

Estandarización XML-JATS. Un modelo para la publicación en acceso abierto

Julio Zetter Patiño¹

Resumen

Actualmente, la transición a la era digital se ha convertido en uno de los temas más comentados y abordados en distintos rubros del universo de la producción de información. Varias propuestas, tanto de acceso abierto como comerciales, han emergido como opciones para lograr una transición integral de textos científicos al plano digital. Entre esas opciones surgió el estándar NISO-JATS, que utiliza para su construcción el metaleguaje XML como vocabulario.

El estándar XML-JATS está enfocado principalmente en la descripción de documentos ligados a los procesos de edición y publicación de revistas científicas. Contiene las instrucciones que regulan la constitución de, por ejemplo, editoriales, artículos originales, artículos de revisión, ensayos clínicos, cartas/respuestas, erratas, y todo lo que tenga que ver con el contenido específico y general de este tipo de publicaciones.

Esta ponencia tendrá como objetivo principal dar a conocer, a grandes rasgos, el contexto en el que se han desarrollado las políticas de acceso abierto, las cuales han propiciado que se generen esquemas de metadatos para la catalogación y preservación de información digital. También se hablará del papel que está jugando en esta dinámica político-editorial el estándar XML-JATS. Para lograr esto, se esbozarán la estructura y propiedades del esquema de metadatos XML-JATS para la catalogación y publicación de artículos científicos, al igual que se expondrán ejemplos reales sobre su aplicación y alcance.

Palabras clave: Acceso abierto; XML (eXtensible Markup Language); JATS (Journal Article Tag Suit); Metadatos, Edición digital; Revistas científicas.

Abstract:

Currently the transition to the digital era has become one of the topics most discussed and addressed in different areas that cover the production of information. Several proposals, both open access and commercial, have emerged as options to achieve a complete transition of scientific texts to the digital plane. Among these options, there is the NISO-JATS standard, which uses the XML meta-language as a vocabulary for its construction.

The XML-JATS standard is mainly focused on the description of documents that are linked to the editing and publication processes of scientific journals. It contains the instructions that regulate the constitution, for example, of editorials, original articles, review articles,

¹ Julio Zetter Patiño es licenciado en Sociología por parte de la Universidad Autónoma Metropolitana-Unidad Xochimilco (México) y maestro en Bibliotecología y Estudios de la Información por parte de la Universidad Nacional Autónoma de México. Actualmente, labora en la Universidad Nacional Autónoma de México como parte de la Dirección General de Bibliotecas/Departamento de Bibliografía Latinoamericana en donde se desempeña como miembro del equipo de coordinadores editoriales de la base de datos y hemeroteca virtual SciELO México. Su línea de investigación es el uso práctico de metadatos para la edición y publicación científica en acceso abierto. CE: juliozp@dgb.unam.mx.

clinical trials, letters/answers, errata, and everything that has to do with the specific and general content of this type of publications.

The main objective of the paper is to make known the context in how the open access policies have been developed and how they have led to the creation of metadata schemas for cataloging-preservation of digital information and the role of the standard XML-JATS that is playing in this political-editorial dynamic.

For this, the structure and properties of the XML-JATS metadata schema will be outlined for the cataloging and publication of scientific articles as well to show real examples of its application and scope.

Key words: Open access; XML (eXtensible Markup Language); JATS (Journal Article Tag Suit); Metadata; Digital publishing; Scientific journals.

Introducción

Actualmente, la *transición a la era digital* se ha convertido en uno de los temas más comentados y abordados en distintos rubros del universo que abarca la producción de información. Esta transición va acompañada de cierto miedo o *incertidumbre*, derivado del desconocimiento de políticas e innovaciones que se han venido desarrollando en el campo de las ciencias de la información.

Como punto de vista personal, la transición a lo digital *ya no es una opción*: es una obligación. Las revistas científicas, como fuentes de información, ya no están exentas de verse involucradas en estos cambios. Diversos actores están dentro del circuito de producción y todos son igualmente importantes para que los resultados cumplan exigencias y estándares de calidad necesarios para lograr una credibilidad científica.

Actualmente, la premisa a seguir es que *la información científica producida con ayuda de recursos públicos sea gratuita y accesible para toda la sociedad*. Esto ha ido aún más allá al hacer posible su diseminación global por medio de la Internet.

La globalización y apertura de la información no se vuelven una realidad por el simple hecho de estar dentro de la web. Han sido necesarios debates, declaraciones, acuerdos y estatutos de la misma envergadura para proponer y generar directrices que posibiliten el acceso libre de datos dentro de los distintos cúmulos de información.

Un resultado directo de haber establecido acuerdos de cooperación científica, es la creación de estándares o modelos para la organización, recuperación y publicación en acceso abierto, es el caso del XML-JATS.

Elementos medulares que componen y conforman el modelo XML-JATS

Los *metadatos*, por definición, describen las características específicas de la información que los documentos poseen, y que son necesarios para su recuperación y su publicación digital. Al proceso de identificación de metadatos se le llama *etiquetación* (o *metaetiquetación*) y emplea elementos conocidos como etiquetas (*tags*). Cada una de estas etiquetas actúa como descriptor, que puede ser susceptibles de normalización.

Para sistematizar estas etiquetas es necesario contar con *esquemas de metadatos* (*schema*). Éstos pueden definirse como un conjunto de metadatos con elementos y reglas, a los cuales se les ha asignado un propósito en particular. Los esquemas de metadatos responden a una serie de necesidades con las que, por ejemplo, cierta disciplina se identifica. Puede haber metadatos tanto generales como específicos. Por ejemplo:

- Esquemas generales: Dublin Core Metadata Element Set
- Esquemas específicos: JATS (Journal Article Tag Suite)

Prácticamente, puede haber tantos esquemas de metadatos como disciplinas, campos o necesidades para organizar y recuperar cierta información existan (por ejemplo, elementos matemáticos, objetos culturales, taxonomías biológicas, comida, entre otros).

El modelo XML-JATS responde a las necesidades de recuperación y publicación en el ambiente digital y de Acceso Abierto. Está conformado, como su nombre lo indica, por dos elementos correlacionados entre sí: el *metalenguaje XML* y el *esquema de metadatos JATS*.

Metalenguaje XML

Para que los metadatos puedan ser funcionales necesitan un *metalenguaje* que pueda objetivarlos digitalmente con el fin de que las máquinas sean capaces de interpretarlos y de que los cruces semánticos y sintácticos sean posibles. Los *metalenguajes*, en este sentido, no son otra cosa que un conjunto de reglas "que sirven para codificar un documento mediante un conjunto de etiquetas" (Leiva 2008, 96). Los metalenguajes "suministran una mayor semántica y estructuración de los documentos, permitiendo más opciones de trabajo con los objetos (datos) y los metadatos" (Senso y Rosa-Piñero 2003, 101).

Los metalenguajes más utilizados son el SGML (Standard Generalized Markup Language 1986) y el XML (eXtensible Markup Language), que se volvió recomendación del W3C² en febrero de 1998. Este último derivó del primero.

En el XML, la codificación de metadatos pretende ser más sencilla y más flexible que en su antecesora. Estas cualidades pueden ser una gran ventaja sobre el SGML, aunque también pueden convertirse en un serio problema si se carece de una estandarización, ya que cualquier persona puede crear sus propios conjuntos de etiquetas para describir cualquier cosa. Algunas de las características principales del metalenguaje XML son las siguientes:

- Es un formato sumamente flexible, que puede moldearse con facilidad y que prácticamente puede ajustarse a cualquier necesidad de estructuración de metadatos, se trate o no de documentos con carácter científico.
- Utiliza códigos que pueden ser interpretados tanto por máquinas como por humanos.
- Fue desarrollado con el fin de transportar datos sin depender de la visualización de los documentos³.
- El conjunto de instrucciones XML no hace nada por sí sola. Se necesita, por fuerza, un estándar que decreta cuál será la forma en la que cierto conjunto de etiquetas adquirirá un sentido y un propósito.

Esquema de metadatos JATS

JATS Journal Article Tag Suite (Paquete de Etiquetas para Artículos Científicos) está basado en un paquete de etiquetas que estableció en el año 2000 la National Library of Medicine (NLM) de Estados Unidos. Ésta, a través del National Center for Biotechnology Information (NCBI), desarrolló un paquete de publicaciones en XML para la entonces experimental base de datos PubMed Central (PMC) (Huh 2014,100). El esquema actual nació en agosto de 2012. Su versión estable más reciente se publicó en el año 2015.

² Consorcio formado en 1994, que emite recomendaciones y estándares para el correcto funcionamiento de la web.

³ En contraste, el metalenguaje HTML está enfocado en la forma en que se van visualizar los datos en cuanto a la estética de la publicación.

Desde su creación, el JATS fue respaldado por los miembros de un conjunto de comités que provinieron de distintas ramas disciplinarias, tanto de instituciones públicas como privadas. La constitución de estos comités tuvo destacadamente carácter universal. Hubo miembros de la American Library Association (ALA), American Psychological Association (APA), CrossRef, EBSCO Information Services, Library of Congress, ProQuest, Public Library of Science (PLoS), SAGE Publications, OCLC Online Computer Library Center, Microsoft Corporation (NISO, 2015, p. IV-V), por mencionar algunas instituciones involucradas en su desarrollo. Podemos apreciar que la mayoría de las personas aquí mencionadas pertenecen a ramas que atienden el área de la bibliotecología, ciencias de la información, editoriales, revistas de acceso abierto, así como desarrolladores que están involucrados en el campo de la web semántica y catálogos digitales.

El estándar está compuesto por tres módulos (DTD's específicas) principales de etiquetas:

- Archiving and Interchange Tag Set (módulo de etiquetas para archivamiento e intercambio)
- Journal Publishing Tag Set (módulo de etiquetas para publicaciones científicas)
- Article Authoring Tag Set (módulo de etiquetas para escribir artículos)

Inicialmente el esquema surgió del campo médico. Después se aplicó a la constitución de artículos de todo tipo de disciplinas. La metodología de estructuración de los documentos XML construidos con este estándar responde al modelo de formación de artículos de ciencias médicas y exactas. Por ejemplo, puede sea aplicada en la necesidad de resúmenes y palabras clave en artículos originales o la identificación de secciones normalizadas en los artículos como Introducción, Metodología, Discusión, Resultados y Conclusiones (conocida como estructura IMRyD), entre otros elementos.

El JATS utiliza como vía de codificación de metadatos al metalenguaje XML, de ahí que identificamos a este modelo conjunto como XML-JATS. Actualmente, este modelo no sólo está siendo utilizado para la codificación de documentos de las ciencias médicas, sino de los documentos que están ligados a los procesos de edición y publicación de revistas científicas de todas las disciplinas. En realidad, contiene las instrucciones que regulan la constitución de, por ejemplo, editoriales, artículos originales, artículos de revisión, ensayos

clínicos, cartas/respuestas, erratas, y todo lo que tenga que ver con el contenido específico y general de este tipo de publicaciones.

Cabe destacar que tanto el metalenguaje XML como el esquema JATS son proyectos desarrollados en código abierto, así como con base en políticas sustentadas por el acceso abierto. Al estar en acceso abierto, este modelo puede ser modificable para cubrir ciertas necesidades particulares que cada base de datos que la utilice pudiera tener. Por ejemplo, permite la inclusión de otros esquemas de metadatos, como el MathML (para descripción de fórmulas matemáticas) o modelos para la construcción de tablas, como el XHTML.

Podemos mencionar algunos puntos importantes que el modelo XML-JATS se enfoca en atender:

- La descripción
- La preservación
- La publicación de documentos.

En la cuestión de la descripción el modelo utiliza tres componentes principales: elementos, atributos y valores.

```
<kwd-group xml:lang="es">  
  <title>Palabras clave:</title>  
  <kwd>seguridad alimentaria</kwd>  
  <kwd>comunidades rurales</kwd>  
  <kwd>desigualdad</kwd>  
  <kwd>grupos desplazados</kwd>  
</kwd-group>
```

En cuanto a la parte de preservación, se puede decir que en el modelo:

- El contenido está apartado de su presentación; es decir, la información puede quedar preservada dentro de las etiquetas, sin necesidad de que dependa de un recipiente en específico (la información puede plasmarse hasta dentro de un bloc de notas).

- Al ser de Acceso Abierto, la información tampoco va a depender de *software* específicos o de su compatibilidad con diferentes Sistemas Operativos.

Publicación:

- Los documentos construidos con lenguaje XML pueden ser fácilmente interpretados por hojas de estilo que estructuren estéticamente la información y le den distintos diseños.
- La flexibilidad del XML permite la estructuración de sus datos en otros formatos de publicación como PDF, HTML o ePub.
- Una de las características más importantes del JATS versión 1.0 (2012) es que permite la identificación y aceptación de todos los caracteres de los lenguajes que se utilizan en todo el mundo. (Huh 2014)

¿Porque el uso de metadatos estandarizados es importante para el acceso abierto?

Los metadatos, además de soportar todos los elementos necesarios para descubrir fuentes de información de manera efectiva, poseen el rol adicional de informar el estatus de un recurso disponible en acceso abierto (UNESCO 2015). Esto es importante, por ejemplo, para el campo de minería de datos o el intercambio de información entre personas y sistemas.

Es de suma importancia tanto para los profesionales de la información como para la comunidad de usuarios en general, conocer la situación en la que los documentos en acceso abierto pueden ser consultados, reproducidos y utilizados. Por ejemplo, si todos los artículos científicos en un repositorio digital dado utilizan las mismas etiquetas y las mismas reglas, se pueden hacer búsquedas de manera consistente, utilizar *un solo software* para gestionar dicho repositorio, y preservar el contenido intelectual de los años por venir (NCBI 2012).

Por ejemplo, la información sobre el manejo de licencias y derechos está registrada dentro de los *metadatos administrativos* de los documentos:

```
<permissions>
  <license license-type="open-access"
    xlink:href="http://creativecommons.org/licenses/by-nc/4.0/" xml:lang="es">
    <license-p>Este es un artículo publicado en acceso abierto bajo una licencia
      Creative Commons</license-p>
  </license>
</permissions>
```

Los sistemas de almacenamiento, repositorios, directorios, bases de datos, etcétera, recuperan esta información de manera automática, lo que hace posible su catalogación adecuada, dependiendo de los tipos de licencia que posean.

Algunos proyectos que utilizan metadatos del modelo XML-JATS son:

- Hemeroteca SciELO (Abarca toda la red Scielo, 16 países y bases de datos especializadas por disciplinas)
- Repositorio Redalyc

Journals:

- PlosOne
- JoVE (Video Journal)
- RIO (Research Ideas and Outcomes)

La organización multidisciplinaria y multinacional JATS FOR REUSE se enfoca, principalmente, en generar recomendaciones para la reutilización de contenido académico y científico marcado con el estándar JATS.

Reflexión final

El editor, como uno de los actores responsables de garantizar la calidad de las fuentes de información que produce, es el que procura el trabajo desde que se concibe hasta que se expone a la comunidad científica. Su labor involucra distintos campos disciplinarios, los cuales convierten a esta profesión en uno de los desempeños interdisciplinarios más complejos dentro del campo de la producción del conocimiento.

Las funciones que ejercen los editores van desde establecer temas, supervisar, administrar, revisar, calificar, publicar y difundir. A ellas se suma ahora suma la tarea de implementar

tecnologías de la información y comunicación en dichos procesos. Con miras hacia el futuro para la producción científica, no está muy claro qué tan conveniente y sostenible es este modelo, por lo menos en México.

Entonces creo que el XML-JATS es una de las pocas opciones que existen. Tiene la capacidad de impulsar la modernización de revistas que necesiten actualizar sus procesos y su manera de presentar la información, a muy bajos costos. Otorga también la posibilidad de insertar a las revistas en el intercambio global de conocimientos, gracias a su naturaleza de sustentarse en el acceso abierto y, al mismo tiempo, de articular sus procesos con otros gestores editoriales, gestores bibliográficos y bases de datos que aprovechen y maximicen los recursos producidos.

Para exponer una escena posible, los propios autores de artículos deben ir pensando la manera en la que su información va a ser asimilada por la web semántica, y cómo ésta podrá lograr un mayor impacto tanto para la difusión como para la preservación de sus manuscritos, dentro de este contexto.

Por ejemplo, los autores, como productores iniciales y principales de la cadena científica de la información, podrán, desde que conciben las primeras líneas de sus manuscritos, pensar que su investigación poseerá metadatos, cuya correcta identificación marcará la diferencia entre ser leído o no ser leído. Es decir, podrán pensar más allá: en posicionar su trabajo y nutrir a su disciplina. Pensar en que su trabajo impactará en su localidad y también en el sistema global, porque aquí todos ya somos parte de un todo.

Referencias

Huh, Sun. 2014. "Journal Article Tag Suite 1.0: National Information Standards Organization standard of journal extensible markup language". *Sci Ed.* 1 (2): 99-104. Publication Date (Web): 2014 August 18 (Training Material) doi: <https://doi.org/10.6087/kcse.2014.1.99>.

Jats for reuse. Acceso el 27 de abril de 2018. <https://jats4r.org/>.

Leiva, Isidoro Gil. 2008. *Manual de indización: teoría y práctica*. España: Ed. Trea.

National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM). 2012. Acceso el 24 de octubre de 2017. <https://dtd.nlm.nih.gov/>.

National Information Standards Organization (NISO). (2015). JATS: Journal Article Tag Suit, versión 1.1. NISO: Baltimore, Maryland, U.S.A.

Senso, J. A., y A. de la Rosa-Piñero. 2003. "El concepto de metadato. Algo más que descripción de recursos electrónicos". *Ciência da Informação* 32(2).

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2015. *Interoperability and Retrieval*. Paris: UNESCO.