

Mesa de preservación en repositorios

10 de agosto de 2022

Seguimiento y actividades de preservación digital en un RI

Presentador: De Giusti Marisa R.



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)



EDUCACIÓN
PÚBLICA
Y GRATUITA



UNIVERSIDAD
NACIONAL
DE LA PLATA



Repositorios institucionales

- Los repositorios institucionales son depositarios de la producción de una institución, cualquiera sea su tipología de acuerdo a lo que la propia institución determine.
- Los repositorios tienen como misión reunir, catalogar, preservar, gestionar, difundir y hacer público ese patrimonio como fuente de conocimiento para toda la sociedad, asegurando el acceso, en principio a perpetuidad.
- Los repositorios deben tener una política de preservación elaborada desde la institución y un plan de preservación que permita cumplir con dicha política a pesar de los desafíos y cambios que puedan surgir.

Preservación digital

La preservación digital es el conjunto de estrategias, procesos y técnicas que dan respuesta a los problemas que plantea la conservación de los materiales digitales y de los medios (hardware y software) que se emplean para su almacenamiento y consulta, y que están derivados fundamentalmente de la obsolescencia provocada por la rápida renovación tecnológica y por la inestabilidad de los soportes. Estas técnicas son muy variadas y responden a diferentes situaciones y líneas estratégicas (copias de seguridad, copia de datos en soportes durables, migración, replicación, emulación, etc.), aunque en general están destinadas a mantener los objetos digitales y sus características de acceso a largo plazo. [Directrices UNESCO](#)

Preservación digital en un RI

- Significa ni más ni menos que asegurar el acceso y la legibilidad de los contenidos gestionados en el repositorio para toda la comunidad a la que sirve ese repositorio durante un plazo previsto o de forma indefinida, según corresponda.
- Implica acciones coordinadas durante todo el ciclo de vida de los objetos digitales, sus medios y la infraestructura que lo contiene.
- Supone unos criterios de selección de plazos previstos de mantenimiento de las obras.
- Supone acciones continuas que pueden ser incrementales.
- Lo guía un plan.
- No solo implica acciones de naturaleza tecnológica.

Lo que se espera de un repositorio: áreas y funciones

Modelo OAIS. ISO 14721: ajustar las funciones del repositorio



El centro de todo: el IP del modelo OAIS

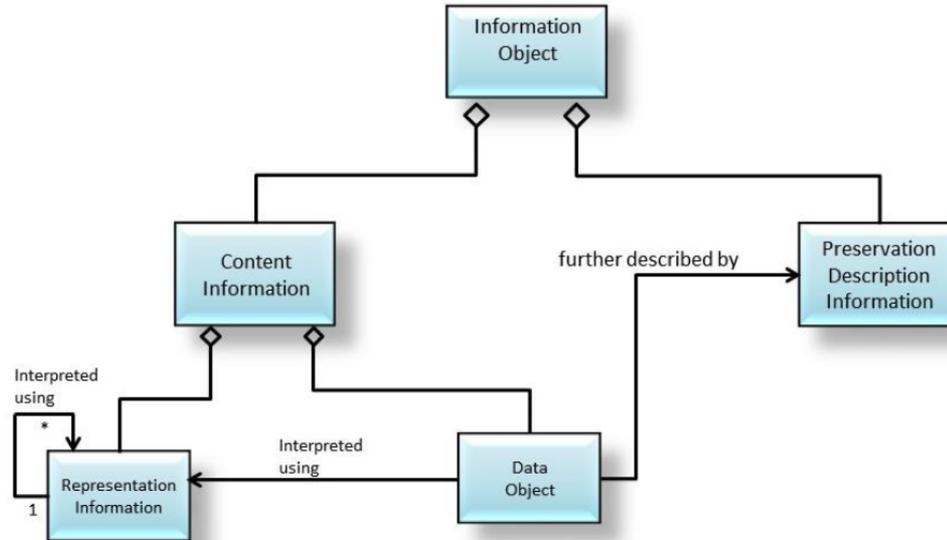


Figure 4-14 Example of an Information Object made up of Content Information and PDI [Management Council of the Consultative Committee for Space Data Systems \(CCSDS\). \(2019\). OAIS final v3 draft with changes wrt OAISv2 20190924-rl.docx](#). P.4-41

El centro de todo: el IP del modelo OAIS

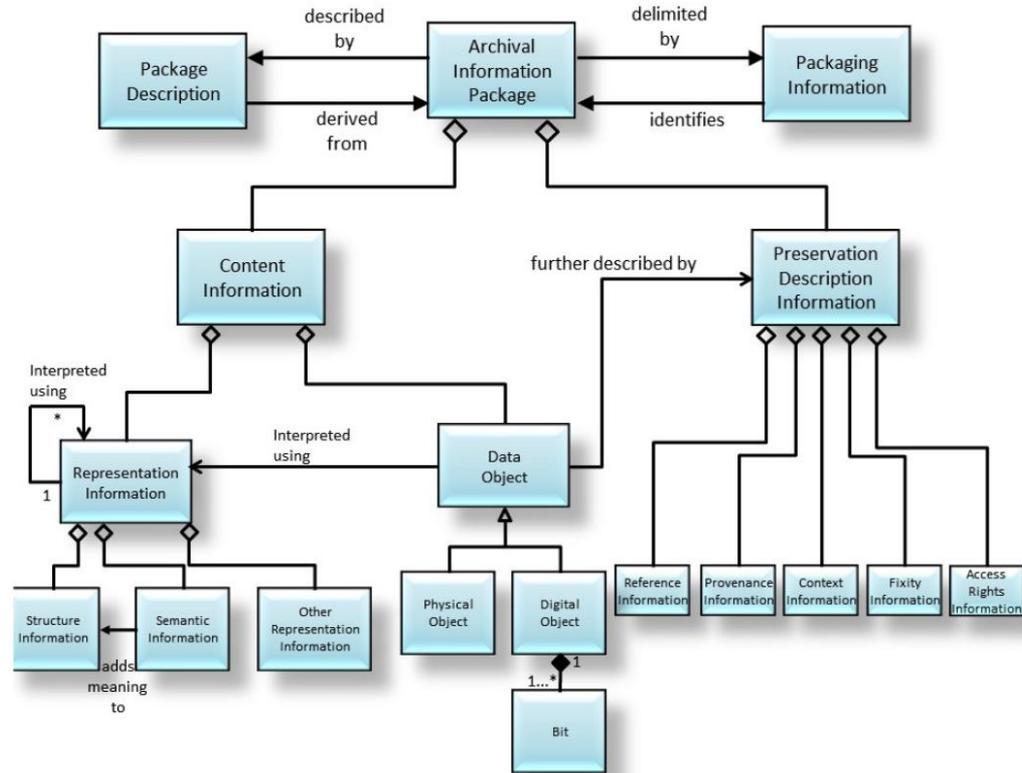


Figure 4-19: Archival Information Package (Detailed View) and its associated Package Description and Packaging Information [Management Council of the Consultative Committee for Space Data Systems \(CCSDS\). \(2019\). OAIS final v3 draft with changes wrt OAISv2 20190924-rl.docx](#). P.4-43

¿Cuándo vigilar y controlar?: durante todo el ciclo de vida

Fases:



OD y metadatos de preservación

Si hay un cambio debe saberse **quién** lo efectuó

Debe mantenerse en el repositorio de manera **segura**

Deben guardarse las relaciones que vinculen al objeto con otros

El repositorio debe tener los derechos suficientes para sostener el **acceso** al objeto



Debe conocerse su **creador**

Debe poder ser **localizado** y **entregado** al usuario

Su soporte deber ser **compatible** con los sistemas actuales

Las estrategias de **emulación** y **migración** requieren datos sobre los objetos originales y sus entornos

Autenticidad
Mediante la documentación de su procedencia

Política de preservación digital: abordaje sugerido para su realización

1. La política de preservación digital de un repositorio debe tener como principal referencia la política institucional de acceso abierto. La cual puede estar en consonancia con legislación a nivel nacional (Argentina) sobre publicaciones y datos.
2. Debe establecer la preservación a largo plazo de los contenidos del repositorio.
3. En lo posible debe hacer referencia a una guía de evaluación de los procesos de preservación en el repositorio (p.e. NDSA) incluso aludir a otros marcos de referencia, por ejemplo recomendaciones de COAR.
4. Debe contener un apartado de “Alcances y propósitos”.

Política de preservación digital: abordaje sugerido para su realización

5. Un capítulo de objetivos.
6. Un capítulo de marco legal.
7. Un capítulo de riesgos.
8. Un capítulo sobre sostenibilidad financiera y técnica.
9. Un apartado sobre seguimiento y revisión de la política.



El documento final debería contar con la aprobación de la institución.

Tener un plan de preservación

- El plan debe exponer los motivos, principios y sobre qué contenidos va a centrarse para garantizar la conservación, el acceso y la comprensión a largo plazo de esos fondos.
- Debe identificar necesidades y prioridades y aportar un cronograma pormenorizado que muestre la distribución de las tareas en el periodo de vigencia del plan.
 - El primer paso es identificar los contenidos en digital (qué prima en el repositorio).
 - En qué formatos y versiones viene después.
 - Diseñar los procedimientos de prevención de riesgos y desastres.
 - Diseñar los procedimientos que indiquen las acciones sobre los ODs según tipología y su frecuencia determinando claramente los responsables.
 - Hacer un cronograma y marcar el estado de las acciones.

Plan de preservación: estructura

1. Situación actual del repositorio: análisis de activos.
2. Definición del plan de preservación:
 - a. Ámbito
 - b. Objetivos: p-e uso de estándares internacionales, metadatos...
 - c. Procedimientos
3. Estrategias:
 - a. Modelo de gestión
 - i. Recursos humanos
 - ii. Financiación
4. Diseño: dimensión de la colección, requisitos del sistema de almacenamiento, digitalización, metodología.
5. Ejecución y difusión.

Autoevaluación: NDSA

- **Aspectos/ Áreas:**

1. Sobre el almacenamiento: copias y localización
2. Sobre la no alteración de los archivos y la integridad de los datos.
3. Sobre la seguridad de la información: quién ha hecho qué con los contenidos.
4. Metadatos.
5. Formatos.

- **Niveles de cumplimiento de cada área (complejidad creciente)**



Matriz de auto-evaluación NDSA

Área Funcional	Nivel			
	Nivel 1 - (Conocer su contenido)	Nivel 2 - (Proteger su contenido)	Nivel 3 - (Controlar su contenido)	Nivel 4 - (Mantener su contenido)
Almacenamiento	<p>Tener dos copias completas en ubicaciones separadas</p> <p>Documentar todos los medios de almacenamiento donde este almacenado el contenido</p> <p>Poner el contenido en soportes de almacenamiento estables</p>	<p>Tener tres copias completas con al menos una copia en una ubicación geográfica distinta</p> <p>Documentar el almacenamiento y medios de almacenamiento, indicando los recursos y las dependencias que estos requieren para funcionar</p>	<p>Tener al menos una copia en una ubicación geográfica con amenaza de desastre diferente a las otras copias</p> <p>Tener al menos una copia en un medio de almacenamiento de diferente tipo</p> <p>Rastrear la obsolescencia del almacenamiento y los medios</p>	<p>Tener al menos tres copias en ubicaciones geográficas distintas, cada una con una amenaza de desastre diferente</p> <p>Maximizar la diversificación del almacenamiento para evitar puntos únicos de falla</p> <p>Tener un plan y realizar acciones para abordar la obsolescencia del hardware, software y medios de almacenamiento</p>
Integridad	<p>Verificar que la información de integridad se ha proporcionado con el contenido</p> <p>Generar información de integridad si esta no ha sido proporcionada con el contenido</p> <p>Se verifica virus en todo el contenido; se aísla el contenido en cuarentena según sea necesario</p>	<p>Verificar la información de integridad al mover o copiar contenido</p> <p>Usar bloqueadores de escritura cuando se trabaja con medios originales</p> <p>Hacer una copia de seguridad de la información de integridad y almacenar una copia en una ubicación separada del contenido</p>	<p>Verificar la información de integridad del contenido en intervalos fijos</p> <p>Documentar los procesos y resultados de verificación de información de integridad</p> <p>Realizar una auditoría de la información de integridad bajo demanda</p>	<p>Verificar la información de integridad en respuesta a eventos o actividades específicas</p> <p>Reemplazar o reparar el contenido dañado según sea necesario</p>
Control	<p>Se determinan los agentes humanos y de software que deben estar autorizados para leer, escribir, mover y eliminar contenido</p>	<p>Documentar a los agentes humanos y de software autorizados para leer, escribir, mover y eliminar contenido y aplicar estos</p>	<p>Mantener los registros (logs) y se identifica a los agentes humanos y de software que realizaron acciones sobre el contenido.</p>	<p>Se realizan revisiones periódicas de acciones / registros (logs) de acceso</p>
Metadatos	<p>Crear un inventario de contenido, documentando también la ubicación de almacenamiento actual de estos</p> <p>Hacer una copia de respaldo del inventario y se almacena al menos una copia por separado</p>	<p>Almacenar suficientes metadatos para saber cuál es el contenido (esto podría incluir alguna combinación de aspectos administrativos, técnicos, descriptivos, de preservación y estructurales)</p>	<p>Determinar qué estándares de metadatos aplicar</p> <p>Encuentra y completa los vacíos en sus metadatos para cumplir con esos estándares</p>	<p>Registrar las acciones de preservación asociadas con el contenido y cuándo ocurren esas acciones Implementa los estándares de metadatos elegidos</p>
Contenido	<p>Documentar los formatos de archivo y otras características de contenido esenciales, incluido cómo y cuándo fueron identificados</p>	<p>Verificar los formatos de archivo y otras características de contenido esenciales</p> <p>Establecer relaciones con los creadores de contenido para fomentar la elección sostenible de archivos</p>	<p>Monitorear la obsolescencia y los cambios en las tecnologías de las que depende el contenido</p>	<p>Realizar migraciones, normalizaciones, emulación y actividades similares que garanticen el acceso al contenido</p>

Ejemplo

Sobre el almacenamiento: copias y localización

Nivel 1 (más básico): tener dos copias completas de la información por separado y en servidores propios (no en discos de manera aislada).

Nivel 2 : tres copias y una en una localización geográfica diferente.

Nivel 3: tres copias y una en una localización geográfica diferente con previsión de desastres distinta. Controlar el proceso de obsolescencia del almacenamiento y el soporte.

Nivel 4: como mínimo tres copias en tres localizaciones geográficas diferentes, cada una con previsión de desastres distinta. (Alguna puede ser en la nube). Disponer de un plan integral para mantener datos y metadatos accesibles en los sistemas y soportes.



En paralelo con la acción documentar el sistema elegido (en el nivel que sea).
Escribir un plan.

Matriz de evaluación NDSA para CIC Digital

	Nivel 1 (Proteja sus datos)	Nivel 2 (Conozca sus datos)	Nivel 3 (Controle sus datos)	Nivel 4 (Repare sus datos)	Puntaje
Almacenamiento	Tener dos copias completas en ubicaciones separadas	Tener tres copias completas con al menos una copia en una ubicación geográfica distinta	Tener al menos una copia en una ubicación geográfica con amenaza de desastre diferente a las otras copias	Tener al menos tres copias en ubicaciones geográficas distintas, cada una con una amenaza de desastre diferente.	0/4
	Documentar todos los medios de almacenamiento donde esté almacenado el contenido	Documentar el almacenamiento y medios de almacenamiento, indicando los recursos y las dependencias que estos requieren para funcionar	Rastrear la obsolescencia del almacenamiento y los medios.	Maximizar la diversificación del almacenamiento para evitar puntos únicos de falla	
	Poner el contenido en soportes de almacenamiento estables		Tener al menos una copia en un medio de almacenamiento de diferente tipo	Tener un plan y realizar acciones para abordar la obsolescencia del hardware, software y medios de almacenamiento	
No alteración de archivos e integridad de los datos	Verificar que la información de integridad se ha proporcionado con el contenido	Verificar la información de integridad al mover o copiar contenido	Verificar la información de integridad del contenido en intervalos fijos	Comprobar la integridad de todo el contenido en respuesta a situaciones o actividades específicas.	1/4
	Generar información de integridad si esta no ha sido proporcionada con el contenido	Usar bloqueadores de escritura cuando se trabaja con medios originales	Documentar los procesos y resultados de verificación de información de integridad	Verificar la información de integridad en respuesta a eventos o actividades específicas	
	Se verifica virus en todo el contenido; se aísla el contenido en cuarentena según sea necesario	Hacer una copia de seguridad de la información de integridad y almacenar una copia en una ubicación separada del contenido	Realizar una auditoría de la información de integridad bajo demanda	Reemplazar o reparar el contenido dañado según sea necesario	



Cumple satisfactoriamente



Cumple parcialmente



No cumple

Matriz de evaluación NDSA para CIC Digital

Seguridad de la información	Se determinan los agentes humanos y de software que deben estar autorizados para leer, escribir, mover y eliminar contenido	Documentar a los agentes humanos y de software autorizados para leer, escribir, mover y eliminar contenido y aplicar estos cambios	Mantener los registros (logs) y se identifican a los agentes humanos y de software que realizaron acciones sobre el contenido.	Se realizan revisiones periódicas de acciones / registros (logs) de acceso	1/4
Metadatos	Crear un inventario de contenido, documentando también la ubicación de almacenamiento actual de estos Hacer una copia de respaldo del inventario y se almacena al menos una copia por separado	Almacenar suficientes metadatos para saber cuál es el contenido (esto podría incluir alguna combinación de aspectos administrativos, técnicos, descriptivos, de preservación y estructurales)	Determinar qué estándares de metadatos aplicar Encuentra y completa los vacíos en sus metadatos para cumplir con esos estándares	Registrar las acciones de preservación asociadas con el contenido y cuándo ocurren esas acciones Implementa los estándares de metadatos elegidos	3/4
Formatos de archivos	Documentar los formatos de archivo y otras características de contenido esenciales, incluido cómo y cuándo fueron identificados	Verificar los formatos de archivo y otras características de contenido esenciales Establecer relaciones con los creadores de contenido para fomentar la elección sostenible de archivos	Monitorear la obsolescencia y los cambios en las tecnologías de las que depende el contenido	Realizar migraciones, normalizaciones, emulación y actividades similares que garanticen el acceso al contenido.	4/4
Puntaje global	3/5	2/5	3/5	1/5	9/20



Cumple satisfactoriamente



Cumple parcialmente



No cumple

Acciones iniciales:

Revisión de prácticas y procedimientos

- Copias: de qué y en dónde.
- Documentos escritos: cuáles.
- Formatos utilizados.
- Autorizaciones, tipos de usuarios y permisos
- Control de agentes y cambios sobre los bitstreams: metadatos.
- Metadatos usados: administrativos + técnicos, de preservación (PDI).

Acciones iniciales:

Revisión de contenidos y formatos

- Realizar un perfilamiento básico de los contenidos del repositorio, entendido en el estricto sentido de saber las grandes tipologías o algo como “súperclases!": texto, video, audio y datos.
 - En cada caso saber “qué aspectos de la información se deben preservar”.
- En función de las cantidades de cada qué, establecer un orden de prioridad que puede ser un plan de acción que comience por lo que hay menos (que se resuelve rápido) o por lo que hay más. Observar que esto es una definición política-administrativa-de recursos.
- Después hacer un perfilamiento detallado para cada formato (DROID por ejemplo).

Acciones de preservación - Resguardo de datos

- **Escribir un plan de riesgos sobre los datos:** ciberseguridad, falla hardware, errores humanos, desastres naturales, conflictos bélicos/atentados, etcétera.
 - ISO 16363. Drambora (pero ya sin mantenimiento).

Definir el requerimiento de copias de seguridad en lo que respecta a cantidad, rotación y ubicación de las mismas.

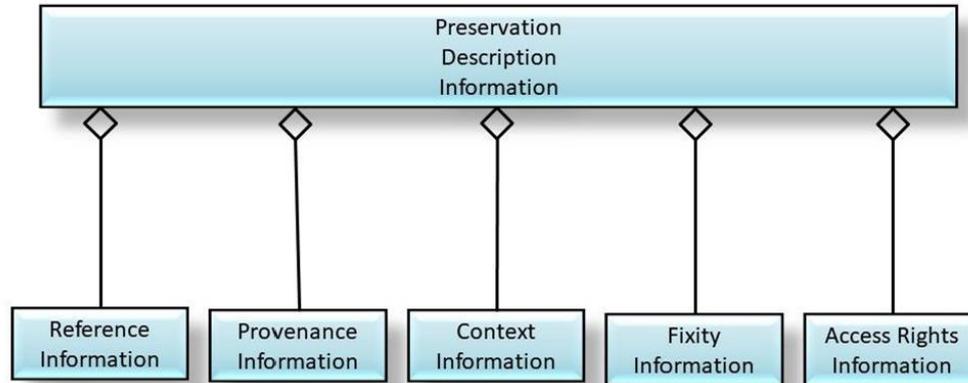
- Se debe definir qué resguardar, cómo, dónde y por cuánto tiempo.
- Copias en localizaciones de fallas distintas
- Copias en la nube en cumplimiento con legislación nacional
- Requerimiento de encriptación.

Acciones de preservación - Documentación

- **Documentación de procesos de backups**
 - ¿Cómo se realizan los backups? ¿Qué herramientas se utilizan? ¿Cuál es el procedimiento?
 - Definir qué hacer ante cada potencial situación de amenaza sobre los datos
- **Manuales de carga**
 - Formato correcto de cada metadato
 - Consideraciones a la hora de seleccionar autores, materias, etc.
- **Documentación del proceso de digitalización de documentos**
- **Documentación de las tareas de mantenimiento y desarrollo**
 - Se suele usar un sistema de tickets (trac)
- **Inventario del equipamiento del repositorio**
 - Plan para el control de obsolescencia de los equipos y su eventual reemplazo

Acciones de preservación - Integridad

- Análisis de checksum sobre el contenido original
- Checksum sobre los backups
- Automatización de estos procesos - tareas de curation



Acciones de preservación - Control de cambios

- Dejar registrado en metadatos quien modifico un archivo o los metadatos de un ítem (provenance).
- Definir distintos permisos sobre los metadatos y objetos digitales para los distintos tipos de usuarios (resuelto en DSpace con permisos y grupos).
- Tener documentados los permisos de los usuarios sobre los archivos
- Tener versiones de los objetos digitales ante una modificación

dc.description.provenance	Submitted by Nancy Martini (nancymartini@quimica.unlp.edu.ar) on 2020-11-19T13:42:16Z workflow start=Step: SeDiCiLevelReview - action:claimaction No. of bitstreams: 1 Tesis doctoral Nancy Martini 2020 .pdf: 13328499 bytes, checksum: fda83eca57a2cb81d91189118344beb (MD5)	en
dc.description.provenance	Step: SeDiCiLevelReview - action:editaction Approved for entry into archive by Analía Pinto(aprumiante@gmail.com) on 2020-11-19T14:17:11Z (GMT)	en
dc.description.provenance	Made available in DSpace on 2020-11-19T14:18:06Z (GMT). No. of bitstreams: 1 Tesis doctoral Nancy Martini 2020 .pdf-PDFA.pdf: 13856207 bytes, checksum: 877fd2b4dedae1da315caa854a27c7e2 (MD5) Previous issue date: 2020	en
dc.description.provenance	Submitted by Nancy Martini (nancymartini@quimica.unlp.edu.ar) on 2022-07-04T11:24:22Z workflow start=Step: SeDiCiLevelReview - action:claimaction No. of bitstreams: 3 Tesis doctoral Nancy Martini 2020 .pdf-PDFA.pdf: 13856207 bytes, checksum: 877fd2b4dedae1da315caa854a27c7e2 (MD5) Tesis doctoral Nancy Martini 2020 .pdf-PDFA.pdf.txt: 560515 bytes, checksum: 3cf2f3164bbcd4ef4723b9da7a5df333 (MD5) Tesis doctoral Nancy Martini 2020 .pdf-PDFA.pdf.jpg: 3801 bytes, checksum: 622108bfd8a2c072d9e9e0c095e1652c (MD5)	en
dc.description.provenance	Step: SeDiCiLevelReview - action:editaction Approved for entry into archive by Analía Pinto(aprumiante@gmail.com) on 2022-07-04T13:12:49Z (GMT)	en

Acciones de preservación - Metadatos

- Separar el almacenamiento de los metadatos del objeto digital
- Almacenar metadatos de administración y de preservación (ej PREMIS)
- Uso de identificadores persistentes
 - Para el ítem y para el OD
- Formatos estándares
- Uso de vocabularios controlados
- Uso de metadatos técnicos (como los que arroja exiftool)

Acciones de preservación - Control de contenido

- Perfilamiento de archivos para revisión de formatos por problemas de obsolescencia. (Raramente se hace).
- Actualización de versiones entre formatos.
- Transformación/migración de un formato a otro.
- Análisis de virus.
- Definir un listado de los formatos aceptados.

Conclusiones

- **Alcances de NDSA y otras posibilidades**
 - NDSA es interesante porque permite la autoevaluación y porque es relativamente simple. De lo mencionado precedentemente en cuanto acciones necesarias para asegurar la PD, todas las descritas por este estándar son de naturaleza técnica.
 - Hay acciones que exceden a los objetos digitales y la infraestructura, por ejemplo acciones que hacen a lo organizacional, que no cubre NDSA, básicamente ahí debe escalarse a una norma como la ISO 16363 que además permite certificar el repositorio pero es bastante compleja.
 - Se debería además, trasladar la sintaxis del diccionario de datos PREMIS tanto a agentes como a eventos para asegurar una buena trazabilidad..

Conclusiones

- Definir documentación prioritaria
 - procedimientos de backups
 - plan de recuperación de desastres
 - políticas de contenidos y preservación
- Controlar ejecución de backups
 - Validar que sean completos
- Mejorar trazabilidad
 - En DSpace se hace pero de manera parcial
- Mejorar metadatos
 - Costo de adaptar PREMIS para tener metadatos de preservación adecuados y de la implementación de versionado para los OD
- Importancia de la política de formatos y de preservación
 - Los formatos actualizados impiden obsolescencia
 - Es necesario que además de tener la capacidad de almacenar los distintos formatos haya medios para reproducir o visualizar los mismos

Otros trabajos del equipo

Esta presentación es un resumen de los aspectos tratados en:

De Giusti, M. R., Lira, A. J., & Tettamanti, S. (2022). Taller sobre prácticas aplicadas a la preservación digital en un repositorio institucional. In Evento conmemorativo de los 10 años de la Red Brasileña de Servicios de Preservación Digital-CARINIANA del Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)(Brasilia, Brasil, 11 y 14 de julio de 2022).

Numerosos trabajos vinculados a preservación digital se encuentran en la colección de PREBI-SEDICI en el repositorio institucional:
<http://sedici.unlp.edu.ar/handle/10915/25293>

Ver en especial:

Bodero Poveda, E. M., De Giusti, M. R., & Morales, C. (2022). Preservación digital a largo plazo: estándares, auditoría, madurez y planificación estratégica. Revista Interamericana de Bibliotecología, 45.

De Giusti, M. R. (2020). Preservación digital: normas, prácticas y acciones recomendadas desde un repositorio institucional. II Encuentro de Preservación digital (México, 2020).

De Giusti, M. R. (2019, May). Workshop 3: Avaliação da gestão da preservação digital nas instituições. In IV SINPRED-VI Encontro de Rede Cariniana. Red Brasileña de Servicios de Preservación Digital (Brasilia, 2019).

Referencias

National Digital Information Infrastructure and Preservation Program (NDIIPP) of the Library of Congress. (n.d.-b). Levels of Digital Preservation. National Digital Stewardship Alliance - Digital Library Federation. Retrieved April 25, 2022, from <https://ndsa.org/publications/levels-of-digital-preservation/>

Levels of Preservation Revision Working Group, Kussmann, C., National Digital Stewardship Alliance (NDSA), Graham, W., Atkins, W., Reich, A., & Walker, P. (2019, October). 2019 LOP Implementation Guide and Working Definitions. <https://osf.io/nt8u9/>

NDSA Levels of Preservation Assessment Subgroup. (2019). Using the Levels of Digital Preservation as an Assesment Tool. <https://doi.org/10.17605/OSF.IO/QGZ98>

Leija, David; Térmens, Miquel. (2019). Traducción de Niveles de Preservación Digital NDSA 2019: Traducción al Español de Versión 2.0. APREDIG - Asociación Iberoamericana de Preservación Digital.

Curation System—DSpace 6.x Documentation—LYRISIS Wiki. (s. f.). Recuperado 9 de agosto de 2022, de <https://wiki.lyrasis.org/display/DSDOC6x/Curation+System#CurationSystem-BitstreamFormatProfiler>

I ENCONTRO DA REDE BRASILEIRA DE REPOSITÓRIOS DIGITAIS

TEMÁTICA: REPOSITÓRIOS DIGITAIS E CIÊNCIA ABERTA

09 a 11/08/2022 | Horário: 9h às 13h

¡Muchas gracias!

Marisa R. De Giusti

marisa.degiusti@sedici.unlp.edu.ar

Presentación disponible en la colección: <http://sedici.unlp.edu.ar/handle/10915/25293>



Esta obra está bajo una [Licencia Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/)
Atribución-NoComercial-CompartirIgual 4.0 Internacional



EDUCACIÓN
PÚBLICA
Y GRATUITA



UNIVERSIDAD
NACIONAL
DE LA PLATA



CIC COMISIÓN DE
INVESTIGACIONES CIENTÍFICAS