Nanoinformatics 2015

Enabling Successful Discovery and Applications

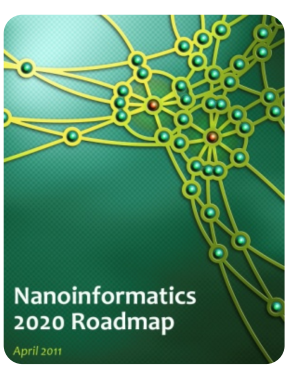
Tutorial Session

Welcome!

Nanoinformatics

A series of workshops by the broad nanoinformatics community

- 2007 Exploratory Nanoinformatics meeting
- 2009 Nanomanufacturing Summit Panel
- Nanoinformatics 2010
- Nanoinformatics 2020 Roadmap Published (April 2011)
- Nanoinformatics 2011: Q-SARs and MinChar
- Nanoinformatics 2012 Summer Workshop: Tools and Resources
- Nanoinformatics 2013: Informatics for Nanomanufacturing
- Nanoinformatics 2015: Enabling Successful Discovery & Applications



DATA – TOOLS - SHARING

Nanoinformatics

Tutorial:

A Broad Introduction to Science Informatics and Lab Automation Strategies

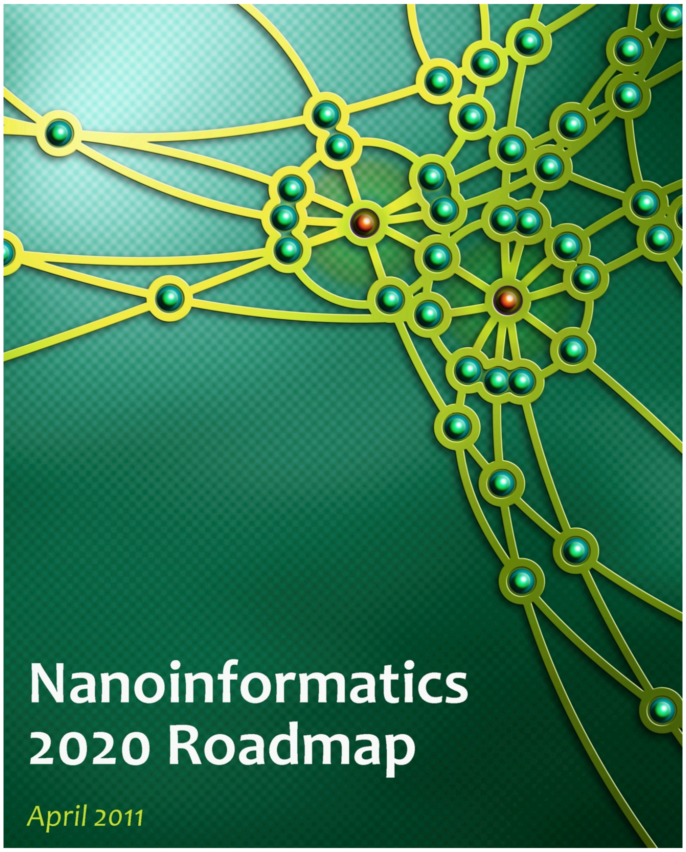
Mark Tuominen
Director National Nanomanufacturing Network
Professor of Physics
University of Massachusetts Amherst

Nanoinformatics

Nanoinformatics

Nanoinformatics is the science and practice of determining which information is relevant to the nanoscale science and engineering community, and then developing and implementing effective mechanisms for collecting, validating, storing, sharing, analyzing, modeling and applying that information.

- from ***Nanoinformatics 2020 Roadmap***



2020 Roadmap – An initial comprehensive overview and outlook on nanoinformatics

Contents
- Definition
- Vision
- Current activities
- Crosscutting issues
- Opportunities
- Future projects

http://eprints.internano.org/607/

Nanoinformatics

Purpose? For understanding and application

Physical Properties

Applications Development

Modeling and Simulation

Materials

Engineering

Manufacturing

EHS

Education

Biological Interactions

Business

It is important to recognize the different perspectives and agendas of diverse domains

Science Informatics

Information — Automatic

In the broadest sense, we try to—as much as currently possible—automate working with the information associated with science.

- To more broadly leverage the information already gathered.
- To reduce uncertainty and assist in decision-making
- To better automate processes that build science.
- To save time and money.
- To help assure safety for people and the environment.
- To build in greater intelligence into the entire system.

"In theory, there is no difference between theory and practice. But in practice, there is."

- Jan van de Snepscheut or Yogi Berra or perhaps someone else

Mark Tuominen – Jan. 26, 2015

Experimentation

Question — *Curiosity, need*

↓

Design Experiment — *An attempt to address question*

↓

Measurements — *Acquiring data*

↓

Analyze — *Find trends, correlations (reduction); building models or connecting to existing ones*

↓

Understanding and/or Application — *Sharing: service and value to society*

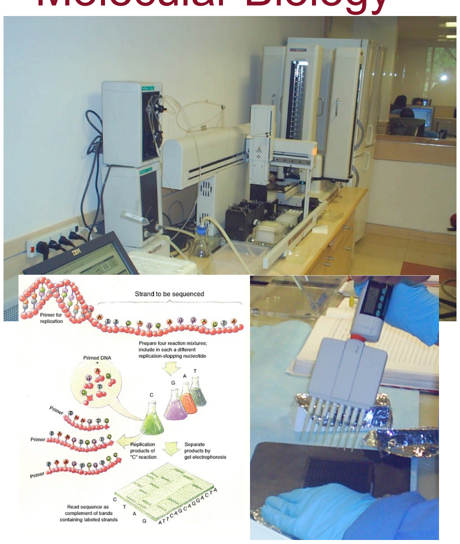
Once in a while things actually work this way; the real world can be more convoluted.

Mark Tuominen – Jan. 26, 2015

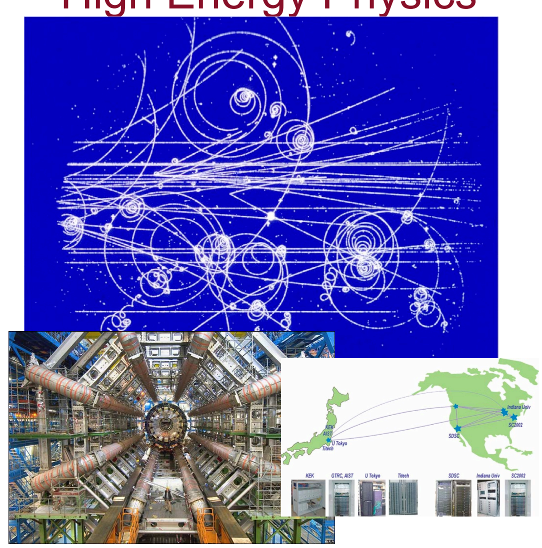
Other Science Informatics Examples

- Informatics is the study and application of information technologies to improve the efficiency and impact of scientific research

Molecular Biology



High Energy Physics



Each research area has its own distinctive characteristics which impact the design of science informatics tools to support it

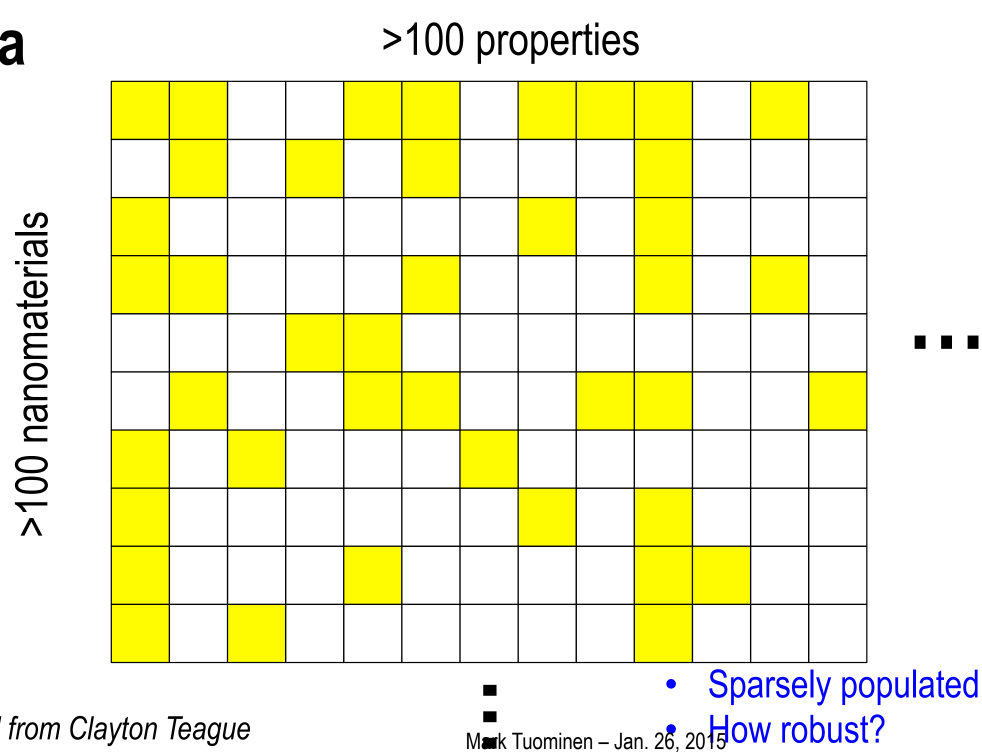
Nanoinformatics

- the science and practice of determining which information is relevant to the nanoscale science and engineering community, and then developing and implementing effective mechanisms for collecting, validating, storing, sharing, analyzing, modeling and applying that information.

Data—Tools—Sharing

Nanoinformatics

Data



- Sparsely populated
- How robust?

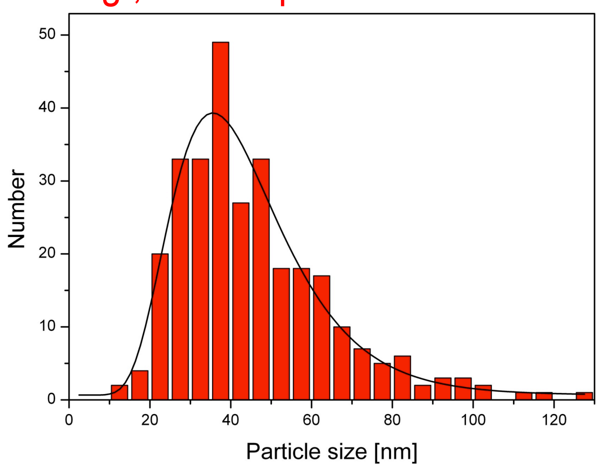
Adapted from Clayton Teague

Nanoinformatics

Statistical Robustness

Absolutely essential in many specific use cases, but not all.

e.g., Si nanoparticle size



- Full distribution is more informative than solely mean value
- Inter-laboratory studies (ILS) can reveal degree of reproducibility
- For some materials and some properties, relative fluctuations are greater at the nanoscale

Nanoscale Research Letters 2012, **7**:76 doi:10.1186/1556-276X-7-76

Nanoinformatics

Data & Sharing

Where is this data?

How reliable is it?

Nanoinformatics

Data & Sharing

Metadata

Context: the story behind the data, the important details
Essential for search and provenance

Data

The target of the study; answers to the research question (represented in numbers, if possible)

Nanoinformatics

Data & Measurement

Please keep this in mind:

Resist the temptation to artificially place greater emphasis on the aspects that are easily measured compared to those that are not.

(Remember the overarching purpose of the study. The aspects that matter still matter, regardless of whether you can measure them easily or not.)

Design of Experiment (DoE)

The experiment seeks to answer the question, to find out what matters.

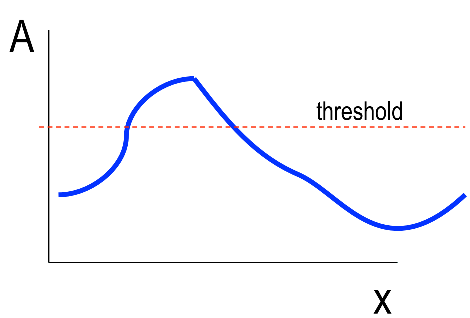
- Single variable – all other things held constant (!)
- Multiple variables (independent or interacting) – factorial or combinatorial design (e.g. optimization)

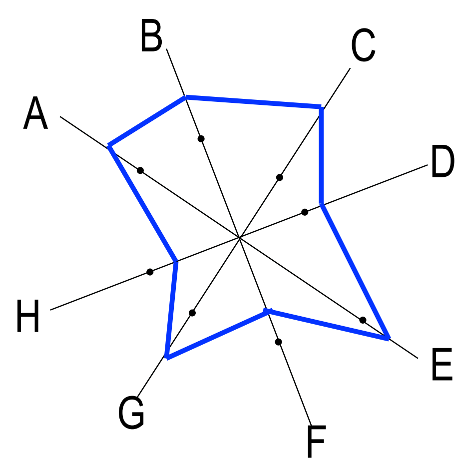
Allows exploration, discovery, and the building of predictive models (extensibility, leveraged value)

Models tested against real-world, experimental data are the most useful to society – the very definition of science

Optimization

- with respect to one or several concurrent properties

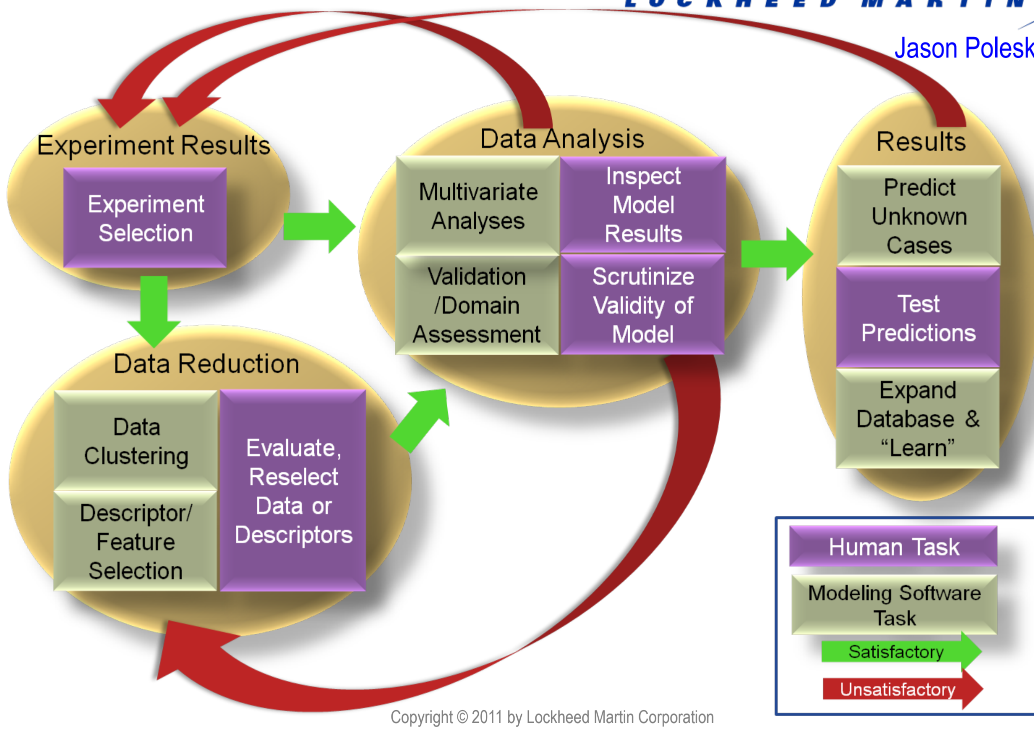




Example: Informatics Methods Applied to Process Development

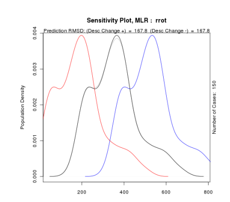
LOCKHEED MARTIN

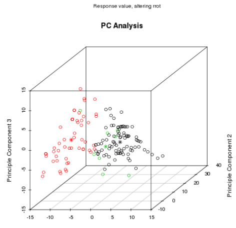
Jason Poleski, et al.

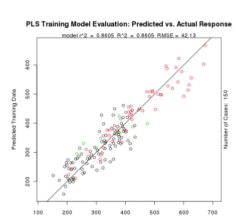


Informatics Tool Developed by Lockheed Martin

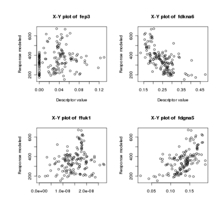
Jason Poleski, et al.

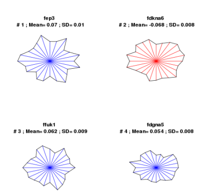


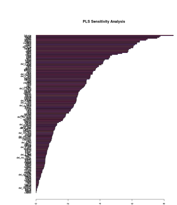




- The "Nanotechnology Material Data – Mining, Modeling and Management (NMD-M3)" Tool.
- Applications of the **NMD-M3 Tool**:
  - Analyzing trends in data sets, e.g. product performance (drug efficacy), treatment efficacy
  - Determining inter-measurement relationships and dependencies, medical data trends/analysis
  - Creating virtual systems in a matter of seconds
  - Comparing resulting system properties side-by-side
- Successfully demonstrated the benefits of the tool on various nanomaterial experiments
- Has over 10 analysis algorithms that run in series or in parallel to predict results based on input numerical data, the next set of experiments (configurations)
- Significant visualization techniques to provide the user with insights that are not clearly apparent







Saves time and money on development efforts by creating virtual configurations that focus future efforts more efficiently

Visualization and Visual Analytics

Aides our perception of trends, correlations, anomalies, etc.

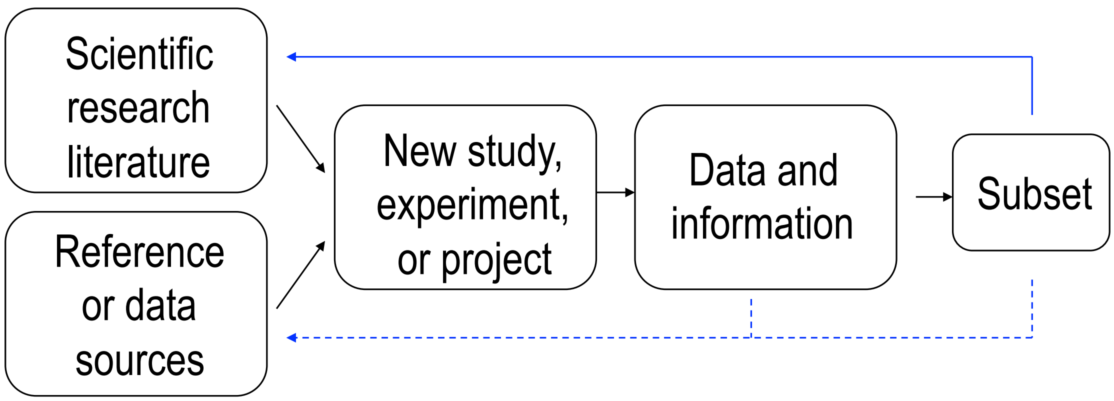
To reduce data to discover what matters; to address questions

Workflow and Tools

Machine People

There is great value to be gained in using and developing informatics tools that systematize, search, learn, and predict.

The Evolution of a Standard Model of Research and Dissemination

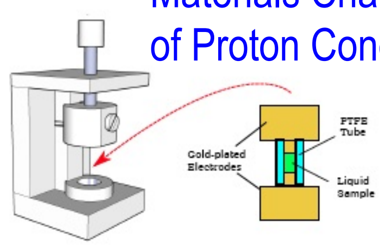


More facile and direct lab-to-dissemination approaches needed, while maintaining the value of peer review. Access to data.

Mark Tuominen – Jan. 26, 2015

A Lab Automation Example - Built on an Informatics Strategy

Materials Characterization of Proton Conductors



Impedance Spectroscopy

Charge transport properties can be probed with *impedance spectroscopy*:

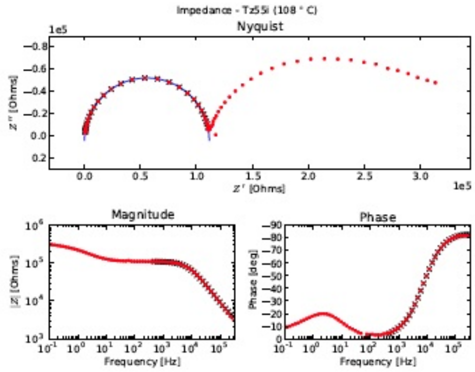
- place material in *blocking electrode* cell
- sweep small amplitude AC voltage over range of frequencies
- measured current response (magnitude and phase-shift) $\Longrightarrow$ *complex impedance function*

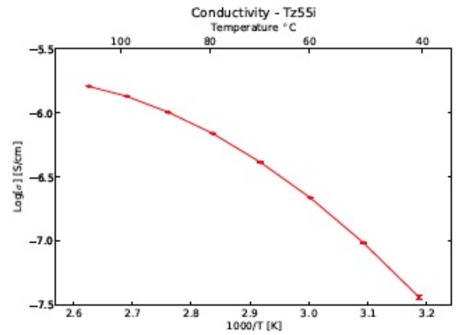
$$\tilde{Z}(\omega) \equiv \frac{V_0 e^{i(\omega t)}}{I_{0(\omega)} e^{i(\omega t - \phi(\omega))}} = \frac{V_0}{I_0(\omega)} e^{i\phi(\omega)} \equiv Z' + Z'' j$$

- analyze data with circuit models

Craig Versek, PhD thesis 2013, UMass Amherst

Material Characterization





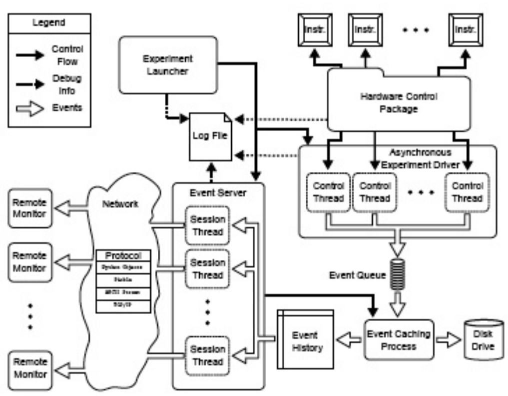
Conductivity Determination

At high frequencies samples' impedance spectra can be modeled as an $R \parallel C$ circuit

- fast current switching $\Longrightarrow$ no charge gradients
- bulk DC resistance in parallel to cell capacitance
- fit data with circle (or constant value in $|Z|$ plateau)
  - $R \approx Z'$ when $Z'' \approx 0$ (or in plateau $R \approx |Z|$)
  - $\Longrightarrow$ DC conductivity $\sigma = \frac{L}{A} \cdot \frac{1}{R}$

Craig Versek, PhD thesis 2013, UMass Amherst

Charge Transport Materials Characterization Platform



Automat - Experimental Automation and Informatics Framework/Toolkit

Using the open source Python scripting language we developed a pilot "middleware" framework

- serves as foundation and reusable code-base for more specialized user applications
- provides application programming interfaces (APIs), but not user interfaces
- simplifies the challenges of multiple asynchronous device control

Craig Versek, PhD thesis 2013, UMass Amherst

Charge Transport Materials Characterization Platform

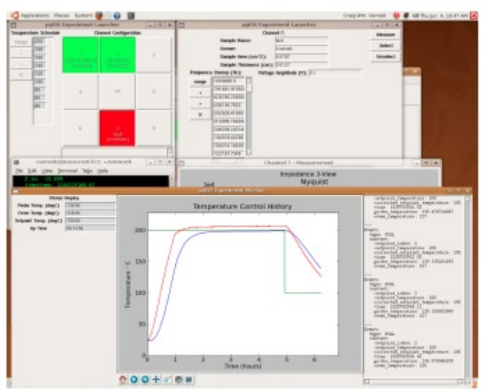
pyEIS - Informatics Tools for Impedance Spectroscopy

Building on top of the *Automat* framework, we developed a custom suite of software applications to give enhanced informatics capabilities to a new customized equipment setup

- developed in Python on Ubuntu Linux, but portable to other platforms
- modular software and hardware design philosophy
- 8 multiplexed sample channels, temperature control, vacuum or humidified air environment
- impedance and other electrical measurements

Craig Versek, PhD thesis 2013, UMass Amherst

Charge Transport Materials Characterization Platform



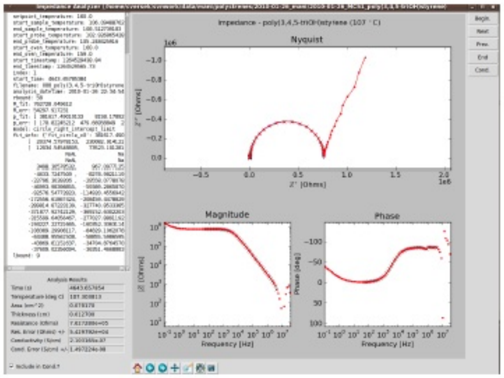
pyEIS - Experiment Launcher

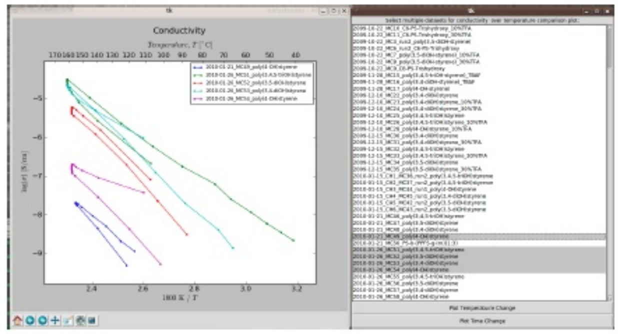
User application for configuring and running experiments

- simple graphical user interface
- extensive metadata collection and preservation
- remote monitoring of the experiment state
- on-the-fly data file caching into organized file structures
- data exported in *Zplot* compatible format

Craig Versek, PhD thesis 2013, UMass Amherst

Charge Transport Materials Characterization Platform





pyEIS - Impedance/Conductivity Analysis Tools

User applications for rapidly analyzing large volumes of data

- quick, interactive curve fitting to reduce data by the batch
- rapid plotting and comparison of samples, using data-discovery tools
- data can be exported to spreadsheet formats

Craig Versek, PhD thesis 2013, UMass Amherst

Nanoinformatics

Science Informatics

Data—Tools—Sharing

Service and value to society

National Nanomanufacturing Network

NSF Grant CMMI-1025020

CHM