CONFERENCIA INTERNACIONAL
BIREDIAL-ISTEC '22
3-7 de octubre 2022
COSTA RICA

Evaluación de CIC Digital a través de NDSA Levels

Santiago Tettamanti; De Giusti, Marisa R.; Lira, Ariel J.

PREBI-SEDICI - UNLP
CESGI, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

04 Octubre 2022

Contexto

CIC Digital

- Repositorio Institucional de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina
- Preservar y dejar accesible en abierto toda la producción científico-tecnológica de la CIC
- En la actualidad cuenta con alrededor de 9.000 ítems
- https://digital.cic.gba.gob.ar/

CIC DIGITAL
Repositorio Institucional

Contexto

¿Qué es NDSA?

- Alianza Nacional para la Custodia Digital (National Digital Stewardship Alliance)
- Consorcio de organizaciones comprometidas con la conservación de la información digital
- Creado en 2010 por la Library of Congress
- 267 instituciones miembro

NDSA LEVELS

- Levels of Digital Preservation creado en 2013
    - Iniciativa de NDSA
- Ayudar a instituciones a evaluar su sistema de preservación digital

NDSA LEVELS

Matriz sobre la cual se definen **tareas** para asegurar la **preservación y el acceso** a largo plazo.

Cinco categorías distintas:

- Almacenamiento
- Integridad de los datos
- Control de la información
- Metadatos
- Contenido

Cuatro niveles de cumplimiento progresivos.

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 qué 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 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
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 garanticen el acceso al contenido

Traducción de Niveles de Preservación Digital NDSA 2019: Traducción al Español de Versión 2.0. (Leija, D; Térmens, M, 2019).

NDSA LEVELS - Matriz de evaluación

- Es progresiva:
  - Las acciones en el primer nivel son requisitos necesarios para los niveles superiores
    - O son actividades más urgentes a lograr
- A diferencia de los métodos tradicionales
  - Detectan las **acciones** a implementar
  - Facilidad de aplicación
    - Las auditorías tradicionales suelen ser caras y de una complejidad alta
  - Permite la **autoevaluación**
    - No se requiere personal experto en preservación
  - Ideal para **pequeños y medianos repositorios**

NDSA LEVELS - Matriz de evaluación

Sin embargo…

- Ofrece una **visión simplificada** de las tareas de preservación
- No un listado exhaustivo de las mismas como en Nestor, TRAC o ISO 16363.
- Visión optimista de la situación real
- Ignoran otras facetas claves que sí cubren otros métodos (ej, ISO 16363)

Evaluación de CIC Digital

Metodología

- Se marcó para cada tarea o punto si se cumplía o no con lo propuesto.
- Con un color las respuestas afirmativas, con otro los puntos completos de manera parcial, y con otro las negativas.
- Se sumaron puntos por cada respuesta afirmativa y se obtuvo un puntaje total.

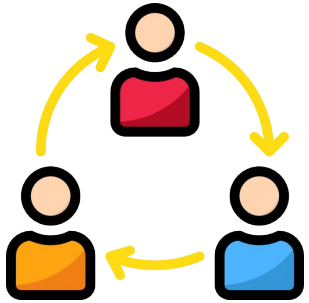
Evaluación de CIC Digital

Metodología

La evaluación fue realizada de **manera conjunta** por los distintos **equipos** de CIC Digital.

- Equipo de infraestructura del repositorio
- Equipo de desarrollo
- Equipo administrativo
- Encargados de la gestión del repositorio

En repositorios con poco personal una misma persona puede pertenecer a varios equipos.



Matriz de evaluación 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 aisla 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	

Matriz de evaluación para CIC Digital

Seguridad de la información	Se determinan los agentes humanos y de software qué 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	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	Registrar las acciones de preservación asociadas con el contenido y cuándo ocurren esas acciones Implementa los estándares de metadatos elegidos	3/4
	Hacer una copia de respaldo del inventario y se almacena al menos una copia por separado		Encuentra y completa los vacíos en sus metadatos para cumplir con esos estándares		
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	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
		Establecer relaciones con los creadores de contenido para fomentar la elección sostenible de archivos			
Puntaje global	3/5	2/5	3/5	1/5	9/20

Resultados

CIC Digital cumple casi con la mitad de las recomendaciones de la matriz

- 9 de las 20 recomendaciones fueron marcadas como completadas
- El cumplimiento es menor al avanzar en los niveles

Hay muchas mejoras posibles

- Propuesta de mejora por cada categoría

Propuesta de mejora

Almacenamiento y localización geográfica

- Backup off site (ej, en la nube), documentación de procesos y soportes y plan de gestión de riesgos

No alteración de archivos e integridad de los datos

- Tarea de curación para el análisis de virus y chequeo de integridad ante eventos

Seguridad de la información

- Mejorar el seguimiento y el registro de cambios en los ítem, bundle de preservación.

Metadatos

- Mantener un histórico o historial de cambios en metadatos, metadatos de preservación (PREMIS)

Formatos de archivos

- La ejecución periódica de esta tarea de curation para el chequeo de formatos

Conclusión

La matriz de autoevaluación de **NDSA**

- Recomendable para repositorios que no poseen los recursos para una auditoría externa experta.
- **Primer paso** de evaluación en el camino hacia la certificación en preservación
- **No** es un reemplazo a los métodos existentes (normas ISO u otros sistemas de auditoría)
    - Siguiente etapa a la que debería apuntar un repositorio.
- Ideal para el caso de CIC Digital
    - Repositorio de tamaño mediano
    - Nunca se le había realizado ningún análisis en términos de preservación digital

Conclusión

CIC Digital

- Cumple de manera satisfactoria gran parte del nivel 1
- Tiene varias falencias en el resto de los niveles
    - Especialmente documentación de procesos, recursos, planeamiento de la seguridad y riesgos.
- Mucho para hacer en el corto plazo
    - Plan de preservación y de riesgos, copias de seguridad off site, chequeos periódicos de formatos.

CONFERENCIA INTERNACIONAL
BIREDIAL-ISTEC '22
3-7 de octubre 2022
COSTA RICA

¡Muchas gracias!

PREBI SEDICI
prebi.unlp.edu.ar sedici.unlp.edu.ar

EDUCACIÓN PÚBLICA Y GRATUITA

UNIVERSIDAD NACIONAL DE LA PLATA

CESGI
Centro de Servicios en Gestión de Información
cesgi.cic.gba.gob.ar

COMISIÓN DE INVESTIGACIONES CIENTÍFICAS