

ESTUDIOS E INFORMES

REPOSITORIOS

2019



INFORME SOBRE LA EVALUACIÓN DEL ESTADO
DE LA PRESERVACIÓN DE LOS REPOSITORIOS
REBIUN 2018



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

**INFORME SOBRE LA EVALUACIÓN DEL ESTADO
DE LA PRESERVACIÓN DE LOS REPOSITORIOS
REBIUN 2018**

**REPORT ON THE EVALUATION OF THE STATE
OF PRESERVATION OF THE REPOSITORIES**

REBIUN Línea 3 (3er. P.E.) Grupo de Repositorios



Documento bajo licencia Creative Commons



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

Este informe ha sido elaborado por los miembros de la Acción 5 del Grupo de Repositorios de REBIUN (2018).

Cristina Azorín

Universitat Autònoma de Barcelona. cristina.azorin@uab.cat

José Manuel Barrueco

Universitat de València. jose.barrueco@uv.es

Cristal Martínez

Universidade de Santiago de Compostela. cristal.martinez@usc.es

Miquel Térmens

Universitat de Barcelona. termens@ub.edu

REBIUN, Red de Bibliotecas Universitarias Españolas, es una comisión sectorial de la Conferencia de Rectores de las Universidades Españolas (CRUE) desde 1998. El Grupo de Trabajo de Repositorios de REBIUN fue creado en 2011 con la finalidad de potenciar los repositorios de contenidos y datos de investigación y docencia institucionales, y aprovechar las tecnologías e interoperabilidad para impulsar nuevos servicios de valor añadido. Su coordinador es Ciro Lluca (Universitat Oberta de Catalunya).

20 de marzo de 2019

Resumen

El presente informe trata de determinar en qué medida los repositorios de investigación de las universidades miembros de REBIUN están llevando a cabo procesos de preservación digital de sus contenidos. Se engloba dentro de la acción 5 del Grupo de Repositorios para el año 2018 y en concreto dentro del objetivo 5.1 *Realización de una encuesta e informe sobre el estado de la preservación digital en los repositorios españoles.*

Para determinar el estado de la preservación digital en los repositorios se eligieron, como punto de referencia, los niveles establecidos por la National Digital Stewardship Alliance (NDSA) y se intentó determinar en qué medida los repositorios se adaptaban a los mismos. Esta correlación se estableció a través de una encuesta en línea pensada para que respondiesen los administradores de repositorios. Las respuestas se recibieron durante el mes de junio de 2018 con una tasa de respuestas del 90% (52 encuestas respuestas de 58 repositorios REBIUN contactados).

La conclusión principal es que no se están aplicando las medidas técnicas de preservación al nivel en el que se comprometen públicamente los repositorios en cuanto a la preservación de los materiales almacenados. Hay una ausencia prácticamente total de planes de preservación y las técnicas de preservación se están llevando a cabo de manera incipiente y parcial.

Palabras clave

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

Introducción

REBIUN/FECYT en su *Guía de Evaluación de repositorios*¹ define los repositorios como “aquel conjunto de servicios prestados por las universidades o centros de investigación a su comunidad para recopilar, gestionar, difundir y preservar su producción científica digital a través de una colección organizada, de acceso abierto e interoperable”. En esta definición se establece, entre las diferentes funciones que deben cumplir los repositorios, la acción de preservación de la producción científica almacenada. Esta función de preservación se atribuye de forma sistemática en toda la literatura publicada sobre el tema. El *Plan S*², aparecido recientemente, la sitúa incluso como una función estratégica al establecer que la producción se debe depositar en repositorios certificados para su conservación a largo plazo.

También parece claro, a partir de la literatura, que los gestores de repositorios han prestado más atención a las funciones de recopilación y difusión que a la preservación. Los repositorios son depósitos de documentos que necesitan aumentar el número de documentos y demostrar su utilidad a corto plazo para fomentar el uso por parte de los investigadores, el apoyo de su propia institución y el respaldo de las instituciones financiadoras. Frente a esta necesidad urgente se ha olvidado o relegado la función de preservación. La novedad de los contenidos hace que no se haya considerado un problema acuciante.

El grupo de trabajo de repositorios empezó a trabajar en temas de preservación en el año 2016 cuando se añadió la acción 5 –Acción 5: Sistema de preservación digital de contenidos de repositorios REBIUN–, que se concretó en un análisis interno de implementación de un sistema de preservación digital de las tesis doctorales alojadas en los repositorios REBIUN.

Con objeto de demostrar hasta qué punto el análisis de necesidades fue correcto, desde el grupo de repositorios de REBIUN se puso en marcha una nueva acción en

¹ Barrueco Cruz, José Manuel; Andrés Rodríguez, Aurelia; Rico Castro, Pilar; [et al.]. Guía para la evaluación de repositorios institucionales de investigación. 38 pág. <<https://recolecta.fecyt.es/sites/default/files/contenido/documentos/2017GuiaEvaluacionRecolectaFECYT.pdf>> [Consulta: 8 marzo 2019].

² Science Europe. Plan S: *10 Principles*. Recuperado de <https://www.coalition-s.org/10-principles/> [Consulta: 8 marzo 2019].

2018. Es importante establecer y conocer el nivel de cumplimiento de los repositorios españoles con el mandato de preservación a largo plazo de los contenidos digitales almacenados, más allá de establecer el compromiso dentro de sus políticas institucionales de acceso abierto. Se trata de determinar si existen políticas y planes de preservación específicos, así como el desarrollo de actividades de preservación concretas.

Para llevar a cabo el estudio se solicitó la ayuda de Miquel Térmens (Universitat de Barcelona) como reconocido especialista en preservación digital en nuestro país. Miquel se incorporó al grupo de trabajo de REBIUN desde el comienzo de la acción. Se tomaron como referencia los niveles establecidos por la National Digital Stewardship Alliance (NDSA) y se intentó determinar en qué medida los repositorios se adaptaban a los mismos. La NDSA se funda en 2010 como un consorcio de instituciones comprometidas con la preservación de objetos digitales³. Sus actividades se centran en la difusión de buenas prácticas y en la mejora de la formación de los profesionales.

En 2013 NDSA publicó una tabla, conocida con el nombre de *NDSA Levels of Digital Preservation*⁴, que permite auditar de forma simple en qué estado se encuentra la preservación de un determinado sistema digital.

NDSA Levels establece cuatro niveles de preservación y cinco grandes áreas o campos de actuación. Para cada campo se establecen una serie de preguntas a las que se debe contestar de forma afirmativa o negativa, si el repositorio cumple o no. Se intenta que el lenguaje utilizado sea sencillo y las definiciones de las preguntas concretas para que puedan ser respondidas por cualquier gestor con unos conocimientos técnicos mínimos, aunque en algunos casos puede ser necesaria la intervención de los servicios informáticos. Véase la tabla 1.

³ Miquel Térmens; David Leija: "Auditoría de preservación digital con NDSA Levels". *El profesional de la información*, v. 26, n. 3 (2017), p. 447-456.

⁴ NDSA. Levels of Digital Preservation. <http://www.digitalpreservation.gov/ndsas/activities/levels.html> [Consulta: 8 marzo 2019].

Metodología

La recogida de datos se basó en una encuesta en línea a cumplimentar por los responsables de los repositorios de las bibliotecas universitarias que forman parte de REBIUN.

El primer paso fue determinar el universo de trabajo. Existen 76 universidades miembros de REBIUN (50 de ámbito universitario público y 26 privado), además del CSIC. Se constató que no todas las universidades disponían de repositorio institucional, por lo que se recurrió a directorios de referencia para identificar el conjunto de universidades a las que se iba a mandar la encuesta. A través del Directorio de repositorios institucionales REBIUN⁵ y los repositorios registrados en RECOLECTA⁶ se identificaron 58 universidades con repositorio.

En segundo lugar, se adaptó el enunciado de las preguntas de los niveles NDSA a las circunstancias propias del entorno español y se añadieron textos de ayuda contextual. La encuesta resultante fue probada en dos rondas sucesivas por los miembros del grupo con el fin de asegurar que la redacción de las preguntas era clara, exacta y no daba lugar a diferentes interpretaciones por parte de los gestores que las habrían de responder.

La implementación se realizó utilizando el software Lime Survey, que pudo ser usado gracias a la colaboración de la Facultad de Biblioteconomía y Documentación de la Universidad de Barcelona.

⁵ Directorio de repositorios institucionales REBIUN: <http://direcrebiun.ulpgc.es/> [Consulta: 8 marzo 2019].

⁶ Repositorios registrados en RECOLECTA: <https://recolecta.fecyt.es/repositorios-recolectados> [Consulta: 8 marzo 2019].

Tabla 1: NDSA Levels of Digital Preservation

AREAS	Nivel 1 (Proteja sus datos)	Nivel 2 (Conozca sus datos)	Nivel 3 (Controle sus datos)	Nivel 4 (Repare sus datos)
Almacenamiento y localización geográfica	<p>Dos copias completas que no estén unidas.</p> <p>Para datos en soportes heterogéneos (discos ópticos, discos duros, etc.) quitar el contenido del soporte y ponerlo en vuestro sistema de almacenamiento.</p>	<p>Como mínimo tres copias completas.</p> <p>Como mínimo una copia en una localización geográfica distinta.</p> <p>Documentar su(s) sistema(s) de almacenamiento y soportes de almacenamiento y lo que usted necesite para usarlos.</p>	<p>Como mínimo una copia en una localización geográfica con una amenaza de desastres diferente.</p> <p>Controlar el proceso de obsolescencia de su(s) sistema(s) de almacenamiento y de sus soportes.</p>	<p>Como mínimo tres copias en localizaciones geográficas con amenazas de desastres diferentes.</p> <p>Disponer de un plan integral preparado para mantener los ficheros y los metadatos accesibles en los actuales soportes o sistemas.</p>
No alteración de ficheros e integridad de los datos	<p>Comprobar la integridad de los ficheros en el momento de la ingesta si sus valores han sido proporcionados junto con el contenido.</p> <p>Crear la información de integridad si no fue proporcionada junto con el contenido.</p>	<p>Comprobar la integridad de todas las ingestas.</p> <p>Usar dispositivos con escritura bloqueada cuando se trabaje con los soportes originales.</p> <p>Comprobar virus en contenido de alto riesgo.</p>	<p>Comprobar la integridad del contenido a intervalos regulares.</p> <p>Mantener registros de la información de integridad; realizar auditoría bajo demanda.</p> <p>Capacidad para detectar datos corrompidos.</p> <p>Comprobar virus en todo el contenido.</p>	<p>Comprobar la integridad de todo el contenido en respuesta a situaciones o actividades específicas.</p> <p>Capacidad para reemplazar o reparar datos corrompidos.</p> <p>Asegurar de que ninguna persona tiene acceso de escritura a todas las copias.</p>
Seguridad de la información	<p>Identificar quien ha leído, escrito, movido o eliminado la autorización a ficheros concretos.</p> <p>Restringir quien tiene este tipo de autorizaciones a ficheros concretos.</p>	<p>Documentar las restricciones de acceso de los contenidos.</p>	<p>Mantener registros de quien ha realizado qué acciones con los ficheros, incluyendo acciones de borrado y preservación.</p>	<p>Realizar auditorías de los registros.</p>
Metadatos	<p>Disponer de un inventario del contenido y de su localización en el almacenamiento.</p> <p>Asegurar una copia de seguridad separada del inventario.</p>	<p>Almacenar metadatos administrativos.</p> <p>Almacenar metadatos de las transformaciones y registrar las incidencias.</p>	<p>Almacenar metadatos estándar técnicos y descriptivos.</p>	<p>Almacenar metadatos estándar de preservación.</p>
Formatos de ficheros	<p>Cuando usted puede participar en la creación de archivos digitales fomentar el uso de un conjunto limitado de formatos abiertos y conocidos de ficheros y de códecs.</p>	<p>Disponer de un inventario de los formatos de ficheros usados.</p>	<p>Monitorear los problemas de obsolescencia de los formatos de ficheros.</p>	<p>Realizar migraciones de formatos, emulaciones o actividades similares si es necesario.</p>

Siguiendo las directrices de REBIUN, la encuesta se envió en primer lugar, a título informativo, a los directores de bibliotecas en formato PDF. A continuación, se envió el enlace a la encuesta en línea y las correspondientes instrucciones a los correos electrónicos que se ofrecen como contacto en las páginas web de los repositorios. Durante los meses de mayo y junio de 2018 se abrió la fase de respuestas de la encuesta. Se hizo un seguimiento de recordatorio a través del correo y finalmente un contacto telefónico, se consiguió una tasa de respuestas del 90% (52 de 58).

Resultados

En la siguiente tabla se ofrecen los resultados, el porcentaje de cumplimiento agregado para todas las instituciones de los niveles NDSA.

		Nivel 1	Nivel 2	Nivel 3	Nivel 4
1	Almacenamiento y localización geográfica	0,89	0,42	0,60	0,16
2	Alteración de ficheros e integridad de los datos	0,30	0,59	0,40	0,45
3	Seguridad de la información	0,86	0,00	0,61	0,14
4	Metadatos	0,42	0,59	0,81	0,44
5	Formatos de ficheros	0,88	0,64	NA	0,14
Media por niveles:		0,65	0,49	0,53	0,30

Como era de esperar, el nivel que mayor grado de cumplimiento alcanza es el 1, el más básico. Casi todas las instituciones disponen de dos copias de seguridad no unidas físicamente y tienen un control o conjunto definido de formatos de ficheros que se pueden depositar en el repositorio.

En el lado opuesto está el nivel 4, con un índice de cumplimiento del 0,30. Con la realización de auditorías como el resultado más bajo, seguida de la realización de migraciones de formatos. Quizás este último punto, debido a la corta vida aún de los repositorios y a que el formato más extendido en el ámbito científico sea el pdf, no ha dado lugar a la necesidad de llevar a cabo estas actividades de migración.

En el punto más negativo, con 0% de cumplimiento, tenemos la documentación de las restricciones de acceso a los contenidos, en el nivel 2. En este apartado la tabla NDSA pide que las restricciones de derechos que se aplican a las contraseñas usadas por el personal estén documentadas, es decir, que exista un documento que indique los derechos de cada nivel de contraseña y qué personal dispone de ellas. Según se deduce de la encuesta, los repositorios estudiados en un 86% aplican restricciones a los diferentes perfiles con contraseña pero ninguno lo tiene documentado. La consecuencia de esto puede ser que en caso de producirse un incidente de seguridad no sea posible probar legalmente qué personas concretas tenían la capacidad para realizar una determinada acción sobre los objetos digitales.

Es de destacar el bajo nivel (0,16) de instituciones que declaran disponer de 3 copias en 3 localizaciones separadas con riesgos distintos (nivel 4 de Almacenamiento). Esta es una de las acciones que pueden reportar más seguridad a los datos, pues asegura la disponibilidad de copias ante cualquier eventualidad, pero al mismo tiempo también es sin duda la acción de mayor coste económico. Desgraciadamente parece que las restricciones económicas que han soportado las instituciones académicas durante los últimos años han impedido el desarrollo de políticas avanzadas de copias de seguridad.

El uso de dispositivos de almacenamiento con escritura bloqueada (nivel 2) es muy bajo, pero tiene una explicación: la gran mayoría de los repositorios estudiados no ingresan objetos digitales a partir del volcado masivo de datos previamente almacenados en dispositivos de terceros como discos duros externos, por el contrario son ingresados principalmente por autoarchivo web o por transferencias internas de los administradores. No existe por tanto la necesidad de prevenir grabaciones accidentales.

Se observa que el nivel 3 presenta un índice de cumplimiento mayor que el nivel 2; esto es una anomalía. Si las acciones de preservación se aplicaran de forma planificada, por lógica, las acciones de niveles inferiores se aplicarían antes que las de los niveles superiores, pues estos tienen un carácter acumulativo y más especializado respecto a los primeros. La realidad no sigue esta pauta, pues tres bloques de acciones del nivel 3 presentan un mayor cumplimiento global que los respectivos bloques de acciones del nivel 2.

Los siete mejores resultados obtenidos son:

Comprobar la integridad de todas las ingestas	52
Mantener dos copias completas que no estén unidas físicamente	45
Fomentar el uso de un conjunto limitado de formatos abiertos	45
Restringir las autorizaciones para eliminar o modificar ficheros concretos	44
Identificar quien ha leído, escrito, movido o eliminado ficheros concretos	39
Almacenar metadatos estándar técnicos y descriptivos	39
Mover el contenido de datos en soportes heterogéneos (discos ópticos, discos duros, etc.) a otro soporte dentro del sistema de almacenamiento del repositorio	38

En el lado opuesto los siete peores resultados son:

Almacenar metadatos de las transformaciones y registrar las incidencias	8
Realizar auditorías de los registros	8
Realizar migraciones de formatos, emulaciones o actividades similares si es necesario	7
Comprobar la integridad de los ficheros en el momento de la ingesta si sus valores han sido proporcionados junto con el contenido	7
Documentar las restricciones de acceso de los contenidos	6
Usar dispositivos con escritura bloqueada cuando se trabaje con soportes originales	5
Mantener tres copias en localizaciones geográficas con amenazas de desastres distintas	3

Como ya hemos comentado anteriormente, cabe destacar dos preguntas relacionadas con la ingesta masiva de contenidos: *Comprobar la integridad de los ficheros en el momento de la ingesta si sus valores han sido proporcionados junto con el contenido* y *Usar dispositivos con escritura bloqueada cuando se trabaje con soportes originales*. La baja tasa de respuestas puede deberse a que los repositorios no suelen hacer cargas sistemáticas desde soportes originales, sino que la vía normal de ingesta es a través del depósito de los investigadores. La ingesta masiva es más propia de colecciones patrimoniales, donde los ficheros son generados por empresas externas y se suministran a través de soportes externos al repositorio.

Conclusiones

El estudio permite llegar a las siguientes conclusiones:

- Sólo 2 repositorios de 52 cuentan con planes de preservación aprobados y en funcionamiento.
- Las técnicas de preservación digital se están aplicando de forma incipiente y parcial en la mayoría de repositorios.
- Se detectan graves problemas en cuanto al número de copias de seguridad y su ubicación, las comprobaciones de la integridad de los ficheros y la capacidad para realizar auditorías en el caso de detectarse incidencias de seguridad.
- No se está aplicando a nivel técnico el compromiso público que se reconoce en cuanto a la preservación de los materiales almacenados.

Bibliografía

- Dryden, Jean (2011). "Measuring Trust: Standards for Trusted Digital Repositories". *Journal of Archival Organization*, v. 9, n. 2, p. 127-130. <http://dx.doi.org/10.1080/15332748.2011.590744> [Consulta: 8 marzo 2019].
- Faundeen, John (2014). "Building Trust: NDSA Levels of Digital Preservation". En: *Digital Preservation*, 2014. http://www.digitalpreservation.gov/meetings/documents/ndiipp14/Faundeen_NDSALevels.pdf [Consulta: 8 marzo 2019].
- Goethals, Andrea (2013). "An Example Self -Assessment Using the NDSA Levels of Digital Preservation". En: *CAIB Ipres*, 2013. <https://benchmarkdigitalpreservation.files.wordpress.com/2013/09/caiw2013goethals.pdf> [Consulta: 8 marzo 2019].
- Harvard Library (2014). *The New DRS: Plan for Metadata Migration Harvard Library & Library Technology Services*. Harvard Library. <http://slideplayer.com/slide/3854725/> [Consulta: 8 marzo 2019].
- Levels (2013). *Levels of digital preservation*. <http://ndsa.org/activities/levels-of-digital-preservation/> [Consulta: 8 marzo 2019].
- Phillips, Megan; Bailey, Jefferson; Goethals, Andrea; Owens, Trevor (2013). "The NDSA Levels of Digital Preservation: An Explanation and Uses". En: *Archiving*, 2013.

http://digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf [Consulta: 8 marzo 2019].

Ross, Seamus; McHugh, Andrew (2006). "The Role of Evidence in Establishing Trust in Repositories". *D-Lib Magazine*, v. 12, n. 7/8. <http://www.dlib.org/dlib/july06/ross/07ross.html> [Consulta: 8 marzo 2019].

Self-assessment (2016). "Self-assessment as digital preservation training aid". *DART Blog*. 18/04/2016. <https://dart.blogs.ulcc.ac.uk/2016/04/18/self-assessment-training-aid-dptp/> [Consulta: 8 marzo 2019].

Térmens, Miquel; Leija, David (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> [Consulta: 8 marzo 2019].

Ying, William (2013). "ARTstor Shared Shelf Preservation Plan Based on the NDSA Levels of Digital Preservation". En: *Digital Preservation 2013 Meeting*. <http://www.digitalpreservation.gov/meetings/documents/ndiipp13/Ying.pdf> [Consulta: 8 marzo 2019].

Listado de instituciones que respondieron a la encuesta

CSIC

UNED

Universidad Autónoma de Madrid

Universidad Camilo José Cela

Universidad Carlos III de Madrid

Universidad Complutense de Madrid

Universidad de Alcalá

Universidad de Alicante

Universidad de Almería

Universidad de Burgos

Universidad de Córdoba

Universidad de Cantabria

Universidad de Extremadura

Universidad de Huelva

Universidad de Jaén

Universidad de La Laguna

Universidad de Las Palmas de Gran Canaria

Universidad de León

Universidad de Murcia

Universidad de Navarra

Universidad de Oviedo

Universidad de Salamanca

Universidad de Sevilla

Universidad de Valladolid

Universidad de Zaragoza

Universidad del País Vasco / Euskal Herriko Unibertsitatea

Universidad Europea de Madrid

Universidad Francisco de Vitoria

Universidad Internacional de Andalucía

Universidad Internacional de La Rioja

Universidad Miguel Hernández de Elche

Universidad Pública de Navarra

Universidad Politécnica de Cartagena
Universidad Politécnica de Madrid
Universidad Pontificia de Salamanca
Universidad Rey Juan Carlos
Universidade da Coruña
Universidade de Santiago de Compostela
Universidade de Vigo
Universitat Autònoma de Barcelona
Universitat de Barcelona
Universitat de Girona
Universitat de les Illes Balears
Universitat de Lleida
Universitat de València
Universitat de Vic - Universitat Central de Catalunya
Universitat Jaume I
Universitat Oberta de Catalunya
Universitat Politècnica de Catalunya
Universitat Politècnica de València
Universitat Pompeu Fabra
Universitat Rovira i Virgili

