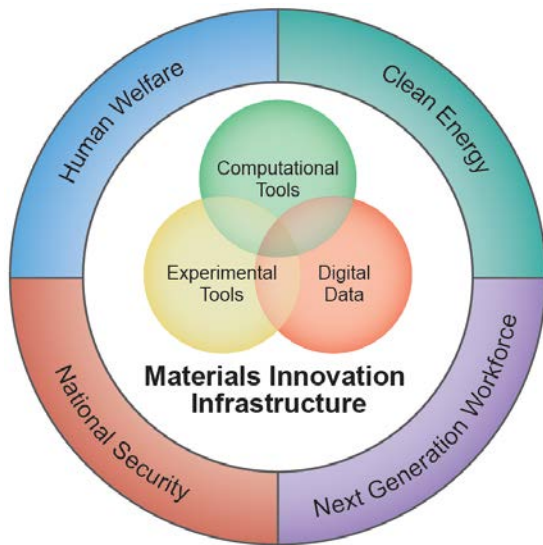


# Materials Genome Initiative & Artificial Intelligence @ NIST



A. Gilad Kusne [aaron.kusne@nist.gov](mailto:aaron.kusne@nist.gov)

James Warren, NIST MGI Director

National Institute of Standards & Technology

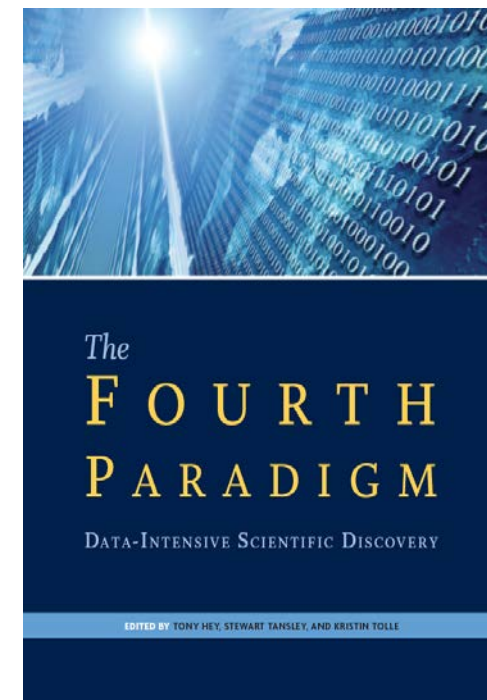
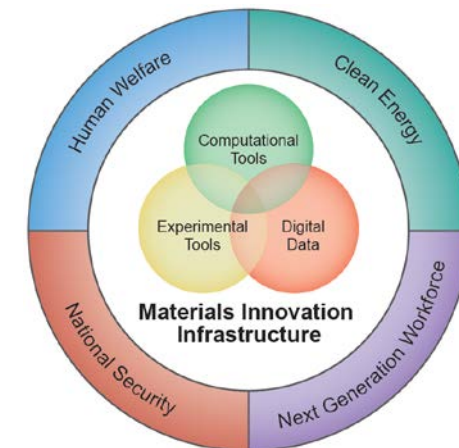
**NIST**  
**National Institute of  
Standards and Technology**  
U.S. Department of Commerce

# MGI + AI?

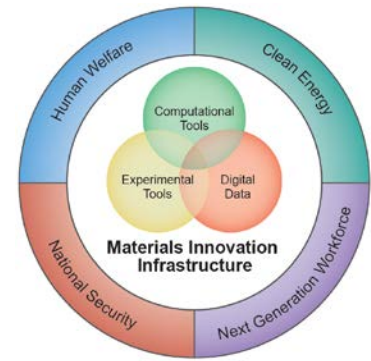
## What is MGI?

- MGI Goal – Accelerate materials development from Discovery to Deployment
- MGI – Key part: Build data science infrastructure to Enable Data Science & in particular AI
- MGI Outcome – Realizing the 4<sup>th</sup> Paradigm: Data-Intensive Scientific Discovery & Development

Make it easier for industry & academia to do materials data science.



# MGI + NIST

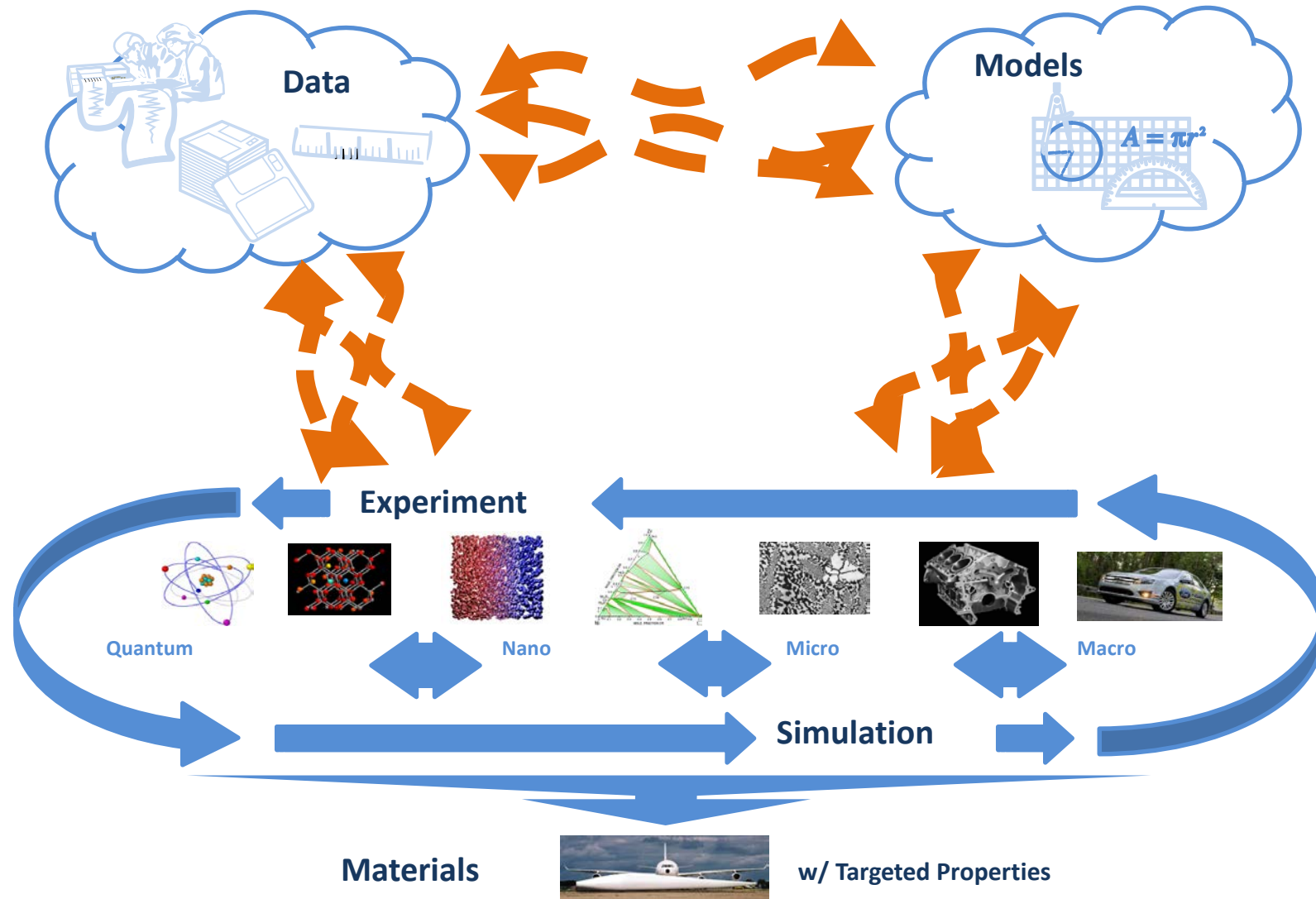


- MGI Subcommittee, Committee on Technology, National Science & Technology Council (First meeting 2012)
- Member Agencies (10): NIST, DOE, DOD, NSF, NASA, NIH, US Geological Survey, National Nuclear Security Administration, DARPA, and Office of Management and Budget
- National Institute of Standards & Technology
  - The US National Metrology Lab
  - Develop consensus standards to support international trade and commerce

**NIST**

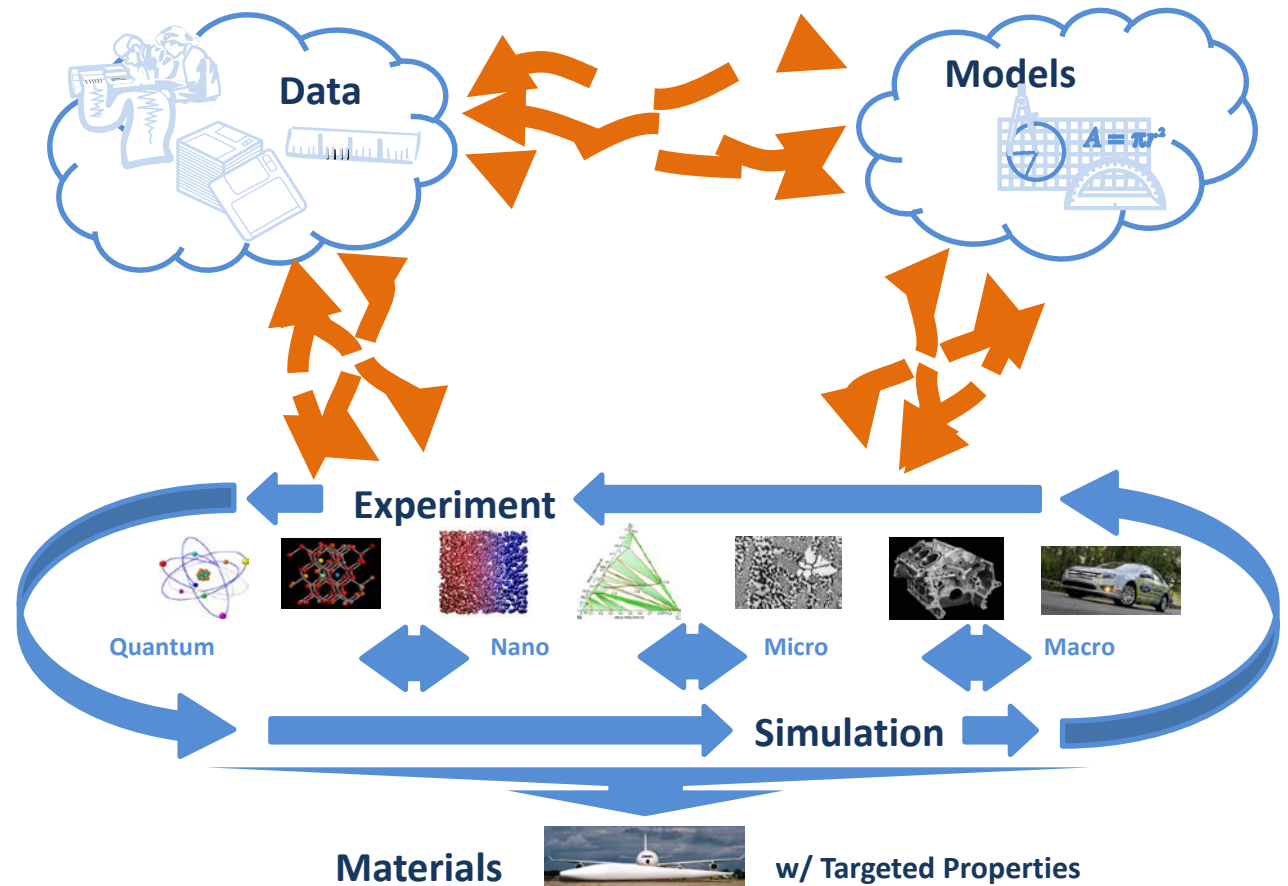


# MGI: Facilitating Data Science



# MGI: Facilitating Data Science

- Bank it: Data Ingestion & Repositories
- Share it: Standards for data & metadata
- Find it: Data Discovery
- Check it:
  - Curation
  - Uncertainty Quantification in Data / Models



# Office of Data and Informatics

## SRD

- continue existing SRD distribution
- Quality Framework
- SRD Modes
- assess external need
- new product ideas
  - SRMDS
  - data streams
  - alternative delivery methods
- Open Data Initiative
- Open Govt Directive
- Data.gov

## Research Data

- deal w/ data deluge
- provide advice to MML bench staff
- gather best practices
- interpret external rules & regulations
- reduce redundancy
- promote cooperation and coherent action
- manage changes in scholarly publishing
- coordinate with
  - WERB
  - Library
  - JResNIST

## Lead/Liaison

- partner with ITL
- represent MML
  - NIST committees
  - NSTC & IWGs
  - NIH, NSF, DOE
  - other Fed Govt
  - Research Data Alliance (RDA)
- data standards
- champion proposals
  - budget initiatives
  - IMS
  - inter-agency, RDA

## Data Science

*The 4<sup>th</sup> paradigm?*

- will it stand next to
  - theoretical
  - experimental
  - computational
- Cloud
- Statistical Learning
- Big Data
- Knowledge Discovery
- very fast growing
- *many* new jobs
- new degrees/depts

NIST Center for  
Excellence in Advanced Materials

---

## Center of Hierarchical Materials Design



Peter Voorhees, Gregory Olson | Begum Gulsoy *Northwestern University*  
Juan De Pablo | *University of Chicago*



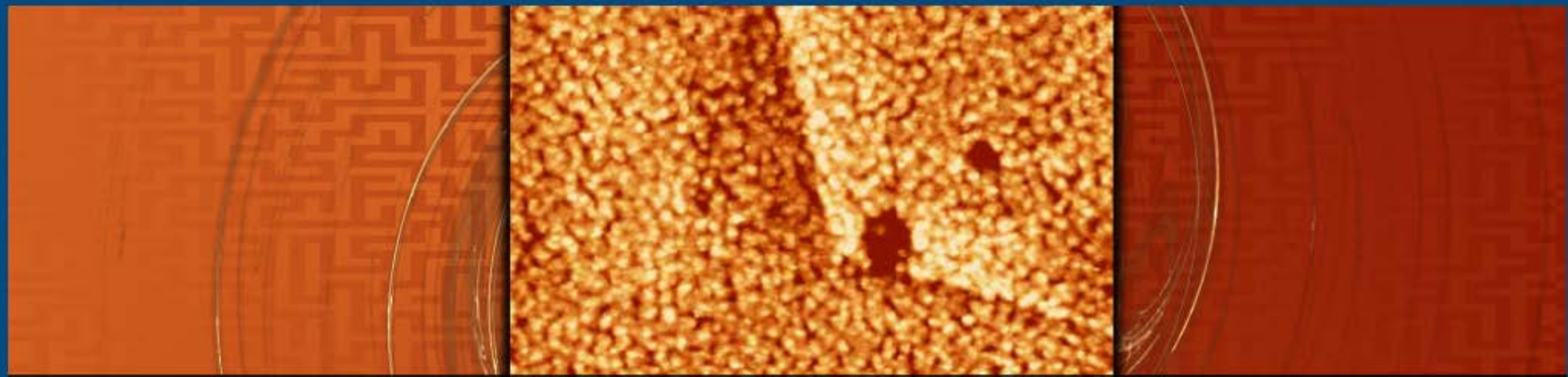
# NIST Materials Genome Initiative

## Gateway to Materials Genome Information

LOGIN | **NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

 Search

- HOME
- PROJECTS
- ABOUT
- MATERIAL RESOURCE CENTERS
- FEDERAL MATERIALS GENOME INITIATIVE
- ONLINE TOOLS
- CONTACT



### Pt on Au

Scanning tunneling microscope image shows ultrathin film layer of platinum deposited on gold.



- Download Key Reports
- External Stakeholders
- Project Owners
- MGI Projects by Category

### CHiMaD

Center for Hierarchical Materials Design (CHiMaD) is a NIST-sponsored center of excellence for advanced materials research focusing on developing the next generation of computational tools, databases and experimental techniques in order to enable the accelerated design of novel materials and their integration to industry, one of the primary goals of the Materials Genome Initiative (MGI).

### Quick Navigation by Project Category

- High Throughput Materials Science
- Data and Model Dissemination
- Data Capture





# NIST Materials Genome Initiative

## Gateway to Materials Genome Information

LOGIN |

HOME

PROJECTS

ABOUT

MATERIAL RESOURCE CENTERS

FEDERAL MATERIALS GENOME INITIATIVE

ONLINE TOOLS

CONTACT

[Download Key Reports](#) | [External Stakeholders](#) | [Project Owners](#) | [MGI Projects by Category](#)

### NIST Projects Supporting the MGI:



#### MGI Projects by Category

▶ [Data and Model Dissemination](#)

▶ [Software Tools](#)

▶ [Data Capture](#)

▶ [UQ, Data Quality, Improved Models](#)

▶ [High Throughput Materials Science](#)

▶ [Use Cases](#)

▶ [CHiMaD Pages](#)

▶ [Markup Tools and Workflow](#)

▶ [Facilities](#)

#### MGI Project Categories

- [High Throughput Materials Science \(5\)](#)
- [Data and Model Dissemination \(17\)](#)
- [Data Capture \(8\)](#)
- [Markup Tools and Workflow \(3\)](#)
- [Software Tools \(11\)](#)
- [UQ, Data Quality, Improved Models \(9\)](#)
- [Use Cases \(4\)](#)
- [Facilities \(3\)](#)
- [CHiMaD Pages \(3\)](#)

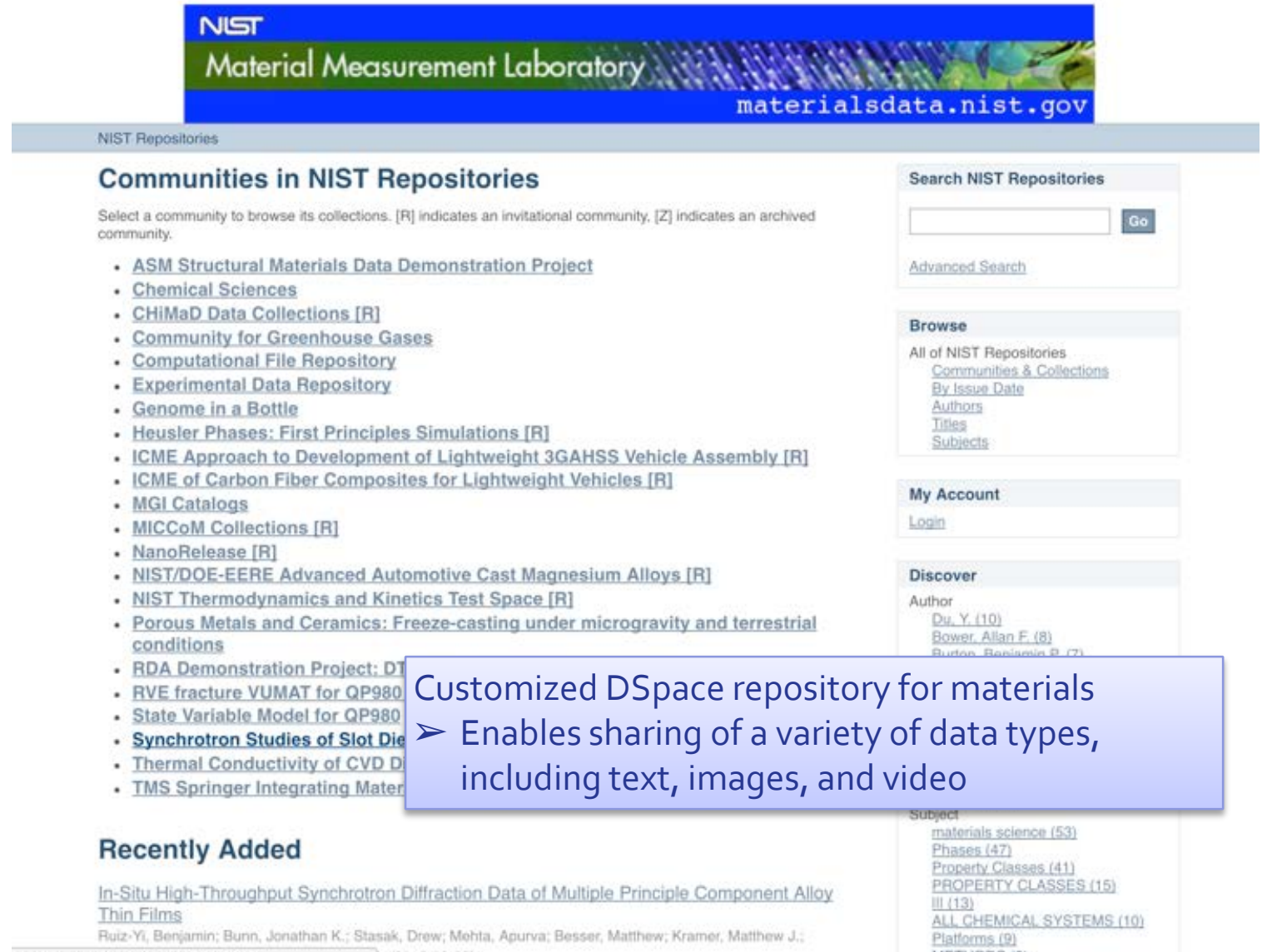
# MGI @ NIST

- MGI: Facilitating data science
  - Standards for data, metadata, & uncertainty
  - Repositories
  - Data Discovery
  - Uncertainty in Models
- MGI Examples / Prototypes
  - Force Field Calculations
  - Autonomous Phase Mapping

# Repositories

- Bank it

<https://materialsdata.nist.gov/>



NIST  
Material Measurement Laboratory  
materialsdata.nist.gov

NIST Repositories

## Communities in NIST Repositories

Select a community to browse its collections. [R] indicates an invitational community, [Z] indicates an archived community.

- [ASM Structural Materials Data Demonstration Project](#)
- [Chemical Sciences](#)
- [CHiMaD Data Collections \[R\]](#)
- [Community for Greenhouse Gases](#)
- [Computational File Repository](#)
- [Experimental Data Repository](#)
- [Genome in a Bottle](#)
- [Heusler Phases: First Principles Simulations \[R\]](#)
- [ICME Approach to Development of Lightweight 3GAHSS Vehicle Assembly \[R\]](#)
- [ICME of Carbon Fiber Composites for Lightweight Vehicles \[R\]](#)
- [MGI Catalogs](#)
- [MICCoM Collections \[R\]](#)
- [NanoRelease \[R\]](#)
- [NIST/DOE-EERE Advanced Automotive Cast Magnesium Alloys \[R\]](#)
- [NIST Thermodynamics and Kinetics Test Space \[R\]](#)
- [Porous Metals and Ceramics: Freeze-casting under microgravity and terrestrial conditions](#)
- [RDA Demonstration Project: DT](#)
- [RVE fracture VUMAT for QP980](#)
- [State Variable Model for QP980](#)
- [Synchrotron Studies of Slot Die](#)
- [Thermal Conductivity of CVD D](#)
- [TMS Springer Integrating Mater](#)

Search NIST Repositories

Advanced Search

Browse

All of NIST Repositories

- [Communities & Collections](#)
- [By Issue Date](#)
- [Authors](#)
- [Titles](#)
- [Subjects](#)

My Account

Login

Discover

Author

- [Du, Y. \(10\)](#)
- [Bower, Allan F. \(8\)](#)
- [Burton, Benjamin P. \(7\)](#)

Subject

- [materials science \(53\)](#)
- [Phases \(47\)](#)
- [Property Classes \(41\)](#)
- [PROPERTY CLASSES \(15\)](#)
- [III \(13\)](#)
- [ALL CHEMICAL SYSTEMS \(10\)](#)
- [Platforms \(9\)](#)

Recently Added

[In-Situ High-Throughput Synchrotron Diffraction Data of Multiple Principle Component Alloy Thin Films](#)

Ruiz-Yi, Benjamin; Bunn, Jonathan K.; Stasak, Drew; Mohta, Apurva; Besser, Matthew; Kramer, Matthew J.;

Customized DSpace repository for materials

- Enables sharing of a variety of data types, including text, images, and video

**Data Citation:**  
**Campbell, Carelyn; Zhao, J-C; Henry, M. F.**  
**Examination of Ni-base superalloy diffusion couples containing multiphase regions (2014-04-02)**  
<http://hdl.handle.net/11256/22>

Digital Identifier

**Affiliation:** National Institute of Standards and Technology, Metallurgy Division, Gaithersburg, MD 20899-8555, USA  
 General Electric Company, GE Global Research, 1 Research Circle, Niskayuna, NY 12309, USA  
**Contact Email:** carelyn.campbell@nist.gov

**Primary Publication Citation:**  
 Materials Science and Engineering A 407 (2005) 135–146  
<http://dx.doi.org/10.1016/j.msea.2005.07.016>

Related Work

**Related Publications by Author:**  
 Campbell CE, Boettinger WJ, Katiner UR (2002) Development of a diffusion mobility database for Ni-base superalloys. Acta Mater 50:775-792 DOI: [http://dx.doi.org/10.1016/S1359-6454\(01\)00383-4](http://dx.doi.org/10.1016/S1359-6454(01)00383-4)

Campbell CE, Zhao JC, Henry MF (2004) Comparison of experimental and simulated multicomponent Ni-base superalloy diffusion couples. J Phase Equil Dif 25 (1):6-15. DOI: <http://dx.doi.org/10.1361/10549710417966>

**Abstract:**  
 Four Ni-base superalloy diffusion couples with multiphase regions were studied. The diffusion couples contained single-phase (gamma ), two phase( gamma +MC carbide) and three-phase ( gamma + gamma prime+MC carbide) regions. Measured average composition profiles were in good agreement with the diffusion simulation predictions. The measured and predicted phase fraction profiles showed similar trends; however, there were some discrepancies in the predicted position of the gamma +gamma prime + MC/ gamma +MC boundary. Phase fraction profiles and optical metallography were used to determine the type and direction of the moving phase region boundaries.

Search

Search  
 This Item  
 Advanced

Browse

All of these  
 Collections  
 Subjects  
 Titles  
 Authors  
 This Collection  
 Subjects  
 Titles  
 Authors



My Account

Login

**Funding Agency & Award No.:**  
 This work was supported by the Defense Advanced Research Project Agency (DARPA) under the accelerated Insertion of Materials (AIM) Program (Grant number F33615-00-C-5215) with Dr. L. Christodoulou as the project manager and Dr. Rollie Dutton as the project monitor. The authors would like to express their appreciation to N. Saunders for the use of his thermodynamic database for Ni alloys and to Louis A. Poluso, Rebecca Casey, Mitch Hammond and Karen Denike for their experimental support.

Data files

**Files in this item**

	<b>Name:</b> R95-R88-expsimul.txt <b>Size:</b> 25.32Kb <b>Format:</b> Text file <b>Description:</b> Experimental and simulate composition profiles for the R88/R95 diffusion couple at 1150 C for 1000 h	<a href="#">View/Open</a>
	<b>Name:</b> r95r88-1000h-labe ... <b>Size:</b> 58.76Mb <b>Format:</b> TIFF image <b>Description:</b> Micrograph of R95/R88 diffusion couple after 1000 h at 1150 C	<a href="#">View/Open</a>



The following license files are associated with this item:  
 • Creative Commons  
**This item appears in the following Collection(s)**  
 • Diffusion Data

Offer licenses with attribution 3.0



 Except where otherwise noted, Universal

**Related items**  
 Showing items related by title, author, creator and subject.

**Further Studies on the Nickel-Aluminum System. I. The  $\beta$ -Ni<sub>2</sub>Al<sub>3</sub> Phase Fields**  
 Taylor, A; Doyle, N.J. (1972-01-31)  
 New lattice parameter and density results have been obtained for alloys in the  $\beta$ -NiAl and  $\beta$ -Ni<sub>2</sub>Al<sub>3</sub> phase fields of the nickel-aluminum system. The lattice parameter of the  $\beta$ -NiAl phase (CsCl-type) falls linearly from ...

**Elemental vacancy diffusion for fcc and hcp structures**  
 Angsten, Thomas; Mayeshiba, Tam; Wu, Henry; Morgan, Dane (2014-08-08)  
 This work demonstrates how databases of diffusion-related properties can be developed from high-throughput ab initio calculations. The formation and migration energies for vacancies of all adequately stable pure elements ...

# Data Discovery

# Search for resources: Materials Resource Registry

- Find it

Materials Resource Registry Curation System

141.142.209.111:8181/explore/

Search for Resources

DFT \*

All text is searched

4 results

All Resources Organizations Data Collections Datasets Services Informational Sites Software

Custom View

Change Custom View

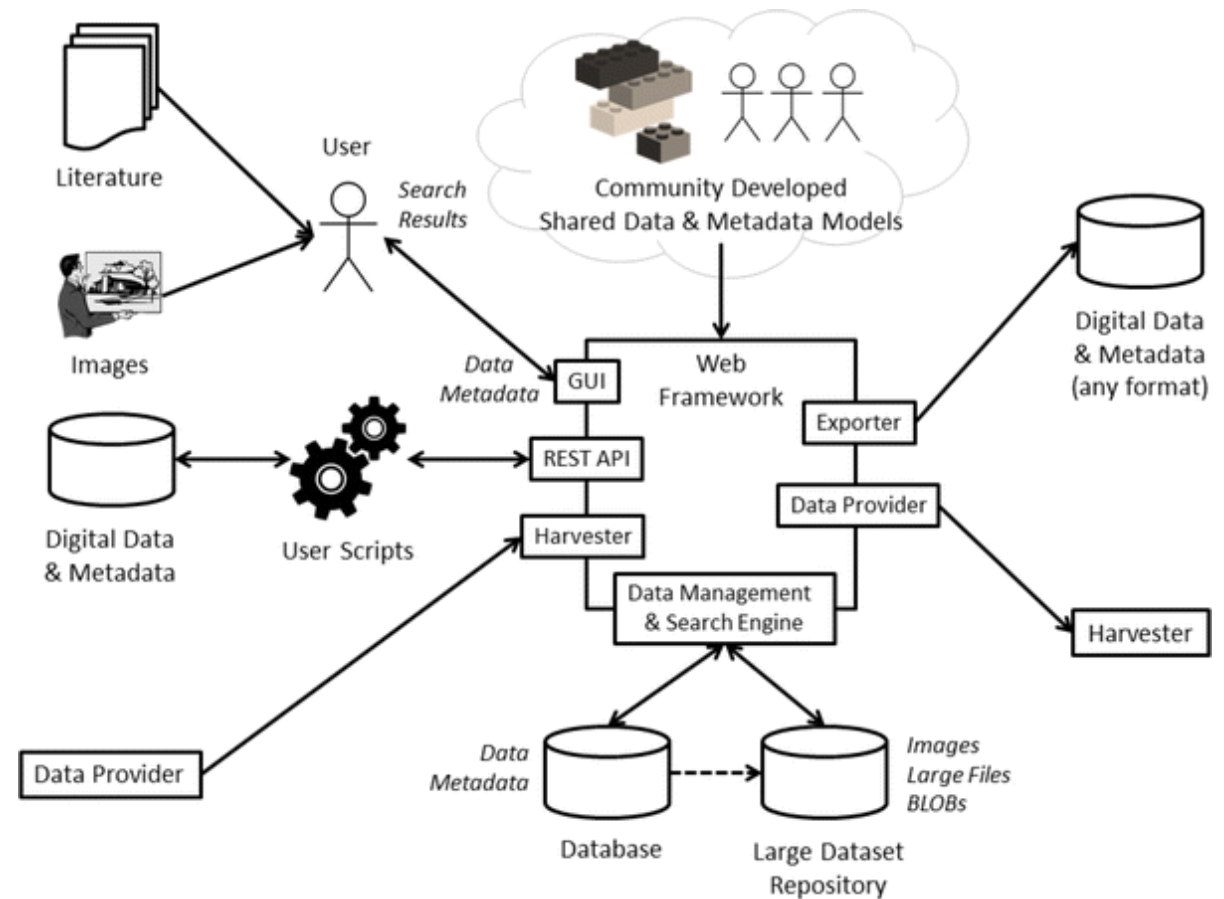
Resource Type:

- All Resources
- Organization
- Data Collection
- Repository
- Project Archive
- Database
- Dataset
- Service
- Informational Site
- Software

MPInterfaces	Resource Details	Go To
title	MPInterfaces	
description	MPInterfaces is a python package that enables high throughput Density Functional Theory (DFT) analysis of arbitrary material interfaces (ligand capped nanoparticles, surfaces in the presence of solvents and hetero-structure interfaces) using VASP, VASPsol, LAMMPS, materialsproject database as well as their open source tools and a little bit of ASE.	
subject	Python, Density Functional Theory (DFT), materials interfaces, surfaces, VASP, VASPsol, LAMMPS, MaterialsProject, ASE	
referenceURL	<a href="http://henniggroup.github.io/MPInterfaces/">http://henniggroup.github.io/MPInterfaces/</a>	
AFLOW	Resource Details	Go To
title	AFLOW	
description	Aflow is a globally available database of 647,815 material compounds	

# Materials Data Curation System

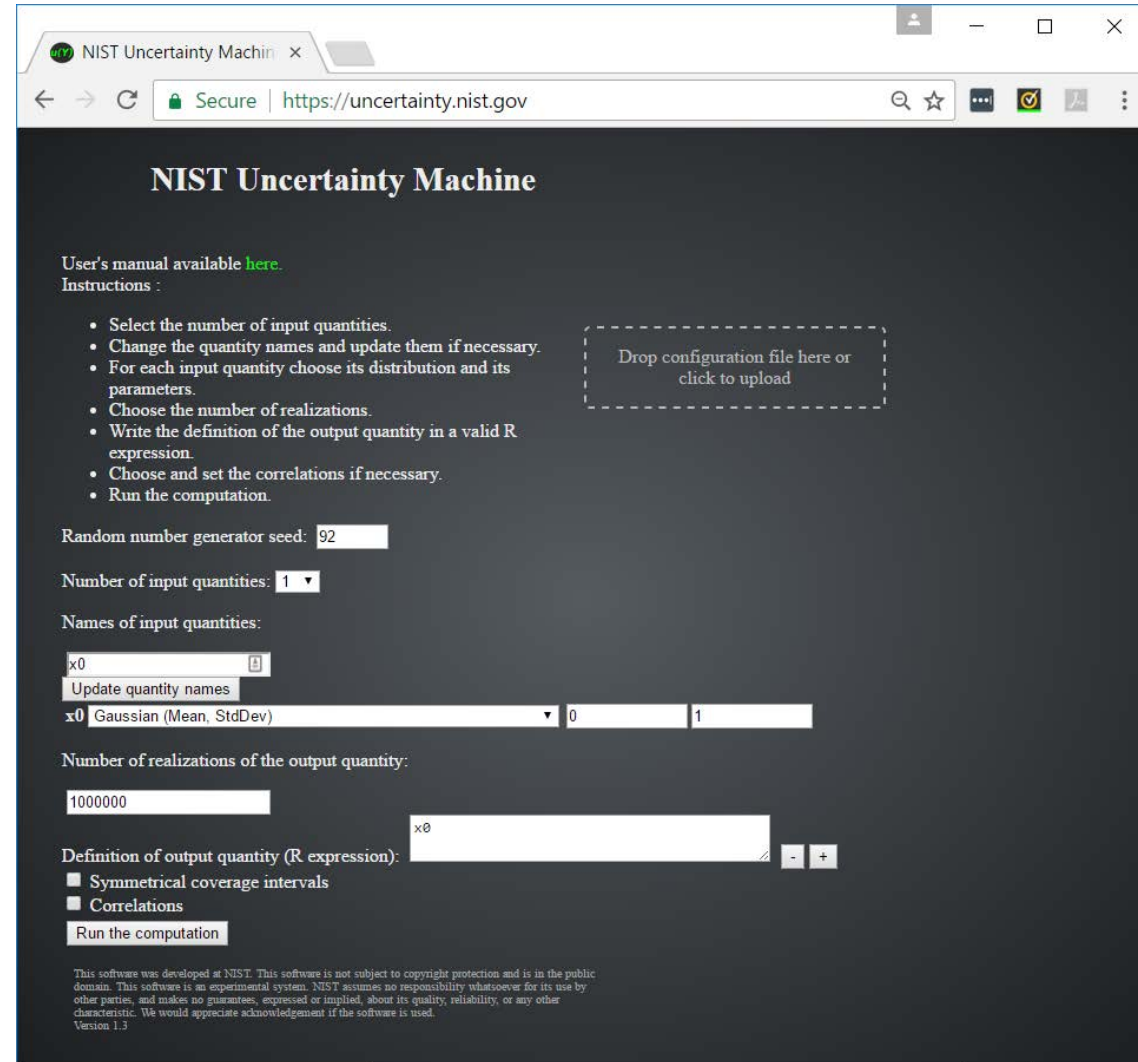
- Bank it: “smart” data ingestion tools & repository
- Share it: tools to automatically convert data into and out of standard formats.
- Find it: search via Website or API
- Federated search



Dima (JOM 2016) Informatics Infrastructure for the Materials Genome Initiative.

# Model Uncertainty

- (Check it)
- Molecular Dynamics
- Finite Element Analysis
- Thermodynamics (CALPHAD)
- etc.
- NIST Uncertainty Machine
  - Use probabilistic models to assess uncertainty – e.g. confidence, or credible intervals.



The screenshot shows the NIST Uncertainty Machine web interface in a browser. The page title is "NIST Uncertainty Machine". Below the title, there is a link to the user's manual and a section for instructions. The instructions list several steps: selecting the number of input quantities, changing quantity names, choosing distributions and parameters, selecting the number of realizations, writing the output quantity definition, setting correlations, and running the computation. The interface includes several input fields: a text box for the random number generator seed (set to 92), a dropdown for the number of input quantities (set to 1), a text box for the names of input quantities (set to x0), a dropdown for the distribution (set to Gaussian (Mean, StdDev)), a text box for the number of realizations (set to 1000000), and a text box for the definition of the output quantity (set to x0). There are also checkboxes for "Symmetrical coverage intervals" and "Correlations", and a "Run the computation" button. A dashed box on the right side of the page contains the text "Drop configuration file here or click to upload". At the bottom of the page, there is a disclaimer and the version number 1.3.

NIST Uncertainty Machine

User's manual available [here](#).

Instructions :

- Select the number of input quantities.
- Change the quantity names and update them if necessary.
- For each input quantity choose its distribution and its parameters.
- Choose the number of realizations.
- Write the definition of the output quantity in a valid R expression.
- Choose and set the correlations if necessary.
- Run the computation.

Random number generator seed: 92

Number of input quantities: 1

Names of input quantities:

x0

Update quantity names

x0 Gaussian (Mean, StdDev) 0 1

Number of realizations of the output quantity:

1000000

Definition of output quantity (R expression): x0

Symmetrical coverage intervals

Correlations

Run the computation

This software was developed at NIST. This software is not subject to copyright protection and is in the public domain. This software is an experimental system. NIST assumes no responsibility whatsoever for its use by other parties, and makes no guarantees, expressed or implied, about its quality, reliability, or any other characteristic. We would appreciate acknowledgement if the software is used.

Version 1.3



# Materials Data Facility

Streamlined and automated data sharing,  
discovery, access, and analysis

- Bank it
- Share it
- Find it
- Use it

Ian Foster ([foster@uchicago.edu](mailto:foster@uchicago.edu))<sup>1,2</sup>,  
Ben Blaiszik<sup>1,2</sup> ([blaiszik@uchicago.edu](mailto:blaiszik@uchicago.edu)),  
Jonathon Gaff<sup>1</sup>, Logan Ward<sup>1</sup>, Kyle Chard<sup>1</sup>, Jim  
Pruyne<sup>1</sup>,  
Rachana Ananthakrishnan<sup>1</sup>, Steven Tuecke<sup>1</sup>

Michael Ondrejcek<sup>3</sup>, Kenton McHenry<sup>3</sup>, John Towns<sup>3</sup>  
University of Chicago<sup>1</sup>, Argonne National Laboratory<sup>2</sup>, University of Illinois at Urbana-  
Champaign<sup>3</sup>

[materialsdatafacility.org](http://materialsdatafacility.org)  
[globus.org](http://globus.org)



Materials Genome Initiative



## Publication

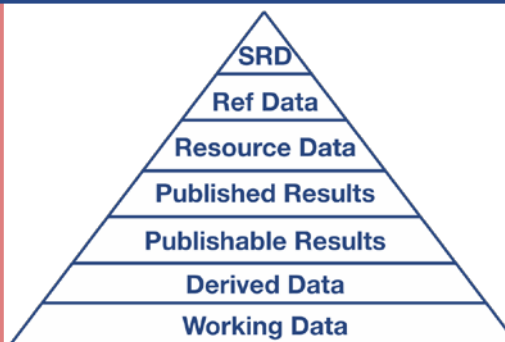
## REST APIs

## Discovery

- Bank it
- Share it
- Find it
- Use it

- Data ingestion
- Identify datasets with persistent identifiers (e.g., DOI or Handle)
- Describe datasets with appropriate metadata and provenance
- Verify dataset contents over time
- Handle big (and small) data: We have already ingested datasets with > 1.5 M files and > 1.5 TB in size

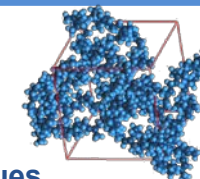
### Materialsdatafacility.org



- Search, query, and access datasets in modern ways
- Automatically index flexible metadata and harvest file contents
- Provide simple user interfaces (c.f., Google and Amazon)

### PPPDB

- Extract key polymer properties from literature via natural language processing and crowdsourcing
- Build interfaces to explore curated  $\chi$  and other property values
- Includes 375 polymer-polymer values and 1,014 polymer-solvent values



# STREAMLINE AND AUTOMATE: FOUR KEY STEPS

- Simplify data publication, regardless of size, type, and location
- Automate data and metadata ingest, to enable capture of many valuable materials datasets
- Enable unified search of disparate materials data sources
- Deploy APIs to foster community development, data creation, and data consumption

# Sharing Data and Tools

A New Model for Materials Genome  
Initiative - Driven Research:  
“High-throughput Experimental Materials  
Collaboratory (HTEMC)”

Zachary T. Trautt<sup>1</sup>, Andriy Zakutayev<sup>2</sup>,  
Martin L. Green<sup>1</sup>, John Perkins<sup>2</sup>

<sup>1</sup>National Institute of Standards and Technology (NIST)

<sup>2</sup>National Renewable Energy Laboratory (NREL)

*[martin.green@nist.gov](mailto:martin.green@nist.gov)*

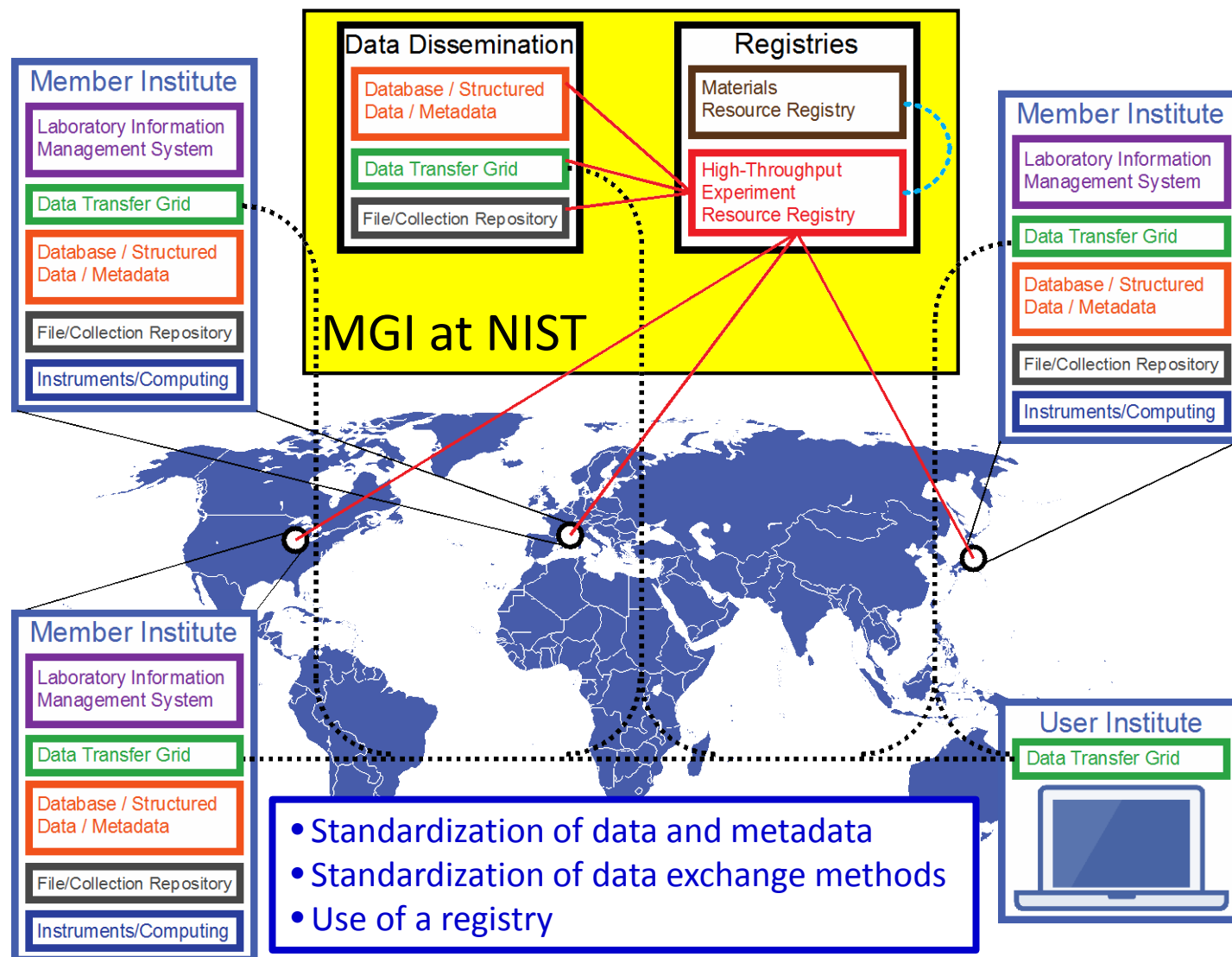
# Collaboratory

**Collaboratory:** a 1989 neologism (William A. Wulf, Computer Scientist at University of Virginia):

“...defined by...a ‘center without walls,’ in which the nation’s researchers can perform their research without regard to physical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, .... accessing information in digital libraries.”

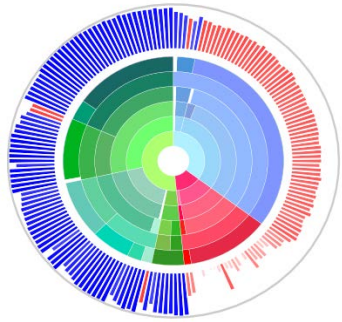
- The HTEMC would consist of:
  - An integrated, delocalized network of high-throughput synthesis and characterization tools
  - A best-in-class materials data management platform, consisting of NIST (and other) software

# Platform Integration Model



Education





# Annual Machine Learning for Materials Research: Bootcamp and Workshop

- Host: University of Maryland
- Dates: **TBD (~June / July)**
- Location: University of Maryland, College Park, MD

<https://nanocenter.umd.edu/events/mlmr/>

The event will introduce materials researchers from industry, national laboratories, and academia to machine learning theory and tools for rapid materials data analysis.

## Bootcamp

Three days of lectures and hands-on exercises covering a range of data analysis topics from data pre-processing through advanced machine learning analysis techniques. Example topics include:

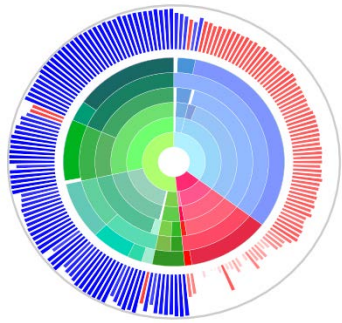
- Identifying important features in complex/high dimensional data
- Visualizing high dimensional data to facilitate user analysis.
- Identifying the fabrication 'descriptors' that best predict variance in functional properties.
- Quantifying similarities between materials using complex/high dimensional data

The **hands-on exercises** will demonstrate practical use of machine learning tools on real materials data.

Attendees will learn to analyze a range of data types from scalar properties such as material hardness to high dimensional spectra and micrographs.

## Workshop

Talks by top researchers in the field as well as open discussions in which attendees can discuss their data analysis needs with experts.



# Annual Machine Learning for Materials Research: Bootcamp and Workshop

<https://nanocenter.umd.edu/events/mlmr/>

MACHINE LEARNING FOR MATERIALS RESEARCH: BOOTCAMP		
<b>DATA FUNDAMENTALS DATA PREPROCESSING TUE 6/28</b>	<b>SUPERVISED LEARNING THEORY, DATA, ALGORITHMS WED 6/29</b>	<b>UNSUPERVISED LEARNING THEORY &amp; ALGORITHMS THU 6/30</b>
Filtering: Noise Smoothing Background Subtraction Feature Extraction <ul style="list-style-type: none"> <li>▪ Cross-correlation Wavelets</li> <li>▪ Edges</li> <li>▪ Closed Boundaries</li> <li>▪ Shapes</li> </ul>	Data Handling <ul style="list-style-type: none"> <li>▪ Cross Validation</li> <li>▪ Prediction</li> </ul> Algorithms <ul style="list-style-type: none"> <li>▪ Regularized Least Squared</li> <li>▪ Support Vector Machines</li> <li>▪ Neural Networks</li> <li>▪ Decision Trees &amp; Ensemble Learning</li> <li>▪ Genetic Programming</li> </ul>	Theory Similarity Measures Latent Variable Analysis Spectral Unmixing Matrix Factorization Clustering
MACHINE LEARNING FOR MATERIALS RESEARCH: WORKSHOP		
<b>WORKSHOP SESSION 1 THU 6/30 EVENING</b>	<b>WORKSHOP SESSION 2 FRI 7/1 MORNING</b>	

## Organizers

**A. Gilad Kusne**  
National Institute of Standards & Technology  
Materials Measurement Science Division



**Daniel Samarov**  
National Institute of Standards and Technology  
Information Technology Laboratory



**Alexei Belianinov**  
Oak Ridge National Laboratory  
Center for Nanophase Materials Sciences



**Tim Mueller**  
John Hopkins University  
Department of Materials Science & Engineering



**Stefano Ermon**  
Stanford University  
Department of Computer Science



**Ichiro Takeuchi**  
University of Maryland, College Park  
Department of Materials Science & Engineering



# MGI Examples

- Force Field Calculations (Logan Ward, Northwestern Univ)
- Autonomous Phase Mapping

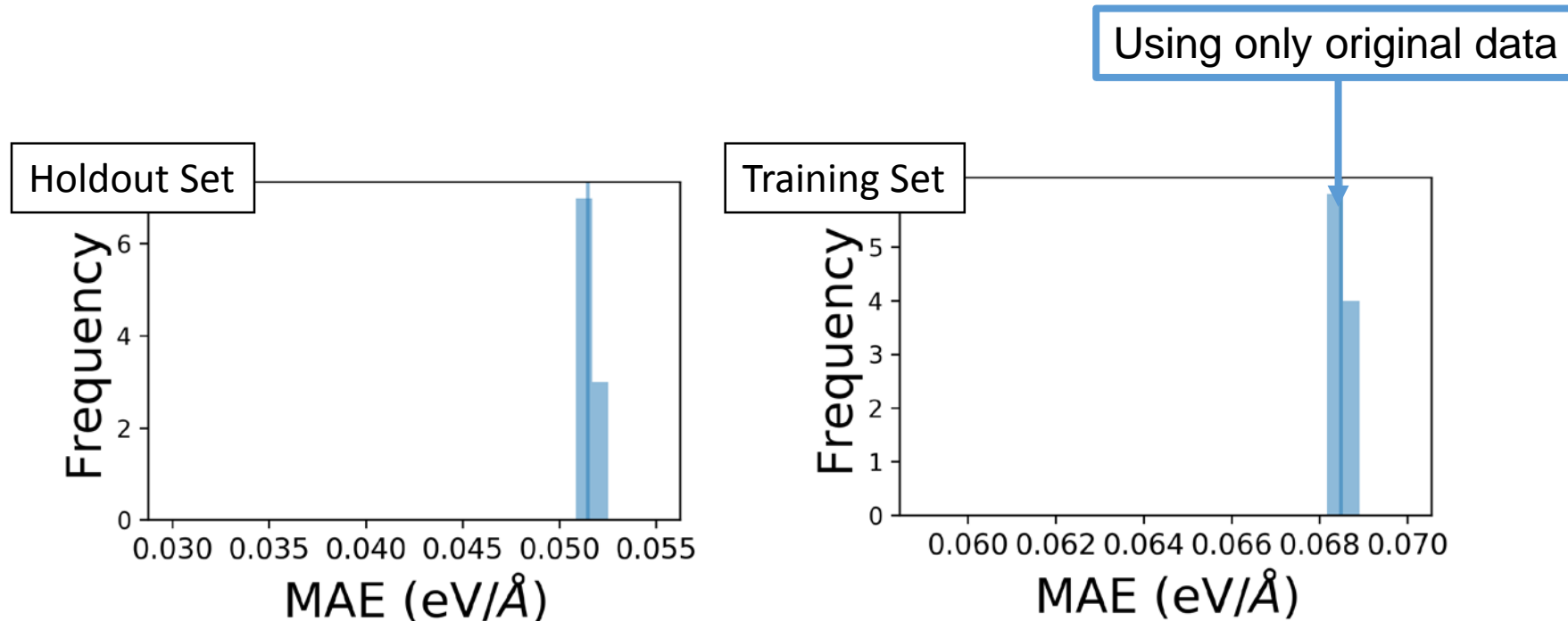
# Building a machine learning model using MDF

**Example:** Building force-field potentials from different datasets

**Data resources:** 3 DFT datasets with Aluminum data

1 dataset from [khazana.uconn.edu](http://khazana.uconn.edu), 2 datasets from [materialsdata.nist.gov](http://materialsdata.nist.gov)

**Result:** Improved performance by integrating data sources



MAE – Mean absolute error

**Method:** Botu *et al.* [JPCC.](#) (2017)

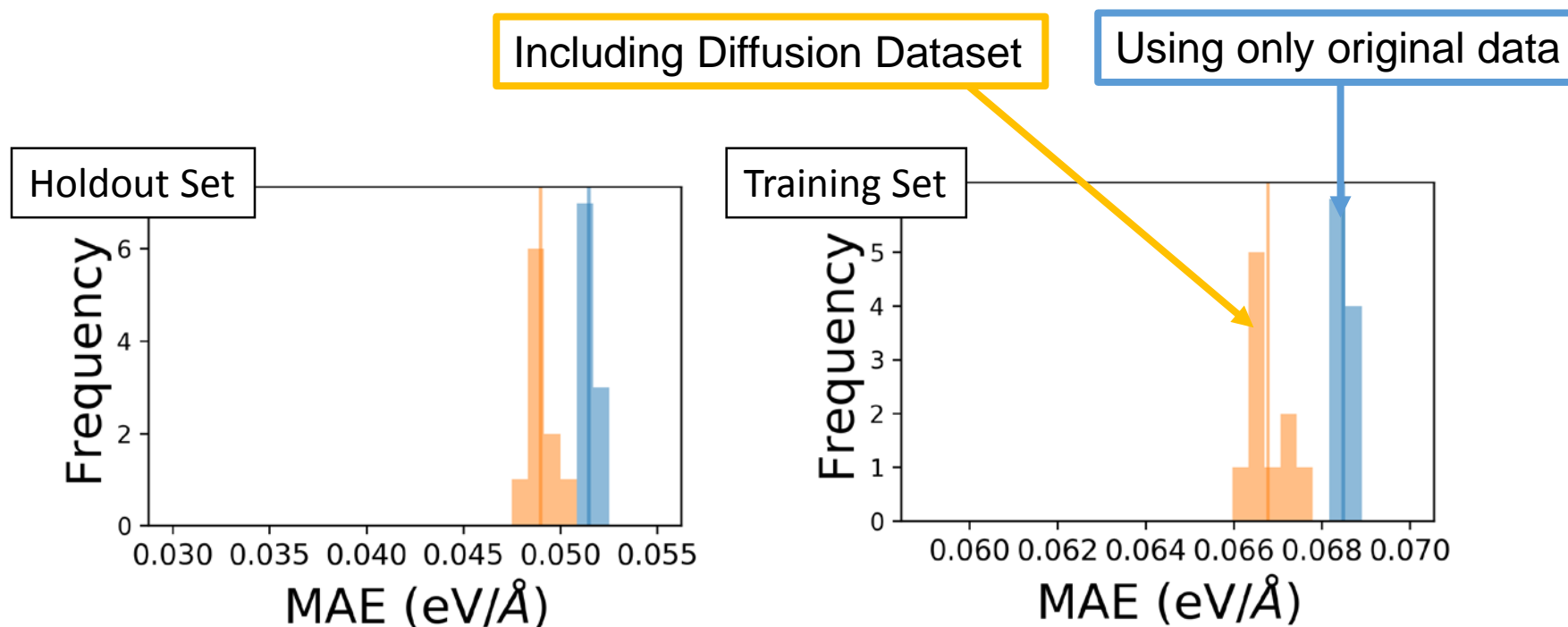
# Building a machine learning model using MDF

**Example:** Building force-field potentials from different datasets

**Data resources:** 3 DFT datasets with Aluminum data

1 dataset from [khazana.uconn.edu](http://khazana.uconn.edu), 2 datasets from [materialsdata.nist.gov](http://materialsdata.nist.gov)

**Result:** Improved performance by integrating data sources



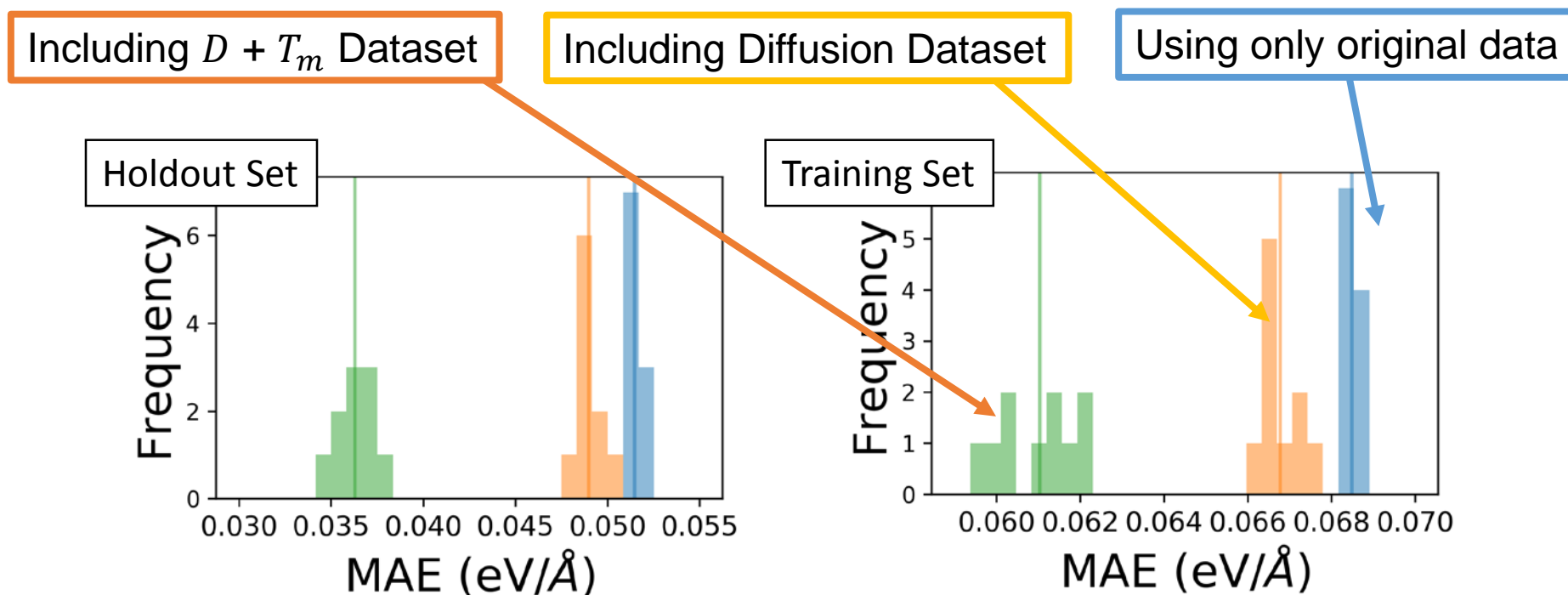
# Building a machine learning model using MDF

**Example:** Building force-field potentials from different datasets

**Data resources:** 3 DFT datasets with Aluminum data

1 dataset from [khazana.uconn.edu](http://khazana.uconn.edu), 2 datasets from [materialsdata.nist.gov](http://materialsdata.nist.gov)

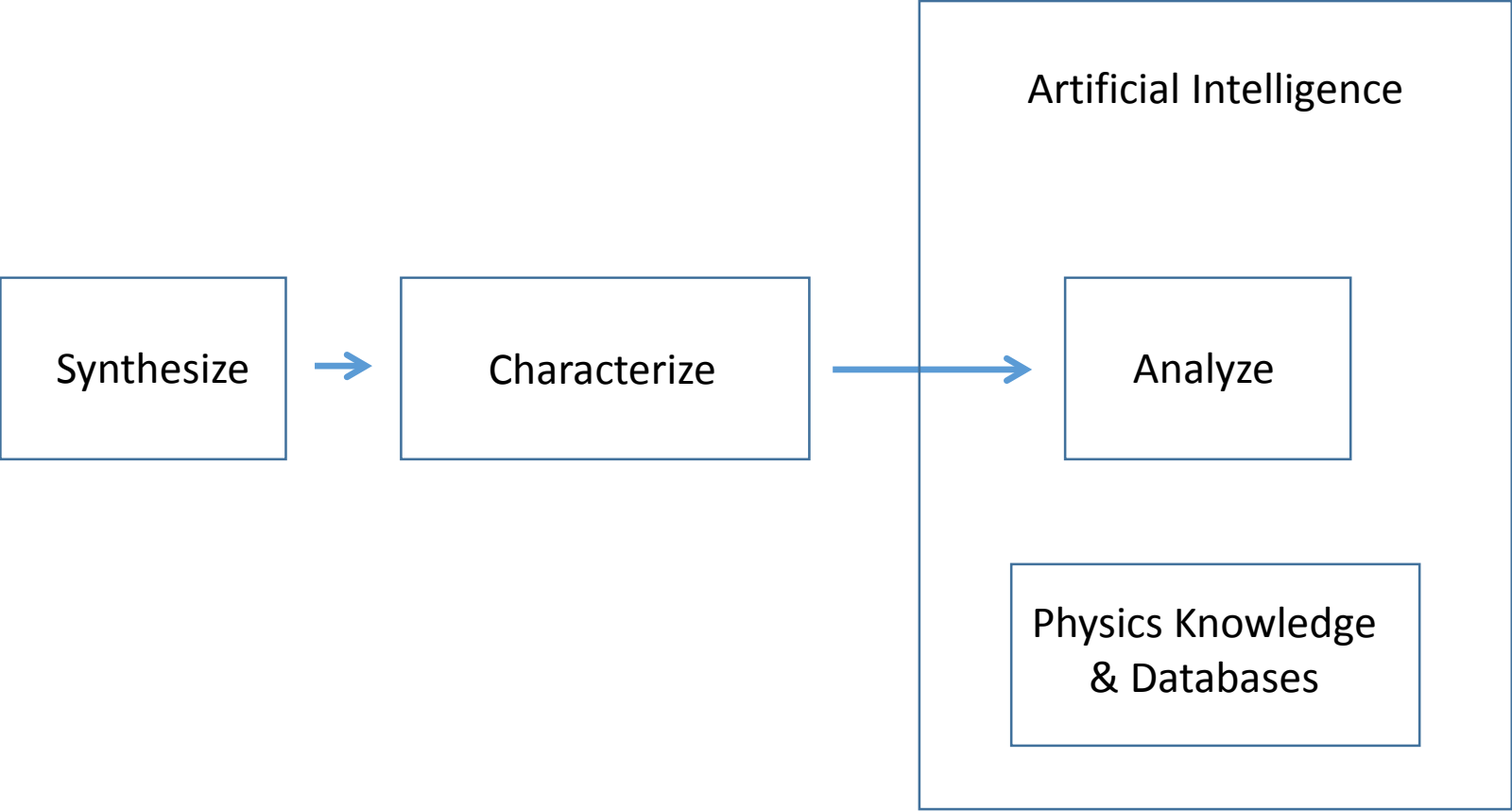
**Result:** Improved performance by integrating data sources



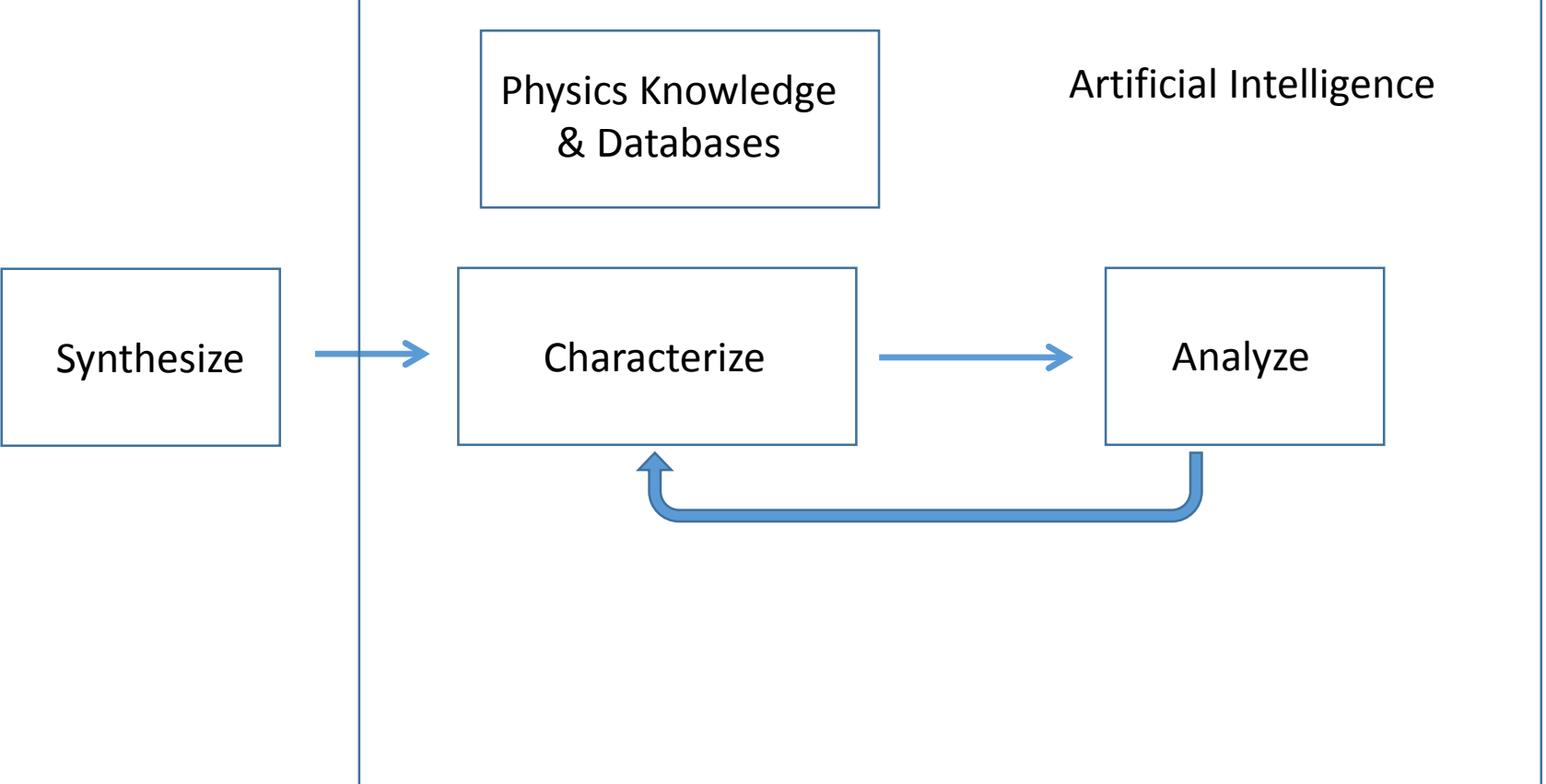
**Better performance in original application: No new DFT calculations**

Method: Botu *et al.* [JPCC](#). (2017)

# AI for Analysis



# Autonomous Metrology

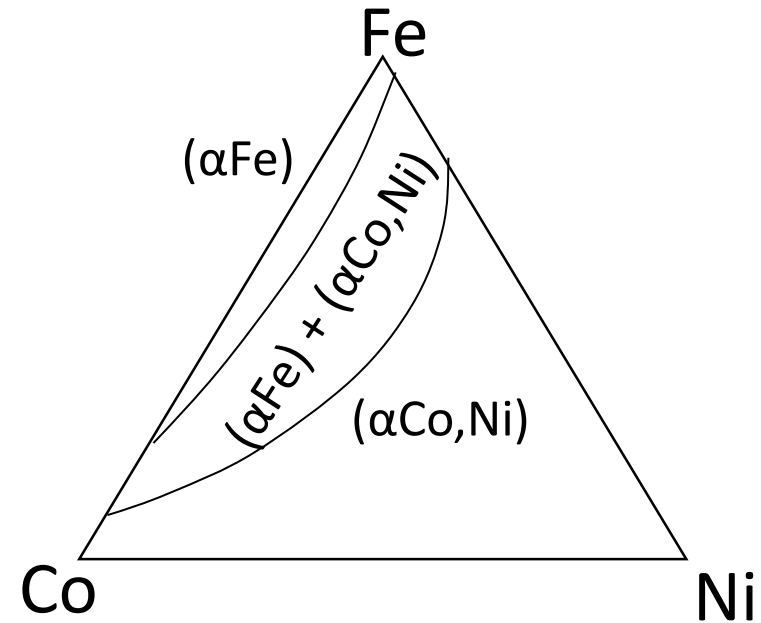




# Using AI to Identify Structural Phase Maps

## Structural Phase Map

- Structure as a function of fabrication parameters (e.g. composition, temperature, pressure, etc.)
- Use map to predict structure of new materials.
- Structure is good predictor of important properties.

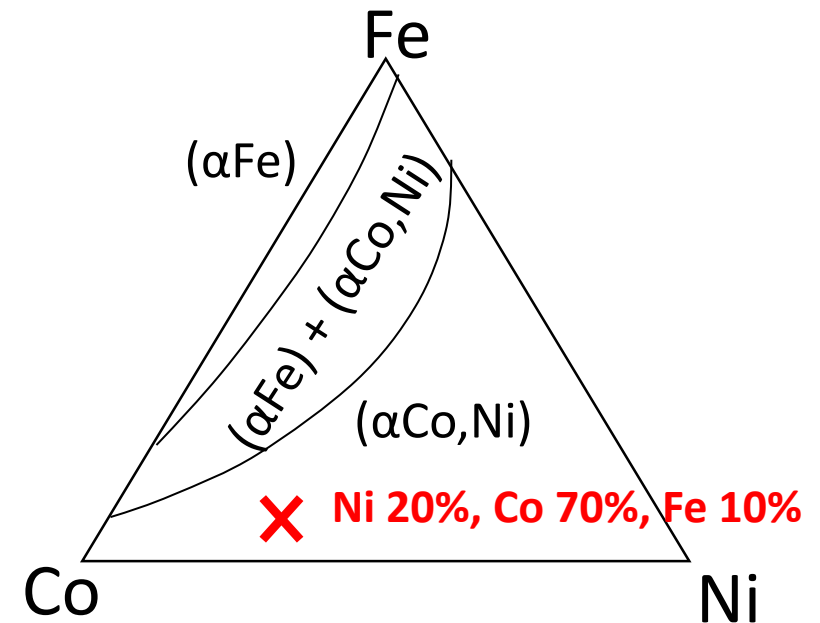


Phase Equilibria in Iron Ternary Alloys (1988) #60

# Structural Phase Mapping: Edisonian Approach

Traditional / Edisonian Approach:

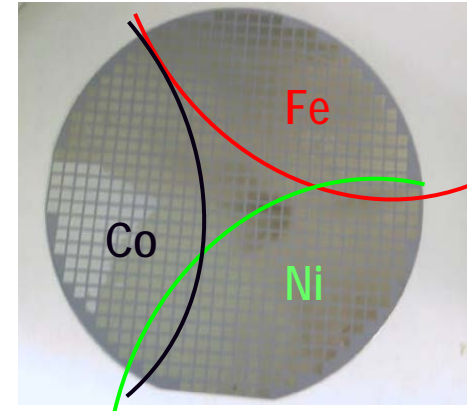
- Fabricate sample
- Measure structure
- Point placed on phase diagram
- Repeat
- This process takes years.



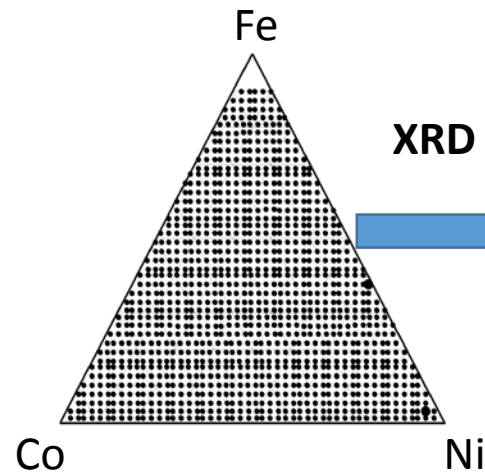
Phase Equilibria in Iron Ternary Alloys (1988) #60

# Phase Mapping: High-Throughput Approach

- Fabricate hundreds-thousands of samples -> HiTp Synthesis
- Measure all samples -> HiTp Characterization
- Rapid phase mapping -> Machine Learning

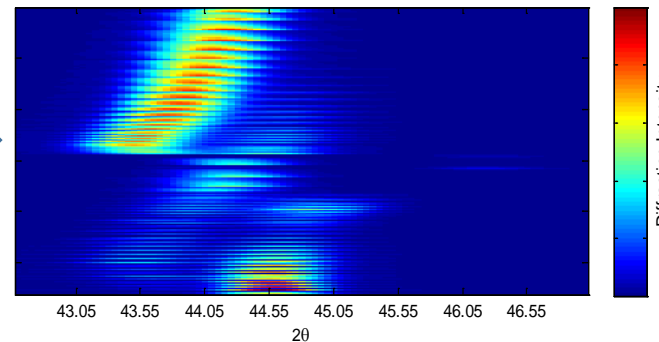


Combi Library for Ternary Spread



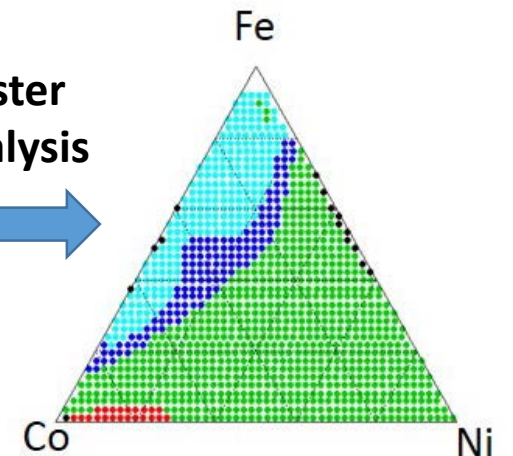
XRD

Diffraction Patterns



Cluster Analysis

Estimated Phase Map



APL Materials (2016)  
Composition–structure–  
property mapping in high-  
throughput experiments:  
Turning data into knowledge

# Phase Mapping: High-Throughput Approach

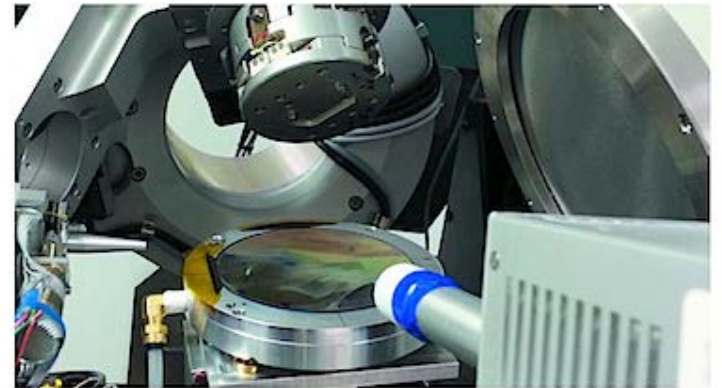
- Measurement is a time / resource sink
- For wafer of 500+ samples:
- In Lab: Takes weeks-months
- Synchrotron: Takes 5+ Hours (Every second counts)



Mn-Ni-Ge library  
535 samples



Bruker D8  
30 Minutes per sample  
2 weeks!



Stanford Synchrotron Radiation  
Lightsource  
30 seconds per sample  
4.5 hours

# Autonomous Metrology: Motivation

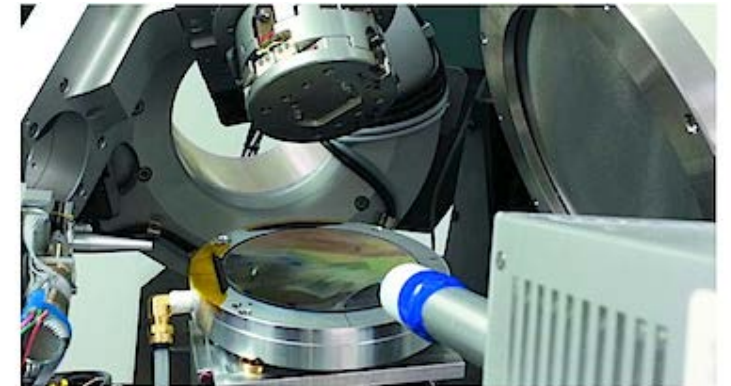
Why use AI to just analyze data? Put it on control of the equipment!

Instead of measuring all the samples, measure only the ones that count -> AI for optimal experiment design

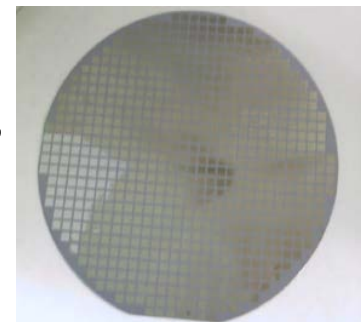
- Minimum measurements -> Maximum knowledge
- Save on worker hours and instrument time.
- Start it up and let it run.
- Minimize human bias: experiment design, execution, data analysis
- Replaced with traceable algorithmic bias
- Democratize Science
  - Simplify equipment use
  - Collaboratory



Bruker D8  
30 Minutes per sample  
2 weeks!



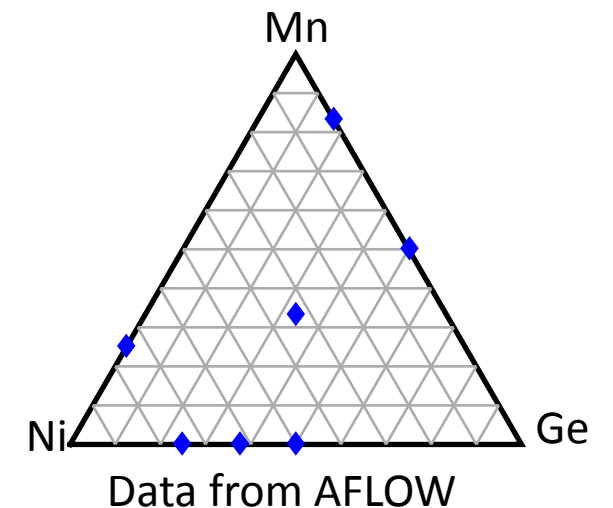
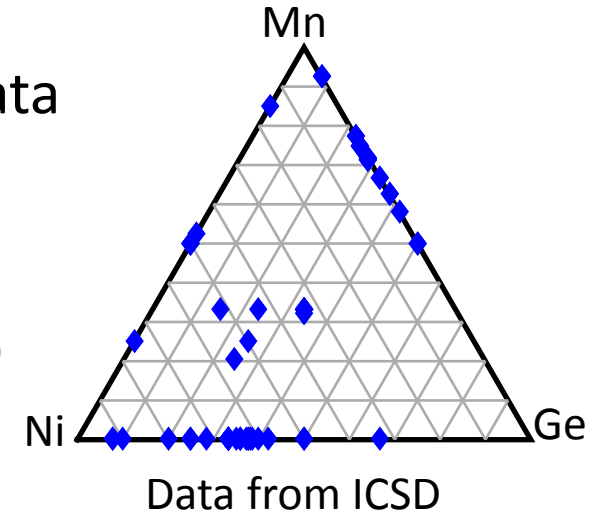
Stanford Synchrotron Radiation  
Lightsource  
30 seconds per sample  
4.5 hours



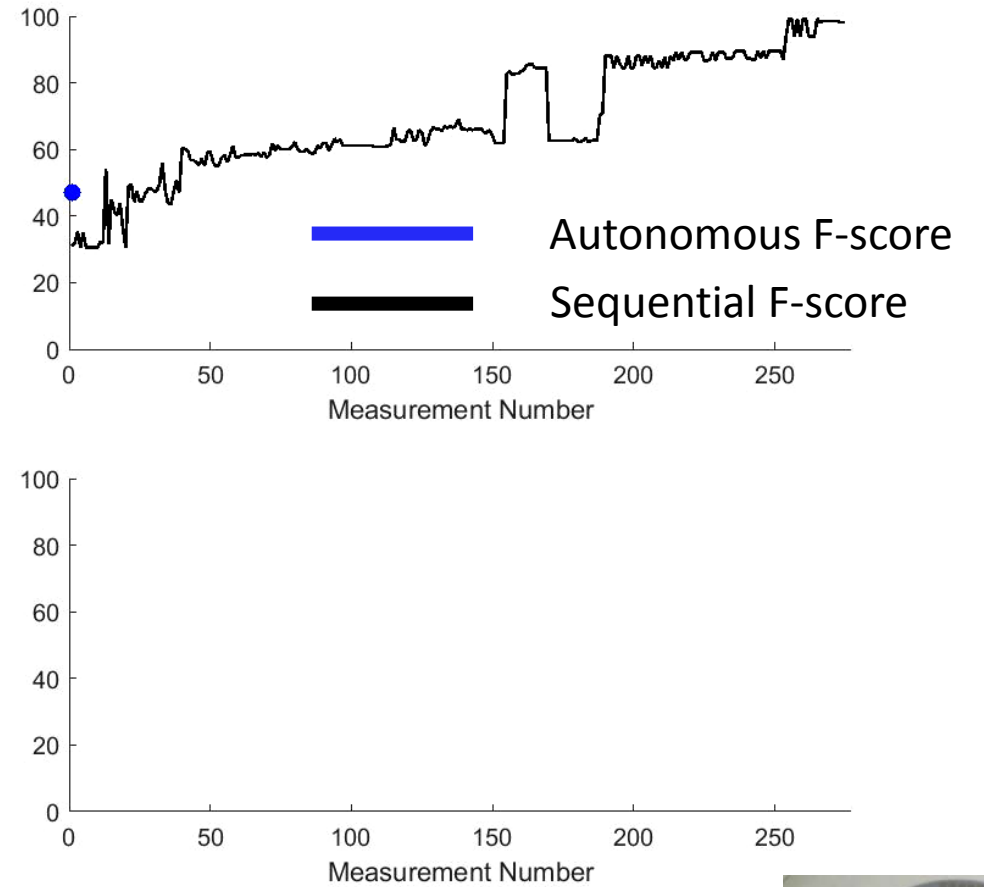
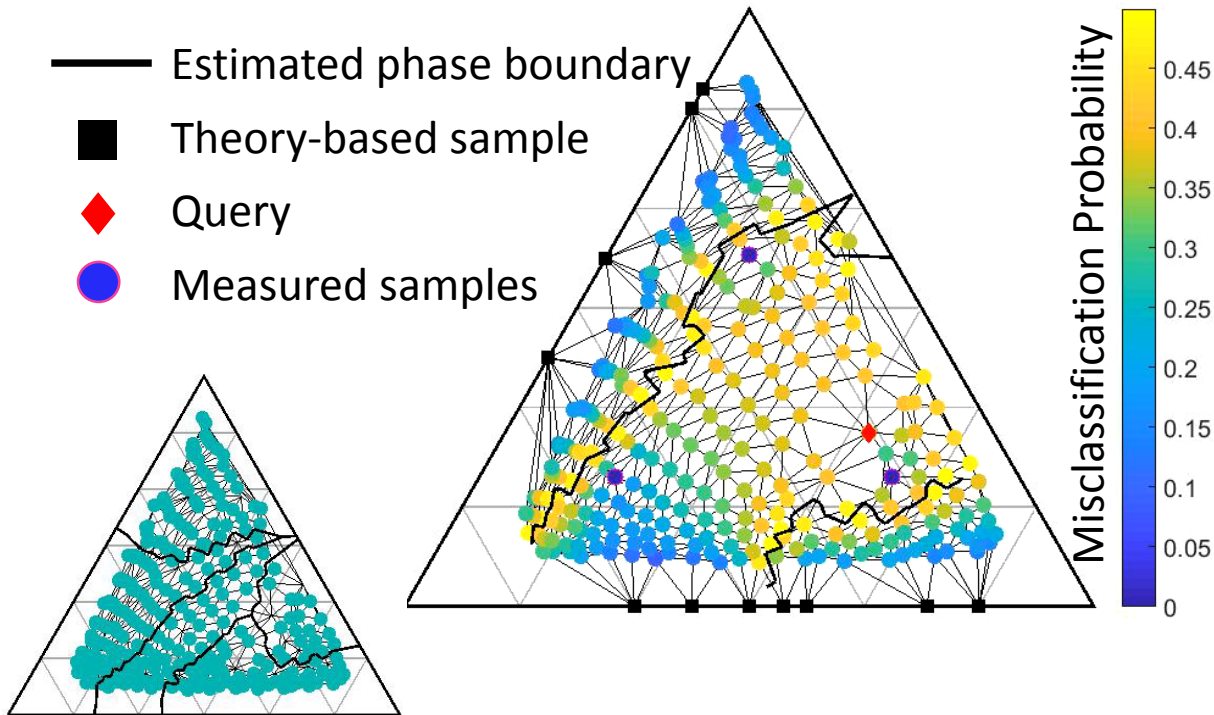
Mn-Ni-Ge library  
535 samples

# Autonomous Phase Mapping: MGI + AI

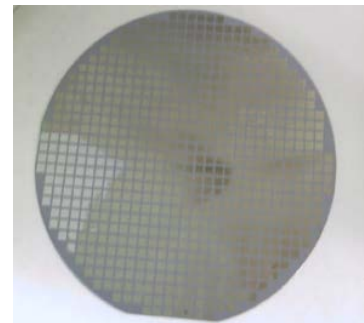
- For Optimal Experiment Design, AI needs access to prior data + physics theory
- AI Interface with Databases
  - Import pertinent data and metadata
  - AFLOW (DFT), Inorganic Crystal Structure Database (Experimental)
- Theory built in (e.g. Gibbs Phase Rule, X-ray diffraction)
  - Constraint Programming
  - Access to physics modeling software
- AI Interface with Equipment
  - X-ray diffraction systems
  - Bruker D-8, SLAC HiTp X-ray diffraction system
  - Data ingestion tools
- Automatic data storage, MDCS standards



# Autonomous Phase Mapping

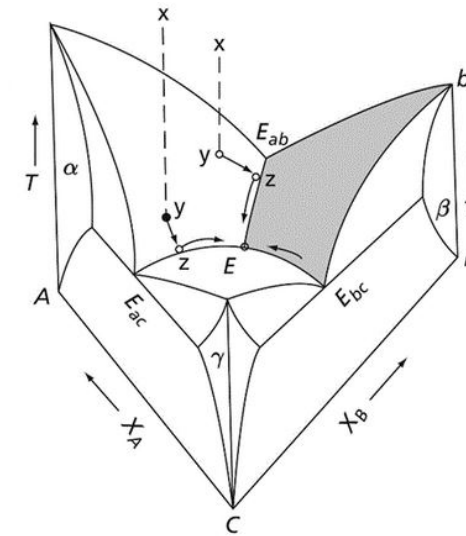
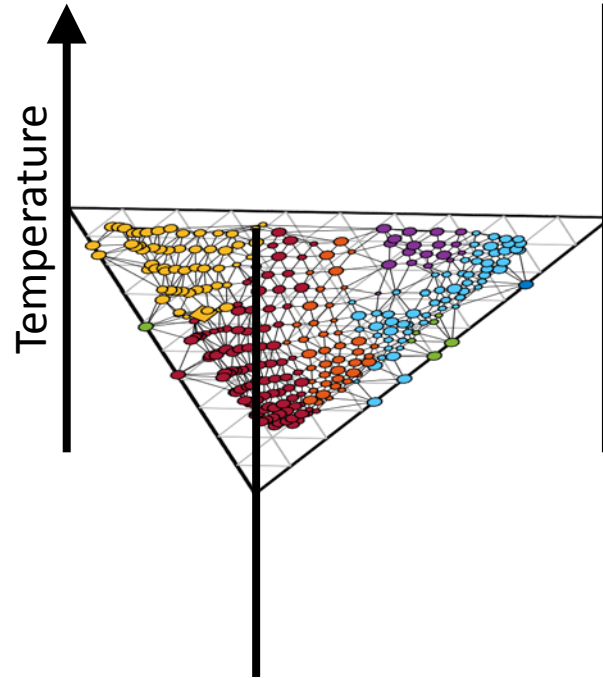
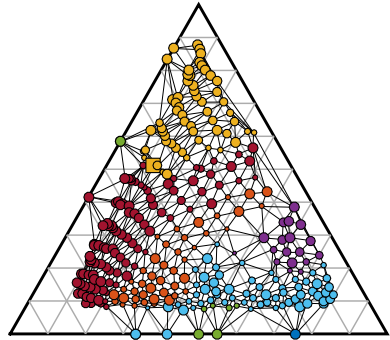


**AI is controlling X-ray diffraction systems at SLAC & in the lab!**



# Autonomous Temperature Phase Mapping

- Developing 2 systems for autonomous composition & temperature phase mapping.
- Minimum measurements for maximum knowledge.



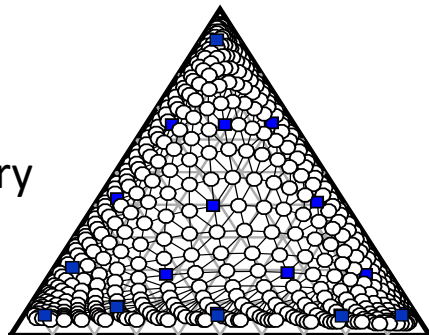
Wikipedia: ternary\_cooling.jpg



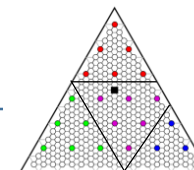
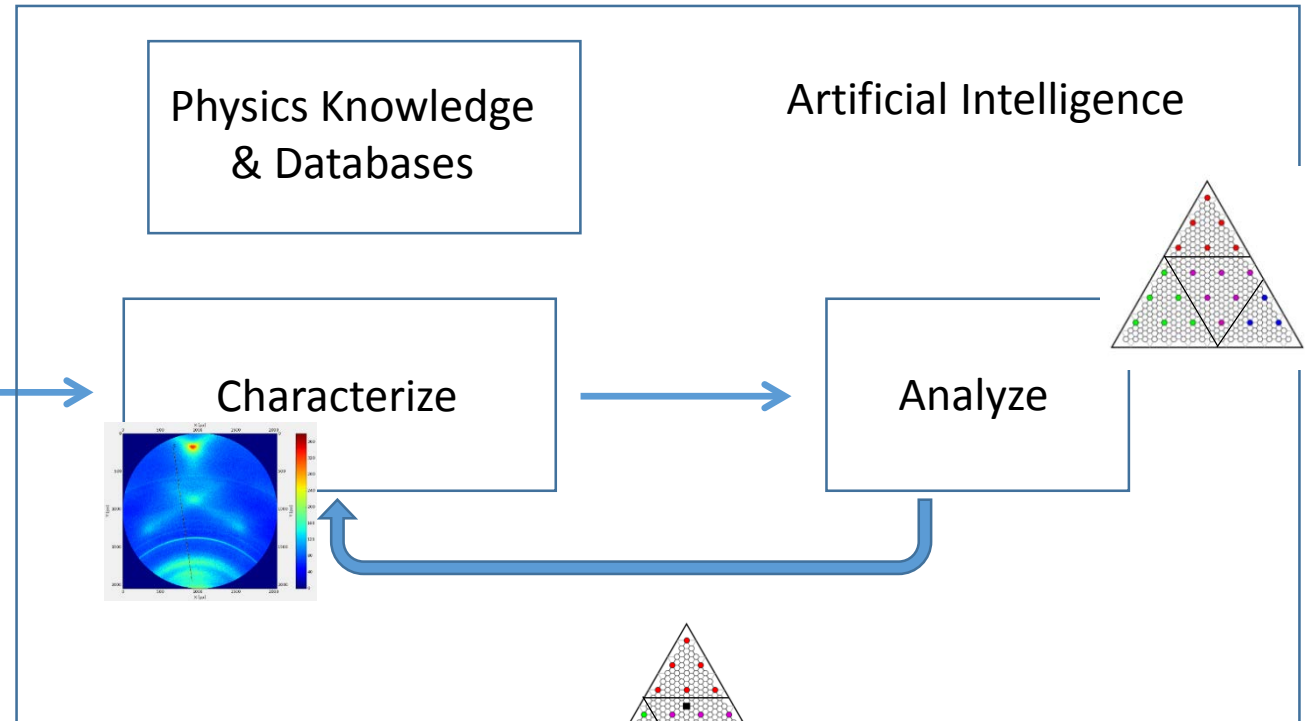
# Autonomous Metrology

- Past: AI is given a pool of samples (100s-1,000s).

Test case:  
Combinatorial Library

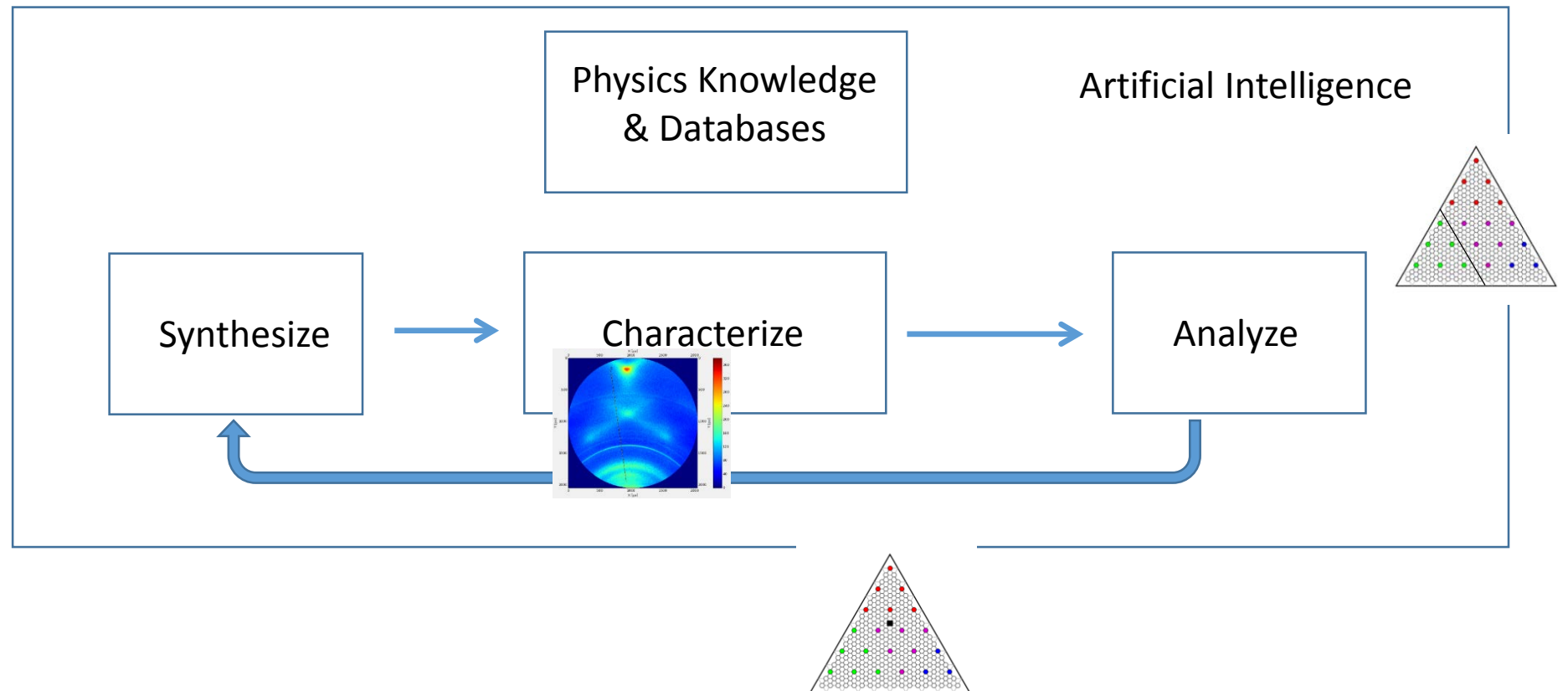


Synthesize



# Autonomous Materials Science

- Current: Place AI in control of Synthesis.



# MGI + AI: Contacts

- MGI – Jim Warren (MGI Director), [james.warren@nist.gov](mailto:james.warren@nist.gov)
- Materials Data Curation System – Zach Trautt, [zachary.trautt@nist.gov](mailto:zachary.trautt@nist.gov)
- Materials Data Facility – Ian Foster, [foster@uchicago.edu](mailto:foster@uchicago.edu)
- Materials Resource Registry – Chandler Becker, [chandler.becker@nist.gov](mailto:chandler.becker@nist.gov)
- Collaboratory – Martin Green, [martin.green@nist.gov](mailto:martin.green@nist.gov)
- Autonomous Metrology / Lab – A. Gilad Kusne, [aaron.kusne@nist.gov](mailto:aaron.kusne@nist.gov)

Questions?

