REXAN – A Rapid Experimental Analysis Framework

MATHEW THOMAS, KERSTIN KLEESE VAN DAM
Content

- Motivation and background of REXAN
- Usage Example of the REXAN framework
- Conclusion
The goal of PNNL’s Chemical Imaging Initiative is to move from post experimental analysis to *in situ* manipulation of samples at the molecular level.

This is to be achieved with a combination of molecular-scale multi-modal imaging tools and concomitant data handling and analysis methods, providing the level of control enabling rational design and synthesis of new chemical, biological, and materials systems.
Overall Goals For REXAN

- Moving towards real time analysis
- Ability to deal with large data volumes (TBs)
- Enable multi-modal analysis
- Provide open framework
Pipelines are made up of recurring components

REXAN is creating a library of reusable, highly optimized versions of those components.

Community Repository to use and share
REXAN - Analysis Component Categories

- Appropriate lossless and lossy compression algorithms
- High compression ratio with low computational overhead

- Reduce noise and smooth data
- Reconstructions will contain the most significant information, are feature-accentuated

- Accurate re-construction of high volume data in real time
- Combine correlation functions with parallelized filtered back projection
REXAN - Analysis Component Categories

- **Segmentation & Feature Association**
  - Scalable differential operators and primitives that can be combined at runtime for application-specific chemical signature and feature recognition

- **Visualization & Analysis**
  - Enabling accurate localized comparisons between experimental datasets from different chemical imaging techniques at high resolutions

- **Registration**
  - Real time, remote, in-situ, high data volume 2,3+4D visualization
PNNL Deployment Framework for REXAN

- **Velo** – collaborative data management and analysis infrastructure – links data, tools, workflows, compute and storage resources for seamless usage.

- **MeDICI** - Flexible creation of data intensive workflows.

- Managing complex and intensive data exchange as well as rapid integration of data sets spanning different spatial and temporal scales.
PNNL Deployment Framework for REXAN

- **Semantic Framework**

- **REXAN repository** - Formal characterizations of the methods, instruments, analysis processes and associated data products and their connection with each other

- **Visko** – Detection of possible workflows from semantically described analytical and visualization components, ability to compare and execute any of those pipelines
Use Case 1

REXAN Deployment – Near Real Time Analysis of High-Resolution Mass Spectrometry Data
Background

• **Mass Spectrometric Imaging (MSI)** generates spatial maps that can be used to visualize the location and quantity of molecules.

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.
Previous Method

Limitations of Previous Analysis Methods

- Basic analysis 8-10 hours for 40 min.
- Could not view more than 100KB at a time – experiment produces 40GB+
- Many different tools and manual steps
- Limited functionality
- No instant feedback available during data acquisition

LTQ/ Orbitrap scans line then Xcalibur saves line

More lines?

Yes
No

Splice lines into smaller files using Xcalibur, due to Firefly limitations

Load files into Firefly and enter setup parameters ($m/z$; time values)

Open image in BioMap and determine $m/z$ values of interest

Load images into image editor and stitch segments together

Crop background out and fix aspect ratio, then save image

Image obtained using BioMap
Now: Near Real Time Analysis of Mass Spectrometry Data during Acquisition

**MSI QuickView – Analysis in seconds**

Integrated analysis environment, immediate analysis of each scan, utilizes HPC and server side visualization for statistical analysis of results:

- Interim analysis to assist in validating the usefulness of the particular experiment
- Unlimited data volumes
- 3D visualization of multilayer sectioning data
- Integrated data management
Near Real Time Analysis of Mass Spectrometric Data during Acquisition

The optical image of a rat brain

Acquire data for Line 34 (marked) of the optical image using the instrument

Data for all 34 Lines is displayed in MSI QuickView

Raw File saved onto the computer
**MSI Quickview – Immediate Analysis**

- **Analyze and Visualize data *during* experiment**
  - Abundance vs. m/z spectrum
  - Abundance vs. time spectrums
  - Line Scans across tissue samples
  - Heat maps (user-selected m/z value)
  - Image processing algorithms

---

Display several abundance vs. m/z spectrums using user set values

The intensity spectrum for a line scan across the sample (left). Corresponding pixels values are displayed (right).

Displays intensity vs. time & intensity vs. m/z spectrum for user-selected region.
MSI Quickview – Imaging Options

Original Image – m/z 820.5

- Normalized
- Aligned
- Rows/Lines Removed
- User set scaled limits
- Smoothed
- User set intensity specific colormap
Interactive visualization of extreme scale data, uses client/server + parallel processing model

Arbitrary processing of data set with filters such as subsetting, contouring, clipping.

Views can be saved and reused.

Animation provided for useful stepping through m/z slices.

Publication quality views savable to multiple formats.

3D View in Paraview
The data transformation approach for data reduction and classification is demonstrated. ‘L’ and ‘S’ denote ‘lines’ and ‘scans’ respectively. 3D data is reduced to 2D, followed by the optional application of an external software called Decon Tools to further reduce the dimensionality. Principal component analysis (PCA) and Linear Discrimination Analysis (LDA) can then be applied to the data.
• Uses MATLAB Mobile app
  • Single line commands for the user
  • Execute on remote desktop
  • Results displayed on hand held device
  • Works with IPad, Iphone, Ipod and Android.

• Offers data management from your mobile device

An example command line entry from MSI QuickView
REXAN Components Used

- Data Reduction
- Decon Tools (PNNL)
- Data Classification
- Data transformation
- PCA
- Clustering

- Normalization
- Alignment
- Scaling Limits
- Smooth Image
- Remove Lines
- Overlay Images
- Personal Colormap
- Line Scan
- Filters....
- Matrix Manipulation
- Heat Maps
Behind the Scene of MSI Quickview

User Environment

Compute Cloud

Processing Pipelines

MeDICi

Control Workflows

Vis Tools

Xcalibur™

MSI Quick View

Shared High Performance File System

Reusable Components

Shared High Performance File System

VELO Knowledge Management System

Provenance Metadata

Metadata

Instruments

Import

SSH Tunnel

Client Remote API
Velo’s – Knowledge Capture, Management and Sharing Approach

- Registry of tools, instruments, components, data sets, etc.
- Collaborative project management and shared workspaces
- Execution environment for tools, workflows and models
- Seamless workflow integration with provenance capture
- Graph-based knowledge storage
- Semantic-powered search and discovery
Use Case 2

Chemical Imaging of Microbial Biofilms (X-Ray Tomography Data)
Previous Analysis of X-ray Tomography Data

Image source
- Scanning transmission x-ray microscopy (STXM) images
- Electron Microscopy (EM) images

Limitations of previous analysis methods
- Need to down-sample data to 32-bit limits
- Could visualize only 2D slices
- Manual segmentation of features
- Analysis took days of labor-intensive work

Main Concerns
- Low laptop memory (travelling scientists)
  - Datasets over 30 GB

2D slice of an OS Biofilm
New Toolkit: Visualization of X-ray Tomography Data in Minutes
3D Viewer: Feature Visualization in Minutes

- Quick view of the dataset
  - Downsample (x, y or z direction)

- Utilize functions of multiple software
  - Automatic loading and data transfer
    - 3D Viewer (MATLAB)
    - ImageJ (uses MIJ plugin)
    - Paraview
    - DDV

Comparison of 2D and 3D Images at different resolutions
3d Viewer: Feature Visualization in Minutes

- Subset selection to load data at full resolution
- Play/pause/scroll through 3D volume or slices of data
- Isosurface rendering
- Full resolution movie of slices for selected ROI in a sample
- Input/output data in multiple formats
3D Viewer: Segmentation Routines

**Segmentation Routines**

- Identify channels in biofilms

  ![OS GelBond](image)
  ![The channels through the biofilm separated](image)

- Color code the image based on thickness

  ![SXT Data](image)
  ![Segmentation based on volume of objects](image)
REXAN Components Used

- Data Reduction
- Normalization
- Filters…
- Scaling Limits
- Smooth Image
- Segment
- Isosurface
- 2D/3D Slicer

- Save Image/Video
- Image Processing…

Visualization
- Paraview
- MATLAB
- Image J
- DDV
Behind the Scenes of 3D Viewer

- Transfer of data between MATLAB, ImageJ, Paraview and DDV from within a single interface

- Format conversion
  - tiff, hdf, st, rec, jpeg, gif, png, bmp

Visualization Tool ‘Visus’ is also embedded within the framework
Use Case 3

Analysis of X-Ray Plant Root CT Data
New Analysis Methods for X-ray CT Data

- Imaging using EMSL CT
- 16 GB dataset
  - 2000x2000x2000
  - 31 micron resolution

- Limitation of Previous Analysis Methods
  - Data difficult to handle, visualize, and analyze
  - No detailed analysis method available

Virtual cross section of soil with root fragments
New Analysis Methods for X-ray CT Data

New Analytics Tool:
Rapidly connected modular image and volume processing elements
  - Subset data to select region of interest
  - Background normalization to enhance signal
  - Interactive connected threshold
  - Generate isosurface of root
  - Smooth isosurface
  - Calculate local feature size using PNNL LaGrit
  - Visualize root surface in Paraview, colored by feature size

Components Used: Normalization, smooth, isosurface, PNNL LaGrit
Visualization: Paraview
Use Case 4

Statistic-based Reconstructions
Use Case – Statistic-based Reconstructions

Task

- Generate High Resolution Image of microstructure using
  - Low Resolution Optical image of large region
  - Electron Microscopy of small subset at high resolution
  - Statistical association of images

New Key Components

- AMSFSA: Adaptive Multiple Super Fast Simulated Annealing
  - Multithreaded solution optimization algorithm
Use Case 5
Multimodal Nano-scale Image Analysis
Multimodal Nano-Scale Image Analysis

Instrument used
- Scanning Transmission X-ray Microscopy (STXM)
- Transmission Electron Microscopy (TEM)

Tasks
- Overlay STXM data with fluorescence data
- Particle detection, measurement, localization
  - Titanium, ceria Particles (STXM)
  - ZnO particles (TEM)

Challenges in analyzing the data
- Correcting the alignment in the dataset
- Large and small vesicles are joined together
- Separate noise from the particles

Limitations of previous approach
- Image J does not detect all particles
- Volocity too aggressive at separating particles
A semi-automated algorithm that will give user multiple options to determine the best boundaries

Semi-automated Options

- Uses functionalities from MATLAB and Image J
- Simultaneous transfer of data between MATLAB and Image J
- Overlay boundaries
- Draw new boundaries
- Join broken boundaries
- ANDing, ORing or subtracting boundaries
- Save multiple images

The interface for the particle detection software
Summary
Project Flow Similarities

Data Collection
- High Resolution Mass Spectrometry
  - LTQ/Orbitrap
  - Xcalibur

X-Ray Tomography Data
- STXM
- X-Ray

Multimodal Nano-scale Analysis
- STXM
- Fluorescence

X-Ray CT Data
- CT

Graphical Feedback
- MSI QuickView
- Paraview
- 3D Viewer
- Paraview
- ImageJ
- DDV
- 2D Viewer
- ImageJ
- 3D Viewer
- Paraview
- Visus

Image Processing
- Normalization
- Alignment
- Scaling Limits
- Smooth
- Remove Lines
- Overlay Images
- Personal Colormap
- Line Scan
- Watershed
- Centroid Detection
- Particle Counter
- ImageJ Tools...
- Filters...
- Personal Colormap
- Boundary Detection
- Normalization
- Connected threshold
- Isosurface
- PNNL LaGrit
- Feature Size

Data Reduction
- PCA
- Clustering
- Decon Tools
- Downsampling
- None
- Downsampling
Results

- Component concept works – we are able to reuse developed components across different analytical pipelines
- Pipeline design accelerated – from months to days were extended reuse is possible
- Scalability – Components scale in terms of time to solution and data volumes
- Leverage – Ability to leverage HPC and optimization expertise through easy to understand single purpose components
- Future – Publish REXAN 1.0 next summer, seeking other contributors to component libraries
Planned REXAN Repository

- Component registration via semantic metadata
- Dynamic queries list the available analysis pipelines for a given data set/instrument
- Ability to download components to aid analysis pipeline development
- Ability to add components to enhance and extend existing pipelines or create new ones
- Ability to register and download full pipelines
- User service – automatic creation and execution of pipelines on user data
Acknowledgement

The Team

**Computational**
- Kerstin Kleese Van Dam
- James Carson
- Abigail Corrigan
- Daniel Einstein
- Zoe Guillen
- Senthil Kabilan
- Andrew Kuprat
- Carina Lansing
- Dongsheng Li
- Yan Liu
- Paulo Pinheiro da Silva
- Seun Ryu

**Microbial Biofilm**
- Matthew Marshall
- Erin Miller

**Nano-scale imaging**
- Galya Orr
- Craig Szymanski

**Mass Spec**
- Brandi Heath
- Ingela LaneKoff
- Julia Laskin

**Plant Root CT**
- Nancy Hess