Ejemplo de dimensiones, métricas e indicadores para evaluar la calidad de los metadatos

Dimensión	Métrica	Indicador
Suficiencia	Todos los registros tienen al menos título, autor y palabras clave.	Porcentaje. Valor esperado: 100%
Precisión	Los identificadores del documento (DOI), la revista (ISSN) y autor (ORCID) aparecen en el documento y en el registro.	Porcentaje. Valor esperado: 90%
Consistencia	Los autores se asientan siempre por apellido.	Porcentaje. Valor esperado: 100%

Fuente: Elaboración del autor.

Las dimensiones, métricas e indicadores pueden variar según los objetivos de la evaluación, quizá para algunas instituciones sea más sencillo determinar únicamente qué tan completos están sus registros respecto a los de sistemas similares. En todo caso, las tres dimensiones actúan como niveles de calidad interrelacionados, el máximo nivel de calidad se obtiene cuando los metadatos son consistentes, hecho que solo se logra si antes cumplen el nivel más básico, que es la suficiencia y el intermedio, es decir, la precisión (Bellini y Nesi, 2013). Los estudios más profundos pueden incluir o hasta proponer métricas más elaboradas, como las recopiladas por Tsiflidou y Manouselis (2013): *entropía*, basada en la cantidad de información que incluye un campo; *Metadata multi-linguality for free text elements (Language Attribution)*, que se emplea para medir la frecuencia con la que un idioma es utilizado en los campos no controlados, como son títulos, resúmenes y palabras clave.

A lo largo de esta revisión teórica, se ha visto que la evaluación de la calidad puede ser tan sencilla como calcular porcentajes o exhaustiva como estudiar aspectos específicos del texto introducido en ciertos campos. En realidad, todas las perspectivas son válidas si se aplican a colecciones que requieren cumplir esas particularidades. Uno de los aspectos más relevantes a considerar es la presentación de los resultados, los métodos visuales son útiles, aunque hace falta explotarlos dentro del ámbito de la indización y evaluación de metadatos.

Método

En esta investigación se desarrolló el tablero interactivo MetaMetrics, que permite estimar la calidad de los metadatos de revistas publicadas en OJS partiendo de las dimensiones propuestas por Park (2009): suficiencia, precisión y consistencia. Los datos obtenidos para la construcción de esta herramienta se recolectaron de las plataformas OJS de 716 revistas editadas en instituciones de la Red de Macrouiversidades de América Latina y el Caribe; solo aquellos que tuvieran OJS se tomaron en cuenta. El listado de revistas en general está disponible en <https://bit.ly/3zUGKMs>.

Para la recolección de metadatos se utilizó el módulo *OAI Harvester Records* y la hoja de estilo XML *OAI-DCtoMARC* del software *MarcEdit 7*. Esta última se adaptó para recolectar todos los elementos Dublin Core disponibles en cada registro, además de programarla para extraer solo aquellos registros cuya fecha de publicación (*dc:date*) fuera 2021; dicha hoja está disponible en <https://bit.ly/3BEJZZH>. En total, se recolectaron más de 26 mil registros, de los cuales se excluyeron aquellos cuyo título hiciera alusión a meros listados del comité editorial, anuncios de eventos o presentaciones del número; se dejaron artículos y reseñas de libros, por ser tipo de documentos definidos por OpenAIRE. Al final, se conformó una muestra de 26 453 registros. El conjunto íntegro de registros se puede consultar a través de <https://bit.ly/3Qgle99>.

Para manipular los datos se utilizó como herramientas Excel y Google Sheets. Primero, se cargaron en Excel para separarlos en columnas y hacer una selección de campos; debido a que los campos repetibles podían variar en cantidad de revista a revista se decidió trabajar solamente con el primer campo de cada campo repetible en los registros. Luego, se importaron los datos en las hojas de Google para analizar las cadenas de texto mediante expresiones regulares. Las expresiones regulares abreviadas, como *regex*, definen secuencias de caracteres que se utilizan para buscar dentro de los datos; en Google una de las fórmulas que permite trabajar con este tipo de expresiones es REGEXMATCH. La aplicación de estas a los registros se realizó con base en las recomendaciones de ortografía, semántica y sintaxis de los *OpenAIRE Guidelines for Literature Repositories v3*; en la tabla 2 pueden encontrarse los criterios tomados en cuenta. La elección de dichos lineamientos se hizo porque al observar los registros de sistemas OJS se encontró que los elementos *dc:type* y *dc:rights* se apegan a estos.

Tabla 2. Elementos del registro bibliográfico OpenAIRE y sus características

Tipo de elemento	Elemento	Semántica		Sintaxis
		Ortografía	Valores	
Obligatorio	Título	Los nombres propios y títulos inician con mayúscula	Texto libre	Un campo por título
	Creador	Los nombres propios inician con mayúscula	Texto libre	Apellido(s), Nombre(s)
		Se omiten los grados o títulos académicos		Un campo por título
	Nivel de acceso	No aplica	Access Terms Vocabulary	info:eu-repo/semantics/[acceso]
	Fecha de publicación	No aplica	Números	YYYY-MM-DD o solo YYYY
	Tipo de documento	No aplica	Publication Type Terms	info:eu-repo/semantics/[tipo]
	Identificador del recurso	No aplica	Uso de identificador persistente	No aplica
Obligatorio si existe en el documento	Identificador del proyecto*	No aplica	Term Grant Agreement	info:eu-repo/grantAgreement/[Funder]/[Funding Program]/[ProjectID]
	Fecha de fin de embargo*	No aplica	Números	info:eu-repo/date/embargoEnd/<YYYY-MM-DD>
	Temas	Se puede usar mayúsculas y minúsculas, pero de forma consistente	Texto libre o un número de clasificación	Un campo por tema o cada tema delimitado por punto y coma
	Descripción	Los nombres propios, inicios de oración y después de punto, se escriben con mayúscula.	Texto libre	Un campo por descripción
	Editorial	Los nombres propios inician con mayúscula	Texto libre	Para instituciones editoras: Entidad editora. Subentidad

Tipo de elemento	Elemento	Semántica		Sintaxis
		Ortografía	Valores	
Recomendables	Identificador alternativo*	No aplica	Identificadores sugeridos por OpenAIRE	info:eu-repo/semantics/altIdentifier/<scheme/<identifier>
	Condiciones de licencia	No aplica	Licencia Creative Commons o texto libre	URL de Creative Commons o © [Propietario derechos], año
	Referencias bibliográficas*	No aplica	Identificadores persistentes	info:eu-repo/semantics/reference/<scheme/<identifier>
	Referencia a conjuntos de datos*	No aplica	Identificadores persistentes	info:eu-repo/semantics/reference/<scheme/<identifier>
	Colaborador**	Los nombres propios inician con mayúscula	Texto libre	Apellido(s), Nombre(s) Un campo por colaborador
	Versión del documento	No aplica	Valores sugeridos por OpenAIRE	info:eu-repo/semantics/<versión>
	Formato	No aplica	Internet Media types (MIME)	application/<formato MIME> Un campo por formato
	Fuente	No aplica	Guidelines for Encoding Bibliographic Citation Information	Revista volumen (número), pp-pp. (año)
	Idioma	No aplica	ISO 639-3	Un campo por idioma
	Cobertura	No aplica	Vocabulario controlado	Un campo por cobertura
Audiencia	No aplica	U.S. Department of Education	Un campo por audiencia	
Opcional	Relación	No aplica	Identificador persistente	Un campo por relación

Fuente: Elaboración del autor.

Nota: Para el elemento Tipo de documento, el valor Texto libre solo es válido si existe un valor Publication Type Terms. El elemento Identificador del proyecto no se evalúa. El elemento Fecha de fin de embargo no se evalúa.

Debido a que REGEXMATCH devuelve valores de FALSO o VERDADERO se anidó con la fórmula IFS, que permite utilizar REGEXMATCH como condición y definir un texto a manera de resultado si se cumple o no la condición de la regex. Los valores asignados para el análisis de esta investigación fueron:

- Cumple: el elemento evaluado cumple con la ortografía, semántica y sintaxis que señalan los OpenAIRE.
- Error sintaxis: el elemento evaluado presenta problemas de sintaxis (dos títulos en un mismo campo, puntuación incorrecta, caracteres basura u otros similares).
- Error semántica: el elemento evaluado utiliza valores no definidos por el vocabulario sugerido por OpenAIRE para campos controlados o está vacío.
- Error ortografía: el elemento evaluado utiliza mayúsculas en todo el contenido del campo. Este valor solo aplica al dc:title, dc:creator y dc:description.
- Error de sintaxis-semántica: el elemento evaluado presenta problemas de sintaxis (más de un dato en un campo que es repetible, error de codificación de idioma en el campo *dc:language*) y el dato asignado no corresponde a

un vocabulario controlado de OpenAIRE para campos controlados. En general, aplica para *dc:rights*, *dc:date*, *dc:type* y *dc:audience*.

- Error de ortografía-semántica: el elemento evaluado presenta problemas de ortografía (uso de mayúsculas) y el idioma de su escritura no corresponde con el declarado en el campo *dc: language*. En general, aplica para *dc:title*, *dc:description* y *dc:type*.
- No cumple: el elemento evaluado no satisface la sintaxis, semántica ni ortografía que señalan los OpenAIRE.

Tras aplicar las expresiones regulares y verificar su funcionamiento, la hoja de Google se conectó a un libro de Excel. En este programa se construyeron tablas dinámicas que actuaran como fuente secundaria del tablero, pero que se actualizaran al realizar algún cambio desde Google.

Resultados

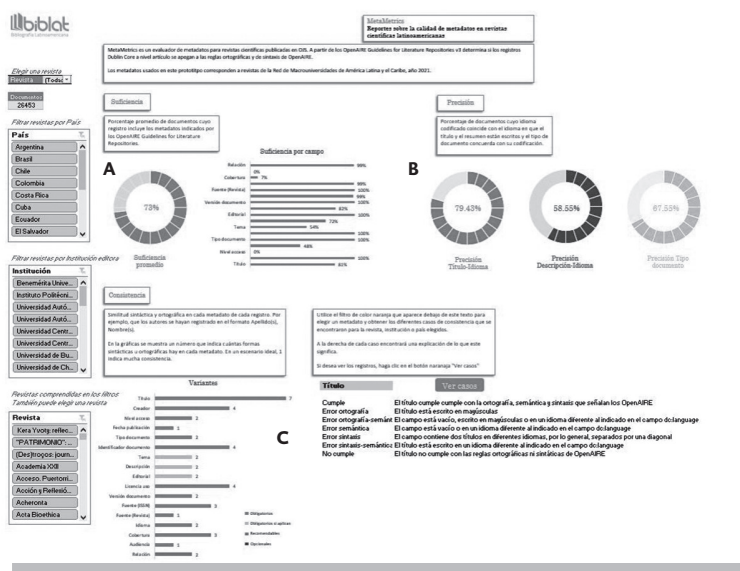
Tras aplicar la metodología descrita para esta investigación, se obtuvo un tablero interactivo en Excel al que se dio el nombre de MetaMetrics. Se trata de una herramienta que muestra los porcentajes de suficiencia (inclusión de 17 metadatos indicados por Dublin Core),

precisión (correspondencia del idioma codificado en los campos *dc:language* con el idioma en que se escribieron el título y el resumen del documento catalogado, además de la correspondencia del tipo de documento codificado en los campos *dc:type*) y la consistencia (número de formas diferentes en que se asentaron los metadatos partiendo de las reglas ortográficas, de sintaxis y semántica de los OpenAIRE); dimensiones de calidad sugeridas en diversos estudios. Para obtener una copia del tablero consulte el siguiente enlace <https://1drv.ms/u/s!ArRTOto1Jo1ujHYbUsGT6lzO3wTO?e=qwM5kv>, haga clic derecho sobre el único archivo que aparece y enseguida elija la opción Descargar.

La fuente de datos de MetaMetrics es una hoja de cálculo de Google donde vienen todas las columnas y celdas con fórmulas utilizadas para evaluar los metadatos. Si se quisiera añadir revistas nuevas sería necesario cosechar previamente sus metadatos a nivel artículo, vaciarlos al final de la hoja y modificar los rangos de las fórmulas. Luego, en MetaMetrics debe actualizarse la consulta de los datos mediante el menú Datos de Excel.

La ilustración 1 muestra la vista general de MetaMetrics. Otras características del tablero se irán presentando a continuación mediante ilustraciones con ejemplos particulares.

Ilustración 1. Vista General de MetaMetrics

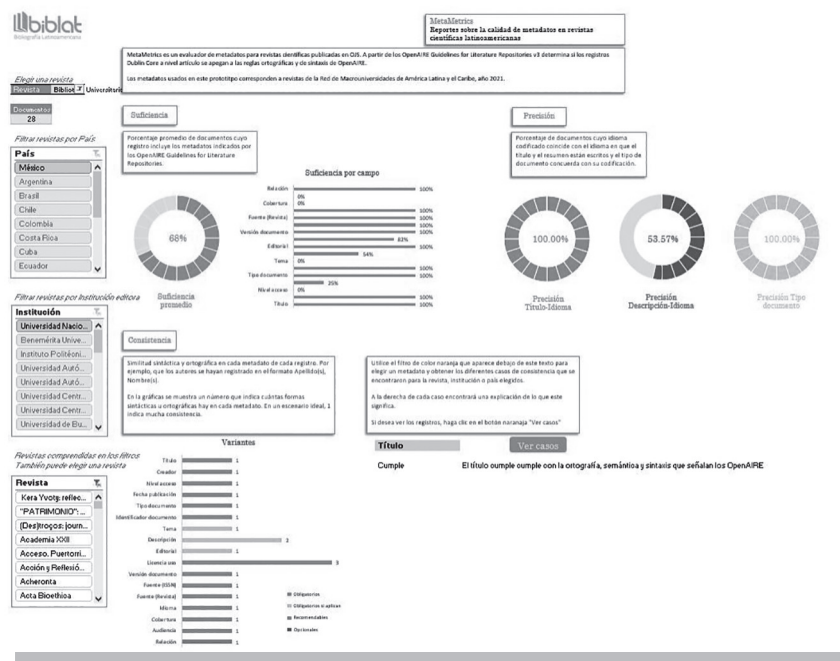


Fuente: Elaboración del autor.

La vista general de MetaMetrics muestra la suficiencia, precisión y consistencia de todos los registros recolectados, esto resulta útil cuando se tiene una colección completa que se desea someter a evaluación (Király, 2019). En esta investigación, esa colección corresponde a las revistas de la Red de Macrouiversidades de América Latina y el Caribe; así, por ejemplo, podríamos decir que según la ilustración 1, las revistas de esta Red tienen registros con el 73% de los campos de los OpenAIRE (gráfica A, ilustración 1) y que, además, cerca del 80% tienen un idioma codificado en dc:language que se corresponde con el idioma en que está escrito el título (gráfica B de la ilustración 1). Además, a nivel consistencia, el título es el campo donde hubo mayor número de formas diferentes de asentarlos (gráficas C de la ilustración 1).

Estos porcentajes pueden aplicarse a una revista en particular o bien a un conjunto de revistas de un país o institución editora. Para ello MetaMetrics cuenta con una serie de filtros, en la ilustración 1 se encuentran del lado derecho, debajo del logo de BIBLAT. La ilustración 2 muestra una vista de filtro para la revista *Biblioteca Universitaria*.

Ilustración 2. Vista de la revista Biblioteca Universitaria

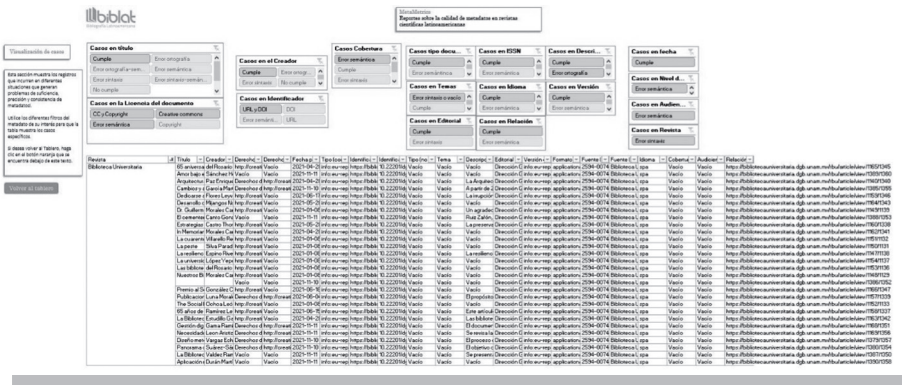


Fuente: Elaboración del autor.

Al elegir la revista *Biblioteca Universitaria*, en el área de filtros se activaron únicamente el país (México) y la institución editora de la revista seleccionada (Universidad Nacional Autónoma de México). Esta publicación tiene un 68% de suficiencia, su correspondencia de idioma codificado en título y resumen es del 100% mientras que tiene una gran consistencia para todos sus metadatos al tener una sola forma de asentarlos, excepto para la Descripción (resumen) y el identificador del documento.

Ahora bien, estos porcentajes y cifras se complementan con la función Visualización de casos que incluye MetaMetrics. Se trata de una tabla dinámica que muestra los registros y el tipo de incidencias por cada campo en forma de filtro; para acceder a dicha función se debe utilizar un botón llamado Ver casos. La ilustración 3 ejemplifica la funcionalidad para el caso de la misma revista, *Biblioteca Universitaria*.

Ilustración 3. Casos de la revista Biblioteca Universitaria

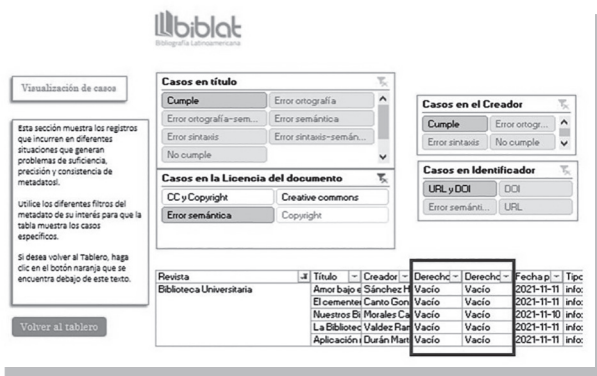


Fuente: Elaboración del autor.

Mediante la Visualización de casos es posible elegir una incidencia en uno o más campos y ver el registro en sí para determinar dónde está el error. Por ejemplo, para el caso de la revista *Biblioteca Universitaria* existe error de sintaxis en el campo que señala la licencia bajo la cual se publicaron los artículos (dc:rights), si se filtra por ese campo la tabla mostrará los registros donde ocurre la incidencia y podrá apreciarse a qué se debió el error.

En la ilustración 4 se observa que el error se debió a que los campos de Derechos/Licencia están vacíos. Si un sistema de indexación recolecta los metadatos de los artículos de esta revista, posiblemente se obtendría un error de sintaxis o los registros con esta incidencia quedarían fuera de la recolección.

Ilustración 4. Filtro de casos en campo de Licencia de la revista Biblioteca Universitaria



Fuente: Elaboración del autor.

Esta posibilidad de revisar los registros es el mejor valor agregado de MetaMetrics. Los datos generados también tienen la posibilidad de utilizarse para otro tipo de estudios sobre la publicación periódica y crear conjuntos de datos (*datasets*) que ayuden a entender la dinámica del documento científico a nivel metadatos (Isenberg *et al.*, 2017); quienes dirigen una revista pueden identificar con facilidad en qué documentos hubo errores durante la catalogación, mientras que los servicios de revistas indizadas, como BIBLAT, podrán dar una mejor retroalimentación a los editores.

Discusión

El propósito de la investigación fue mostrar el desarrollo y características de MetaMetrics. La herramienta proporciona una evaluación simple de los valores asentados en los registros Dublin Core de artículos científicos incluidos en plataformas OJS, teniendo como criterios de evaluación, por un lado, tres de las dimensiones de calidad más concurridas en los estudios sobre el tema (Park, 2009; Bellini & Nesi, 2013; Kim *et al.*, 2019), y por el otro, el estándar *OpenAI-RE Guidelines for Literature Repositories v3*. En contraparte, MetaMetrics carece de otras dimensiones consideradas en las fuentes académicas (*usabilidad, actualización, nivel de interoperabilidad o de descubrimiento*), tampoco agrupa las dimensiones por aspecto de calidad (Elouataoui, 2022, p.115-116), puntos que pueden ser áreas de mejora en una próxima versión.

Hasta el momento MetaMetrics está disponible como archivo Excel, por lo que su grado de usabilidad es pequeño y únicamente a nivel local (la versión web está aún en desarrollo); por otro lado, su funcionamiento solo ha sido probado con metadatos Dublin Core recolectados directamente de OJS y para revistas latinoamericanas, por lo que utilizarlo en otros contextos puede requerir ajustes no considerados en esta investigación.

Conclusiones

El resultado de esta investigación se concibe como un prototipo funcional para identificar los errores ortográficos, de sintaxis y semántica en los registros bibliográficos Dublin Core de revistas científicas gestionadas en Open Journal System, según el estándar de descripción *OpenAIRE Guidelines for Literature Repositories v3*. En un solo archivo, cuyos datos pueden actualizarse desde Google Sheets, es posible obtener un panorama general de qué tanto las revistas o conjunto de revistas (en función del país o institución editora) cumplen con lo establecido por los OpenAIRE a nivel de los metadatos de los documentos publicados.

Entre las mejoras (y el trabajo futuro) para MetaMetrics se encuentra la optimización de los procesos para ingresar metadatos nuevos, posiblemente mediante el desarrollo de un código que permita recolectar los registros y almacenarlos en la hoja de cálculo, esto rompería con la dependencia actual del prototipo hacia el cosechador de MarcEdit 7.5.

Hasta el momento, MetaMetrics está listo para producir reportes de revistas específicas. Al ser un archivo de hoja de cálculo no se necesitan requerimientos técnicos más allá de tener el software Excel 2013 o una versión pos-

terior y una cuenta Google para acceder a la fuente de datos de Google Sheets, en caso de que se desee añadir metadatos de una revista no considerada en la muestra.

Lo complejo en el estudio de la calidad de los metadatos es que estos se crean para satisfacer las necesidades de información que tiene una comunidad de usuarios y no necesariamente las de un usuario que utiliza los metadatos en sí. Por ejemplo, el tipo de documento puede ser poco relevante para quien simplemente quiere resolver una inquietud a partir del contenido, pero de suma importancia al momento de recolectar selectivamente metadatos de otros sistemas. Así, las bibliotecas y otros sistemas de información quedan en la disyuntiva de si es conveniente ser o no exhaustivos en el seguimiento de un estándar, o si incluir ciertos metadatos en el registro pero ocultarlos de la interfaz de búsqueda y habilitarlos únicamente para su eventual cosecha.

Al momento de idear soluciones para asegurar la calidad, conviene dejar de lado la idea de juzgar como buena o mala una descripción bibliográfica; en su lugar, optar por ofrecer a quienes son responsables de los metadatos un panorama lo más completo posible de las principales incidencias o errores al momento de registrar los elementos bibliográficos de un recurso, contribuyendo a la mejora continua y al establecimiento de mejores prácticas.

Agradecimientos

El autor agradece las facilidades otorgadas por la Mtra. María Guadalupe Trinidad Argüello Mendoza para emplear la lista de revistas que conformaron la muestra de esta investigación, así como las motivaciones para escribirla. También a la Mtra. Patricia Garrido Villegas por su apreciación en el orden de los elementos del tablero de datos. ■

REFERENCIAS

- BELLINI, E. y NESI, P. (2013). Metadata Quality Assessment Tool for Open Access Cultural Heritage Institutional Repositories. En P. Nesi y R. Stantucci (eds.), *Information Technologies for Performing Arts, Media Access, and Entertainment: Second International Conference, eclap 2013 Porto, Portugal, April 2013. Revised Selected Papers*. Springer. https://doi.org/10.1007/978-3-642-40050-6_9
- DUSHAY, N. y HILLMANN, D. I. (2003). Analyzing Metadata for Effective Use and Re-Use. *International Conference on Dublin Core and Metadata Application*, 161-170. <https://dcpapers.dublincore.org/pubs/article/view/744>
- ELOUATAOUI, W., EL ALAOUI, I. y GAHI, Y. (2022). Metadata Quality in the Era of Big Bang and Unstructured Content. En Y. Maleh, M. Alazab, N. Gherabi, L. Tawalbeh, A. A. Abd El-Latifs (eds.), *Advances in Information, Communication and Cybersecurity. IC12C 2021. Lecture Notes in Networks and Systems*, vol 357. Springer, Cham. https://doi.org/10.1007/978-3-030-91738-8_11
- GOOGLE. (s.f.). Directrices para el uso de expresiones regulares. <https://support.google.com/a/answer/1346938>
- HARPER, C. A. (2016). Metadata Analytics, Visualization, and Optimization: Experiments in statistical analysis of the Digital Public Library of America (DPLA). *Code4lib Journal*, (33). <https://journal.code4lib.org/articles/11752>
- HUGHES, B. (2004). Metadata Quality Evaluation: Experience from the Open Language Archives Community: 7th International Conference on Asian Digital Libraries, ICADL 2004, Shanghai, China, December 13-17, 2004, Proceedings. En Z. Chen, H. Chen, Q. Miao, Y. Fu, E. Fox y E. Lim (eds.), *Digital Libraries: International Collaboration and Cross-Fertilization* (pp. 320-329). Springer. https://doi.org/10.1007/978-3-540-30544-6_34
- ISENBERG, P., HEIMERL, F., KOCH, S., ISENBERG, T., Xu, P., STOLPER, C. D., SEDLMAIR, M., CHEN, J., MÖLLER, T. y STASKO, J. (2017). Vispubdata.org: A Metadata Collection About IEEE Visualization (VIS) Publication. *IEEE Transactions on Visualization and Computer Graphics*, 23(9), 2199-2206. <http://doi.org/10.1109/TVCG.2016.2615308>
- JACKSON, A. S., HAN, M. J., GROETSCH, K., MUSTAFOFF, M. y COLE, Timothy W. (2008). Dublin Core Metadata Harvested Through OAI-PMH. *Journal of Library Metadata*, 8(1), 5-21. <https://www.researchgate.net/publication/32962578>
- KIM, J., YAKEL, E. y FANIEL, I. M. (2019). Exposing Standardization and Consistency Issues in Repository Metadata Requirements for Data Deposition. *College & Research Libraries*, 80(6), 843-875. <https://doi.org/10.5860/crl.80.6.843>
- KIRÁLY, P. et al. (2019). Evaluating Data Quality in Europeana: Metrics for Multilinguality. En E. Garoufallou, F. Sartori, R. Siatiri, M. Zevras (eds.), *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science*, vol 846. Springer, Cham. https://doi.org/10.1007/978-3-030-14401-2_19
- NICHOLS, D. M., MCKAY, D. y TWIDALE, M. B. (2008). A Lightweight Metadata Quality Tool. En *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries (JCDL '08)* (pp. 25-28). Association for Computing Machinery. <https://doi.org/10.1145/1378889.1378894>
- OCHOA, X. y DUVAL, E. (2009). Automatic evaluation of metadata quality in digital Repositories. *International Journal on Digital Libraries*, 10(2-3), 67-91. <https://doi.org/10.1007/s00799-009-0054-4>
- PALAVITSINIS, N. (2013). *Metadata Quality Issues in Learning Repositories* [Tesis de doctorado, Universidad de Alcalá]. e_Bu@h. Biblioteca Digital Universidad de Alcalá. <http://hdl.handle.net/10017/20664>

- PARK, J. R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly*, 47(3-4), 213-228. <http://dx.doi.org/10.1080/01639370902737240>
- RAMOS ECLEVIA, M., VINZON, M. R., BALBAS, Y. A. y JANIO, R. V. (2019). Creative Ways of Demonstrating Your Value Using a Library Dashboard. *Qualitative and Quantitative Methods in Libraries (QQL)*, 8(4), 535-546. <http://qqml-journal.net/index.php/qqml/article/view/586>
- Shankaranarayanan, B. Z. (2021). Enhancing decision-making with quality metadata. *Journal of System and Information Technology*, 23(2), 199-217. <http://dx.doi.org/10.1108/JSIT-08-2020-0153>
- SHREEVES, S. L., KNUTSON, E. M., STVILIA, B., PALMER, C. L., TWIDALE, M. B. y COLE, T. W. (7 de abril de 2005). *Is "Quality" Metadata "Shareable" Metadata? The Implications of Local Metadata Practices for Federated Collections* [Ponencia]. ACRL Twelfth National Conference, Minneapolis, Minnesota, Estados Unidos de América. <https://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/pdf/shreeves05.pdf>
- TANI, A., CANDELA, L. y CASTELLI, D. (2013). Dealing with metadata quality: The legacy of digital library efforts. *Information Processing & Management*, 49(6), 1194-1205. <https://doi.org/10.1016/j.ipm.2013.05.003>
- TSIFLIDOU, E. y MANOUSELIS, M. (2013). Tools and Techniques for Assessing Metadata Quality. En E. Garoufallou y J. Greenberg (eds.), *Metadata and Semantics Research: 7th International Conference* (pp. 99-110). Springer. https://doi.org/10.1007/978-3-319-03437-9_11

