



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

Gabriela Pérez Caviglia¹

¹Universidad de la República/Comisión Sectorial de Enseñanza/gabriela.perez@cse.udelar.edu.uy

Tema: Formatos de publicação XML / ePub / etc

RESUMEN

En este artículo se describe el procedimiento para utilizar el XML/JATS de un contenido académico, resultante del procesamiento con la metodología SciELO, para generar de forma casi automática una versión HTML de dicho contenido utilizando el plugin eLife eLens para OJS.

Palabras clave: XML/JATS, HTML, eLens, OJS, SciELO

RESUMO

Este artigo descreve o processo de utilização da XML/JATS de um arquivo académico resultante do metodología SciELO, para gerar automaticamente uma versão HTML do conteúdo em forma quase automática 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 XML resultante de la marcación JATS para SciELO para generar versiones HTML de los contenidos científicos utilizando el plugin eLife eLens de OJS.

El Journal Article Tag Suite (JATS) es un estándar de 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, 2013).

Varios artículos han abordado ya la metodología para la preparación de contenidos para su inclusión en la Scientific Electronic Library Online (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 Redalyc, SciELO y 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).

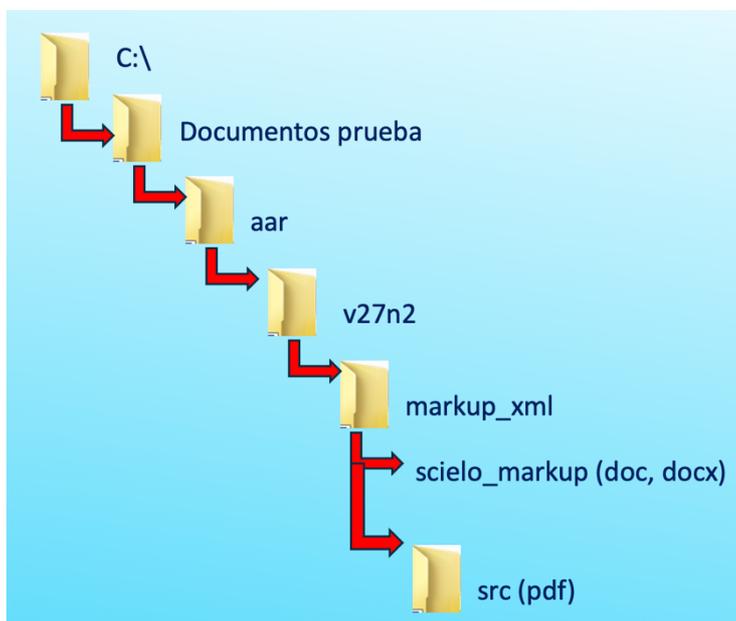
A continuación, abordaremos paso a paso el proceso que debe realizarse para la reutilización de XML/JATS de SciELO en OJS con eLens eLife.

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



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)

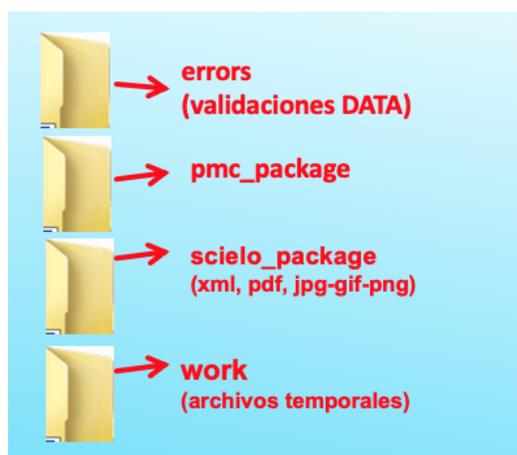
Paso 4. Preparación del documento: consiste en arreglar el documento según los requerimientos de SciELO, para comenzar la marcación, como lo son el orden del DOI, sección a la que pertenece el documento, títulos del artículo, autores, afiliaciones, tamaño de la letra para los títulos de las secciones 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



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.	☑
▶ Web Feed Plugin	This plugin produces RSS/Atom web syndication feeds for the current issue.	☑
▶ Recommend Articles by Author	This plugin inserts a list of articles by the same author on the article abstract page.	☑
▶ HTML Article Galley	This plugin provides rendering support for HTML Article Galleys.	☑
▶ Acron Plugin	This plugin attempts to reduce the dependance of the application on periodic scheduling tools such as 'cron.'	☑
▶ Usage Statistics	Present data objects usage statistics. Can use server access log files to extract statistics.	☑
▶ PDFJS PDF Viewer	This plugin uses the pdf.js PDF viewer to embed PDFs on the article and issue galley view pages.	☑
▶ 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.	☑
▶ Google Scholar Indexing Plugin	This plugin enables indexing of published content in Google Scholar.	☑
▶ TinyMCE Plugin	This plugin enables WYSIWYG editing of textareas using the TinyMCE content editor.	☑
▶ eLife Lens Article Viewer	This plugin provides rendering support for JATS XML galleys using eLife Lens.	☑
▶ Dublin Core Indexing Plugin	This plugin embeds Dublin Core meta tags in article views for indexing purposes.	☑

Edición del archivo XML para arreglar las rutas de las imágenes

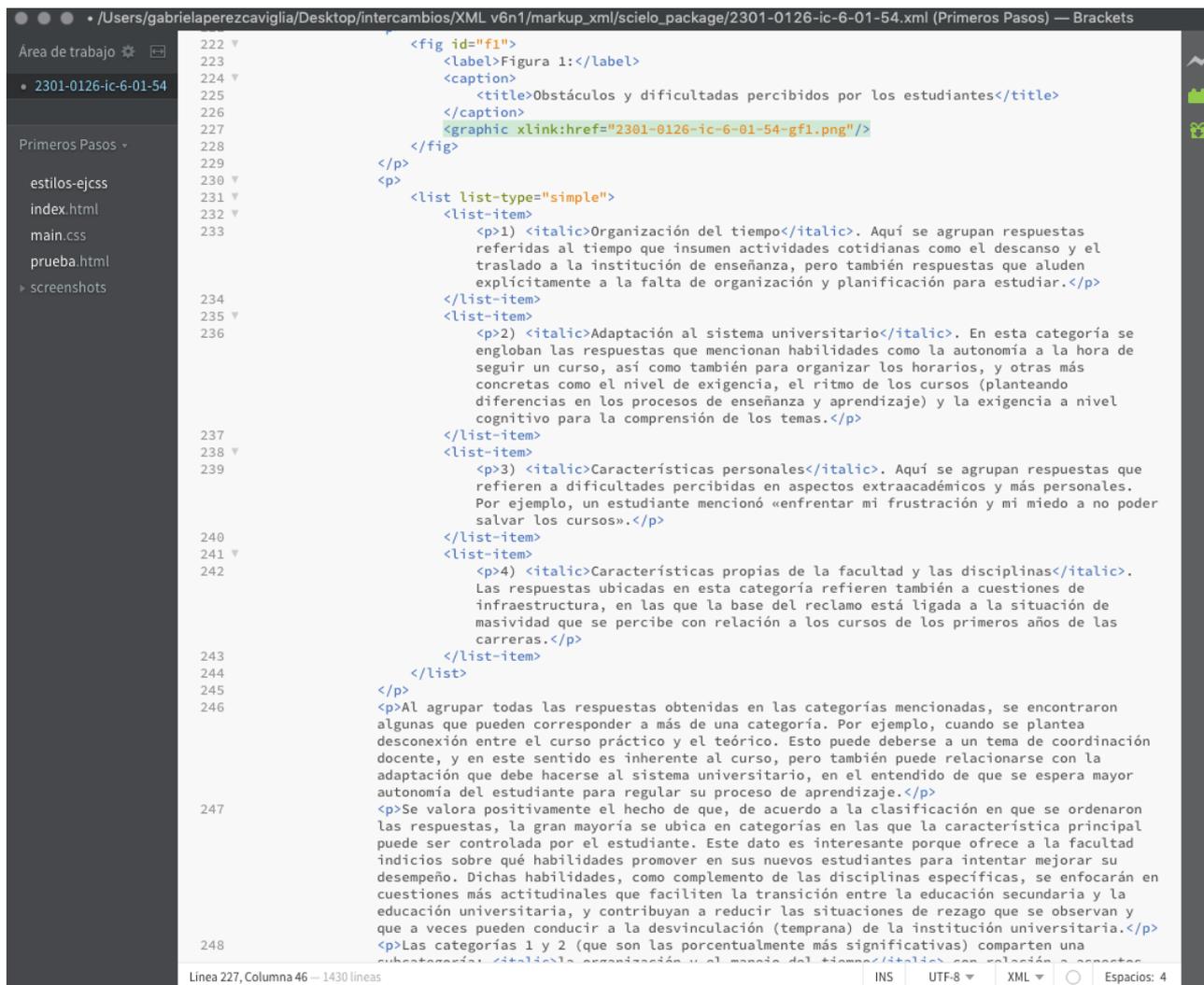
En el archivo XML de SciELO, los enlaces a las imágenes son relativos. Si usamos el XML sin editarlo las imágenes no se despliegan, pues no las encuentra.

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*¹ que sirve para editar un variado número de documentos (xml, html, css, php, etc.)

¹ Más información y página de descarga de Brackets: <http://brackets.io/>

Gráfico 4: Archivo XML en editor Brackets



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

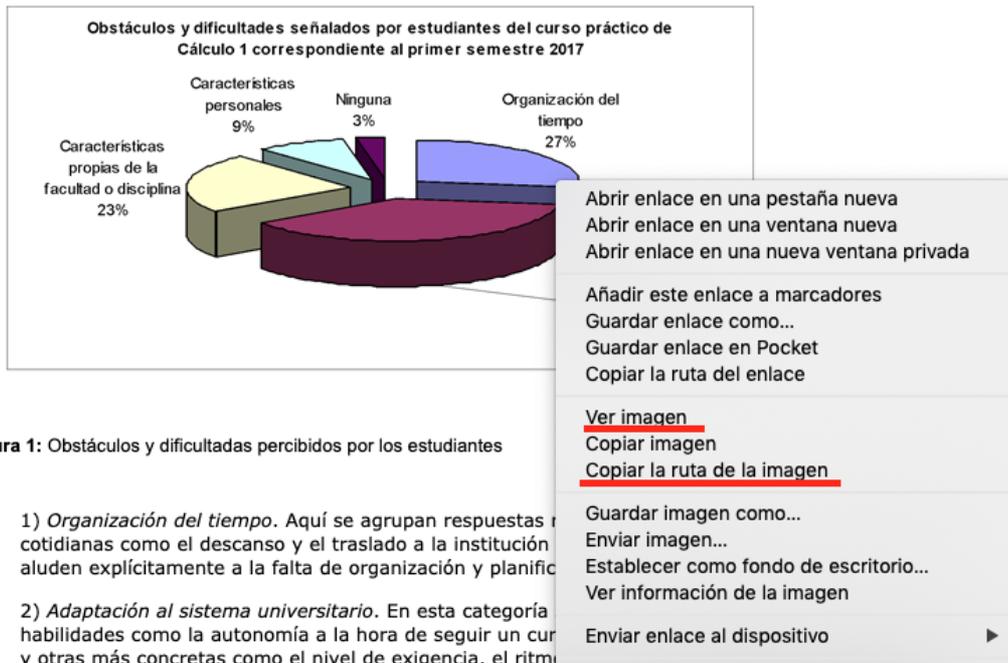
```
<graphic xlink:href="2301-0126-ic-6-01-54-gf1.png"/>
```

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

Se analizaron las respuestas de 373 estudiantes (una muestra estadística representativa) que realizaron *el TIPE*. Se trabajó cualitativamente con la información relevada sobre la percepción de los estudiantes frente a los obstáculos y dificultades que enfrentan al ingreso (primer semestre) de las carreras de ingeniería. La pregunta disparadora que se empleó fue la siguiente: «¿Cuáles desafíos ha enfrentado Ud. hasta el momento en sus estudios universitarios (dificultades u obstáculos ocasionadas por cursos, estudios, situaciones personales, etc.)?».

Al registrar las respuestas, se encontró que en términos generales los estudiantes identifican más de un obstáculo y dificultad. La diversidad de respuestas obtenidas se agrupó en cuatro categorías principales. En la [figura 1](#) se muestra el porcentaje de cada una.



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

Gestión del número: Vol. 6 Núm. 1 (2019): InterCambios. Dilemas y Transiciones de la Educación Superior (Junio 2019)

Tabla de contenidos Datos del número Galeradas del número Identificadores Ayuda

Ordenar

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
 - Envío** Eliminar
- ▶ Alfabetización científica en la formación en ciencias de la naturaleza

Reseñas

- ▶ Rafael Porlán: Enseñanza Universitaria. Cómo mejorarla.

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 6.

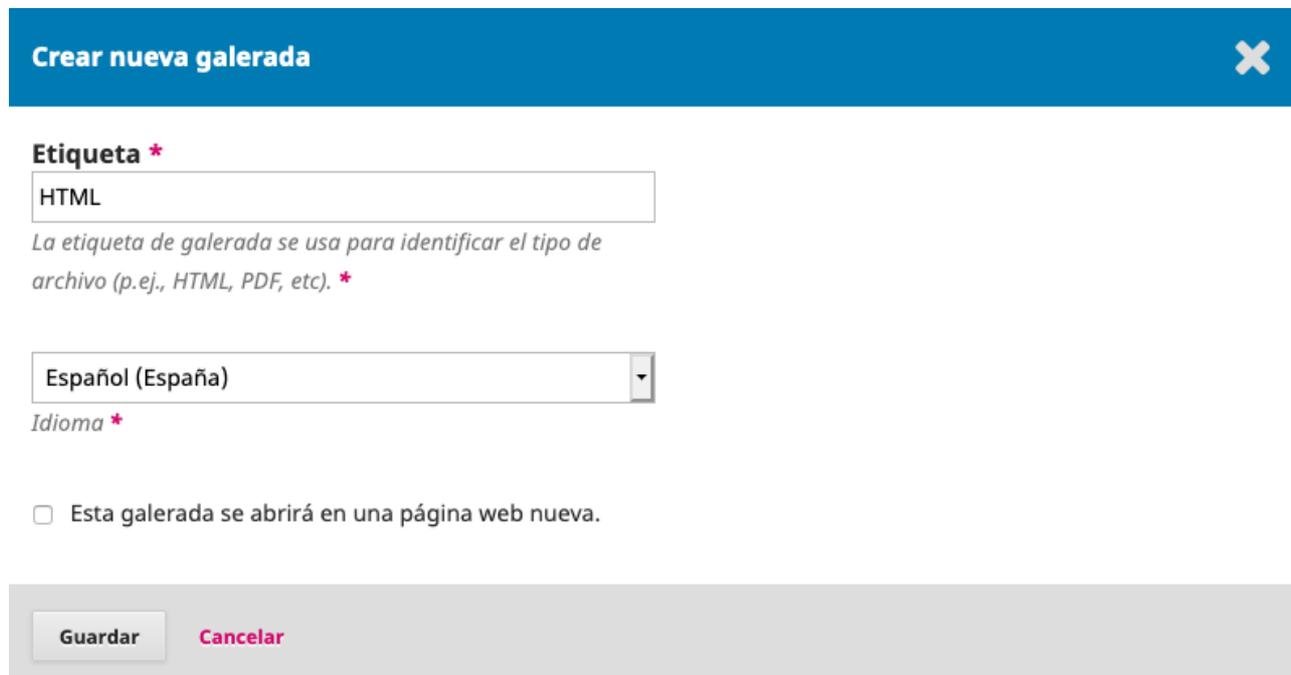
Gráfico 6: Añadir Galerada

Galeradas **Añadir galerada**

- ▶ PDF

En el apartado “Galeradas” hacemos clic en Añadir galerada, como nombre de etiqueta ponemos HTML, seleccionamos el idioma y hacemos clic en “Guardar” (Gráfico 7)

Gráfico 7: Creando una nueva galerada

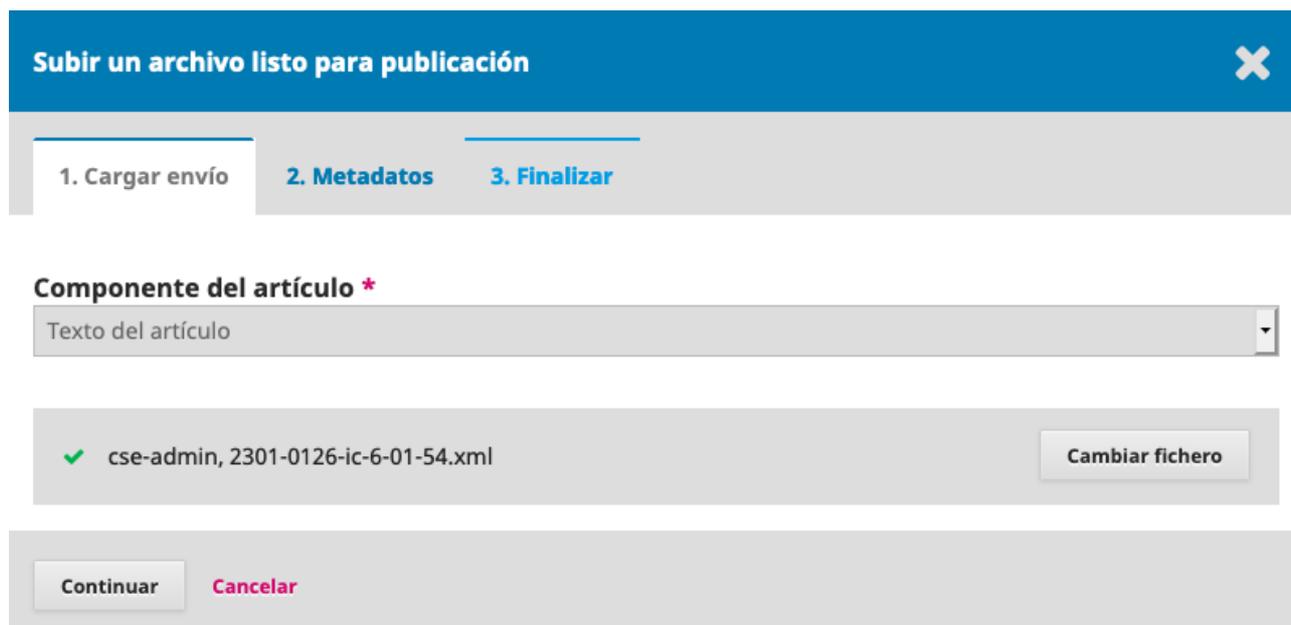


The screenshot shows a form titled "Crear nueva galerada" with a close button (X) in the top right corner. The form contains the following elements:

- Etiqueta ***: A text input field containing "HTML". Below it is a note: "La etiqueta de galerada se usa para identificar el tipo de archivo (p.ej., HTML, PDF, etc). *"
- Idioma ***: A dropdown menu showing "Español (España)".
- Esta galerada se abrirá en una página web nueva.
- Buttons: "Guardar" and "Cancelar".

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 8), hacemos clic en “Continuar”

Gráfico 9: Cargando el XML en la galerada



The screenshot shows a form titled "Subir un archivo listo para publicación" with a close button (X) in the top right corner. The form contains the following elements:

- Componente del artículo ***: A dropdown menu showing "Texto del artículo".
- Progress indicator: Three steps are shown: "1. Cargar envío" (active), "2. Metadatos", and "3. Finalizar".
- File upload area: A green checkmark, the filename "cse-admin, 2301-0126-ic-6-01-54.xml", and a "Cambiar fichero" button.
- Buttons: "Continuar" and "Cancelar".

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

<https://lens.elifesciences.org/about/#info/all>

The screenshot displays the eLife Lens interface. On the left, there is a sidebar with the title "eLife Lens: A novel way of seeing content" and a list of authors: Ivan Grubic, Michael Aufreiter, Oliver Buchtala, Graham Nott, Rebecca Close, Samo Korosec, Ian Hamilton, and Ian Mulvany. Below the authors, there are sections for "Abstract", "Main Text", and "Motivation". The "Abstract" section contains text about the tool's purpose and how it works. The "Main Text" and "Motivation" sections are partially visible. On the right, there is a video player titled "Video 1." showing a video titled "Introducing eLife Lens." Below the video player, there is a section titled "Figure 1." which displays JSON metadata. The JSON is organized into four sections: "JSON Backbone", "Content Nodes", "Figure Nodes", and "Annotations".

eLife Lens: A novel way of seeing content

Ivan Grubic Michael Aufreiter Oliver Buchtala Graham Nott
Rebecca Close Samo Korosec Ian Hamilton Ian Mulvany

Abstract

eLife Lens provides a novel way of looking at content on the web (see [Introduction post](#)). It is designed to make life easier for researchers, reviewers, authors and readers. For example, have you tried to look at a figure in an online article, while simultaneously trying to see what the author says about the figure? You end up jumping all around the article, losing track of what you were looking for in the first place. The reason for this is that most online research articles are published in a fixed digital version of the original paper. With eLife Lens, we take full advantage of the dynamic nature of HTML combined with javascript ([Video 1](#)).

Main Text

Motivation

Working with digital documents has been difficult, primarily because they come in presentation-centric formats with the goal of appearing identical across multiple devices for printing purposes. Content today, however, is no longer being printed out readily. Instead, it is being read on a variety of platforms, from computers to tablets and other mobile devices. Although optimized for print, when presented with different screen sizes and other device inconsistencies traditional digital publishing formats lack key features that could otherwise improve readers' ability to focus on an author's arguments.

Video 1.

introducing eLife Lens

Introducing eLife Lens.

Watch Ian Mulvany from eLife demonstrating Lens.

Figure 1.

```
JSON Backbone
{
  "id": "introducing_lens",
  "nodes": [ { ... } ],
  "properties": { ... },
  "views": [ {
    "content": ["textintro"],
    "figures": ["imagefig1"]
  } ]
}

Content Nodes
{
  "type": "text",
  "id": "textintro",
  "content": "Lens is an alternative way to view a research article (Figure 1).",
}

Figure Nodes
{
  "type": "image",
  "id": "imagefig1",
  "label": "Figure 1.",
}

Annotations
{
  "annotation1": {
    "type": "figure_reference",
    "id": "annotation1",
    "target": "introducing_lens"
  }
}
```

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., AUFREITER M., BUCHTALA O., NOTT G., CLOSE R., KOROSSEC S., HAMILTON I., y MULVANY I., (consultado en junio 2019). eLife Lens: A novel way of seeing content. Recuperado de <https://lens.elifesciences.org/about/#toc/all>

HERRERA, L. A., FERRACUTTI, V. M., MARTÍNEZ, F. A., y MITZIG, N. L. (2011, octubre 27). El Acceso Abierto para la sustentabilidad de las publicaciones de la Universidad Nacional del Sur. Jornada Virtual Acceso Abierto Argentina 2011. Biblioteca Electrónica de Ciencia y Tecnología - Ministerio de Ciencia, Tecnología e Innovación Productiva / CAICYT-CONICET / OPS-OMS / SPARC. [Archivo de video] Recuperado de <https://www.youtube.com/watch?v=h-lqf9mDqU8>.

NCBI, National Center for Biotechnology Information, U.S. National Library of Medicine. (2013). JATS: Journal Article Tag Suite. Recuperado de <http://jats.nlm.nih.gov/>

PACKER, A. L., SALGADO, E., ARAUJO, J., AQUINO, L., ALMEIDA, R., SANTOS, J., LUCENA, S., SOARES, C. M. (4 de abril de 2014). ¿Por qué XML? SciELO en Perspectiva. [Mensaje en un blog] Recuperado de <http://blog.scielo.org/es/2014/04/04/porque-xml/>.

ROZEMBLUM, C., BAVA, L., UNZURRUNZAGA, C., y BANZATO, G. (noviembre de 2014). Costos y beneficios de la inclusión de revistas universitarias en bases de datos de Acceso Abierto. En 12ª Jornada sobre la Biblioteca Digital Universitaria. "La producción científica en la actividad universitaria". Universidad de La Plata, Salta, Argentina. Recuperado de <http://www.unsa.edu.ar/biblio/jbdu/gacetillas/JBDU.pdf>.