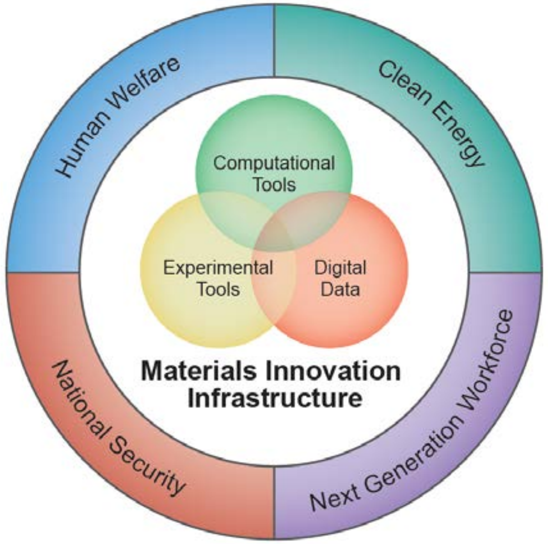
Materials Genome Initiative & Artificial Intelligence @ NIST



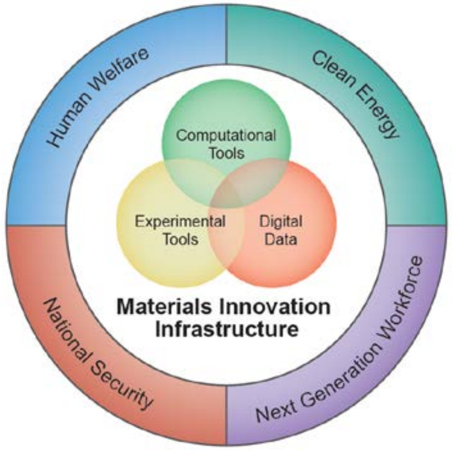
A. Gilad Kusne aaron.kusne@nist.gov

James Warren, NIST MGI Director

National Institute of Standards & Technology

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

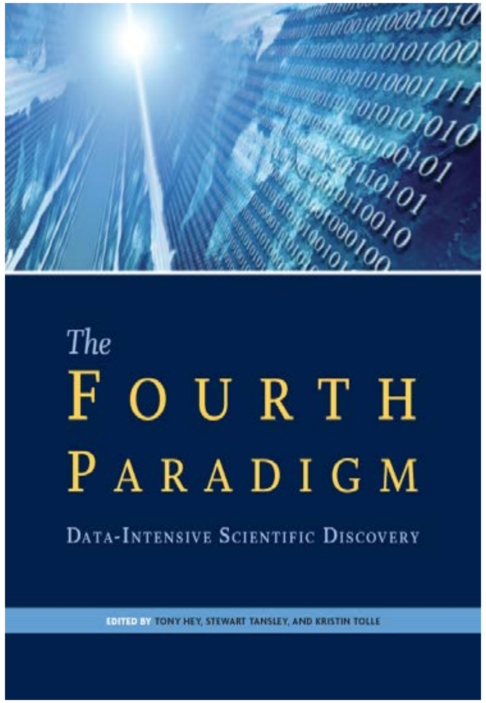
MGI + AI?

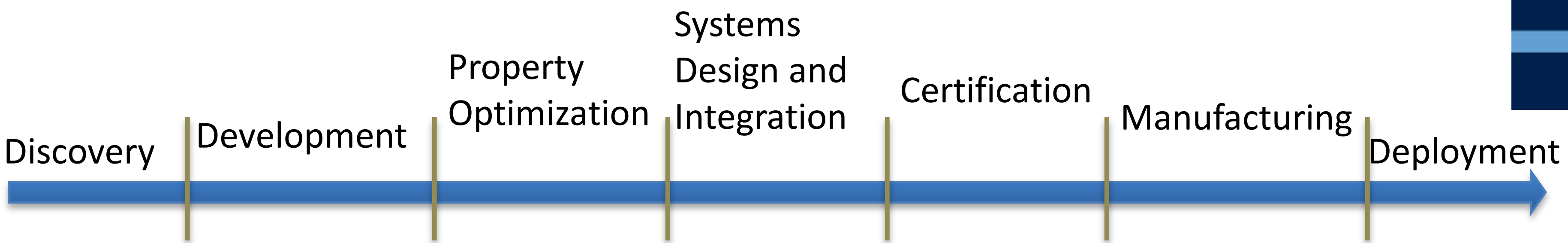


What is MGI?

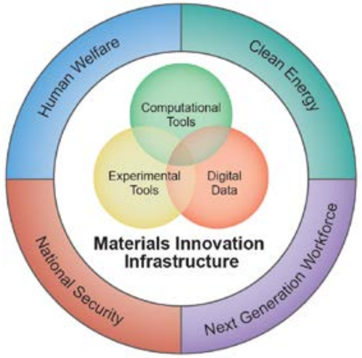
- MGI Goal – Accelerate materials development from Discovery to Deployment
- MGI – Key part: Build data science infrastructure to Enable Data Science & in particular AI
- MGI Outcome – Realizing the 4th Paradigm: Data-Intensive Scientific Discovery & Development

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





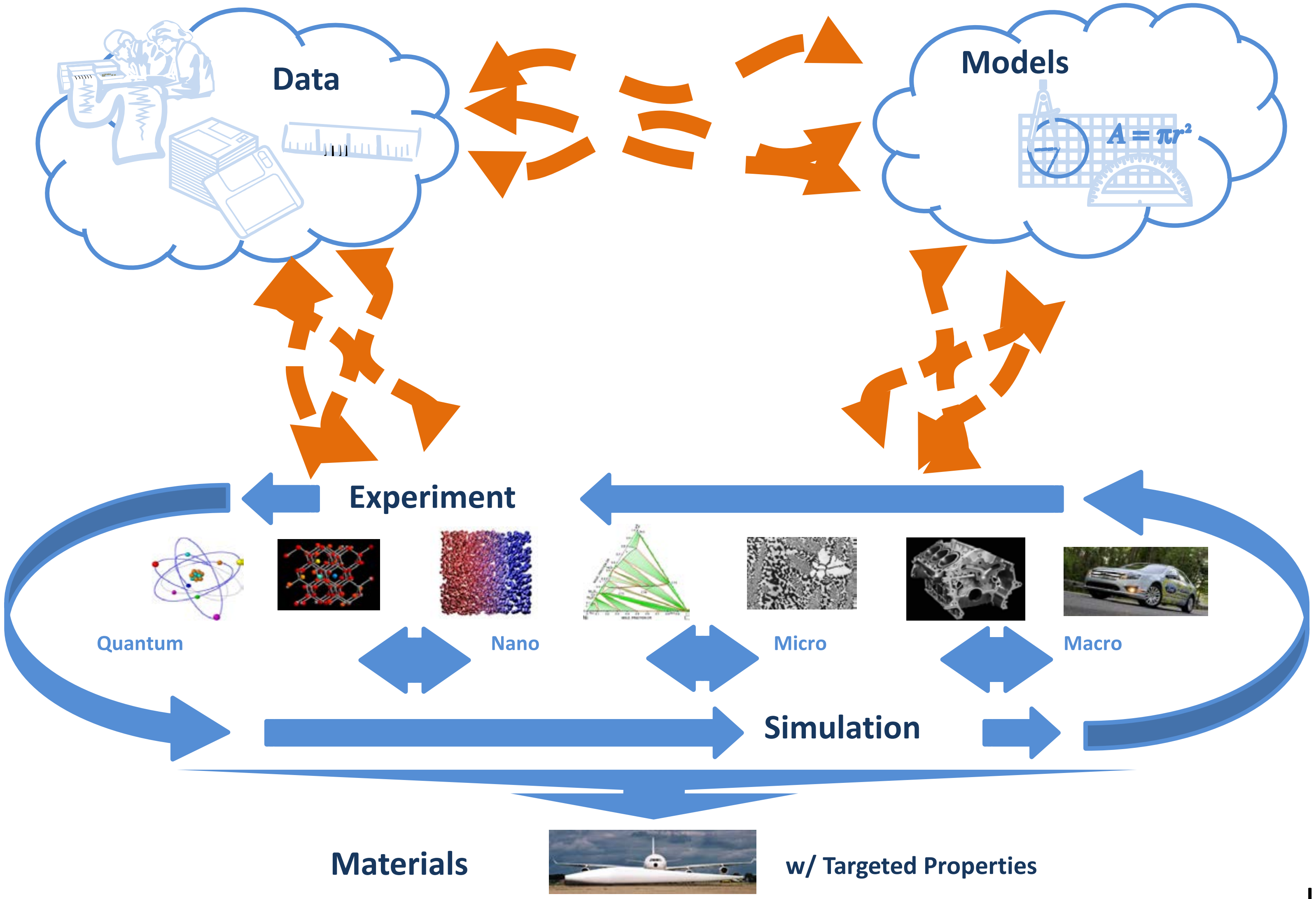
MGI + NIST



- 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

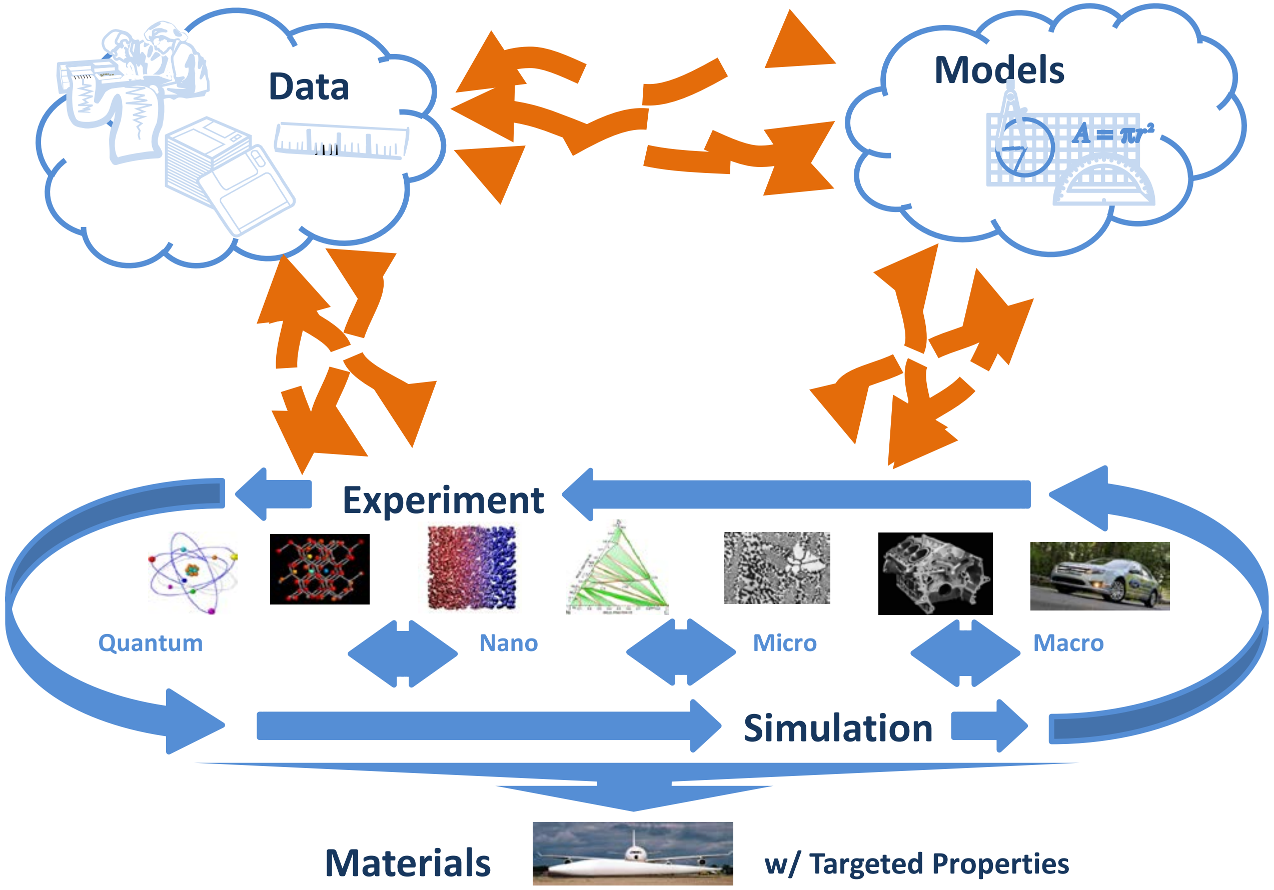
MGI: Facilitating Data Science



James Warren, NIST

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



James Warren, NIST

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 4th 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

THE UNIVERSITY OF CHICAGO

NORTHWESTERN UNIVERSITY

Argonne NATIONAL LABORATORY

QuesTek INNOVATIONS LLC

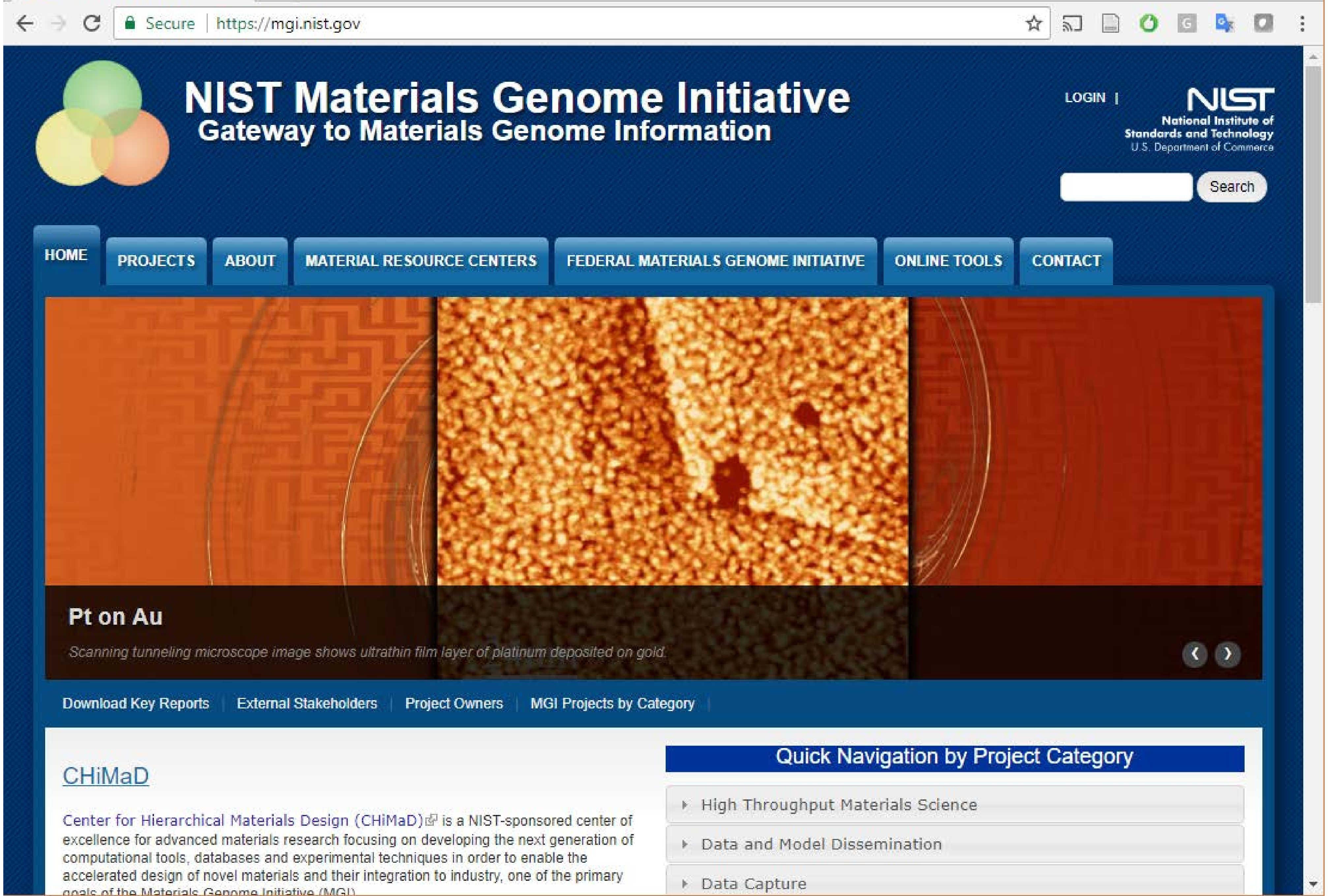
ASM INTERNATIONAL

FAYETTEVILLE STATE UNIVERSITY

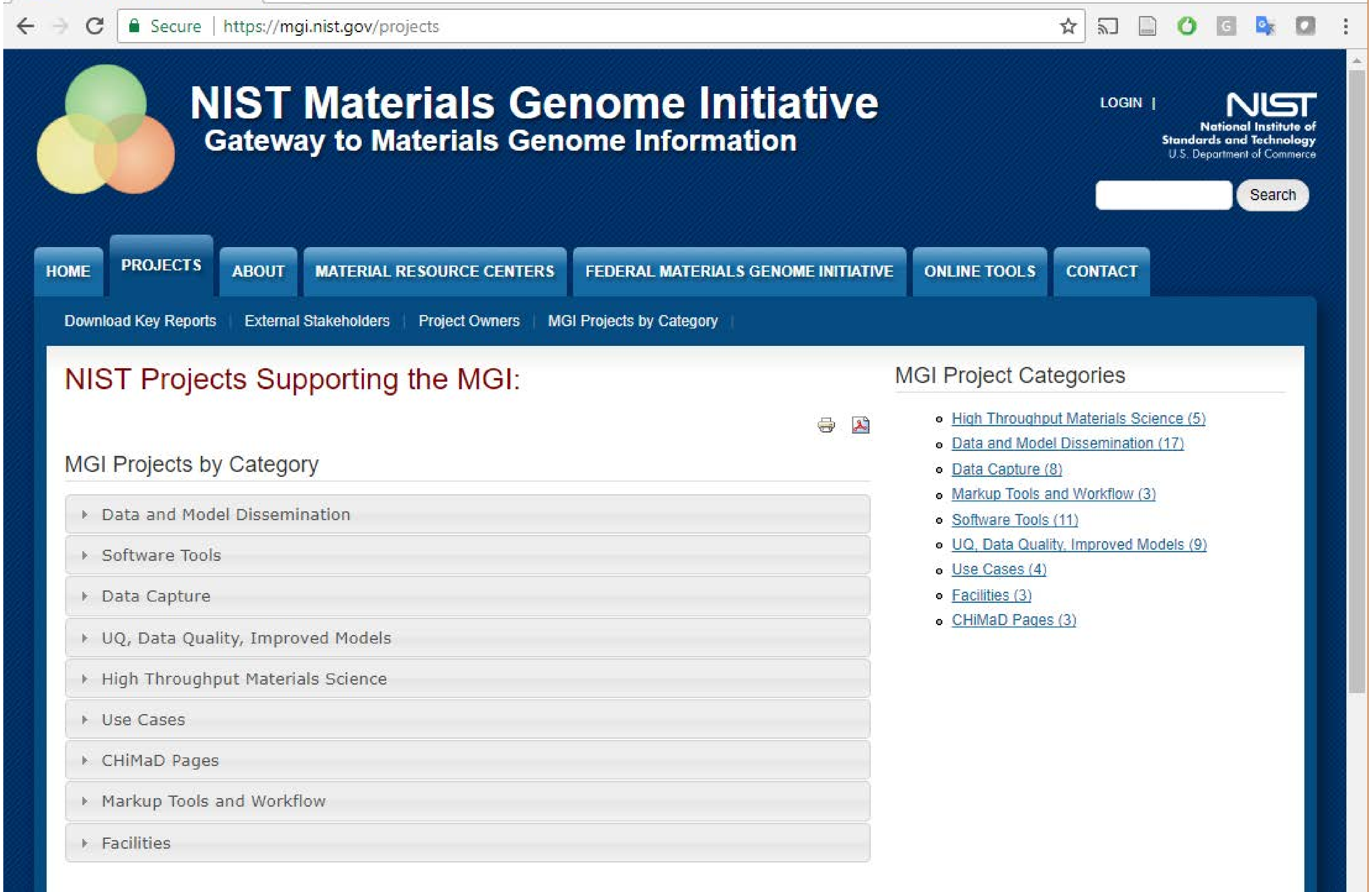
Peter Voorhees, Gregory Olson | Begum Gulsoy *Northwestern University*
Juan De Pablo | *University of Chicago*

CHiMaD

https://mgi.nist.gov



https://mgi.nist.gov



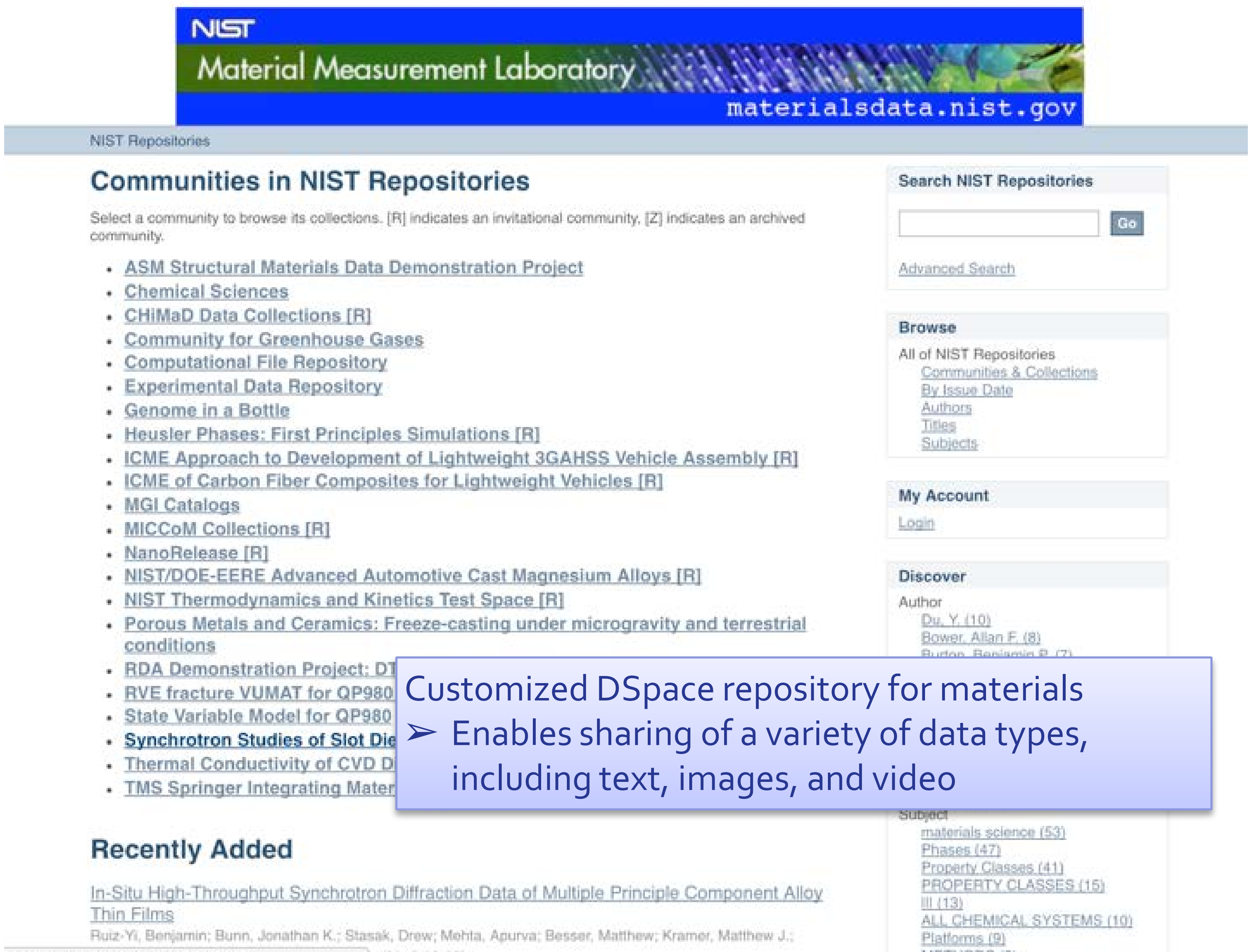
MGI @ NIST

- 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

https://materialsdata.nist.gov/

- Bank it



Customized DSpace repository for materials

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

Login

NIST Material Measurement Laboratory

materialsdata.nist.gov

NIST Repositories → Experimental Data Repository → Diffusion Data → View Item

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, Kattner 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

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.

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. Peluso, Rebecca Casey, Mitch Hammond and Karen Denike for their experimental support.

Data files

Files in this item

Name: R95-R88-expsimul.txt View/Open
Size: 25.32Kb
Format: Text file
Description:
Experimental and simulate composition profiles for the R88/R95 diffusion couple at 1150 C for 1000 h

Name: r95r88-1000h-labe ... View/Open
Size: 58.76Mb
Format: TIFF image
Description:
Micrograph of R95/R88 diffusion couple after 1000 h at 1150 C

The following license files are associated with this item:

- Creative Commons

Offer licenses with attribution 3.0

This item appears in the following Collection(s)

- Diffusion Data

Related items

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

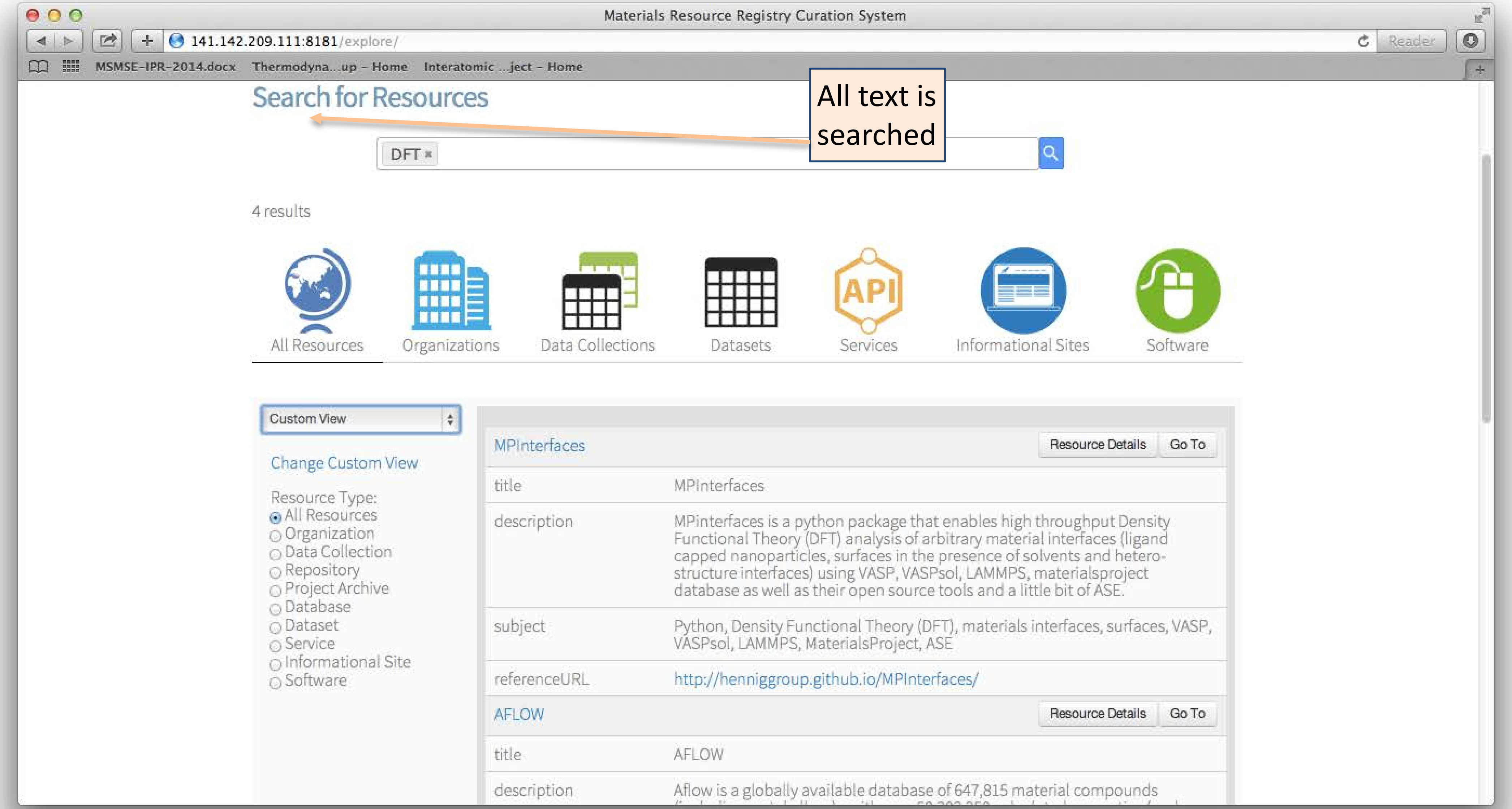
Further Studies on the Nickel-Aluminum System. I. The β-Ni2Al3 Phase Fields
Taylor, A; Doyle, N.J. (1972-01-31)
New lattice parameter and density results have been obtained for alloys in the fl-NiA1 and 6-Ni2A13 phase fields of the nickel-aluminum system. The lattice parameter of the fl-NiAl phase (CsCl-type) 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

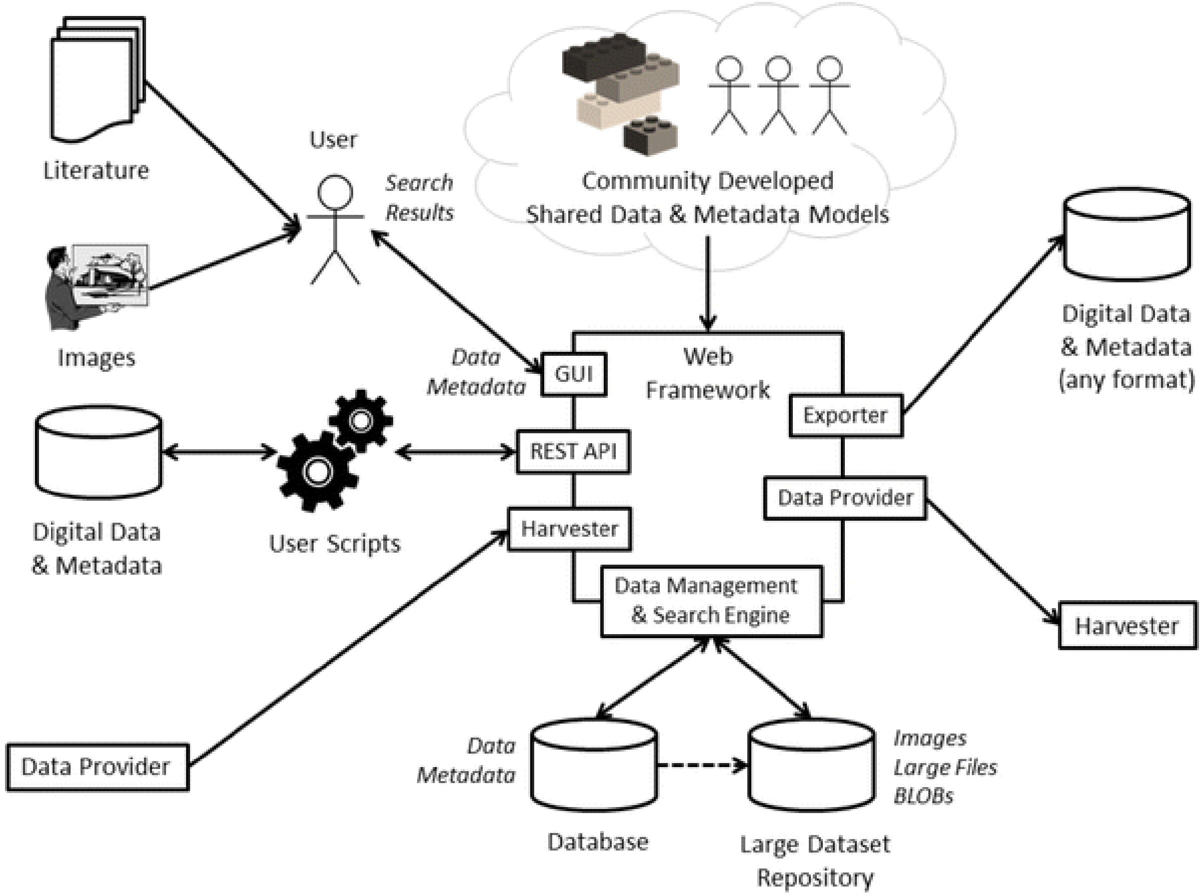
Search for resources: Materials Resource Registry

- Find it



Materials 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

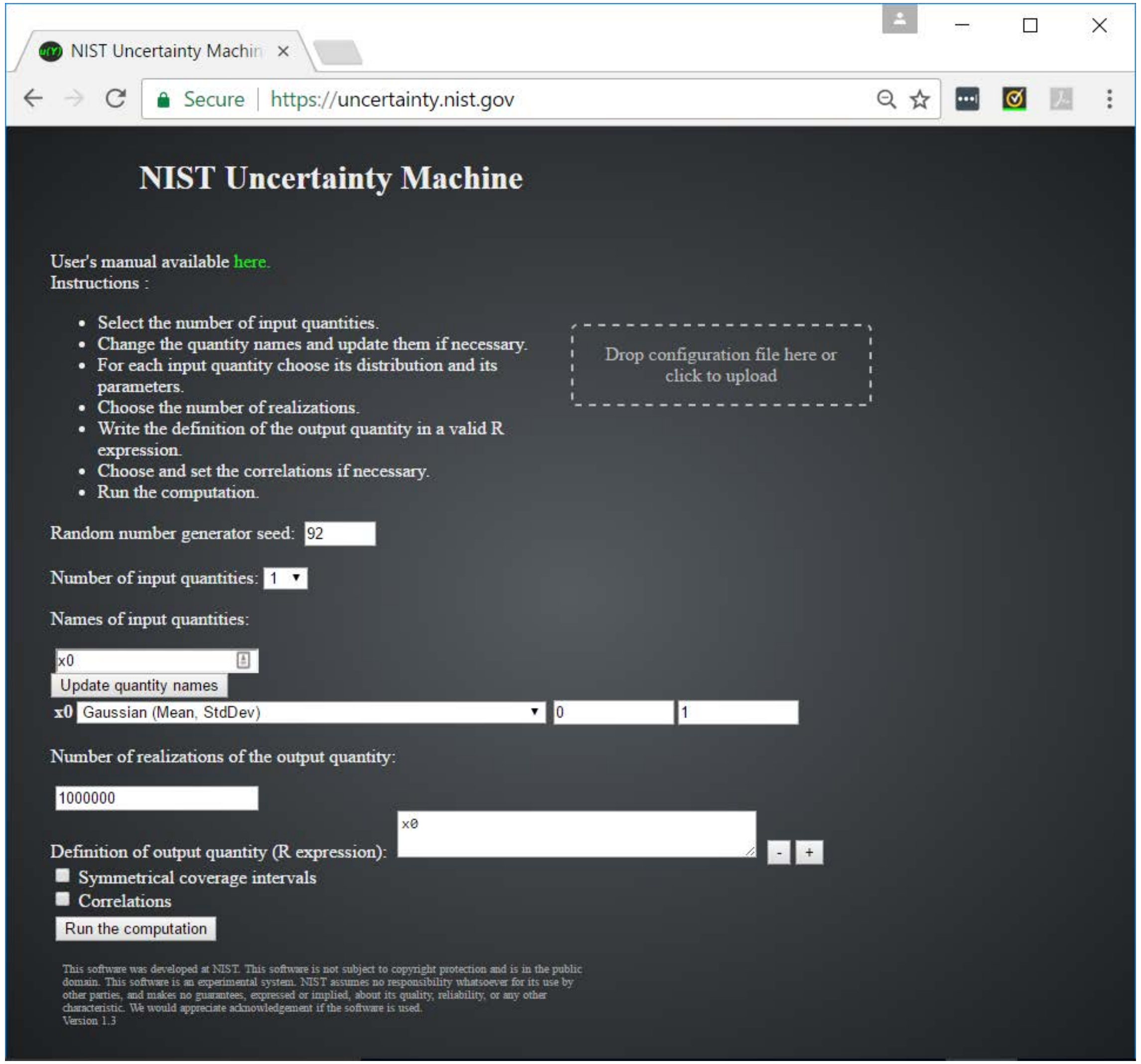


Dima (JOM 2016) Informatics Infrastructure for the Materials Genome Initiative.

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.



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

Materials Data Facility

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

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

Michael Ondrejcek[3], Kenton McHenry[3], John Towns[3]
University of Chicago[1], Argonne National Laboratory[2], University of Illinois at Urbana-Champaign[3]

materialsdatafacility.org
globus.org

Materials Genome Initiative

NIST National Institute of Standards and Technology U.S. Department of Commerce · CHiMaD Center for Hierarchical Materials Design · The University of Chicago · Northwestern University · Illinois University of Illinois at Urbana Champaign · Argonne National Laboratory · National Data Service Participant

STREAMLINE AND AUTOMATE

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

Publication

REST APIs

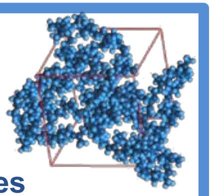
Discovery

- 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 χ and other property values
- Includes 375 polymer-polymer values and 1,014 polymer-solvent values

NIST National Institute of Standards and Technology U.S. Department of Commerce | CHiMaD Center for Hierarchical Materials Design | The University of Chicago | Northwestern University | Illinois University of Illinois at Urbana-Champaign | Argonne National Laboratory

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[1], Andriy Zakutayev[2],
Martin L. Green[1], John Perkins[2]

[1]National Institute of Standards and Technology (NIST)
[2]National Renewable Energy Laboratory (NREL)

martin.green@nist.gov

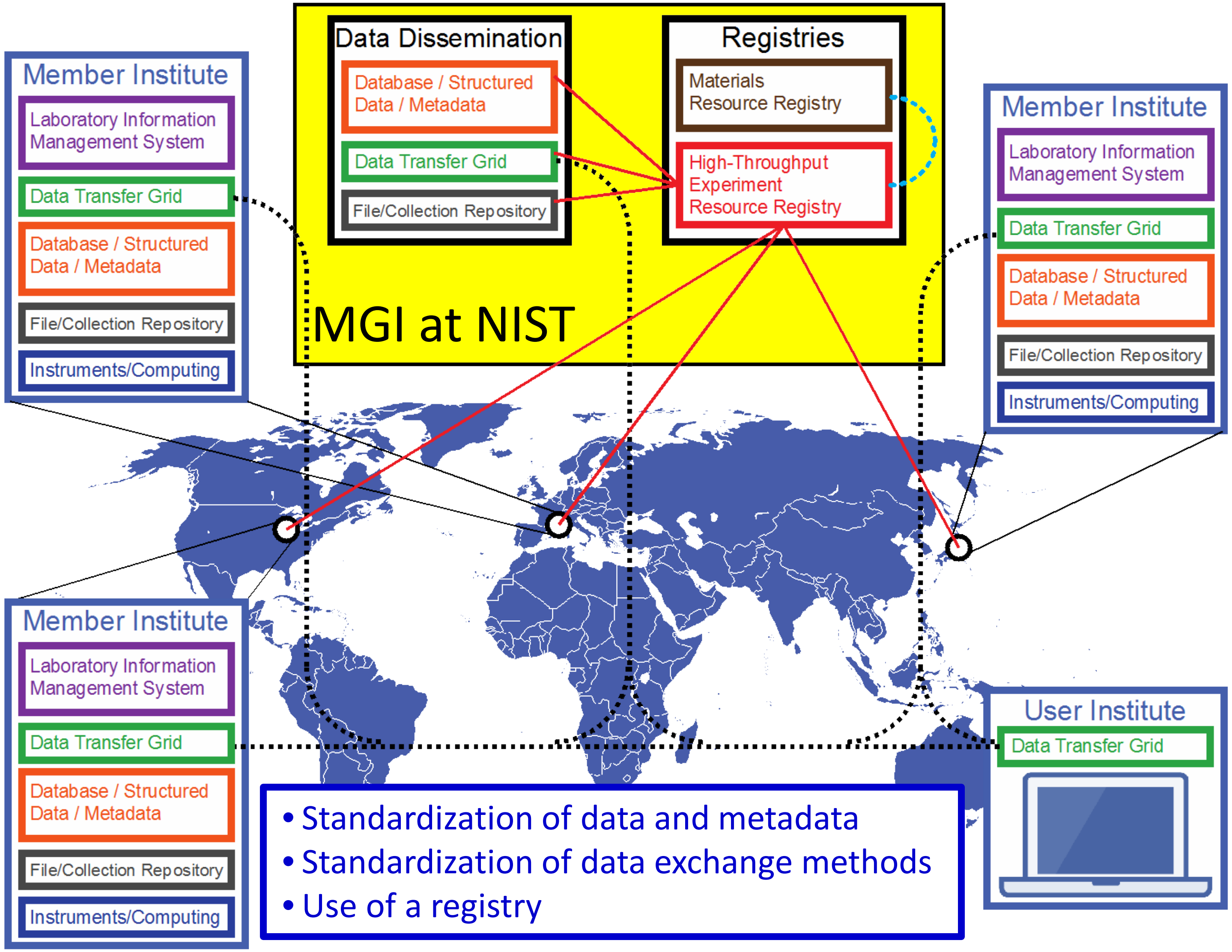
Collaboratory

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."

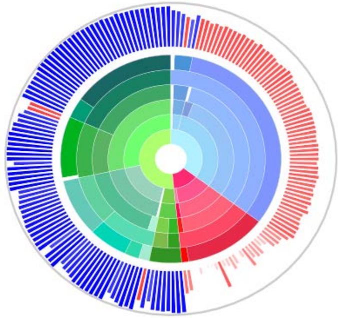
- The HTEMC would consist of:
  - An integrated, delocalized network of high-throughput synthesis and characterization tools
  - A best-in-class materials data management platform, consisting of NIST (and other) software

Platform Integration Model



- Standardization of data and metadata
- Standardization of data exchange methods
- Use of a registry

Education



Annual Machine Learning for Materials Research: Bootcamp and Workshop

- Host: University of Maryland — **https://nanocenter.umd.edu/events/mlmr/**
- Dates: **TBD (~June / July)**
- Location: University of Maryland, College Park, MD

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:

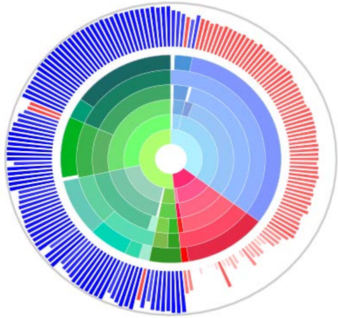
- Identifying important features in complex/high dimensional data
- Visualizing high dimensional data to facilitate user analysis.
- Identifying the fabrication ‘descriptors’ that best predict variance in functional properties.
- Quantifying similarities between materials using complex/high dimensional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

NIST | UNIVERSITY OF MARYLAND | STANFORD UNIVERSITY | OAK RIDGE National Laboratory | JOHNS HOPKINS UNIVERSITY



Annual Machine Learning for Materials Research: Bootcamp and Workshop

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

MACHINE LEARNING FOR MATERIALS RESEARCH: BOOTCAMP		
DATA FUNDAMENTALS DATA PREPROCESSING TUE 6/28	SUPERVISED LEARNING THEORY, DATA, ALGORITHMS WED 6/29	UNSUPERVISED LEARNING THEORY & ALGORITHMS THU 6/30
Filtering: Noise Smoothing Background Subtraction Feature Extraction ▪ Cross-correlation Wavelets ▪ Edges ▪ Closed Boundaries ▪ Shapes	Data Handling ▪ Cross Validation ▪ Prediction Algorithms ▪ Regularized Least Squared ▪ Support Vector Machines ▪ Neural Networks ▪ Decision Trees & Ensemble Learning ▪ Genetic Programming	Theory Similarity Measures Latent Variable Analysis Spectral Unmixing Matrix Factorization Clustering

MACHINE LEARNING FOR MATERIALS RESEARCH: WORKSHOP	
WORKSHOP SESSION 1 THU 6/30 EVENING	WORKSHOP SESSION 2 FRI 7/1 MORNING

Organizers

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

Alexei Belianinov
Oak Ridge National Laboratory
Center for Nanophase Materials Sciences

Daniel Samarov
National Institute of Standards and Technology
Information Technology Laboratory

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

NIST UNIVERSITY OF MARYLAND STANFORD UNIVERSITY OAK RIDGE National Laboratory JOHNS HOPKINS UNIVERSITY

MGI Examples

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

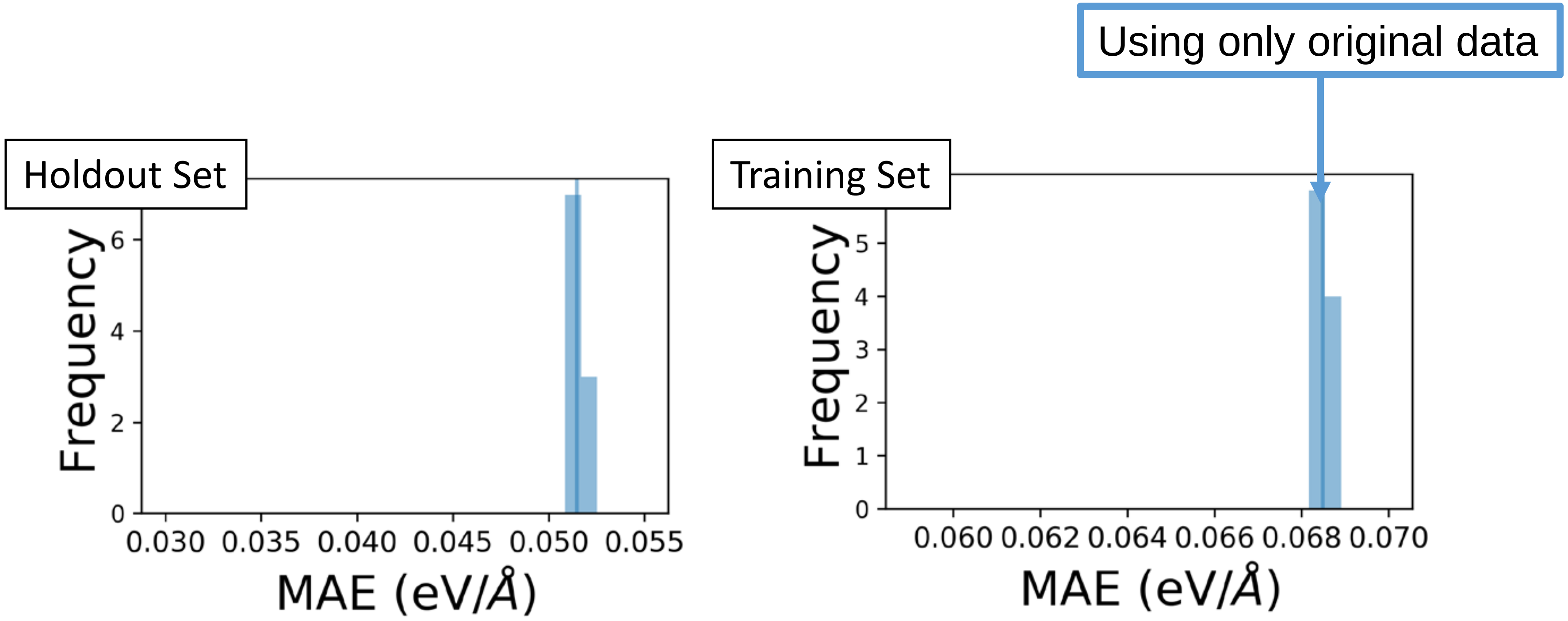
Building a machine learning model using MDF

Example: Building force-field potentials from different datasets

Data resources: 3 DFT datasets with Aluminum data

1 dataset from khazana.uconn.edu, 2 datasets from materialsdata.nist.gov

Result: Improved performance by integrating data sources



MAE – Mean absolute error

Method: Botu *et al.* JPCC. (2017)

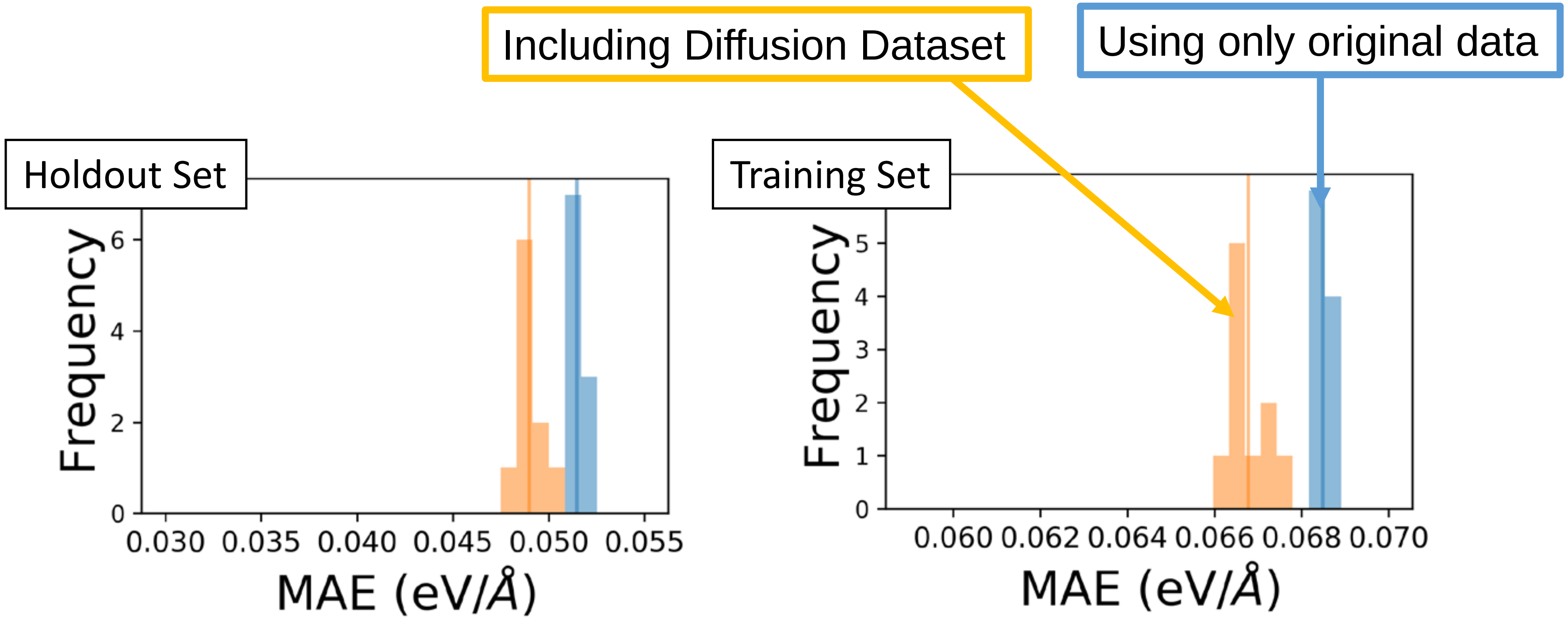
Building a machine learning model using MDF

Example: Building force-field potentials from different datasets

Data resources: 3 DFT datasets with Aluminum data

1 dataset from khazana.uconn.edu, 2 datasets from materialsdata.nist.gov

Result: Improved performance by integrating data sources



Method: Botu *et al.* JPCC. (2017)

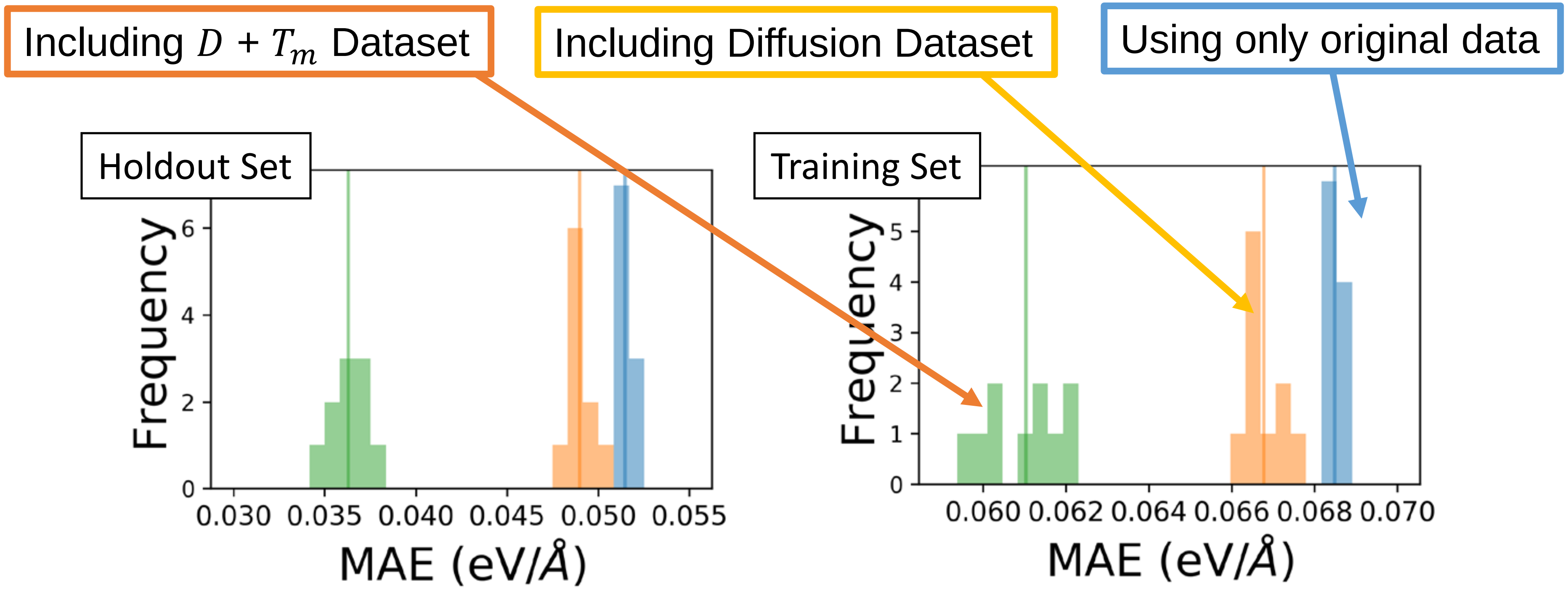
Building a machine learning model using MDF

Example: Building force-field potentials from different datasets

Data resources: 3 DFT datasets with Aluminum data

1 dataset from khazana.uconn.edu, 2 datasets from materialsdata.nist.gov

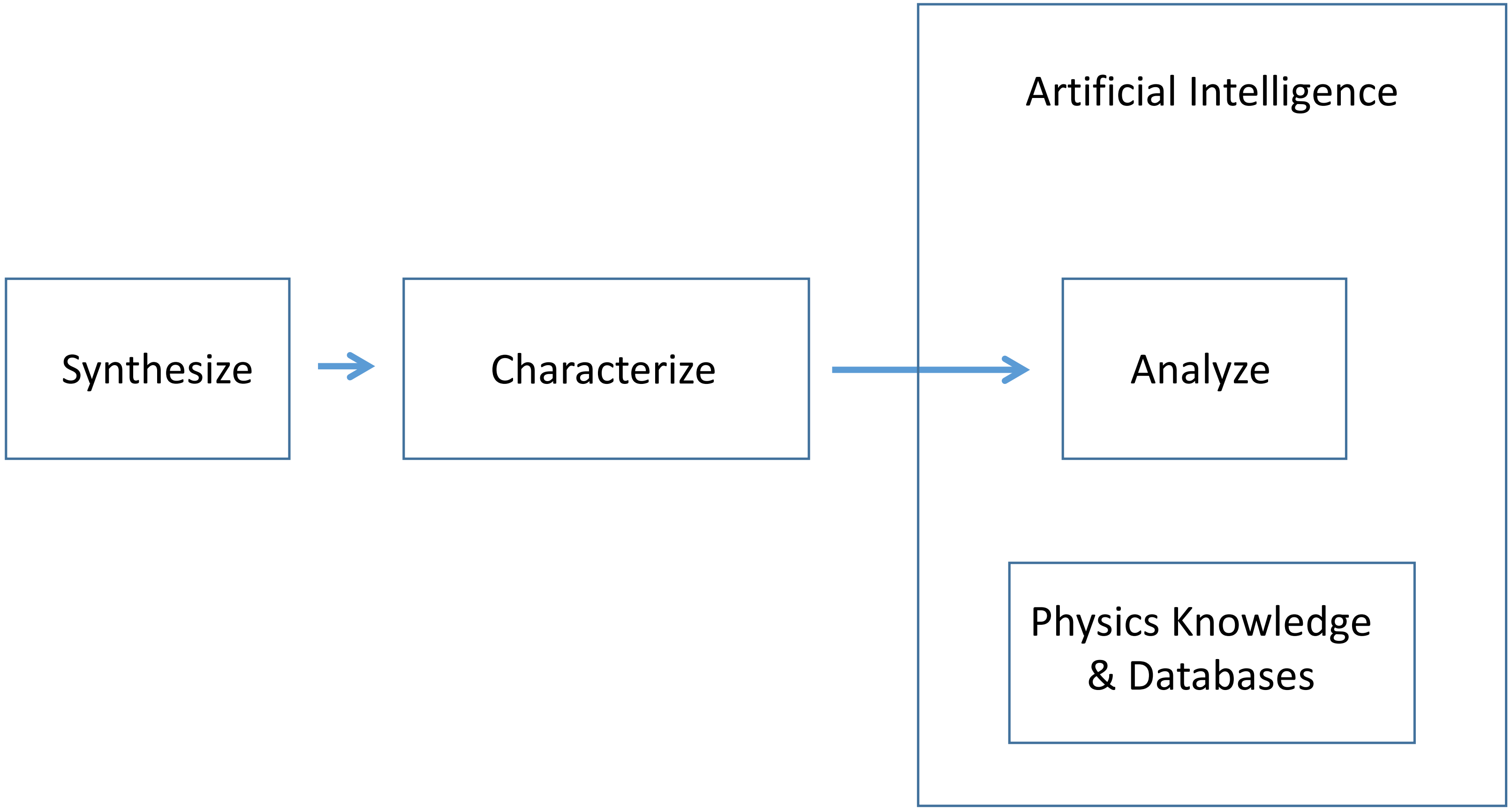
Result: Improved performance by integrating data sources



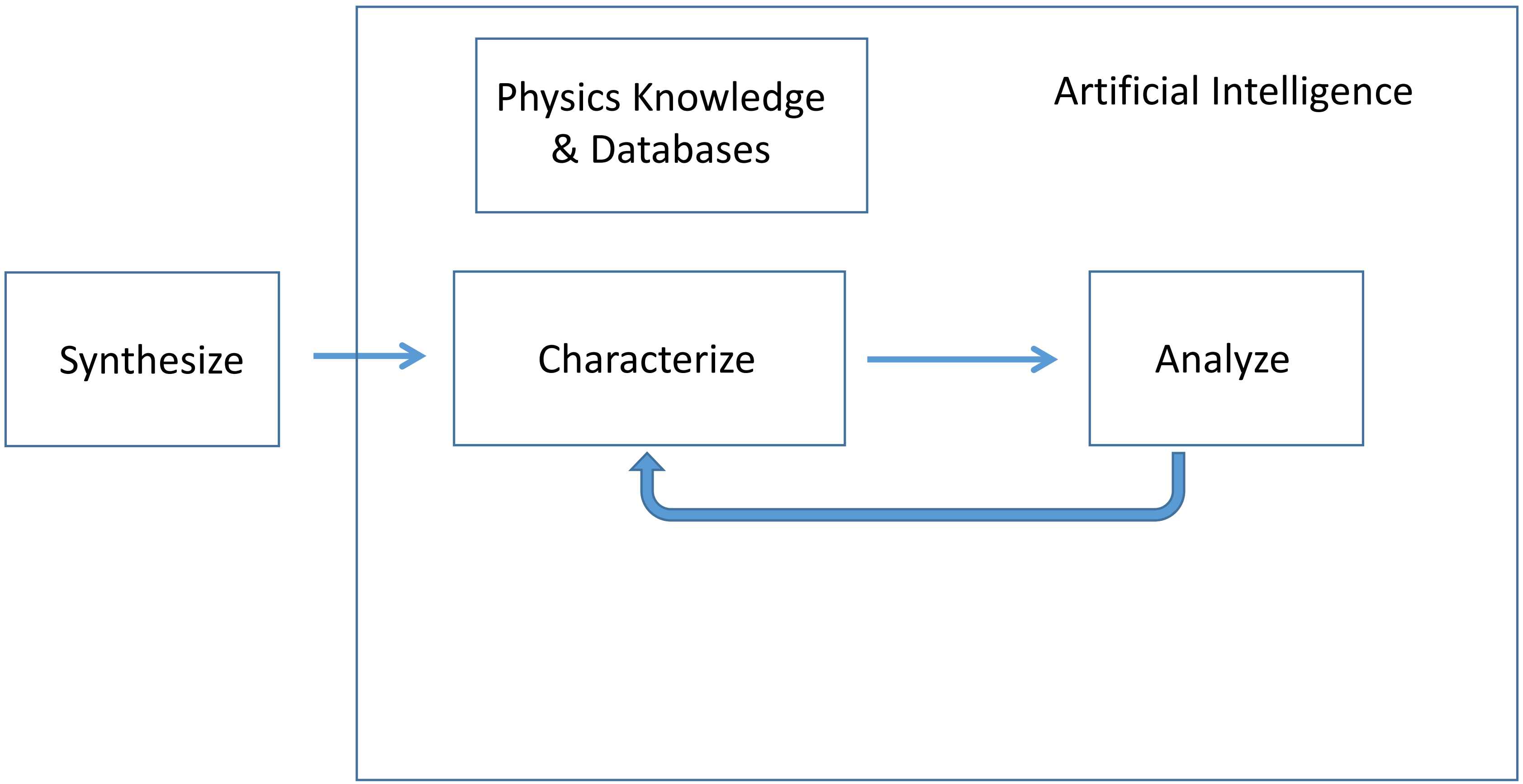
Better performance in original application: No new DFT calculations

Method: Botu *et al.* JPCC. (2017)

AI for Analysis



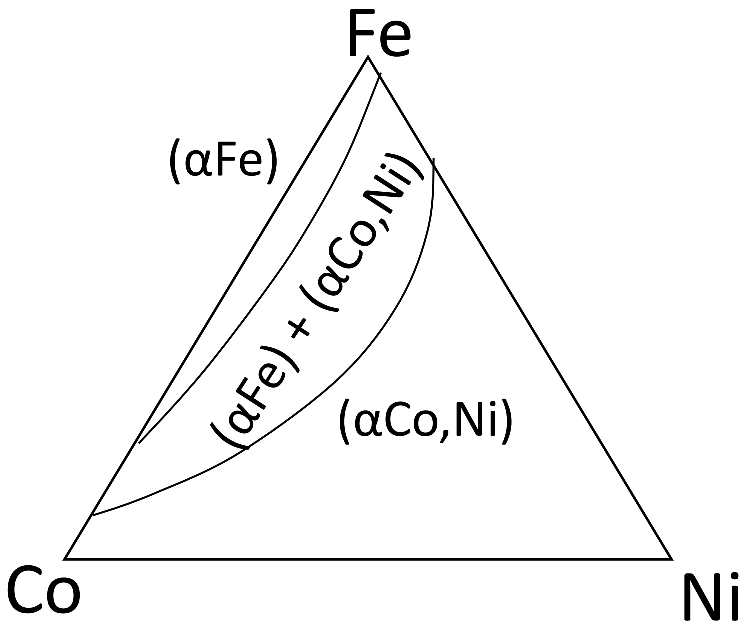
Autonomous Metrology



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.

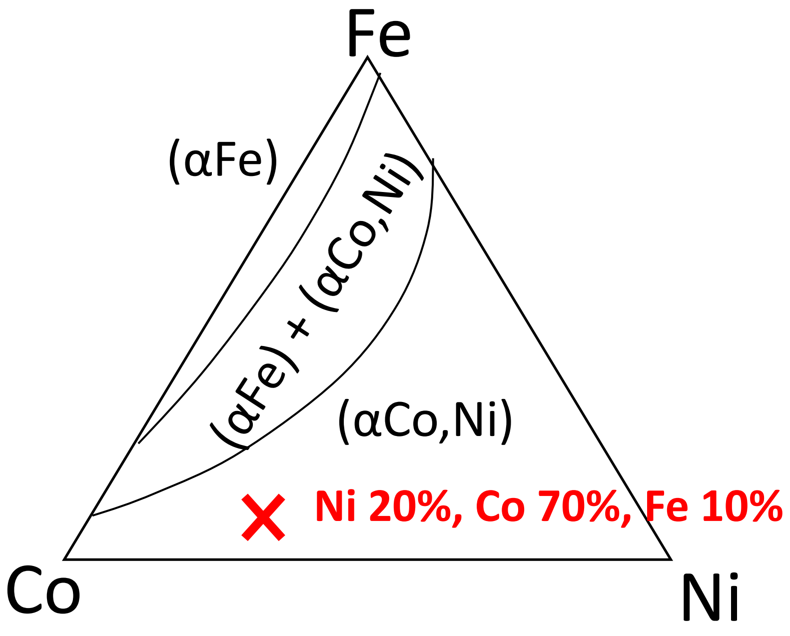


Phase Equilibria in Iron Ternary Alloys (1988) #60

Structural Phase Mapping: Edisonian Approach

Traditional / Edisonian Approach:

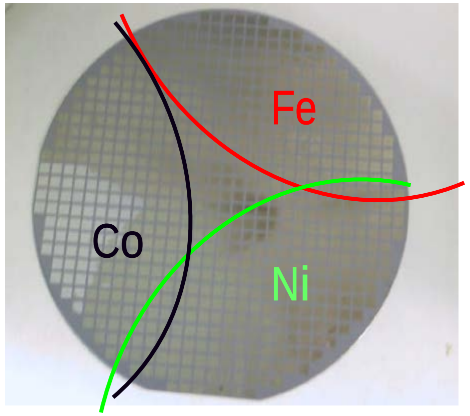
- Fabricate sample
- Measure structure
- Point placed on phase diagram
- Repeat
- This process takes years.



Phase Equilibria in Iron Ternary Alloys (1988) #60

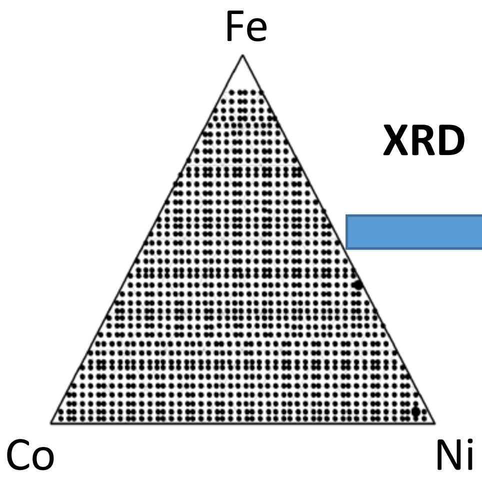
Phase Mapping: High-Throughput Approach

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



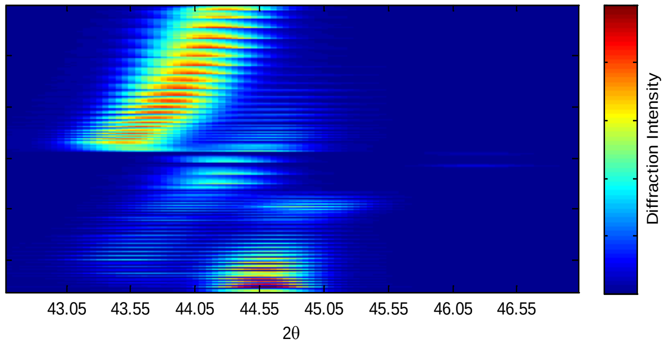
APL Materials (2016)
Composition–structure–property mapping in high-throughput experiments: Turning data into knowledge

Combi Library for Ternary Spread



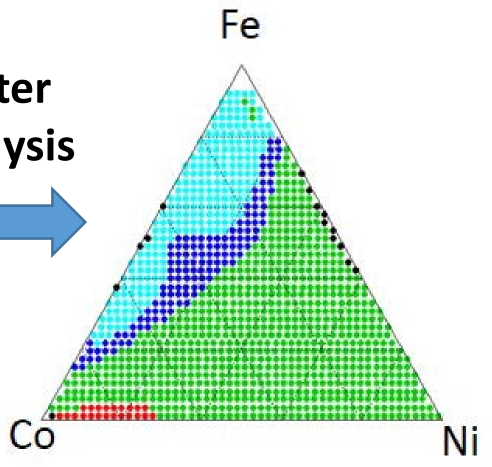
XRD

Diffraction Patterns



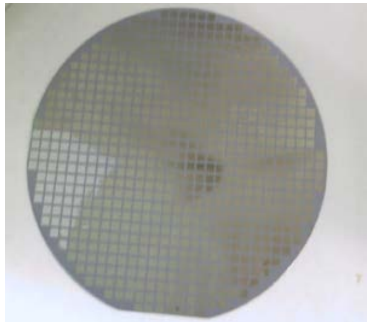
Cluster Analysis

Estimated Phase Map



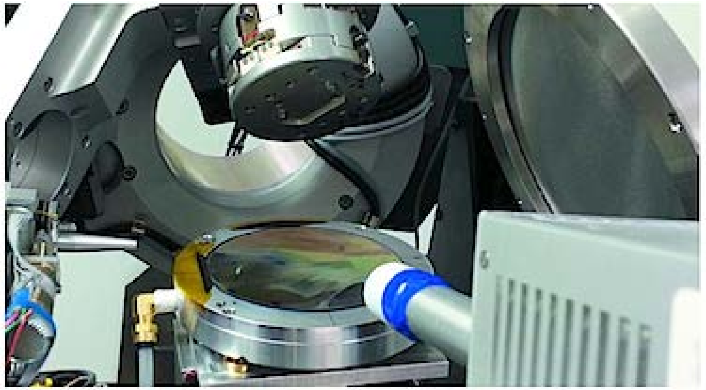
Phase Mapping: High-Throughput Approach

- 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

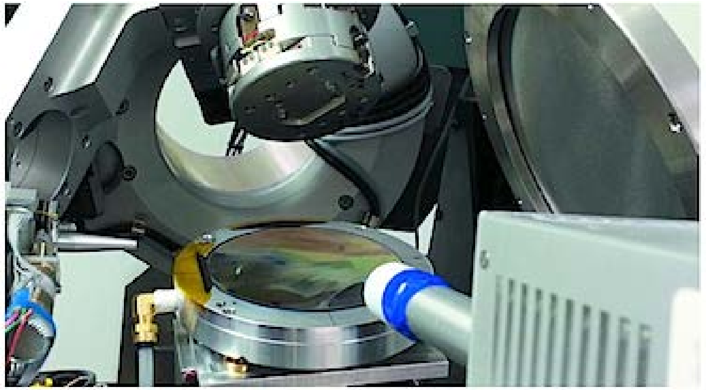
Autonomous Metrology: Motivation

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

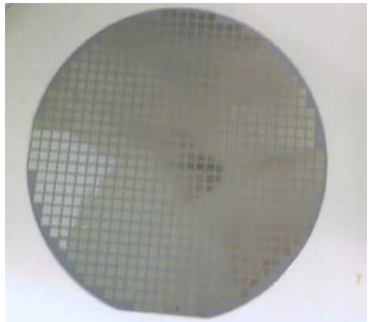
Instead of measuring all the samples, measure only the ones that count -> AI for optimal experiment design

- 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!



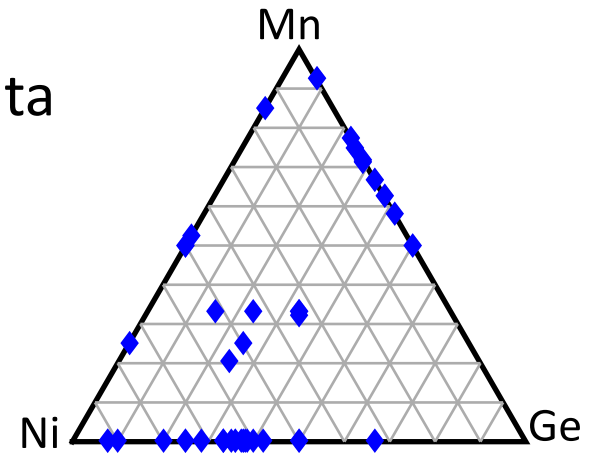
Stanford Synchrotron Radiation Lightsource
30 seconds per sample
4.5 hours



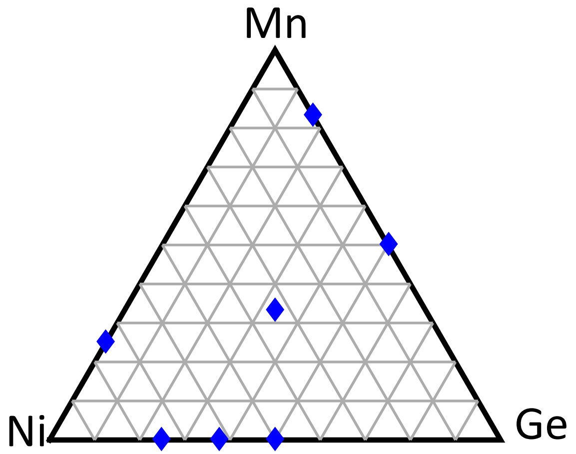
Mn-Ni-Ge library
535 samples

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

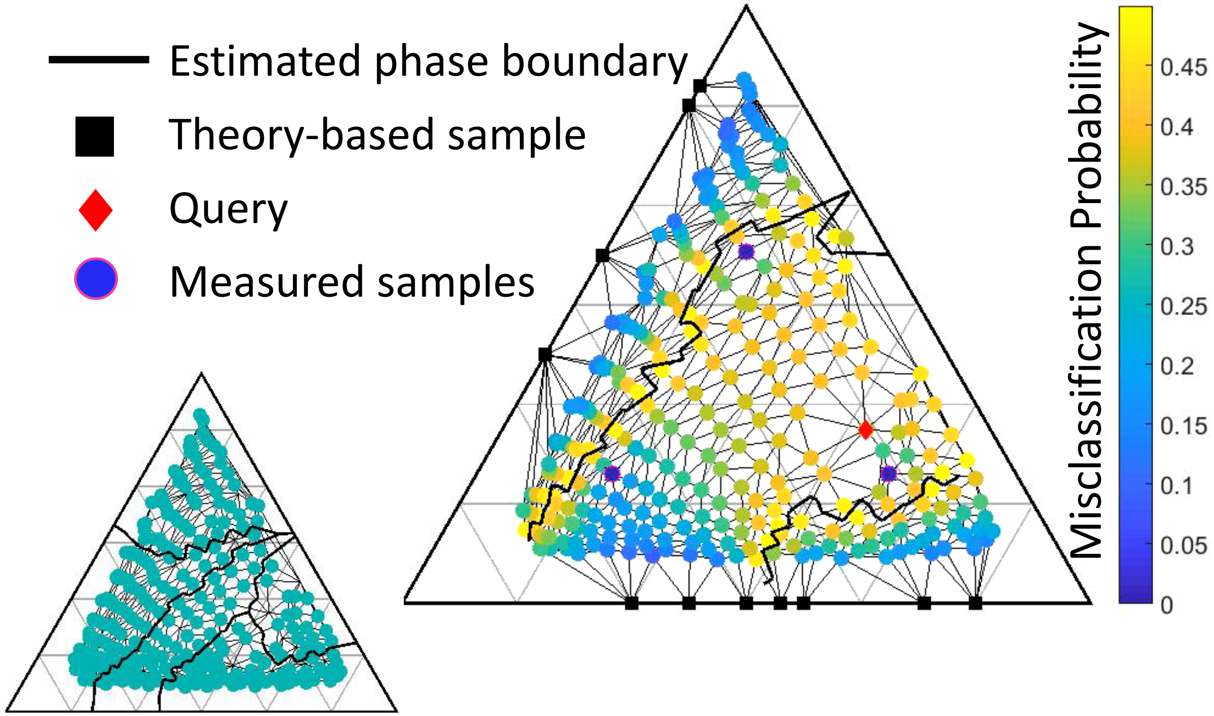


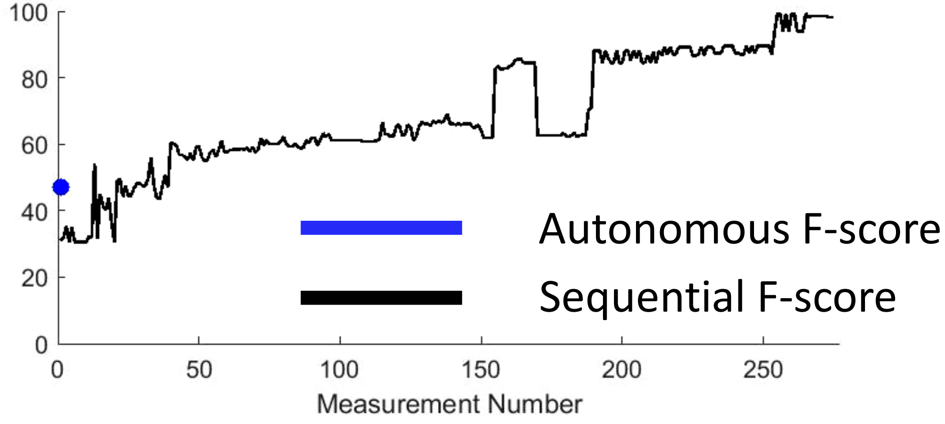
Data from ICSD

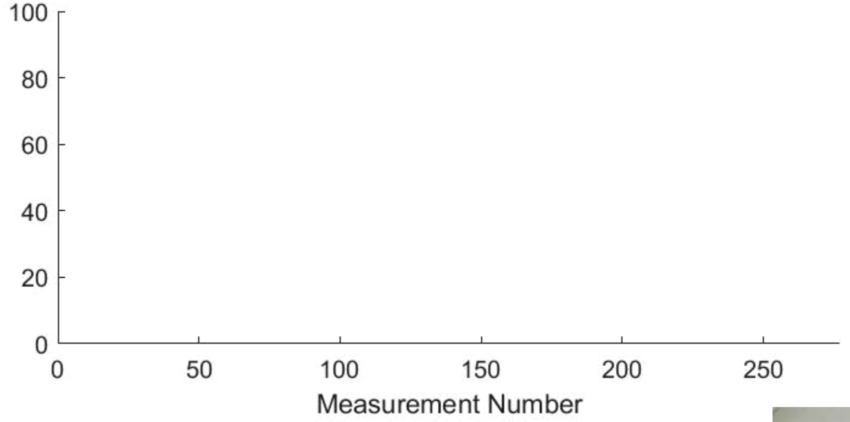


Data from AFLOW

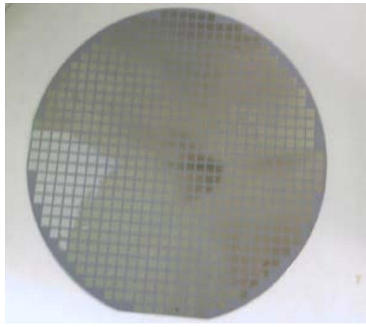
Autonomous Phase Mapping





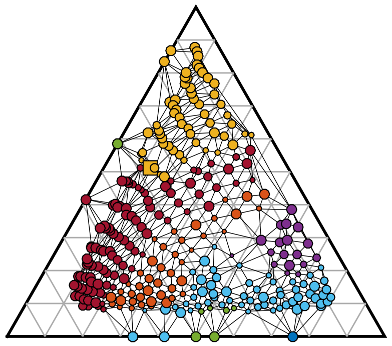


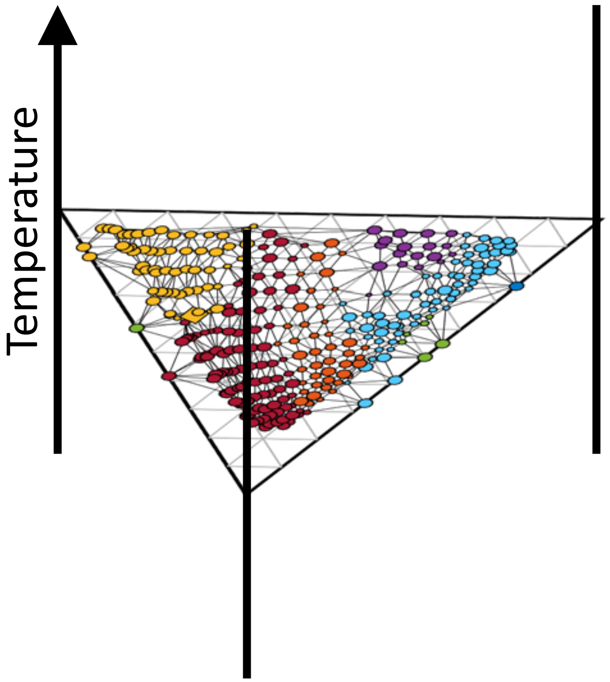
AI is controlling X-ray diffraction systems at SLAC & in the lab!

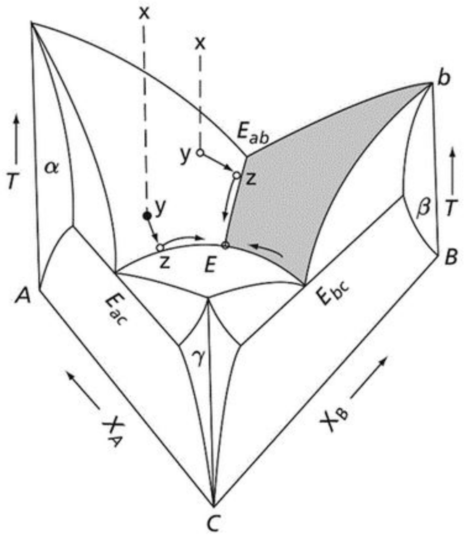


Autonomous Temperature Phase Mapping

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



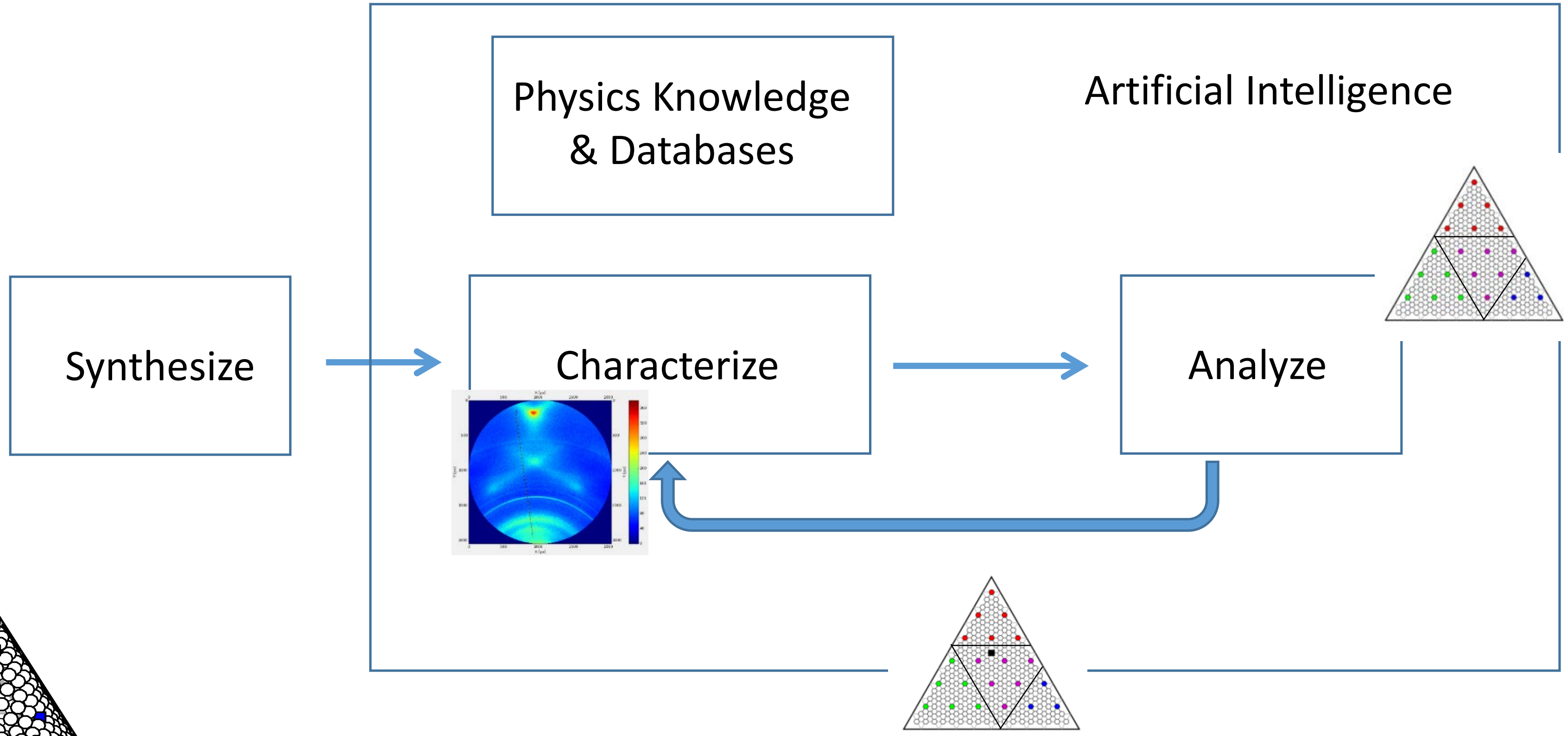




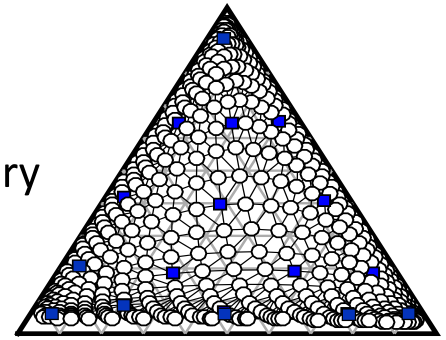
Wikipedia: ternary_cooling.jpg

Autonomous Metrology

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

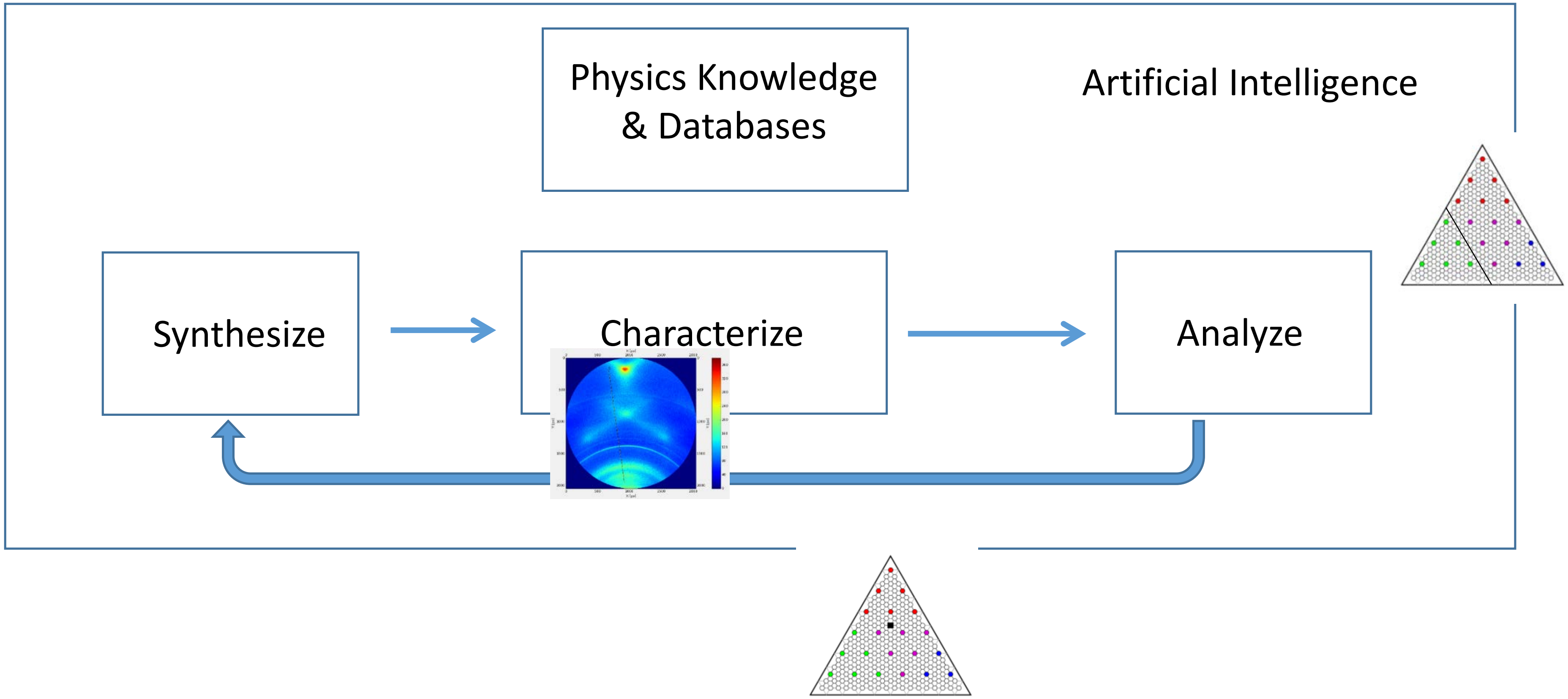


Test case:
Combinatorial Library

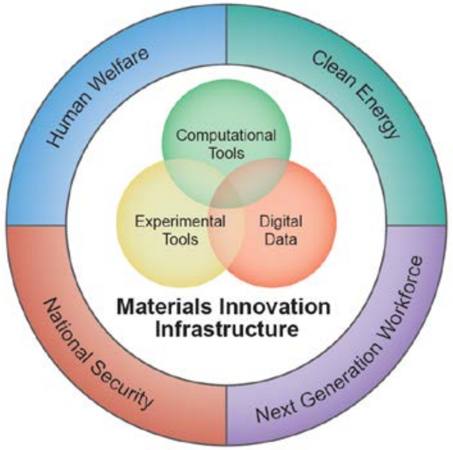


Autonomous Materials Science

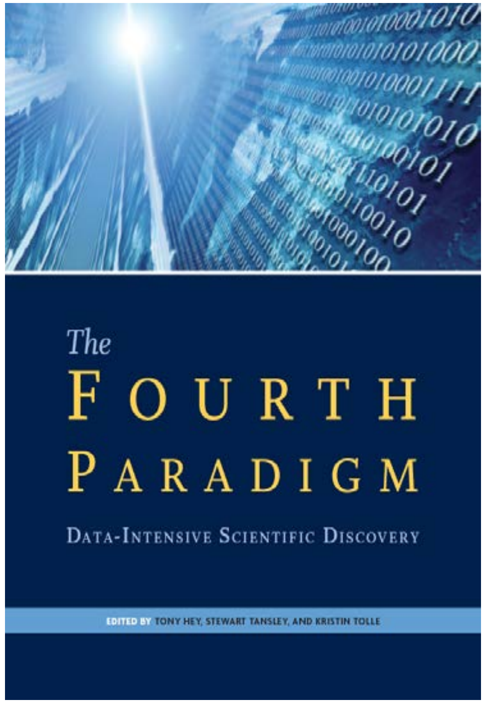
- Current: Place AI in control of Synthesis.



MGI + AI: Contacts



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



Questions?