

The NREL High Throughput Experiments for Materials (HTEM) Database A Prototype Project-Specific Analytics Database Enabling the Application of Machine Learning to Experimental Data

John Perkins, Andriy Zakutayev, Caleb Phillips, Jacob Hinkle, Robert White, Kristin Munch, Marcus Schwarting john.perkins@nrel.gov

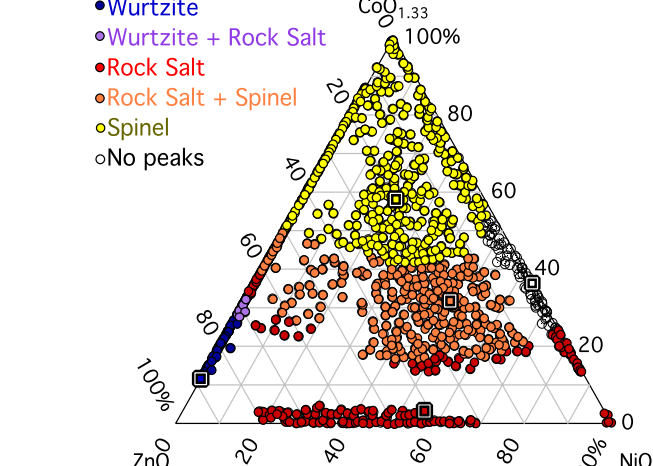
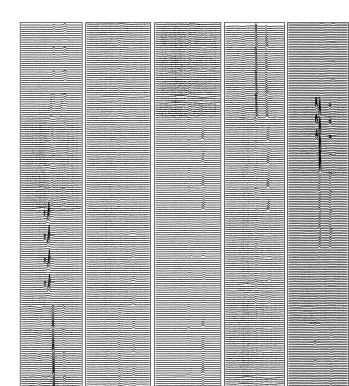
Overview

In the process of accelerating our analysis of large volumes of x-ray diffraction data, we have created a prototype data analysis ecosystem for creating project specific data analytics tools. We leverage existing NREL data harvesting capabilities and open source machine learning code libraries to create custom user-friendly analysis tools.

The Data Analysis Need: Rapid Analysis of XRD data

880 Unanalyzed XRD Patterns (Data)

Structure Phase Map (Knowledge)



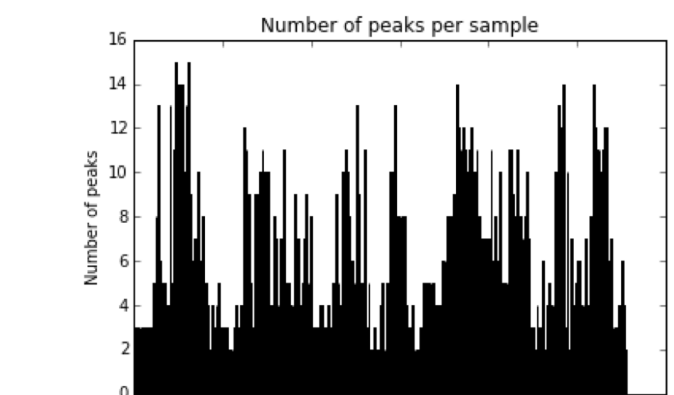
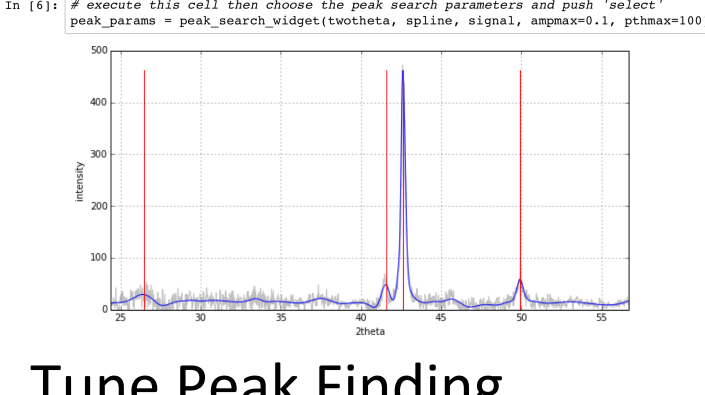
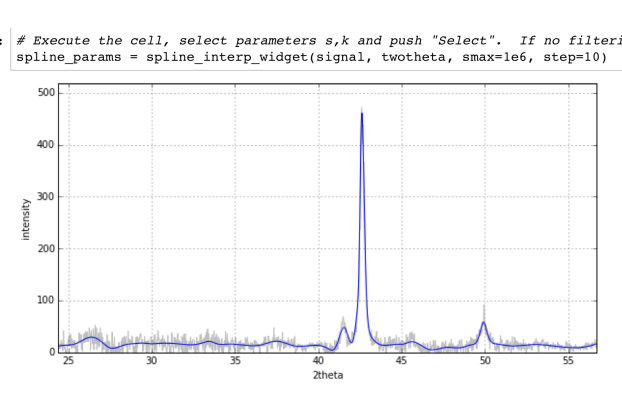
Pieces of the Solution

Peak Identification in Background Subtracted Data

1. Noise Removal

2. Find Peaks

3. Apply To Full Set

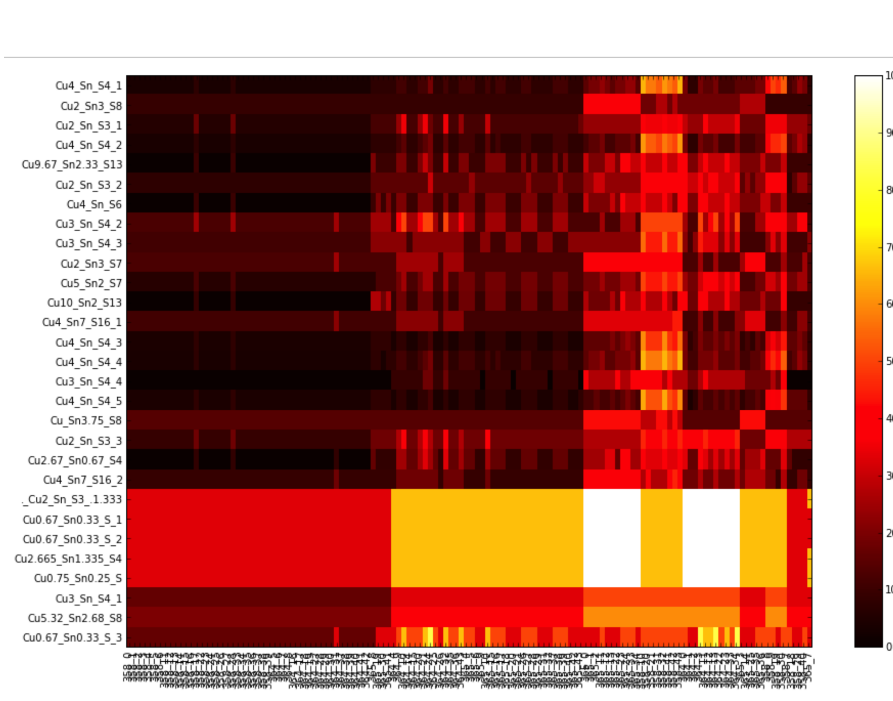
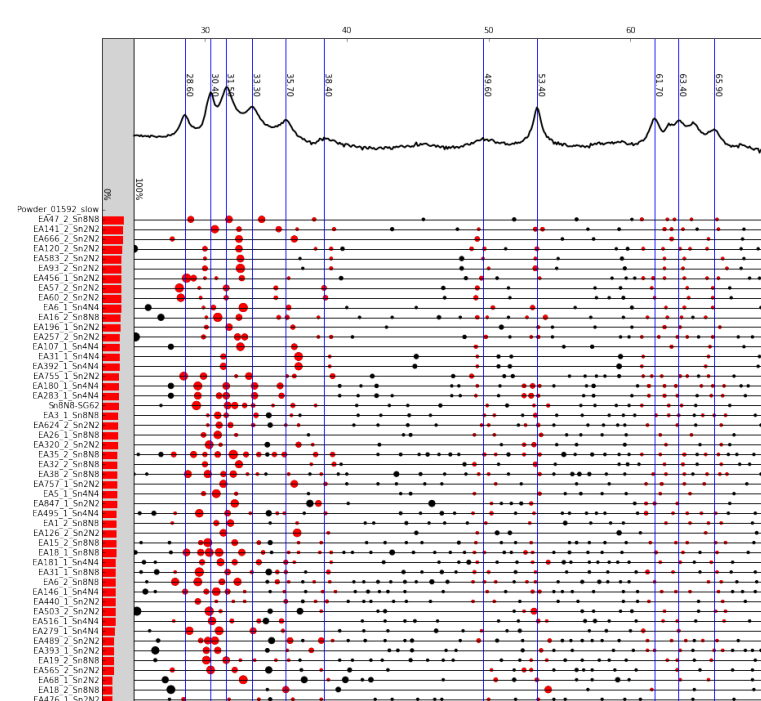


Tune Peak Finding
- continuous wavelet transformation

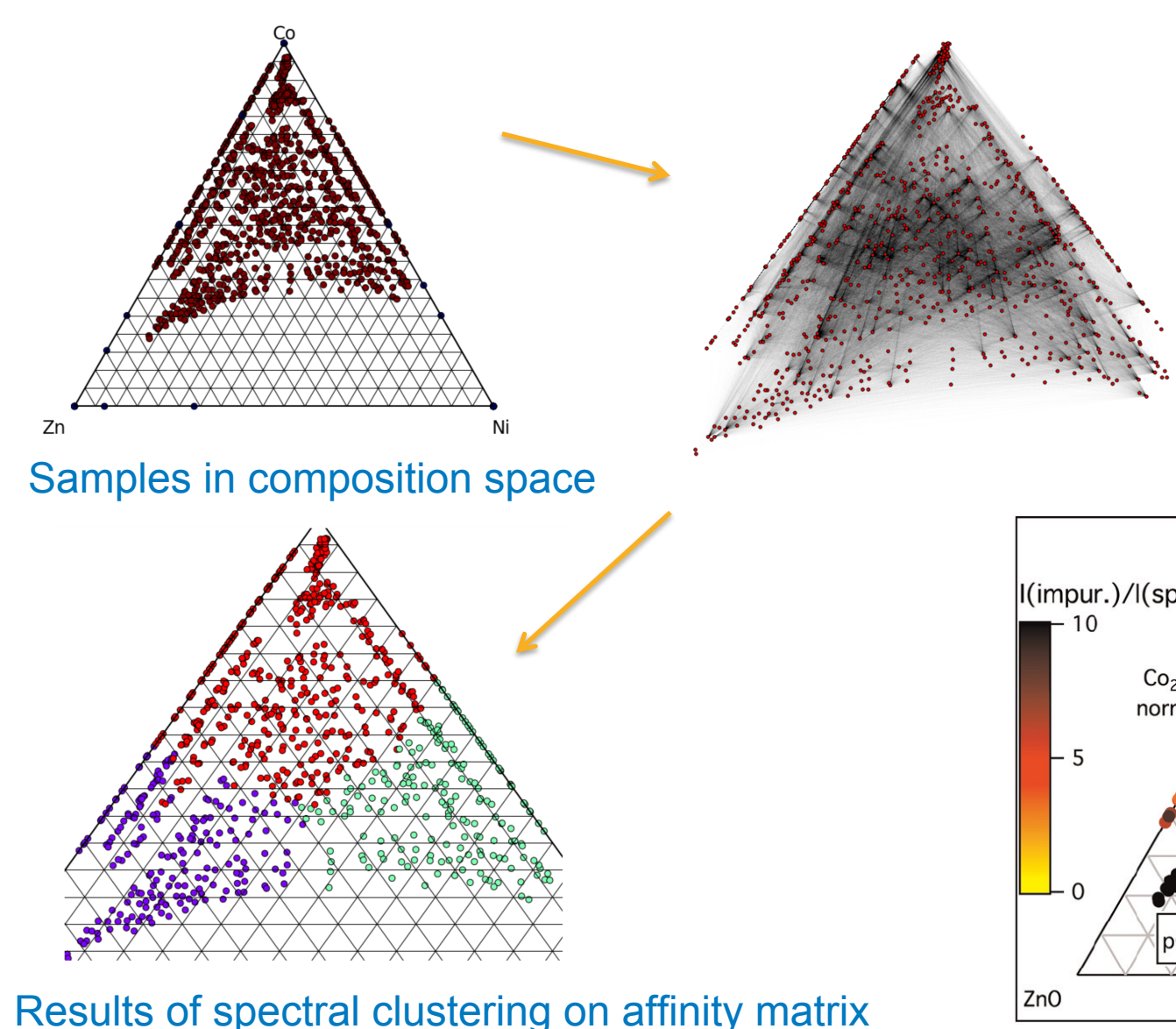
Peak Match Visualization

"Hit" Map
1 sample, many ref

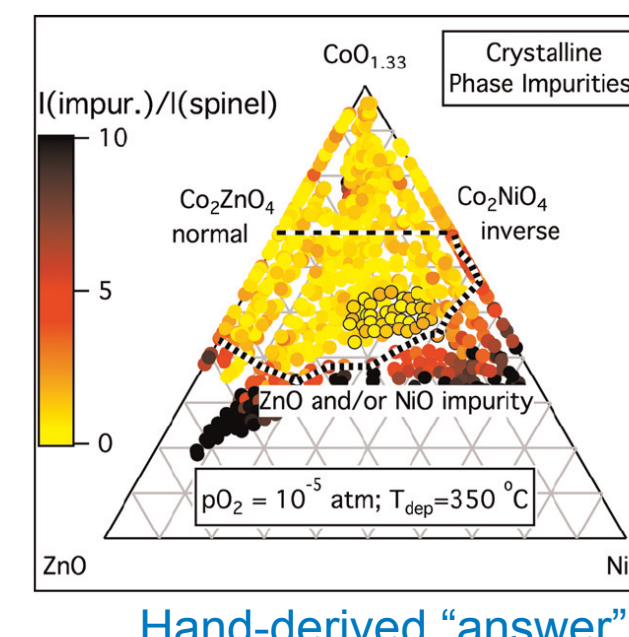
"Heat" Map
many sample, many ref



Spectral Clustering



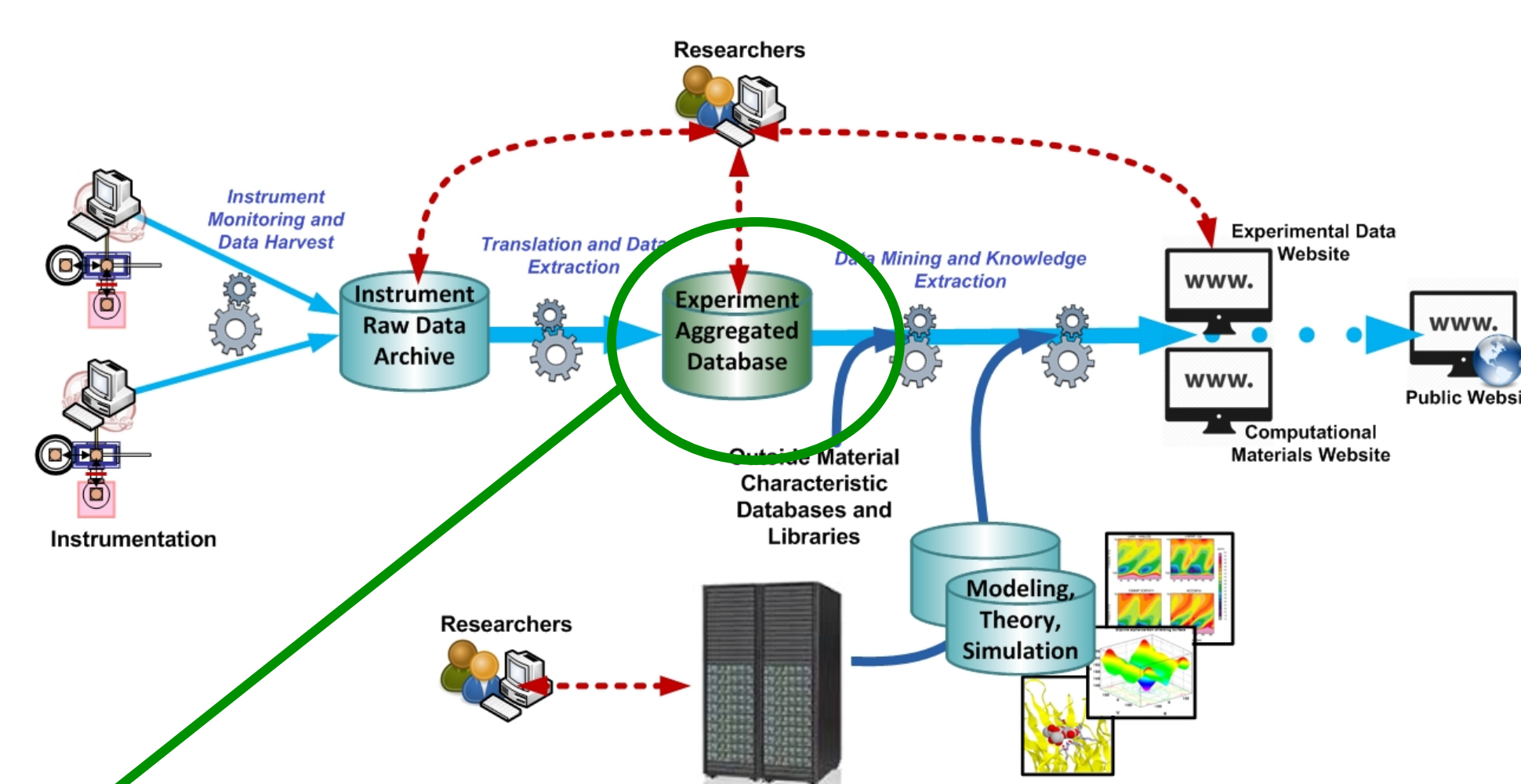
Samples as nodes in graph whose edges are weighted by "affinity"



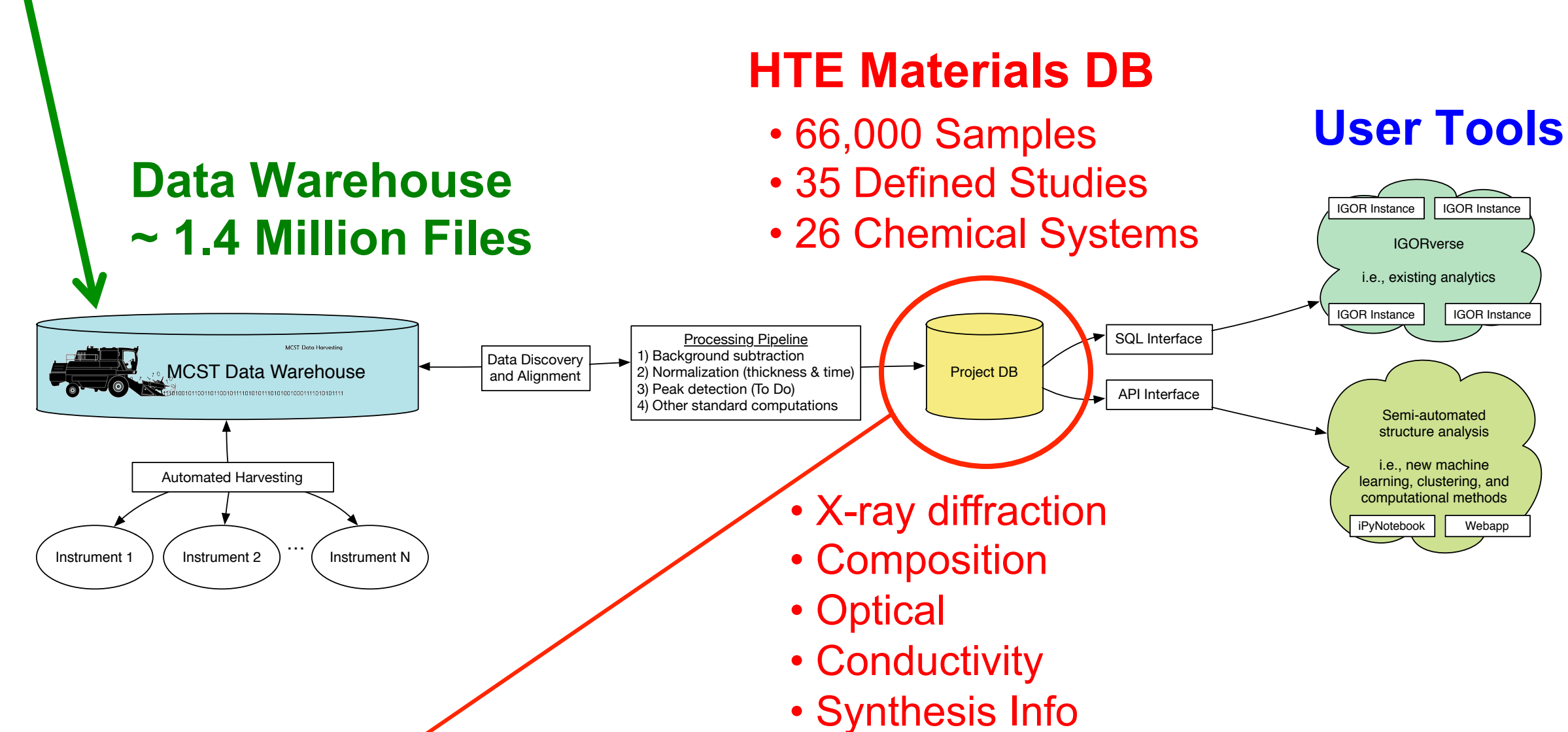
Hand-derived "answer"

Data Ecosystem for Data Analytics

NREL Data Harvesting System Architecture



Project Specific Analysis Databases

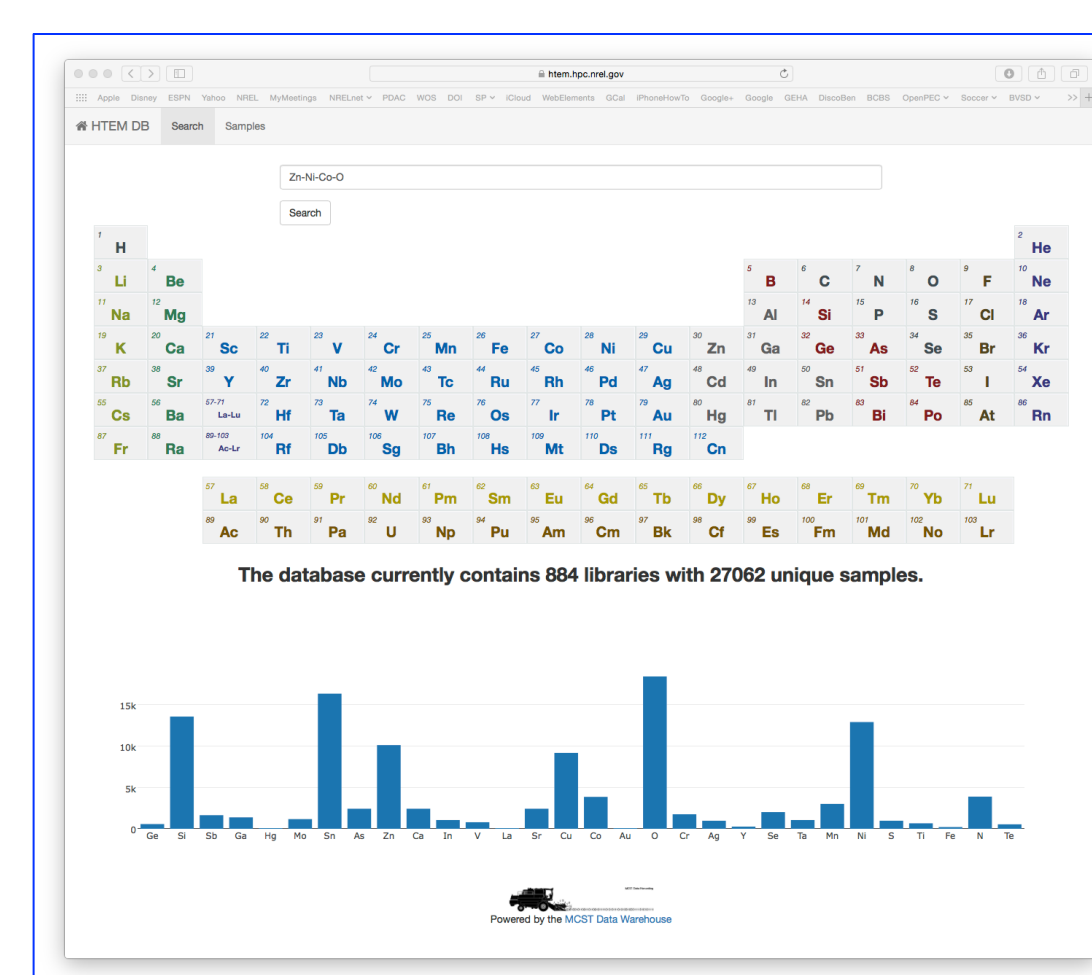


HTE Materials DB
• 66,000 Samples
• 35 Defined Studies
• 26 Chemical Systems

User Tools

- X-ray diffraction
- Composition
- Optical
- Conductivity
- Synthesis Info

Web Interface



SQL Interface

```
SELECT * FROM samples WHERE ...
```

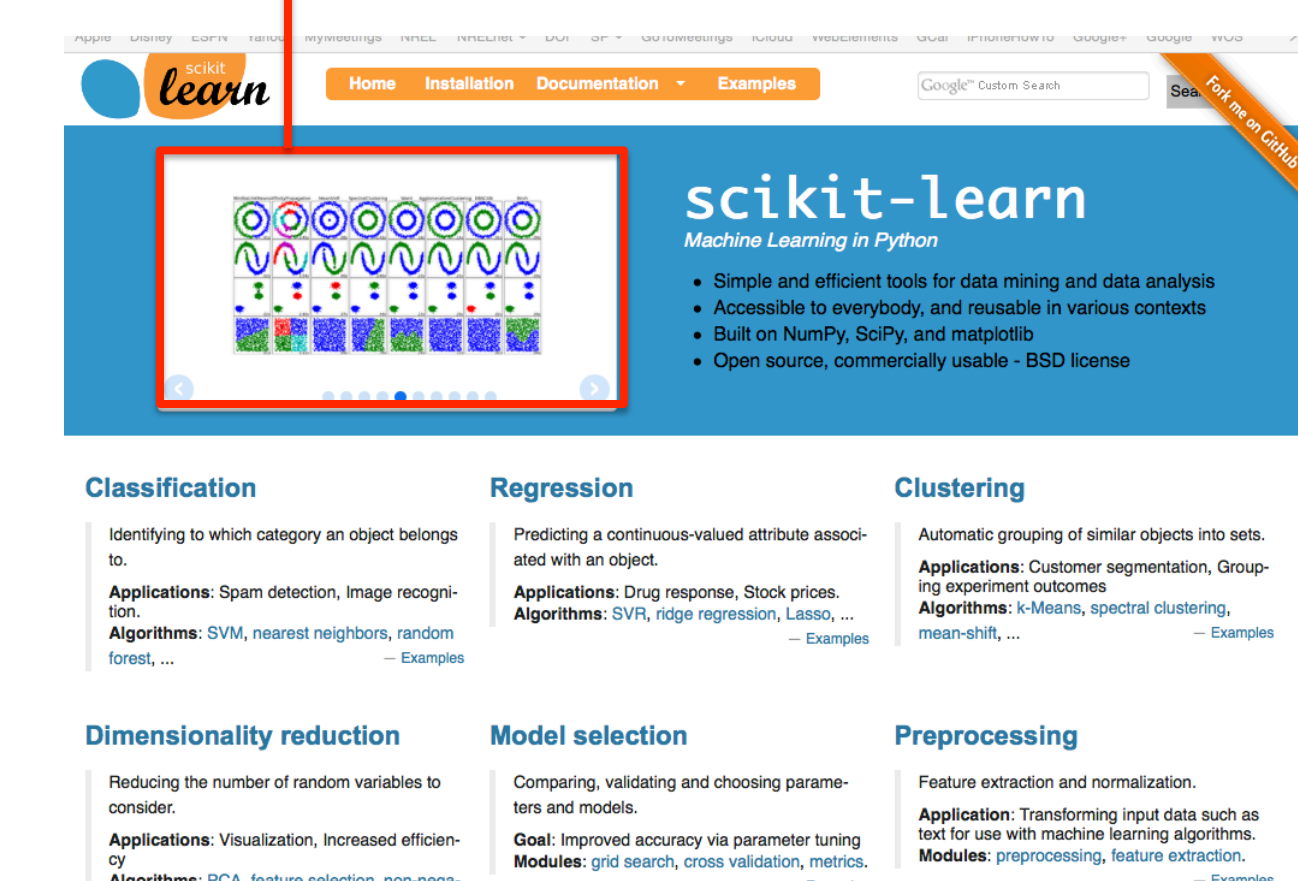
API Interface

```
xrd-api
Welcome to the NREL Materials Structure API
This is a programmatic front-end to the PDAC XRD database...
```

User Tool Development

Leverage Open Source Machine Learning Resources

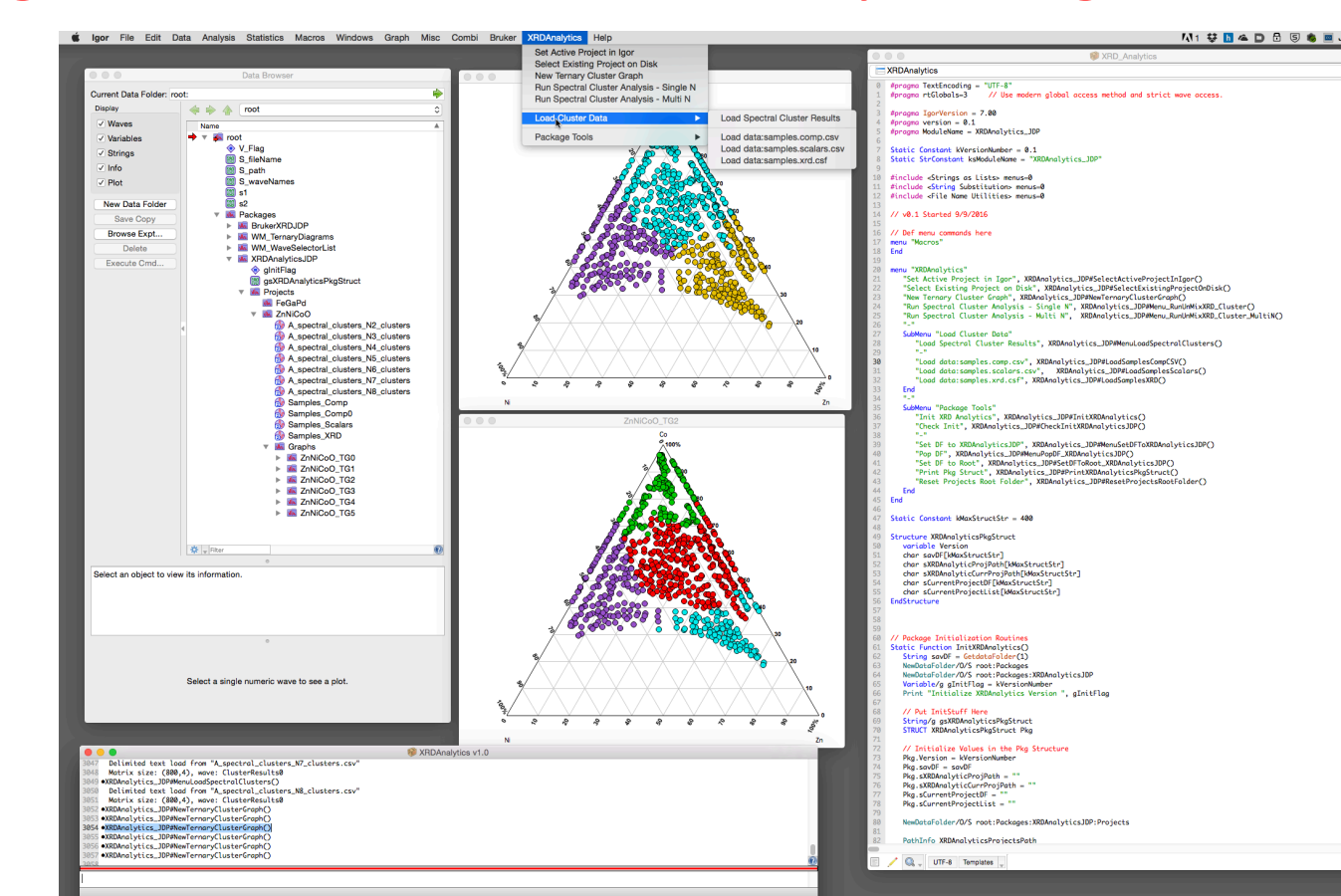
e.g. many clustering algorithms pre-coded



- Includes variety of machine learning tasks
- Always growing
- Free

Create Custom Python Package: *unmix_xrd*

- Custom python library & command-line program
 - use routines in scikit learn
 - can be run from Terminal or Python Notebook
 - single line commands run custom analysis
- e.g. `$ unmix_xrd cluster --method spectral --o ./out/G_ --n 4 ./data`
- Includes peak detection, clustering, sparse unmixing, visualization
- Pulls data from public & private databases resources
 - materialsproject.org, materials.nrel.gov, HTEM
- Extensible
- Can be called from commercial analysis programs
 - integrate into existing analysis workflow
 - easily create custom user analysis tools
- e.g. Menu Driven Cluster Analysis in Igor Pro



Implemented via simple system calls to *unmix_xrd* package

```
unmix_xrd_cluster_v4
function unmix_xrd_cluster_v4(project, slist, nClusters)
    string sProject
    variable nClusters
    string sFunctionName = "unmix_xrd_cluster_v4"
    string sPyWB, sProjectDirectory, sStdOut
    variable sStdIn
    sPyWB = IsFile(sFunctionName)
    // Move to the correct directory
    sProjectDirectory = sProjectDirectory + sProject
    sStdOut = "out" + sProjectDirectory
    sPyWB = AddME(sStdOut)
    // Run the function
    print(sStdOut, "unmix_xrd_cluster = " + sProjectDirectory + "/" + sFunctionName + ".py")
    sStdOut = "unmix_xrd_cluster = " + sProjectDirectory + "/" + sFunctionName + ".py"
    // Send it all to the system
    val = Run(sFunctionName(sPyWB))
end
```

Standard Initialization Section

Purpose Specific Call to unmix_xrd python package

Standard Termination and Run Section

The Take Away

- Custom Data Tools Can Accelerate Research
- Automated Processing, Visualizations, Easy Access, Machine Learning, ...
- Affordable by Leveraging Existing Prototypes and Open Source Resources