# Explicación de Scripts de Automatización de Archivematica

## Pasos previos

1. Configurar el demonio RSYNC en el servidor DSpace. Usa el puerto 873
   1. https://www.atlantic.net/hipaa-compliant-cloud-storage/how-to-setup-rsync-daemon-linux-server/
2. Instalar un cliente Python en el servidor DSpace y Archivematica, de preferencia Miniconda que tenga una versión de Python 3.6.
3. Instalacion del cliente Archivematica en el servidor
   1. Puede seguir las instrucciones de la página oficial:
      1. https://www.archivematica.org/es/docs/archivematica-1.6/admin-manual/installation/installation/#installation
   2. O seguir estás instrucciones más condensadas:
      1. https://github.com/research-technologies/hull\_provisioning/wiki/Install-Archivematica
4. Habilitar el RTS Replication Task Suite en DSpace
   1. https://wiki.duraspace.org/display/DSPACE/ReplicationTaskSuite
5. Bajar el zip del repositorio que tienen dos carpetas

## Instrucciones

Dentro de la carpeta del repositorio hay dos carpetas

### Dspace Server

Los archivos de DSpace Server tienen la funcionalidad de cosechar todo el contenido del repositorio y en caso poder llevar un historial de cambio en los ITESM. Usan los archivos siteaudit.log y sitebackup.log para realizar sus tareas.  
Para usarlos, deben cambiar las rutas en el archivo SH y .PY donde está instalado el repositorio y donde quieren que se alojen siteaudit.log y sitebackup.log para que el script en Python los pueda leer y procesar.

El archivo para iniciar la cosecha del repositorio es RTSdspace.sh, este archivo se puede configurar como tarea periódica usando crontab.

Nota: Recordar que se debe tener Python 3.6 instalado y permiso de ejecución del archivo sh. Todo es por el usuario ROOT.

### Automation-tools-Archidspace-master

Esta carpeta tiene tres carpetas internas:

1. Carpeta automation-tools
   1. Git del proyecto con las instrucciones de instalación y configuración:
      1. github.com/artefactual/automation-tools
   2. De forma general la configuración de esta carpeta es de la siguiente manera:
      1. Instalación y creación de carpetas para funcionamiento
         1. <https://github.com/artefactual/automation-tools#installation>
      2. Creación de flujo de trabajo automático en la interfaz archivematica y posterior copiado del archivo defaultProcessingMCP.xml a la carpeta automation-tools/transfers/pre-transfer/ con la regla para evitar cuarentena para transferencias de DSpace.
         1. <https://github.com/artefactual/automation-tools#setting-processing-rules>
         2. <https://github.com/artefactual/automation-tools#tips-for-ingesting-dspace-exports>
      3. Configuracion de los Transfer-source-helpers (abajo).
      4. Pruebas de funcionamiento con el script DSpaceArchi.sh
2. MaintenceScripts: Dentro de esta carpeta están scripts de mantenimiento de archivematica
   1. Delete-registries.sh – se usa para limpiar el Dashboard de todas las transferencias y depósitos terminados correctamente. Se usa en caso de que se cosechen muchos ítems que ocasionen una tardanza en la carga de la página principal.
   2. RestartServices.sh – se usa para reiniciar los servicios de Archivematica, desde la base de datos hasta el servidor web ngix
   3. StopServices.sh - Para todos los servicios de archivematica.
3. Transfer-source-helpers
   1. Colocar los archivos dspace-transfer-src\* en la carpeta /opt/archivematica/transfer-source-helpers
   2. Colocar el archivo dspace-transfer-src.conf en /etc/archivematica/automation-tools/
      1. Este archivo de debe modificar los campos:
         1. Rsync\_password - Password del usuario Rsync de la maquina DSpace
         2. Rsync\_path – User@host::path/del/DirectrorioRTS de la maquina DSpace
         3. transfe r\_source\_dir – Directorio donde se iniciará el depósito de la transferencia que haga el script es el path Transfer Source del Archivematica Storage Service
   3. Git de la documentación oficial:
      1. https://github.com/artefactual-labs/transfer-source-helpers/

Y dos archivos:

1. DSpaceArchi.sh – Es el archivo iniciador de la cosecha desde Dspace a Archivematica. No hace falta modificar mucho, solo revisar si la línea 10 apunta al directorio correcto donde se encuentra su archivo **main.py.**
2. main.py – El archivo main es el encargado de hacer todo el trabajo de cosecha hacia archivematica. En ese archivo se deben modificar los siguientes campos:
   1. --user (Usuario del DashBoard)
   2. --api-key (Key del usuario de DashBoard)
   3. --ss-user (Usuario del SS)
   4. --ss-api-key (Key del usuario del SS)
   5. --transfer-source (UUI de la carpeta transfer source, esta se obtiene en el Storage Service)
   6. --am-url http://localhost (Url del Dashboard)
   7. --ss-url <http://localhost:8000> (URL del Storage Service)