GUÍA PARA REALIZAR COPIAS DE SEGURIDAD DEL REPOSITORIO INSTITUCIONAL DE LA UNSA

INTRODUCCIÓN

El Repositorio Institucional de la UNSA, reconoce que la preservación de los contenidos es el conjunto de principios, políticas, normas y estrategias diseñadas para asegurar que un objeto digital permanezca accesible, inteligible y usable a través del tiempo y de los cambios tecnológicos. Por tal motivo, el Repositorio Institucional expresa su compromiso en hacer disponibles los contenidos de forma permanente y tomar las medidas de preservación (tales como migraciones) necesarias para garantizar el acceso a los mismos.

OBJETIVO

El objetivo del presente documento es establecer la guía para realizar copias de seguridad, tanto del software sobre el que funciona el repositorio, los metadatos y los documentos digitales.

ALCANCE

La guía es aplicable al personal encargado y/o responsable de gestión de la Dirección Universitaria de Gestión de la Información.

BASE NORMATIVA

- Ley N° 30220 – Ley Universitaria.
- Estatuto Vigente de la Universidad Nacional de San Agustín de Arequipa
- Directrices para repositorios institucional de la Red Nacional de Repositorios Digitales de Ciencia, Tecnología e Innovación de Acceso Abierto (RENARE)
- Reglamento del Registro Nacional de Trabajos conducentes a Grados y Título – RENATI – aprobado por Resolución del Consejo Directivo N° 174–2019–SUNEDU/CD

DEFINICIONES Y ABREVIATURAS

Copia de seguridad– (backup):

Se define como backup o copia de seguridad, la actividad de resguardar de forma segura la información contenida en un medio de almacenamiento de origen (disco duro) a un medio de almacenamiento de destino de diferente tipo (otro disco duro, servidor de backup, memoria USB, CD, DVD, ZIP, etc.).

Archivos digitales

Son todos aquellos archivos digitales registrados, cargados en el Sistema de Registro de Investigación Docente de la UNSA

Base de datos

Una base de datos es un sistema informatizado cuyo propósito principal es mantener información y hacer que esté disponible en el momento requerido. Esta información es persistente dentro del sistema, es decir, una vez introducida en él, se mantiene hasta que el usuario decida eliminarla

Medios de almacenamiento de datos

Un dispositivo de almacenamiento de datos es un dispositivo para grabar o almacenar información (datos). La grabación se puede hacer usando virtualmente cualquier forma de energía. Un dispositivo de almacenamiento puede guardar la información y procesarla, o ambas. Un dispositivo que únicamente guarda la información es un dispositivo de grabación. Dispositivos que procesan la información (equipo de almacenamiento de datos) pueden tener acceso a un medio extraíble (portable) separado o a un componente permanente para almacenar y recuperar la información.

DSPACE

DSpace es un software de código abierto que provee herramientas para la administración de colecciones digitales, y comúnmente es usada como solución de repositorio bibliográfico institucional. Soporta una gran variedad de datos, incluyendo libros, tesis, datos de investigación, etc.

UGINF

Unidad de Gestión de la Información.

Asistente Técnico UGINF

Personal encargado de realizar las actividades de copias de seguridad.

RESPONSABLES

El Asistente Técnico UGINF, es responsable del resguardo de las copias de seguridad de archivos digitales, carpetas y bases de datos.

DESARROLLO

Actividad	Responsable	Descripción de la Actividades
1. Acceder al servidor del REPOSITORIO	Asistente Técnico DUGINF	El Asistente Técnico de la UGINF, accede al servidor del DSPACE a través de las credenciales correspondientes.
2. Realizar la copia de seguridad de archivos digitales y carpetas del DSPACE	Asistente Técnico UGINF	Se realiza las copias de seguridad de las carpetas del Repositorio: ▪ [dspace]/[dspace–source], si no se ha realizado cambios. ▪ [dspace]/assetstore ▪ [dspace]/config ▪ [dspace]/log ▪ [dspace]/webapps (las que tienen relación con el funcionamiento del Repositorio) y, ▪ Carpetas en las que se ha realizado cambios Nota 1: En el caso de la carpeta *assetstore*, las copias serán de preferencia de tipo incremental. Nota 2: En el caso de las demás carpetas, serán de manera mensual o antes de realizar cambios en nivel correspondiente.

3. Realizar la copia de seguridad de la base de datos	Asistente Técnico UGINF	Realiza la copia de seguridad de la base de datos del Dspace. El que deberá ser de manera diaria.
4. Guardar las copias de seguridad en medio de almacenamiento de destino	Asistente Técnico UGINF	Realiza el copiado de los archivos de copias de seguridad (base de datos y carpetas) en una unidad de disco externa y 2 respaldos diferentes en línea (u otro disco duro, servidor de backup, memoria USB, CD, DVD, ZIP, etc.)
E. Actualizar el inventario de copias de seguridad	Asistente Técnico UGINF	Registra la actividad en el inventario de copias de seguridad del Anexo 1 Nota: de manera mensual el Asistente Técnico realiza el informe de inventario realizado o según necesidad de información.

ANEXO 1

REGISTRO DE INVENTARIO DE COPIAS DE SEGURIDAD

Nro	Denominación de Ia información	Fecha y hora	Tipo de Información[1]	Ubicación		Tipo de Backup[2]	Tamaño (MB)	Medio de almacenamiento
				Nombre del servidor	IP			

[1] Tipo (archivo/carpeta, Base de datos, Aplicativo, Programa fuente, archivo configuración, etc)

[2] Tipo de Backup/respaldo: Completo, incremental, diferencial

ANEXO 2
PROTOCOLO DE RESPALDO DE BASE DE DATOS

1. Participantes y responsables

Actores	Observaciones
Asistente Técnico UGINF	Desarrollador directo o responsable de la tarea a poner en producción

2. Actividades

Durante esta tarea el desarrollador realizan estas actividades:

- Generación de respaldo SQL

  Será necesario indicar la fecha en la que se generó el respaldo

  fecha=`date +%Y%m%d`

  La nomenclatura para el nombre del backups será:

  archivo="dbunsa_NOMBRESISTEMA_backup_$fecha.sql"

  La nomenclatura para el nombre de la carpeta en la que se guardaran los backups será:

  backups_NOMBRESISTEMA

- Puesta en automatización de respaldo.

  Para la automatización de este proceso deberá crearse un archivo bash y ejecutarlo a nivel de sistema operativo todos los días con un CRON.

  Crontab es un archivo en nuestro Linux: Cron es el administrador de procesos, Cron lee en un archivo de texto plano llamado Crontab en donde se guardan una lista de comandos creados por el usuario ha ser ejecutados; la sintaxis para crear una de estas tareas programadas en muy sencilla:

```
* * * * * comando ha ser ejecutado
- - - - -
| | | | |
| | | | -------Día de la semana (0 - 7)
| | | ---------- Mes (1 - 12)
| | -------------Día del mes (1 - 31)
| ---------------- Hora (0 - 23)
                   Minuto (0 - 59)
```

  Formato básico de una tarea en el Crontab: Agregar tareas al Crontab
  Básicamente consta de dos mitades:

- El tiempo en el cual se ejecutará la tarea del Crontab; el mismo está compuesto por operadores y resulta de lo más versátil:
  - Cada minuto: En intervalos de entre 0 a 59.
  - Cada hora: En intervalos de entre 0 a 23.
  - Cada día: En intervalos de entre 0 a 31.
  - Cada mes: En intervalos de entre 0 a 12 (0==12 y 12 == Diciembre).
  - Cada día de la semana: En intervalos de entre 0 a 7 (0==7 y 7 == domingo).
- El comando BASH

  Ejemplo:

```
#!/bin/bash

fecha=`date +%Y%m%d`
archivo="dbunsa_siri_backup_$fecha.sql"
mysqldump --user=root --password=root siri_unsa > backups_siri/$archivo
&& \
echo "Respaldo realizado exitosamente el `date`" >> backups_siri/log.txt
```

En consideración esta tarea debe realizarse todos los días a las 23 horas

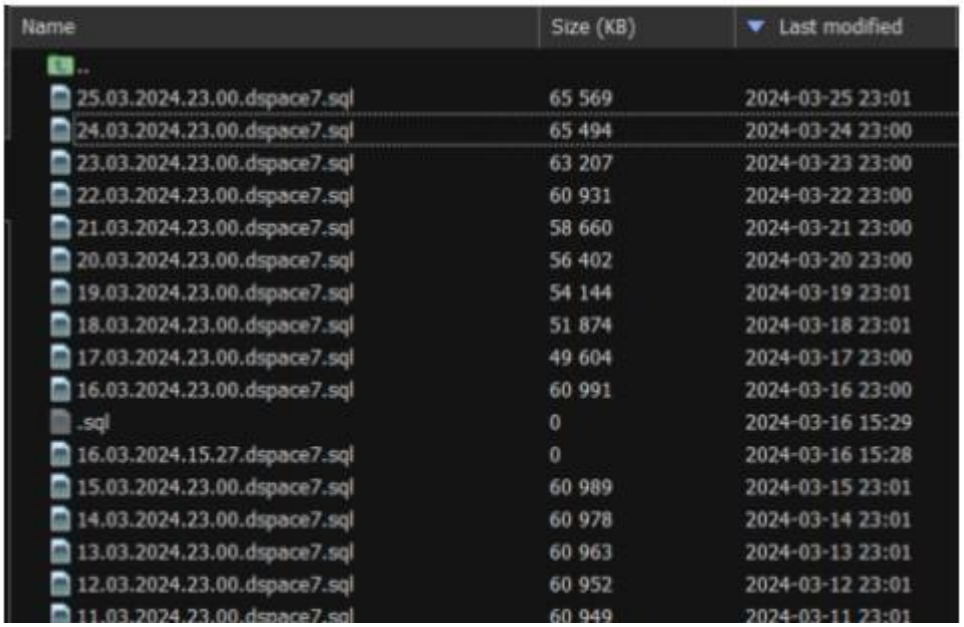
```
siri@siri:~$ crontab -l
# Edit this file to introduce tasks to be run by cron.
#
# Each task to run has to be defined through a single line
# indicating with different fields when the task will be run
# and what command to run for the task
#
# To define the time you can provide concrete values for
# minute (m), hour (h), day of month (dom), month (mon),
# and day of week (dow) or use '*' in these fields (for 'any').#
# Notice that tasks will be started based on the cron's system
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h  dom mon dow   command
0 23 * * * siri /home/siri/createbackups.sh
siri@siri:~$
```

- Documentación de proceso instalado.

Deberá generarse un documento que haga mención al proceso verificado y puesto en funcionamiento.

- Limpieza de servidor

  Cada 15 días deberá recuperarse los backups generados y almacenarse en la respectiva carpeta DRIVE



Name	Size (KB)	Last modified
..		
25.03.2024.23.00.dspace7.sql	65 569	2024-03-25 23:01
24.03.2024.23.00.dspace7.sql	65 494	2024-03-24 23:00
23.03.2024.23.00.dspace7.sql	63 207	2024-03-23 23:00
22.03.2024.23.00.dspace7.sql	60 931	2024-03-22 23:00
21.03.2024.23.00.dspace7.sql	58 660	2024-03-21 23:00
20.03.2024.23.00.dspace7.sql	56 402	2024-03-20 23:00
19.03.2024.23.00.dspace7.sql	54 144	2024-03-19 23:01
18.03.2024.23.00.dspace7.sql	51 874	2024-03-18 23:01
17.03.2024.23.00.dspace7.sql	49 604	2024-03-17 23:00
16.03.2024.23.00.dspace7.sql	60 991	2024-03-16 23:00
.sql	0	2024-03-16 15:29
16.03.2024.15.27.dspace7.sql	0	2024-03-16 15:28
15.03.2024.23.00.dspace7.sql	60 989	2024-03-15 23:01
14.03.2024.23.00.dspace7.sql	60 978	2024-03-14 23:01
13.03.2024.23.00.dspace7.sql	60 963	2024-03-13 23:01
12.03.2024.23.00.dspace7.sql	60 952	2024-03-12 23:01
11.03.2024.23.00.dspace7.sql	60 949	2024-03-11 23:01

3. Archivos

Bajo la misma lógica se respaldan archivos en caso los hubiera, usando un script automático incremental y comprimiéndolos.

4. Integridad de información

Se implementan scripts automáticos para para Curador de Base de Datos y Checker Sum: Se habilitaron herramientas automáticas para el curado de la base de datos y la verificación de la suma de comprobación (checksum) para garantizar la integridad y seguridad de los archivos almacenados en nuestro repositorio.

```
# m h  dom mon dow   command
00 22 * * * /usr/local/sbin/update-ngxblocker -e repositorio@unsa.edu.pe
*/59 * * * * /home/deskop/scripts/dspace-verify-handle.sh
0 8,20 * * * /home/dspace7/scripts/restart-nginx.sh
0 23 * * * /home/dspace/createbackups.sh
0 0 * * * /home/dspace7/dspace-app/bin/dspace oai import -o > /dev/null
0 4 * * * /home/dspace7/dspace-app/bin/dspace curate -q admin_ui > /dev/null
0 4 * * * /home/dspace7/dspace-app/bin/dspace checker -l -p > /dev/null
0 5 * * 0 /home/dspace7/dspace-app/bin/dspace checker-emailer > /dev/null
0 1 1 * * /home/dspace7/dspace-app/bin/dspace cleanup > /dev/null
01 0 1 * * find /home/dspace7/dspace-app/log/*.log.* -mtime +30 -exec rm {} \;
```

5. Carpeta DRIVE

La carpeta de documentación del proyecto debe contener una subcarpeta exclusiva para el almacenamiento de backups

Ejemplo: La carpeta dependiendo del sistema almacena el historial de backups