

Esta es una transcripción natural del [Podcast APREDIG EP6](#) – Los repositorios de difusión digital institucional en el 2020 y los límites para cubrir la norma 14721 – Sistema abierto de información de archivo (OAIS) para preservación digital. (Entrevista a Marisa De Giusti – Directora de SEDICI y Presidenta de ISTEAC) | Argentina.

*Bienvenidos al podcast sobre preservación digital de APREDIG, la Asociación Iberoamericana de Preservación Digital. Hoy hablaremos con la doctora Marisa De Giusti, quien es directora del repositorio institucional central de la Universidad Nacional de La Plata en Argentina, SEDICI, y presidenta del ISTEAC, el Consorcio Iberoamericano para la Educación en Ciencia y Tecnología.*

—*Marisa, bienvenida al podcast de APREDIG y gracias por darnos esta entrevista. ¿Qué es el SEDICI de la Universidad Nacional de La Plata?*

—El SEDICI es el repositorio central de la Universidad Nacional de La Plata, que alberga más de 73 mil recursos, de tipologías muy diferentes, sumado a lo tradicional, las tesis, las tesinas, los libros, capítulos de libros, artículos; contamos con contenidos de la Red de Museos de la Universidad Nacional de La Plata (UNLP), entonces tenemos objetos, por ejemplo del Museo Azzarini, dedicado a instrumentos musicales de la universidad o piezas muy antiguas, incluso producciones que pertenecen a instituciones que han acordado convenios con la UNLP para albergar su producción en el repositorio. En los últimos tiempos, lo que hemos tenido que ir agregando es la parte de datos. Y también la universidad está involucrada con lo que tiene que ver con transparencia activa, que sería la contraparte pública de lo que es el acceso abierto frente a nuestras áreas más académicas y científicas. Y como está involucrada en transparencia activa, todo lo que tiene que ver con concursos, convenios, todo lo que tiene que ver con documentación, resoluciones, ordenanzas que tiene la universidad, digestos, etcétera, etcétera se incorporan al repositorio. Con lo cual el repositorio deviene en una herramienta de la administración de la universidad. Una herramienta importante para la visibilidad de la universidad. Pone una buena parte de la visibilidad que tiene la institución y significa un trabajo ingente por la diversidad del tipo de producciones que tenemos que alojar, todo un desafío.

—*¿Qué tipo de servicios debe ofrecer un repositorio de difusión digital institucional en el 2020?*

—Básicamente te diría que, primero que nada, los repositorios se tienen que adaptar a tener, a proveer acceso a una gran variedad de recursos y cuando uno dice acceso, también dice la posibilidad de visualización correcta, de utilización, de reuso, de una variedad de recursos de tipologías diferentes: artículos, tesis, preprints, conjuntos de datos, papers, planos georeferenciados, software, etcétera. Todos los repositorios de la nueva generación deben estar centrados en los recursos y no en los metadatos de los recursos, y eso significa una gran variante. Significa una gran variante en muchas cosas. También en cuestiones como la interoperabilidad y muchos otros aspectos. Entonces, la Confederación Internacional de Repositorios de Acceso Abierto (COAR) está pidiendo que los repositorios se establezcan como redes entre ellos. Donde existan conexiones que sean como especie de enlaces bidireccionales, resultado de la interacción de esos recursos. Como si vos tuvieras objetos, objetos enlazados que deben referirse entre unos y otros de una manera transparente,

identificarse correctamente. Esos recursos también deben ser aptos para el procesamiento por computadoras, para permitir justamente la utilización y reutilización a partir de servicios que establezcan esos repositorios entre sí, teniendo que hacer poco desarrollo extra sobre eso. COAR está planteando y la comunidad está aceptando, que hablemos de plataformas más activas que puedan aceptar otras cosas, versionados, comentarios, actualizaciones, y revisiones por pares que es un planteamiento que ya llevaba adelante COAR hace un tiempo. Y esto también plantea, en los aspectos de interoperabilidad, como ya lo está viendo, me imagino, un cambio del tradicional protocolo OAI-PMH que es un protocolo que lleva quince años y que fue fantástico cuando se inició, a un cambio del protocolo para un protocolo tipo ResourceSync, que puedan utilizar los repositorios y que permita mantener una sincronización entre esos repositorios y por tanto la posibilidad de este enlace entre estos objetos, y demás. Y otras tecnologías, que yo no te digo que estas tecnologías ya no se estén usando en algunos casos, por ejemplo signposting, que te permite mostrar no solamente el recurso sino las relaciones que tiene ese recurso o por ejemplo pensar en protocolos que trabajen en aplicaciones que sirvan para realizar anotaciones arriba de los objetos digitales, comentarlos etcétera, etcétera. Entonces fijate que es un panorama complejo en algunos aspectos, porque cuando uno empieza a buscar no es que haya muchos repositorios que estén implementando en este momento un protocolo de interoperabilidad como ResourceSync, pero hacia eso tenemos que ir.

—¿Cómo se gestiona la preservación digital en un repositorio de difusión digital institucional como SEDICI?

—Bueno, nosotros cuando, digamos, pensamos en preservación ha sido un aprendizaje, voy a decir la experiencia que tenemos, lo cual no es, no va más allá de eso, la experiencia que tenemos. Pero, digamos, si nosotros pensamos en preservación, lo primero que pensamos es tener en cuenta la ISO 14721. Es decir, para nosotros es una guía. No importa cuándo o hasta cuánto te vayas animando. Digamos, tenemos que pensar la estructura de la unidad de información que movemos dentro del repositorio, tal cual como la define la 14721. Es decir, un paquete de información que tiene información de contenido, de representación de ese contenido, e información estrictamente de preservación y también descriptiva. Es decir, el paquete de información en su estructura, que sería tal cual lo entiende la norma. Información propia de contenido e información específica de preservación, lo que la norma llama la PDI. Es decir, de un objeto que está guardado en tu repositorio, de un contenido, tenés que tener la procedencia, que no sólo describe la fuente de ese contenido, sino los procesos que se realizan sobre el objeto digital. La historia de ese objeto digital, los cambios, las versiones. quien hizo ese cambio, etcétera, etcétera. El contexto, es decir, las relaciones que existen entre ese objeto digital y otra fuente de información o contenidos que estén en el repositorio o externos, no importa. También tenés que saber tener una identificación única de ese contenido, es decir, un repositorio hoy no se concibe sin tener un identificador persistente, puede ser un handle, puede ser un DOI, etcétera. Y tenés que tener también seguridad de que esa información no va a cambiarse, no va a alterarse. Es decir, eso lo tenés que tener siempre en mente. Y toda la información relativa a derechos. Que no sólo engloban lo que tiene que ver con el derecho de la difusión del propio repositorio, sino qué tipos de licenciamiento hacés,

típicamente las licencias Creative Commons, que refieren a los usos que el autor permite sobre esa obra. Eso, digamos, es la configuración que uno tiene en la cabeza del contenido como lo va a mover. Y esa información, ese paquete y ese IP, como lo llama la norma, el paquete de información tiene que estar adecuadamente descrito por una información que lo identifique y que nos permita, por supuesto, localizarlo rápidamente y todo lo demás. Entonces uno piensa dos cosas: esto del paquete información, pensar la información que tiene dentro del repositorio, que debe tener los elementos que se solicitan para un paquete de información según la norma y después que dentro del repositorio, tu repositorio tiene que tener, en lo posible, una funcionalidad lo más cercana posible a la funcionalidad que se describe y justamente en la norma en lo que uno da en llamar el modelo abstracto. Que sería la norma, describe seis entidades, como vos sabés muy bien, y las funciones que tiene que tener cada una de esas entidades. El ingreso o ingesta o ingest, el archivo en la administración, el acceso, la gestión de datos y el planeamiento en la preservación. Y ahí viene la cosa y ahí viene el problema, David, porque, digamos, uno tiene bastante bien todo lo que tiene que ver con la ingesta, es decir pensando los software. Yo digo, me paro y pienso en DSpace, supónete que es en lo que tenemos implementado en SEDICI y CIC digital, que son los dos repositorios que gestionamos. Está bárbaro, o sea, no tiene problemas en los términos de la ingesta, lo mismo en lo relativo a la funcionalidad de la administración, vos podés personalizar muchas cosas, se trabaja muy bien. Archivo también, más allá de que vas a tener que pensar todas otras cuestiones vinculadas a la preservación que son las copias, estas cosas que uno puede ver separadamente o prever el archivo por cualquier desastre que pueda ocurrir, fuego, inundación, etcétera, pero eso también lo tenés bien. También el acceso, porque, hoy en día, vos no podés pensar un repositorio sin una interfaz que le confiera a los usuarios, cualquiera de estos sean, y a la administración misma, un acceso amigable. Digamos, no podés dejar de pensar en eso. Y también la gestión de los datos es algo que se logra, porque si vos describís la información que tenés en el repositorio podés permitir una buena gestión de esos datos y a la vez esto va a permitir una buena preservación, si se agregan otro tipo de metadatos como puede ser los PREMIS. Pero, digamos, en la parte descriptiva, si cataloga bien el recurso, en la interfaz, podés brindar lo que se llama “faceting”, es decir podés permitir filtrado para que la persona que entra y busca algo pueda ir discriminando fácilmente la pertinencia de lo que está o de lo que quiere localizar. Pero el gran problema, justamente, tratando de seguir esta norma, es cuando uno se encuentra con lo que sería la entidad que se llama “preservation planning” o planeamiento de la preservación y la funcionalidad que requiere. Porque esa entidad en particular es la que requiere que vos tengas un planeamiento correcto en la preservación. Que tengas bien, digamos, cualquier anuncio, alerta respecto del formato que se te están por quedar obsoletos, que vos tengas un plan de contingencias, migraciones o emulaciones o lo que se necesite. Y que todo esté correctamente pensado en relación a tener un plan de preservación que pudiera eventualmente “ingestarse” en el repositorio y pudiera ir conduciendo las acciones de ese módulo, pensando esa entidad, pensando en términos de 14721, esto se hace bien difícil de llevar adelante. Es decir, todos los aspectos del paquete de información los tenés en un repositorio normal bien establecido, me entendés, todo eso lo tenés. Lo que se te hace más difícil es poder cumplir con esta parte extra, de tener bien pensado el planeamiento de la preservación, las acciones de preservación y demás. Más allá de que puntualmente en un

repositorio hay algunas tareas que se hacen automáticamente con “cron jobs”. Suponete, vos chequeás que nadie haya cambiado el archivo, la integridad, digamos. Tenés un identificador persistente, bueno, hay determinadas cosas que podés mejorar con el uso del diccionario de datos, por ejemplo PREMIS, etcétera. Pero planear la preservación y tener alertas y tener un montón de cosas que significarían, digamos, una preservación más evolutiva, por así decirlo, o cambiante en el tiempo o creciente o mejorada, eso es lo que se hace más difícil.

*Marisa De Giusti, directora del repositorio institucional central de la Universidad de La Plata, SEDICI, y presidenta del Consorcio Iberoamericano para la Educación en Ciencia y Tecnología ISTECS, muchas gracias por participar en esta entrevista para APREDIG, la Asociación Iberoamericana de Preservación Digital. Yo soy David Leija, nos escuchamos en una próxima entrega de este podcast sobre preservación digital en Iberoamérica.*