

# La programación de la hoja de estilo XSL para la versión HTML de un artículo científico a partir de un archivo XML JATS: una propuesta

Cuauhtémoc Jiménez Pérez<sup>1</sup>

## Resumen

Las diferentes instancias calificadoras de revistas de investigación señalan la necesidad de contar, para su lectura, con diferentes formatos gráficos de los artículos, por lo cual la inclusión del HyperText Markup Language (HTML) se considera indispensable. Se tiene identificada la posibilidad de automatizar la conversión de archivos en eXtensible Markup Language (XML) a HTML por medio de las hojas de estilo en eXtensible Stylesheet Language (XSL). En este trabajo (parte de un proyecto para la revista *Científica* del Instituto Politécnico Nacional de México), se planea una propuesta para la codificación de una hoja de estilo XSL, empleando un archivo XML etiquetado en el estándar Journal Article Tag Suite (JATS) por medio de Marcalyc (de Redalyc). Se considera la estructura del JATS, que divide a los artículos en tres etiquetas: *front*, *back*, *body*, para utilizar el modelo en espiral de Boehm como el ciclo de desarrollo de *software* pertinente para esta programación. Éste es un ciclo que se repite hasta que el sistema o programa está finalizado. Por tanto, al dedicarle una iteración a cada etiqueta del estándar (tres iteraciones en total), en una cuarta, se obtiene la integración de las etiquetas con sus contenidos, que permite mostrar el artículo en HTML. Se emplea el *software* NetBeans 8.2 (producto gratuito sin restricciones de uso) que permite tanto la programación en XSL como la transformación de XML a HTML. Se espera que este proyecto sirva a las publicaciones que busquen iniciar la automatización de sus procesos editoriales de un modo similar.

Palabras clave: Revista científica; Programación; eXtensible Stylesheet Language; eXtensible Markup Language; Journal Article Tag Suite; HyperText Markup Language.

## Abstract

There is a growing tendency nowadays to evaluate journals in the digital environment, so it is necessary to include a HyperText Markup Language (HTML) version of the papers. As is known, an eXtensible Markup Language (XML) document can be turned into an HTML by using an *eXtensible Stylesheet Language* (XSL). In this document (part of a project in the journal *Científica* of the Instituto Politécnico Nacional of Mexico), how to codify a simplified XSL is shown using an XML in Journal Article Tag Suite (JATS) standard (obtained from Marcalyc framework, developed by Redalyc). Having the three JATS tags: *front*, *back*, *body*; it is appropriate to implement Boehm's spiral model for programming.

---

<sup>1</sup> Es licenciado en Lengua y Literaturas Hispánicas por la Universidad Nacional Autónoma de México (UNAM) e ingeniero en desarrollo de *software* por la Universidad Abierta y a Distancia de México (UnADM). Colabora como editor técnico en la revista *Científica* del Instituto Politécnico Nacional (IPN) de México. Es egresado de los diplomados "Los procesos en la edición de revistas" (Cámara Nacional de la Industria Editorial Mexicana) y "Edición y evaluación de revistas científicas en América Latina" (UNAM).

NetBeans Integrated Development Environment allows to write XSL, also performs the transformation from XML to HTML. We hope this contribution will be useful for editorial teams.

Key words: Journal; Programming; eXtensible Stylesheet Language; eXtensible Markup Language; Journal Article Tag Suite; HyperText Markup Language.

El Instituto Politécnico Nacional en México, a través de la Sección de Estudios de Posgrado e Investigación de la Escuela Superior de Ingeniería Mecánica y Eléctrica (Unidad Zacatenco), publica desde 1997 una revista de investigación con artículos referentes a las especialidades en ingeniería que se imparten ahí: *Científica* (ISSN 1665-0654). Desde el comienzo, ha sido una publicación periódica impresa. En el proceso de edición técnica se realizan las actividades pertinentes para obtener la publicación (Jiménez 2012).

En la actualidad, las revistas se encuentran ante la transición de los modelos de publicación que se habían diseñado para obtener productos impresos. Sin embargo, las exigencias han cambiado de manera acelerada, pues los diferentes organismos que las evalúan centran sus parámetros de calificación, ahora, en el ámbito digital.

Para hacer referencia a los requerimientos sobre los cuales trabajan hoy los equipos editoriales de las revistas, obsérvese del catálogo Latindex (El Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal) los criterios 31 y 32 en la *Nueva lista de características de calidad editorial* (Latindex 2017, 10):

31. Uso de protocolos de interoperabilidad

La revista debe incorporar protocolos de interoperabilidad que le permitan ser recolectada por otros sistemas de distribución...

*Nota de aplicación: Debe indicar qué protocolo de interoperabilidad utiliza y la dirección electrónica para acceder a él. Además de verificar que cada artículo esté descrito con metaetiquetas. Es obligatorio el uso de una norma de metaetiquetas.*

32. Uso de diferentes formatos de edición

Para calificar, la revista debe incluir más de un formato de edición para el despliegue y lectura de los textos completos.

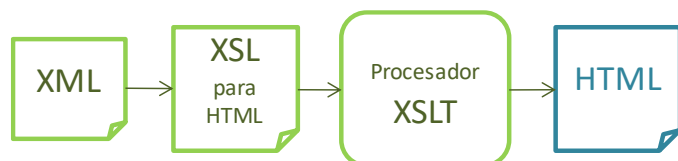
*Nota de aplicación: La revista puede elegir los formatos de su preferencia según el público lector. Cada documento debe tener más de un formato. Se busca ampliar el uso de PDF, el uso de HTML y otros formatos que faciliten el hipertexto y la lectura.*

En este contexto, se tiene identificada la posibilidad de automatizar la conversión de archivos en eXtensible Markup Language (XML) a HyperText Markup Language (HTML), por medio de las hojas de estilo eXtensible Stylesheet Language (XSL) (World Wide Web Consortium 2017).

El lenguaje de metamarcado XML es un formato que, por medio de etiquetas, permite el intercambio de información o datos en Internet (World Wide Web Consortium: 2015). Por otro lado, XSL es un grupo de reglas que permite la transformación de documentos originalmente creados en XML. Está conformado por dos partes: la primera, las *XSL Transformations* (XSLT), propiamente son las instrucciones que transforman los documentos; y la segunda, *XSL Path Language*, el lenguaje empleado por XSLT para hacer referencias a las etiquetas y otorgar una presentación específica. Así, una hoja de estilo XSL especifica el formato de presentación en detalle de un documento XML para su difusión. Esto es desarrollado por el equipo de trabajo del W3C XSLT (World Wide Web Consortium 2017).

El flujo de esta tecnología se muestra en la Figura 1, donde se obtiene el formato electrónico HTML de manera automática, a partir de un formato electrónico XML al emplear tanto los documentos XSL como un procesador XSLT.

**Figura 1. Proceso de los archivos XML y XSL para la obtención de HTML**



Por otro lado, el estándar Journal Article Tag Suite (JATS) define un conjunto de etiquetas XML (elementos y atributos) para marcar artículos científicos. Actualmente, es la *National Information Standards Organization* la entidad que se ha encargado de perfilar la norma vigente (NISO, 2015). El estándar determina tres dimensiones en los elementos: preservación (*archiving*), difusión (*publishing*) y autoría (*authoring*), con doscientas setenta

y tres etiquetas y ciento treinta y seis atributos que no necesariamente se emplean en cada artículo, pues algunos elementos son obligatorios y otros facultativos.

En este sentido, tanto la Scientific Electronic Library On-Line (SciELO México, 2017) como la Red de Revistas Científicas de América Latina y el Caribe, España y Portugal (Redalyc, 2016) han promovido el uso de este estándar internacional. Sin embargo, cada organismo se reserva la facultad de emplear determinados elementos facultativos del estándar; es decir, la aplicación del estándar no es uniforme.

En el seno de la revista *Científica* del IPN, se promueven los cambios pertinentes para enfrentar los nuevos retos que plantea la creación de formatos electrónicos. Un primer paso es abordar la obtención del HTML para los artículos de la revista.

Se conoce que, para este fin, RedALyC desarrolló su plataforma Marcalyc (Redalyc: 2017), un sistema que realiza tanto el marcado XML (en JATS) como la transformación para HTML (entre otros formatos). Sin embargo, se requirió un equipo de desarrollo de, aproximadamente, diez programadores durante dos años (Becerril 2017). Por tanto, no se consideró una opción el desarrollo de un sistema semejante, sino que más bien se optó por la implementación de determinadas prácticas propuestas por ese modelo dentro del proceso de edición técnica en operación.

Esta parte del proyecto, se inicia con la pregunta: ¿Qué solución se puede dar, con conocimientos de la ingeniería de *software*, para producir los artículos de la revista *Científica* en HTML? La principal característica de éste ha sido el desarrollo del código de una hoja de estilo en XSL (con la inclusión de las instrucciones en Cascading Style Sheets, CSS) para la automatización del proceso de edición. Así, se muestran los pormenores de este proceso de programación como una propuesta asequible, dadas ciertas condiciones para su ejecución por parte de los equipos editoriales.

## **Método**

### **Participantes:**

- Supervisor del proyecto. Es quien otorga visto bueno a los avances del proyecto en la entidad (el editor técnico, en este caso, pero puede ser la directora o coordinadora; es decir, alguien del equipo editorial que esté familiarizado con el proceso de publicación).

- Encargado de la página web. Determina los parámetros del HTML y reconoce las características o necesidades del sitio web de la revista.
- Programador. Codifica la hoja de estilo en XSL, con conocimientos de los lenguajes de programación, de CSS y, desde luego, de HTML.

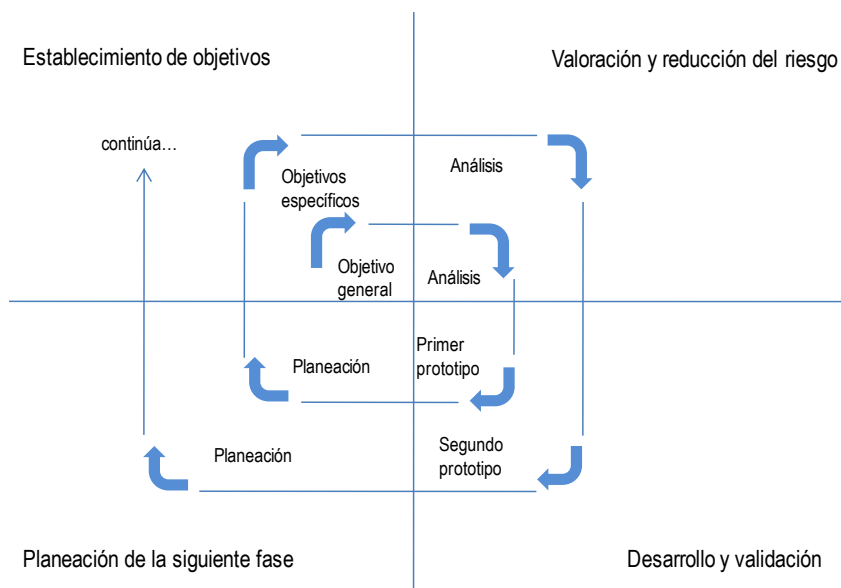
## **Recursos**

- *Hardware*. El equipo de cómputo debe contar, al menos, con memoria RAM de 4 GB, además de conexión permanente a internet.
- *Software*. El entorno de desarrollo integrado (*Integrated Development Environment*) NetBeans IDE en su versión 8.2 (Oracle, 2017, producto gratuito sin restricciones de uso) permite la codificación y uso de archivos en XML, XSL y HTML. Además, incluye la aplicación para las transformaciones XSLT.
- La plataforma en línea Marcalyc (Redalyc 2018) permite el etiquetado en XML (y la descarga) del contenido de los artículos con el estándar JATS (desde luego, se pueden emplear archivos JATS de los artículos elaborados en otros procedimientos).

## **Modelo de proceso de *software***

El modelo en espiral de Boehm considera que un proceso de *software* centrado en el riesgo (Sommerville 2011, 48-50) se representa precisamente como una espiral que conforma ciclos o iteraciones con cuatro pasos: a) establecimiento de objetivos, b) valoración y reducción de riesgo, c) desarrollo y validación, y d) planeación (véase la Figura 2).

**Figura 2. Modelo en espiral de Boehm**



Fuente: Sommerville (2011, 49).

Se utilizó este modelo de proceso de *software*, debido a que, a la luz de los procesos editoriales, los errores en los documentos finales se asumen como un problema en la comunicación de los resultados científicos, por lo cual se considera que el enfoque sobre reducción del riesgo o error (Cervantes y Gómez 2012) aporta un valor al proceso mismo de edición técnica —que presenta también iteraciones al realizar ciclos de revisión de los escritos.

### **Estándar de calidad de software IEEE Std 829-2008**

Se consideró usar el Standard for Software and System Test Documentation (IEEE Std 829-2008) para integrar elementos de calidad al proceso. Este documento establece un marco para las pruebas y validación de los procesos o actividades en el desarrollo u operación de *software* (Institute of Electrical and Electronics Engineers 2008, 2). Se centró la atención para los procedimientos, su ejecución y, como era de esperarse, los resultados. Además, para integrar un instrumento que le diera solidez al desarrollo, se consideró al producto que se puede validar al final del proceso: el artículo en HTML.

Para los casos de los documentos en HTML, existen lineamientos editoriales basados en diferentes normas de los datos que conforman un artículo científico (Delgado *et al.*, 2006). En estos lineamientos, cada elemento es valorado en tres rubros: inclusión (I), presentación (P) y ubicación (U). Para la *inclusión*, se valora que el elemento obligatorio aparezca en el

documento. Para la *presentación*, que esté escrito según las normas pertinentes. Para la *ubicación*, que se encuentre en el lugar necesario, si así lo determina la norma. Esta forma de evaluar tiene sentido para la implementación que se buscaba, ya que los elementos de las etiquetas en el archivo XML debían aparecer, ser legibles y estar bien escritos ,además de estar posicionados en el lugar adecuado dentro del archivo HTML. En la Tabla 1 se observa la lista de cotejo para los elementos de un artículo, empleada en la revista. La lista está basada en los parámetros propuestos por Delgado *et al.* (2006) que se encuentran entre los números 459 y 484. En ella se han vuelto a numerar estos parámetros para un total de quince elementos para revisar en cada uno de los artículos.

**Tabla 1. Lista de cotejo normalizada para revisión de los datos en artículos científicos, basada en los parámetros de calidad de las revistas científicas de Delgado *et al.* (2006) <sup>2</sup>**

| P <sub>o</sub> | P <sub>r</sub> | Elemento  | I | P | U |
|----------------|----------------|---|---|---|---|
|                |                | Título del artículo                                   |   |   |   |
| 459            | 1              | Título del artículo                                   |   |   |   |
| 461            | 2              | Traducción del título                                 |   |   |   |
|                |                | Autoría: identificación (existencia de)               |   |   |   |
| 462            | 3              | Nombre(s) y apellido(s) completos de los autor(es)    |   |   |   |
| 463            | 4              | Filiación institucional                               |   |   |   |
| 465            | 5              | Correo electrónico                                    |   |   |   |
|                |                | Resumen, palabras clave                               |   |   |   |
| 470            | 6              | Resumen   |   |   |   |
| 473            | 7              | Traducción resumen                                    |   |   |   |
| 474            | 8              | Palabras clave, descriptores                          |   |   |   |
| 475            | 9              | Traducción palabras clave, descriptores               |   |   |   |
|                |                | Identificación bibliográfica                          |   |   |   |
| 478            | 10             | Referencia bibliográfica normalizada                  |   |   |   |
|                |                | Fechas de recepción, revisión y aceptación            |   |   |   |
| 482            | 11             | Fecha de recepción                                    |   |   |   |
| 484            | 12             | Fecha de aceptación                                   |   |   |   |
|                |                | Identificación de la revista en las páginas del texto |   |   |   |
| 430            | 13             | Número de volumen                                     |   |   |   |
| 431            | 14             | Número de fascículo                                   |   |   |   |
| 432            | 15             | Periodo cubierto por el fascículo                     |   |   |   |

## Proceso

Para la codificación de la hoja de estilo se retoman, primero, los cuatro pasos del ciclo en espiral de Boehm: a) establecimiento de objetivos, b) valoración y reducción del error, c)

<sup>2</sup> P<sub>o</sub>: número de parámetro en el documento original. P<sub>r</sub>: número de parámetro de esta lista. I: inclusión. P: presentación. U: ubicación.

desarrollo y validación, y d) planeación. Además, se considera la división del artículo en tres secciones principales: 1) página de identificación (primera página) [etiqueta <front> en JATS], 2) referencias [etiqueta <back> en JATS] y, 3) desarrollo [etiqueta <body> en JATS]. Así, al integrar estas actividades en el proceso para la primera iteración se obtuvo:

### **Primera iteración**

#### *Establecimiento de objetivos*

- Describir el estándar JATS de la página de identificación en un artículo.
- Programar la hoja de estilo XSL para HTML de las etiquetas contenidas en el <front> del JATS con los lineamientos editoriales de la revista *Científica*.

#### *Valoración y reducción del error*

Determinar el formato de salida de las etiquetas en el HTML con las instrucciones de cada elemento en CSS.

#### *Desarrollo y validación*

- Codificación de las instrucciones con el lenguaje XSLT para HTML y del respectivo CSS de cada etiqueta.
- Revisión de cada elemento con el instrumento de validación seleccionado con los criterios del estándar IEEE Std 829-2008.

#### *Planeación*

- Revisión, en consenso con los participantes del proyecto, del HTML resultante.
- Propuesta de mejora para la segunda iteración, donde se valoren los avances.

Como se puede inducir, las actividades descritas se repitieron en el mismo orden para las dos etiquetas restantes, <back> y <body>. Se terminó en una cuarta iteración, considerando el artículo completo (véase la Figura 3).

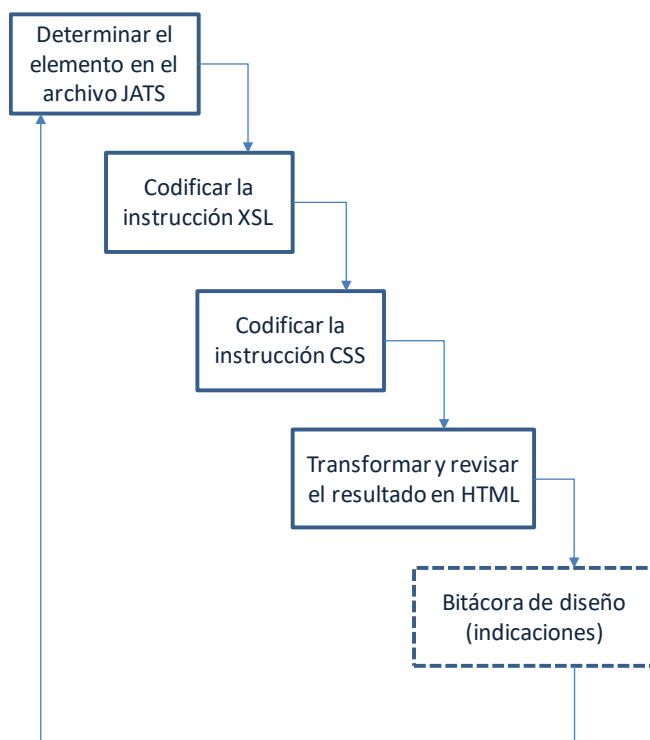
**Figura 3. Proceso de programación con cuatro iteraciones (en cada una se realiza el ciclo de desarrollo de *software* de Boehm)**





Este proceso de codificación, observado desde la perspectiva de una etiqueta (véase la Figura 4), se resume en cuatro actividades principales: determinar el elemento en el archivo JATS, codificar la instrucción XSL, codificar la instrucción CSS, transformar y revisar el resultado en HTML. Se añadió, en el desarrollo del proyecto, una quinta actividad: la elaboración de una bitácora de diseño, que pretende convertirse en una descripción detallada de cómo es tratada cada etiqueta, y que servirá para realizar las modificaciones requeridas posteriormente. No obstante, si una revista ya cuenta con una versión en HTML establecida, dicha tarea puede ser motida, de modo que sea posible continuar con el flujo de actividades propuesto. Como se mencionó anteriormente, nuestra revista no contaba con una primera versión en el formato mencionado.

**Fig. 4. Actividades del proceso de programación en cada etiqueta**



En la Figura 5 se muestra el ejemplo de los códigos del ISSN de la revista, su etiquetado en XML (como una subetiqueta de <journal-meta>, que a su vez está dentro de <front>), su "llamado" en el código XSL por medio de la ruta 'article/front/journal-meta/issn', así como su diseño en CSS por medio del atributo 'cornisa' y, finalmente, el resultado en el HTML. Cabe mencionar que se trata de tres archivos diferentes (XML, XSL y CSS) que, por medio de la aplicación, dan como resultado la creación automatizada del HTML.

**Figura 5. Vistas parciales de los códigos XML, XSL y CSS para la obtención del ISSN de la revista en la versión HTML**



## Resultados

Se obtuvieron, en las primeras versiones de los artículos, las hojas de estilos XSL y CSS, que trasformaron el contenido completo de ocho textos en XML JATS, con cincuenta instrucciones (sentencias) en XSL, y veinte en el CSS, con la omisión desde el archivo JATS del etiquetado en detalle (de fórmulas y tablas). Además, se redactaron tres documentos breves: una revisión sucinta del JATS, una descripción del proceso de trasformación y una bitácora de diseño.

## Conclusiones

Esta propuesta incide de manera directa en cómo se editan y producen los contenidos de una revista de investigación. Es un paso importante para la gestión de artículos en el ámbito digital. La versión HTML obtenida es susceptible de mejoras, en la medida que se incorporen más herramientas de las transformaciones XSLT.

Se espera que este procedimiento le sirva a las publicaciones que busquen iniciar la automatización de sus procesos editoriales. Además, la prospectiva se relaciona con el desarrollo de la edición electrónica en sí, ya que es posible codificar los archivos XSL respectivos para otros formatos, atendiendo a un desarrollo semejante.

## Referencias

Becerril, A. 2017. “Mesa 4. Retos de las revistas electrónicas. Repositorios y datos abiertos” [video, 49:22-50:00]. Conferencia impartida en el *Encuentro de editores de revistas científicas en ciencias sociales: Oportunidades y desafíos*. 17 de noviembre. Acceso el 04-08-2018. [https://www.youtube.com/watch?v=gzu\\_aJfG0Bo](https://www.youtube.com/watch?v=gzu_aJfG0Bo).

Cervantes Ojeda, J., Gómez, M. C. 2012. “Taxonomía de los modelos y metodologías de desarrollo de software más utilizados”. *Universidades* núm. 52: 37-47. Acceso el 4 de agosto del 2018. <http://www.redalyc.org/pdf/373/37326902005.pdf>.

Delgado, E., Ruiz, R., Jiménez, E. 2006. *La edición de revistas científicas; directrices, criterios y modelos de evaluación*. Granada: Universidad de Granada. Acceso el 4 de agosto del 2018. <https://www.revistacomunicar.com/pdf/2011-04-Delgado.pdf>.

Institute of Electrical and Electronics Engineers. 2008. *IEEE Standard for Software and System Test Documentation [IEEE Std 829™-2008]*. USA: Software & Systems Engineering Standards Committee.

Jiménez, C. 2012. *La edición técnica de la revista Científica* [Informe de actividad profesional]. México: Universidad Nacional Autónoma de México. Acceso el 4 de agosto del 2018. <http://132.248.9.195/ptd2015/anteriores/filosofia/0723689/Index.html>.

Latindex. *Nueva lista de características de calidad editorial. Latindex Catálogo*. México, Universidad Nacional Autónoma de México. (2017). Acceso el 4 de agosto del 2018. <https://drive.google.com/file/d/0B21qLJD16Sx3OWJTV3h1Q2gzb2M/view>.

National Information Standards Organization (NISO). 2015. *JATS: Journal Article Tag Suite, version 1.1*. USA: National Information Standards Organization. Acceso el 4 de agosto del 2018. <https://www.niso.org/publications/z3996-2015-jats>.

Oracle. 2017. “NetBeans IDE”. Oracle. <https://netbeans.org/index.html>.

RedALyC. 2016. “XML JATS en Redalyc”. UAEM Redalyc 3.0. Acceso el 4 de agosto del 2018. <https://xmljatsredalyc.org/>.

RedALyC. 2017. “Marcalyc. Sistema de marcación XML-JATS.” UAEM Redalyc 3.0. Acceso el 4 de agosto del 2018. <http://marcalyc.redalyc.org/>.

SciELO México. “Evaluación, admisión y permanencia de las revistas en la colección SciELO México”. SciELO México. Acceso el 4 de agosto del 2018. <http://www.scielo.org.mx/avaliacao/EvaluacionAdminPermanencia.htm>.

Sommerville, I. 2011. *Ingeniería de Software*. México: Pearson.

World Wide Web Consortium. 2017. “The Extensible Stylesheet Language Family (XSL).” W3C. Acceso el 4 de agosto del 2018. <https://www.w3.org/Style/XSL/>.

World Wide Web Consortium. 2015. “Extensible Markup Language (XML).” W3C. Acceso el 4 de agosto del 2018. <https://www.w3.org/XML/>.