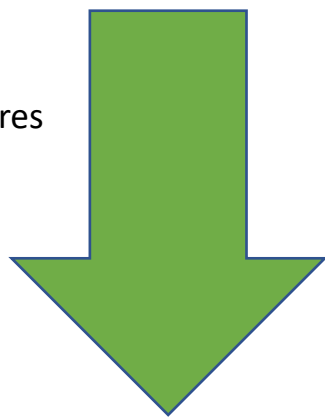


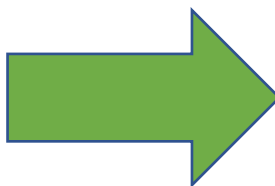
**Initial library
construction**
~165k structures
(notebook 0)



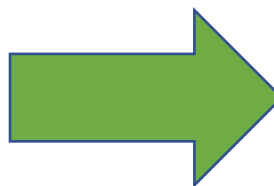
Library updates
~100-200 new
structures / week
(notebook 1)



Local File Library
120 GB



**Parse Files into
DataFrame**
(notebooks 0/1)
163,507 rows



Clean DataFrame
(notebook 2)
Standardization of entries
Removal of some rows
~15k rows dropped

Pandas DataFrame
Shape = 149,772 x 9

Data
Quality
Predictions



Linear Regression ML
(notebook 3)
RidgeCV Model
Score = 0.003