The Materials Genome Initiative (and NIST)

James A Warren
Technical Program Director for Materials Genomics
Material Measurement Laboratory
National Institute of Standards and Technology

The Materials Genome Initiative (and NIST)

James A Warren

Technical Program Director for Materials Genomics

Material Measurement Laboratory

National Institute of Standards and Technology

Science advances one funeral at a time -Max Planck

The Materials Genome Initiative (and NIST)

James A Warren

Technical Program Director for Materials Genomics

Material Measurement Laboratory

National Institute of Standards and Technology

Science advances one funeral at a time -Max Planck

Take time to stop and smell the data -Nate Silver

Outline/Goals

- 21 month perspective about the National effort
- Data, Culture and what's coming
- Share info about the NIST effort
  - Pilots
  - Supporting Activities
  - External Engagement

A comment on audiences

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

*To help businesses discover, develop, and deploy new materials twice as fast, we're launching what we call **the Materials Genome Initiative**. The invention of silicon circuits and lithium ion batteries made computers and iPods and iPads possible, but it took years to get those technologies from the drawing board to the market place. We can do it faster.*

-President Obama
Carnegie Mellon University, June 2011

To help businesses discover, develop, and deploy new materials twice as fast, we're launching what we call ***the Materials Genome Initiative****. The invention of silicon circuits and lithium ion batteries made computers and iPods and iPads possible, but it took years to get those technologies from the drawing board to the market place. We can do it faster.*

-President Obama

Carnegie Mellon University, June 2011

TWO YEAR ANNIVERSARY COMING!

The Materials Genome Initiative

Goal: to decrease the cost and time-to-market by 50%

1. Develop a Materials Innovation Infrastructure

2. Achieve National goals in energy, security, and human welfare with advanced materials

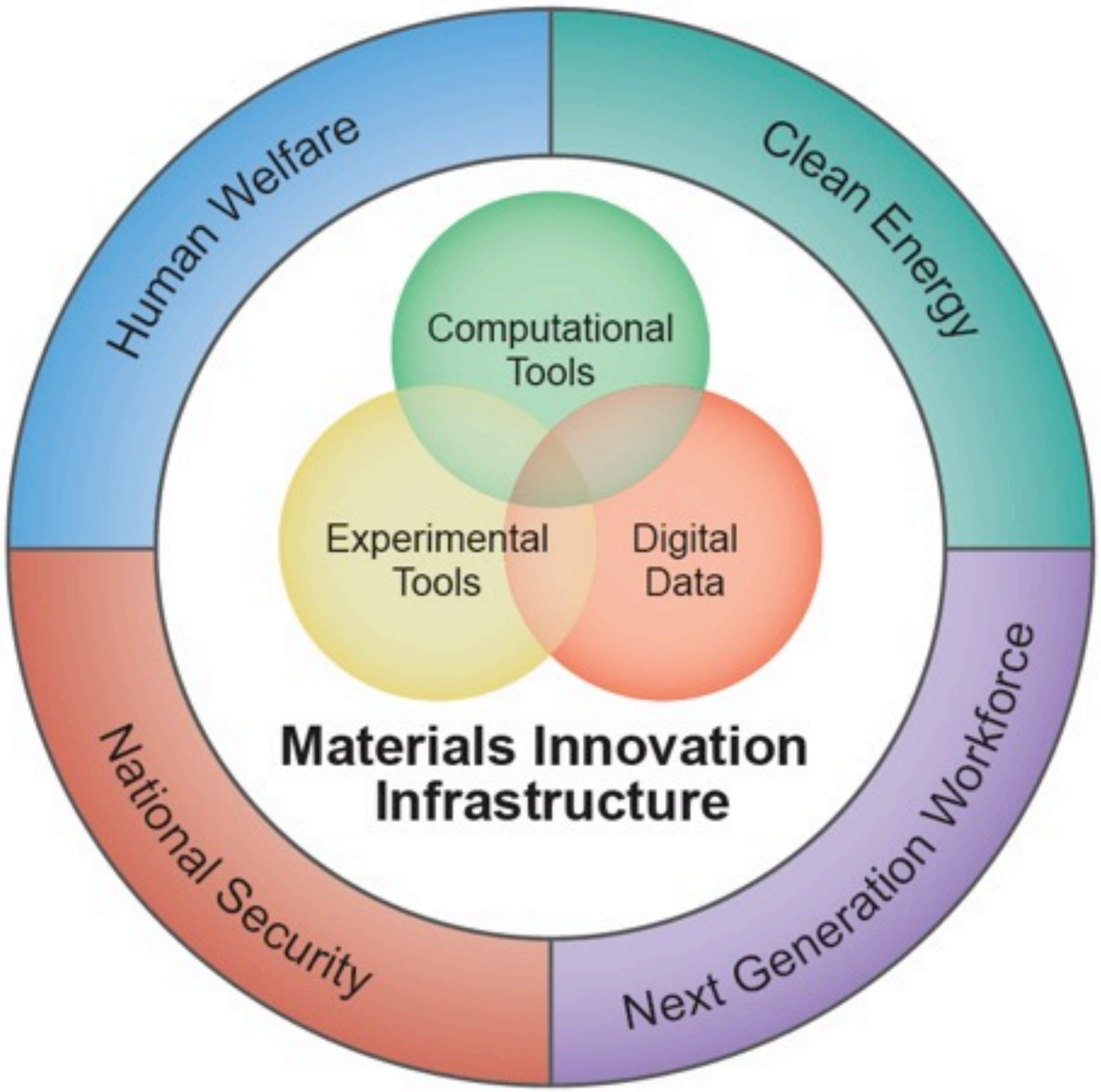
2. Equip the next generation materials workforce

http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf
http://www.wtec.org/sbes-vision/RDW-color-FINAL-04.22.10.pdf
http://www.wtec.org/sbes/SBES-GlobalFinalReport.pdf
National Research Council. (2008). *Integrated Computational Materials Engineering.* Washington, DC: The National Academies Press.

http://nvlpubs.nist.gov/nistpubs/ir/2012/NIST.IR.7898.pdf

The Materials Innovation Infrastructure

Three-pronged approach to advance materials design

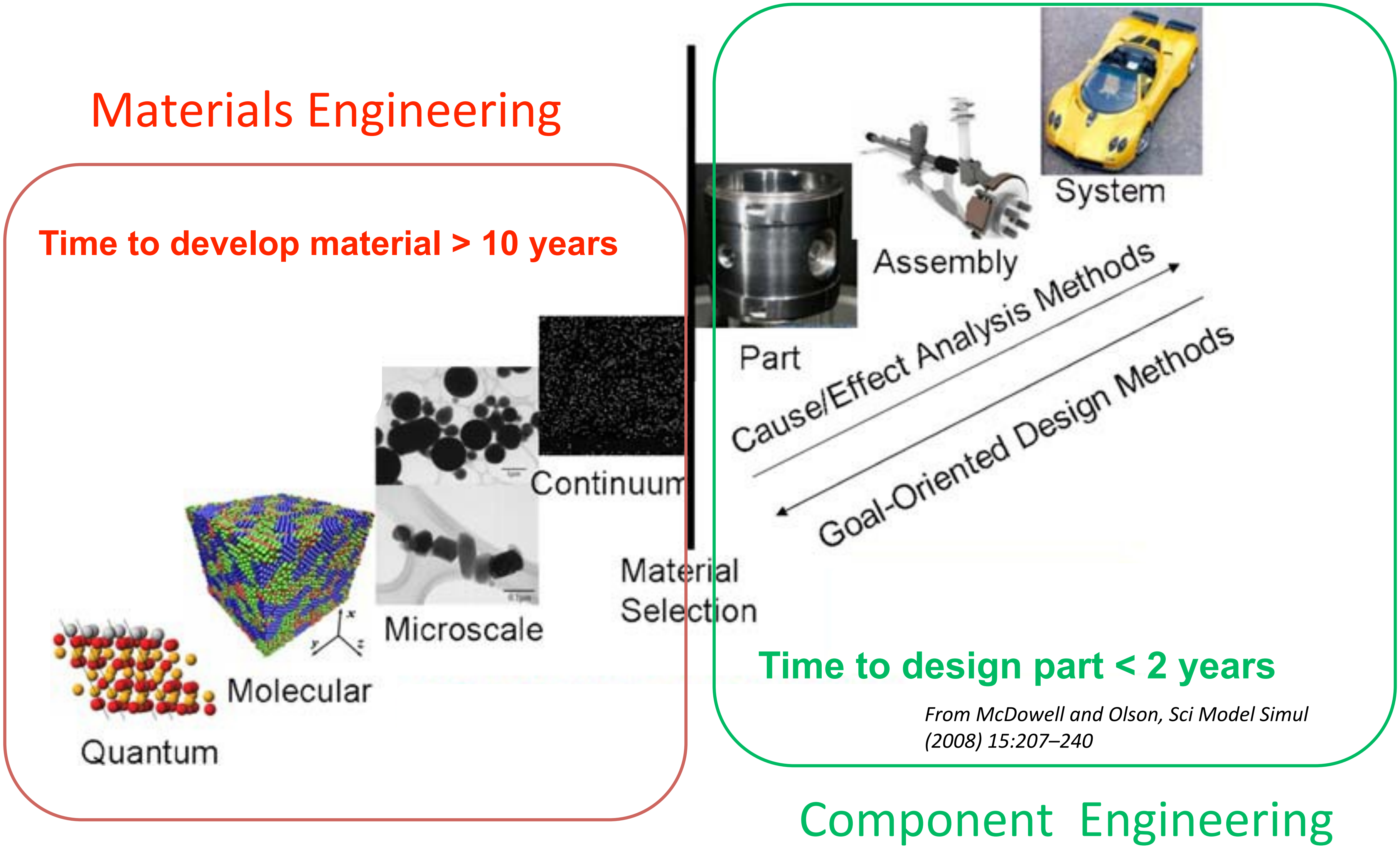


"The Materials Genome Initiative will develop new integrated computational, experimental, and data informatics tools. These software and integration tools will span the entire materials continuum, be developed using an open platform, improve best-in-class predictive capabilities, and adhere to newly created standards for quick integration of digital information across the materials innovation infrastructure."

Materials Genome Initiative for Global Competitiveness

Acceleration of Advance Materials Development

Disconnect between time to design a part and time to develop the material out which to make that part



From McDowell and Olson, Sci Model Simul (2008) 15:207–240

Accelerating the development of advanced materials is critical for achieving global competitiveness

Why an MGI?

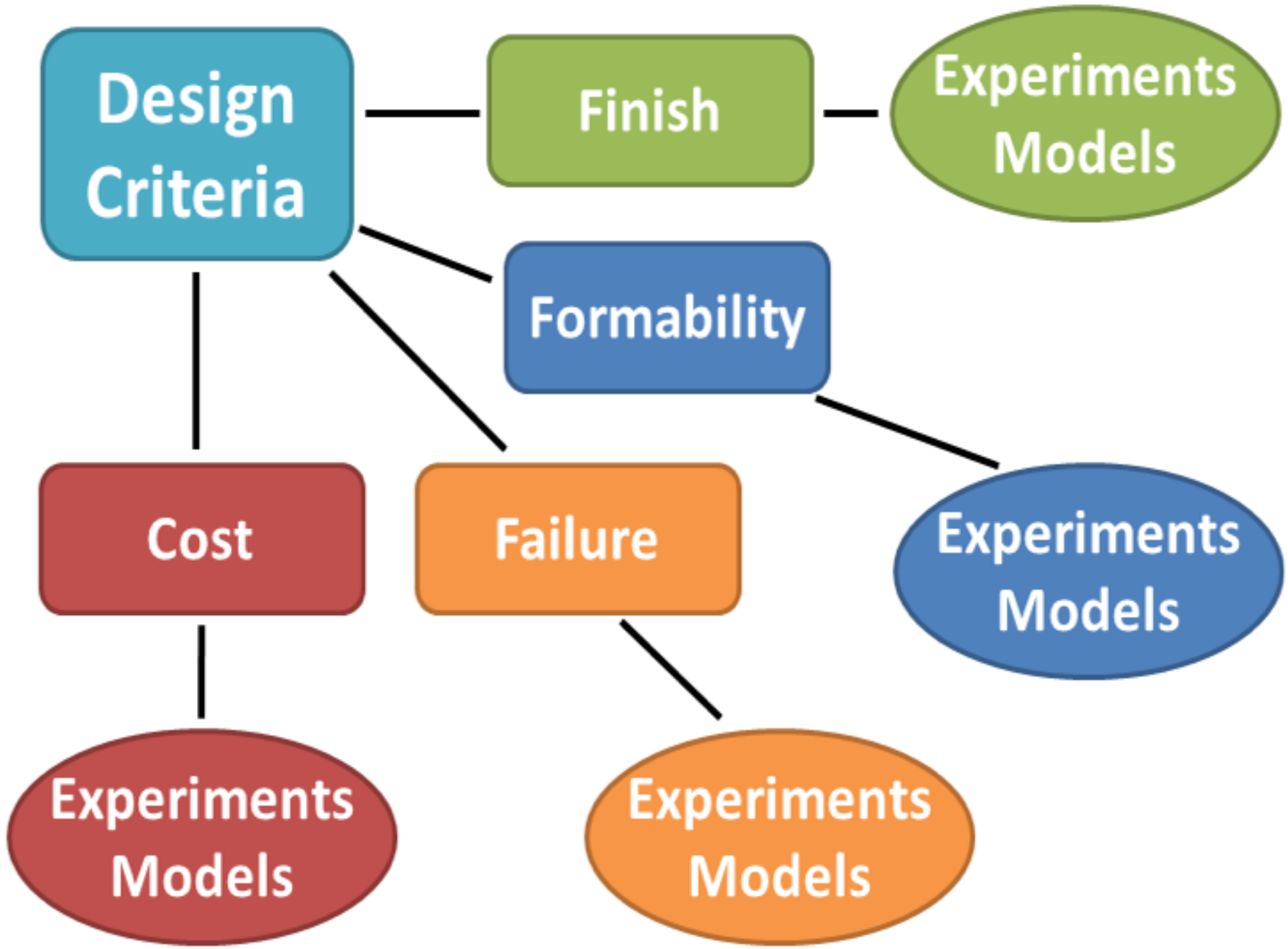


Figure 1. Schematic of how the design criteria for a given material dictate the needed material properties and thus define the needed experiments, models and data.

The MGI is a direct effort to enable the creation of new materials by leveraging the growth in computational power with a commensurate focus on modeling and simulation

1986!

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

The MGI is a direct effort to enable the creation of new materials by leveraging the growth in computational power with a commensurate focus on modeling and simulation

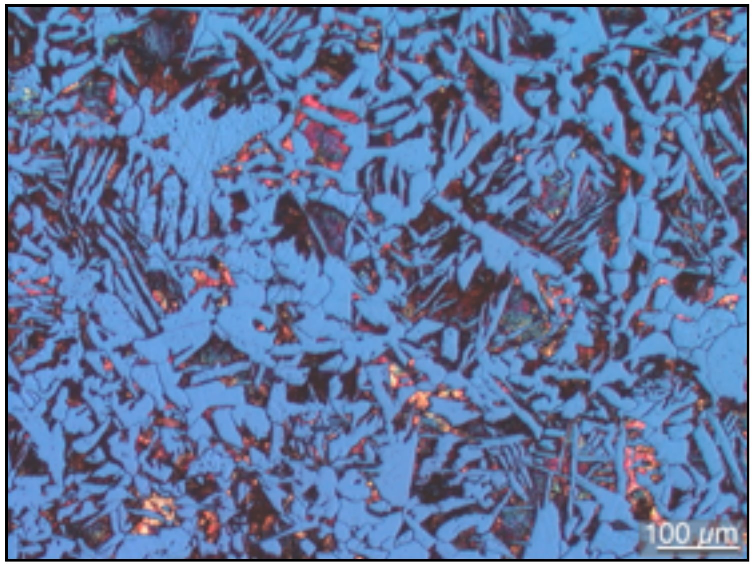
1986!

Why an MGI Now?

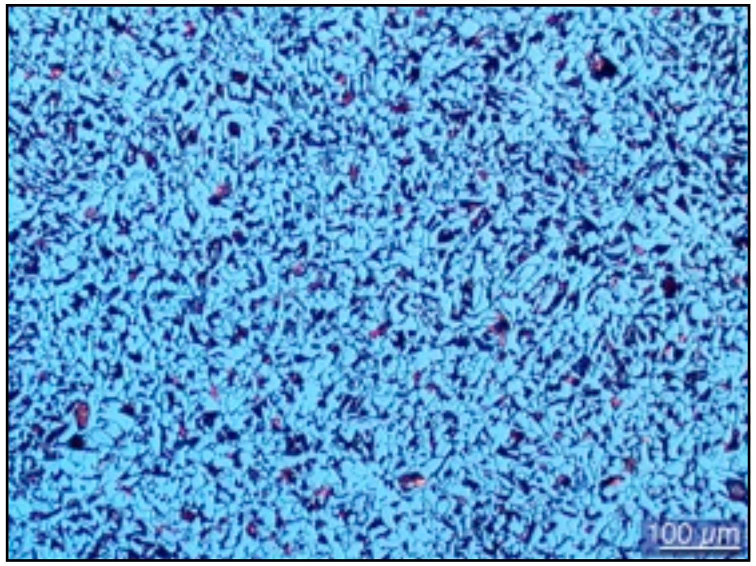
Materials Are Complicated Systems
Modeling is a Challenge

- Advanced materials are complex: multi-component and multi-phase
- Without adequate modeling, informatics and data exchange, the development of next generation materials using empirical approaches is bogged down by their complexity
- The Materials Genome Initiative seeks to advance materials design capabilities to promote faster, cheaper

Alloy cooled from 300 °C



Alloy cooled from 800 °C



- Composition and processing affect properties
- Phases change as a function of processing
- Microstructures consist of mixtures of multiple material phases
- Finer microstructure results in a much stronger alloy

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

Federal Programs

The federal government has announced several programs and funding opportunities related to the Materials Genome Initiative. Learn more about these activities below:

Federal Funding Announcements

FY 2012

Department of Energy:

- Funding announcement for the development of lightweight materials and awardees.
- Funding announcement for predictive theory and modeling for advanced materials.
- Funding announcement for Scientific Discovery through Advanced Computing (SciDAC) partnerships.

Department of Defense:

- Funding announcement for the Enterprise for Multiscale Research of Materials program, led by the Army Research Laboratory (ARL).
- Funding announcement for the Air Force Research Laboratory (AFRL) Center of Excellence for Integrated Computational Material Science and Engineering of Structural Materials
- Funding announcement for the Air Force Research Lab (AFRL) Center of Excellence in advanced organic composites
- Beta release of the Ab-Initio Electronic Structure Library (AFLOWLIB), maintained by Duke University in partnership with the Office of Naval Research, with open access to over 17,000 compounds derived from the Inorganic Crystal Structure Database and over 160,000 binary alloys

National Institute of Standards and Technology:

- Funding announcement for the NIST Advanced Materials for Industry program.

National Science Foundation:

- Funding announcement for the FY 2012 NSF Designing Materials to Revolutionize and Engineer our Future (DMREF) program.

NIST
onal Institute of
and Technology
nent of Commerce

Nanoinformatics, the NKI-NSI and the MGI

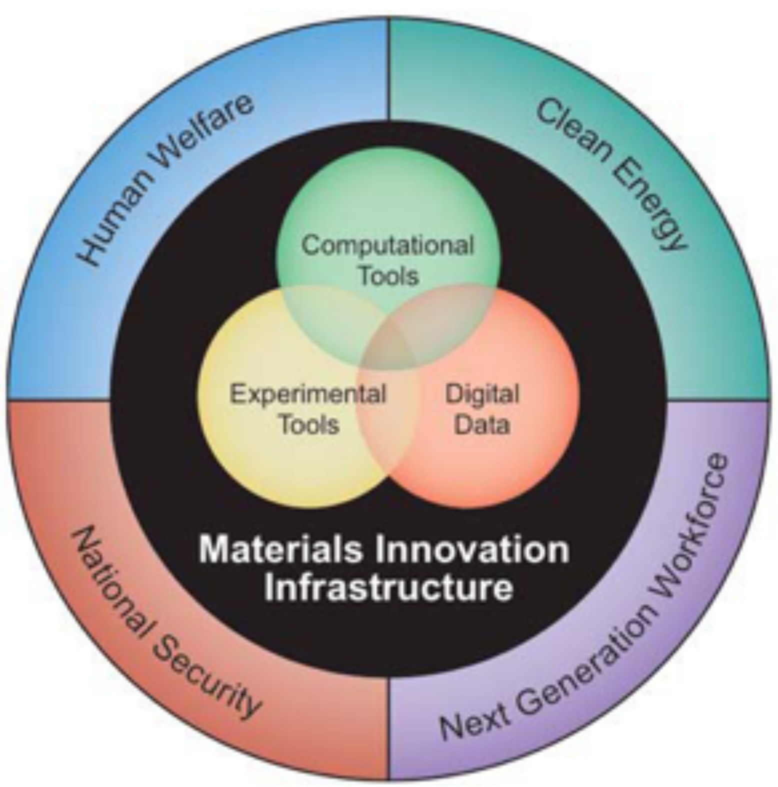
National Institute of Standards and Technology
U.S. Department of Commerce

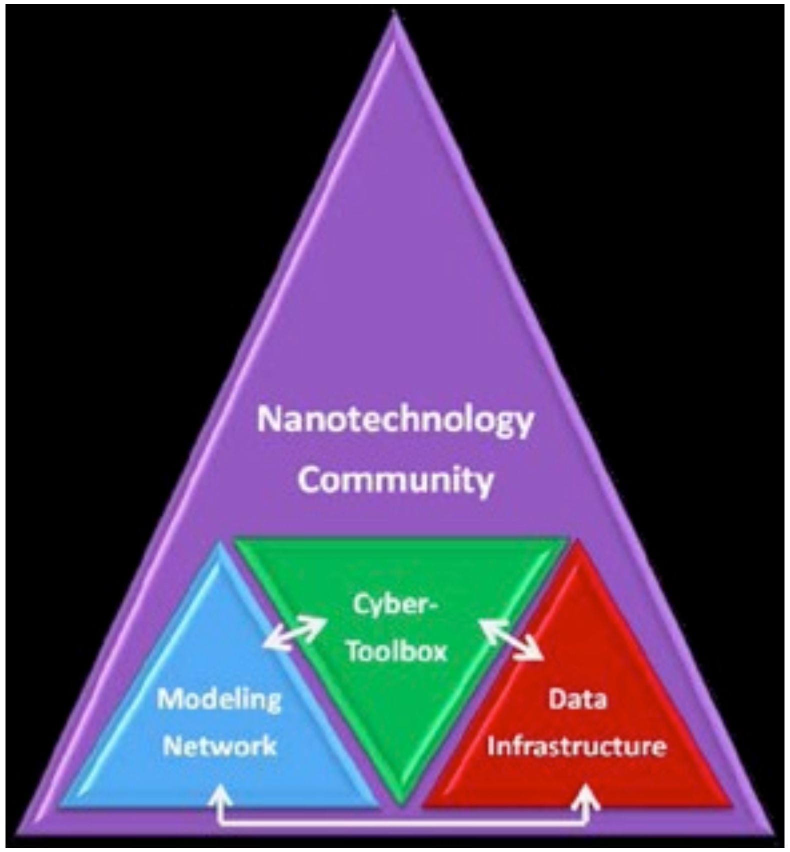
"The Nanotechnology Knowledge Infrastructure developed by ten agencies will stimulate the development of models, simulation tools, and databases that enable the prediction of specific properties and characteristics of nanoscale materials.

Also approaches, protocols, and standards developed through MGI activities may be initially explored, tested, or evaluated specifically for nanomaterials under NKI efforts.

The cross fertilization between NNI and MGI will yield broader knowledge dissemination and can be facilitated by the NKI effort."

Fact Sheet on Progress on the Materials Genome Initiative
Executive Office of the President, May 14, 2012





NIST and MGI

2010

NIST, DOE, NSF, and DOD work with OSTP to prepare MGI Whitepaper

2011

NIST develops Initiative, $4M allocated for internal programs

2012

04/12: NSTC Subcommittee (CoT) on the Materials Genome Initiative formed (NIST, DOE, DOD, NSF, NASA, NIH, USGS, NNSA, DARPA, NSA, OMB)

6/12: MGI White House Kickoff Event www.whitehouse.gov/mgi

2013

2/13:
- Subcommittee co-chairs announced -L. Locascio (NIST), H. Kung (DOE)
- Strategic planning kicked off
- Focus teams formed: Data, Code Curation, Industry Outreach, Transformational Challenges

Multi-Agency Efforts

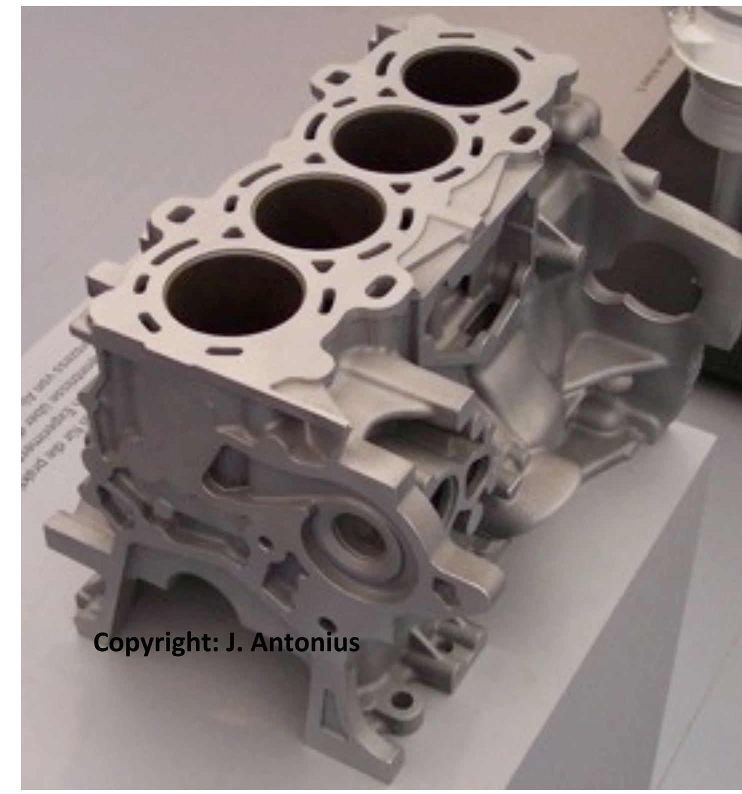
- NSTC Subcommittee (CoT) on the Materials Genome Initiative formed (First meeting 4/12)
- Membership now includes NIST, DOE, DOD, NSF, NASA, NIH, USGS, NNSA, DARPA, NSA, and OMB.
- Coordination with the NNCO and new Nanotechnology Knowledge Infrastructure (NKI) Signature Initiative.
- Coordination with RDA, NITRD and ACI (NSF)
- MGI White House Kickoff Event (Discussed Later)
- New focus teams: Data, Code Curation and Technical Gaps, Industry Outreach

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

FY12 NIST Budget Initiative: Advanced Materials For Industry (4M FY12)

NIST will further enable advanced materials by working with our partners in other government agencies, academia and industry to develop

- Computational tools, validated databases, data assessment tools and techniques, and standards
- Reference models and simulations
- Mechanisms for exchange of information, Consortia to determine consensus standards for data exchange



Copyright: J. Antonius

Courtesy: U.S. ARMY

Formulating the NIST Role in MGI

Some background and further thoughts

NIST's Mission

To promote U.S. innovation and industrial competitiveness by advancing

measurement science,

standards, and

technology

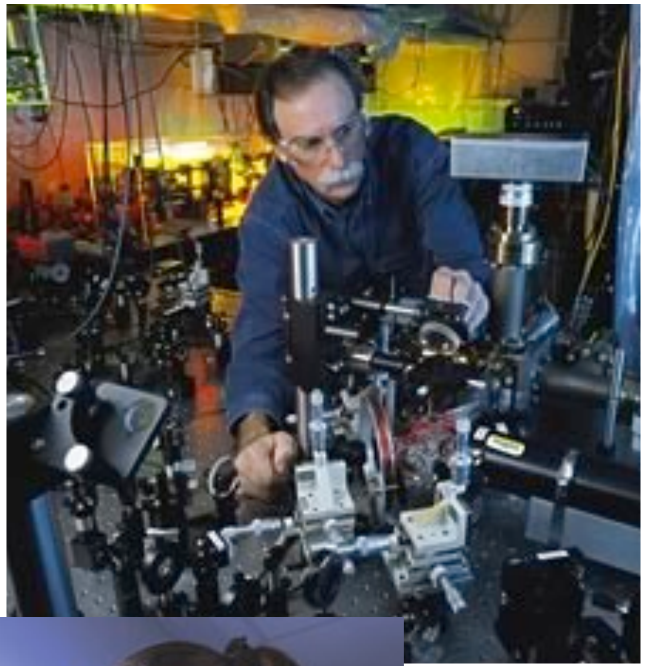
in ways that enhance economic security and improve our quality of life

©Robert Rathe

NIST at a Glance

Major Assets

- ~ 2800 employees ~(50/50 technical/admin)
- ~ 2600 associates and facilities users
- ~ 1600 field staff in partner organizations (Manufacturing Extension Partnership)



© Geoffrey Wheeler

Major Programs

- NIST Laboratories
- Baldrige Performance Excellence Program
- Hollings Manufacturing Extension Partnership

©Robert Rathe

NIST Products and Services

© Robert Rathe

Measurement Research

- ~ 2,200 publications per year

Standard Reference Data

- ~ 100 different types
- ~ 6,000 units sold per year
- ~ 226 million data downloads per year

Standard Reference Materials

- ~ 1,300 products available
- ~ 30,000 units sold per year

Calibration Tests

- ~ 18,000 tests per year

Laboratory Accreditation

- ~ 800 accreditations of testing and calibration laboratories

The Problem with Materials Science Data

A simple example to illustrate the key questions, gaps, obstacles to realizing materials by design

The Problem with Materials Science Data

A simple example to illustrate the key questions, gaps, obstacles to realizing materials by design

Jim states the obvious? Or causes trouble

Q: What is Data?
A: Standard Answer

- Data: result of measurement
  - Physical: experimental result, uncertainty
  - Virtual: simulation result, uncertainty
- Metadata: information describing the measurement process
  - Physical: experimental setup, settings
  - Virtual: explicit underlying model, software*, input parameters

Q: What is Data?
A: Standard Answer

- Data: result of measurement
  - Physical: experimental result, uncertainty
  - Virtual: simulation result, uncertainty
- Metadata: information describing the measurement process
  - Physical: experimental setup, settings
  - Virtual: explicit underlying model, software*, input parameters

* simulation software is data!

Traditional Approach Example

- Measure a diffusion coefficient (perhaps by tracking the root mean displacement of tracer particles)
- record values is some sort of table (perhaps Excel)
- perhaps publish
- publication contains metadata

What the #$&%^ is a diffusion coefficient?

$$\frac{\partial c}{\partial t} = D\nabla^2 c$$

- It's only a model (not reality)
- It's an equation
- What you think you're measuring is ***NOT* the diffusion coefficient**
- You are measuring the (maybe) root mean displacement of particles and **assuming the diffusion equation is true**.

All Models are Wrong, but some are Useful -- George Box

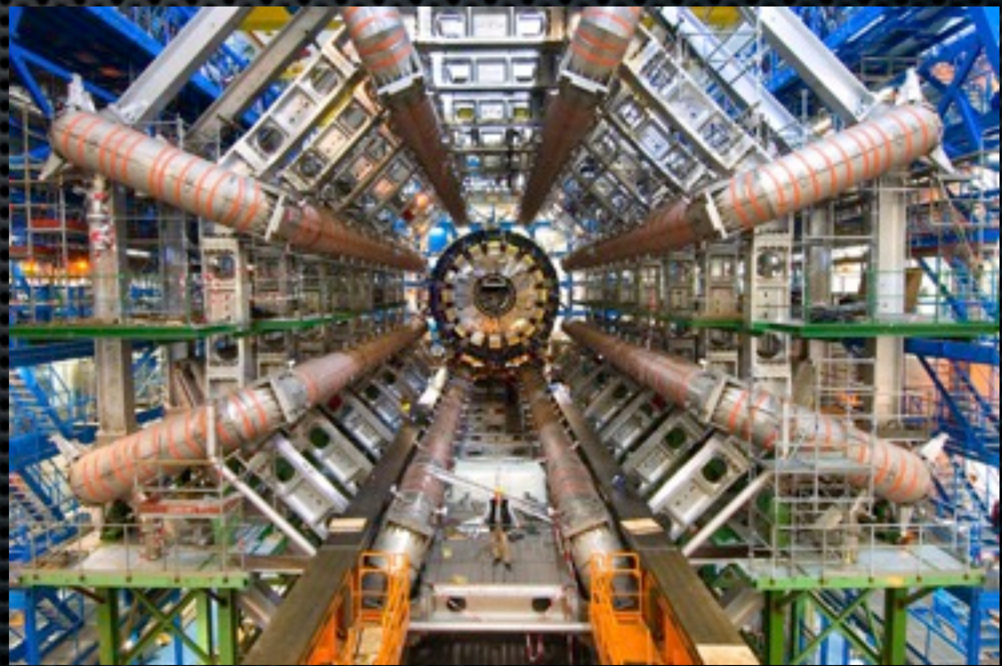
Old School, Computation

- Write down Equations
- “Solve them” (verify please)
- Analyze, make graphs, compute numbers of interest (results)
- Publish results

We can do much much better!

Better Models = Less Data

- Take the LHC as an example
- Data produced at 1PB/sec !
- Reduced data saved: 300 MB/s
- That's a darn good model
- Other end : Biology?
- Materials: In the middle



- But you need to know the model to make sense of the data!

What is Data “2.0” ?

Science is characterized by the iteration of experiment and models, yielding higher fidelity with lower uncertainty.

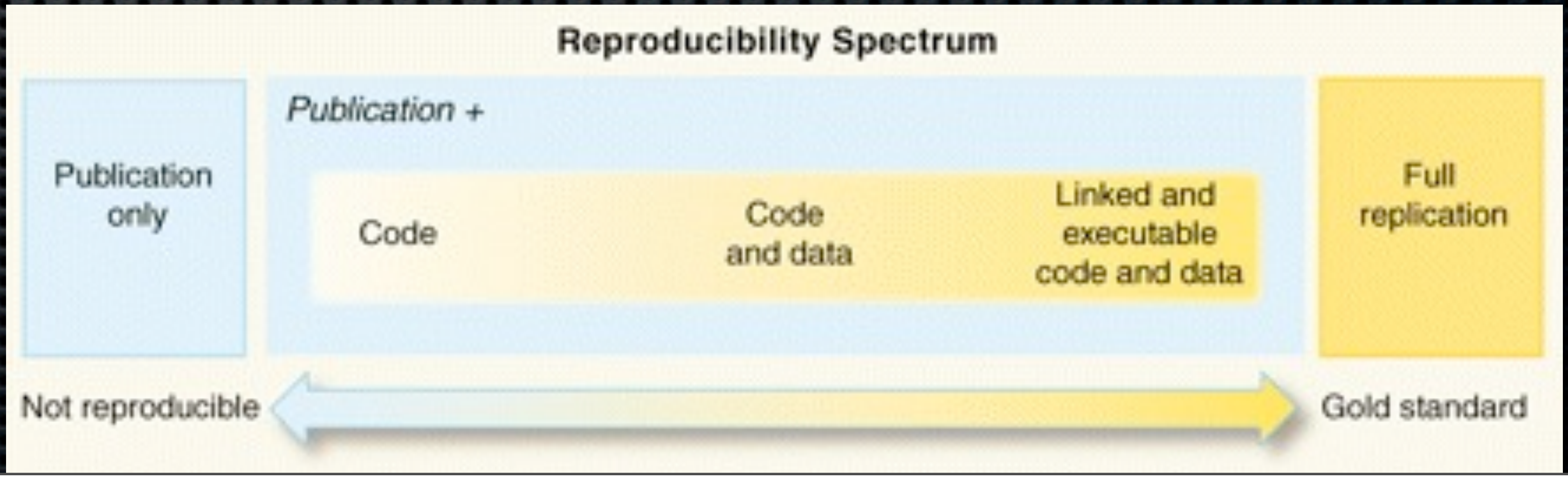
The measurement or computation of a quantity (data) is generally *meaningless* without the associated quantifying model that defines both the data and its uncertainties.

Thus dissemination of data is ideally **the dissemination of the following information:**

1. **(1) Measured quantities,**
2. **(2) Associated quantifying models, and**
3. **(3)** ***Raw data, including the protocols (and equipment) by which it was obtained***

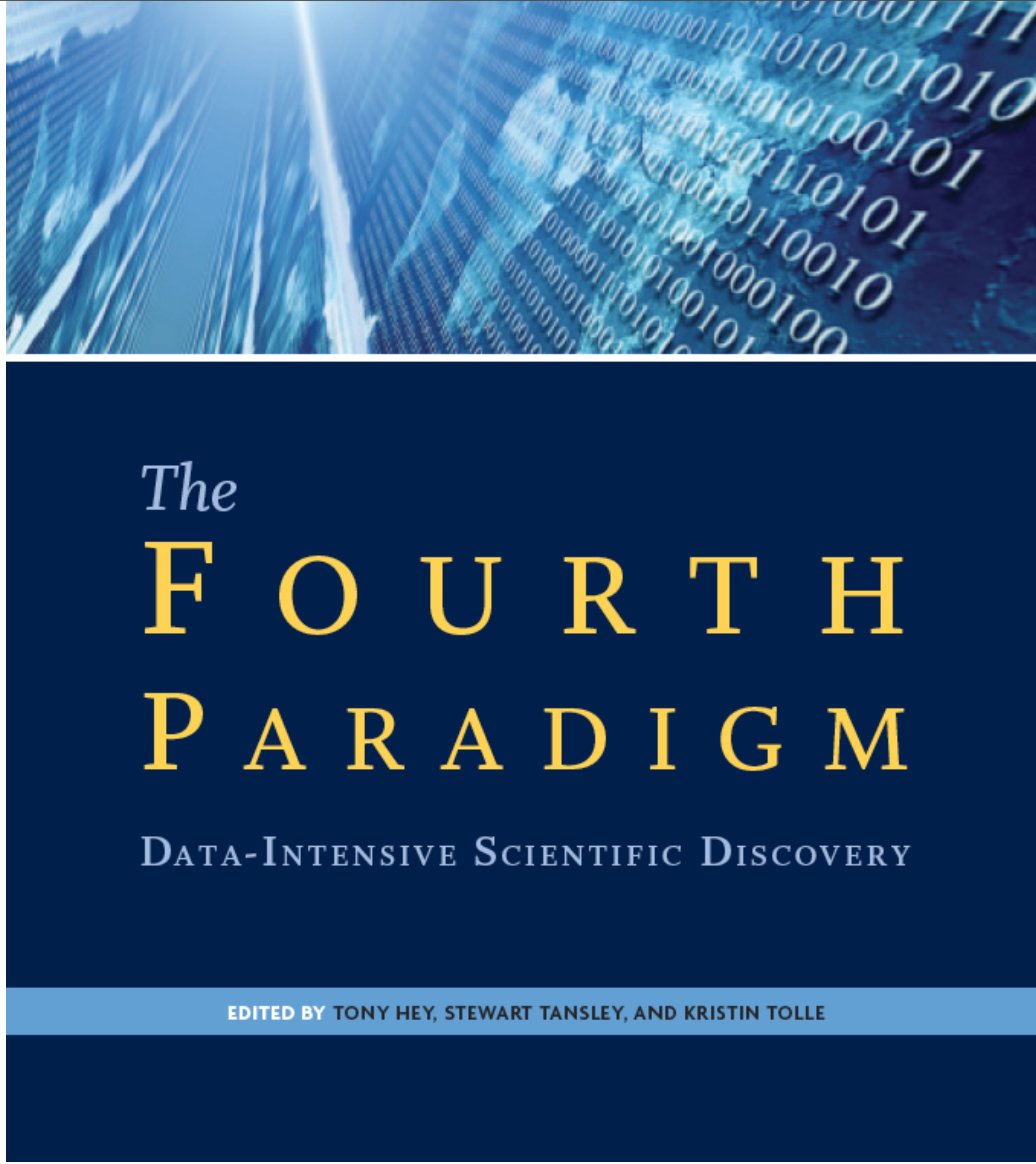
Without these descriptions (expressible through software), the data becomes divorced from meaning, making interpretation “difficult.”

see R. Peng *Science* 2 December 2011: 1226-1227



A Comment on Data Driven Discovery for MGI

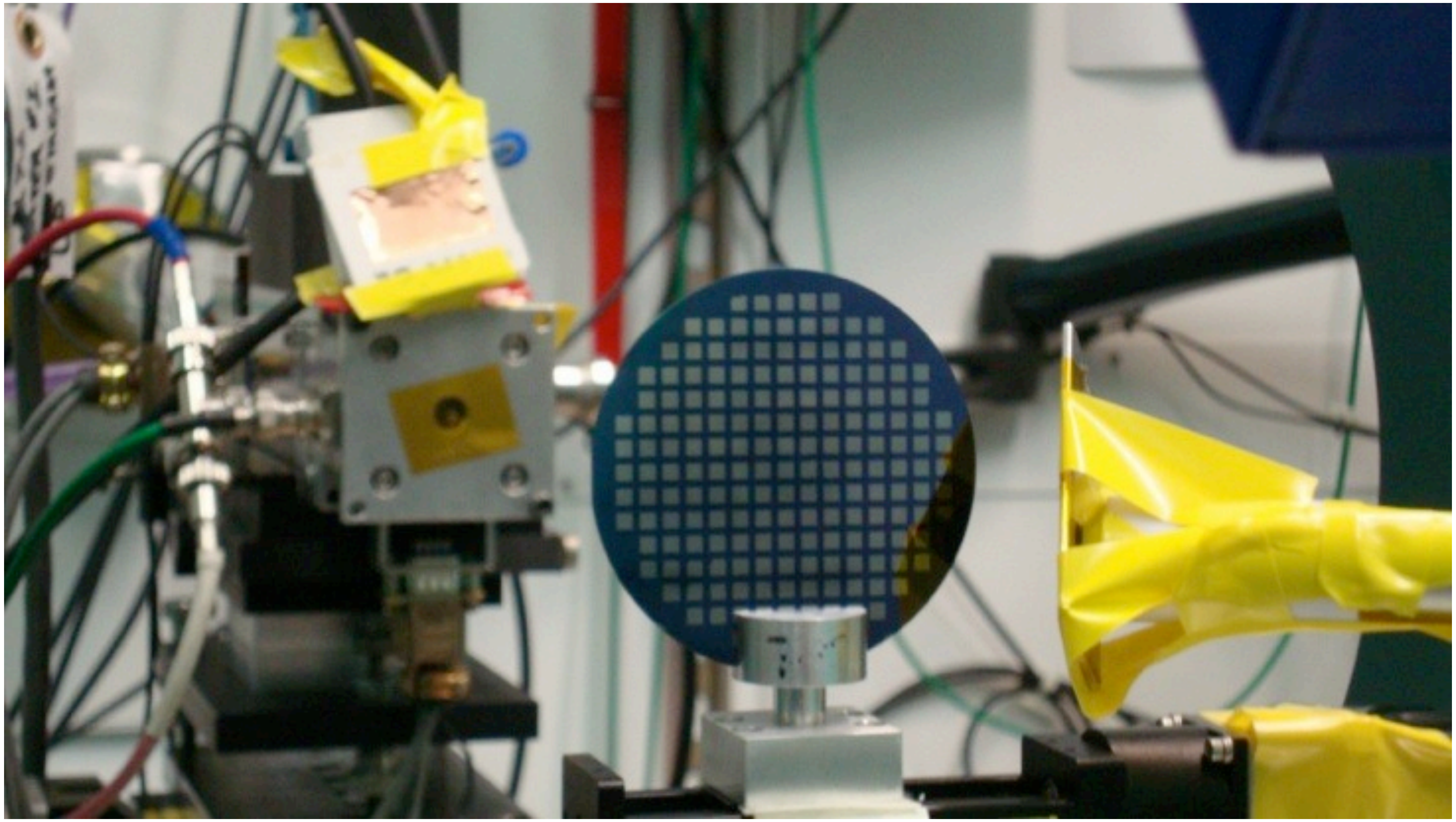
- Biology community comfortable with this
- Physics community finds this approach “weird”
- Materials Science is in the middle
- Presents unique challenges for Computer Science and informatics research


The
FOURTH
PARADIGM
DATA-INTENSIVE SCIENTIFIC DISCOVERY
EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

It's all about mapping structural information

Synchrotron diffraction set up at Beamline 7-2 (SSRL)

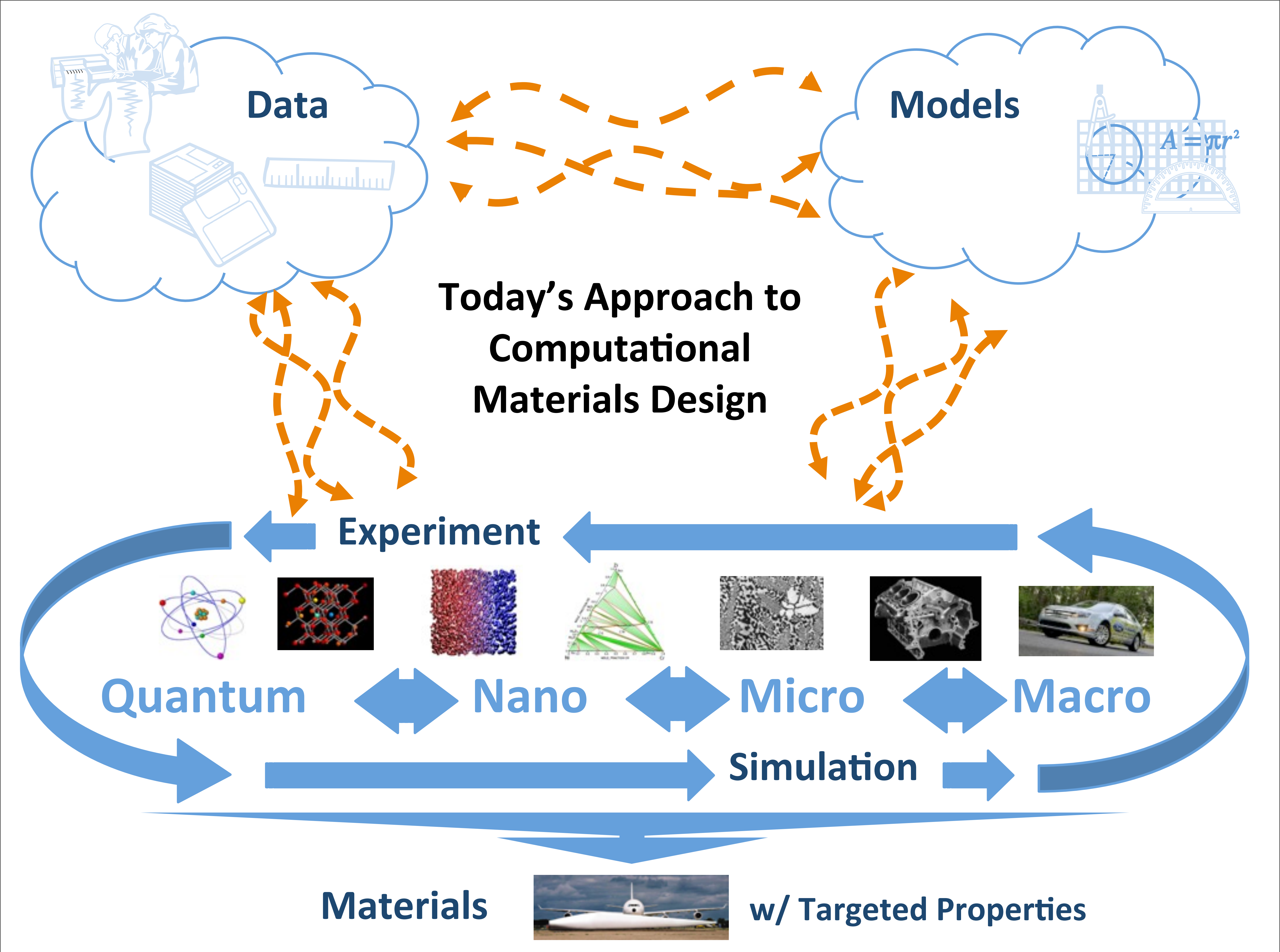
The entire 3" wafer (300 spots) can now be measured in 2 hrs

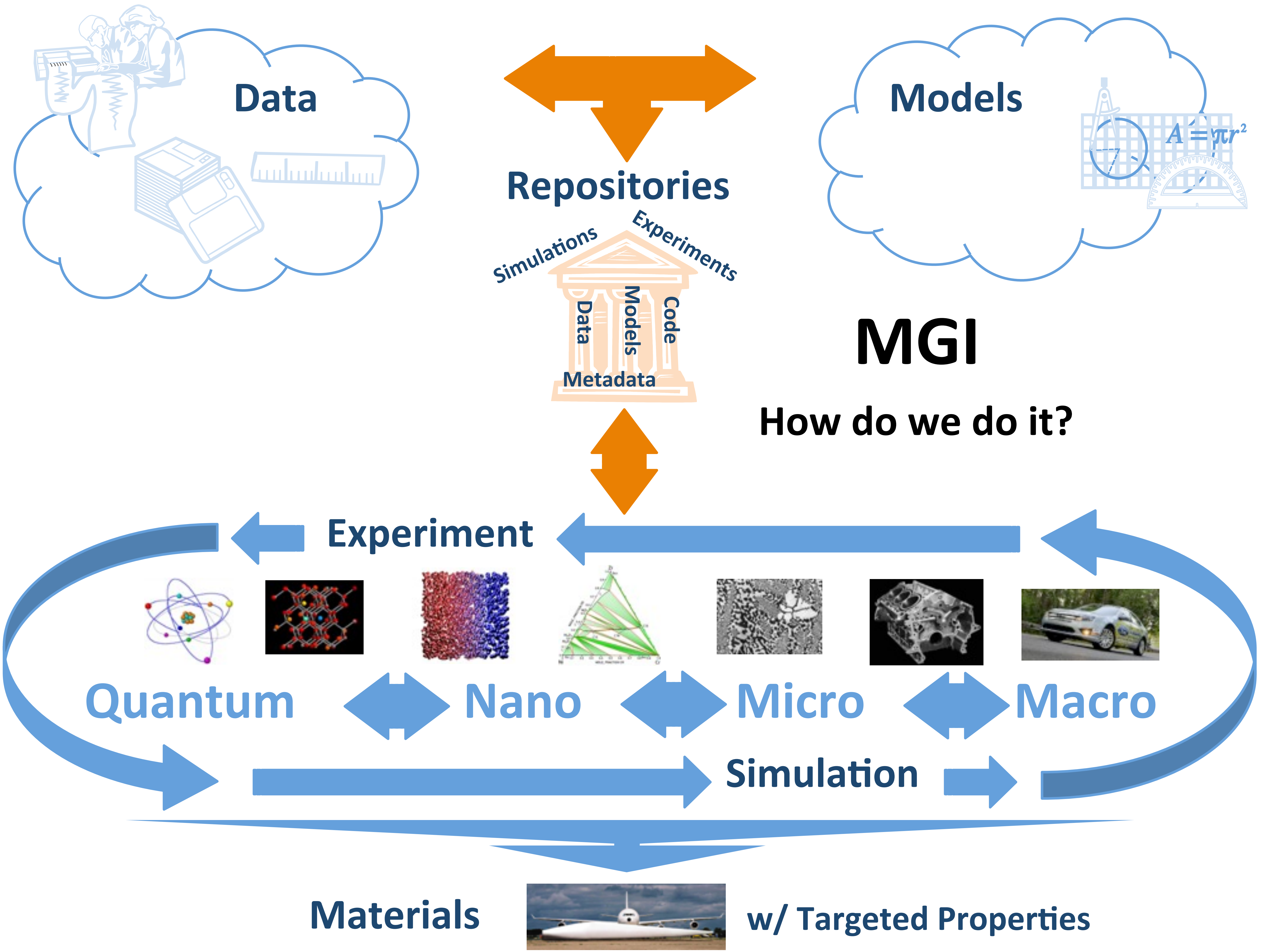


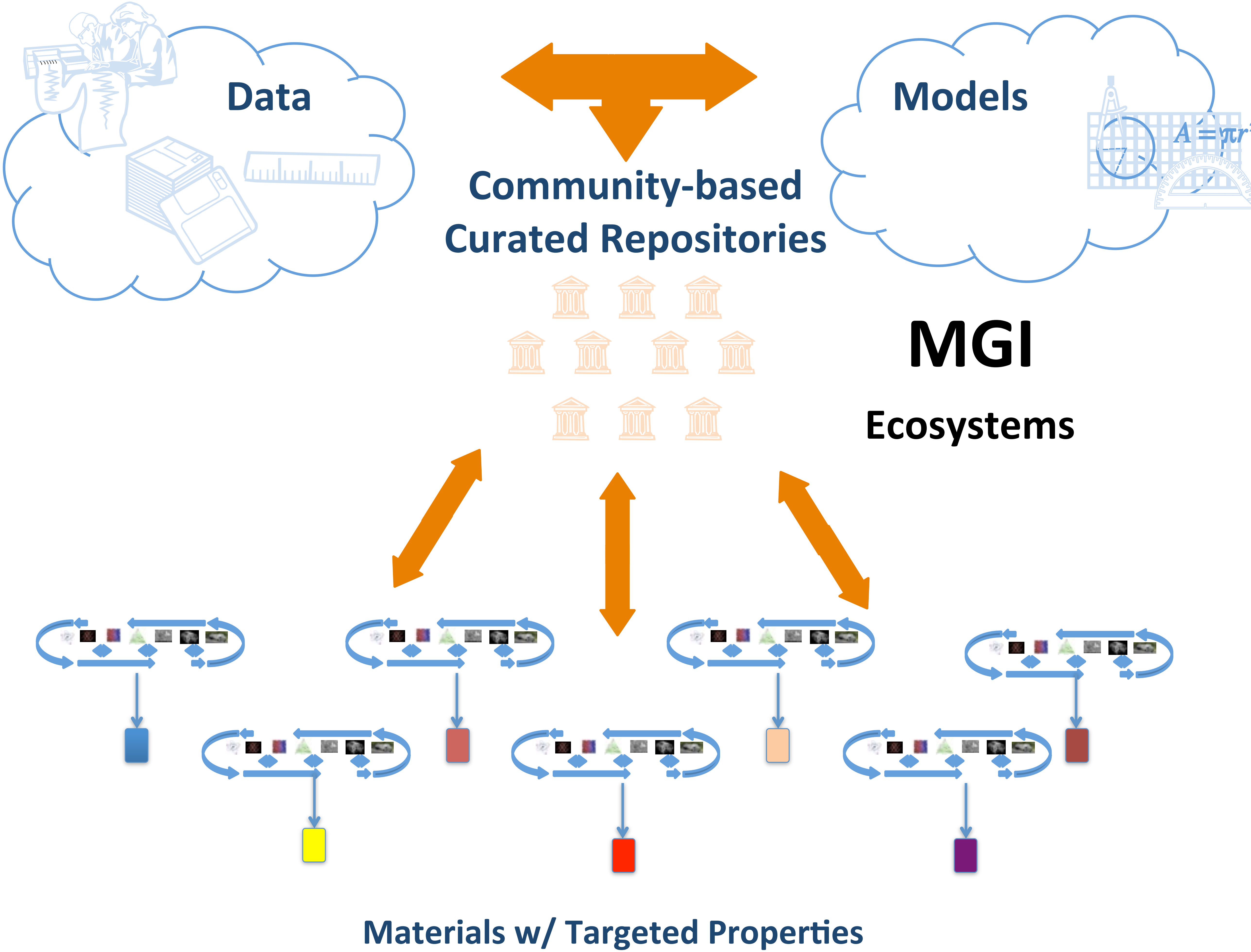
Consequences of the traditional approach

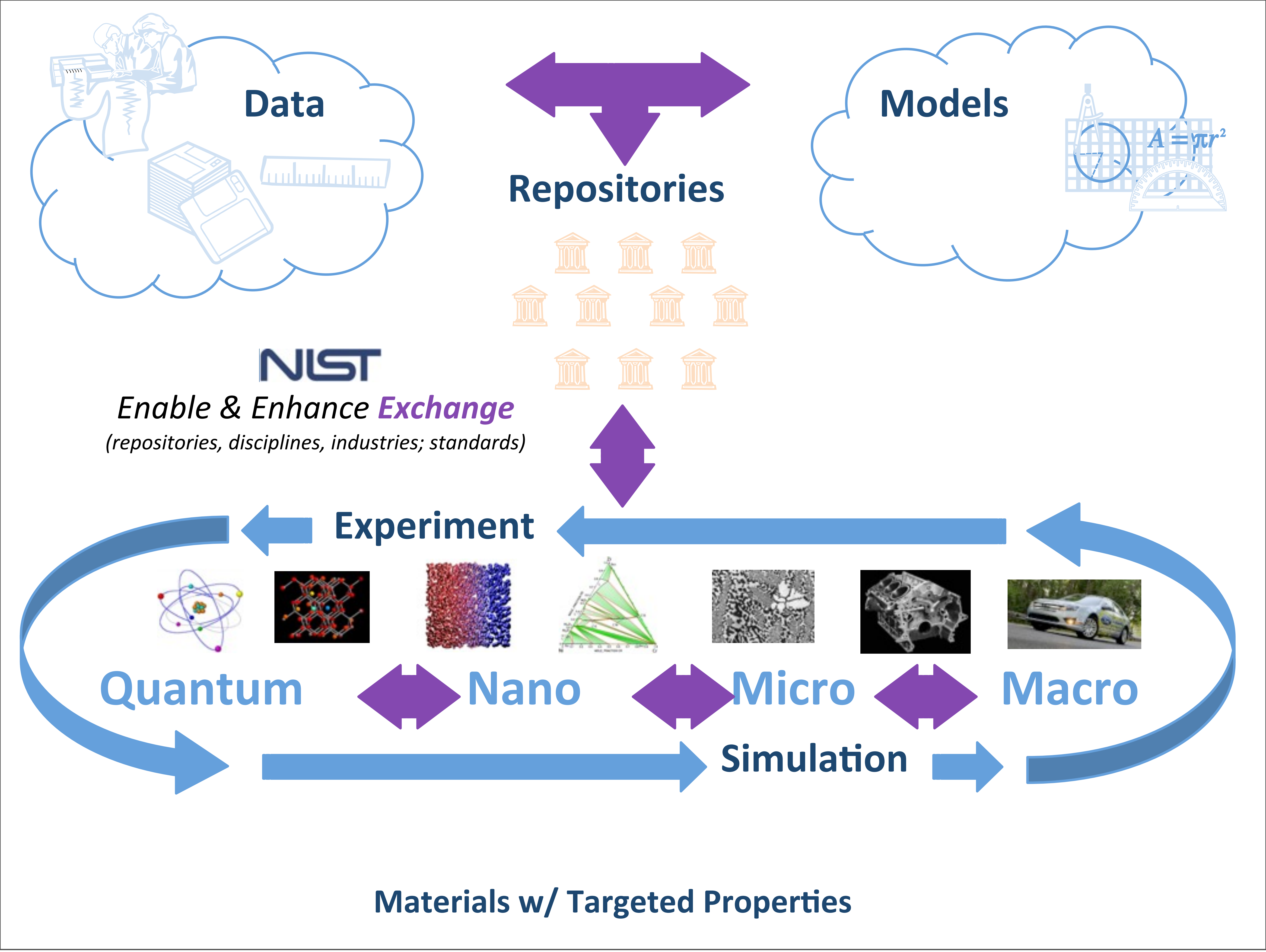
- High barrier to adoption of methods and results
- Extra expense due to duplication of effort
- Lost opportunities to enable discovery & further science

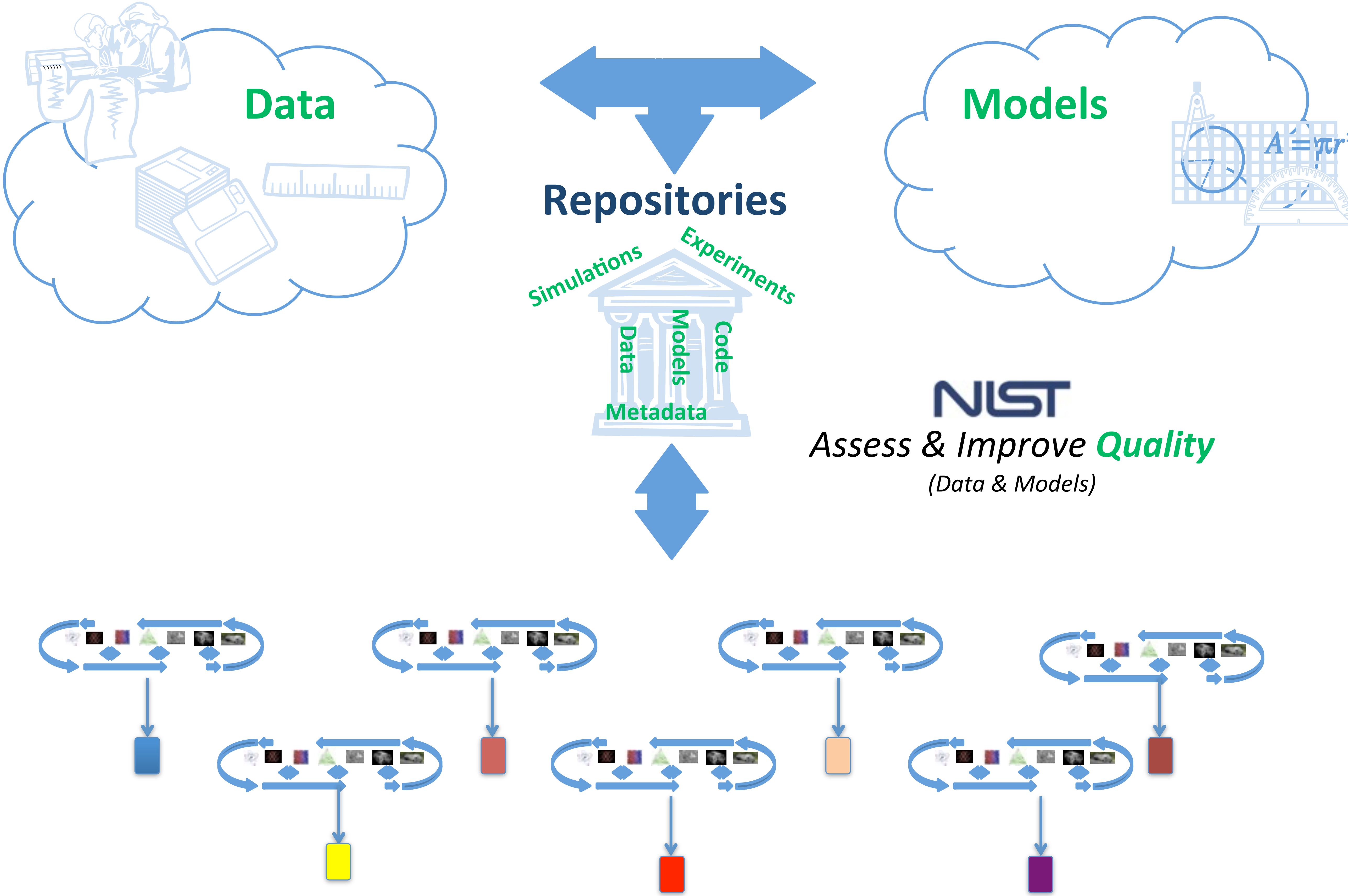
So Now What?


Data
Models
$A = \pi r^2$
Today's Approach to Computational Materials Design
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties

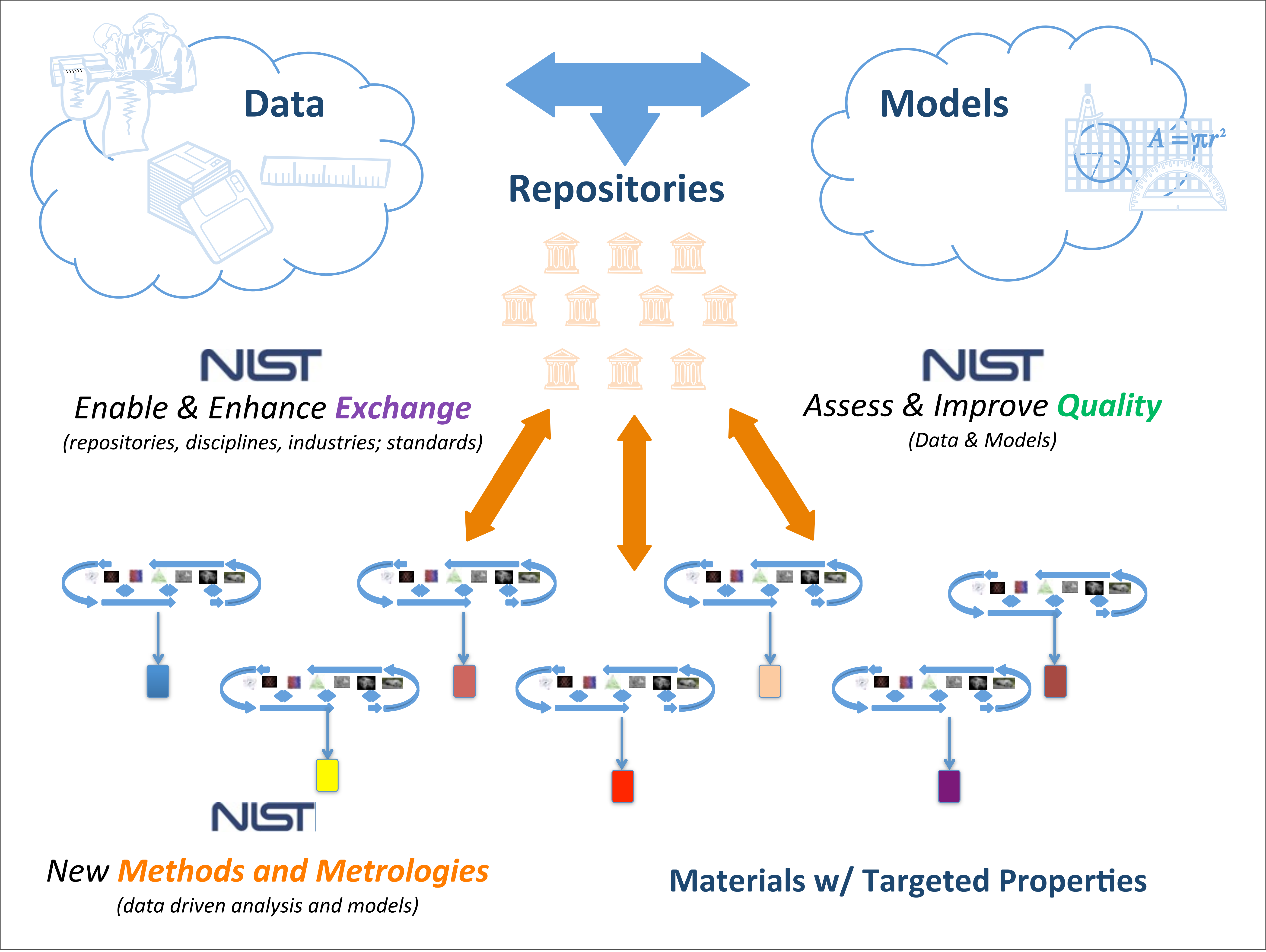

Data
Models
$A = \pi r^2$
Repositories
Simulations
Experiments
Data
Models
Code
Metadata
MGI
How do we do it?
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties


Data
Models
$A = \pi r^2$
Community-based
Curated Repositories
MGI
Ecosystems
Materials w/ Targeted Properties


Data
Repositories
Models
$A = \pi r^2$
NIST
Enable & Enhance ***Exchange***
(repositories, disciplines, industries; standards)
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials w/ Targeted Properties



Materials w/ Targeted Properties


Data
Repositories
Models
$A = \pi r^2$
NIST
Enable & Enhance Exchange
(repositories, disciplines, industries; standards)
NIST
Assess & Improve Quality
(Data & Models)
NIST
New Methods and Metrologies
(data driven analysis and models)
Materials w/ Targeted Properties

Short Summary

To foster widespread adoption of the MGI Paradigm both across and within the multitude of materials development ecosystems

NIST is establishing essential data exchange protocols and the means to ensure the quality of materials data and models

Yielding new methods, metrologies, and capabilities necessary for accelerated materials development

*Enable & Enhance **Exchange***

- Develop and deploy repositories
- Develop and disseminate materials informatics infrastructure
  - Enable data discovery through tools and standards
  - Capture data from scientific workflows and archival sources
  - Engage with stakeholders to determine needs and disseminate best practices
- Integrate across length and time scale
- Build and Test infrastructure through Pilots

*Assess & Improve **Quality***

- Validate Experiments and Models
- Verify Model accuracy
- Quantify Uncertainty
- Quantify Sensitivity
- Define Next Generation of Experiments and Models

New ***Methods and Metrologies***

- Develop Data Driven Materials Science
- Integrate with Modeling Expertise
- Achieve targets in Materials by Design/ICME

White House Event

Tom Kalil, OSTP
Gerbrand Ceder, MIT
Pat Dehmer, DOE

Alan Taub, GM (Ret)
Ed Seidel, NSF
Dave Turek, IBM

Scott Fish, US Army
Chuck Romine, NIST
Eric Isaacs, Argonne NL

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

Follow-on NIST Workshop on Building the Materials Innovation Infrastructure: Data & Standards

- Goals:
  - To start a broad conversation with thought leaders in the materials community
  - To scope the data/informatics challenges
- Stats
  - 130 participants (120 Attended WH event)
  - 23 Companies
- Outcomes
  - Report completed http://nvlpubs.nist.gov/nistpubs/ir/2012/NIST.IR.7898.pdf
  - Identification of the critical data/infrastructure challenges for specific applications
  - Synthesize crosscutting data challenges across length scales and application areas
  - Identify short and long term wins
  - Identify leaders and associated teams to work specific problems
  - Develop Web 2.0 resources to enable collaborations
  - Roll out of (http://www.mgidata.org/Home/)

- Further MGI Workshops under development
  - Building the Materials Innovation Infrastructure: Uncertainty
  - Follow-on Workshop in Summer to outline the Data Infrastructure for MGI
  - Grand Challenge Workshop

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

NIST Supports New Collaboration Modalities

New White House Push

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

http://whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Ideation Challenge

How do we establish, operate and maintain the hardware, software, and personnel needed to house and serve this data in a manner that is:

- Cost efficient
- Enables rapid web-based discovery and retrieval of information
- Ensures that the data retrieval has both high precision and recall.
- Scalable
- Standards-based

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

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

ional Institute of and Technology tment of Commerce

EERE Collaboration

U.S. Department of Energy

Office of Energy Efficiency and Renewable Energy

Fiscal Year 2013 Vehicle Technologies Program

Wide Funding Opportunity Announcement

Funding Opportunity Announcement (FOA) Number: DE-FOA-0000793
FOA Type: FINAL
CFDA Number: 81.086

Dissemination of Data and Results
In support of the President's Materials Genome Initiative, and to ensure that the results supported by this AOI can make the broadest impact, awardees are required to disseminate the results of **their work through infrastructure and methods identified by the National Institute of Standards and Technology (NIST)**. NIST will provide data schemas and informatics tools in accordance with the specific data types generated by the project; for example tracer, intrinsic and chemical diffusivity data; diffusion couple data; and phase transformation data from differential scanning calorimetry, differential thermal analysis, continuous cooling transformation data, and isothermal cooling transformation data. In addition to the specific tools for kinetic data, a variety of other data platforms will be offered. Specific file repositories will be provided for CALPHAD assessment files, first-principles files, and interatomic potentials (*http://www.ctcms.nist.gov/potentials/*). In addition to these specific file repositories, a general file repository platform will be established for all other data, which cannot be captured by the previously mentioned tools. In addition, dissemination of results via publication in peer-reviewed journals will be encouraged. Additionally, applications must describe how such data will be valuable in the development of high performance magnesium casting alloys

NIST Coordination

- Implement and Test Infrastructure through Pilots
- Outreach: Convene Workshops, engage with Industry, speak at conferences and other workshops
- Identify critical gaps and develop solutions to address these infrastructure holes

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

Pilots

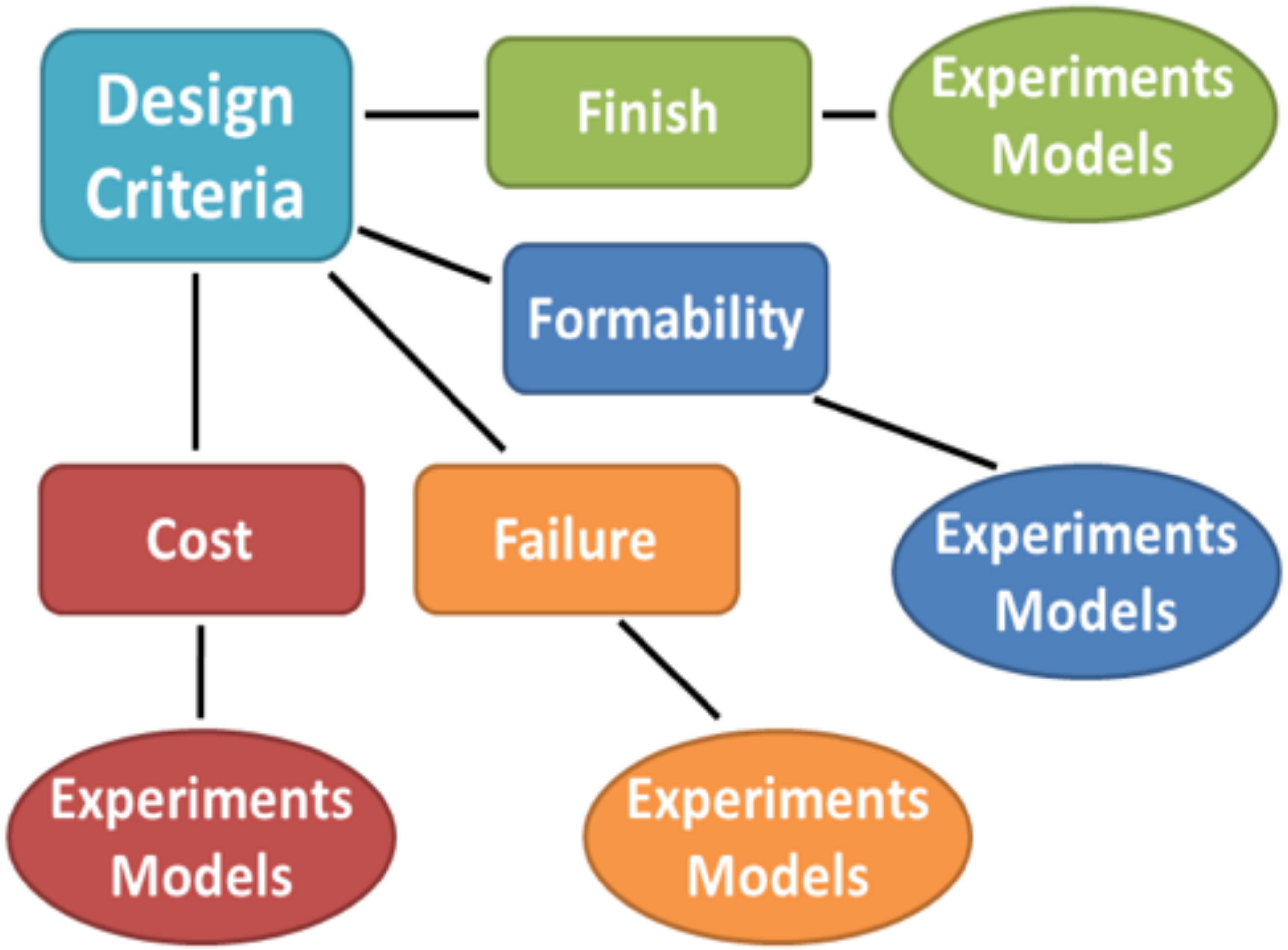
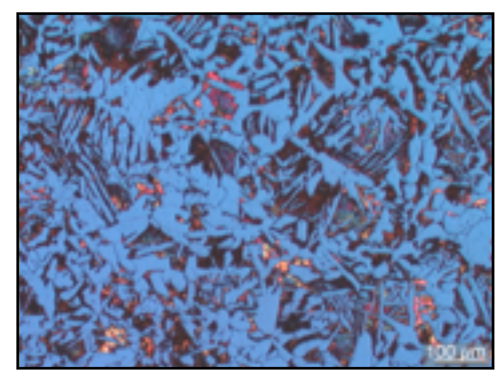


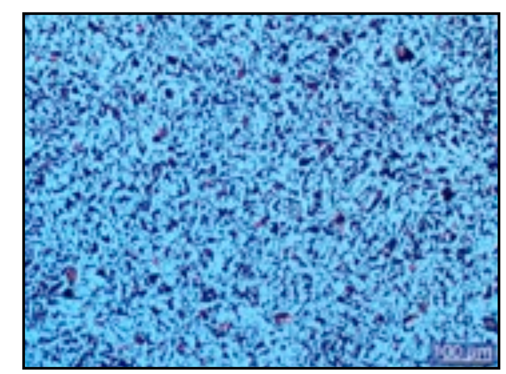
Figure 1. Schematic of how the design criteria for a given material dictate the needed material properties and thus define the needed experiments, models and data.

Processing

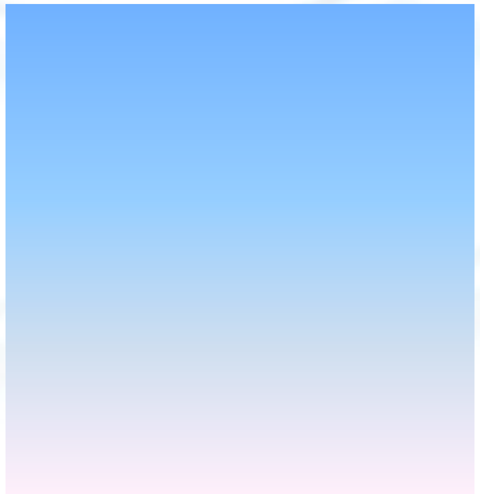
Annealed at 300C

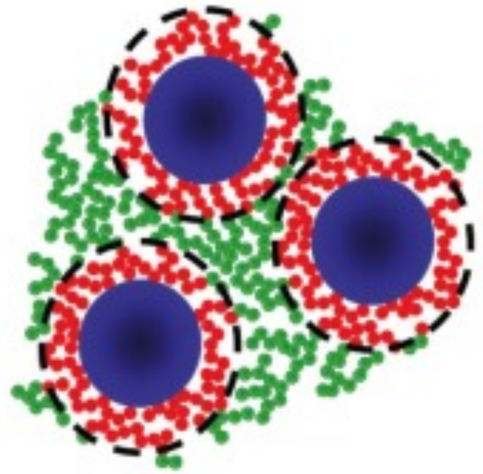


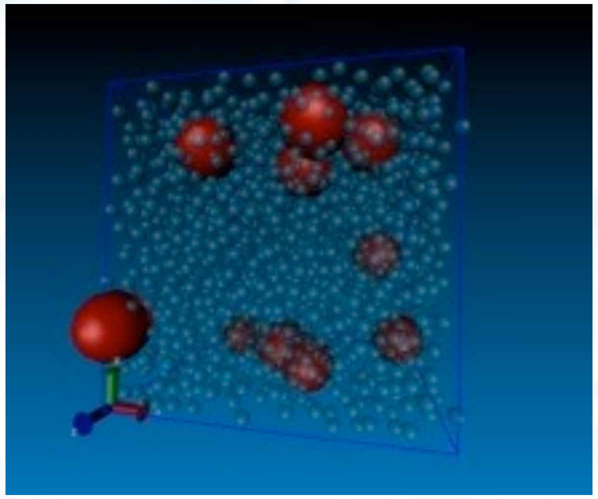
Annealed at 800C



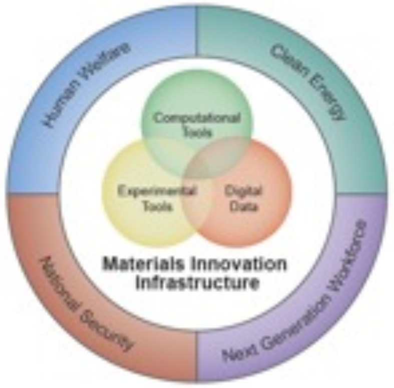
Properties







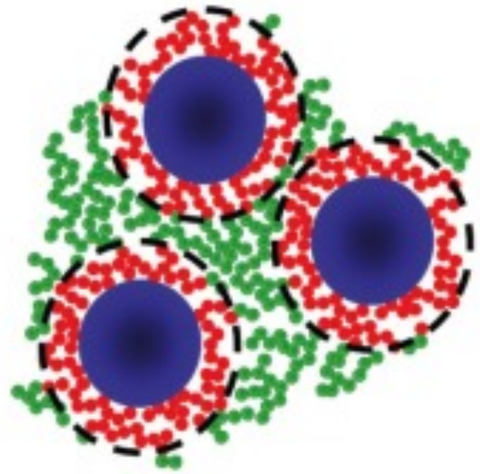
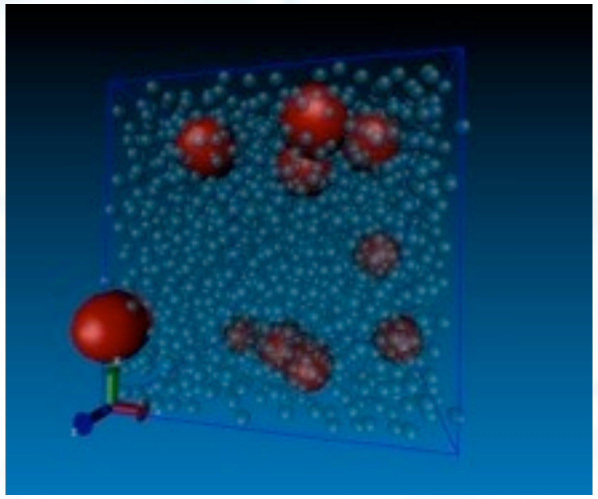
Advanced Composites Pilot for the Materials Genome Initiative (MGI)



Polymers Theory, Modeling and Simulation Group
Fred Phelan, Eric Lin, Jack Douglas, Martin Chiang (MML)
Ed Garboczi, Nick Martys (EL)

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

Polymers Division
MATERIAL MEASUREMENT LABORATORY


Advanced Functional Materials

- Materials Measurement Division, Materials Science and Engineering Division and Physical Measurement Laboratory

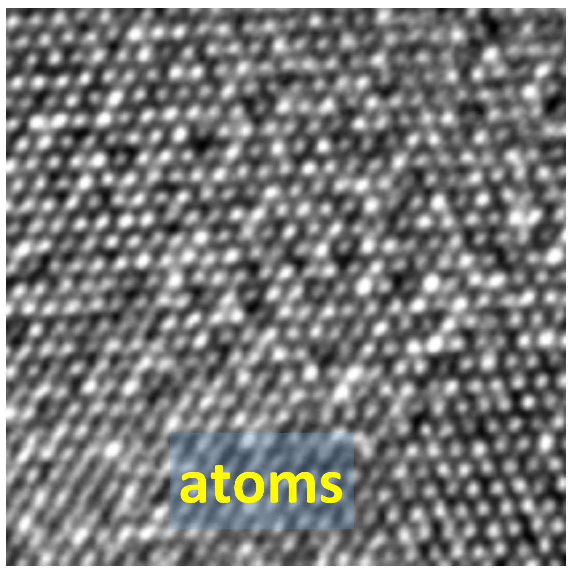
NIST
National Institute of Standards and Technology
U.S. Department of Commerce

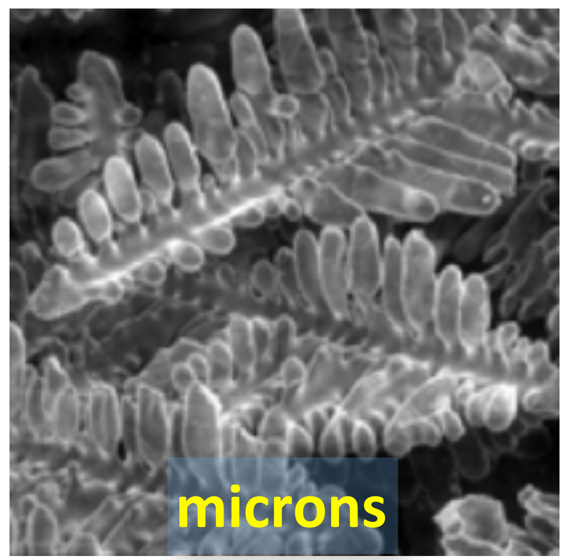
MATERIAL
MEASUREMENT
LABORATORY

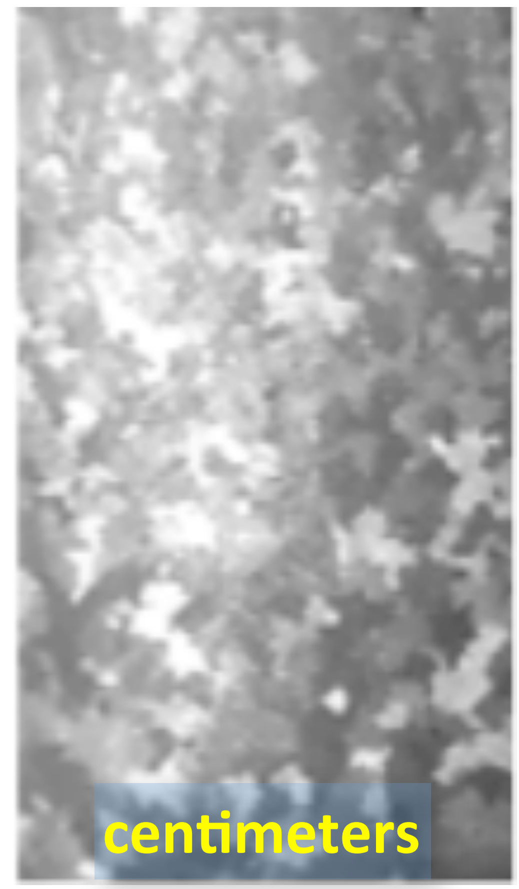
Pilot 3: Structural Alloys

Microstructure is the key to all properties !!

•Processing creates the microstructure – Question: how to make the physical model?
•Predict phases?
•Predict phase distribution?
•How will we store and manage information?
•What data infrastructure is needed?





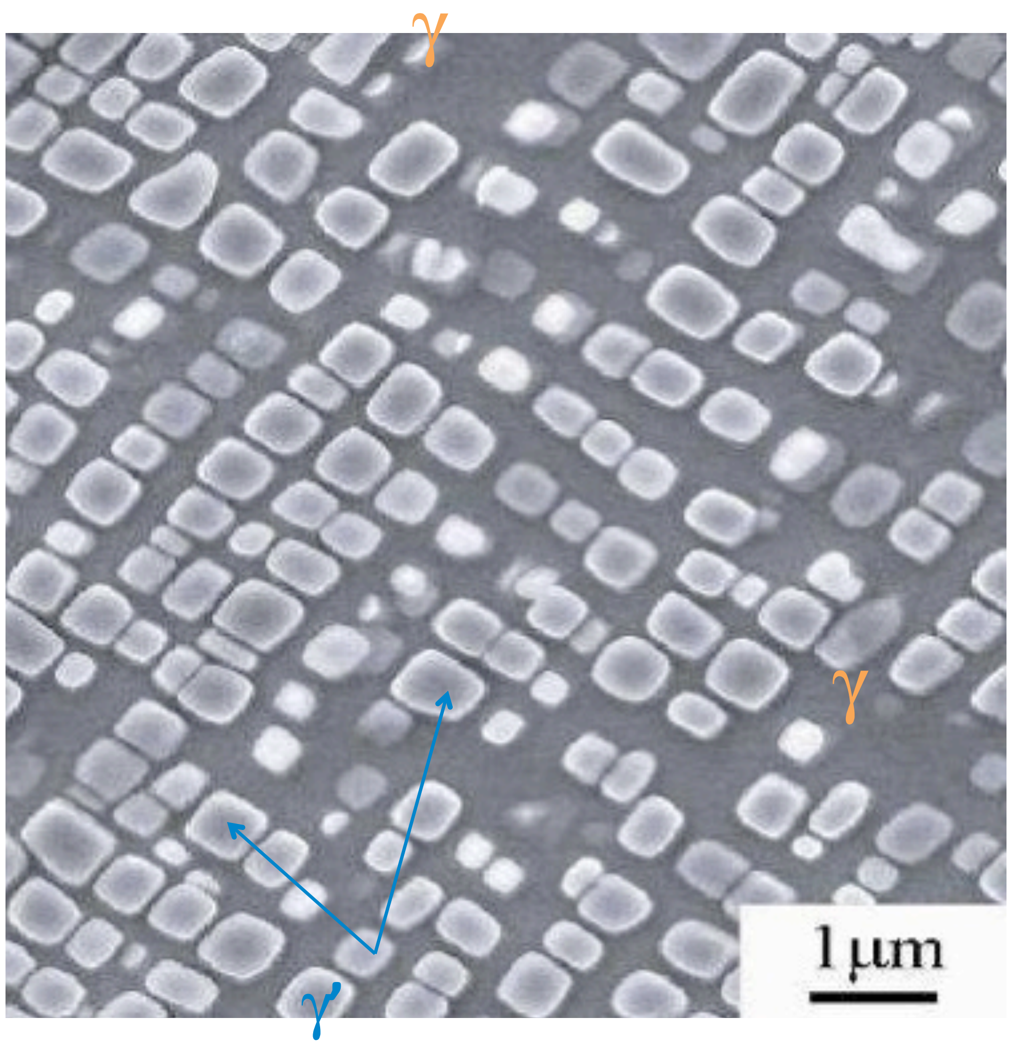


•Microstructure establishes properties
– Question: How to make the physical model?
•How does structure influence properties?
•Collaborations between ITL and MML with Northwestern, Penn State, Carnegie
with experimental, computational, informatics and uncertainty analysts.

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

MGI Test Case-Energy Efficient Turbines and Cars: Co-based γ-γ′ Superalloys and Lightweight Alloys

- Superalloys are typically Ni-Al-based, and consist of up to 10-12 elements
- Used in turbine blades for aircraft engines and energy production
- Properties: High temperature strength, creep resistance
- Structure: Disordered FCC γ-phase matrix with ordered γ′-phase precipitates as cubes
- In 2006, Sato et al.* discovered an analogous γ-γ′ microstructure in ternary Co-W-Al
  - Potentially higher temperature application; i.e. improved efficiency
  - Ideal prototype system for building MGI infrastructure
    - Establish the essential processing-structure-property relationships for building links between computational tools, experiment, and engineering design
    - Development of the protocol for efficient workflow management, data sharing, analysis, and application required for the MGI infrastructure



K. Ishida, *MRS Proc.*, **1128**, (2009) U06-06

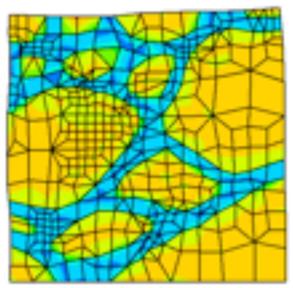
* J. Sato, et al., *Science*, **312**, (2006) 90

Macroscopic Materials Properties

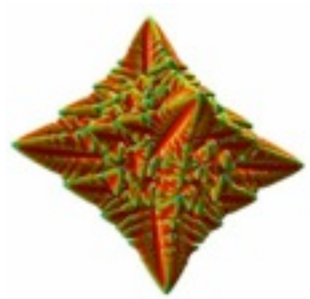
Structural Alloys Pilot

MATERIAL PROPERTIES

Collaborations
Universities
Industry
MML (3 Divisions)
ITL

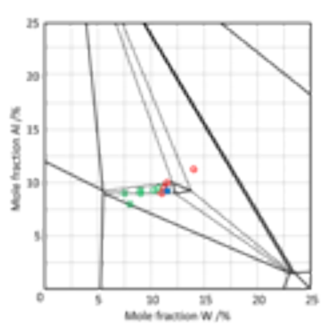


Structure-Property Relationships
Computation, Theory, Expt

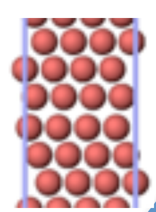


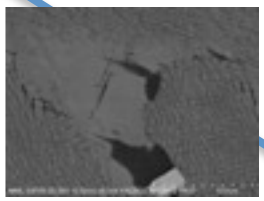
Processing-Structure Relationships
Phase-field, FE, etc.

Crosscuts
Repositories
Workflow Tools
Standards
Protocols
Uncertainty
Dissemination



Composition Dependent Databases
CALPHAD, Diffusion Mobility





HT 900 °C, 1000 h

Atomic Structure and Properties

Atomic Scale Models
Ab initio, Molecular Dynamics, Monte Carlo, Embedded Atom

Experimental Inputs
Crystal Structure, Diffusion, Thermochemical Data

STOP!