



Marcalyc: software para la marcación XML JATS para las revistas científicas de acceso abierto diamante

Marcalyc: software for XML JATS tagging for open access scholarly journals published in Ibero-America

Arianna Becerril-García

Universidad Autónoma del Estado de México, México

abecerrilg@uaemex.mx

 <https://orcid.org/0000-0003-0278-8295>

Eduardo Aguado-López

Universidad Autónoma del Estado de México, México

eaguadol@uaemex.mx

 <https://orcid.org/0000-0002-4322-9556>

Alejandro Macedo García

Universidad Autónoma del Estado de México, México

amacedog@uaemex.mx

 <https://orcid.org/0000-0002-8480-5740>

RESUMEN:

La transición de las revistas científicas hacia el XML JATS es muy reciente. La propuesta de Redalyc busca contribuir a la sustentabilidad de la publicación científica de acceso abierto diamante, dotando al editor de conocimiento y tecnología. Marcalyc, es el *software* desarrollado por Redalyc, disponible sin costo alguno para las revistas indexadas en este sistema. Implementa un modelo para la marcación XML y genera los formatos de lectura *PDF*, *HTML*, *ePUB*, visor inteligente y visor móvil. A la fecha, 1082 revistas indexadas en Redalyc han transitado al *XML JATS*. Estos resultados representan un beneficio directo para las instituciones editoras y una de las aportaciones al acceso abierto por parte de Redalyc. El presente trabajo describe la metodología, diseño, desarrollo y resultados obtenidos de Marcalyc.

PALABRAS CLAVE: Redalyc, *XML JATS*, Revistas científicas, Acceso abierto.

ABSTRACT:

The transition of the scholarly journals to XML JATS is very recent. The proposal of Redalyc seeks to contribute to the sustainability of the Diamond Open Access publications, providing the editor with knowledge and technology. Marcalyc, is the software developed by Redalyc, available at no cost to the journals indexed by this system. It implements a model for XML markup and generates the following output formats: PDF, HTML, ePUB, intelligent reader and mobile reader. Currently, 1082 journals indexed by Redalyc have adopted XML JATS. These results represent a direct benefit for publisher institutions and one of the contributions of Redalyc to the Open Access initiative. This paper describes the methodology, design, development and results of Marcalyc.

KEYWORDS: Redalyc, *XML JATS*, Scholarly journals, Diamond Open access.

1. INTRODUCCIÓN

Históricamente, las publicaciones científicas no han podido seguir el ritmo del avance de la tecnología. Ya McKnight (1997, p. 184) apuntaba, desde hace un par de décadas: “hay un cuerpo establecido de investigación relacionada con el diseño de la interfaz de usuario. Sin embargo, los editores han ignorado esta

Recepción: 15 Febrero 2023 | Aceptación: 22 Marzo 2023 | Publicación: 03 Abril 2023

Cita sugerida: Becerril-García, A., Aguado-López, E. y Macedo García, A. (2023). Marcalyc: software para la marcación XML JATS para las revistas científicas de acceso abierto diamante. *Palabra Clave (La Plata)*, 12(2), e179. <https://doi.org/10.24215/18539912e179>



investigación a gran escala, prefiriendo producir revistas electrónicas que dependen de interfaces existentes como Adobe Acrobat”. El autor analizó el éxito de tales interfaces para respaldar tareas comunes de los usuarios, tales como navegar, hojear y leer, concluyendo que es necesaria mayor investigación sobre la aceptabilidad humana.

Así los formatos *HTML* y *PDF* fueron adoptados por la industria editorial como las tecnologías para transitar al ámbito electrónico.

Hitchcock, Carr & Hall (1997) justifican este hecho argumentando que dado que el *HTML* es el lenguaje de formato de la *Web* y que *Postscript* era el estándar de facto para copias impresas desde páginas generadas por computadora, era inevitable que los formatos basados en estos estándares dominaran la producción de revistas electrónicas. *Adobe* desarrolló *Acrobat PDF*, formato que puede producirse a partir de *Postscript* para mejorar la transmisión y distribución de documentos, y que incluía la innovación en términos de las capacidades de compresión de archivos para reducir su tamaño. Es así que, desde entonces y hasta ahora, el *PDF* y el *HTML* han llegado a dominar la producción de publicaciones en línea.

El lenguaje de marcado de hipertexto, mejor conocido como *HTML*, es el lenguaje de marcado principal de la *World Wide Web* (World Wide Web Consortium [W3C], 2018b). Su rápida propagación se debió a la simplicidad de su uso y a la tolerancia a los errores en el etiquetado por parte de los navegadores. Es un lenguaje heredado del *SGML* (World Wide Web Consortium [W3C], 1995), para expresar formato y enfrentar los retos de presentación en la web.

Sin embargo, tanto el *HTML* como el *PDF*, no proveen una forma transparente de procesar el texto completo de las publicaciones científicas ya que combinan datos de formato y presentación con el contenido, de ahí la necesidad de contar con una tecnología que separe el contenido del formato.

El *Extensible Markup Language* (*XML*) nace en su primera versión en 1998 (World Wide Web Consortium [W3C], 2018a) como un formato que provee un conjunto de reglas simples y un método uniforme para describir e intercambiar datos estructurados. Permite describir la estructura y semántica, no el formato en el que se presenta la información.

El *XML* está desempeñando un papel cada vez más importante en el intercambio de una amplia variedad de datos en la *Web* y en otros lugares (World Wide Web Consortium [W3C], 2016). Es un metalenguaje con el cual se pueden crear otros. Cada lenguaje creado a partir del *XML* está diseñado para un propósito específico. Estos lenguajes creados hacen uso de vocabularios.

Un vocabulario para *XML* se expresa a través de un *Schema* o *DTD* (*Document Type Definition*) que formaliza la sintaxis, los conceptos y su jerarquía para ser implementados bajo las reglas del *XML*, siendo útil además para verificar si un *XML* está bien construido y con ello garantizar que la información publicada cumpla con las reglas establecidas. Así como ejemplifican Murray-Rust & Rzepa (2006), una <molécula> o un <organismo> deben ser procesados exactamente de la misma forma sin importar el autor, el lector o el *software* de procesamiento; quien lo implemente debe adherirse a las especificaciones y debe comportarse de una forma predecible.

Entre los vocabularios se encuentran múltiples ejemplos como *MathML* para expresiones matemáticas; *CML*, el lenguaje de marcado para química; *MusicXML*, para música; *WML*, para aplicaciones inalámbricas; *GML* para geografía, y muchos más, incluido *JATS* (*Journal Article Tag Suite*) para la escritura de artículos científicos.

Los antecedentes de *JATS* se remontan a los desarrollos del Centro Nacional para la Información Biotecnológica (NCBI) de la Biblioteca Nacional de Medicina de Estados Unidos (NLM), la cual creó un formato para expresar las citas llamado NLM DTD desde la versión 1.0 hasta la 3.0 publicada en 2008 (National Center for Biotechnology Information [NCBI], 2018).

En 2012, NISO presentó a *JATS* v1.0 como la adaptación de *NLM DTD* v3.1, convirtiéndose en el estándar para marcaje de artículos de revista de publicaciones académicas (National Center for Biotechnology Information [NCBI], 2012).

El objetivo del estándar NISO JATS es definir un conjunto de elementos y atributos *XML* que describen el contenido y los metadatos de artículos de revistas, con la intención de proporcionar un formato común en el que los editores y archivos puedan intercambiar contenido de revistas (National Information Standards Organization, 2015b).

La transición de las revistas científicas en Iberoamérica hacia el *XML*, en específico al estándar JATS, es muy reciente. Por una parte, *SciELO* (*Scientific Electronic Library Online*) promovió el uso de esta tecnología desde 2012, pero comenzó su adopción en la operación de todas sus revistas a partir del año 2015. Las revistas de ciencias de la salud de esta plataforma comenzaron a adoptar la innovación a partir de 2014 (Packer *et al.*, 2014).

Por otra parte, Redalyc (Red de Revistas Científicas de América Latina y el Caribe, España y Portugal),¹ en 2015 se sumó a la adopción del estándar *JATS* con una estrategia basada en la profesionalización de los editores, proporcionando herramientas y conocimientos para hacer el etiquetado *XML* un proceso sostenible con el objetivo de mantener la naturaleza académica de las publicaciones científicas de acceso abierto de la región.

El presente trabajo describe la propuesta tecnológica de Redalyc para la adopción del estándar *JATS* con el formato *XML* para las revistas de acceso abierto editadas en la región iberoamericana.

2. METODOLOGÍA PROPUESTA POR REDALYC PARA LA EDICIÓN DE PUBLICACIONES CIENTÍFICAS DE ACCESO ABIERTO DE IBEROAMÉRICA

La propuesta de Redalyc en torno a la adopción del *XML* busca contribuir a la sustentabilidad de los proyectos editoriales de acceso abierto de Iberoamérica, dotando al editor de conocimiento y *software* para llevar a cabo la publicación electrónica. En 2016 Redalyc lanzó la primera versión de Marcalyc, una herramienta de marcación que permite generar archivos *XML* de acuerdo al estándar *JATS*, así como generar un conjunto de archivos de lectura en diversos formatos.

La incorporación del modelo de publicación basado en *XML JATS* ofrece diversas ventajas y otorga valores agregados al producto editorial: reduce el tiempo de procesamiento, da mayor visibilidad a cada artículo en la web, produce diversos formatos de salida (*HTML*, *PDF*, *ePub*) y posibilita la impresión con el formato de la revista (Aguado-López, Becerril-García & Chávez-Ávila, 2016).

Marcalyc utiliza algoritmos e inteligencia artificial para automatizar una parte importante del proceso de etiquetado *XML*, con lo que ofrece un proceso eficiente donde el usuario supervisa lo inferido automáticamente en lo concerniente a autores, resúmenes, palabras clave, secciones del artículo y, sobre todo, citación.

La adopción de este modelo de publicación trae a las revistas científicas diversas ventajas, entre ellas:

- Cumplimiento de los estándares internacionales de estructuración de artículos científicos.
- Generación automática de los siguientes formatos de lectura para cada artículo publicado: *PDF*, *ePUB*, *HTML*, visor inteligente y visor móvil.
- Interoperabilidad con Redalyc, *SciELO* y otras plataformas.
- Aprovechamiento de lectores multiplataforma como el provisto por *OJS*.
- Interoperabilidad con los agregadores de contenido.
- Retomar el control del proceso y los productos editoriales.
- Garantía de la preservación digital
- Generación automáticamente la versión para impresión.

Marcalyc fue diseñado para que el usuario “marcador” sea un miembro del equipo editorial y no requiere conocimiento informático, con la finalidad de que la institución editora tenga la capacidad de hacerlo *in-house* y pueda evitar la tercerización que en muchos casos eleva los costos de publicación sustancialmente.

Marcalyc traduce la complejidad que conlleva la aplicación del estándar JATS a una interfaz gráfica simple con las siguientes características:

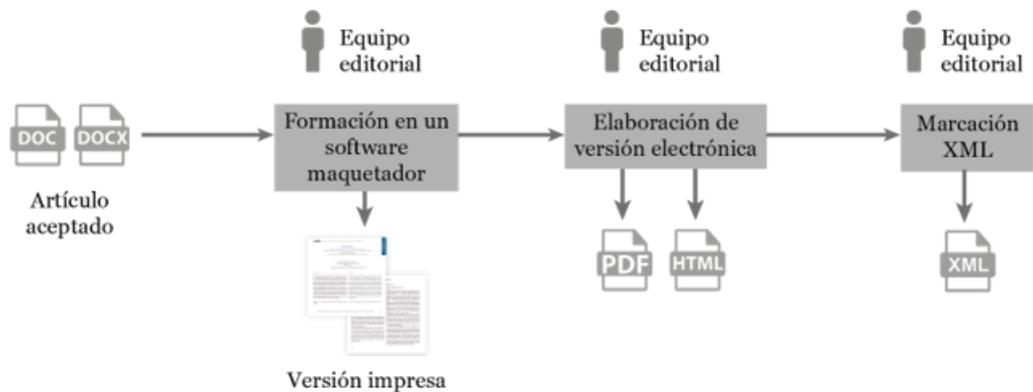
1. Traducción al español de todas las etiquetas del estándar JATS.
2. Sustitución de la acción “copiar / pegar” por “seleccionar texto/soltar” que minimiza los clics en el etiquetado de cada elemento.
3. Manejo de jerarquía y anidación de etiquetas a través de ventanas.
4. Inferencia automática de los siguientes estilos de citación: *ABNT, AMA, APA, ASA, CHICAGO, HARVARD, IEEE, ISO 690 (2007), MLA, NLM y Vancouver*.
5. Inferencia automática de metadatos del artículo.
6. Sugerencia automática de etiquetas acorde a la selección del texto.
7. Aparición de atributos de acuerdo a la etiqueta que se esté usando.
8. Separación automática de texto en párrafos.
9. Etiquetado automático de tablas.
10. Editor especializado de fórmulas y ecuaciones.
11. Separación de los tres apartados de *JATS*: *front* (metadatos del artículo), *back* (referencias y suplementos), *body* (texto completo).
12. Inferencia automática de notas.
13. Detección de falta de relaciones dentro del texto en elementos como menciones a figuras o citas.
14. Detección de errores de marcación a nivel advertencia o error crítico.

El modelo propuesto por Marcalyc conlleva una innovación en la forma tradicional de generar las publicaciones electrónicas. La figura 1 muestra el proceso que generalmente es usado, en el cual, una vez que se concluye la revisión por pares y se obtiene la versión lista para ser publicada el equipo editorial de la revista se encarga de formar la publicación electrónica con algún software como *Adobe InDesign*® diseñando el formato de salida sujeto muchas veces a la versión impresa aunque la revista ya no se imprima.

A partir del resultado de dicha formación se derivan los productos electrónicos ya sea el *HTML* o el *PDF*, heredando las restricciones de la publicación en papel, como imágenes en blanco y negro, enlaces dentro del contenido, paginación, márgenes, ausencia de elementos multimedia como animación, audio o video, entre otras.

Este proceso comúnmente no incluye la publicación de otros formatos que promueven una mejor lectura como el ePUB o el aprovechamiento de tecnologías como *HTML5* para obtener formatos adaptativos y responsivos que promueven la portabilidad, ubicuidad y que favorecen la lectura en los dispositivos móviles.

FIGURA 1
Modelo tradicional de generación de la versión electrónica



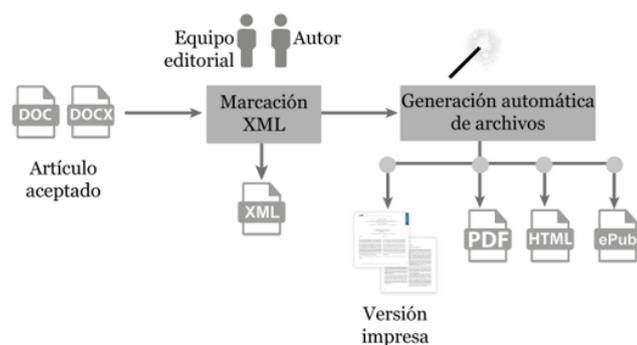
Fuente: elaboración propia.

Este modelo tradicional no aprovecha las potencialidades que brinda el contar con el contenido en XML, archivo a partir del cual es posible generar automáticamente formatos electrónicos de lectura, lo que conlleva una elevación del costo de publicación. Además no ayuda a la visibilidad y el descubrimiento de la revista en la *Web* e ignora los beneficios de las tecnologías de información y comunicación actuales a favor de la accesibilidad, la lectura y el consumo del contenido científico.

El diseño de Marcalyc (mostrado en la figura 2), contempla dos principios fundamentales: 1. Que además del equipo editorial de la revista, el autor pueda participar en el proceso; 2. Que el perfil de usuario de Marcalyc no demande conocimientos especializados en XML o JATS. Estas características permitirán bajar los costos de la edición electrónica de las revistas científicas al brindar la posibilidad que el autor colabore en la generación del XML al momento del envío de la versión final o de que el equipo interno de la revista lo realice evitando la tercerización del proceso.

Con ello Redalyc busca aportar en la sustentabilidad de los proyectos editoriales de acceso abierto, para que permanezcan en manos de la academia y para que eviten implementar modelos comerciales como *cobrar por publicar* (*article processing charges o APC*).

FIGURA 2
Proceso de generación de una revista electrónica a partir del XML



Fuente: elaboración propia.

En este nuevo modelo se inicia la marcación del *XML* y a partir de ella se generan automáticamente todos los elementos que componen la revista electrónica como el *PDF*, *HTML* y *ePUB*, inclusive la versión impresa de la revista si es necesaria. El apartado siguiente muestra la implementación de este modelo en el *software* Marcalyc.

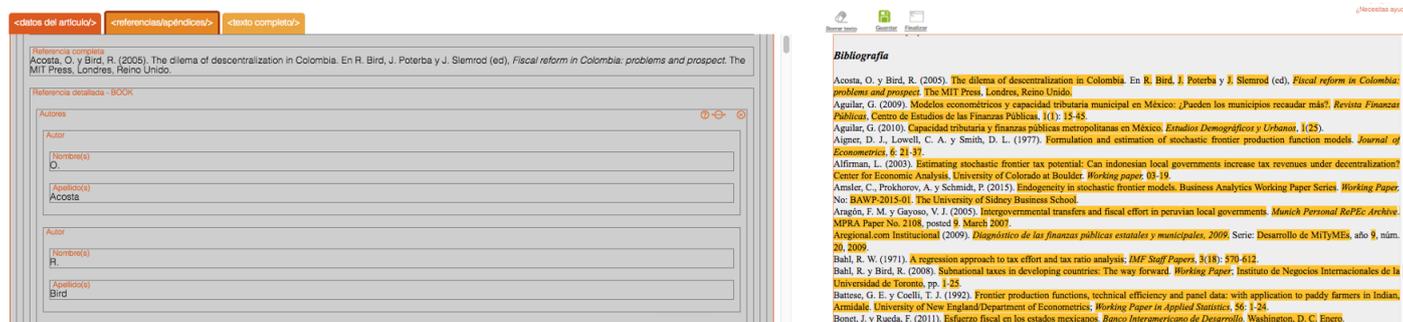
3. MARCALYC, TECNOLOGÍA DESARROLLADA POR REDALYC PARA LA MARCACIÓN XML BAJO EL ESTÁNDAR JATS

3.1. Descripción general

Marcalyc es un desarrollo de *software* que funciona en un ambiente *web* a través de un navegador (*Firefox*, *Chrome*, *Opera* o *Safari* en sus versiones recientes, no es compatible con *Internet Explorer*), probado en múltiples sistemas operativos como *Linux*, *MacOS* y *Windows*; no requiere de ningún *software* o *plugin* adicional para su ejecución.

1. Los archivos XML son generados bajo el conjunto de caracteres Unicode.
2. *MathML*, para expresar e intercambiar información matemática.
3. *XLink*, usado para enlazar recursos en la web.
4. *JATS* (National Information Standards Organization, 2015a).

FIGURA 3
Interfaz de usuario de Marcalyc



Fuente: elaboración propia.

La interfaz de usuario de Marcalyc (Figura 3) fue diseñada para mostrar un área de trabajo permanente (que no abra ventanas o salte a otras aplicaciones o páginas) y consiste en una página dividida en dos paneles y una barra superior. El panel derecho muestra el texto original del artículo, el izquierdo muestra -dividido en tres pestañas- los elementos de *JATS*. El procedimiento para etiquetar se realiza dando clic en alguna caja del panel izquierdo (equivalente a una etiqueta *JATS*) y posteriormente seleccionando el texto del lado derecho, al soltar el ratón automáticamente ese texto se copiará en la etiqueta correspondiente.

La barra superior es sensible al contexto de la marcación y mostrará únicamente los atributos y propiedades correspondientes a la etiqueta que se esté marcando, ayudando así al usuario a ubicar tales datos y a evitar errores comunes de la marcación.

3.2. Metodología del proceso de marcación XML

El diagrama de la figura 4 muestra el proceso que un usuario debe llevar para la marcación de un artículo en Marcalyc. En él se identifican básicamente 4 fases:

1. Preparación para la marcación.
2. Marcación de cabecera o *front*.
3. Marcación de referencias y suplementos finales o *back*.
4. Marcación de texto completo o *body*.

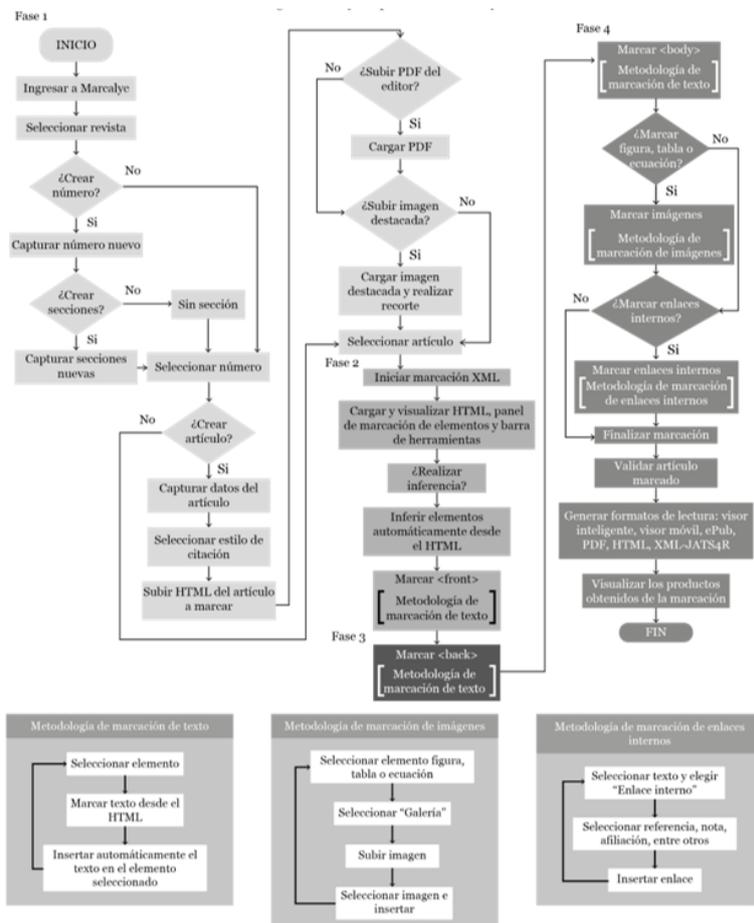
En la fase 1 se elige la revista a la que pertenece el artículo a etiquetar y se capturan datos generales del número en el cual fue publicado. En este punto también se recomienda indicarle al sistema el estilo de citación que sigue el artículo, ello con el fin de optimizar los resultados de la inferencia automática. Finalmente, se carga el archivo correspondiente al artículo que se va a etiquetar. Marcalyc requiere que este sea en formato *HTML* usando la codificación *UTF-8*, el cual deberá de compactarse en un archivo *ZIP* junto con sus imágenes en el caso que las contenga.

La marcación de la cabecera del artículo consiste en el etiquetado de los elementos correspondientes a la sección llamada *front* del estándar *JATS*, como títulos, autores, afiliaciones institucionales, resúmenes, palabras clave, entre muchos otros. En esta fase 2, el sistema realiza las primeras inferencias para identificar dichos metadatos del artículo y con ello realiza un llenado automático de información en las etiquetas correspondientes para que el usuario verifique y complemente.

En la fase 3 se realiza el etiquetado de referencias, suplementos y apéndices. Como se mencionó anteriormente Marcalyc realiza una identificación automática -basada en expresiones regulares- de los elementos de cada referencia si estas se encuentran estructuradas bajo uno de los estilos de citación que el sistema reconozca. De lo contrario se sigue un proceso de marcación con el método de seleccionar/soltar como en la fase 2.

Finalmente, en la fase 4, Marcalyc realiza automáticamente para la marcación del texto completo la separación del mismo en párrafos, el etiquetado de tablas, identificación de notas, así como la detección de falta de relaciones dentro del texto en elementos como menciones a figuras o citas. Además implementa un editor especializado de fórmulas y ecuaciones que permite codificar dichos elementos con el lenguaje *MathML*. Dentro de este etiquetado se deben cargar las imágenes del artículo a través de una galería e ir creando los enlaces internos en el artículo, como son, menciones a las referencias, notas, figuras, entre otros.

FIGURA 4
Metodología para el proceso de marcación XML de Marcalyc

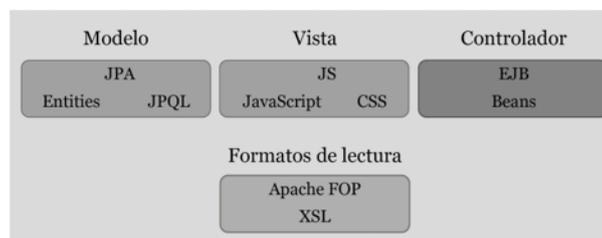


Fuente: elaboración propia.

3.3. Arquitectura de Marcalyc

El sistema de marcación Marcalyc fue desarrollado siguiendo el patrón de arquitectura Modelo-Vista-Controlador, que permite separar los datos y la lógica del negocio de la interfaz de usuario y del módulo gestor de eventos. Para ello se utilizaron diferentes marcos de trabajo *JavaEE*.

FIGURA 5
Arquitectura de Marcalyc



Fuente: elaboración propia.

El modelo se construyó utilizando *Java Persistence API (JPA)*, el cual proporciona un modelo de persistencia de la información para generar mapeos de bases de datos relacionales en Java. Para la vista se utilizó *Java Server Faces (JSF)* junto con *JavaScript* y *CSS*, tecnologías que permitieron lograr la interacción entre los paneles. Y el controlador fue desarrollado con *Enterprise JavaBeans (EJB)* para el manejo de la comunicación entre los módulos.

3.4. Compatibilidad

El archivo XML generado usando la herramienta Marcalyc cumple con las especificaciones del estándar *JATS*, así se puede verificar haciendo uso del validador del estándar. Ahora bien, Marcalyc genera adicionalmente un archivo XML compatible con *JATS for Reuse* mejor conocido como *JATS4R* (*JATS4R Group*, 2018).

JATS4R, una iniciativa formada por un grupo de trabajo donde participan organizaciones como *Crossref*, *PLOS*, *eLIFE*, *Cambridge University Press*, entre otras, establece una serie de recomendaciones a seguir sobre el uso del estándar *JATS* con la finalidad de optimizar la reusabilidad del contenido académico etiquetado con *XML JATS* y con ello hacer más eficientes los procesos de interoperabilidad, intercambio, almacenamiento y recuperación de información.

3.5. Soporte para licencias Creative Commons

Marcalyc cuenta con las licencias *Creative Commons* precargadas en el sistema, de tal forma que el usuario seleccione alguna de ellas, con las diferentes combinaciones de las características a configurar para este tipo de licencias, las cuales son: reconocimiento, uso comercial, obra derivada y compartir igual.

De esta forma no solo los archivos XML incluyen esta información, sino que todos los formatos de salida para lectura como *PDF*, *HTML*, *ePUB*, visor inteligente y visor móvil incorporan el código de la licencia y metadatos respectivos, haciendo estos archivos adecuadamente localizables por buscadores de recursos con licenciamiento. Además, incorpora la opción para agregar otro tipo de licenciamiento en caso de ser necesario.

3.6. Generación formatos de lectura

El sistema incluye un módulo de generación de diferentes formatos de lectura que están disponibles en Redalyc pero que también están preparados para ser descargados para que los editores puedan subirlos al sitio web de la revista.

Para la generación de estos archivos se realiza primeramente una transformación XSLT (Extensible Stylesheet Language Transformation) (Refsnes Data, 2018), la cual permite traducir el XML al formato especificado por cada tipo de archivo para visualización de la información. Posteriormente se adiciona código *HTML5*, *JavaScript* y *CSS* junto con rutinas de *Apache FOP* (Formatting Objects Processor) para generar componentes y eventos que dotan de funcionalidad e interactividad los formatos de salida.

La tecnología de *Apache FOP* se utiliza en Marcalyc para generar el formato *PDF*. Apache FOP es un formateador de impresión para objetos *XSL* (*XSL-FO*) y un formateador independiente de salida. Es una aplicación Java que lee un árbol de objetos de formato y procesa las páginas resultantes a un resultado específico (Apache Software Foundation, 2016). Es así como cada artículo marcado en Marcalyc genera los siguientes seis formatos de salida: *XML JATS* (cumpliendo la especificación *JATS4R*), *PDF*, *HTML*, *ePUB*, Visor inteligente de artículos y Visor móvil de artículos. La tabla 1 resume las principales características de ellos.

TABLA 1
Formatos de salida de Marcalyc

Formato de salida	Versión	Características
<i>XML JATS</i>	<i>JATS</i> versión 1.1d3 Especificación <i>JATS4R</i>	Se descarga en un archivo comprimido ZIP que contiene el archivo XML y los archivos de todas las imágenes del artículo. Este archivo cumple con el validador XML de PMC. ²
<i>PDF</i>	<i>Portable Document Format 1.4</i>	Dos formaciones de texto a elegir. Una formación a una columna y otra a dos. Incluye los datos de la revista, número, volumen y año en la portada, así como todos los datos de identificación del artículo.
<i>HTML</i>	<i>HTML5</i>	<i>HTML5</i> incorpora nuevos elementos y atributos al <i>HTML</i> tradicional, elementos que brindan mayor potencial a las revistas científicas tanto en visualización de contenidos como en recuperación de información. Por ejemplo: <ul style="list-style-type: none"> - Incorpora etiquetas (canvas 2D y 3D, audio, vídeo) con codecs para mostrar contenidos multimedia. - Etiquetas para manejar conjuntos de datos de gran dimensión, generar tablas dinámicas que pueden filtrar, ordenar y ocultar contenido. - Posibilidad de uso de visores de <i>MathML</i> (fórmulas matemáticas) y <i>SVG</i> (gráficos vectoriales). - Las etiquetas permiten describir el significado del contenido, esto permite a los buscadores interpretar mejor la información, lo que incrementa su potencial de descubrimiento y visibilidad.

ePUB	EPUB Content Documents version 3.0.1 (International Digital Publishing Forum™, 2014)	El formato ePUB es de código abierto para leer texto e imágenes. Con esta versión las revistas también pueden introducir audio en los artículos. Se adapta a distintos tamaños de letra y de pantalla, además de aprovechar todos los beneficios de los lectores de libros electrónicos.
Visor inteligente de artículos	HTML5, J2EE, Javascript, CSS	Es una versión web enriquecida para archivos XML-JATS cuyas principales características son: - Permite la lectura del documento completo o por partes, de acuerdo con las preferencias del usuario (secciones, referencias, resúmenes, entre otras). - Acceso directo a servicios ofrecidos por Redalyc y a la ficha biblio-hemerográfica del artículo. - Es responsivo y adaptativo. Se adapta a cualquier dispositivo y pantalla. - Personalización de la apariencia del visor a través de temas predeterminados. - Se permite la navegación entre páginas como un libro electrónico e incluye un buscador de texto dentro del artículo. - Vista previa y enlace a referencias y notas al pie, desde el texto del documento. - Visor de imágenes que permite ampliar y descargar una o todas las imágenes en diferentes formatos.
Visor móvil	HTML5, J2EE, Javascript, CSS	Una versión web adicional diseñada para la lectura en dispositivos móviles. Incluye una forma de navegación entre secciones colapsables a través de barras que favorece la interacción touch de los dispositivos, así como también incluye un comportamiento distinto para las imágenes y acciones del visor inteligente de artículos.

Fuente: elaboración propia.

4. RESULTADOS

La estrategia de Redalyc para la transición hacia esta tecnología ha sido bajo dos vertientes:

1. Desarrollo de Marcalyc y profesionalización de editores, para dotar a las revistas de un sistema de marcación que abone a la sustentabilidad de la revista y transfiera conocimiento a los equipos editoriales.
2. Importación de archivos XML JATS que los editores hayan generado con herramientas distintas a Marcalyc como SciELO u otras, siempre que cumplan con el estándar JATS. Con ello, se busca evitar que la revista realice una marcación más y se aprovecha lo que ya haya trabajado para otras plataformas. Esta opción obliga a Redalyc a realizar un procesamiento adicional que toma el 200%

del tiempo de lo que toma procesar un archivo generado por Marcalyc, tanto en recursos humanos como informáticos. Este costo adicional lo asume Redalyc en aras de contribuir a la sustentabilidad de las publicaciones de acceso abierto de la región.

A la fecha, 1082 revistas indexadas en Redalyc han transitado al *XML JATS* a través de las estrategias descritas, 429 haciendo uso de Marcalyc y 329, de forma híbrida, con Marcalyc y enviando de manera complementaria sus archivos *XML JATS* realizados en otras plataformas.

Además, 324 revistas envían sus archivos *XML JATS* que han sido generados con otras herramientas y son procesados por Redalyc, lo que implica acciones adicionales por parte del equipo interno de Redalyc para importar los *XML* realizados con otros sistemas. Ejemplo de estas implicaciones son:

- Se asigna a cada artículo las claves correspondientes al identificador de la revista y del número, es decir, no se le pide al editor que incluya ningún dato relacionado a Redalyc.
- Se verifica el *ISSN*, así como la nomenclatura de volumen y número de la revista.
- Se comprueba la existencia de las figuras mencionadas en el *XML*.
- Se corrigen errores de marcación (por ejemplo: marcación de notas a pie de página o información de financiamiento).
- Se pasa cada *XML* por los validadores del estándar *JATS* y la recomendación *JATS4R*.

Marcalyc brinda un servicio totalmente gratuito a través de un modelo cooperativo, donde Redalyc brinda la tecnología y la capacitación y los editores realizan el proceso de marcación. Un trabajo colaborativo que ayuda a la sostenibilidad de las revistas de acceso abierto diamante.

CONCLUSIONES Y TRABAJO FUTURO

Con el desarrollo de Marcalyc y la estrategia de apoyo a las revistas científicas de acceso abierto diamante en su transición al *XML JATS*, Redalyc ha logrado beneficiar a cientos de revistas editadas por otro tanto de instituciones, fortaleciendo sus equipos editoriales internos y absorbiendo los costos del desarrollo y el mantenimiento de esta tecnología, para que los editores no tengan que subcontratar la marcación *XML JATS*, una acción a favor de las revistas científicas de acceso abierto diamante.

La adopción del *XML JATS* ha representado un reto para las revistas científicas en general, ya sea del ámbito comercial o académico. Por un lado, debido a que las instituciones editoras no han desarrollado las competencias necesarias para una plena transición digital de las revistas científicas. Por otro lado, aún no es reconocido el potencial de uso de estas tecnologías para la visibilidad y capacidad del contenido científico de ser descubierto y aplicado, para la sostenibilidad de la actividad editorial, eficiencia y costos de producción de revistas y para la evolución de la cultura evaluativa alrededor de las revistas científicas.

El *XML JATS* abre un campo de estructuración de datos para ir más allá de las métricas basadas en citación para valorar la producción científica. Más aún, permite entender la estructura de composición del tejido científico y es en esta justa medida que su aportación debe ser valorada y su implementación debe ser promovida.

REFERENCIAS

- Aguado-López, E., Becerril-García, A. & Chávez-Ávila, S. (2016). *Conectando al Sur con la ciencia global. El nuevo modelo de publicación en ALyC, no comercial, colaborativo y sustentable*. Recuperado de <https://xmljatsredalyc.org/2016/08/12/el-nuevo-modelo-de-publicacion-alyc-xml-jats/>
- Apache Software Foundation. (2016). *Apache™ FOP*. The Apache™ FOP Project. Recuperado de <https://xmlgraphics.apache.org/fop/>

- Hitchcock, S., Carr, L. & Hall, W. (1997). Web journals publishing: a UK perspective. *Serials: the journal for the serials community*, 10(3), 285-299. <https://doi.org/10.1629/10285>
- International Digital Publishing Forum™. (2014, June 26). *EPUB content documents 3.0.1*. International Digital Publishing Forum. Recuperado de <https://idpf.org/epub/301/spec/epub-contentdocs.html>
- JATS4R Group. (2018). *JATS4R*. JATS4R. JATS for reuse. Recuperado de <https://jats4r.org/>
- McKnight, C. (1997). Designing the electronic journal: why bother? *Serials*, 10(2), 184-188. <https://doi.org/10.1629/10184>
- Murray-Rust, P. & Rzepa, H. S. (2006). Scientific publications in XML - towards a global knowledge base. *Data science journal*, 1, 84-98. <https://doi.org/10.2481/dsj.1.84>
- National Center for Biotechnology Information [NCBI]. (2012, September 14). *NLM Journal Archiving and interchange tag suite Version 1.0*. NLM Journal archiving and interchange tag suite. Recuperado de <https://dtd.nlm.nih.gov/1.0/>
- National Center for Biotechnology Information [NCBI]. (2018). *JATS and the NLM DTDs*. Journal article tag suite. Recuperado de <https://jats.nlm.nih.gov/about.html>
- National Information Standards Organization. (2015a, April 14). *JATS: version 1.1d3*. Recuperado de <https://jats.nlm.nih.gov/1.1d3/>
- National Information Standards Organization. (2015b, November 19). *JATS: journal article tag suite, ANSI/NISO Z39.96-2015*. Recuperado de <http://jats.niso.org/1.1/>
- Packer, A. L. *et al.* (2014, April 4). ¿Porqué XML?. *SciELO En Perspectiva*. Recuperado de <https://blog.scielo.org/es/2014/04/04/porque-xml/>
- Redalyc. (2018). *Aprende a usar Marcalyc*. Sistema de Información Científica Redalyc. Recuperado de <http://marcalyc.redalyc.org/ayuda/>
- Refsnes Data. (2018). *XSLT Introduction*. W3Schools. Recuperado de https://www.w3schools.com/xml/xsl_intro.asp
- World Wide Web Consortium [W3C]. (1995, November). *Overview of SGML Resources*. W3C. Recuperado de <https://www.w3.org/MarkUp/SGML/>
- World Wide Web Consortium [W3C]. (2016, October 11). *Extensible Markup Language (XML)*. Recuperado de <https://www.w3.org/XML/>
- World Wide Web Consortium [W3C]. (2018a). *Extensible Markup Language (XML) 1.0 (Fifth Edition) Publication History—W3C*. W3C. Recuperado de <https://www.w3.org/standards/history/xml>
- World Wide Web Consortium [W3C]. (2018b). *W3C HTML*. W3C. Recuperado de <https://www.w3.org/html/>

NOTAS

- 1 Para más información: <https://www.redalyc.org/>
- 2 Validador de PMC: <https://www.ncbi.nlm.nih.gov/pmc/tools/xmlchecker/>