Enabling Re-executable Workflows with Near-real-time Visualization, Provenance Capture and Advanced Querying for Mass Spectrometry Data

MATHEW THOMAS

Pacific Northwest National Laboratory

2016 New York Scientific Data Summit (NYSDS)
Topics

- Nano-DESI Mass Spectrometry experimental setup
  - How is data acquired?
  - Data structure and its complexities
- MSI QuickView Software Toolkit
  - Capabilities
  - Web App
- Workflow components that enable …
  - Re-execution and reproducibility of experiments
    - ProvEn
  - Collaborative querying of experimental data/information
    - ELK Stack (Elasticsearch, Logstash, Kibana)
Near Real Time Analysis of Mass Spectrometric Data during Acquisition

The optical image of a mouse brain

Acquire data for each Line (marked) of the optical image using the instrument

Raw File saved onto the computer

Data for all Lines is displayed in MSI QuickView Real-Time Visualization Tool

Line 1
Line 2
Line 3
Line 4
Line 5
Line 6
Line 7
Line 8
Line 9
Line 10
Line 11
Line 12
Line 13
Line 14
Data Structure

- Mass Spectrometric Imaging (MSI) generates spatial maps that can be used to visualize the location and quantity of molecules.
- Each point in space can have several thousand mass-to-charge (m/z) values.

Heat map (top) and associated spectra at 8 different points (bottom). Each location has m/z and abundance data associated with it. The sums between the m/z limits (red lines) correspond to the particular pixels values for that line on the heat map.
Classification is one of the fundamental methodologies for analyzing mass spectral data.

Major goals of classifying mass-spectrometry data include:
- Automatically group compounds based on their mass spectra.
- To determine correlation between properties of compounds and their mass spectra.

6 different classes from a single dataset are shown as an example.
The Workflow
MSI QuickView – Provenance Capture

MSI QuickView Folder Structure within a dataset

Sample snippet from ProvEn text file

Session 1 – USER 1

Session 2 – USER 2

Proudly Operated by Battelle Since 1945
What does the provenance message look like?
Re-executing Workflows
The Workflow
MSI QuickView Web App
Summary

How is this helping …

- MSI QuickView
  - Near-real time analysis
  - Visualization and Analysis on normal laptops

- ProvEn
  - Experimental reproducibility
  - Re-executable workflows

- ELK Stack
  - Openly available
  - Breaks down complex information
  - Tailored information representation
  - Querying capabilities
Future Work

- Extend MSI QuickView to multi-modal datasets
  - Liquid extraction surface analysis (LESA)
  - Electrochemical Microscopy
  - Fluorescence Microscopy
  - Hyperspectral Microscopy

- Port current desktop application capabilities to the web application

- Automate metadata capture workflow for new datasets into the ELK stack

- Investigate the use of ProvEn Platform as a message database
Team

**MSI QuickView**
Mathew Thomas
Nhuy Van

**MSI Imaging**
Julia Laskin
Son Nguyen

**Provenance**
Todd Elsethagen
Bibi Raju
Eric Stephan

**Acknowledgement**
- Chemical Imaging Initiative – PNNL LDRD
- LungMAP - Grant No. NIH/NHLBI U01HL122703
- NIH R21 – NIEHS Grant No. 1R21ES024229-01
- Integrated End-to-End Performance Prediction and Diagnosis for Extreme Scientific Workflows (IPPD)