The Materials Genome Initiative, Data, Open Science & NIST

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

The Materials Genome Initiative, Data, Open Science & 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, Data, Open Science & 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 Perfect is the Enemy of the Good -Voltaire

Outline/Goals

- 30 month perspective about the National effort
- Articulating our Story
- Data, Culture and what's coming
- Share info about the NIST effort
  - Pilots and Integration
  - Supporting Activities
  - External Engagement

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

Outline/Goals

- 30 month perspective about the National effort
- Articulating our Story
- Data, Culture and what's coming
- Share info about the NIST effort
  - Pilots and Integration
  - 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

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
3. 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 U.S. MATERIALS GENOME INITIATIVE

"...to discover, develop, and deploy new materials twice as fast, we're launching what we call the Materials Genome Initiative"
– President Obama, 2011

Meeting Societal Needs

Advanced materials are at the heart of innovation, economic opportunities, and global competitiveness. They are the foundation for new capabilities, tools, and technologies that meet urgent societal needs including clean energy, human welfare, and national security.

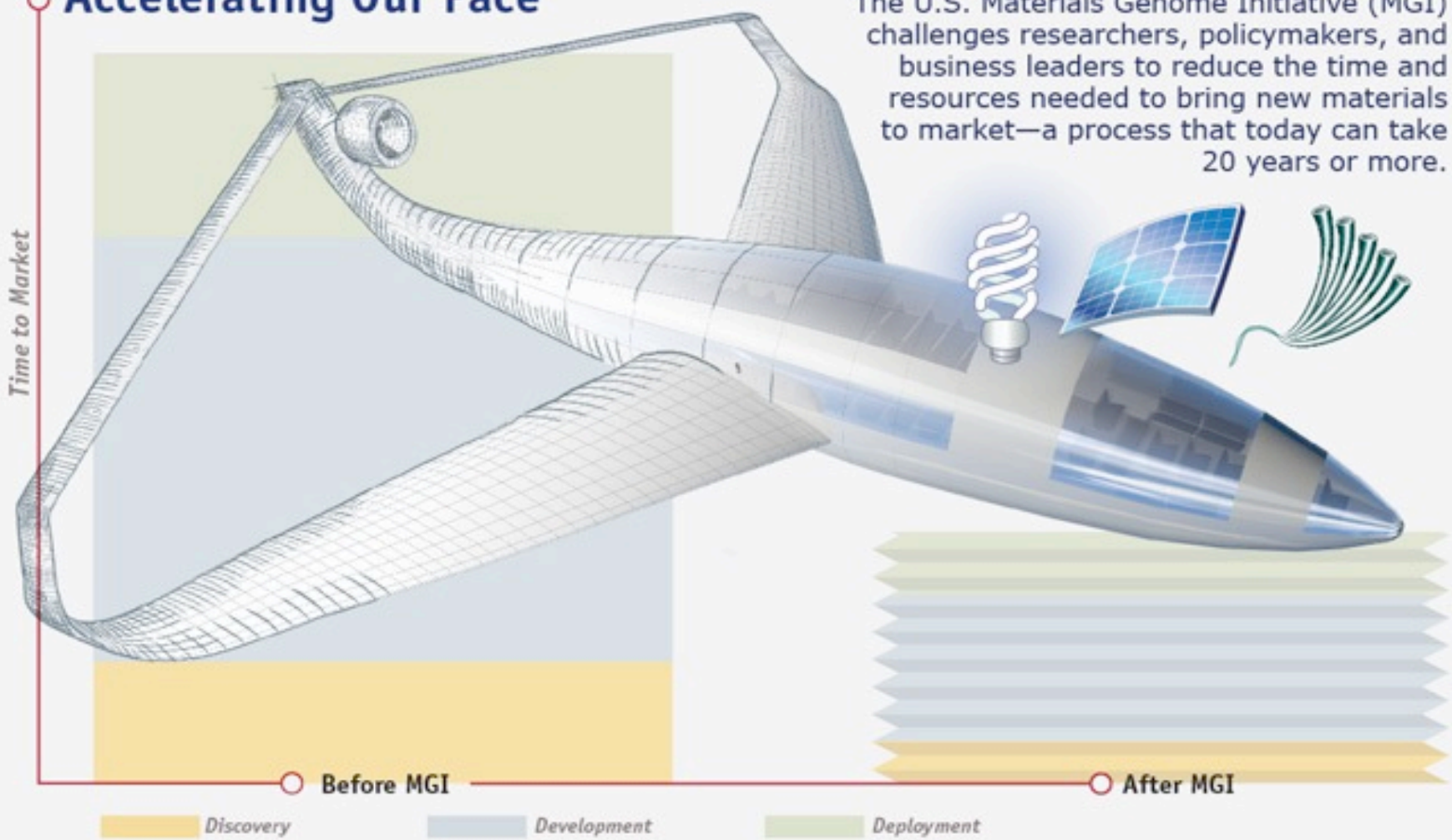
Clean Energy

Human Welfare

National Security

Accelerating Our Pace

The U.S. Materials Genome Initiative (MGI) challenges researchers, policymakers, and business leaders to reduce the time and resources needed to bring new materials to market—a process that today can take 20 years or more.



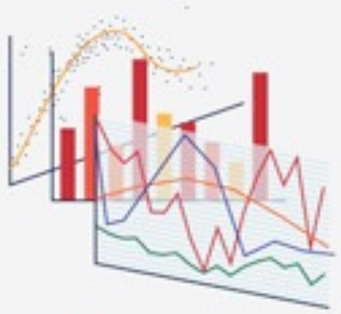
Building Infrastructure for Success

The MGI is a multi-agency initiative to renew investments in infrastructure designed for performance, and to foster a more open, collaborative approach to developing advanced materials, helping U.S. Institutions accelerate their time-to-market.

Computational tools

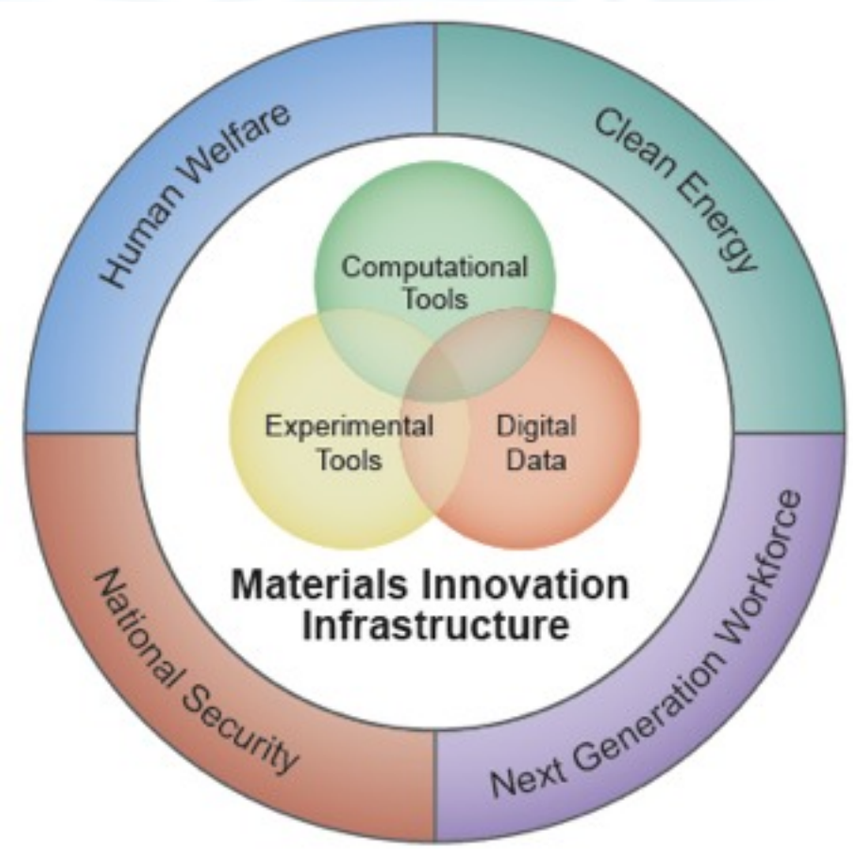
Experimental tools

Collaborative networks



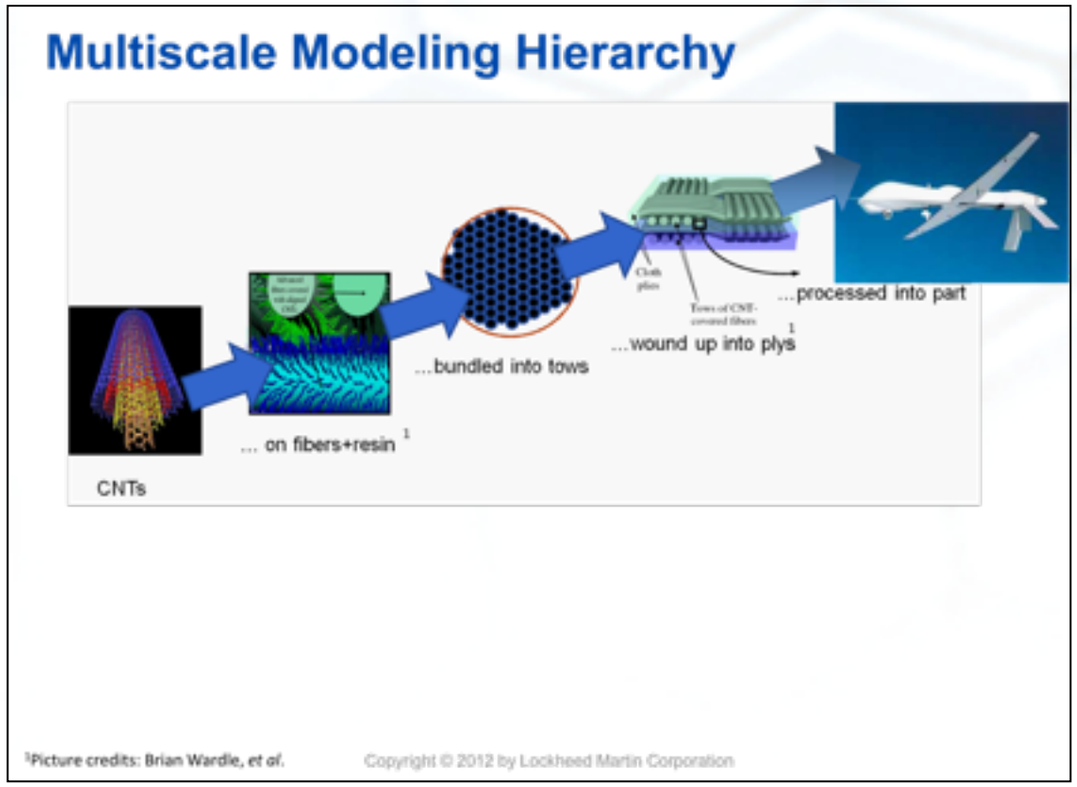
Digital data

Multiple Components to the MGI!



Material "Informatics"
National Infrastructure for Data Sharing and Analysis

Courtesy: BASF



Multiscale Modeling
Atoms to Component Description of Materials

Functioning Infrastructure

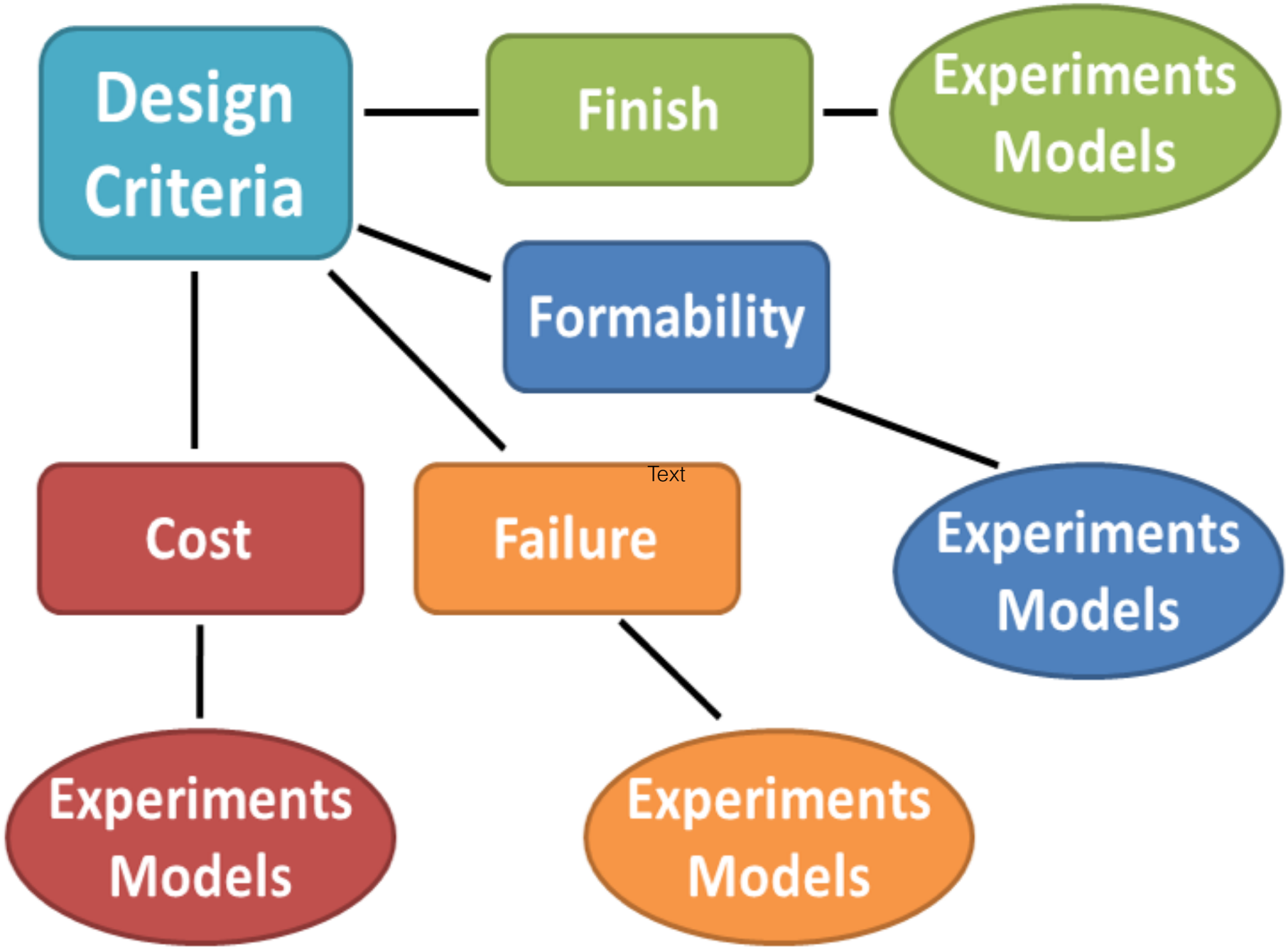


Figure 1. Schematic of how the design criteria for a given material dictate the needed material properties and thus define the needed experiments, models and data.

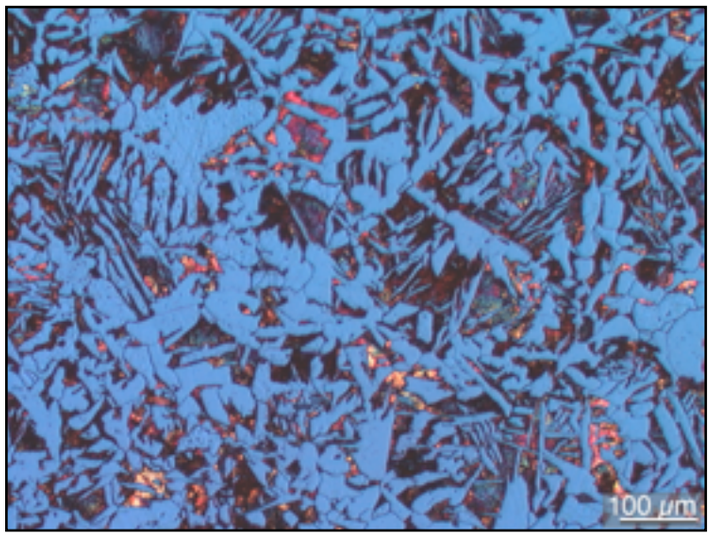
NIST
National Institute of
rds and Technology
epartment of Commerce

Why an MGI Now?

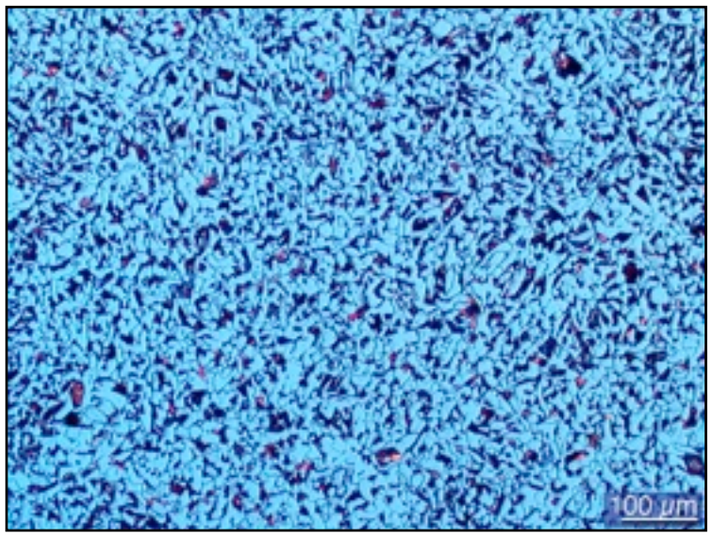
Materials Are Complicated Systems
Modeling is a Challenge

- 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

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

Why Materials Design?

Materials development is a “needle in haystack” problem

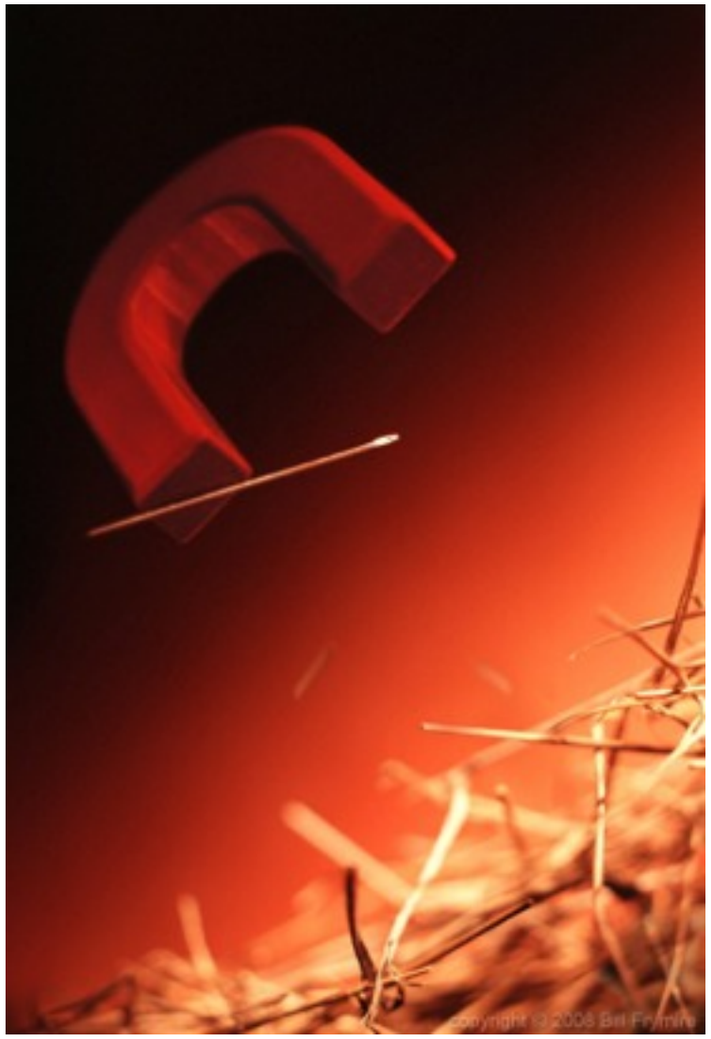
- If an empirical approach is taken to develop new materials, number of possible compositions to evaluate is overwhelming.
  - e.g., superalloys, which often contain 12 components, if one only considers the *current composition ranges* the number of compositions to investigate exceeds $1x10^8$

Why Materials Design?

Materials development is a “needle in haystack” problem

- If an empirical approach is taken to develop new materials, number of possible compositions to evaluate is overwhelming.
  - e.g., superalloys, which often contain 12 components, if one only considers the *current composition ranges* the number of compositions to investigate exceeds $1x10^8$

The right tools:
MGI ?



What is the MGI?

The MGI is building the infrastructure to design and fabricate new materials by leveraging off the enormous growth in computational power with a commensurate focus on integrated modeling, simulation, and data analytics

1986!!

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

What is the MGI?

The MGI is building the infrastructure to design and fabricate new materials by leveraging off the enormous growth in computational power with a commensurate focus on integrated modeling, simulation, and data analytics

1986!!

NIST
National Institute of
ndards and Technology
S. Department of Commerce

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 and Data Workshop
- First Grand Challenge Summit Concluded
- Second (on soft materials) in November
- Goal of Rolling out strategic plan in early 2014 (small shutdown glitch)

NIST
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.

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

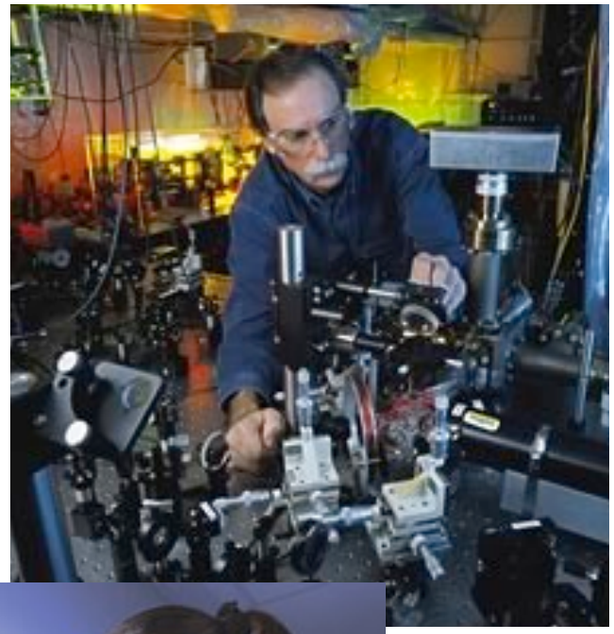
in ways that enhance economic security and improve our quality of life

©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 Relationship between Models and Data and An Evolution towards Open Science

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

The Relationship between Models and Data and An Evolution towards Open Science

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

Hopefully food for thought

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 metadata!

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

The Model is the Central Concept in Science

- Sometimes it's a Theory (essentially accepted as fact)
- Sometimes it's a Hypothesis
- Sometimes it's an unacknowledged assumption
- Simulation = Implementation of a Model in silico
- Experiment = Test of Model in vivo
- Data is the output

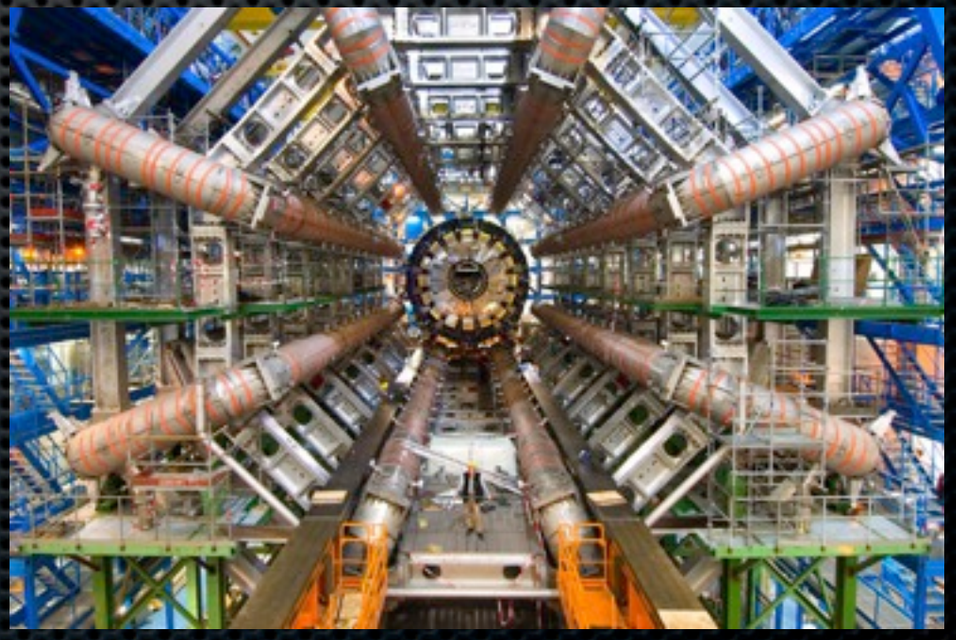
Old School, Computation

- 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 should we be doing?

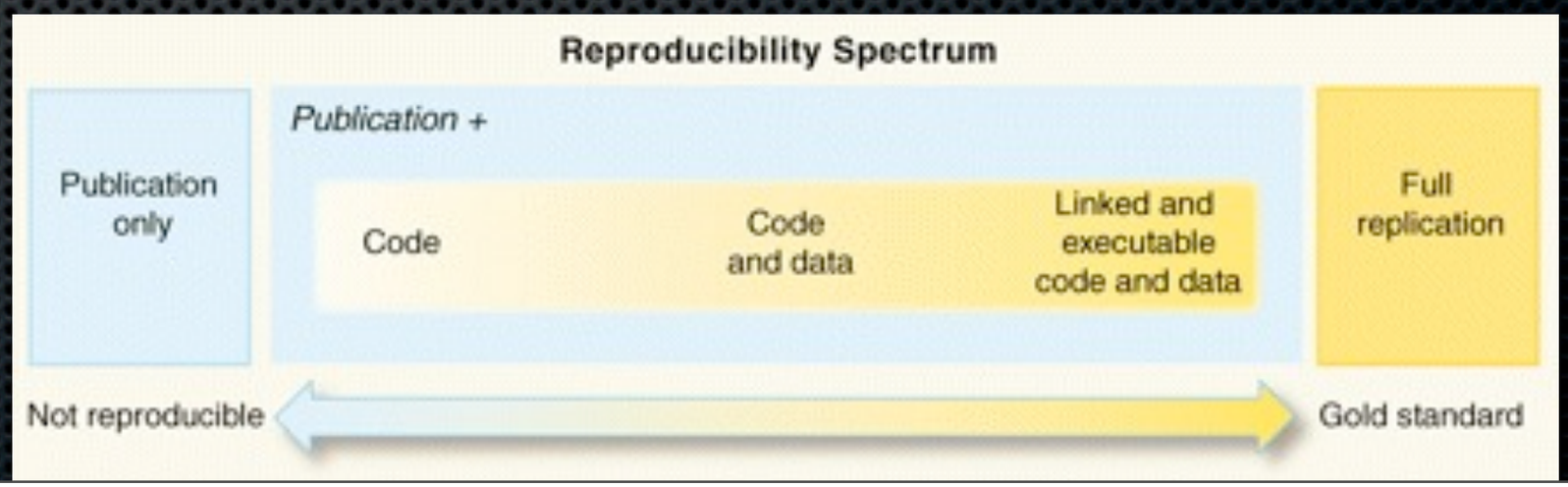
Science is characterized by the iteration of experiment and models, yielding higher fidelity with lower uncertainty.

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

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



What should we be doing?

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

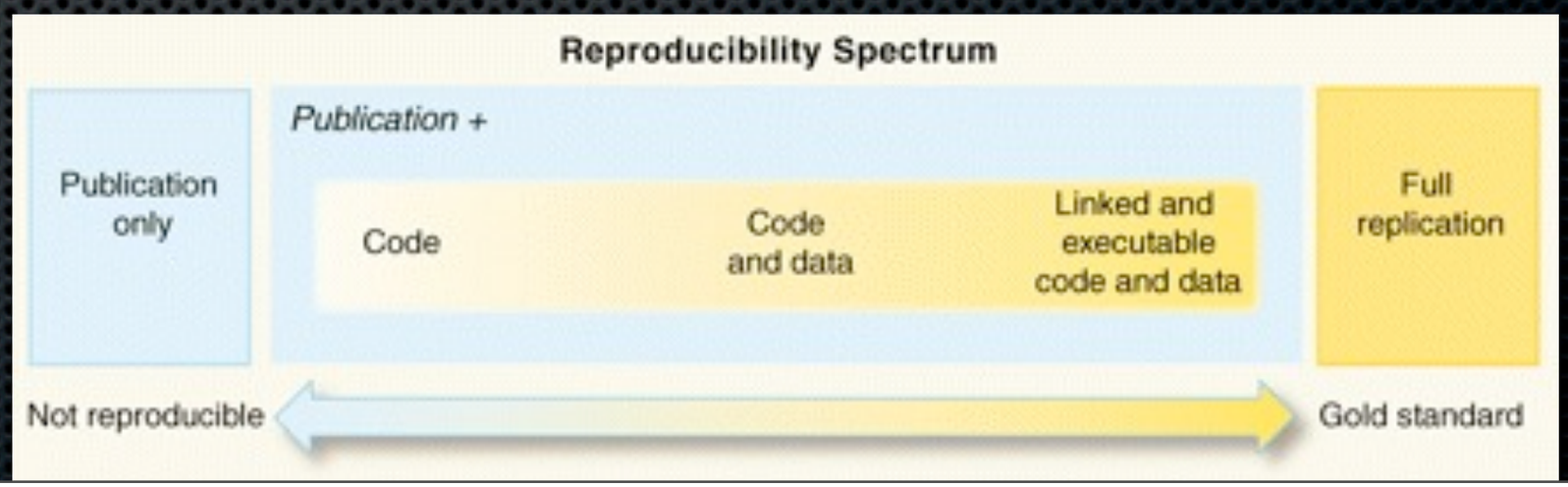
The measurement or computation of a quantity (data) is generally *meaningless* without the associated quantifying model that defines both the data and its uncertainties.

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

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

PUBLISHING

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



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?

What Should NIST Do?

Data Sharing is Important Beyond MGI

OSTP "Public Access" Memo
Feb 22, 2013

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
constraints possible and consistent with law and the
federally funded scientific research are made availab
the scientific community. Such results include peer-

Scientific research supported by the Federal Govern
drive our economy. The results of that research beco
for progress in areas such as health, energy, the envi

Access to digital data sets resulting from federally f
resources and efforts on understanding and exploitin
data underpins the forecasting industry, and making
spawned many biotechnology innovations. In additi
publications and scientific data in digital formats wi
services related to curation, preservation, analysis, a
publications and data for re-use through preservatio
the impact and accountability of the Federal researc
scientific breakthroughs and innovation, promote en
growth and job creation.

OMB "Open Data" Memo
May 9, 2013

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

May 9, 2013

M-13-13

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Sylvia M. Burwell

Officer

ficer

ice of Information and Regulatory Affairs

ging Information as an Asset

ource and a strategic asset to the Federal Government, its
at the Federal Government is taking full advantage of its
s and agencies (hereafter referred to as "agencies") must
s life cycle to promote openness and interoperability, and
Managing government information as an asset will increase
e services, support mission needs, safeguard personal
luable government information.

ible, discoverable, and usable by the public can help fuel

Executive Order
May 9, 2013

For Immediate Release

May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

- - - - - - -

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT
FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles. Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

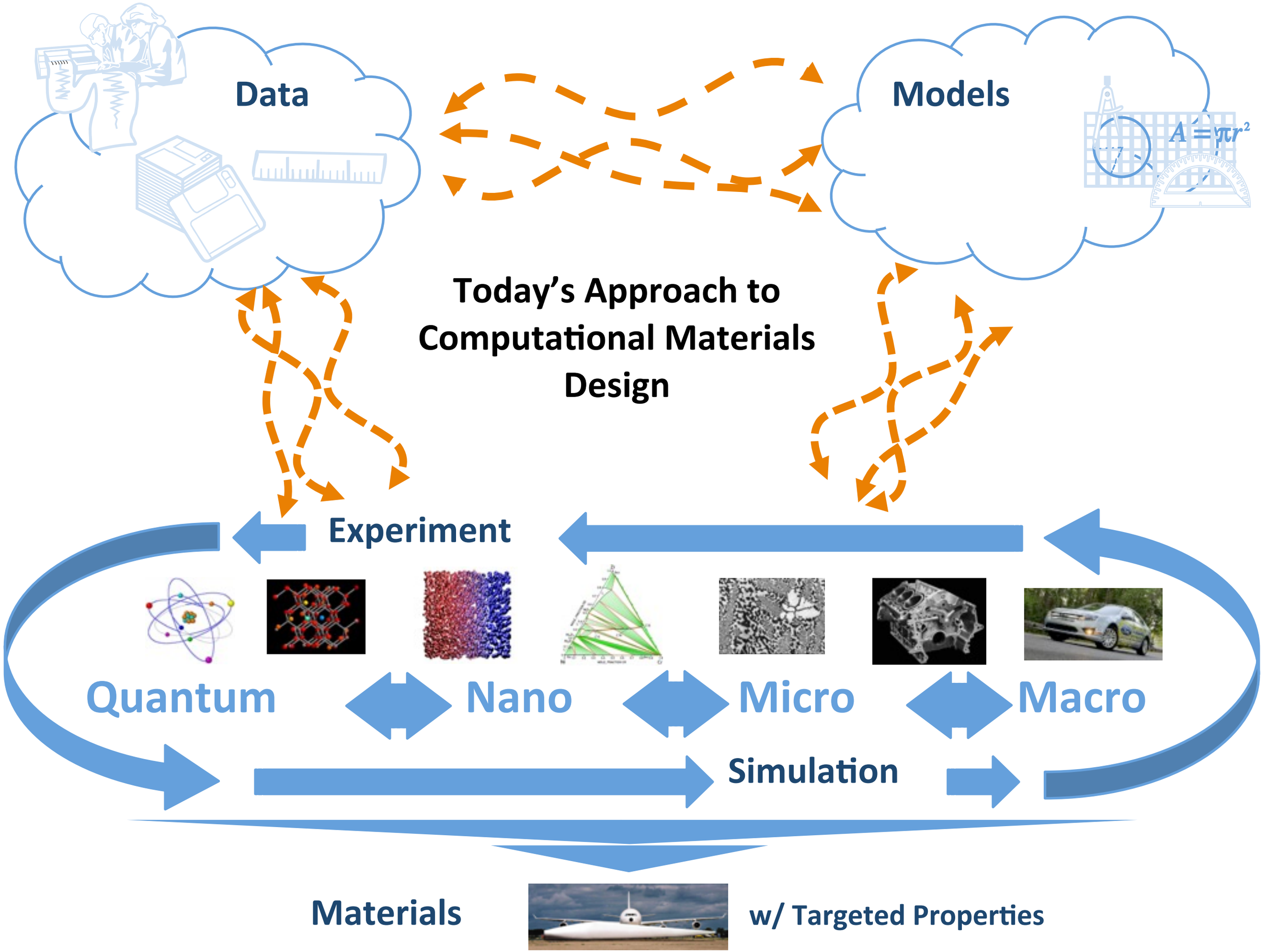
Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

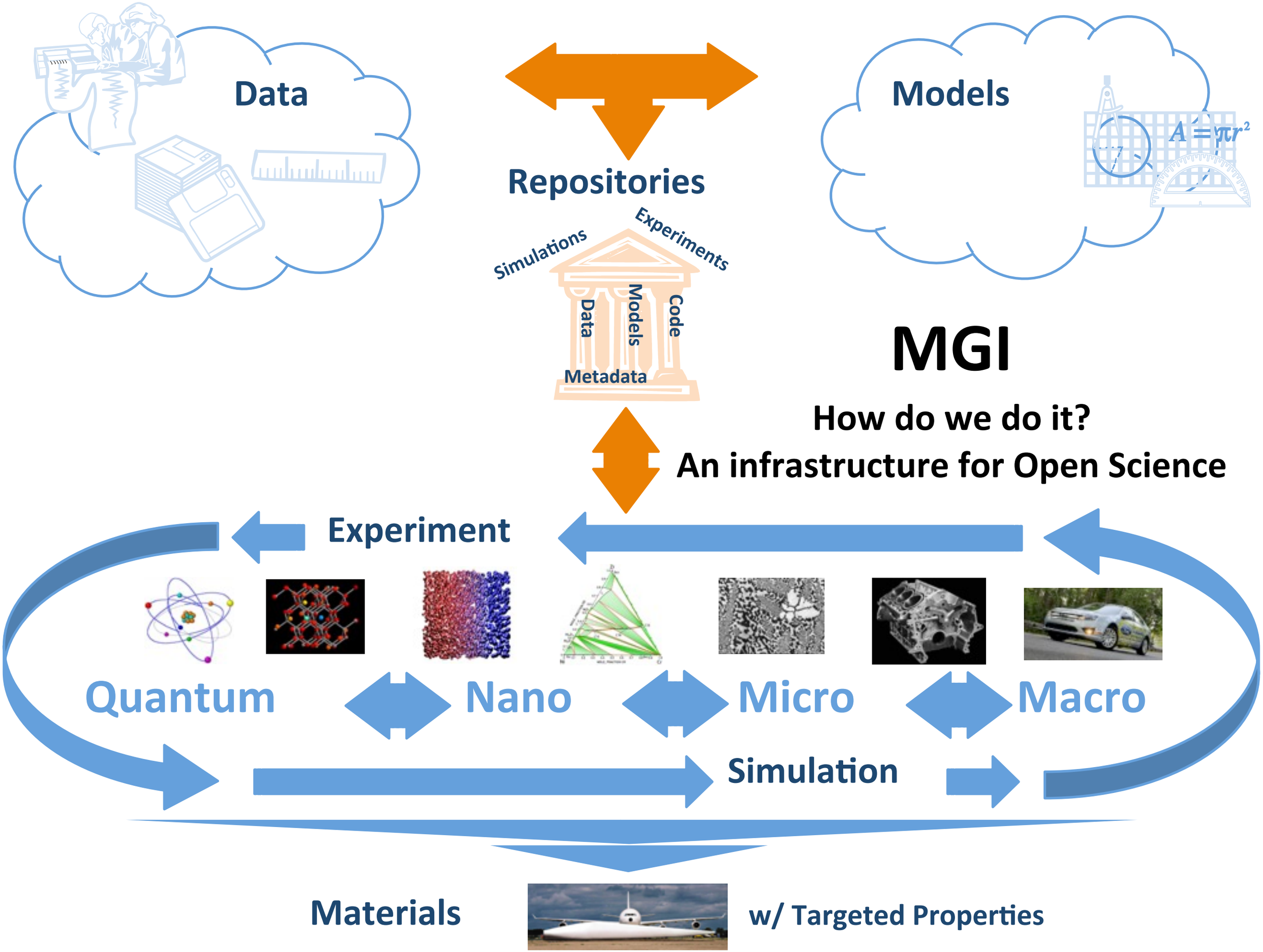
Data Sharing and Management Snafu in 3 Short Acts

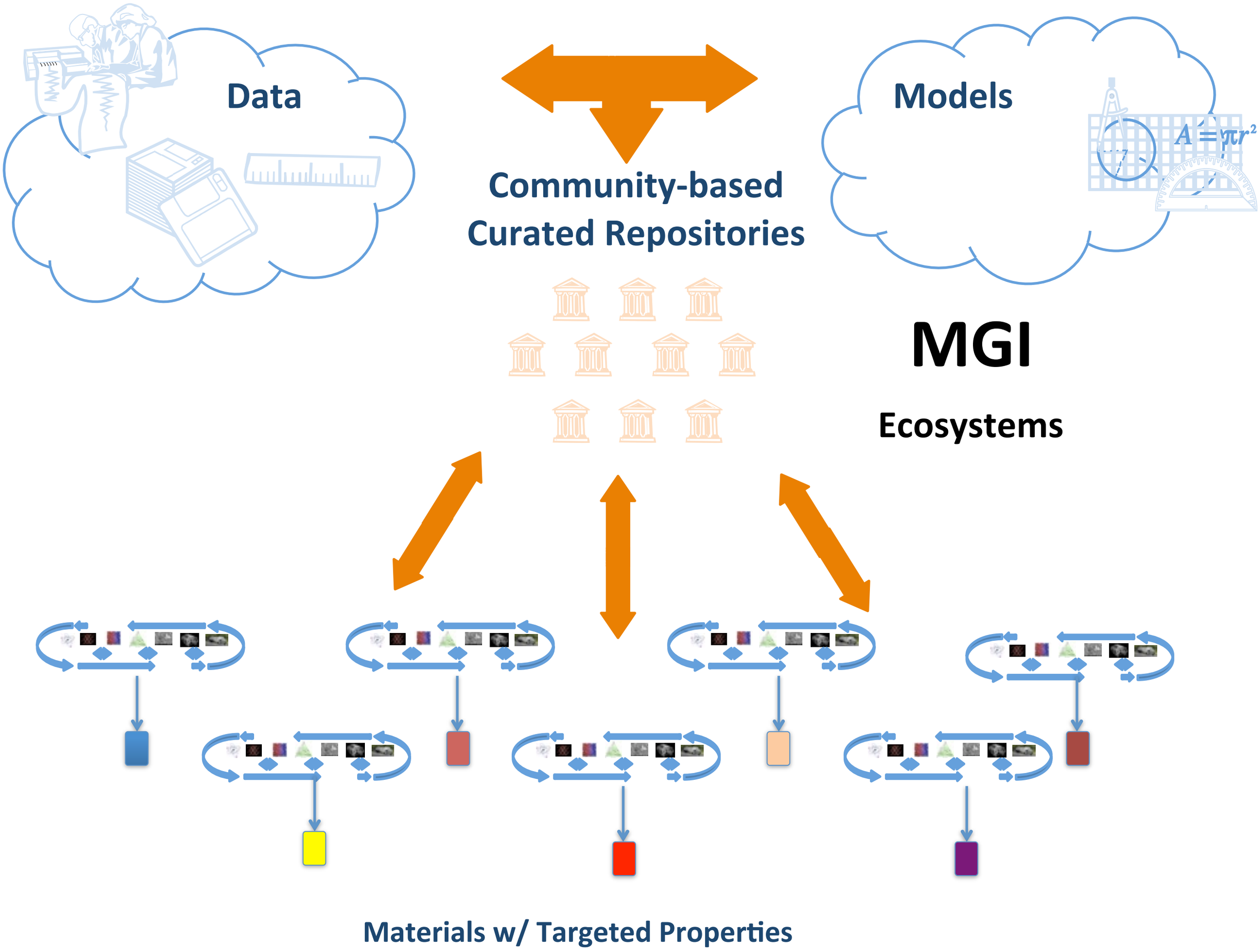
by Karen Hanson, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries

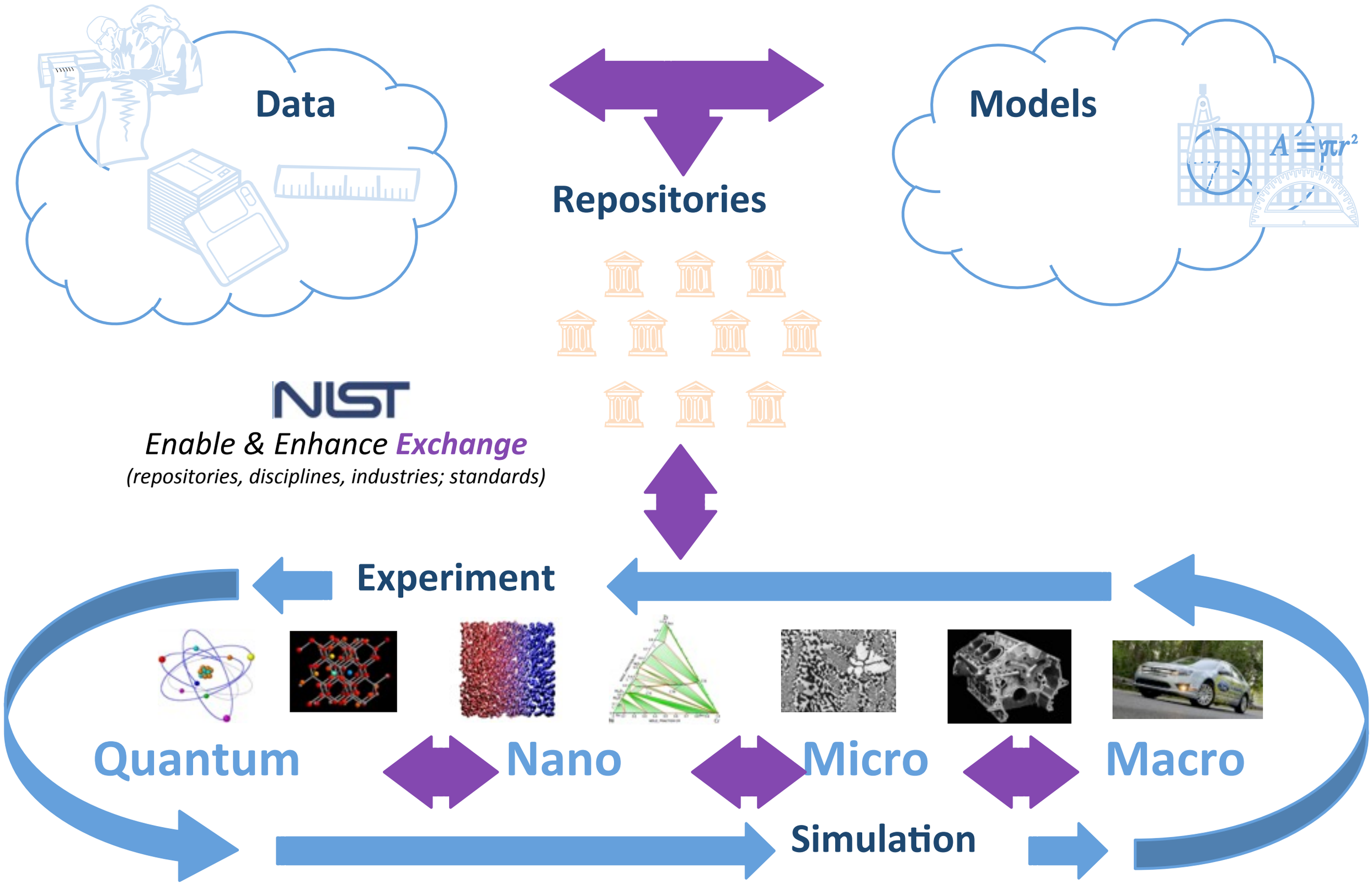
August 3, 2012 (Last Update: December 12, 2012)

xtranormal

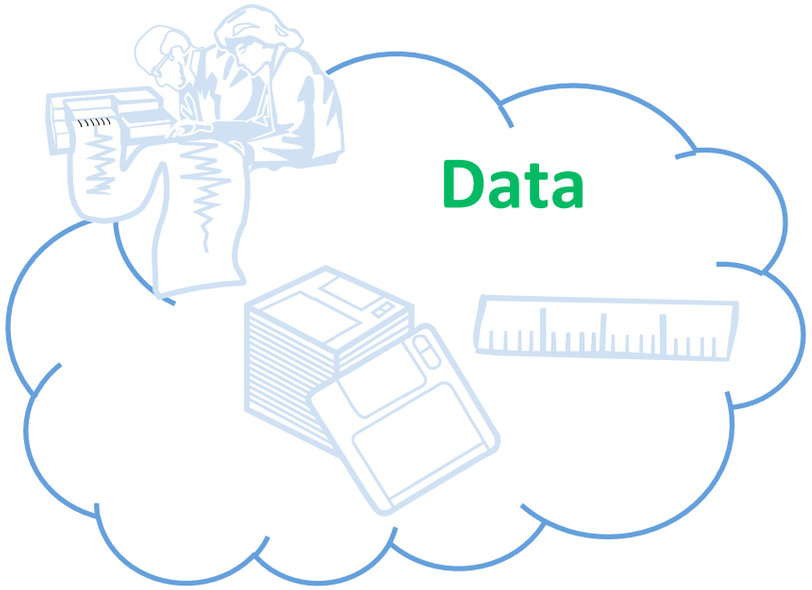

Data
Models
$A = \pi r^2$
Today's Approach to Computational Materials Design
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties

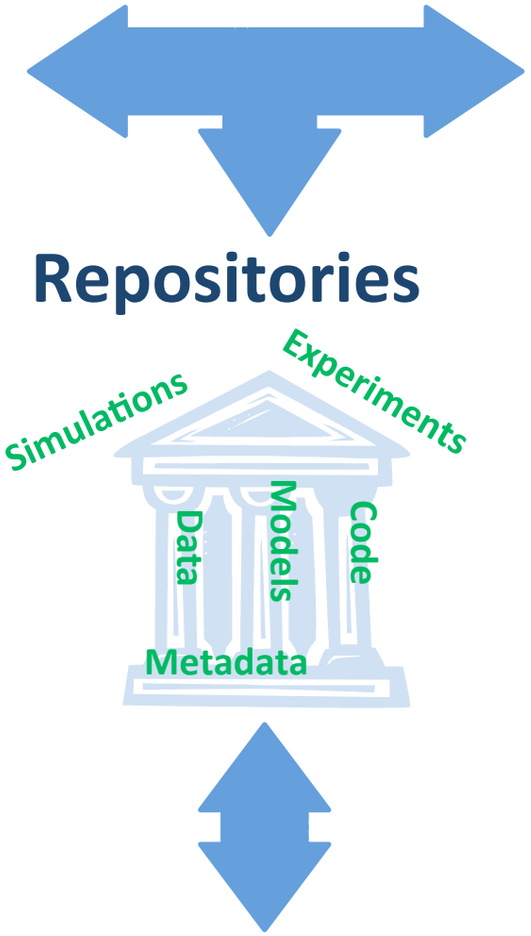

Data
Repositories
Models
$A = \pi r^2$
Simulations
Experiments
Data
Models
Code
Metadata
MGI
How do we do it?
An infrastructure for Open Science
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties

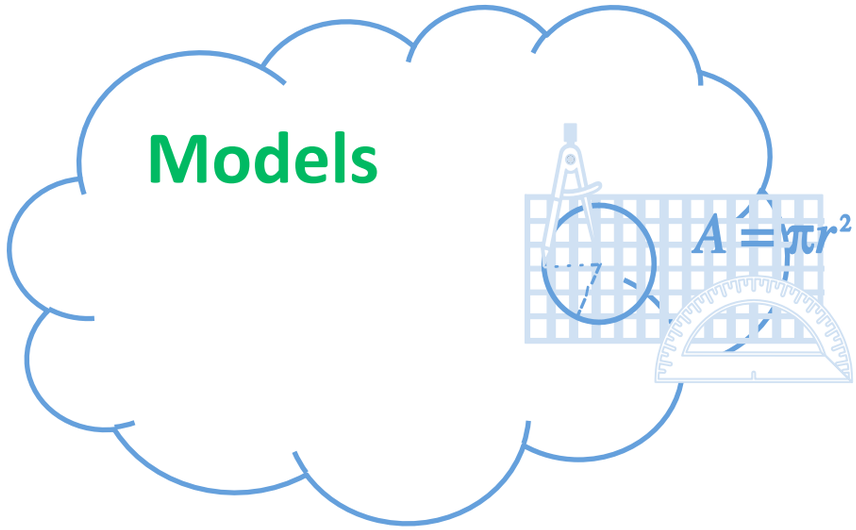

Data
Models
$A = \pi r^2$
Community-based
Curated Repositories
MGI
Ecosystems
Materials w/ Targeted Properties



Materials w/ Targeted Properties



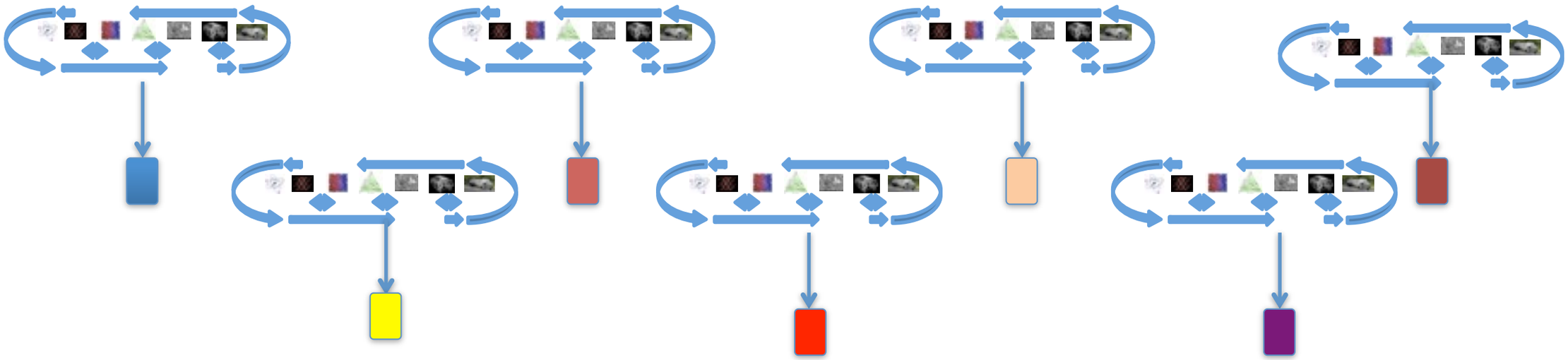




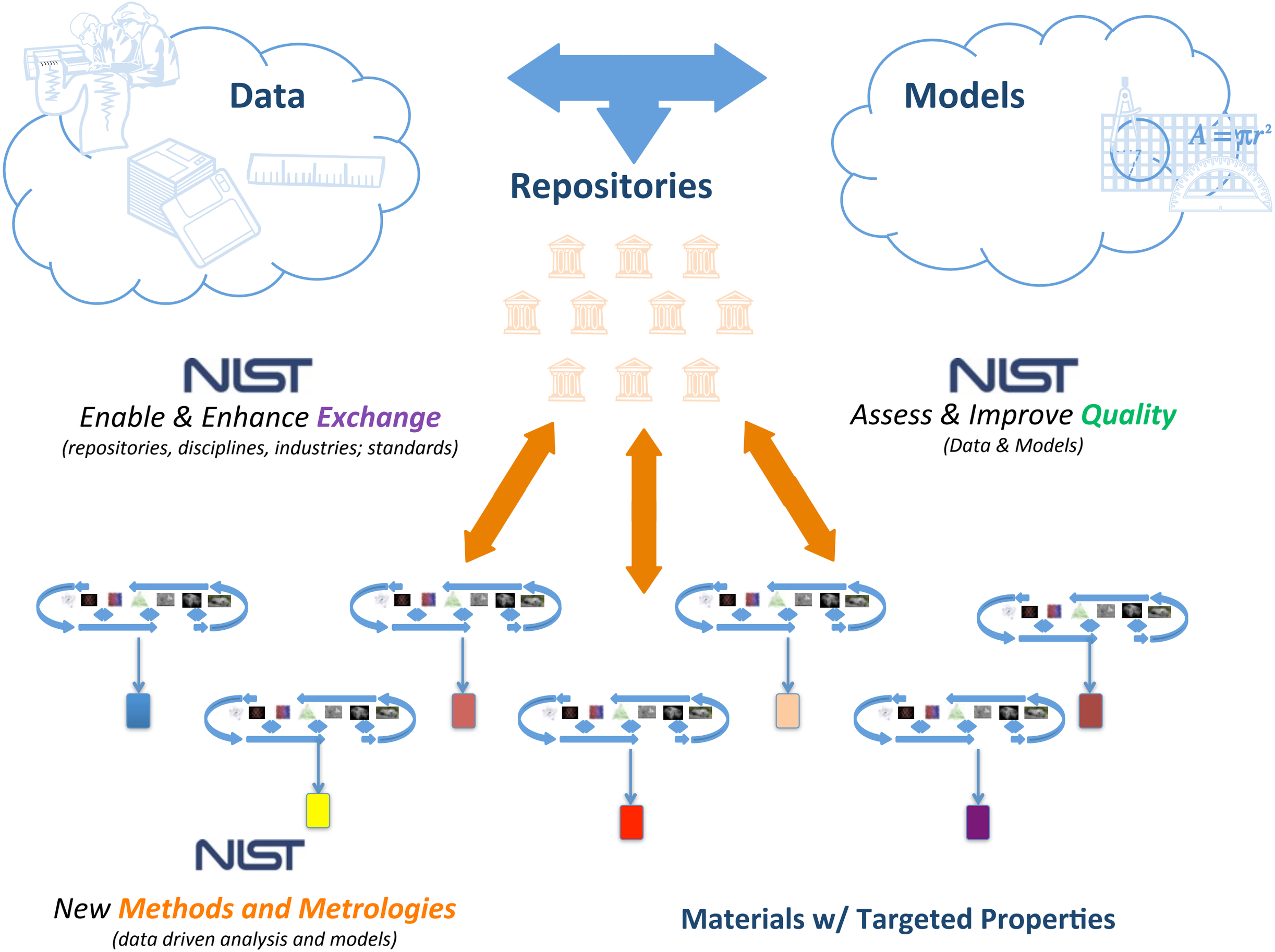
NIST

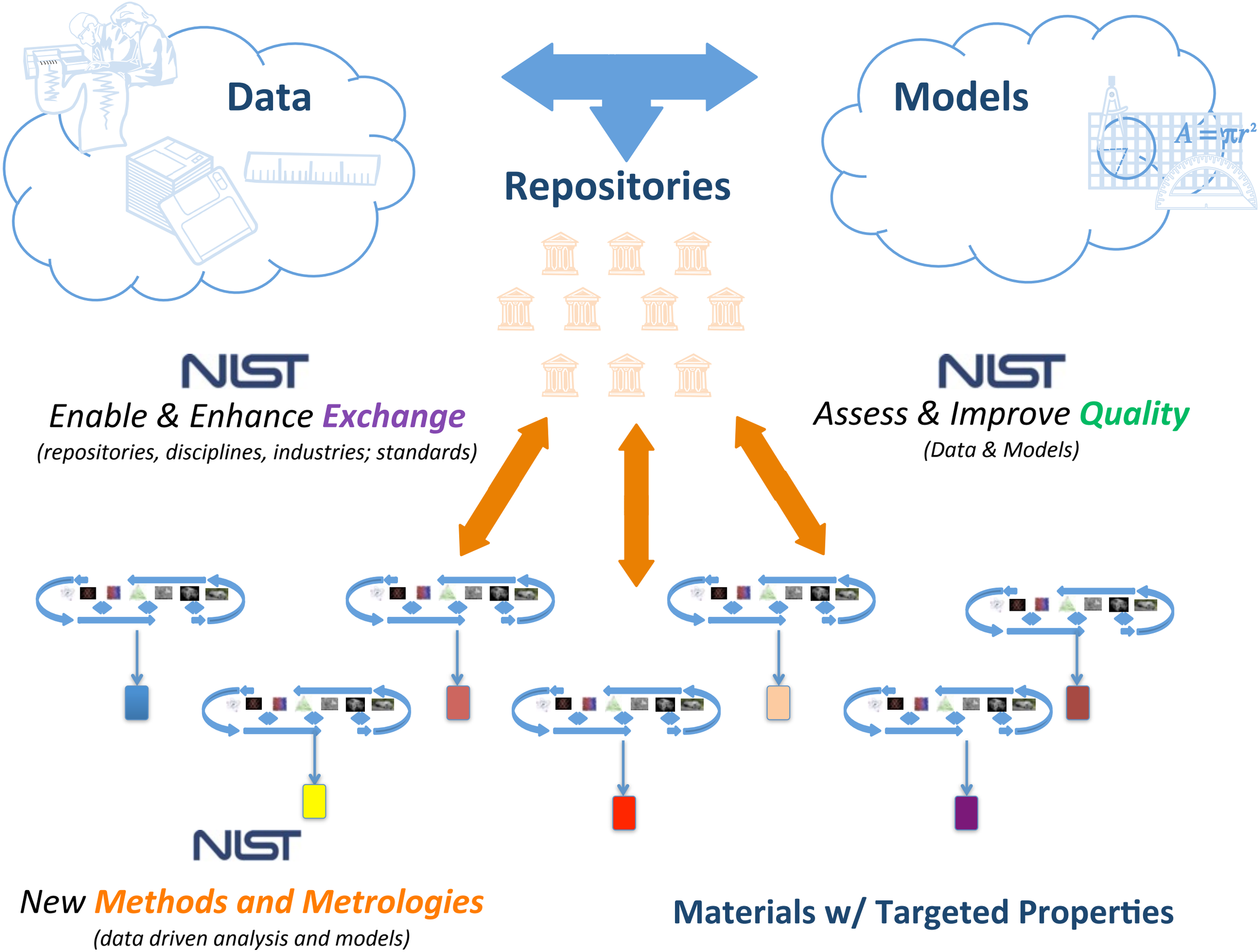
*Assess & Improve **Quality***

(Data & Models)



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


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

SCOPE: Goals of Initiative (9.2 M/yr)

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

Goal 1: NIST establishes *essential materials data and model exchange protocols*

Goal 2: NIST establishes the *means to ensure the quality* of materials data and models

Goal 3: NIST establishes *new methods, metrologies and capabilities* necessary for accelerated materials development.

Enable & Enhance ***Exchange***

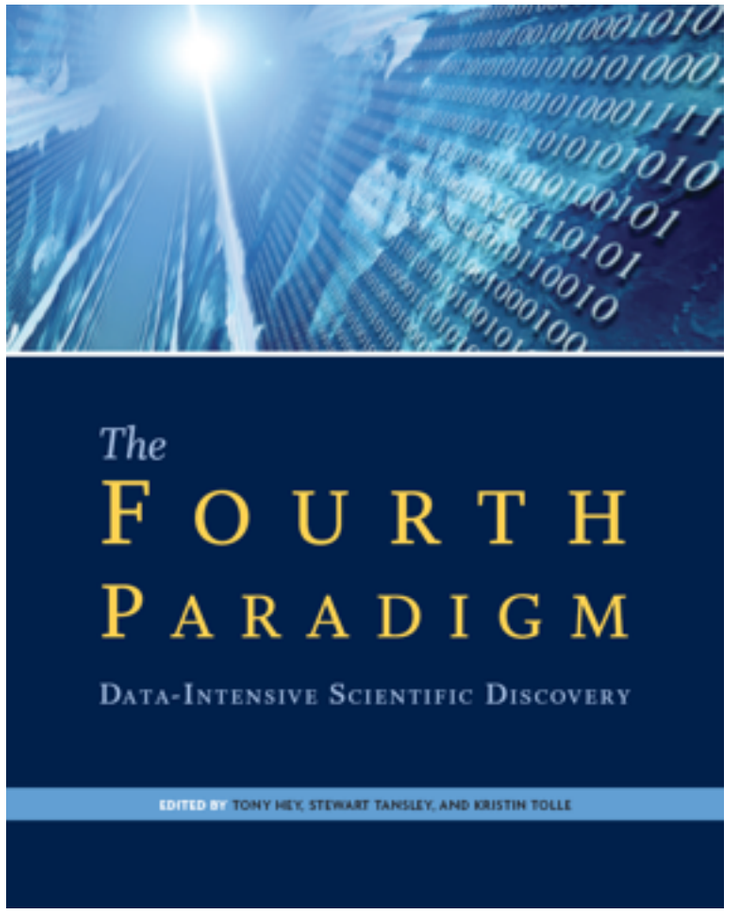
- Develop and deploy repositories
- Develop and disseminate materials informatics infrastructure
  - Enable data discovery through tools and standards
  - Capture data from scientific workflows and archival sources
  - Engage with stakeholders to determine needs and disseminate best practices
- Integrate 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
- Enable Crowdsourced/Open Science
- Achieve targets in Materials by Design/ICME



Another way of telling the story

- Industry needs good data and models
  - How do they get them?(exchange)
  - Can they trust them? (quality)
  - How can they use them best and who has the talent? (new methods)

Convene and Engage

White House Event (May 2012)

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

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 Workshops

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

Grand Challenges: Articulation

- Grand Challenge workshop series to seek stakeholder input on critical industrial problems that MGI should target
- Follow-up Webinars to engage wider community
- Multi-Agency sponsorship (including NSF, DOE, NIST....)
- Keep and eye on www.whitehouse.gov/mgi and www.nist.gov/mgi
- Can we inspire?

NIST Supports New Open Collaboration Modalities

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
National Institute of Standards and Technology
U.S. Department 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

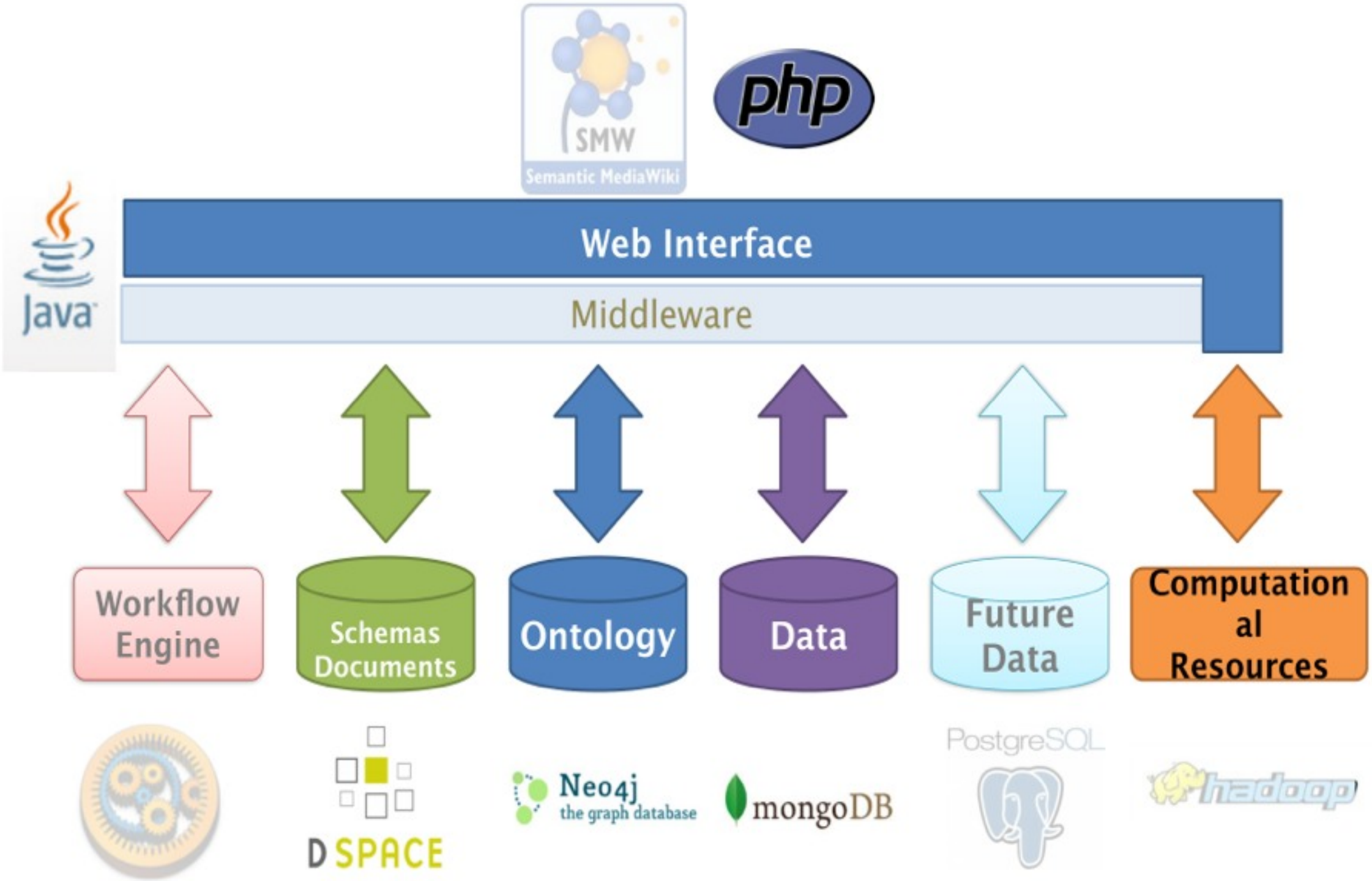
NIST
National Institute of Standards and Technology
U.S. Department of Commerce

NIST Center of Excellence for Advanced Materials

Announcement targeted for late November, 2013

New RDA IG

Planned MGI Architecture



STOP!

www.nist.gov/mgi