



16º Workshop de REBIUN de Proyectos Digitales

7ªs Jornadas OS-Repositorios

11º Coloquio Internacional de Ciencias de la Documentación

Salamanca 25, 26 y 27 de octubre de 2017

Análisis del estado de la preservación digital en los repositorios de las universidades mexicanas

David Leija

Universidad Autónoma de Tamaulipas, México

dleija@uat.edu.mx

Miquel Térmens

Universitat de Barcelona

termens@ub.edu

Resultados de la tesis

David Alonso Leija Román: *Preservación digital distribuida y la colaboración interinstitucional: Modelo de preservación digital para documentos con fines de investigación en universidades de México*. Universitat de Barcelona, 2017.



Ha contado con el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT) de México, del proyecto *El acceso abierto a la ciencia en España: evaluación de su impacto en el sistema de comunicación científica (Plan Nacional, ref. CSO2014-52830-P)*. También ha recibido soporte del GRC *Cultura i continguts digitals: aspectes documentals, polítics i econòmics*.

Sistema público de educación superior en México

| | |
|--|------------|
| Universidades Públicas Federales | 9 |
| Universidades Públicas Estatales | 34 |
| Institutos Tecnológicos Federales | 132 |
| Institutos Tecnológicos Descentralizados | 130 |
| Universidades Públicas Estatales con Apoyo Solidario | 23 |
| Universidades Tecnológicas | 104 |
| Universidades Politécnicas | 50 |
| Universidades Interculturales | 12 |
| Centros Públicos de Investigación | 6 |
| Educación Normal Superior | 261 |
| Otras Instituciones Públicas | 86 |
| Total | 847 |

- 43.000 investigadores
- 12.000 artículos/año

Repositorios

El año 2015 solamente 43 instituciones públicas de educación superior tenían repositorios.

Pero no todos son de acceso abierto...

Según datos de OpenDOAR (2017), México cuenta con 30 repositorios en acceso abierto, de los cuales 19 pertenecen a instituciones públicas. Por tanto solamente tienen repositorio abierto un 3,3% del total de 900 instituciones de educación superior públicas y privadas.

Encuesta

La encuesta se aplicó mediante el sistema Survey Monkey entre el 18 de mayo y el 8 de julio de 2015

| Instituciones de educación superior | Número | Con respuesta | Porcentaje de participación |
|-------------------------------------|--------|---------------|-----------------------------|
| Federales | 9 | 8 | 89 % |
| Estatales | 34 | 28 | 82 % |
| Total | 43 | 36 | 86 % |

Materiales depositados en los repositorios

| Material depositado | Respuestas positivas | Porcentaje |
|----------------------------|----------------------|------------|
| Tesis académicas | 27 | 77 % |
| Libros digitales | 25 | 71 % |
| Artículos de investigación | 21 | 60 % |
| Bases de datos | 17 | 49 % |
| Material educativo | 16 | 46 % |
| Videos | 12 | 34 % |
| Fotografías | 9 | 26 % |
| Audios | 7 | 20 % |
| Otro | 5 | 14 % |

NDSA Levels of Digital Preservation

<http://ndsa.org/activities/levels-of-digital-preservation/>

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|--|---|---|--|---|
| Almacenamiento y localización geográfica | <p>Dos copias completas que no estén unidas.</p> <p>Para datos en soportes heterogéneos (discos ópticos, discos duros, etc.) mover el contenido a otro soporte dentro de su sistema de almacenamiento.</p> | <p>Como mínimo tres copias completas.</p> <p>Como mínimo una copia en una localización geográfica distinta.</p> <p>Documentar el(los) 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 distinta.</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 distintas.</p> <p>Disponga 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 soportes originales.</p> <p>Comprobar virus en contenido de alto riesgo</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>Asegúrese 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 que acciones con los ficheros, incluyendo las acciones de borrado y preservación.</p> | <p>Realizar auditorías de los registros.</p> |
| Metadatos | <p>Inventario del contenido y de su localización en el almacenamiento.</p> <p>Asegurar la 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 fomente 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> |

Resumen de resultados según niveles NDSA

| Nivel NDSA | Número de elementos a controlar | Porcentaje de cumplimiento |
|------------|---------------------------------|----------------------------|
| Nivel 1 | 9 | 35% |
| Nivel 2 | 10 | 26% |
| Nivel 3 | 9 | 25% |
| Nivel 4 | 8 | 27% |

Almacenamiento y localización geográfica

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|--|--|---|--|--|
| Almacenamiento y localización geográfica | <p>Dos copias completas que no estén unidas.</p> <p>Para datos en soportes heterogéneos (discos ópticos, discos duros, etc.) mover el contenido a otro soporte dentro de su sistema de almacenamiento.</p> | <p>Como mínimo tres copias completas.</p> <p>Como mínimo una copia en una localización geográfica distinta.</p> <p>Documentar el(los) 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 distinta.</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 distintas.</p> <p>Disponga de un plan integral preparado para mantener los ficheros y los metadatos accesibles en los actuales soportes o sistemas.</p> |

- Se obtuvo que el 25% de las instituciones tienen 2 copias de seguridad completas y separadas y el 3% tienen 3 copias. Por tanto el 72% de las instituciones o no realizan copias de seguridad o estas se encuentran almacenadas en el mismo servidor, con lo que su eficacia es baja ante un desastre grave.
- Es baja la unificación sistemática de soportes variados (CD-ROM, pendrive, disquete) en un único soporte (31%) o disponer de un plan de migración futura de soportes (28%).

No alteración de archivos e integridad de los datos

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|---|---|--|--|---|
| 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 soportes originales.</p> <p>Comprobar virus en contenido de alto riesgo</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>Asegúrese de que ninguna persona tiene acceso de escritura a todas las copias.</p> |

- El 39% de las instituciones comprobaban su integridad de los ficheros durante la ingesta, pero que luego solamente el 19% continuaba realizando comprobaciones periódicas a lo largo del tiempo. Solamente el 8% guardaban registros de las comprobaciones realizadas.
- En caso de corrupción de ficheros, solamente el 33% declaraba disponer de la capacidad para recuperar los datos originales.
- Únicamente el 33% de las organizaciones podía hacer una comprobación de antivirus en la totalidad de sus fondos digitales.

Seguridad de la información

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|-----------------------------|---|---|--|---------------------------------------|
| Seguridad de la información | Identificar quien ha leído, escrito, movido o eliminado la autorización a ficheros concretos. Restringir quien tiene este tipo de autorizaciones a ficheros concretos. | Documentar las restricciones de acceso de los contenidos. | Mantener registros de quien ha realizado que acciones con los ficheros, incluyendo las acciones de borrado y preservación. | Realizar auditorías de los registros. |

- Sólo el 44% de los repositorios identificaban quien ha leído, escrito, movido o eliminado ficheros concretos. Pero aún eran menos (el 14%) los que podían realizar auditorías en los registros de eventos para identificar qué persona había realizado una determinada acción.
- El 48% de los repositorios no mantenían ningún tipo de restricción en cuanto a la manipulación de los ficheros ni limitaban que empleados lo podían hacer.

Metadatos

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|-----------|--|--|--|---|
| Metadatos | Inventario del contenido y de su localización en el almacenamiento. Asegurar la copia de seguridad separada del inventario. | Almacenar metadatos administrativos. Almacenar metadatos de las transformaciones y registrar las incidencias. | Almacenar metadatos estándar técnicos y descriptivos | Almacenar metadatos estándar de preservación. |

- El 69% de los repositorios declaraban que no disponían de un inventario completo de sus fondos y, por ejemplo, no eran capaces de identificar en que soporte físico concreto estaba almacenado un determinado objeto digital.
- El 61% de las instituciones guardaban metadatos descriptivos o de tipo técnico, pero solamente el 36% también guardaba metadatos de preservación, tipo PREMIS. Únicamente el 6% documentaban en forma de metadatos las transformaciones o migraciones de formatos que habían aplicado a los ficheros.

Formatos de ficheros

| | Nivel 1 | Nivel 2 | Nivel 3 | Nivel 4 |
|----------------------|---|---|--|--|
| Formatos de ficheros | Cuando usted puede participar en la creación de archivos digitales fomentando el uso de un conjunto limitado de formatos abiertos y conocidos de ficheros y de códecs | Disponer de un inventario de los formatos de ficheros usados. | Monitorear los problemas de obsolescencia de los formatos de ficheros. | Realizar migraciones de formatos, emulaciones o actividades similares si es necesario. |

- El 56% de los responsables que contestaron el cuestionario afirmaron que participaban en la toma de decisiones que afectaban al uso de nuevos formatos en documentos creados dentro de su organización. Luego se reducía al 42% aquellos que disponían de inventario de todos los formatos almacenados en su repositorio, y aún bajaba más, al 28%, aquellos que disponían de capacidad para monitorear la posible obsolescencia de estos formatos.

Conclusiones

- Los repositorios son una herramienta de la política científica nacional.
- La permanencia (preservación) de sus contenidos forma parte de esta política.
- Existe un bajo número de repositorios.
- Las actuaciones de preservación de los repositorios son incipientes.
- Es imprescindible que los responsables de los repositorios tomen conciencia del problema y actúen.
- Se constata la existencia de muchas actuaciones parciales o equivocadas.

¡Gracias!

David Leija

Universidad Autónoma de Tamaulipas, México

dleija@uat.edu.mx

Miquel Térmens

Universitat de Barcelona

termens@ub.edu