

Adopción del estándar XML JATS en SciELO España

Cristina Bojo Canales.

Biblioteca Nacional de Ciencias de la Salud,
Instituto de salud Carlos III, Ministerio de Economía, Industria y Competitividad.

cbojo@isciii.es



¿Qué es el XML?

- EL **X**ML (eXtensible Markup Language, o lenguaje de Marcación Extendido) es una **metalenguaje de marcación recomendado por W3C**.
- Creado para ser auto descriptiva, **transportar y almacenar datos**.
- Permite **identificar claramente los elementos bibliográficos** que componen un documento.
- Permite **verificar la estructura y asegurar que cumple con las especificaciones** con que debe contar el contenido de acuerdo con las reglas de formación.
- Lenguaje que puede ser **entendido tanto por humanos como por máquinas**.



¿Por Qué XML?

- El **HTML** fue creado para presentar el texto (se trata de un lenguaje para **exhibición**).
- El XML permite la **separación entre estructura y exhibición**.
- El mismo texto puede utilizarse de **distintas maneras**.
- Mayor **reutilización** del contenido para producir otros formatos (Ej.: ePub).
- Garantiza mayor **compatibilidad** con sistemas o **formatos futuros**.
- Capacidad de **preservación digital** del contenido a **largo plazo**, sin importar los **sistemas operativos o programas de cómputo** que existan en el futuro.
- Facilita la interoperabilidad entre sistemas y bases de datos.
- Abre la posibilidad de ingreso de las revistas de **Ciencias de la Salud en el PMC**.



¿Por Qué XML?

Es un lenguaje que permite definir reglas ue especifica cómo marcar las partes significativas de un texto, incluyendo palabras, frases, números, fórmulas, etc.

En el caso de artículos científicos, estos siguen una determinada estructura: título, autores, afiliaciones, resumen...

XML se usa para estructurar precisamente todos los elementos de los artículos y otros tipos de documentos. Cada elemento está definido por una etiqueta (tag en inglés).

Los textos estructurados son susceptibles de procesamiento exhaustivos por programas de ordenador, permitiendo la extracción de metadatos del y formar/construir su referencia bibliográfica.

Los textos marcados son susceptibles de ser almacenados en bases de datos e intercambiados entre sistemas en la Web.



XML – Posibilita ...

**Exhibición en varios
Dispositivos**



**Salida en diversos
formatos**





JATS: Journal Article Tag Suite

- ✓ **Journal Article Tag Suite (JATS)** es una norma (NISO Z39.96-2012) que **define y describe el conjunto de elementos y atributos** para la identificación y marcación de los elementos de artículos científicos.
- ✓ La norma fue creada por la National Library of Medicine (NLM) para la estructuración de textos completos del repositorio PubMed Central
- ✓ Compatible con la versión 3.0 de **NLM (National Library of Medicine)**, su antecesora.
- ✓ **JATS es flexible**, o sea, la **mayoría de sus elementos no son obligatorios**.

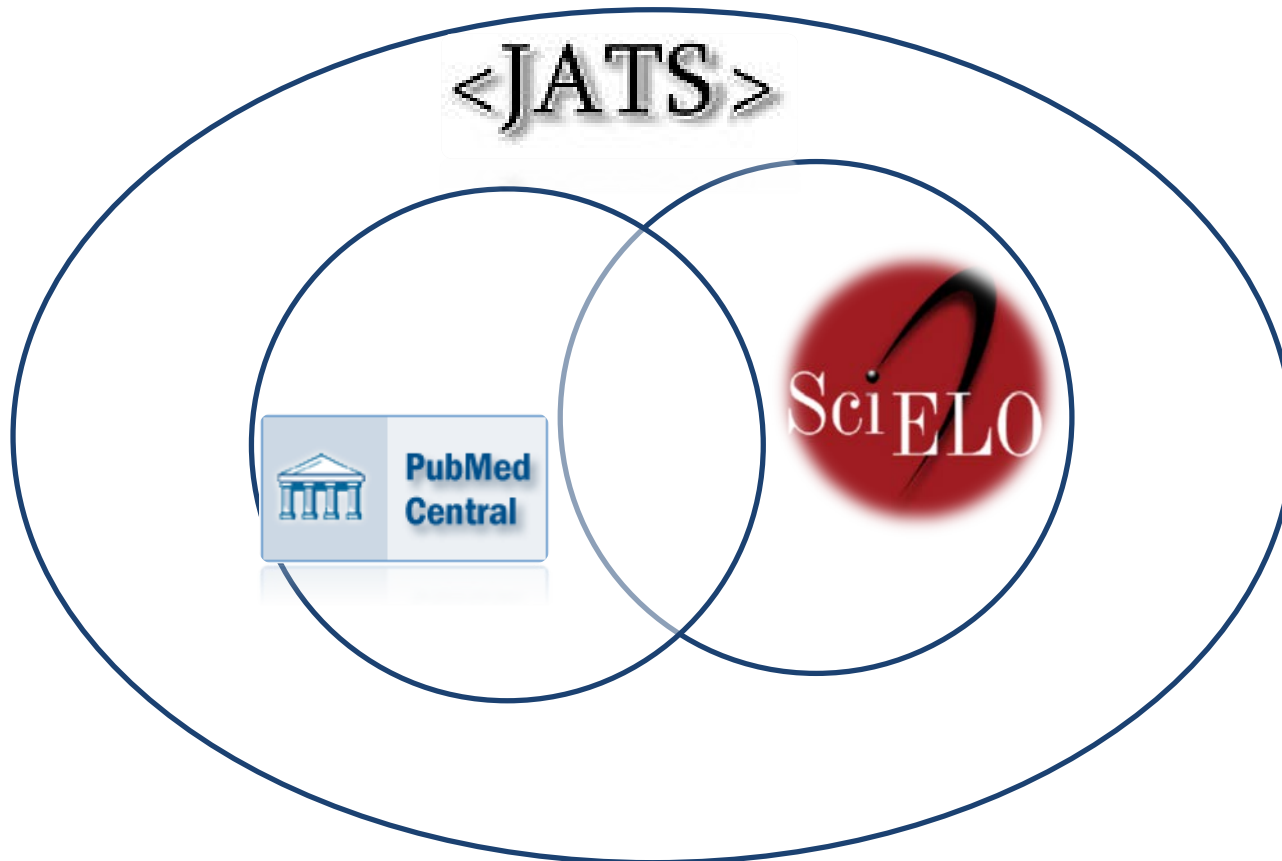


SciELO Publishing Schema (SciELO PS)

- SciELO Publishing Schema (SciELO PS) basado en PMC y JATS.
- El SciELO PS se basa en un subconjunto de JATS y se compone a partir de NISO JATS Journal Publishing DTD en version 1.0 + el Estilo SciELO, que son las reglas que especializan aspectos de la especificación JATS Publishing
- Para atender las necesidades de SciELO, el SciELO PS da prioridad a la identificación de los elementos importantes para generación de los índices bibliométricos y garantiza la interoperabilidad de los datos con otras bases de datos como WOS/PMC...
- Actualmente se encontra en la versión 1.6



JATS, PMC y SciELO PS



Diferencias entre **PMC** y **SciELO PS**



SciELO PS

1. El Objetivo principal es **publicar el texto completo**.
2. Sigue **norma Vancouver** para las referencias.
3. Identificación demasiado **sencilla de las afiliaciones**.
4. **No** es obligatorio identificar datos de **financiación ni registros de ensayos clínicos**.

1. Además de la **publicación** se generan **indicadores de producción** científica.
2. Las revista publicadas adoptan **distintas normas** y no solamente Vancouver.
3. Identificación **detallada de las afiliaciones**.
4. Se identifican datos de **financiación y registros de ensayos clínicos**.

XML SciELO: especificaciones adicionales al JATS

Afiliación

- Se identifican cada uno de los **elementos de la afiliación** lo que permite el análisis detallado del origen de la producción científica indexada y la generación de índices para búsqueda por afiliaciones.

```
<aff id="aff1">  
  <label>1</label>  
  <institution content-type="original"> Departamento de Medicina Preventiva, Salud Pública e Historia de la Ciencia,  
  Facultad Medicina, Universidad Complutense de Madrid.</institution>  
  <institution content-type="normalized">Universidad Complutense de Madrid</institution>  
  <institution content-type="orgdiv2">Departamento de Medicina Preventiva, Salud Pública e Historia de la Ciencia</institution>  
  <institution content-type="orgdiv1">Facultad Medicina</institution>  
  <institution content-type="orgname">Universidad Complutense de Madrid</institution>  
  <country country="ES">Spain</country>  
</aff>
```



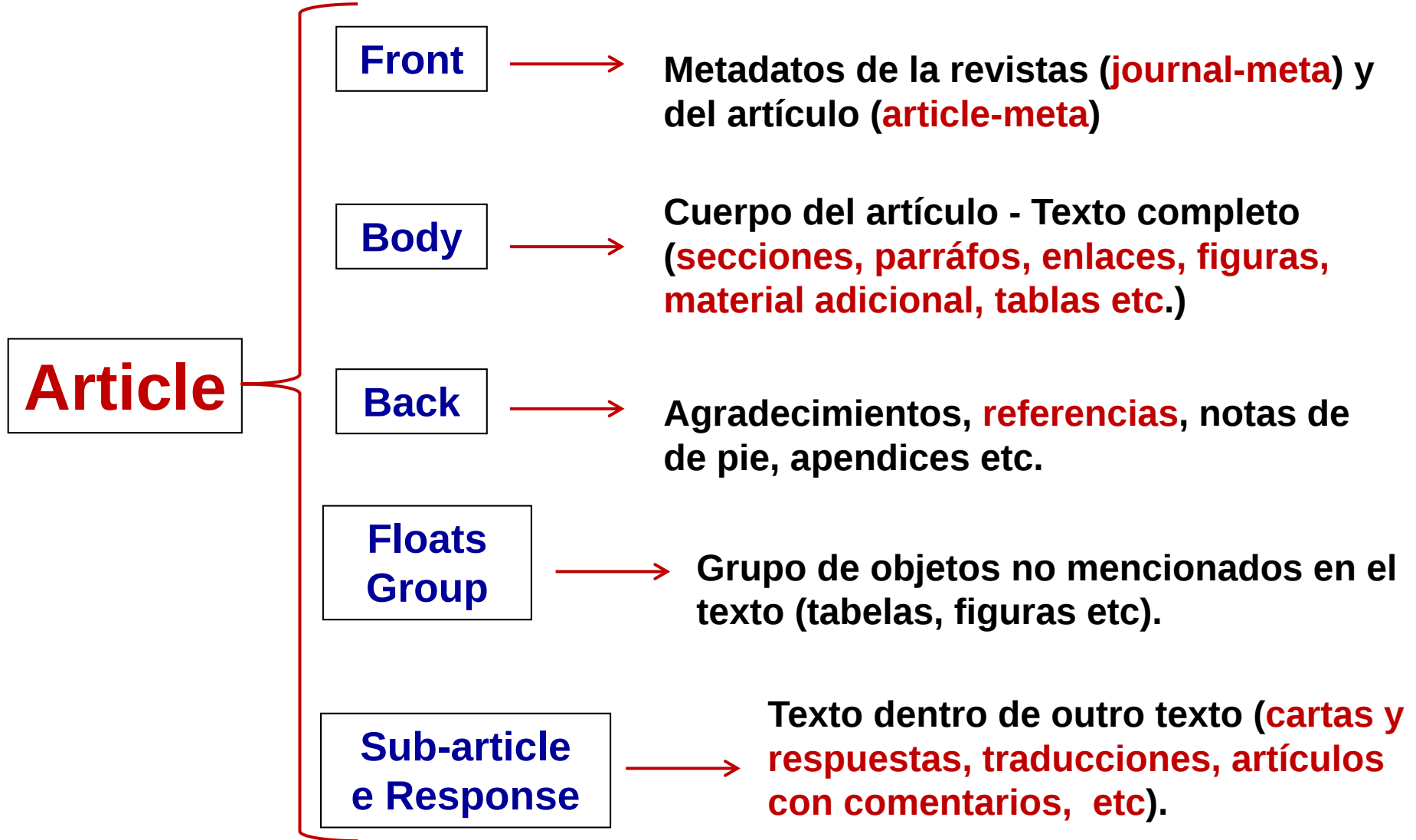
XML SciELO:

especificaciones adicionales al JATS

Datos de Financiación

```
<article>
<front>
...
<article-meta>
...
<kwd-group kwd-group-type="author">...</kwd-group>
<funding-group>
  <award-group>
    <funding-source>Consejo de Desarrollo Científico y Humanístico de la Universidad del Zulia</funding-source>
    <award-id>CC-0627-04</award-id>
  </award-group>
</funding-group>
</article-meta>
</front>
<body>...</body>
<back>
<ack>
  <sec>
    <title>ACKNOWLEDGMENTS</title>
    <p>Research partially funded by Consejo de Desarrollo Científico y Humanístico de la Universidad del Zulia
      (Condes-Luz, Venezuela); Project # CC-0627-04. To Nivia Abreu and Mary Nelly Ramos for their laboratorial support.</p>
  </sec>
</ack>
...
</back>
</article>
```

XML SciELO: estructura



XML SciELO: herramientas

SciELO pone a disposición de sus editores herramientas de generación de XML y validación de archivos

<http://docs.scielo.org/projects/scielo-pc-programs/en/latest/>

- Programa de marcación de textos (Markup)
- Generación de xml: XML Package Maker (XPM)
- Herramienta de validación de estructura online (Style Checker)



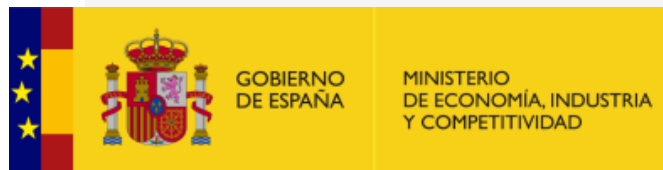
SciELO Style Checker

Utilize esta ferramenta para verificar se um arquivo XML está em conformidade com o Estilo SciELO como descrito no Guia de Identificação de Elementos do SciELO Publishing Schema.

Selecione o arquivo XML e clique em "Validar". Os resultados serão exibidos abaixo.

[Ajuda?](#)

Tamanho máximo: 512,0 KB



SciELO España: situación

De las 43 revistas vigentes en la colección:

- 24 revistas ya publican en xml de forma regular
- 7 revistas están corrigiendo sus xml (han enviado ya pruebas y están añadiendo las correcciones para adaptarlas correctamente al schema)
- 12 revistas no han enviado ningún xml (pero tienen previsto el cambio o están empezando a elaborar el material).

iiii lo estamos
logrando!!!



SciELO España: situación

..... Pero tenemos que seguir aprendiendo y corregir:

- **Afiliaciones:** identificación completa de la entidad donde los autores desarrollan su actividad profesional, incluidos país y ciudad. En orden decreciente y sin acrónimos, siglas ni abreviaturas, en el idioma original.

“Hospital Ramón y Cajal. Servicio de Oftalmología. Unidad de Glaucoma. Madrid, España”.

SciELO España: situación

Algunos ejemplos:

Técnico en Neurofisiología clínica. E-mail marilyn@neuro.ciren.cu

**Doctora en Ciencias. Jefa del Servicio de Neurofisiología clínica

M.E. Vázquez Fernández^a, M.F. Muñoz Moreno^b, A. Fierro Urturi^c, M. Alfaro González^d, M.L. Rodríguez Carbajo^e y L. Rodríguez Molinero^f

^aPediatra. CS Arturo Eyries. Valladolid. España

^bUnidad de Investigación Biomédica. Hospital Clínico Universitario. Valladolid. España

^cPediatra. CS Pisuegra. Valladolid. España

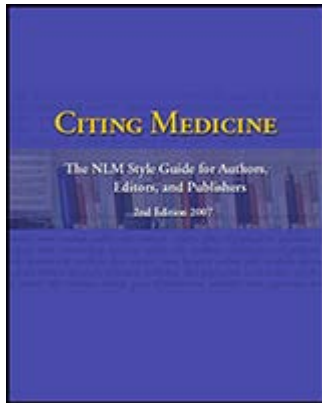
^dServicio de Pediatría. Hospital de Medina del Campo. Medina del Campo, Valladolid. España

^eMIR-MFyC. CS Arturo Eyries. Valladolid. España

^fPediatra. CS Casa del Barco. Valladolid. España

SciELO España: situación

Ojo con las referencias bibliográficas



Citing Medicine, 2nd edition The NLM Style Guide for Authors, Editors, and Publishers
Karen Patrias; Dan Wendling, Technical Editor.
National Library of Medicine, National Institutes of Health, Bethesda (MD): National Library of Medicine (US); 2007

Citing Published Print Documents

Citing Unpublished Material

Citing Audio and Visual Media (audiocassettes, videocassettes,

slides, photographs, etc.)

Citing Material on CD-ROM, DVD, or Disk

<https://www.ncbi.nlm.nih.gov/books/NBK7256/>

SciELO España: situación

Ojo con las referencias bibliográficas

- Errores en los datos numéricos de localización: año, número de páginas...
- Errores en los nombres de los autores. Un mismo autor y un mismo trabajo citado había sido recogido con nombres diferentes (por ejemplo. el trabajo de Pérez AM y Pérez M)
- Faltan datos (año, autor...)
- Referencias presentadas en distinto formato
- Referencias electrónicas que no conducen a los documentos referenciados por incluir vínculos inexactos o incompletos.

SciELO España: situación

- Tablas y Figuras **sin llamada** dentro del texto
- **Fechas** de Aceptado y Recibido
- Declaración de **fuentes de financiación** y **conflictos de interés**
- **Licencia CC**