

# cryoEM Data Acquisition Workflow

This page describes the processes and scripts that provide the transfer of image and meta data from the TEM's.

EPU and SerialEM are installed and running on the TEM machines. Each TEM has the following machines

- temN-control: actual server connected to the TEMs
- temN-k2: server connected to the K2 cameras where data will reside
- temN-support: machine used for remote access from FEL.

Internally to each tem, there is a private network where all machines are interconnected. Across these servers, an 'X:' drive is mounted where data collection is stored.

Operators are expected to utilise the elogbook to control the backend data management systems. This primarily functions as a means to

- Copy and remove (old) data from the local tem servers
- Organise the data onto the large disk subsystems in the data center
- Begin pre-processing pipelines to align and ctf the images

In the data center, an [apache airflow](#) instance runs to manage the workflows required to support data management. It currently runs as a [docker swarm](#) instance on cryoem-daq[01-5] nodes.

On these cryoem-daq nodes, the K2 server's disk is mounted via CIFS. the cryoem-daq nodes also mount the large GPFS filesystem where the data ultimately resides and users can access. As the docker swarm instances are containers, these mountpoints are 'bind' mounted into the airflow instances.

The airflow stack is kept under revision control under [github](#).

Airflow define workflows in DAGs. These are coded in python and provide dependency graphs between tasks. The following table describes the function of each DAG.

DAG	Purpose			
temN-daq.py	Literally monitors the elogbook for the current experiment and sets up the storage and preprocessing pipelines in preparation. Also copies the data from the tem servers to the GPFS file system - and as it goes so, triggers the appropriate new pre-processing task to kick off.			
<experiment name>_<sample id>.py	these DAGs get generated with every new experiment and contain the actual preprocessing pipeline to align, ctf and particle pick each and every image that is triggered from the temN-daq DAG.			
pipeline_single-particle_pre-processing.py	default template DAG for single particle pre-processing. This file is copied to <experiment name>_<sample id>.py when a new experiment starts			