

Importación de artículos en OJS desde Dspace

Eje temático: Interoperabilidad y preservación digital

Tipo de trabajo: póster

Palabras clave: revistas académicas, repositorios institucionales, importación, metadatos

Keywords: academic journals, institutional repositories, import, metadata

Autores

Gonzalo Luján Villarreal (contacto principal). PREBI-SEDICI, Universidad Nacional de La Plata; CESGI, Comisión de Investigaciones Científicas. Email: gonzalo@prebi.unlp.edu.ar. Doctor en Ciencias Informáticas, Docente-Investigador de la UNLP, Subdirector del Centro de Servicios en Gestión de Información (CESGI) de la Comisión de Investigaciones Científicas y docente de la Facultad de Informática de la UNLP. Realiza actividades de investigación y desarrollo en PREBI-SEDICI, y coordina los portales de Revistas, de Congresos y de Libros de la UNLP. Trabaja en las áreas de interoperabilidad, visibilidad web y preservación digital.

Ezequiel Manzur. PREBI-SEDICI, Universidad Nacional de La Plata. Email: ezemanzur@sedici.unlp.edu.ar. Estudiante avanzado de Licenciatura en Sistemas, docente en la Facultad de Informática, y pasante en PREBI-SEDICI. Trabaja en el área de visibilidad institucional e interoperabilidad con repositorios digitales, e integración de repositorios con portales web institucionales.

Dolores García. PREBI-SEDICI, Universidad Nacional de La Plata; CESGI, Comisión de Investigaciones Científicas. Email: dolores.garcia@sedici.unlp.edu.ar. Licenciada en Comunicación Social y pasante de la Comisión de Investigaciones Científicas. Brinda soporte a los equipos editoriales de las revistas de la UNLP y a los organizadores de congresos de la UNLP. Trabaja en gestión de calidad de publicaciones periódicas, indización y formatos para difusión y preservación.

Marisa Raquel De Giusti. PREBI-SEDICI, Universidad Nacional de La Plata; CESGI, Comisión de Investigaciones Científicas. Email: marisa.degiusti@sedici.unlp.edu.ar. Doctora en Informática por la Universidad Nacional de La Plata, Argentina. También es Ingeniera en Telecomunicaciones y Profesora de Letras. Ha dirigido el Servicio de Difusión de la Creación Intelectual (SEDICI) de la UNLP desde 2003, y el repositorio CIC-Digital desde 2014, y el Centro de Servicios en Gestión de Información (CESGI) desde 2016. También coordina el Observatorio Medioambiental La Plata. Es profesora en la Facultad de Informática de la UNLP y sus intereses de investigación incluyen el campo de repositorios digitales, preservación digital y acceso abierto.

Introducción

Los repositorios digitales brindan un espacio para preservar y difundir materiales, ya sea nacidos digitales o digitalizados. Por lo general, los repositorios reciben los objetos, digitales o para digitalizar, desde los mismos productores o generadores: editores de revistas, autores de artículos, tesis, etc. Sin embargo, en algunas ocasiones, el flujo de información se invierte, y el repositorio se convierte en fuente de datos para estos productores de contenidos. Esta situación se da, por ejemplo, cuando una revista académica que existe hace mucho tiempo, y que ya es preservada en un repositorio

institucional, comienza a utilizar un sistema como OJS para la gestión de sus publicaciones. Esta situación se ha repetido en la Universidad Nacional de La Plata, donde revistas con varias décadas de publicaciones deciden incorporarse al Portal de Revistas de la UNLP, y para ello deben incorporar todo su historial de artículos en OJS. El objetivo de este trabajo es presentar una herramienta desarrollada para recuperar e importar todo el historial de una revista desde un repositorio DSpace hacia un portal OJS. Asimismo, se propone una metodología para minimizar el trabajo por parte del equipo del repositorio y para agilizar la carga en OJS por parte del equipo que gestiona las revistas.

Materiales y Metodología

Para realizar este trabajo se estudiaron los formatos de exportación de metadatos disponibles en Dspace, así como también los formatos de importación disponibles en OJS. Para la exportación de metadatos se optó por el formato CSV de DSpace, que permite agrupar en un único archivo todos los metadatos de un número, o incluso de todos los números de una revista. Para la importación, se optó por el formato XML nativo de OJS, tanto por su simpleza como por la disponibilidad de plugins de importación en casi todas las versiones de OJS. Se seleccionaron 4 revistas de la UNLP con un gran historial de publicaciones y disponibles en el repositorio SEDICI: JCST, TE&ET, Económica y Anales JURSOC. Se extrajeron los metadatos de estas revistas desde el repositorio SEDICI, se relevaron los distintos esquemas disponibles en la exportación (DC, esquemas propios) y se analizaron requerimientos de normalización, mapeos y post-procesamiento de algunos metadatos (por ej. URI al documento). También se analizó el formato XML nativo de OJS, y se determinaron los elementos disponibles en la exportación CSV de DSpace que podían aprovecharse para generar este XML.

Resultados

Se desarrolló una aplicación que toma como entrada una exportación CSV y genera como salida uno o más documentos XML. Para simplificar la exportación de metadatos, la aplicación permite un único archivo CSV con todos los números y artículos de la revista (lo que requiere una única exportación por parte del repositorio), y es capaz de segmentar este archivo en subconjuntos de artículos, agrupados por el número en el que fueron publicados. Si bien la exportación del CSV incluye sólo metadatos, la aplicación desarrollada es capaz de conectarse con el repositorio DSpace para obtener también el archivo principal del artículo (el PDF publicado), el cual es codificado e incorporado al XML. Como resultado, la aplicación genera múltiples XML, uno por cada número de la revista, que incluyen tanto metadatos como los archivos PDF. Este XML también soporta metadatos en múltiples idiomas, y permite especificar una sección de la revista para depositar los artículos una vez importados en OJS.

A partir de este desarrollo, se propuso una metodología de trabajo con los equipos editoriales: el equipo que gestiona el Portal de Revistas y la administración del repositorio se encargan de la generación de los CSV y XML. Se entregan a los equipos editoriales varios archivos XML, uno por cada número de la revista, y los editores deben importar uno por uno los archivos, lo que resulta en la incorporación en OJS de varios artículos por cada XML importado. Estos artículos son colocados en una sección especial que permanece oculta para el público en general. Esto permite a los editores revisar y completar los metadatos, asignar los artículos al número y sección correspondiente y, finalmente, publicar el número con todos los artículos importados.

Conclusiones

El desarrollo de esta aplicación ha servido para realizar la importación del historial completo de 4 revistas de la UNLP. Si bien la importación no es automática, sino que requiere un trabajo de revisión de metadatos y en algunos casos de carga de información adicional, esta herramienta permite ahorrar mucho tiempo y, al realizarse la importación por cada número, ayuda a organizar el trabajo. El uso de una sección especial también ha servido para evitar confusiones entre artículos en importación y envíos tradicionales en proceso editorial.

Si bien la herramienta ha sido muy útil, existen aún algunos elementos para incorporar o mejorar. Por un lado, sería interesante sumar más metadatos al XML Nativo para OJS, en particular aquellos relacionados con los autores: institución, título, etc. Si bien estos metadatos no están disponibles en la exportación CSV de DSpace, podría considerarse una conexión con el sistema de gestión de autoridades y así recuperar información más completa en estos casos. También sería deseable que la herramienta considere más de un documento por artículo; en este momento, se recupera la URI de un único documento, a partir del análisis del documento HTML que expone el repositorio para cada recurso, pero existen casos en los que un artículo posee también otros archivos: datos crudos, imágenes en alta calidad, etc. Si estos documentos son expuestos en el repositorio, entonces deberían también incorporarse en el XML nativo para OJS. Por último, la herramienta desarrollada funciona desde la línea de comandos, lo que permitió un desarrollo ágil y un funcionamiento con pocos recursos. Sin embargo, esto limita mucho su uso por parte de usuarios sin conocimientos técnicos. Es por ello que se está trabajando en una interfaz gráfica de usuario (GUI), que permita a cualquier persona cargar el archivo CSV generado desde DSpace, especificar algunos parámetros (sección de destino para los artículos, dividir en múltiples XML, recuperar PDF, etc.) y genere como resultado todos los archivos XML listos para importar a OJS.

Referencias

1. PKP Web Application Library. En línea. <https://github.com/pkp/pkp-lib/>
2. Importing and Exporting Data, PKP Wiki. En línea. https://pkp.sfu.ca/wiki/index.php/Importing_and_Exporting_Data . Accedido el 13 de julio de 2018.
3. Journal of Computer Science and Technology, Facultad de Informática de la Universidad Nacional de La Plata En línea. <http://journal.info.unlp.edu.ar/> . Accedido el 13 de julio de 2018.
4. Revista Tecnología en Educación y Educación en Tecnología, Facultad de Informática de la Universidad Nacional de La Plata. En línea. <http://teyet-revista.info.unlp.edu.ar/> . Accedido el 13 de julio de 2018.
5. Revista Económica. En línea. <https://revistas.unlp.edu.ar/economica> . Accedido el 13 de julio de 2018.
6. Revista Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata. En línea. <https://revistas.unlp.edu.ar/RevistaAnalesJurso> . Accedido el 13 de julio de 2018.
7. DSpace2OJS, espacio Github SEDICI UNLP. En línea. <https://github.com/sedici/dspace2ojs> . Accedido el 13 de julio de 2018.
8. Batch Metadata Editing, DuraSpace Wiki. En línea. <https://wiki.duraspace.org/display/DSDOC5x/Batch+Metadata+Editing> . Accedido el

13 de julio de 2018.