Pacific Northwest
NATIONAL LABORATORY

CoreMS Framework Overview

July 15th, 2020

Yuri E. Corilo

Allison Thompson, Will Kew, Lisa Bramer,

Qian Zhao, Aditi Sengupta, Rosalie Chu,

Lee Ann McCue

U.S. DEPARTMENT OF ENERGY BATTELLE

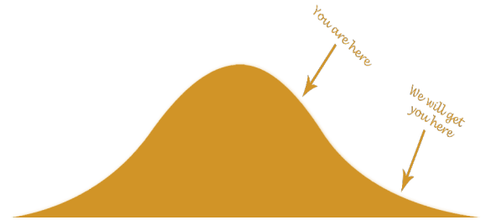
PNNL is operated by Battelle for the U.S. Department of Energy

CoreMS Framework Overview

- **Modern, flexible, stable and sustainable software framework and platform** for mass spectra data-processing workflows, including hyphened methods, i.e., LC-MS, IM-MS targeting small molecules analysis of complex mixtures

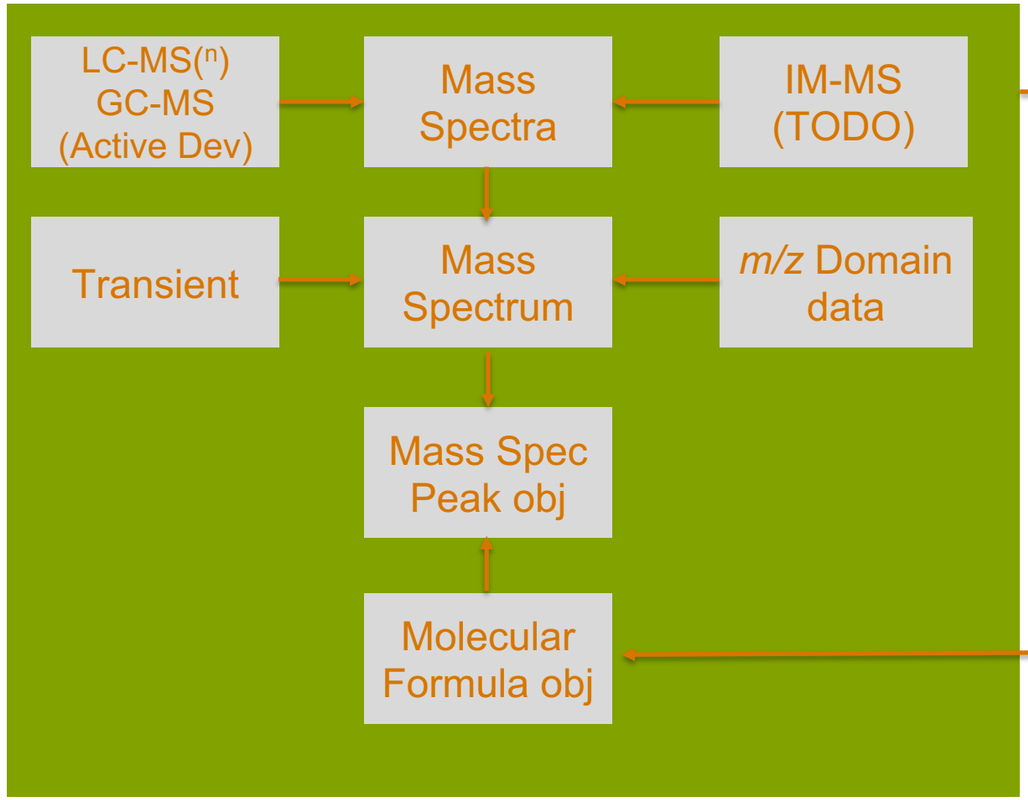
- Comprehensive **molecular formula assignment algorithm** merging state of the art approaches and universal mass spectrometry features

- **Robust confidence metric for formula assignment** enables analysis validation, comparison of algorithms results and ensures analytical reproducibility

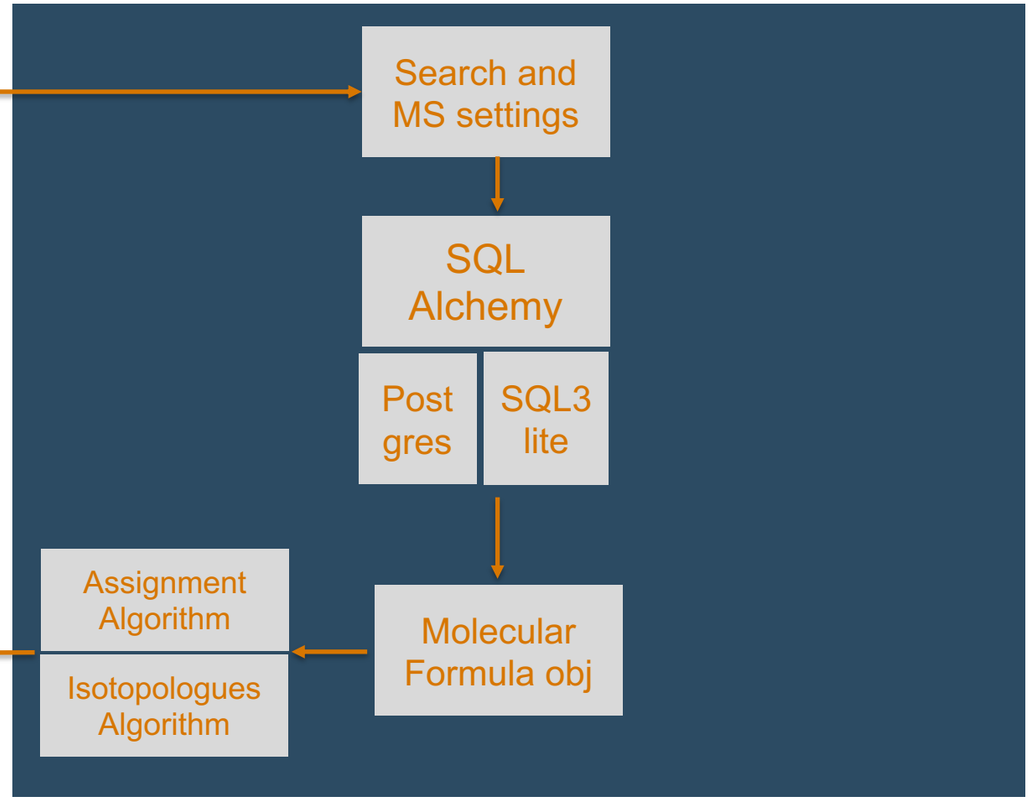


Software Framework

CoreMS Hierarchical Data Structure



CoreMS Molecular Formula Assignment Workflow



For available input, output and data structure types please refer to:
https://github.com/EMSL-Computing/CoreMS/blob/master/README.md

Software Framework

CoreMS Basic Molecular Formula Assignment Script

Input

Signal processing

Molecular Identification

Data analysis and visualization

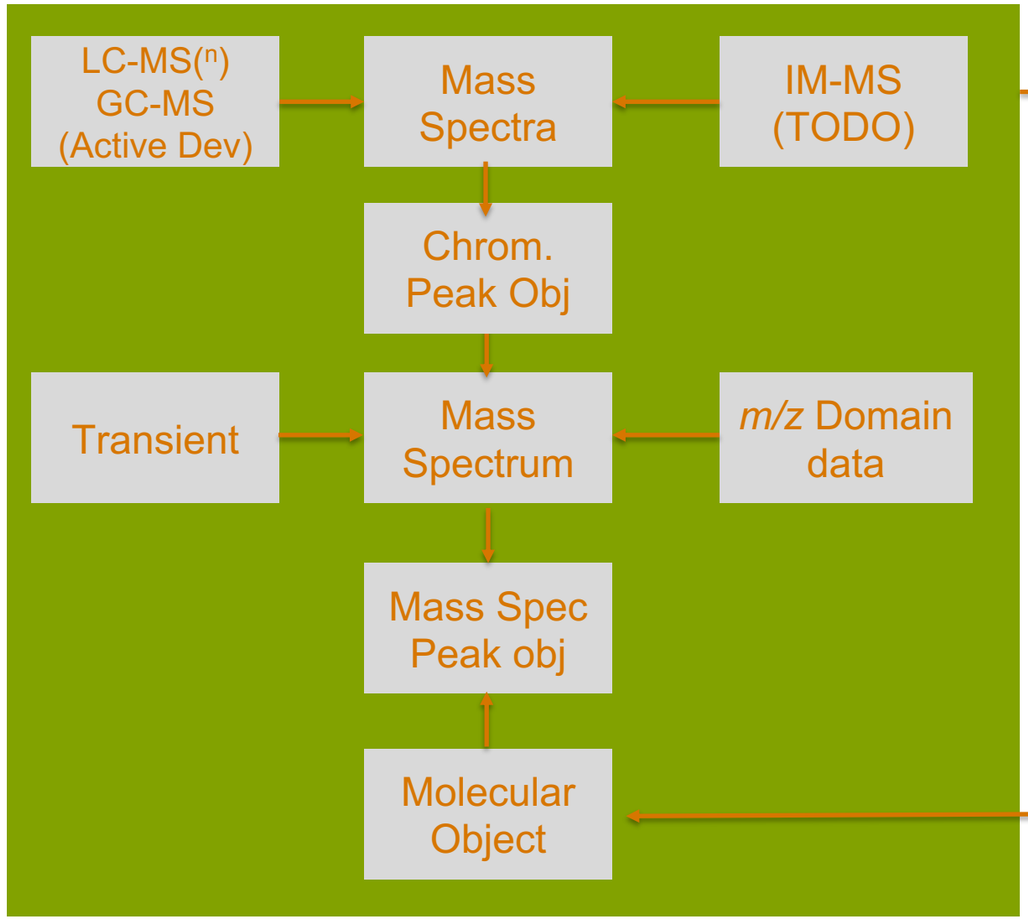
Output

```
file_name = 'neg_esi_srfa_1ppm_test.d'

bruker_reader = ReadBrukerSolarix(file_location)

bruker_transient_obj = bruker_reader.get_transient()

T =  bruker_transient_obj.transient_time

mass_spectrum_obj = bruker_transient.get_mass_spectrum(plot_result=False, auto_process=True)

SearchMolecularFormulas(first_hit=False).run_worker_mass_spectrum(mass_spectrum_obj)

for mspeak in mass_spectrum_obj.sort_by_abundance():

      if  mspeak:
            molecular_formula = mspeak.molecular_formula_lowest_error
            pyplot.plot(mspeak.mz_exp, mspeak.abundance, 'o', c='g')
            for molecular_formula in mspeak:
                  if molecular_formula.is_isotopologue:
                     print (molecular_formula.to_string())
                     print (molecular_formula['13C'])
       else:
             print(mspeak.mz_exp,mspeak.abundance)


mass_spectrum_obj.to_csv("output_file_name")
mass_spectrum_obj.to_excel("output_file_name")
mass_spectrum_obj.to_hdf("output_file_name")
df = mass_spectrum_obj.get_pandas()
```

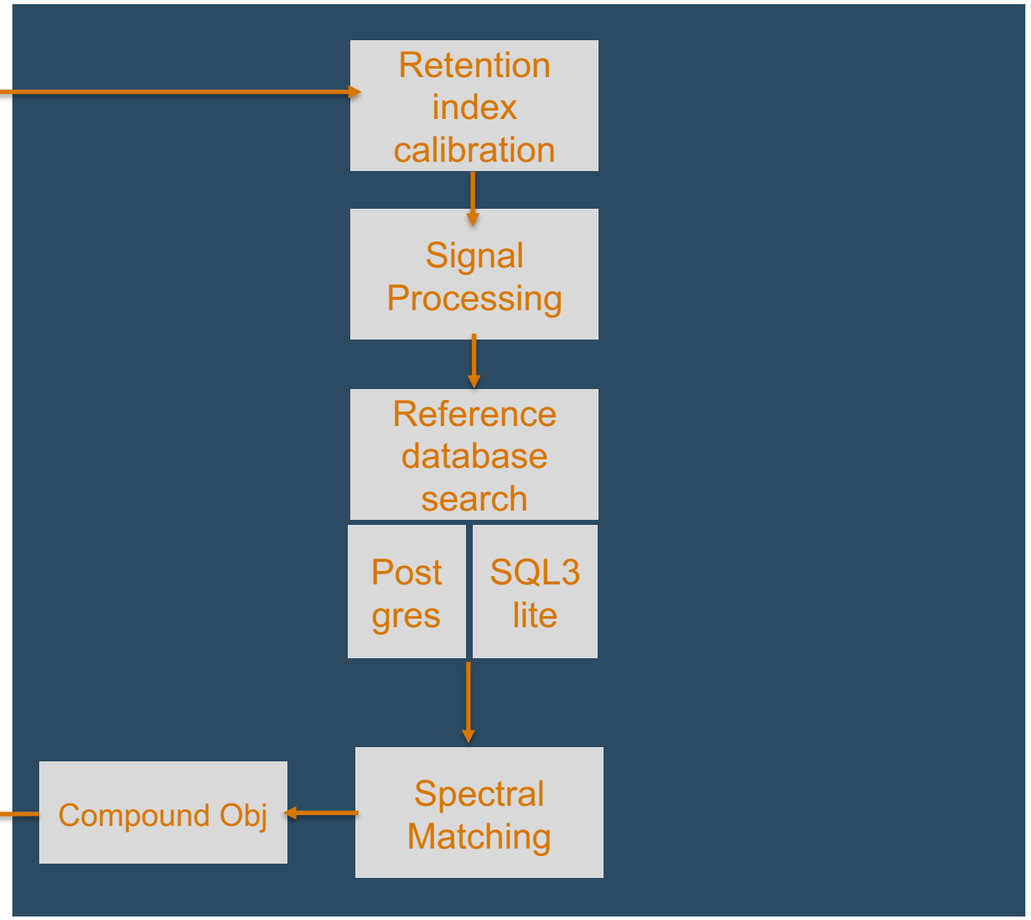
More examples available at:
https://github.com/EMSL-Computing/CoreMS/tree/master/doc/examples

Software Framework

CoreMS Hierarchical Data Structure



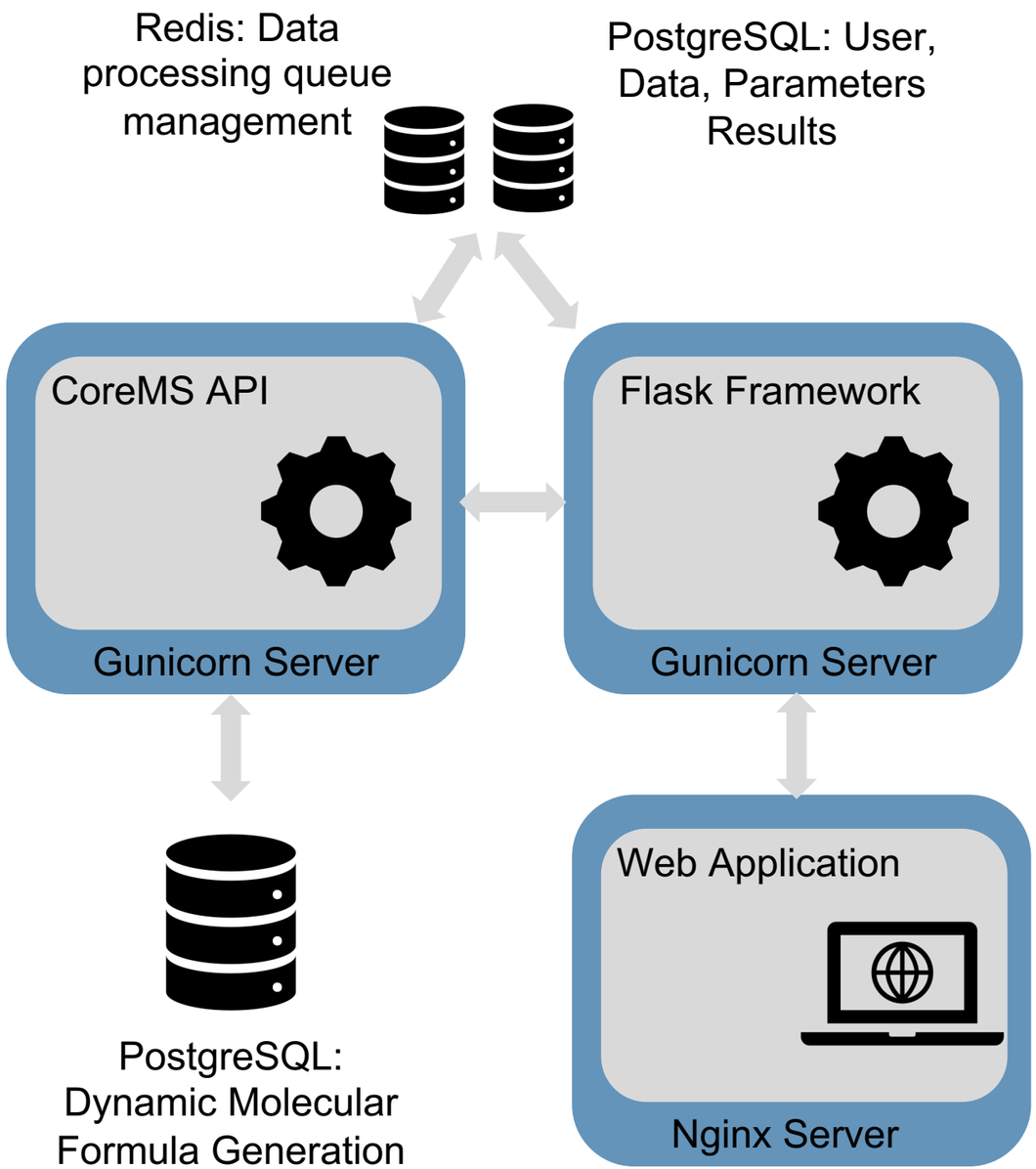
CoreMS GC-MS Workflow



For available input, output and data structure types please refer to:
https://github.com/EMSL-Computing/CoreMS/blob/master/README.md

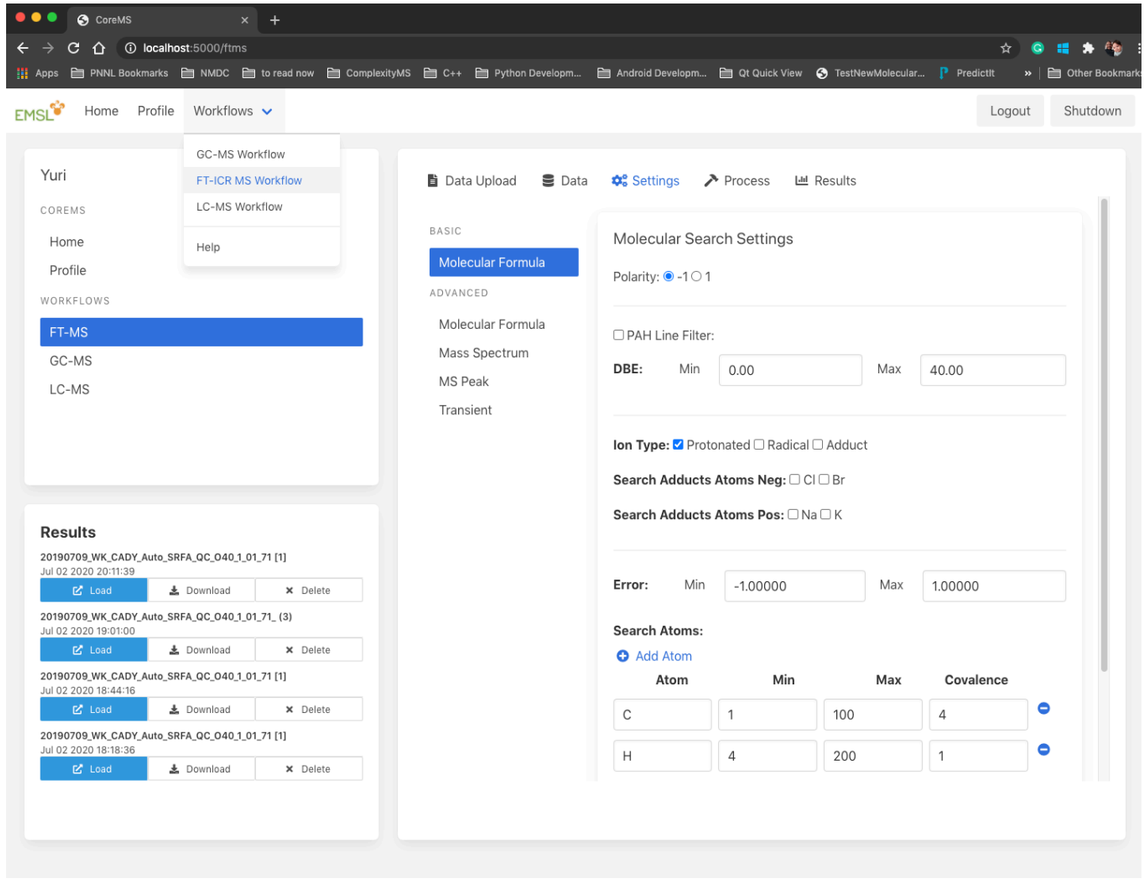
Web Application

- **G**raphical **U**ser **I**nterface (GUI):
  - Back-End:
    - User based temporary storage, and data model for raw, mass-list and reference file data
    - User based data model and storage for results and metadata
    - User based data model and storage of workflow parameters



Web Application

- **G**raphical **U**ser **I**nterface (GUI):
  - Front-End:
    - Web Application framework (HTML, CSS, JavaScript and, jQuery)
    - Data management for temporary raw data storage
    - Data management for workflow parameters, and parameters presets
    - Raw data visualization (Thermo, and Bruker)
    - Workflow result visualization



Pacific Northwest
NATIONAL LABORATORY

Thank you