

Encontro de Usuários de Sistemas de Publicação - SIS PUB



UTILIZACIÓN DEL XML/JATS DE SCIELO PARA GENERAR GALERADAS HTML EN OJS

Gabriela Pérez Caviglia

Universidad de la República
gabriela.perez@cse.udelar.edu.uy

Como citar:

CAVIGLIA, G. P. Utilización del xml/jats de scielo para generar galeradas html en OJS. In: Sispub, 2, 2019, Curitiba. **Anais...** Curitiba: Ibict, 2019, p. 1-12
[HTTP://DX.DOI.ORG/10.18225/SISPUB.2019.9](http://dx.doi.org/10.18225/sispub.2019.9)

RESUMO

Este artigo descreve o processo de utilização da XML/JATS de um arquivo acadêmico resultante do metodologia SciELO, para gerar automaticamente uma versão HTML do conteúdo em forma cuase automatica utilizando o plugin eLife eLens para o OJS.

Palavras-chave: XML/JATS. HTML. ELens. OJS. SciELO.

INTRODUCCIÓN

En el presente trabajo describiremos el proceso que debe realizarse para aprovechar el *Extensible Markup Language* (XML) resultante de la marcación el *Journal Article Tag Suite* (JATS) para *Scientific Electronic Library Online* (SciELO) para generar versiones *Hypertext Markup Language* (HTML) de los contenidos científicos utilizando el *plugin* eLife eLens de *Open Journal system* (OJS).

El JATS es un estándar de *National Information Standards Organization* (NISO) Z39.96-2015, que define un conjunto de elementos y atributos XML para etiquetar el contenido completo de los artículos publicados en revistas científicas, con el fin de describir su estructura, semántica y metadatos. JATS es una continuación de la obra NLM de Archivo e Intercambio DTD iniciada en 2002 por el *National Center for Biotechnology Information* (NCBI).

Varios artículos han abordado ya la metodología para la preparación de contenidos para su inclusión en la SciELO, así como la importancia de este repositorio para visibilizar la producción científica latinoamericana y también se han analizado los aspectos económicos asociados al proceso de marcación de los contenidos para este modelo.

Relacionado con el último punto Rozemblum, Bava, Unzurrunzaga y Banzato (2014) efectuaron un análisis de los costos y beneficios de incluir revistas universitarias en bases de datos de acceso abierto, haciendo una comparación entre Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc), SciELO y *Directory of Open Access Journals* (DOAJ), y obtuvieron como resultado que para un número de revista de 10 contribuciones, SciELO es 54.67 veces más caro que Redalyc y 82 veces más que DOAJ, por lo cual es sumamente beneficioso si podemos capitalizar el trabajo realizado en la marcación JATS para la generación de otro formato digital para nuestros contenidos.

Por otra parte, Herrera, Ferracutti, Martínez y Mitzig (2011) presentan la distribución de los días de procesamiento de 14 números de diferentes revistas usando la metodología de SciELO, los cuales se dividen en días de generación del HTML (7 días en promedio) y días de marcación (7 días en promedio), y obtienen que se necesitan 14 días hábiles en promedio para procesar un número de aproximadamente 12 contribuciones. Con la generación automática de la versión HTML a partir del XML resultante del proceso de marcación se lograría disminuir la cantidad de días de este proceso de trabajo y también su costo, aprovechando el trabajo realizado con la marcación para generar otro aporte a la revista.

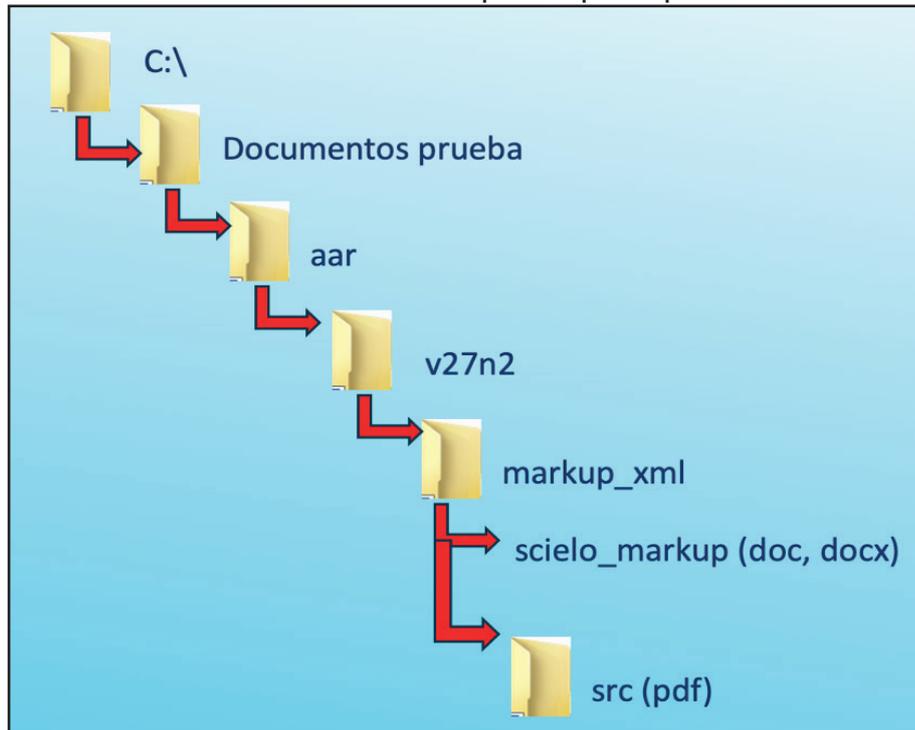
eLife Lens proporciona una forma novedosa de ver el contenido en la web. Permite, ver una figura en un artículo en línea, y simultáneamente ver lo que dice el autor sobre la figura. Se aprovecha al máximo la naturaleza dinámica de HTML combinado con javascript en el renderizado de documentos XML (Grubisic I.; Aufreiter M.; Buchtala O.; Nott G.; Close R.; Korosec S.; Hamilton I.; y Mulvany I., 2019).

A continuación, abordaremos paso a paso el proceso que debe realizarse para la re-utilización de XML/JATS de SciELO en OJS con eLens eLife.

METODOLOGÍA DE TRABAJO

Proceso de marcapeo para ScieELO

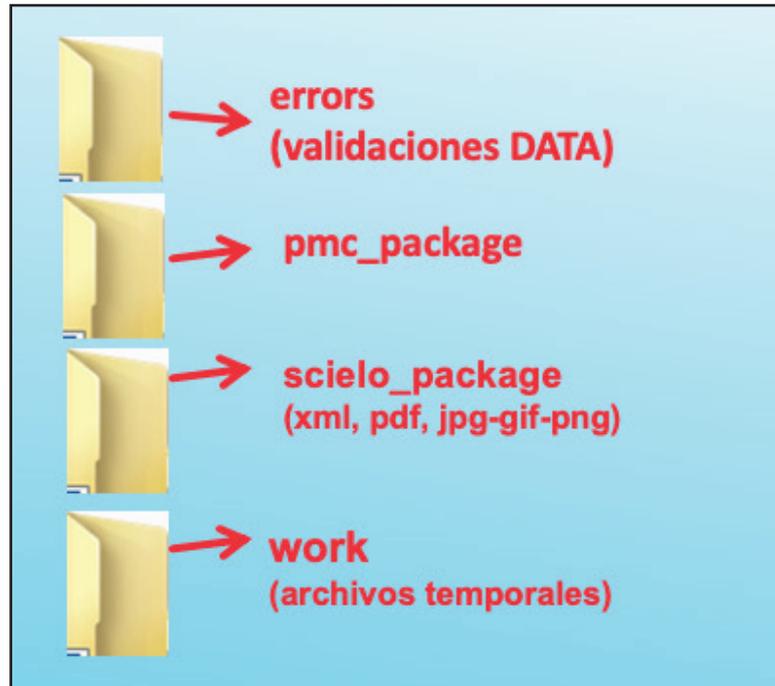
Para comenzar el proceso debemos contar con un XML/JATS. En este artículo vamos a describir el proceso usando el XML resultante de la marcación para SciELO, para la cual se necesita una estructura de carpetas en el disco duro donde se encuentra instalada la suite de marcación, que es una macro de *Microsoft Word* (Gráfico 1). Luego de la preparación de las carpetas, estamos listos para la realización de los pasos de la metodología SciELO:

Gráfico 1- Estructura de carpetas para proceso SciELO.

Fuente: Elaboración propia (2019)

- Paso 1. Creación de la revista usando el Title Manager
- Paso 2. Creación de las secciones (Title Manager)
- Paso 3. Creación del número de la revista por ser procesado (Title Manager)
- Paso 4. Preparación del documento: consiste en arreglar el documento según los requerimientos de SciELO, para comenzar la marcación, como lo son el orden del DOI, sección a la que pertenece el documento, títulos del artículo, autores, afiliaciones, tamaño de la letra para los títulos de las secciones y subsecciones, interlineado tanto en el cuerpo del artículo como en las referencias e inclusión de las imágenes.
- Paso 5. Marcación usando el Markup: consiste en la marcación del texto del documento usando las etiquetas de SciELO. con la finalidad de generar estos documentos en formato XML JATS, según el esquema de SciELO.
- Paso 6. Realización de validaciones y generación del XML: una vez que el documento ha sido completamente marcado, se debe generar el archivo XML, el cual se acompaña con los archivos de imágenes y el documento en formato PDF. (con las validaciones se generarán nuevas carpetas, Gráfico 2)
- Paso 7. Depuración de errores del XML: al generar el archivo XML es posible que existan errores en la marcación del documento, lo que requiere revisar los informes de errores que se generan y depurarlos.

Gráfico 2 - Estructura de carpetas que se generan en el paso 6.



Fuente: Elaboración propia (2019)

Luego de la realización de todos estos pasos, en la carpeta scielo_package vamos a encontrar los archivos XML que debemos usar. Es bueno realizar la galerada HTML luego que el número haya sido subido al repositorio SciELO, pues allí será donde se encontrarán las imágenes asociadas, veremos más adelante las opciones para manejar las rutas hacia ellas.

Habilitación de eLife eLens en OJS 3

Para habilitar el plugin se debe contar con los permisos necesarios. Hecha esta aclaración, se debe ir a Settings-Website-Plugins / Ajustes-Sitio web-Módulos / Configurações-Website-Plugins en el apartado Plugins Genéricos podrá habilitar eLife eLens como se muestra en el Gráfico 3

Gráfico 3 - Habilitación del plugin eLife Lens en OJS 3.

Generic Plugins (21)		
▶ Usage event	Creates a hook that provides usage event in a defined format.	<input checked="" type="checkbox"/>
▶ Web Feed Plugin	This plugin produces RSS/Atom web syndication feeds for the current issue.	<input checked="" type="checkbox"/>
▶ Recommend Articles by Author	This plugin inserts a list of articles by the same author on the article abstract page.	<input checked="" type="checkbox"/>
▶ HTML Article Galley	This plugin provides rendering support for HTML Article Galleys.	<input checked="" type="checkbox"/>
▶ Acron Plugin	This plugin attempts to reduce the dependance of the application on periodic scheduling tools such as 'cron.'	<input checked="" type="checkbox"/>
▶ Usage Statistics	Present data objects usage statistics. Can use server access log files to extract statistics.	<input checked="" type="checkbox"/>
▶ PDF.JS PDF Viewer	This plugin uses the pdf.js PDF viewer to embed PDFs on the article and issue galley view pages.	<input checked="" type="checkbox"/>
▶ Google Analytics Plugin	Integrate OJS with Google Analytics, Google's web site traffic analysis application. Requires that you have already setup a Google Analytics account. Please see the Google Analytics site for more information.	<input checked="" type="checkbox"/>
▶ Google Scholar Indexing Plugin	This plugin enables indexing of published content in Google Scholar.	<input checked="" type="checkbox"/>
▶ TinyMCE Plugin	This plugin enables WYSIWYG editing of textareas using the TinyMCE content editor.	<input checked="" type="checkbox"/>
▶ eLife Lens Article Viewer	This plugin provides rendering support for JATS XML galleys using eLife Lens.	<input checked="" type="checkbox"/>
▶ Dublin Core Indexing Plugin	This plugin embeds Dublin Core meta tags in article views for indexing purposes.	<input checked="" type="checkbox"/>

Fuente: Elaboración propia (2019)

Edición del archivo XML para arreglar las rutas de las imágenes

En el archivo XML de SciELO, los enlaces a las imágenes son relativos. Si usamos el XML sin editarlo las imágenes no se despliegan, pues no las

encuentra.

Para que todo se vea correctamente hay que hacer una pequeña edición en la ruta a de los archivos de imagen.

Para ello puede utilizarse cualquier editor gratuito o simplemente el block de notas. En el gráfico 4 lo que se puede ver es el documento XML desplegado en el editor gratuito Brackets que sirve para editar un variado número de documentos (xml, html, css, php, etc.)

Gráfico 4 - Archivo XML en editor Brackets.

```

222 <fig id="f1">
223   <label>Figura 1:</label>
224   <caption>
225     <title>Obstáculos y dificultades percibidos por los estudiantes</title>
226   </caption>
227   <graphic xlink:href="2301-0126-ic-6-01-54-gf1.png"/>
228 </fig>
229 </p>
230 <p>
231   <list list-type="simple">
232     <list-item>
233       <p>1) <i>Organización del tiempo</i>. Aquí se agrupan respuestas referidas al tiempo que insumen actividades cotidianas como el descanso y el traslado a la institución de enseñanza, pero también respuestas que aluden explícitamente a la falta de organización y planificación para estudiar.</p>
234     </list-item>
235     <list-item>
236       <p>2) <i>Adaptación al sistema universitario</i>. En esta categoría se engloban las respuestas que mencionan habilidades como la autonomía a la hora de seguir un curso, así como también para organizar los horarios, y otras más concretas como el nivel de exigencia, el ritmo de los cursos (planteando diferencias en los procesos de enseñanza y aprendizaje) y la exigencia a nivel cognitivo para la comprensión de los temas.</p>
237     </list-item>
238     <list-item>
239       <p>3) <i>Características personales</i>. Aquí se agrupan respuestas que refieren a dificultades percibidas en aspectos extraacadémicos y más personales. Por ejemplo, un estudiante mencionó «enfrentar mi frustración y mi miedo a no poder salvar los cursos».</p>
240     </list-item>
241     <list-item>
242       <p>4) <i>Características propias de la facultad y las disciplinas</i>. Las respuestas ubicadas en esta categoría refieren también a cuestiones de infraestructura, en las que la base del reclamo está ligada a la situación de masividad que se percibe con relación a los cursos de los primeros años de las carreras.</p>
243     </list-item>
244   </list>
245 </p>
246 <p>Al agrupar todas las respuestas obtenidas en las categorías mencionadas, se encontraron algunas que pueden corresponder a más de una categoría. Por ejemplo, cuando se plantea desconexión entre el curso práctico y el teórico. Esto puede deberse a un tema de coordinación docente, y en este sentido es inherente al curso, pero también puede relacionarse con la adaptación que debe hacerse al sistema universitario, en el entendido de que se espera mayor autonomía del estudiante para regular su proceso de aprendizaje.</p>
247 <p>Se valora positivamente el hecho de que, de acuerdo a la clasificación en que se ordenaron las respuestas, la gran mayoría se ubica en categorías en las que la característica principal puede ser controlada por el estudiante. Este dato es interesante porque ofrece a la facultad indicios sobre qué habilidades promover en sus nuevos estudiantes para intentar mejorar su desempeño. Dichas habilidades, como complemento de las disciplinas específicas, se enfocarán en cuestiones más actitudinales que faciliten la transición entre la educación secundaria y la educación universitaria, y contribuyan a reducir las situaciones de rezago que se observan y que a veces pueden conducir a la desvinculación (temprana) de la institución universitaria.</p>
248 <p>Las categorías 1 y 2 (que son las porcentualmente más significativas) comparten una subcategoría <i>la organización y el manejo del tiempo</i> con relación a aspectos

```

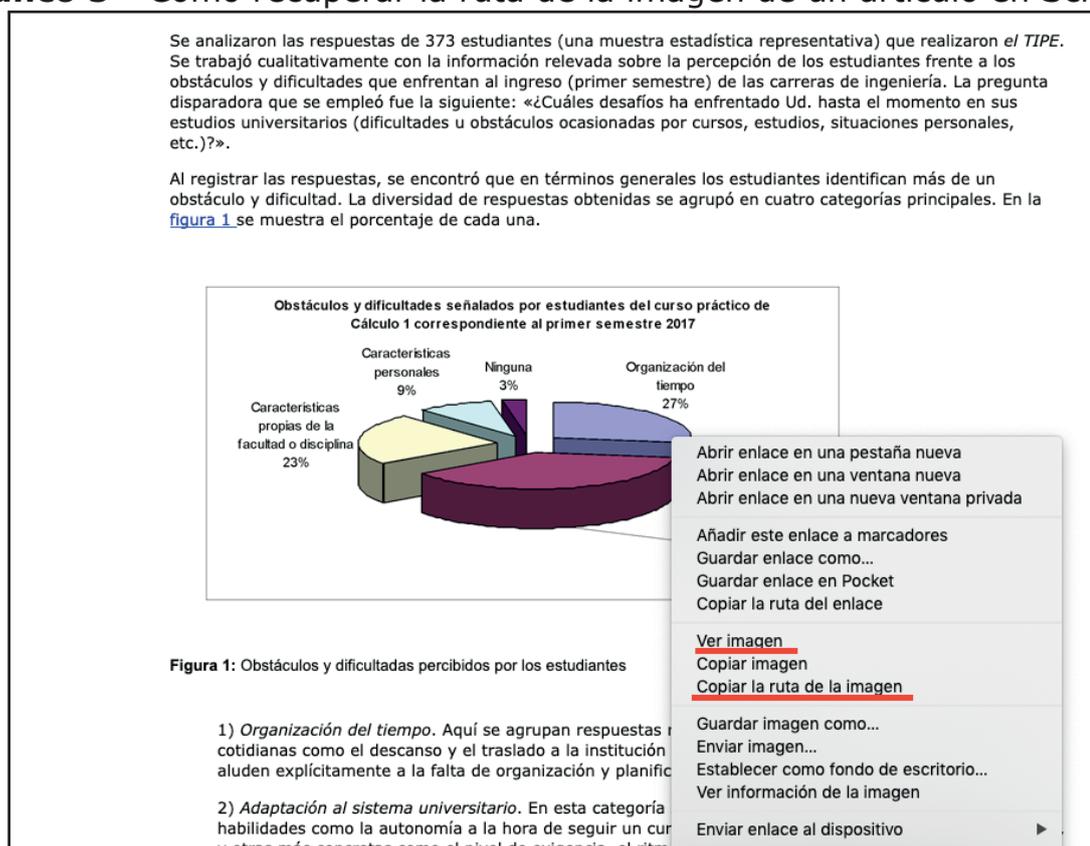
Fuente: Elaboración propia (2019)

Las etiquetas que debemos editar son las <graphic> y se verán de una forma parecida a esta en el editor:

<graphic xlink:href="2301-0126-ic-6-01-54-gf1.png"/>

Aquí tenemos más de una opción, se pueden descargar las imágenes y ponerlas en un servidor local con ayuda del área de informática, y luego construir las rutas, ó utilizar las imágenes que se encuentran en el repositorio de SciELO. En este caso, lo haremos utilizando las imágenes del artículo que se encuentran en el repositorio de SciELO. Para ello, iremos a la página de SciELO, localizaremos el artículo y haremos clic derecho sobre la imagen para recuperar su URL, y podemos hacer clic en "Ver imagen" que nos mostrará la ruta en la barra de direcciones o en "Copiar la ruta de la imagen" que nos copiará la ruta (Gráfico 5).

Gráfico 5 - Cómo recuperar la ruta de la imagen de un artículo en SciELO.



Fuente: Elaboración propia (2019)

Aquí tenemos más de una opción, se pueden descargar las imágenes y ponerlas en un servidor local con ayuda del área informática y referenciar a ese servidor, o utilizar las imágenes que ya están el repositorio SciELO. Esta última opción es la que explicaremos.

La ruta copiada luego de este proceso se verá de un modo parecido a esta:

<http://www.scielo.edu.uy/img/revistas/ic/v6n1/2301-0126-ic-6-01-54-gf1.png>

En rojo se encuentra señalado lo que nos importa, pues esta ruta será la misma para todos los artículos del número, entonces volviendo a la etiqueta graphic, lo que debemos incorporar a ella es lo señalado en rojo, quedando de este modo:

```
<graphic xlink:href="http://www.scielo.edu.uy/img/revistas/ic/v6n1/2301-0126-ic-6-01-54-gf1.png"/>
```

Este procedimiento lo debemos realizar en todas las etiquetas graphic de todos los archivos. Si las tablas se encontraran tratadas como imágenes también deben arreglarse las URL, así como otros formatos que puedan estar enlazados.

Realización de la galerada HTML

Esta parte es muy fácil, teniendo ya acondicionadas todas las rutas de los XML, realizaremos la galerada. Para esto debemos ir al número que deseamos editar (Gráfico 6).

Gráfico 6 - Editando un envío

The screenshot shows a web interface for managing a journal issue. The title bar reads "Gestión del número: Vol. 6 Núm. 1 (2019): InterCambios. Dilemas y Transiciones de la Educación Superior (Junio 2019)". Below the title bar are navigation tabs: "Tabla de contenidos", "Datos del número", "Galeradas del número", "Identificadores", and "Ayuda". A "Ordenar" button is also visible. The main content area is a table of contents with the following sections and items:

- Editorial**
 - ▶ Editorial
- Dilemas y debates**
 - ▶ Pensar la formación universitaria entre disciplinas y entre instituciones
- Investigaciones y experiencias**
 - ▶ Alfabetización informacional y digital mediante b-learning
 - ▶ Estudio de toma de apuntes en un centro de profesores uruguayo
 - ▶ La brecha secundaria-universidad: diagnóstico de los estudiantes de Agronomía en el curso inicial de matemáticas
 - ▶ Cambios y tensiones en las políticas de ingreso en Argentina
 - ▶ Un taller de iniciación a la planificación estratégica para estudiantes de ingreso a carreras de ingeniería
- Reseñas**
 - ▶ Rafael Porlán: Enseñanza Universitaria. Cómo mejorarla.

At the bottom of the 'Investigaciones y experiencias' section, there are two buttons: "Envío" (highlighted with a red box) and "Eliminar".

Fuente: Elaboración propia (2019)

Al editar el envío nos llevará a su flujo de trabajo. Allí podremos añadir la galerada en formato HTML como se muestra en el gráfico 7.

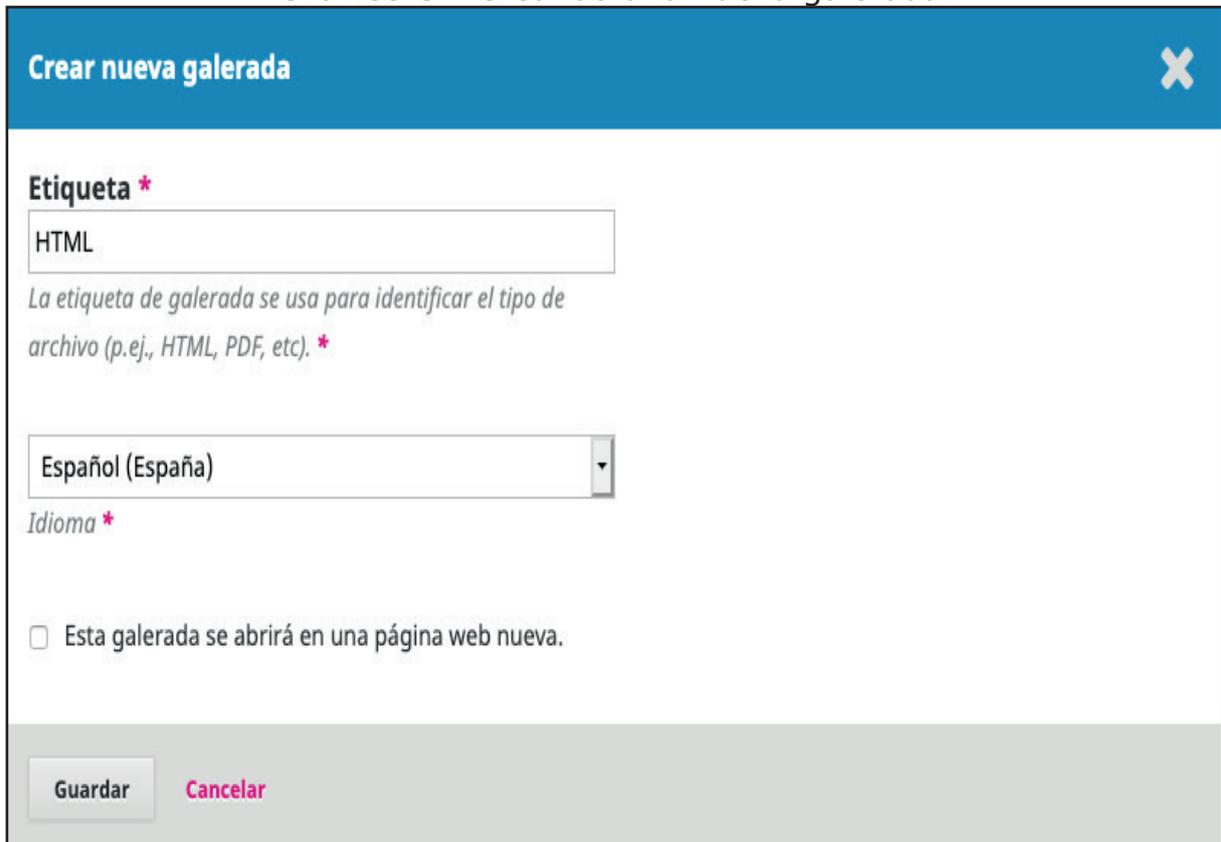
Gráfico 7 - Añadir Galerada.



Fuente: Elaboración propia (2019)

En el apartado "Galeradas" hacemos clic en Añadir galerada, como nombre de etiqueta ponemos HTML, seleccionamos el idioma y hacemos clic en "Guardar" (Gráfico 8)

Gráfico 8 - Creando una nueva galerada.

The image shows a modal window titled 'Crear nueva galerada' with a close button (X) in the top right corner. The form contains the following elements: a text input field for 'Etiqueta *' with the value 'HTML'; a descriptive text below it: 'La etiqueta de galerada se usa para identificar el tipo de archivo (p.ej., HTML, PDF, etc). *'; a dropdown menu for 'Idioma *' with 'Español (España)' selected; a checkbox labeled 'Esta galerada se abrirá en una página web nueva.' which is currently unchecked; and at the bottom, two buttons: 'Guardar' and 'Cancelar'.

Fuente: Elaboración propia (2019)

Luego deberemos cargar el envío y los metadatos. En "Componente de artículo" seleccionamos "Texto de artículo", luego hacemos clic en "Subir fichero" y buscamos en el XML del artículo. Una vez cargado (Gráfico 9), hacemos clic en "Continuar"

Gráfico 9 - Cargando el XML en la galerada.



Subir un archivo listo para publicación

1. Cargar envío 2. Metadatos 3. Finalizar

Componente del artículo *

Texto del artículo

✓ cse-admin, 2301-0126-ic-6-01-54.xml Cambiar fichero

Continuar Cancelar

Fuente: Elaboración propia (2019)

En el apartado de metadatos, podemos editar lo que nos parezca necesario o si todo está ok hacer clic en "Continuar" y luego en "Completar". Con esto ya quedaría la galerada HTML cargada.

RESULTADOS

Los artículos renderizados con eLife Lens se ven del modo que muestra el gráfico 10, manteniendo todas las referencias cruzadas realizadas en el marcapeo (imágenes, tablas, referencias).

Gráfico 10 -Artículo de muestra en el sitio eLife

The screenshot displays the eLife Lens interface. On the left, there is a sidebar with the article title "eLife Lens: A novel way of seeing content" and a list of authors: Ivan Grubisic, Michael Aufreiter, Oliver Buchtala, Graham Nott, Rebecca Close, Samo Korosec, Ian Hamilton, and Ian Mulvany. Below the authors is the "Abstract" section, followed by "Main Text" and "Motivation". The "Motivation" section discusses the challenges of digital documents and the benefits of eLife Lens.

On the right, there is a video player titled "Video 1." showing a video titled "introducing eLife Lens". Below the video player is a section titled "Figure 1." which displays JSON data for "JSON Backbone", "Content Nodes", "Figure Nodes", and "Annotations".

```

JSON Backbone
{ "id": "introducing_lens",
  "nodes": [ ... ],
  "properties": [ ... ],
  "images": [ ... ],
  "content": [ "see intro",
    "figure" ] }

Content Nodes
{ "type": "text",
  "id": "introduction",
  "content": "This is an alternative way to view a research article (Figure 1).",
  "images": [ "see intro",
    "figure" ] }

Figure Nodes
{ "type": "image",
  "id": "image-fig1",
  "label": "Figure 1." }

Annotations
{ "annotations": [ {
  "type": "figure_reference",
  "id": "annotation1",
  "label": "reference" } ] }

```

Fuente: eLife Lens (2019)

CONSIDERACIONES FINALES

Se ha identificado como flujo de trabajo ideal para los editores de libros y revistas comenzar con XML, para luego generar versiones digitales como lo son HTML y PDF (entre otras). Si bien, no es inmediato dar un salto a este flujo de trabajo ideal, es posible hacer un progreso gradualmente hacia él e ir obteniendo beneficios intermedios en el camino, como en este caso.

La versión generada en HTML además permitirá al lector una experiencia más cómoda y adecuada del artículo científico, cuyo contenido es una amalgama de texto, figuras, tablas, videos y referencias que se utilizan para formar el argumento del autor. Además el HTML, es un formato estándar y mucho más adecuado para su lectura en dispositivos móviles que el PDF.

El poder re-utilizar el XML resultante del arduo proceso de marcación para SciELO, capitalizando todas sus capacidades (referencias cruzadas de bibliografía, imágenes, tablas) para generar en forma casi automática, sin necesidad de contar con conocimientos técnicos específicos, una versión HTML de los artículos de una revista; es un beneficio en tiempo y esfuerzo.

REFERENCIAS

GRUBISIC I. et al. **eLife Lens: A novel way of seeing content**. 2019. Disponible em: <<https://lens.elifesciences.org/about/#toc/all>>. Acesso

em: 10 jun. 2019.

HERRERA, L. A. et al. **El Acceso Abierto para la sustentabilidad de las publicaciones de la Universidad Nacional del Sur.** (Argentina), nov. 2011. Disponible em: <<https://www.youtube.com/watch?v=h-lqf9mDqU8>>. Acceso em: 27 out. 2011.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI). **National Library of Medicine.** (2013). Disponible em: <<http://jats.nlm.nih.gov/>>. Acceso em: 15 jun. 2019.

PACKER, A. L. et al. **¿Por qué XML? SciELO en Perspectiva.** 2014. Disponible em: <<http://blog.scielo.org/es/2014/04/04/porque-xml/>>. Acceso em: 30 abr. 2019.

ROZEMBLUM, Cecilia et al. Costos y beneficios de la inclusión de revistas universitarias en bases de datos de Acceso Abierto. In: 12ª JORNADA SOBRE LA BIBLIOTECA DIGITAL UNIVERSITARIA 6 AL 7 DE NOVIEMBRE DE 2014 SALTA, ARGENTINA. **La producción científica en la actividad universitaria.** Universidad Católica de Salta, 2014. Disponible em: <<http://www.unsa.edu.ar/biblio/jbdu/gacetillas/JBDU.pdf>>. Acceso em: 15 jun. 2019.