# **Enriquecimiento de XML JATS: fortaleciendo los metadatos de la publicación científica**

## **Introducción**

En el contexto de la publicación académica en entornos digitales, la calidad operativa de un artículo para los sistemas editoriales e infraestructuras de información no se define únicamente por su contenido textual, sino también por la cantidad y precisión de sus metadatos.

En el contexto de la producción científica digital, la calidad de un artículo no se define únicamente por su contenido textual, sino también por la cantidad y precisión de sus metadatos. Estos metadatos constituyen la base sobre la cual se apoyan los procesos de indexación, interoperabilidad, preservación y evaluación de la producción académica. En este marco, el estándar XML JATS se ha consolidado como un pilar fundamental para la representación estructurada de artículos científicos. Este estándar proporciona un marco robusto y expresivo para modelar la complejidad inherente a los artículos: autores y roles, afiliaciones, identificadores persistentes, información editorial, lenguajes, licencias y otros elementos contextuales.

Sin embargo, la sola adopción del estándar no garantiza, por sí misma, la calidad del resultado final. La experiencia demuestra que gran parte del valor de JATS emerge cuando estos metadatos se encuentran completos, normalizados y coherentemente integrados en el flujo editorial.

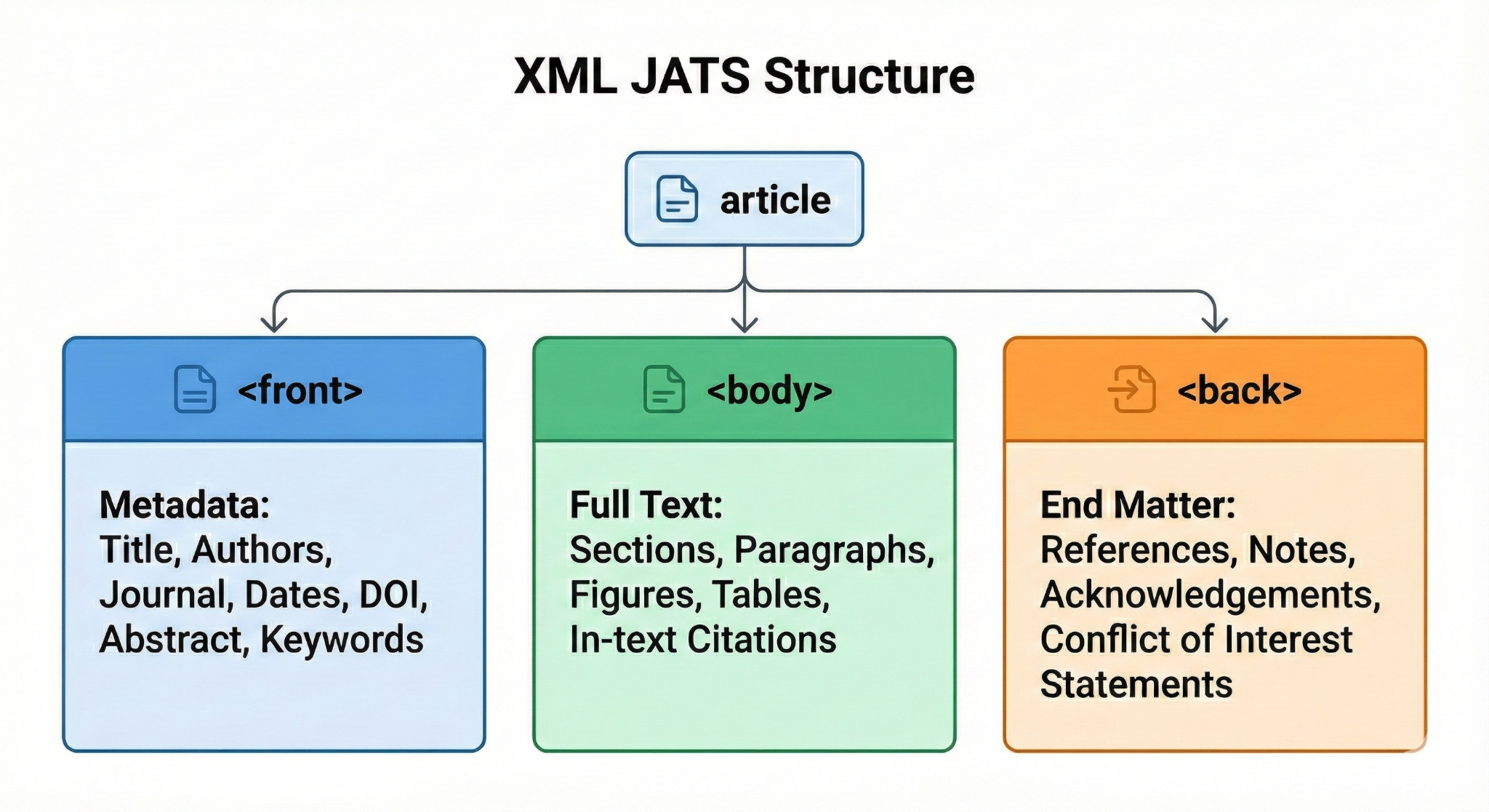
Desde SUMARC, se desarrolló un plugin orientado al **enriquecimiento de XML JATS**, con el objetivo de complementar y completar los metadatos de los trabajos académicos en OJS. Este artículo presenta el marco conceptual que sustenta este trabajo, detalla la importancia del estándar JATS y profundiza en el rol central que cumplen los metadatos en los flujos editoriales contemporáneos, dejando planteada la base para la explicación del funcionamiento técnico del plugin.

## **¿Qué es XML JATS y qué define como estándar?**

XML JATS (Journal Article Tag Suite) es un estándar mantenido por NISO que define un conjunto de elementos y atributos para describir de manera estructurada el contenido y los metadatos de artículos científicos. Su objetivo principal es proporcionar un modelo común que permita representar, de forma semánticamente explícita, tanto la estructura lógica del artículo como la información contextual que lo rodea.

JATS no impone un formato visual ni un estilo editorial específico; en cambio, establece qué información debe y/o puede estar presente y cómo debe estructurarse, dejando libertad y facilitando el proceso para su posterior transformación a distintos formatos de salida. Esta separación entre estructura semántica y presentación es una de las razones centrales de su adopción en sistemas editoriales académicos.

Desde el punto de vista estructural, un documento XML JATS se organiza jerárquicamente en tres grandes secciones principales: **<front>**, **<body>** y **<back>**, contenidas dentro del elemento raíz **<article>**. La sección <front> agrupa los metadatos descriptivos del artículo, como información bibliográfica, autorías, afiliaciones, resúmenes y datos editoriales. El <body> contiene el texto principal del artículo, organizado en secciones y subsecciones que representan la estructura lógica del contenido científico. Finalmente, el <back> reúne los elementos complementarios, como referencias bibliográficas, agradecimientos, apéndices y material adicional. Esta organización jerárquica garantiza una representación explícita y consistente de la información (véase Figura 1).



Organizaciones, repositorios y sistemas de indexación internacionales han adoptado JATS como formato preferente o requerido, convirtiéndose en un estándar operativo de facto para la publicación científica digital.

## **El rol central de los metadatos en un artículo científico**

Los metadatos no constituyen un elemento accesorio del artículo científico, sino que forman parte indisoluble de su definición estructural y semántica. JATS concibe al artículo como una unidad de contenido y contexto, donde este último es vital para la circulación y validación del conocimiento.

Desde una perspectiva técnica, los metadatos hacen que el contenido sea procesable por máquinas; editorialmente, permiten su correcta identificación y evaluación; e institucionalmente, garantizan la interoperabilidad y preservación a largo plazo. Un XML JATS con metadatos completos trasciende la simple visualización, habilitando procesos críticos como:

* Indexación en bases de datos y cumplimiento con plataformas (SciELO, Redalyc, DOAJ).
* Apoyo a la generación de métricas, estadísticas y reportes más completos.
* Generación automática de métricas, estadísticas y reportes.
* Vinculación con identificadores persistentes y perfiles de autor.
* Validación estructural y semántica del contenido.
* Preservación del conocimiento en el largo plazo.

En consecuencia, el enriquecimiento de metadatos es una etapa indispensable del flujo editorial para asegurar la calidad técnica y la proyección institucional de la producción científica. Entre los metadatos más relevantes destacan: autores y roles editoriales, afiliaciones institucionales, identificadores persistentes (DOI, ORCID, etc.), información editorial, idiomas, resúmenes, palabras clave y licencias. Sin ellos, el artículo científico corre el riesgo de convertirse en un objeto digital aislado, invisible para los motores de búsqueda académicos e incapaz de dialogar con la infraestructura global de investigación.

La carencia de estos elementos estructurales rompe la cadena de comunicación científica: impide la correcta atribución de créditos a autores e instituciones, bloquea la generación de métricas fiables y anula la trazabilidad y estabilidad que garantizan los identificadores persistentes. Del mismo modo, la falta de información editorial precisa, sumada a la omisión de resúmenes o licencias claras, obstaculiza la interoperabilidad técnica con bases de datos y cosechadores, haciendo imposible que el contenido sea correctamente indexado, citado o utilizado legalmente, lo que en la práctica condena al trabajo a una menor relevancia a pesar de su calidad intelectual.

## **Estrategia integral de calidad de metadatos**

El enriquecimiento no debe entenderse como una acción aislada o correctiva al final del camino, sino como un eslabón más dentro de la estrategia integral de completitud de metadatos. Al igual que la preservación digital se plantea desde el origen, aquí la calidad de los metadatos se aborda desde el nacimiento del artículo. Sistemas como SUMARC orquestan un conjunto de herramientas y procesos que atraviesan todo el flujo editorial —desde la conversión de DOCX a XML JATS y la generación estructurada de citas, hasta la edición del XML—; en este esquema, el plugin de enriquecimiento actúa para consolidar, normalizar y potenciar ese trabajo previo, asegurando que la información gestionada durante todo el proceso se refleje en el resultado final.

Por otro lado, el enriquecimiento de metadatos en XML JATS adquiere un carácter estratégico que excede la mera corrección técnica del documento. La completitud, normalización y coherencia semántica de los metadatos constituye un factor clave para garantizar la interoperabilidad de las revistas científicas de la Universidad Nacional de La Plata con infraestructuras de publicación e indexación que imponen estándares formales exigentes, entre ellas SciELO.

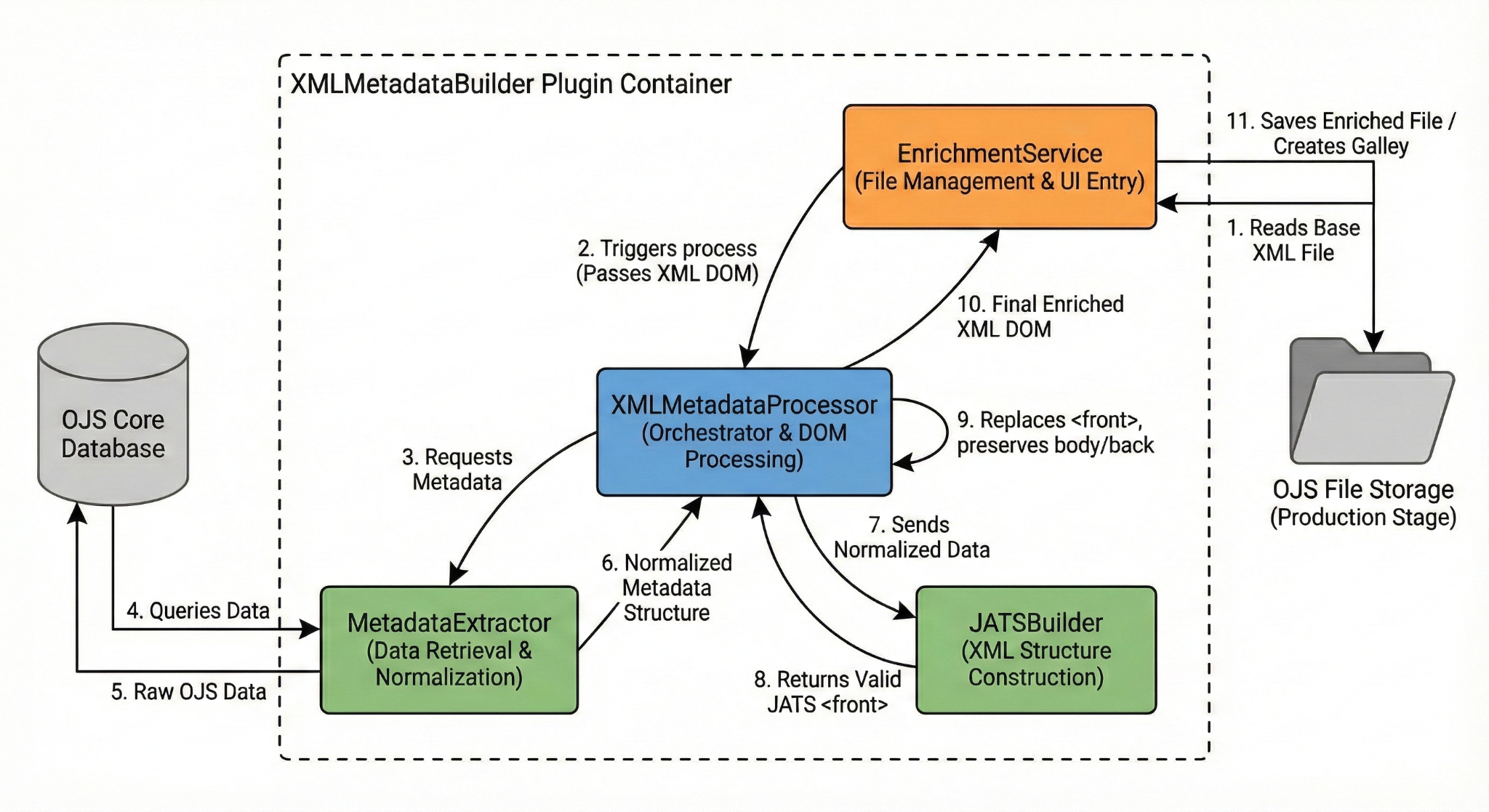
Diversos sistemas de publicación científica concentran buena parte de sus requisitos técnicos precisamente en la presencia explícita y correctamente estructurada de los metadatos. La ausencia, ambigüedad o inconsistencia de esta información suele derivar en observaciones técnicas recurrentes, incrementa la carga de trabajo editorial y dificulta el cumplimiento de los criterios de indexación. En este sentido, disponer de XML JATS enriquecidos permite anticiparse a estas problemáticas y facilita los procesos de validación e integración con plataformas externas, reduciendo fricciones y trabajos posteriores.

El enriquecimiento de metadatos resulta igualmente central para las estrategias de preservación digital. En repositorios digitales, los XML JATS completos facilitan los procesos de ingestión, disminuyen la dependencia de interpretaciones implícitas y aseguran que la información editorial y contextual del artículo permanezca accesible e interpretable a largo plazo, independientemente de los sistemas de visualización o publicación que se utilicen en el futuro.

De este modo, se incorpora un proceso sistemático que no se limita a completar campos faltantes, sino que apunta a normalizar valores, explicitar decisiones editoriales (como fechas de aceptación o revisión) y asegurar consistencia semántica y estructural en el XML JATS, siempre que la información se encuentre disponible en OJS. Esto reduce correcciones manuales, disminuye errores en instancias de validación externa y mejora sustancialmente la calidad y previsibilidad del documento final. El nuevo desarrollo fortalece la interoperabilidad, sostenibilidad y proyección de las revistas científicas, en un ecosistema basado en OJS, estándares abiertos y desarrollo institucional soberano.

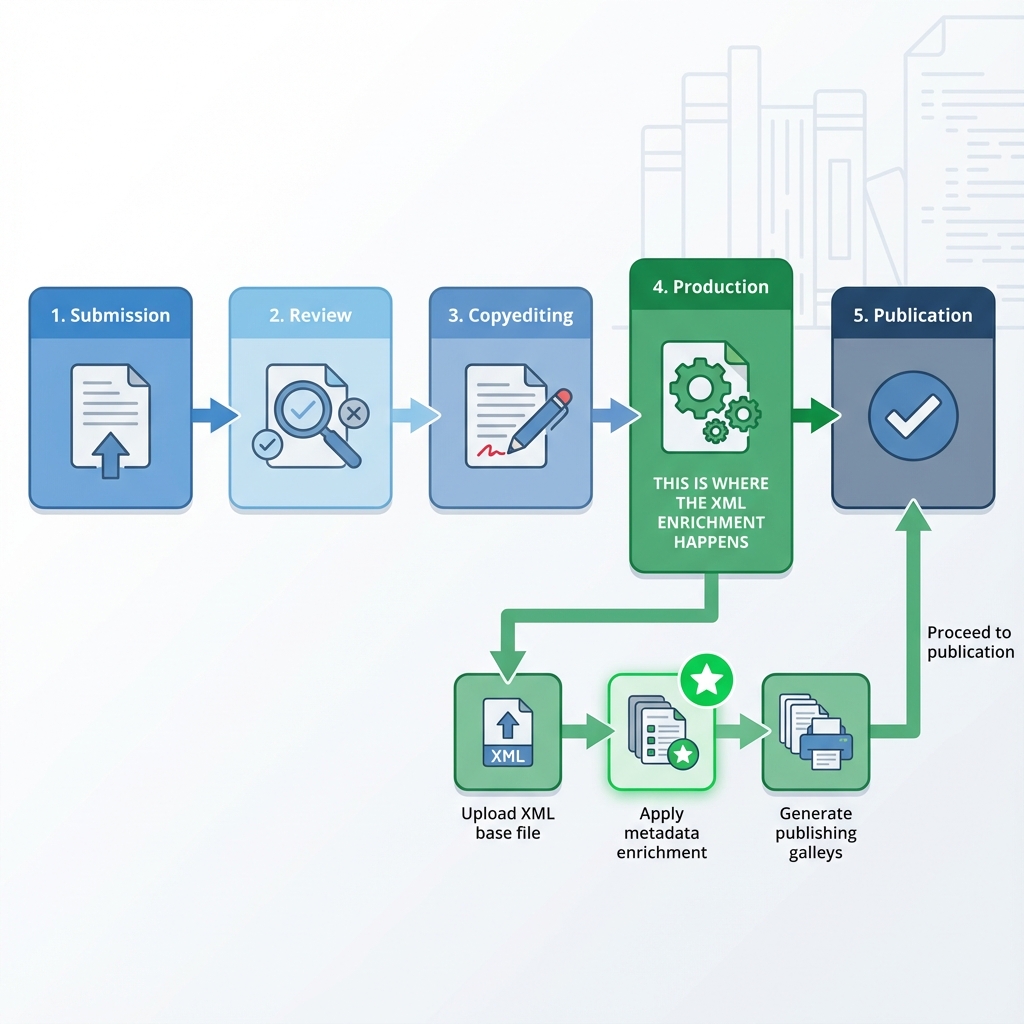
## **Arquitectura general del plugin de enriquecimiento**

El plugin **XMLMetadataBuilder** adopta una arquitectura modular en capas, diseñada para integrarse con Open Journal Systems (OJS), actualmente en su versión 3.4 pero fácilmente adaptable a otras versiones, y para desacoplar claramente la extracción de metadatos, la construcción del XML JATS y la gestión de archivos editoriales. Esta organización facilita el mantenimiento, la extensibilidad del sistema y el cumplimiento consistente de estándares de publicación científica. La **Figura 2** presenta una visión general de esta arquitectura.



### **Integración en el flujo editorial**

El plugin se activa específicamente en la fase de *Producción* (Figura 3) del flujo editorial de OJS, una vez que el artículo ha superado las etapas de revisión y corrección editorial. En este momento, el editor o la persona correspondiente selecciona un archivo XML marcado como *production-ready* desde la pestaña “Enriquecedor XML” en *Publicación*, elige si desea sobrescribir el mismo archivo o no y/o generar una galerada, y ejecuta el enriquecimiento previo a la generación de otros formatos publicables.



### 

### **Componentes principales del plugin**

El plugin se estructura en cuatro componentes centrales que trabajan de manera coordinada siguiendo un flujo claramente definido: extracción de metadatos, construcción del XML, orquestación del proceso y gestión de archivos y galeradas.

#### **MetadataExtractor: extracción y normalización de metadatos**

El componente **MetadataExtractor** es responsable de obtener todos los metadatos relevantes desde OJS y normalizarlos en una estructura intermedia independiente del formato XML. Actúa como puente entre los objetos internos del sistema editorial (revista, artículo, autores, fechas y permisos) y las etapas posteriores del procesamiento. Al concentrar toda la lógica de acceso a los datos de OJS, este componente abstrae su complejidad y produce una representación homogénea que facilita la construcción posterior del XML JATS.

#### **JATSBuilder: construcción del XML conforme a estándares**

El **JATSBuilder** transforma los metadatos normalizados en una estructura XML conforme al estándar JATS Publishing Schema, y en gran medida compatible con perfiles como el utilizado por SciELO. Su responsabilidad es exclusivamente estructural y semántica: construye nodos XML válidos respetando el orden de los elementos, los atributos requeridos y las reglas del estándar. Este componente genera el elemento <front> completo del documento JATS, centralizando la lógica de conformidad y asegurando consistencia entre los documentos producidos.

#### **XMLMetadataProcessor: orquestación del enriquecimiento**

El **XMLMetadataProcessor** cumple el rol de orquestador del proceso de enriquecimiento y opera directamente sobre el documento XML original cargado como un árbol DOM. Coordina la extracción de metadatos y la construcción del nuevo <front>, reemplazando la sección existente por una versión enriquecida y preservando íntegramente el cuerpo del artículo y las referencias bibliográficas del XML original.

#### **EnrichmentService: gestión de archivos y publicación**

El componente **EnrichmentService** encapsula la lógica relacionada con la gestión de archivos dentro de OJS y constituye el punto de entrada desde la interfaz de usuario. Se encarga de leer los archivos XML originales, invocar el procesamiento de enriquecimiento y persistir los nuevos archivos generados en los repositorios del sistema. Asimismo, gestiona la creación de galeradas publicables y la copia de archivos dependientes, integrando el enriquecimiento de forma transparente en el flujo editorial.

### **Síntesis de la arquitectura**

En conjunto, la arquitectura del plugin permite transformar automáticamente los metadatos estructurados de OJS en documentos XML JATS completos, válidos y coherentes, mediante una separación clara de responsabilidades.

acá se podría mostrar snippets del código xml antes y despues de ser enriquecido.

## **Impacto en la interoperabilidad y validación (2da versión propuesta)**

El análisis de los XML JATS generados en el marco del flujo editorial de SUMARC, evaluados mediante el validador provisto por SciELO, permitió identificar que una proporción significativa de los errores y advertencias detectados en los artículos se encontraba asociada a la ausencia o incompletitud de determinados metadatos.

En este contexto, el desarrollo de la nueva herramienta de enriquecimiento representa una mejora sustancial sobre las capacidades previamente disponibles en SUMARC. Los XML JATS enriquecidos muestran una reducción considerable en la cantidad de errores y advertencias reportadas por el validador de SciELO y, al mismo tiempo, un mayor grado de alineación con los lineamientos definidos por JATS4R (JATS for Reuse). Este cumplimiento se manifiesta principalmente en la incorporación explícita, consistente y normalizada de metadatos, lo que incrementa la capacidad de reutilización, intercambio e interpretación automática de los artículos en diversos contextos de consumo y preservación.

No obstante, los resultados también evidencian que el enriquecimiento de metadatos, si bien constituye un factor clave para mejorar la interoperabilidad y facilitar la adecuación a los requerimientos de SciELO, no resulta por sí solo suficiente para garantizar la aprobación automática de todos los artículos. Persisten observaciones técnicas vinculadas a criterios específicos de marcación definidos por organismos como SciELO, que imponen reglas más estrictas y particulares, en algunos casos más allá de lo establecido por el estándar JATS y sus perfiles.

Aún con estas limitaciones, el enriquecimiento sistemático de metadatos introduce una mejora sustancial en la previsibilidad del proceso de validación. La reducción del volumen y la complejidad de las correcciones necesarias impacta directamente en la aceleración de los tiempos de adecuación a los perfiles técnicos exigidos por plataformas como SciELO. Este efecto es especialmente relevante debido a que el proceso de enriquecimiento se encuentra integrado de manera nativa dentro del flujo editorial de OJS a través de SUMARC, lo que evita la dependencia de herramientas externas o de terceros, cuyo funcionamiento interno suele ser opaco para el público general y difícil de auditar o adaptar a necesidades locales.

En esta misma línea, el análisis permitió observar que incluso artículos científicos previamente marcados utilizando herramientas provistas por organismos como SciELO y Redalyc se ven favorecidos por la aplicación del proceso de enriquecimiento. En la **figura X** se puede observar la cantidad de metadatos nuevos en **X** artículos realizados con las herramientas de SciELO y Redalyc que pasaron el proceso de enriquecimiento.

acá iría una figura mostrando la diferencia entre un xml jats generado con scielo, y ese mismo luego de ser enriquecido. podría mostrarse como código, o con una estadísticas de cuántos datos se agregan, etc. los últimos análisis nos dicen que introducimos varios metadatos nuevos, por lo que podemos afirmar que nuestros xml están más completos en cuańto al front.

En promedio, se mejoró un **X%** el número de metadatos. Esto refuerza la idea de que la herramienta no sólo complementa y mejora el marcado existente, sino que aporta un valor diferencial al integrarse directamente en el flujo editorial, potenciando las capacidades de SUMARC y fortaleciendo la interoperabilidad, preservación digital y calidad técnica de los XML JATS generados.

## **Conclusión**

El enriquecimiento de metadatos en XML JATS constituye un componente fundamental para garantizar la calidad, interoperabilidad y visibilidad de los artículos científicos. Más que un aspecto técnico accesorio, los metadatos definen en gran medida qué usos, integraciones y procesos pueden realizarse sobre una publicación dentro del ecosistema académico.

La experiencia desarrollada en SUMARC muestra que contar con herramientas específicas para enriquecer y normalizar metadatos permite fortalecer los flujos editoriales, generar XML JATS más completos y alineados con los requerimientos de organismos como SciELO, y mejorar la integración con infraestructuras de indexación y preservación. En este sentido, el desarrollo institucional de soluciones basadas en estándares abiertos resulta clave para adaptar estos modelos a las prácticas locales de las revistas.

En conclusión, el enriquecimiento sistemático de metadatos no solo mejora la calidad formal de los artículos, sino que constituye una estrategia central para consolidar infraestructuras editoriales sostenibles, interoperables y orientadas al acceso abierto.