

ESTUDIOS E INFORMES

LÍNEA 2

2022



PRESERVACIÓN DIGITAL DE LAS TESIS
EN LOS REPOSITORIOS DE REBIUN



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

PRESERVACIÓN DIGITAL DE LAS TESIS EN LOS REPOSITARIOS DE REBIUN

Informe elaborado por:

Maria Luisa Segundo Martín (URJC) – Coordinadora L2-OG2

Juana Frías Fernández (UAH)

Carolina Corral (UNED)

Maria Dolores Moreno (UM)

Ana Amigot Zardoya (UPNA)

Eva María Navarro Gimena (UPO)



Documento bajo licencia [Creative Commons BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)

REBIUN / Plan Estratégico 2020-2023

Línea 2 Transformación digital y conocimiento abierto

Objetivo general 2. Contribuir a la transformación digital
de las universidades y centros de investigación

Objetivo específico 2. Preservar digitalmente las tesis.

Contenido

INTRODUCCIÓN	2
METODOLOGÍA	3
RESULTADOS	5
Planificación	5
Niveles de preservación digital	6
Almacenamiento	6
Integridad	6
Control	7
Metadatos	7
Contenido	7
Comparativa Informe 2018	8
Preservación de tesis	9
CONCLUSIONES	10
BIBLIOGRAFÍA	11
UNIVERSIDADES PARTICIPANTES EN LA ENCUESTA	12

INTRODUCCIÓN

Los repositorios institucionales en acceso abierto tienen como objetivo la visibilidad y la difusión de la producción científica de sus instituciones y garantizar el acceso y preservación a largo plazo. En esta última cuestión se encuentra enmarcado este informe que expone el estado actual de la preservación digital, es decir, conocer las tareas o actividades de preservación que se realizan y en qué grado en los repositorios REBIUN.

En 2018 el Grupo de Trabajo de Repositorios de REBIUN realizó una encuesta e informe sobre el estado de la preservación de los repositorios digital en el que se concluía que había un largo recorrido por delante, las técnicas de preservación digital se estaban aplicando de forma incipiente y parcial en la mayoría de los repositorios y apenas había planes de preservación aprobados y en funcionamiento¹.

El presente informe ha sido elaborado por el Objetivo 2 sobre Transformación digital de la Línea 2 sobre Transformación Digital y Conocimiento Abierto de Rebiun en el contexto del Plan Estratégico 2020-2023 de REBIUN. El informe recoge los resultados de un cuestionario enviado a las universidades de REBIUN en junio de 2022 que se adaptaba a los nuevos niveles de preservación digital establecidos por la National Digital Stewardship Alliance (NDSA), publicados en 2019, con un doble objetivo: analizar la evolución de la situación de partida obtenida en 2018, para servir como toque de atención a la comunidad universitaria, y centrar la atención en la preservación digital de las tesis doctorales.

La conclusión principal es que los resultados con relación a las medidas técnicas de preservación son algo mejores que en la encuesta del 2018. No obstante, aunque en una gran parte de los repositorios se realizan tareas en materia de preservación, todavía apenas hay políticas aprobadas y planes de preservación, aunque sí en preparación.

Palabras clave

Repositorios institucionales de investigación; preservación digital; políticas institucionales; tesis doctorales; encuestas de evaluación.

¹ Informe sobre la evaluación del estado de la preservación de los repositorios, REBIUN 2018 <<https://hdl.handle.net/20.500.11967/253>>

METODOLOGÍA

La encuesta se realizó en línea entre el 15 de junio y el 29 de julio de 2022, enviándose un recordatorio la segunda semana de julio.

La encuesta, que constaba de 44 preguntas, se envió a los directores de las bibliotecas de las 76 universidades que conforman Rebiun (bibliotecas universitarias, públicas y privadas, miembros de CRUE Universidades Españolas y el CSIC), para que fuera cumplimentada por el personal responsable de los repositorios.

El grupo de trabajo para la elaboración del cuestionario revisó las preguntas de la encuesta del año 2018 y trabajó en su adaptación con los nuevos niveles del NDSA publicados en 2019, la “Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación”² realizado por el Grupo de Repositorios en el año 2020 y el “Informe sobre la evaluación del estado de la preservación de los repositorios Rebiun 2018”³.

² Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación <<http://hdl.handle.net/20.500.11967/634>>

³ Informe sobre la evaluación del estado de la preservación de los repositorios Rebiun 2018. REBIUN. <<http://hdl.handle.net/20.500.11967/253>>

Tabla de niveles de preservación digital NDSA 2019⁴

Área Funcional	Nivel			
	Nivel 1 - (Conocer su contenido)	Nivel 2 - (Proteger su contenido)	Nivel 3 - (Controlar su contenido)	Nivel 4 - (Mantener su contenido)
Almacenamiento	Tener dos copias completas en ubicaciones separadas Documentar todos los medios de almacenamiento donde este almacenado el contenido Poner el contenido en soportes de almacenamiento estables	Tener tres copias completas con al menos una copia en una ubicación geográfica distinta Documentar el almacenamiento y medios de almacenamiento, indicando los recursos y las dependencias que estos requieren para funcionar	Tener al menos una copia en una ubicación geográfica con amenaza de desastre diferente a las otras copias Tener al menos una copia en un medio de almacenamiento de diferente tipo Rastrear la obsolescencia del almacenamiento y los medios	Tener al menos tres copias en ubicaciones geográficas distintas, cada una con una amenaza de desastre diferente Maximizar la diversificación del almacenamiento para evitar puntos únicos de falla Tener un plan y realizar acciones para abordar la obsolescencia del hardware, software y medios de almacenamiento
Integridad	Verificar que la información de integridad se ha proporcionado con el contenido Generar información de integridad si esta no ha sido proporcionada con el contenido Se verifica virus en todo el contenido; se aísla el contenido en cuarentena según sea necesario	Verificar la información de integridad al mover o copiar contenido Usar bloqueadores de escritura cuando se trabaja con medios originales Hacer una copia de seguridad de la información de integridad y almacenar una copia en una ubicación separada del contenido	Verificar la información de integridad del contenido en intervalos fijos Documentar los procesos y resultados de verificación de información de integridad Realizar una auditoría de la información de integridad bajo demanda	Verificar la información de integridad en respuesta a eventos o actividades específicas Reemplazar o reparar el contenido dañado según sea necesario
Control	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	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
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 esos estándares	Registrar las acciones de preservación asociadas con el contenido y cuándo ocurren esas acciones implementa los estándares de metadatos elegidos
Contenido	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 garantizan el acceso al contenido

Levels of Digital Preservation v2.0. - 26 Septiembre 2019.

Copyright © 2019 by National Digital Stewardship Alliance. This work is licensed under a Creative Commons Attribution 3.0 Unported License. | DOI 10.17605/OSF.IO/QGZ98

Traducción: APREDIG

www.apredig.org

⁴ Traducción de Niveles de Preservación Digital NDSA 2019: traducción al español de versión 2.0. <<https://osf.io/egjk8>>
<<http://www.apredig.org/npdnds2019/>>

RESULTADOS

El número de universidades que contestó a la encuesta fue de 48 (63%) frente a las 52 que lo hicieron en el 2018.

Planificación

Las preguntas de este apartado tuvieron en cuenta las acciones de preservación en el repositorio que se plantean en el test de autoevaluación de la Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación⁵ en los apartados de planificación y plan de preservación. En la guía se señala que “la planificación empieza con la redacción y la aprobación de dos documentos: la política de preservación y el plan de preservación. En el primero, la institución u organización declaran su compromiso con la preservación digital de los contenidos del repositorio y hacen explícitas las pautas básicas de su actuación; (...) En el segundo documento se describen los principios tecnológicos y las actuaciones técnicas en los que se basa su estrategia de preservación digital.”

De la encuesta se desprenden los siguientes datos:

Software utilizado en el repositorio	
Dspace	44
Eprints	1
Fedora	1
Otros	2

La institución ha aprobado formalmente una política de preservación	
SI	3
No	33
Está en preparación	12

Existe un plan de preservación en forma de documento aprobado	
SI	1
NO	36
Está en preparación	11

Su institución realiza algunas tareas en materia de preservación	
SI	38
NO	10

En el año 2018 únicamente 2 repositorios contaban con planes de preservación aprobados y en funcionamiento y desde entonces solo otra institución ha aprobado una política y plan de preservación. No obstante, son 12 las universidades que están trabajando en la aprobación de estos documentos.

⁵ Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación <<http://hdl.handle.net/20.500.11967/634>>

Conviene señalar que 10 instituciones han indicado que no realizan tareas en materias de preservación.

Niveles de preservación digital

Según las áreas funcionales, el porcentaje de cumplimiento de los niveles NDSA se refleja en el siguiente cuadro.

Área funcional	Nivel			
	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Almacenamiento	56%	32%	35%	7%
Integridad	58%	20%	25%	73%
Control	96%	62%	98%	87%
Metadatos	41%	85%	77%	12%
Contenido	97%	73%	20%	32%

Almacenamiento

En lo referente al área de almacenamiento, el 98% de las universidades tiene al menos dos copias en ubicaciones separadas, pero solo un 21% tiene tres o más copias en ubicaciones geográficas distintas, cada una con una amenaza de desastre diferente.

El punto más negativo del análisis es el referente al mantenimiento del contenido (nivel 4), solo el 7% de las instituciones cumplen con las exigencias de los niveles NDSA. Por el contrario, el 56% de las instituciones cumplen el nivel 1 referente al conocimiento de los contenidos. Aunque en la mayoría de las universidades se tiene en cuenta la seguridad del almacenamiento a través de las copias, solo un 21% lo tiene documentado, indicando los recursos y las dependencias que estos requieren para funcionar.

Integridad

Casi tres cuartas partes de los repositorios tiene información respecto a la integridad de los ficheros (38 de los encuestados, un 73% del total).

El peor resultado en cuanto a la integridad se da con relación a la protección del contenido (el nivel 2). Sólo 10 de los encuestados respondieron que sí comprueban la integridad (19%). Sólo el 21% de los encuestados utiliza bloqueadores de escritura cuando accede a soportes que llegan de proveedores o usuarios externos.

Por contra, un 73% de los encuestados afirma reemplazar o reparar los datos cuando el contenido ha sido dañado.

Control

En el área funcional de Control se puede observar cómo las instituciones tienen un cumplimiento homogéneo que solo se ve interrumpido en el segundo nivel (Proteger su contenido) a causa de que un menor porcentaje de universidades, el 62% documentan a los agentes humanos y software utilizados para escribir, mover y eliminar contenido.

El resto de los niveles de Control se cumplen satisfactoriamente, ya que un 98% de las instituciones alcanzan el nivel 3 (Controlar su contenido) al mantener los registros logs e identificar a los agentes humanos que realizan acciones sobre el contenido, un 96% alcanzan el nivel 1 (Conocer su contenido) al determinar a las personas autorizadas a leer, escribir, mover y eliminar contenido. En el nivel 4 (Mantener su contenido) un 87% de instituciones realizan revisiones periódicas de acciones y logs de acceso.

Metadatos

En lo referente al área de metadatos encontramos un nivel alto de instituciones que almacenan metadatos técnicos y descriptivos (85%), utilizan y aplican estándares (83%) y disponen de un inventario del contenido de sus repositorios (62%).

El punto más negativo del análisis es el referente a los metadatos de preservación. En la encuesta se preguntó de forma separada por el almacenamiento de metadatos de preservación y solamente el 27% de los repositorios dieron una respuesta positiva.

Contenido

En Contenido observamos que ha habido un gran avance en el mantenimiento del contenido, que ha pasado de un 14% (2018) a un 32% (2022).

El resultado más bajo lo sigue teniendo el nivel 3 (controlar el contenido) ya que sólo 10 de los encuestados tienen debidamente documentados todos los tipos de soportes de almacenamiento

En cuanto a protección y conocimiento del contenido sigue mejorando, especialmente en el nivel 1 que ha pasado de un 88% (2018) a un 97% (2022).

Comparativa Informe 2018

Área funcional	Nivel							
	Nivel 1		Nivel 2		Nivel 3		Nivel 4	
	2018	2022	2018	2022	2018	2022	2018	2022
Almacenamiento	89%	56%	42%	32%	60%	35%	16%	7%
Integridad	30%	57%	59%	20%	40%	25%	45%	73%
Control	86%	95%	0	62%	61%	97%	14%	87%
Metadatos	42%	40%	59%	85%	81%	76%	44%	12%
Contenido	88%	97%	64%	72%	NA	20%	14%	32%
Nivel	65%	69%	49%	54%	53%	50%	30%	42%

En la comparativa de los resultados de las encuestas del 2018 y 2022 se constata que en varios de los niveles hay un descenso del porcentaje de instituciones que cumplen los requisitos de cada una de las áreas funcionales. Entendemos que puede deberse a unas mayores exigencias en la adaptación de la versión 1 a la 2 de la tabla de la NDSA. El cumplimiento menor al avanzar en los niveles es esperable dado que la estructura de la tabla es progresiva y para cumplir los superiores es necesario realizar las actividades de los primeros niveles.

Todos los niveles del Área funcional de control mejoran su cumplimiento en 2022 de forma extraordinaria con respecto a 2018 ya que el nivel 1 y 3 ocupan el primer y tercer mejor resultado de la encuesta global. Las instituciones deben avanzar y abordar más acciones en los niveles 2 y 4.

En varias de las áreas, el nivel 4 de la versión 2 de NDSA plantea una mayor exigencia frente a la versión anterior. Si en la versión anterior se pedía el almacenamiento de metadatos de preservación, ahora se solicita “Registrar las acciones de preservación asociadas con el contenido y cuando ocurren esas acciones implementa los estándares de metadatos elegidos”. Esta exigencia solamente la cumplen el 12% de los repositorios.

Preservación de tesis

Preservación de tesis	Sí (%)	
	Digital	39,58%
Las tesis doctorales de su institución están en formato digital	Parcialmente	60,42%
En su institución existen tesis doctorales que solamente están en formato digital	70,83%	
Se realiza alguna actuación de preservación diferente a la del repositorio con relación a las tesis doctorales en su institución	62,50%	
La institución se ha planteado la implantación de un sistema de preservación digital distribuido de las tesis doctorales	29,17%	
Su institución pertenece a alguna red de preservación digital distribuida (MetaArchive Cooperative, ...)	22,92%	

En el IV Plan estratégico de REBIUN, la Línea 2: Transformación Digital y Conocimiento Abierto se asumió el objetivo de preservación de tesis doctorales⁶ planteando la acción de la elaboración de un documento de recomendaciones para su preservación digital. El grupo de trabajo expuso que el tratamiento de preservación digital en los repositorios de las tesis doctorales era el mismo que el de otro tipo de publicaciones, y que otras posibles acciones de preservación ya no estaban vinculadas a la Biblioteca o podrían estar encaminadas a incorporar las tesis a una red de preservación distribuida. Esta acción ya se estudió en iniciativas anteriores de REBIUN pero se descartó. Por ello, se decidió realizar una encuesta sobre el estado actual de la preservación digital en los repositorios e incluir un apartado de cuestiones sobre las tesis doctorales.

Actualmente hay 11 universidades que están dentro de un sistema de preservación digital distribuida para sus tesis doctorales y 4 universidades se lo han planteado. Todas las universidades que están dentro de una red de preservación digital distribuida son catalanas y depositan sus tesis en el repositorio cooperativo TDX que desde el año 2011 forma parte de la cooperativa MetaArchive.

⁶ IV Plan estratégico de la red de bibliotecas REBIUN 2020-2023 <<https://hdl.handle.net/20.500.11967/856>>

CONCLUSIONES

Pese a que solo 3 instituciones tienen aprobados planes y/o políticas de preservación, otras 12 están trabajando en su elaboración.

En los últimos cuatro años se detecta un descenso, en alguno de los niveles, del porcentaje de instituciones que cumplen los requisitos de cada una de las áreas funcionales. Esto es debido a unas mayores exigencias en la versión 2 de la NDSA.

Es en los metadatos de preservación y en su control donde se encuentra la principal área de mejora para los repositorios españoles y donde se constata que no ha habido avances significativos desde 2018.

La mayoría de las tesis doctorales se preservan en los repositorios institucionales. Las instituciones no han mostrado interés por pertenecer a una red distribuida de tesis.

BIBLIOGRAFÍA

Azorín, C.; Barrueco, J.M.; Martínez, C.; Térmens, M. (2019). Informe sobre la evaluación del estado de la preservación de los repositorios Rebiun 2018. REBIUN. Recuperado de <http://hdl.handle.net/20.500.11967/253>.

Barrueco Cruz, J. M.; Andrés Rodríguez, A.; Rico Castro, P.; [et al.]. (2017). Guía para la evaluación de repositorios institucionales de investigación. <https://recolecta.fecyt.es/sites/default/files/contenido/documentos/2017GuiaEvaluacionRecolectaFECYT.pdf>.

Azorín, C.; Barrueco, J.M.; Bernal, I.; Macías, V.; Marín, R. Martínez, C.; Térmens, M. (2020). Guía para la evaluación de los procesos de preservación en repositorios institucionales de investigación. <http://hdl.handle.net/20.500.11967/634>.

Ferrer, A. (2015). Sobre la conveniencia de diseñar y producir un sistema distribuido de preservación digital para las tesis doctorales generadas en las universidades españolas REBIUN: Línea estratégica 1. Ponencia presentada en el XIV Workshop REBIUN sobre proyectos digitales / VI Jornadas OS-Repositorios Los horizontes de los repositorios Córdoba, 11 a 13 de marzo de 2015. <http://hdl.handle.net/20.500.11967/619>. .

Frías, L. (2017). Preservación digital de tesis doctorales: GT Repositorios: acción 5. Ponencia presentada en el Congreso Ecosistemas del Conocimiento Abierto (ECA 2017). Salamanca. <http://hdl.handle.net/10366/135600>.

Leija Román, D.; Térmens, M. (2017). "Preservación digital distribuida: lecciones de experiencias internacionales". BiD: textos universitaris de biblioteconomia i documentació, núm. 39 (diciembre). <https://bid.ub.edu/39/leija.htm>. DOI: <https://doi.org/10.1344/BiD2017.39.25>.

Leija, D.; Térmens, M. (2019). Traducción de Niveles de Preservación Digital NDSA 2019: traducción al español de versión 2.0. APREDIG -Asociación Iberoamericana de Preservación Digital. <https://osf.io/egjk8> <http://www.apredig.org/npdndsa2019/>.

Levels of digital preservation: versión 2.0 (2019). <https://ndsa.org/publications/levels-of-digital-preservation/>.

Térmens, M.; Leija, D. (2017). "Auditoría de preservación digital con NDSA Levels". El profesional de la información, v. 26, n. 3, p. 447-456. <https://doi.org/10.3145/epi.2017.may.11>.

UNIVERSIDADES PARTICIPANTES EN LA ENCUESTA

- CSIC
- IE University
- UNED
- Universidad a Distancia de Madrid UDIMa
- Universidad Autónoma de Madrid
- Universidad Carlos III de Madrid
- Universidad Católica de Valencia San Vicente Mártir
- Universidad CEU Cardenal Herrera
- Universidad de Alcalá
- Universidad de Alcalá
- Universidad de Alicante
- Universidad de Córdoba
- Universidad de Extremadura
- Universidad de Granada
- Universidad de Huelva
- Universidad de León
- Universidad de Murcia
- Universidad de Navarra
- Universidad de Oviedo
- Universidad de Sevilla
- Universidad del País Vasco / Euskal Herriko Unibertsitatea
- Universidad Francisco de Vitoria
- Universidad Internacional de Andalucía
- Universidad Internacional de La Rioja
- Universidad Miguel Hernández
- Universidad Pablo de Olavide
- Universidad Politécnica de Cartagena
- Universidad Politécnica de Madrid
- Universidad Pontificia Comillas
- Universidad Pontificia de Salamanca
- Universidad Pública de Navarra
- Universidad Rey Juan Carlos

- Universidad San Jorge
- Universidad San Pablo-CEU
- Universidade de Vigo
- Universitat Abat Oliba CEU
- Universitat Autònoma de Barcelona
- Universitat de Barcelona
- Universitat de Girona
- Universitat de les Illes Balears
- Universitat de València
- Universitat de Vic - Universitat Central de Catalunya
- Universitat Jaume I
- Universitat Politècnica de Catalunya (UPC)
- Universitat Politècnica de València
- Universitat Pompeu Fabra
- Universitat Ramon Llull
- Universitat Rovira i Virgili
- Universitat de Lleida

