

# MODS Example Script 1.2

## Your first Harvest

Since MODS is a widely known data format, much of the work has been already done for your harvest. No translation file needs to be created, nor does the workflow need created as the harvester team has completed these steps for you. However, you will need to do the some configuration so the harvest knows where your vivo data is and the records you wish to ingest, and what type of data you are converting to MODS.

- Change directory to example-scripts/example-mods
- If there is not one already, create a subdirectory called input
  - If there is, clear it out.
  - Place the file(s) you wish to import into the input directory
- Edit the runbibutils.conf.xml file
  - Set the inputFormat parameter to match the type of data you are using as input
  - For more information on these parameters and their use, please see [RunBibutils](#)
- Edit the vivo.model.xml file
  - Set the dbURL, dbUser, dbPass, and Namespace
  - For more information on these parameters and their use, please see [Harvester vivo configuration file](#)
- Edit changenamespace-author.conf.xml, changenamespace-authorship.conf.xml, changenamespace-datetime.conf.xml, changenamespace-geo.conf.xml, changenamespace-hyperlink.conf.xml, changenamespace-interval.conf.xml, changenamespacejournal.conf.xml, changenamespace-org.conf.xml, and changenamespace-pub.conf.xml files and set the namespace parameters in each one to be your vivo namespace
  - For more information on these parameters and their use, please see [ChangeNamespace](#)
- Edit the run-mods.sh file and set the HARVESTER\_INSTALL\_DIR= to be the directory you unpacked the harvester in
- Run bash run-mods.sh
- Restart tomcat and apache2. You may also need to force the index to rebuild to see the new data. The index can be rebuilt by issuing the following URL in a browser: <http://your.vivo.address/vivo/SearchIndex>. This will require site admin permission, and prompt you to login if your not already.

## The first run

Three folders will be created

- logs
- data
- previous-harvest

The logs folder contains the log from the run, the data folder contains the data from each run, and the previous-harvest folder contains the old harvest data for use during the update process at the end of the script. While you're testing, I would recommend treating each run as the first run (so no update logic will occur). You can do this by removing the previous-harvest folder before running again.

Inside the data folder, you will find the raw records utilized during the ingest. To see what rdf statements went into VIVO, you can view the vivo-additions.rdf.xml file. Conversely, to view what the harvester removed (because of updated data), you can view the vivo-subtractions.rdf.xml file. This file will be blank on your first run, since you have no previous harvest to compare the incoming data against.