Loading functionality:

Load configurations and parameters and verify their syntax.

Get the min frequency parameter and ignore errors in signals with lower frequency.

Permit subsampling of patients for fast debugging.

Check that all patient ids are numeric. If not: build an Id2NR dictionary in the specific configuration.

Accept patient and signal lists to include or exclude.

Map signal names according to a dictionary from configuration.

Check that all signal names are recognized. Unrecognized signals should be reported in descending frequency order.

Check that all time values and units match the signal type (numeric categorial etc…)

Check that units that do not match signal have a conversion factor to the desired unit. Convert them to the desired unit.

Check that all categorial signal values are recognized and mapped to number in the configuration.

Check there are no inputs before patient birth and after death (warning).

Check there are no inputs outside membership (warning)

Check that every patient has exactly one bdate and gender (we may make a list of signals that are required and should be unique to each patient. For example race or membership)

Present histogram of each signal, and statistics (average, quartiles etc) with some acceptable GUI.

Compare distribution and statistics to same signal from reference repository (optional)

Generate updated configuration

Generate detailed log along processing data (errors, warnings, transformations, count found, hyperparameters, etc.)