The MGI and the Materials Data Infrastructure

James A. Warren
Director, Materials Genome Program
National Institute of Standards and Technology
Executive Secretary, NSTC Subcommittee on the MGI

Scope

- [ ] Define MGI and NIST Role
- [ ] Status of the Infrastructure (ready to go but not even close to done)
- [ ] National Materials Data Network
  - [ ] Meeting the Needs (Human Factors)
- [ ] Preview of MGI 2.0
- [ ] AI as a Driver and transformations in materials R&D


NEW CUYAMA
Population 562
Ft. above sea level 2150
Established 1951
TOTAL 4663

The Materials Genome Initiative

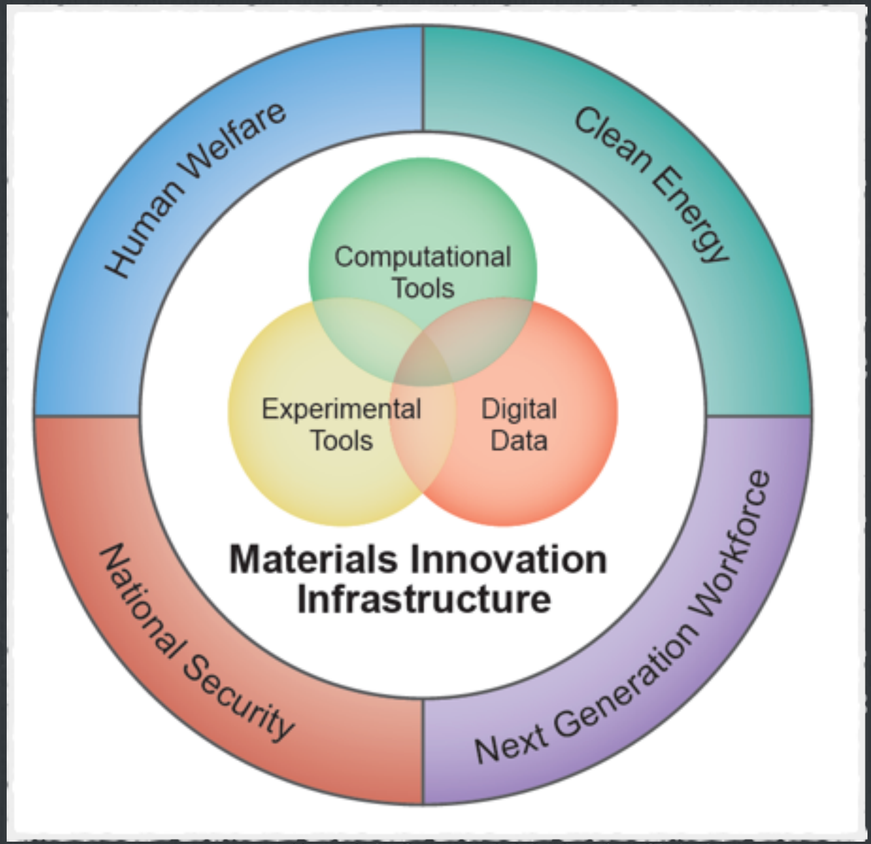
A Multi-Agency Effort

It’s an Initiative

DOE, NSF, NIST, DOD, NASA, FDA, NIH…

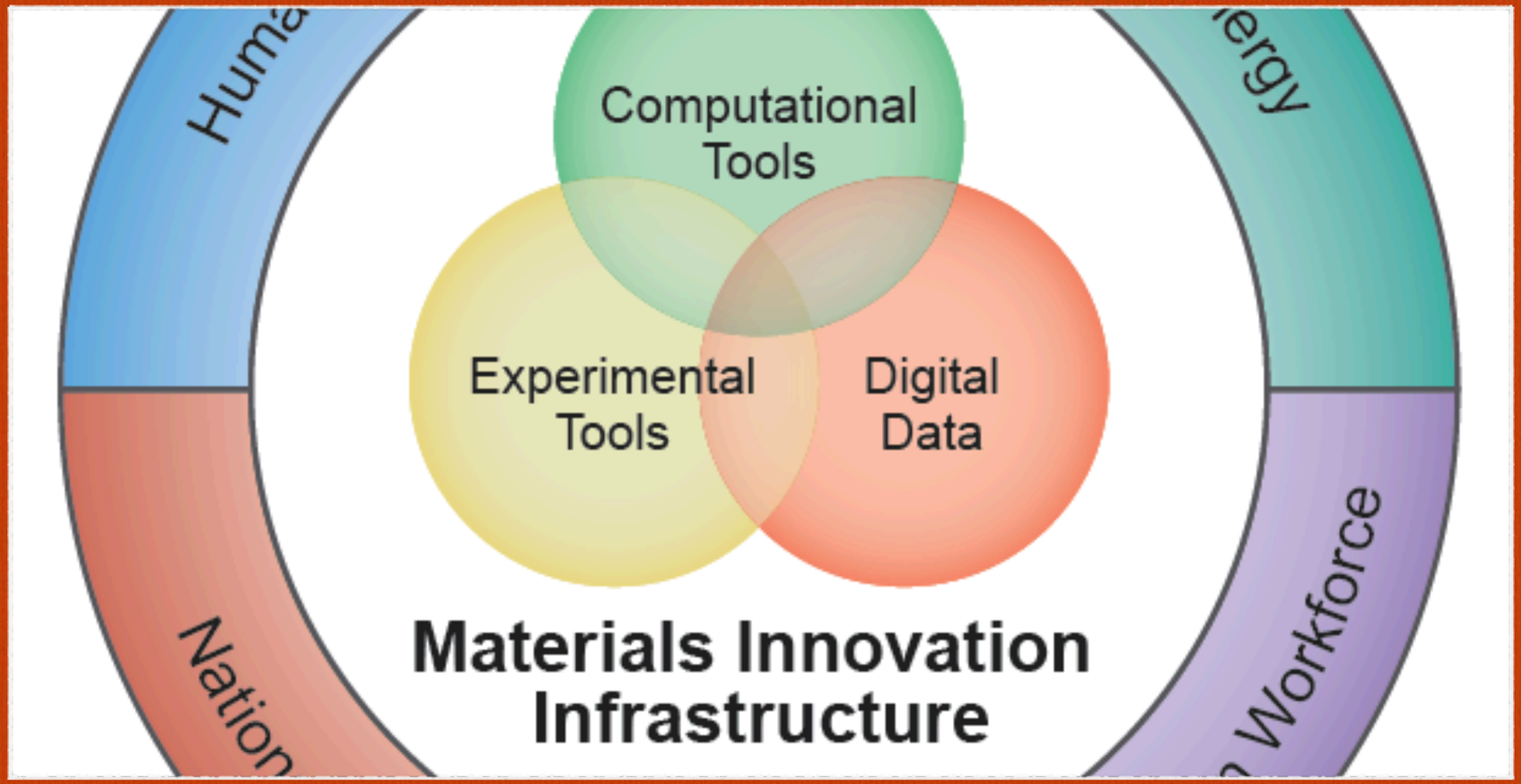
3 years into a new adminstration

To decrease time-to-market by 50% while <$$

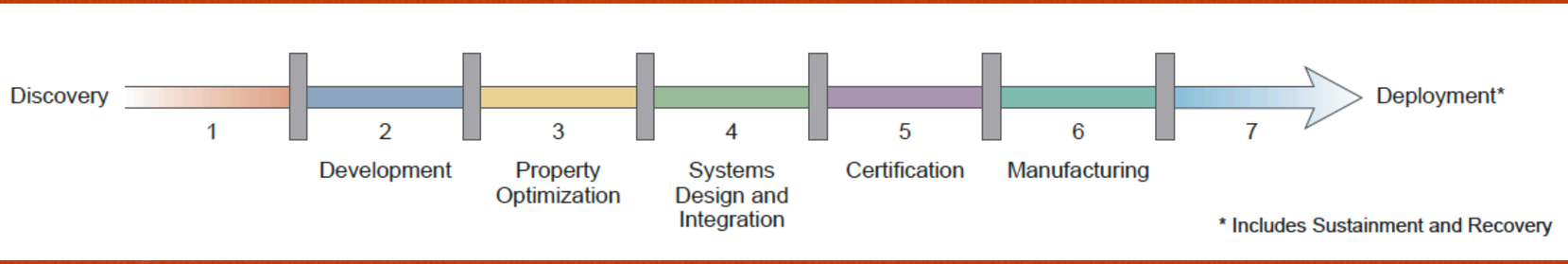


Materials Genome Initiative for Global Competitiveness

- ☐ Develop a Materials Innovation Infrastructure
- ☐ Achieve National goals in energy, security, and human welfare with advanced materials
- ☐ Equip the next generation materials workforce

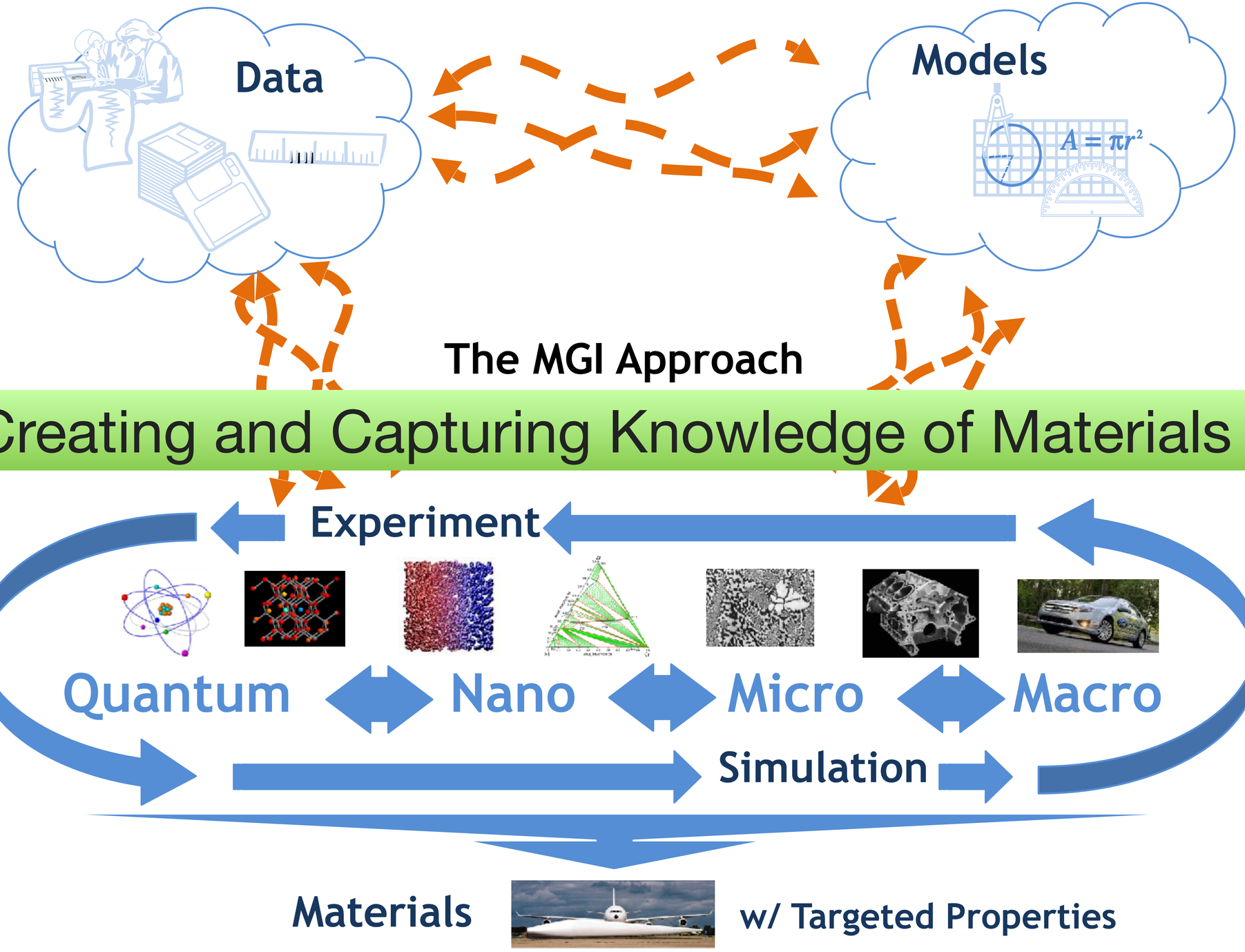


Span the Continuum



The G in MGI

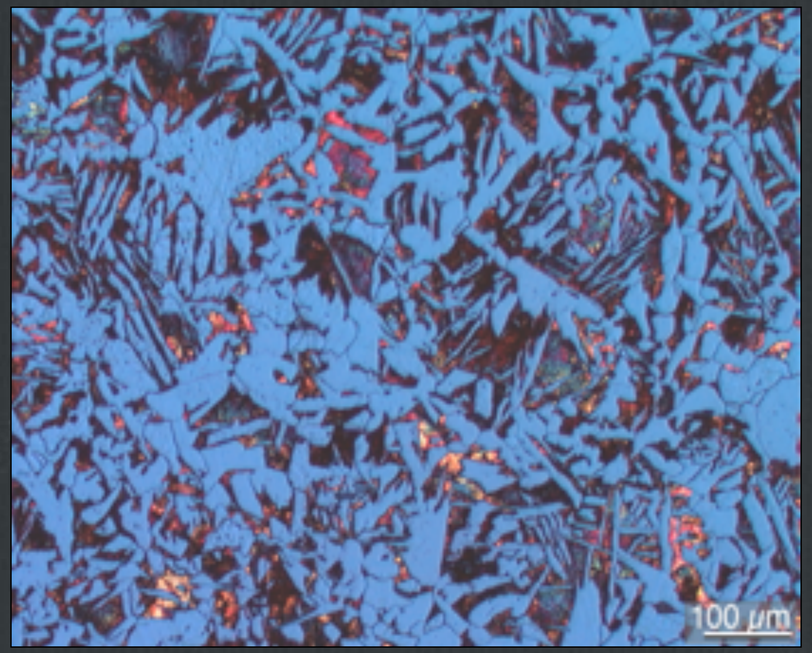
Metaphors


Data
Models
$A = \pi r^2$
The MGI Approach
Creating and Capturing Knowledge of Materials
Experiment
Quantum
Nano
Micro
Macro
Simulation
Materials
w/ Targeted Properties

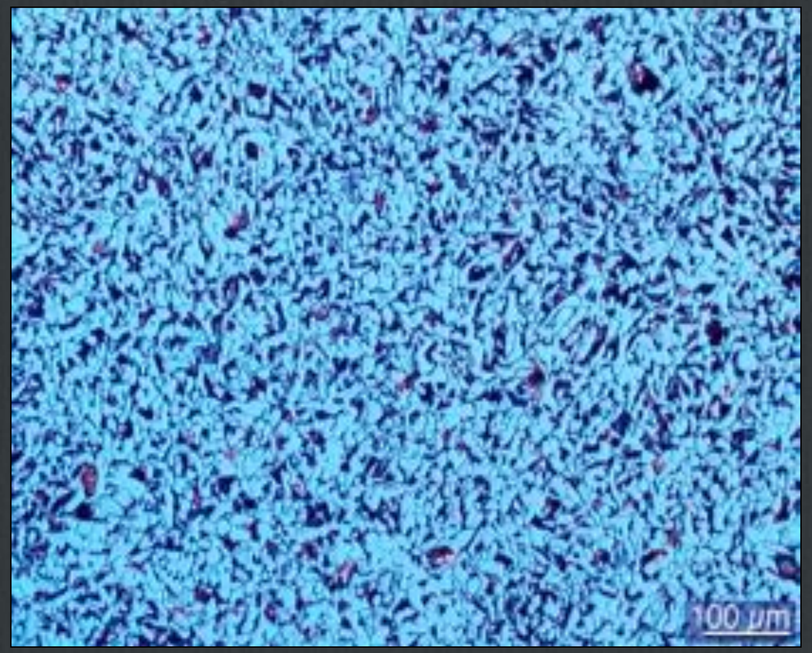
M in MGI

Materials Are Complicated Systems Modeling is a Challenge

Alloy cooled from 300 °C



Alloy cooled from 800 °C



The Decade of MGI?

These ideas are not new

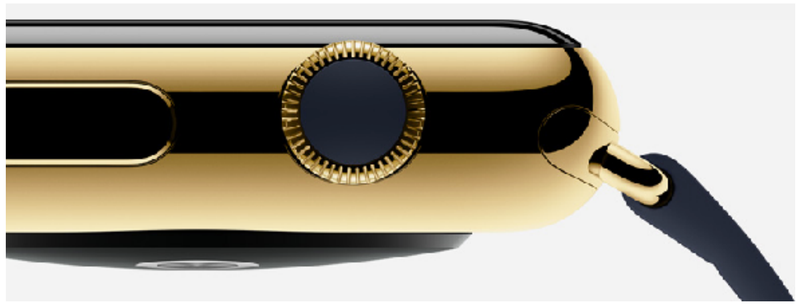
Apple watch

-Announced September 2014

Baseline: 316L Stainless Steel

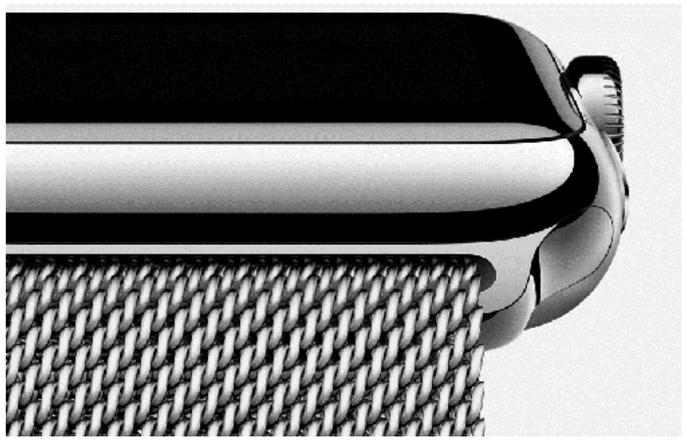
-Cold-forged to 40% harder
-Special purity mirror finish

High Strength 18K Gold



-2X harder

Milanese Loop Alloy



-Custom Magnetic Stainless Steel

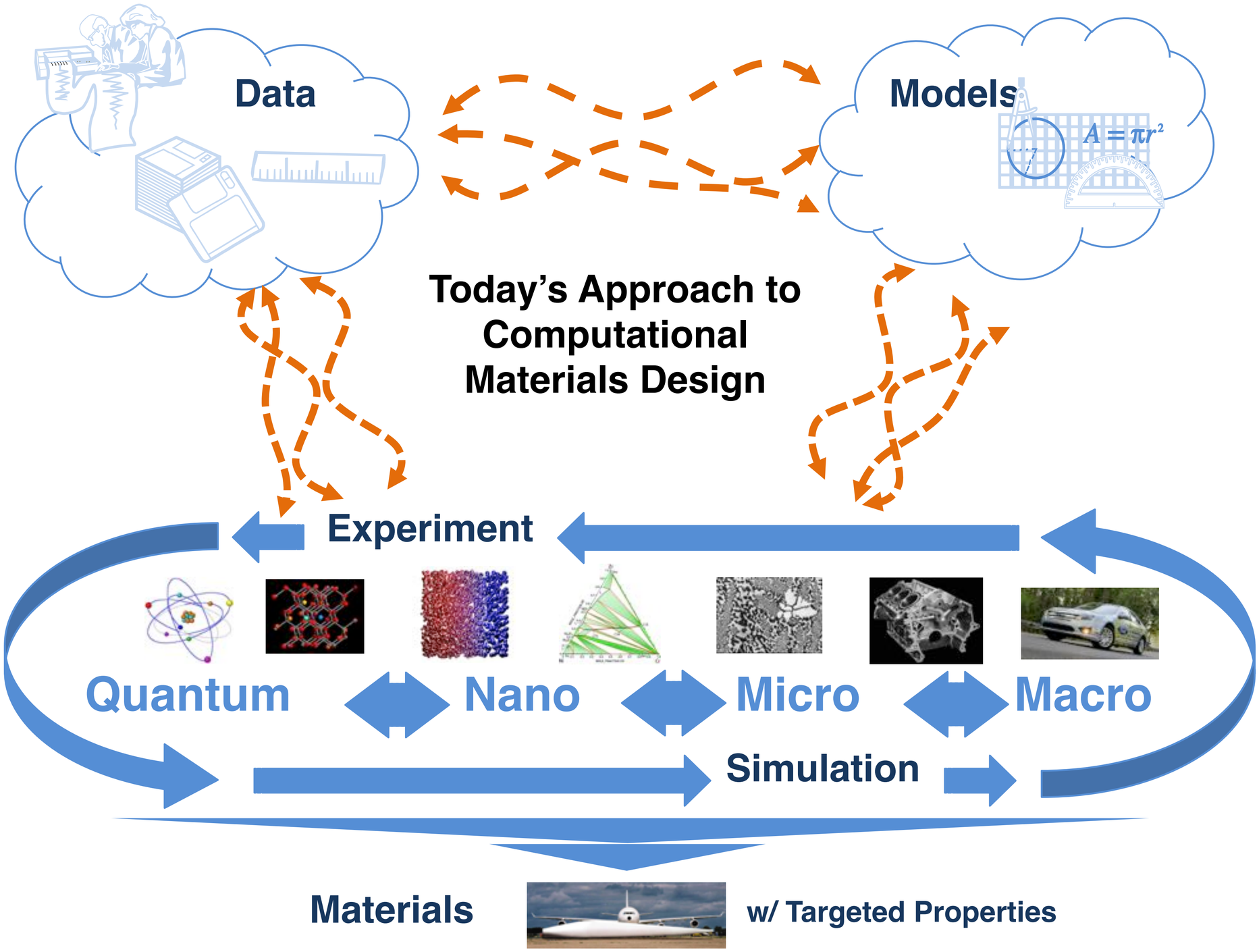
Anodizable 7000 Aluminum

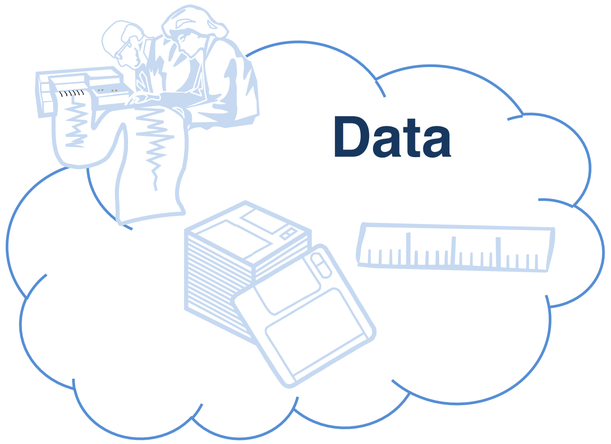
-60% stronger Al
-30% lighter than 316L

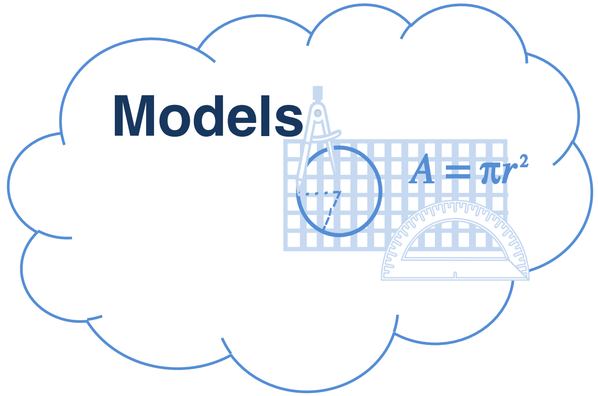
MGI IN SUM

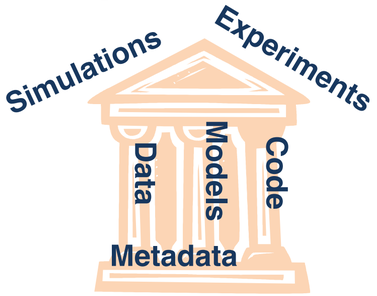
- The MGI Is about improving our ability to design and deploy new materials (faster)
- Need better (or just any) data and models
- The MGI is essentially a direct consequence of our improvements in computational power and associated models, coupled to the disruptive consequences of the Internet.
- There are a limited number of ways to get the “knowledge” that is the fuel for the MGI
  - High Throughput Computations, w/ published data and software
  - High Throughput Experiments
  - Get it from everywhere (Mine the literature, mine published data, if only it were published!)—- Change Publication and feed the models

NIST Role ~2013


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

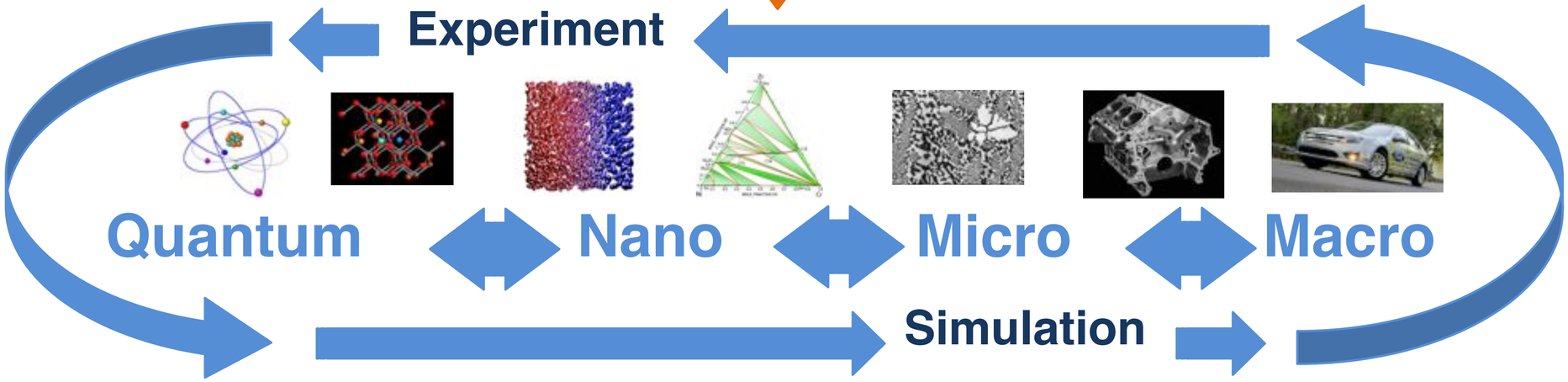


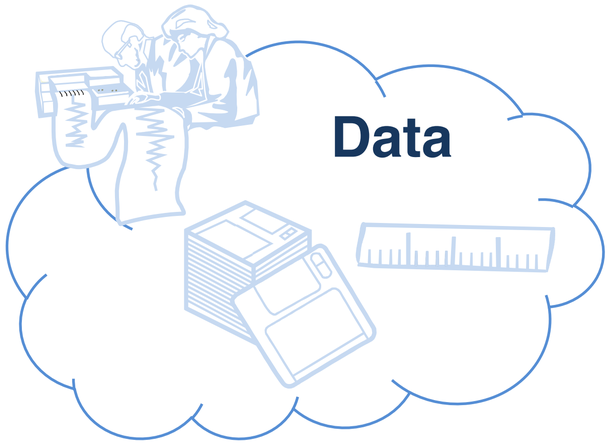




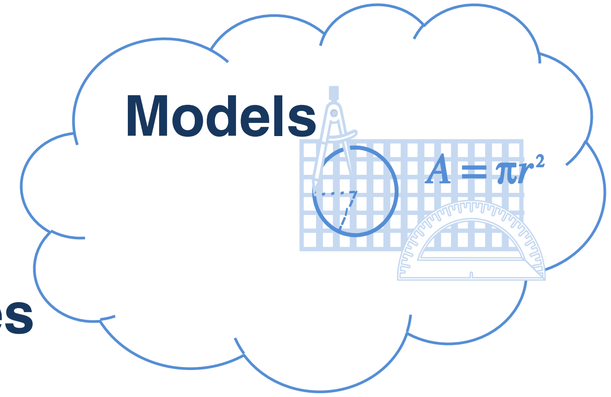
MGI

How do we do it?



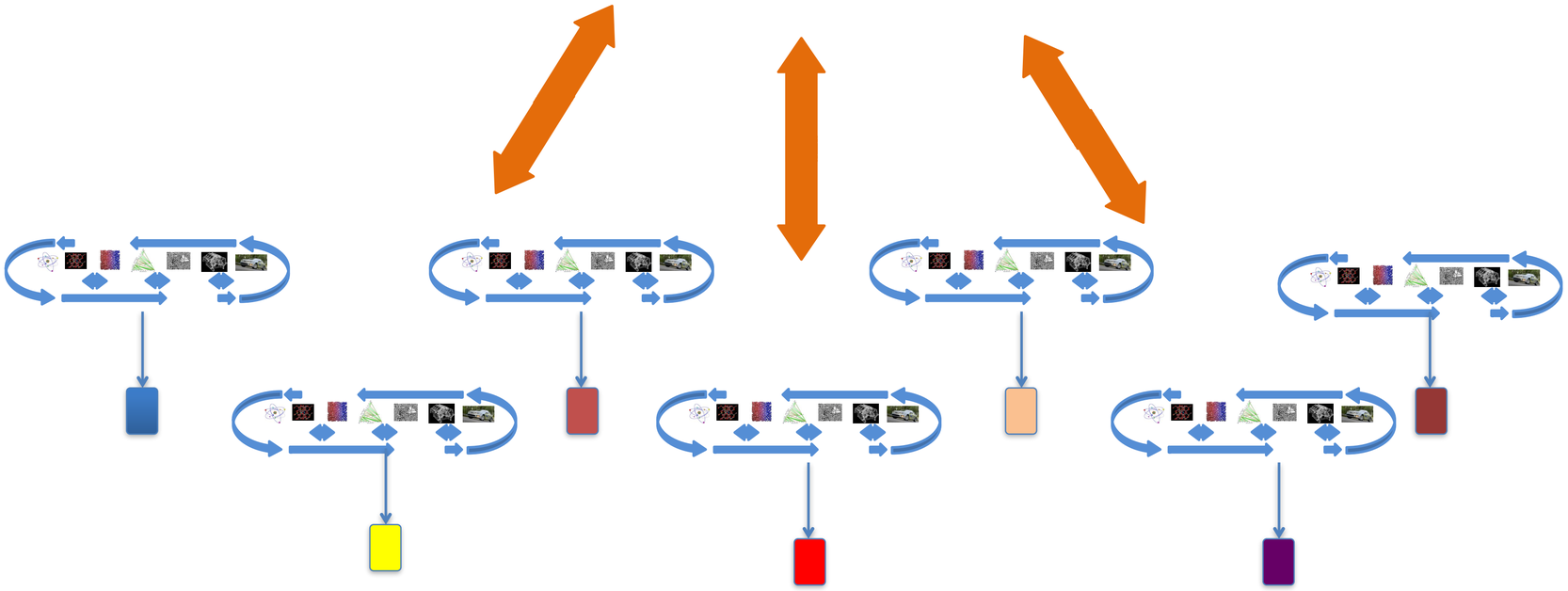


Community-based
Curated Repositories

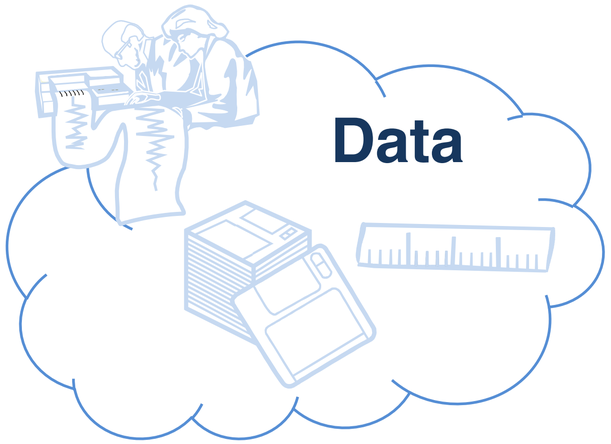


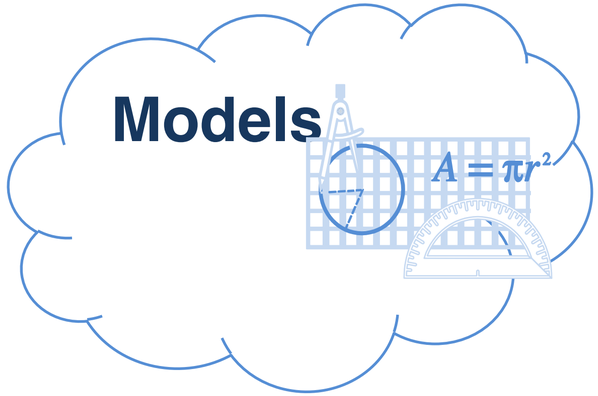
MGI

Ecosystems



Materials w/ Targeted Properties





NIST

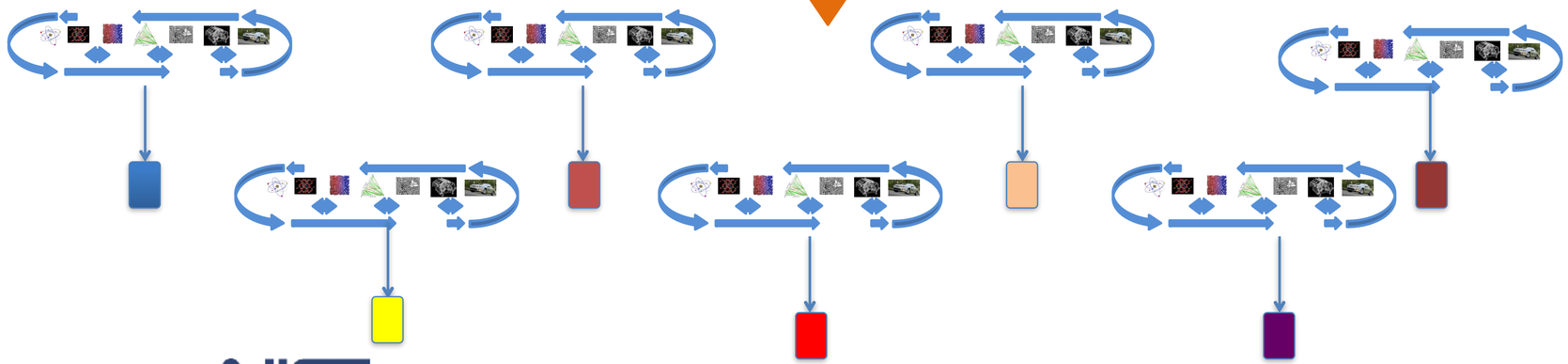
*Enable & Enhance **Exchange***

(repositories, disciplines, industries; standards)

NIST

*Assess & Improve **Quality***

(Data & Models)



NIST

*New **Methods and Metrologies***

(data driven analysis and models)

Materials w/ Targeted Properties

Building a Materials Data Infrastructure

Chuck Ward

TMS A Study Organized by The Minerals, Metals & Materials Society

Mixed Bag

- [ ] Recommendation 1: Strengthen the MDI core in repository, registry, and tool development
- [ ] Recommendation 2: Sustain and grow MDI-dedicated funding programs
- [ ] Recommendation 3: Create, execute, and monitor incentive mechanisms
- [ ] Recommendation 4: Develop demonstration projects and cross-disciplinary community efforts that enhance and accelerate adoption of the MDI
- [ ] ….

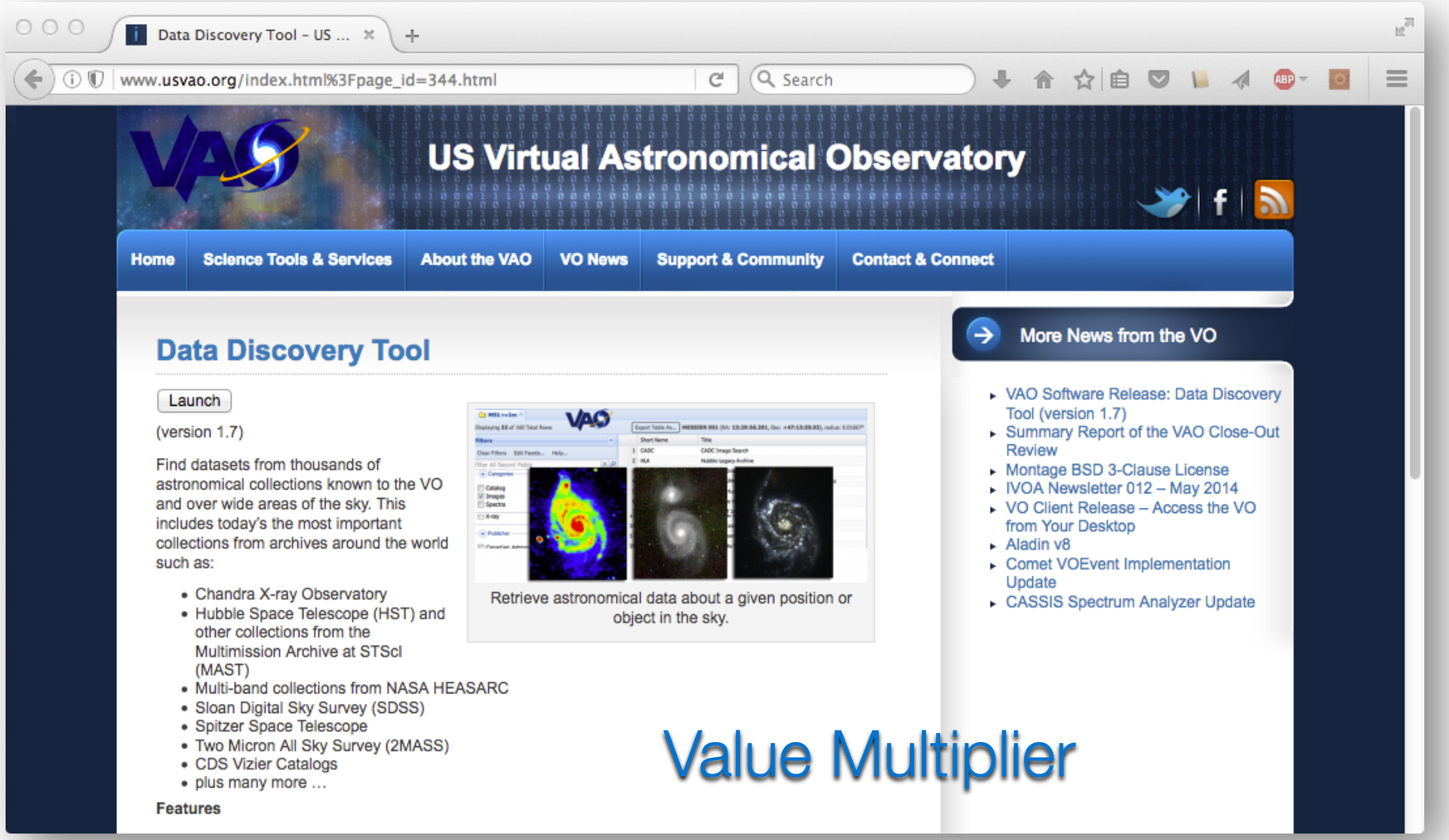
What's stopping us from living the dream? Moving to MGI 2.0

- Incentive failures
  - Ignorance
  - Technical
  - Professional (some thoughts on industrial aspects)

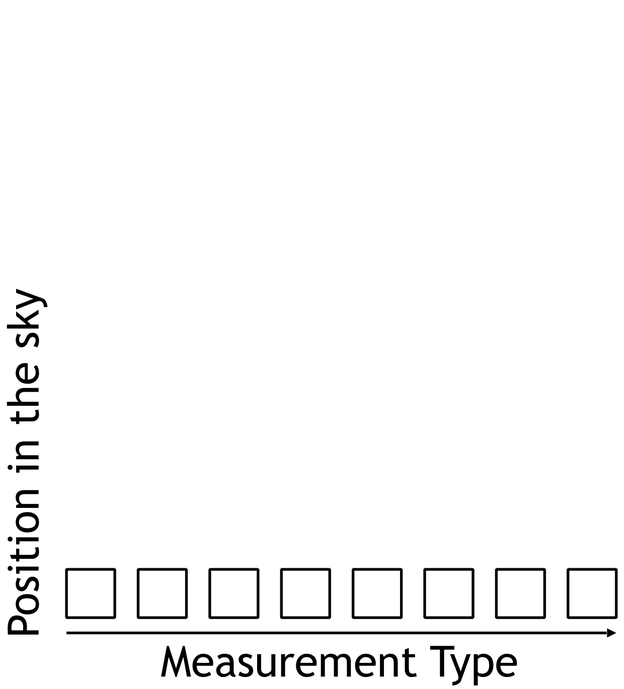
Ignorance

Technical/What Kind of Tools are Needed?

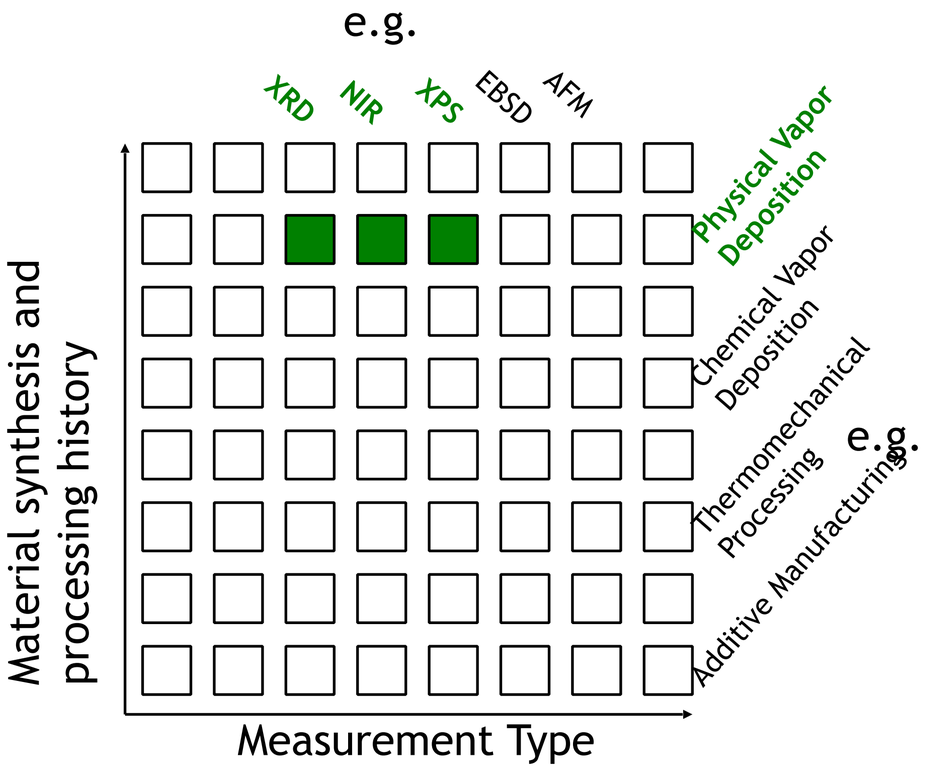
Motivated by the Astronomy Community



Astronomy vs. Material Measurement



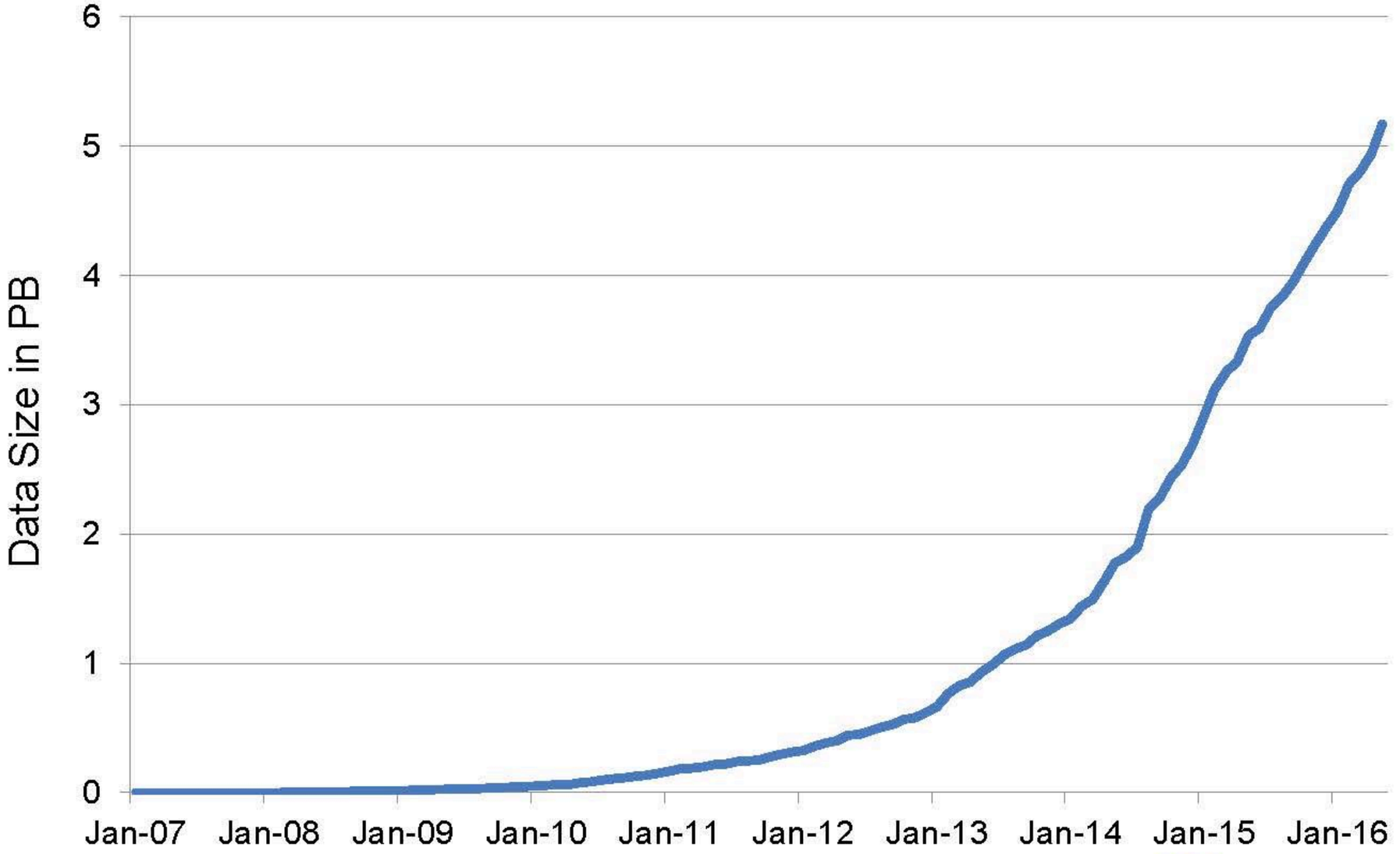
Singular data models



Modular data models

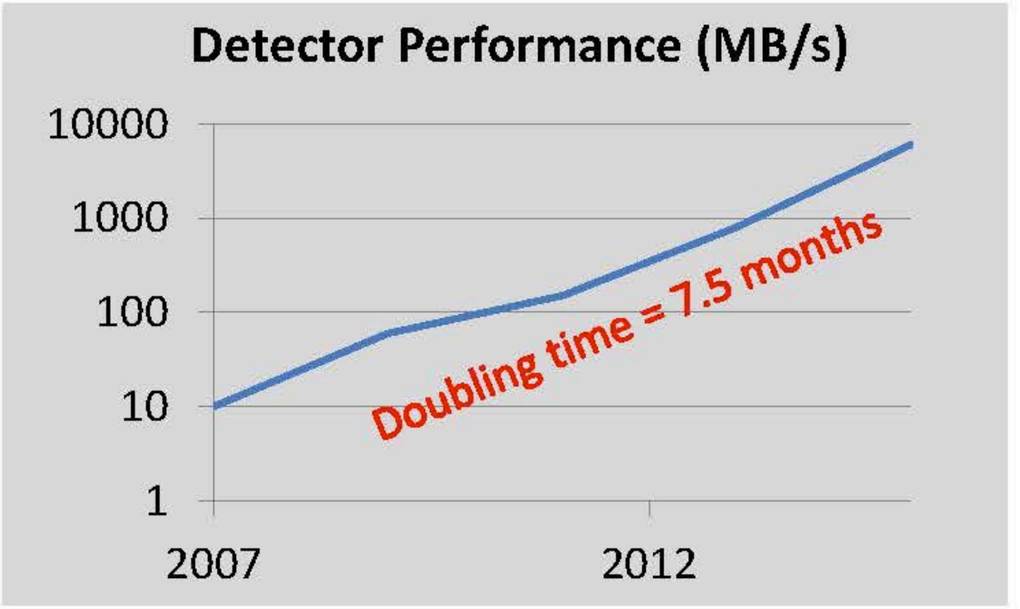
WHAT'S OUR TELESCOPE?

Cumulative Amount of Data Generated By Diamond



diamond

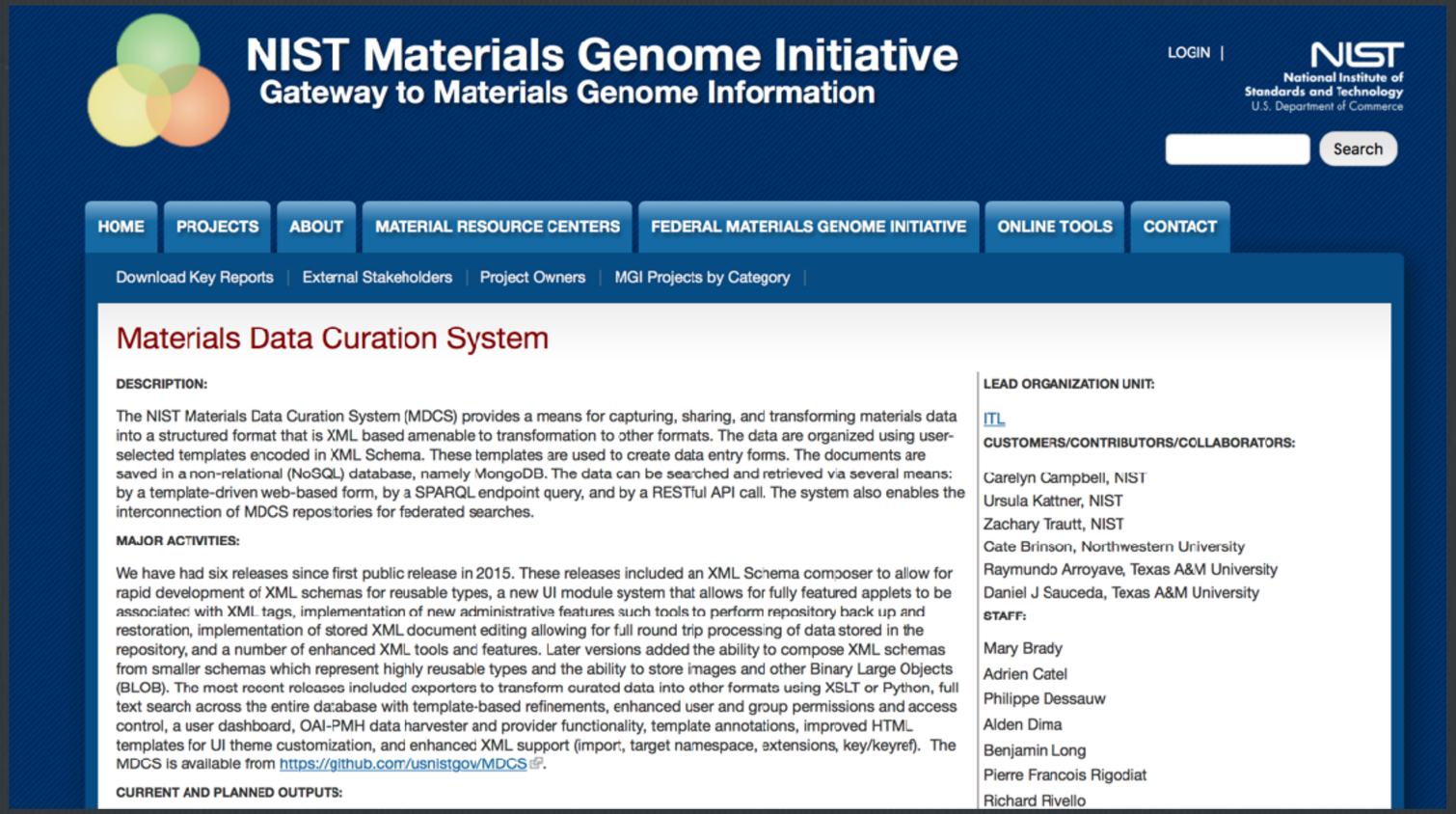
Data Rates



- 2007 No detector faster than ~10 MB/sec
- 2009 Pilatus 6M system 60 MB/s
- 2011 25Hz Pilatus 6M 150 MB/s
- 2013 100Hz Pilatus 6M 600 MB/sec
- 2013 ~10 beamlines with 10 GbE detectors (mainly Pilatus and PCO Edge)
- 2016 Percival detector 6GB/sec

diamond

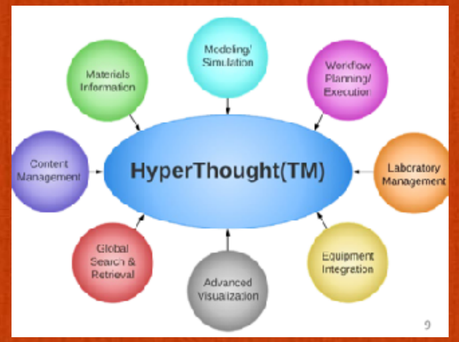
Better living through curation and schemas

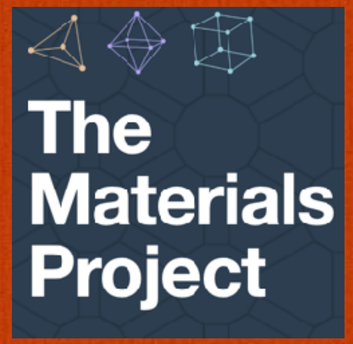


https://github.com/usnistgov/MDCS

Infrastructures

Materials Innovation Platforms

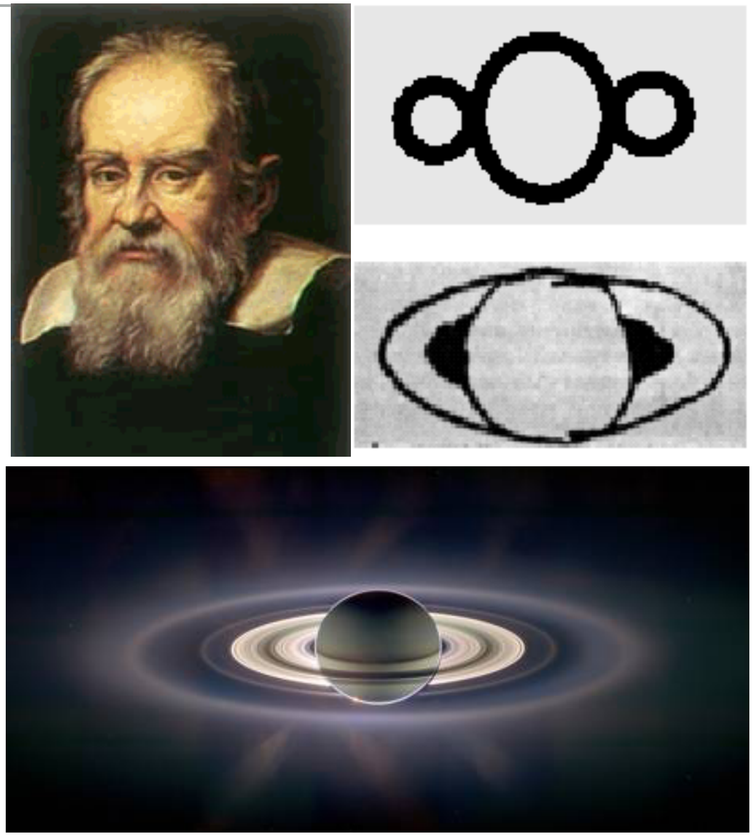




Professional

THE MOONS OF SATURN

Galileo no doubt planned to publish this new discovery in his next book, but in the meantime, how could he preserve his priority and prevent others from claiming the discovery as their own? His solution was to circulate an anagram, s m a i s m r m i l m e p o e t a l e u m i b u n e n u g t t a u i r a s. Others would know that he had discovered something and when he had discovered it, but they would not known what the discovery was. The number of letters in the anagram, 37, was too small to allow him later to fudge and change the solution to describe a discovery made by someone else in the meantime. Before the days of scientific papers (invented in the 1660s) this was an effective (if not always foolproof) method of claiming priority.



We need to both change the rules and give people tools to make that easier

http://galileo.rice.edu/sci/observations/saturn.html

Addressing the real issue

Until people can do science/engineering on these platforms it won't happen

- ☐ Combine data sets
- ☐ Build on prior work
- ☐ Discover inter-relationships
- ☐ Use AI/ML

Fame and Fortune

We need to create a National Materials Data Network

We need to have an ongoing conversation

We need to create a broad academic/ industrial/government community to drive the development of needed infrastructures

Funding

Must come from the community

MaRDaC

Materials Research Data Council

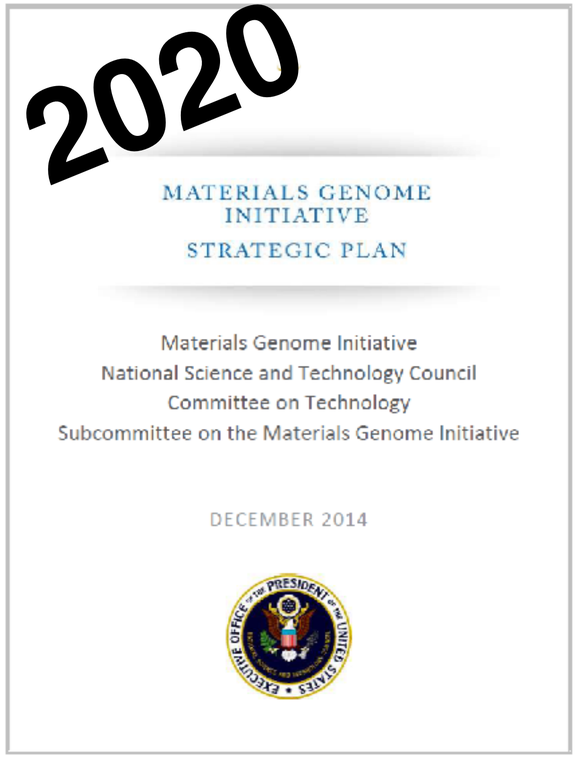
MaRDaC

Mission

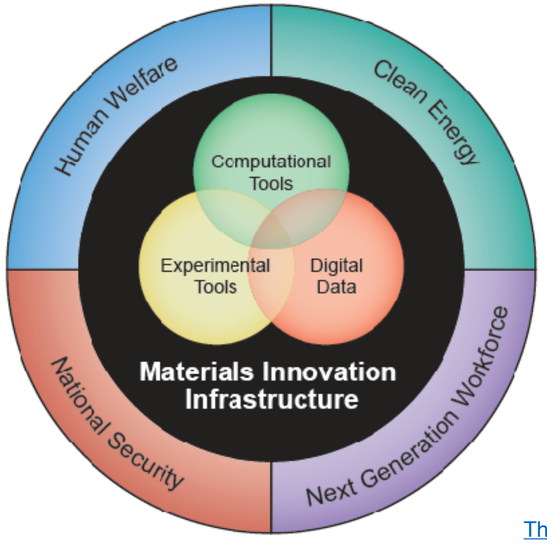
The Materials Research Data Alliance (MaRDa) is a newly formed organization focused on realizing the promise of open, accessible, and interoperable materials data. Each of these elements are aligned with the goals of the Materials Genome Initiative (MGI). MaRDa provides a platform that promotes the convergence of ideas, people, data, and tools to accelerate discovery, enable new insights into materials mechanisms, and lay the foundation for both human-centered and artificial intelligence-assisted approaches to materials design. MaRDa's governing council, MaRDaC, is an elected leadership body that promotes the interests of materials data researchers nationally and internationally and coordinates the efforts of MaRDa.

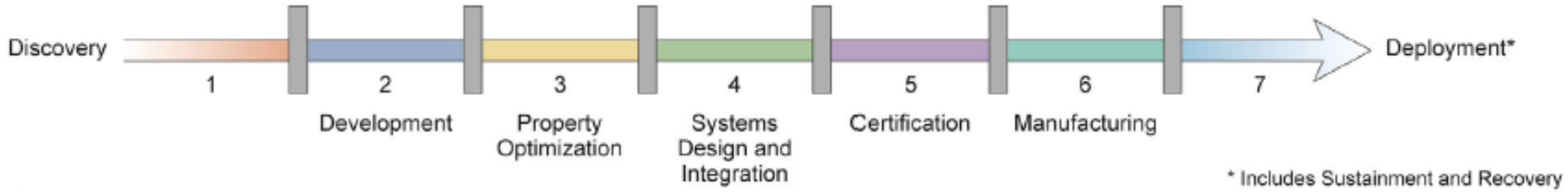
MGI 2.0 Preview

MGI 2.0



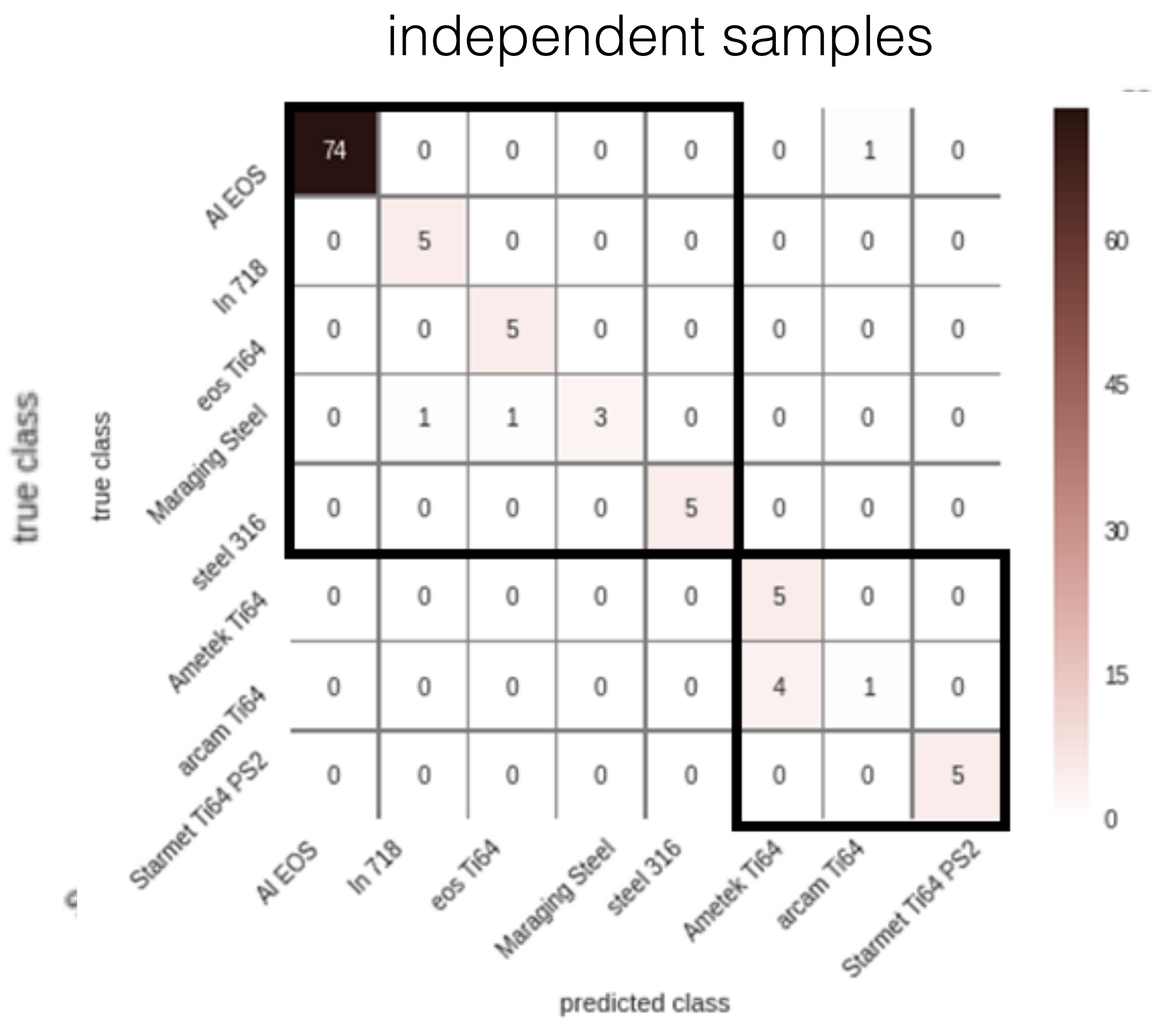
- Deploy the Materials Innovation Infrastructure (MII)
- Harness the Power of Materials Data
- Educate, train and connect the materials R&D workforce





AI/ML is a true enabler and driver for MGI

SEM powder classification

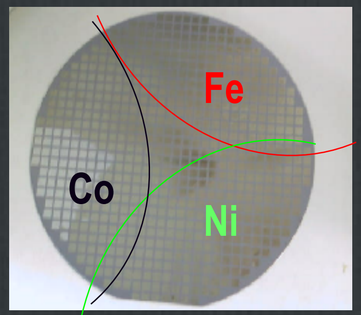


ML and Autonomous Materials Science

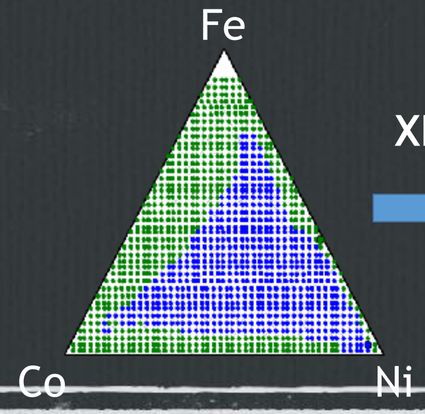
A. Gilad Kusne and Collaborators

Phase Mapping: High-Throughput Approach (APL Materials 2016)

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

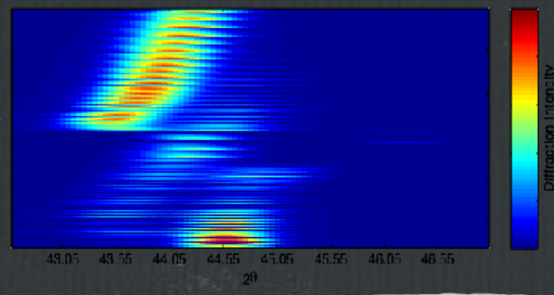


Combi Library for Ternary Spread



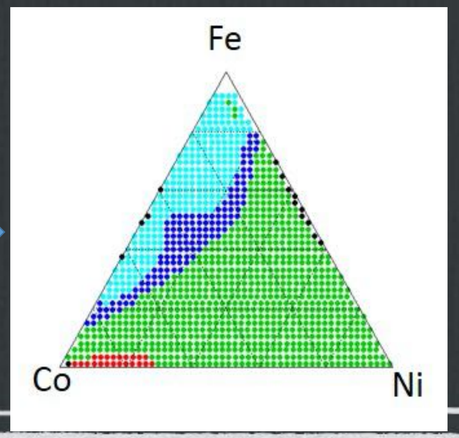
XRD

Diffraction Patterns



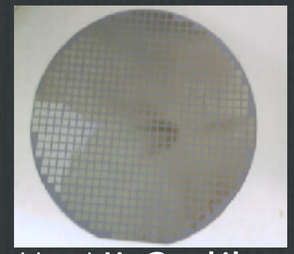
Cluster Analysis

Estimated Phase Map

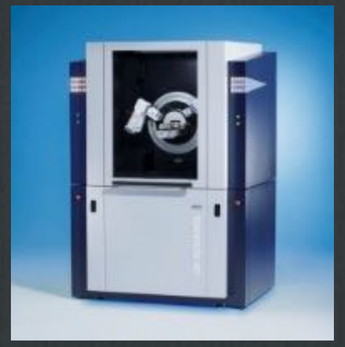


Phase Mapping: High-Throughput Approach

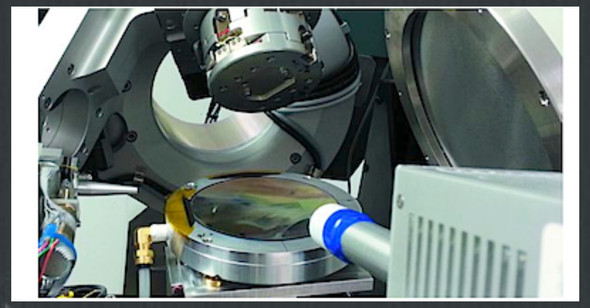
- Measurement is a time / resource sink
- For wafer of 500+ samples:
- In Lab: Takes weeks-months
- Synchrotron: Takes 5+ Hours (Every second counts)



Mn-Ni-Ge library
535 samples



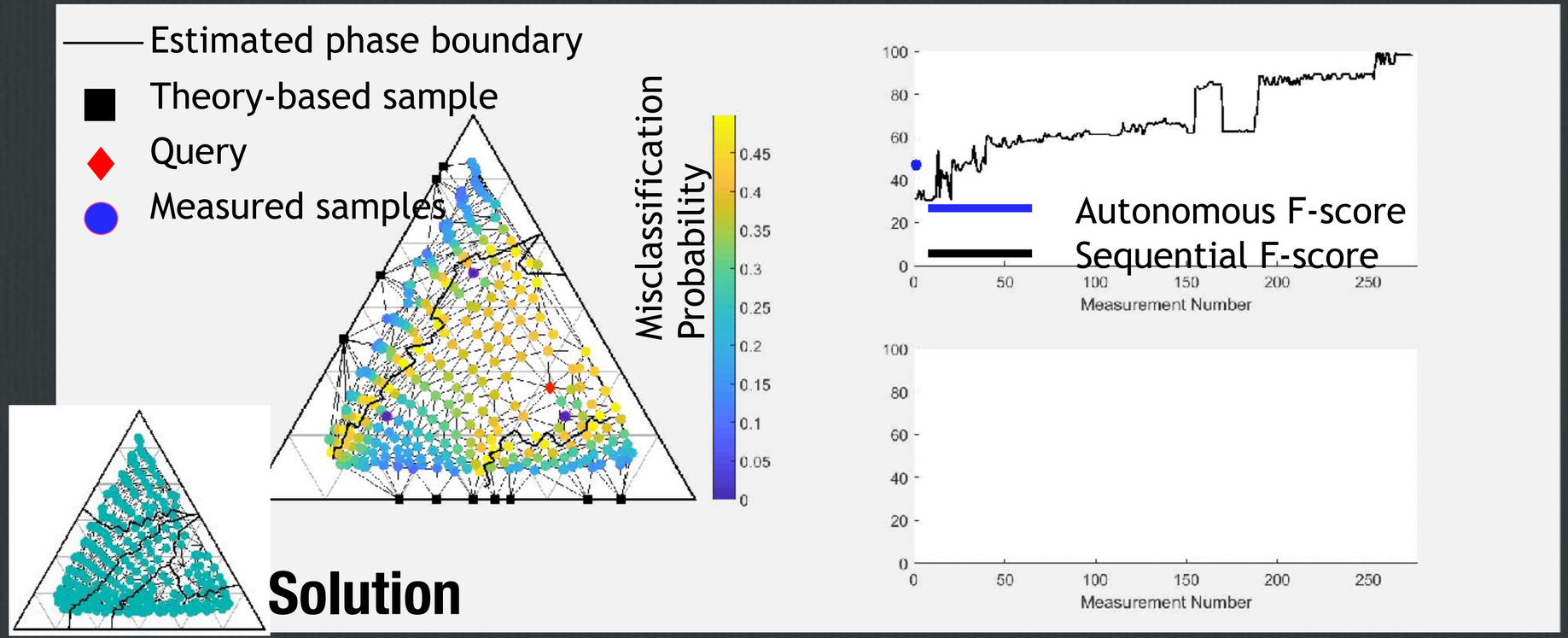
Bruker D8
30 Minutes per sample
2 weeks!



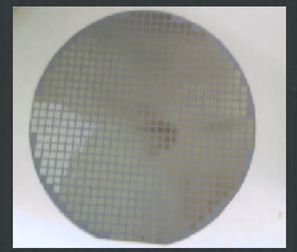
Stanford Synchrotron Radiation Lightsource
30 seconds per sample
4.5 hours

Autonomous Phase Mapping

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

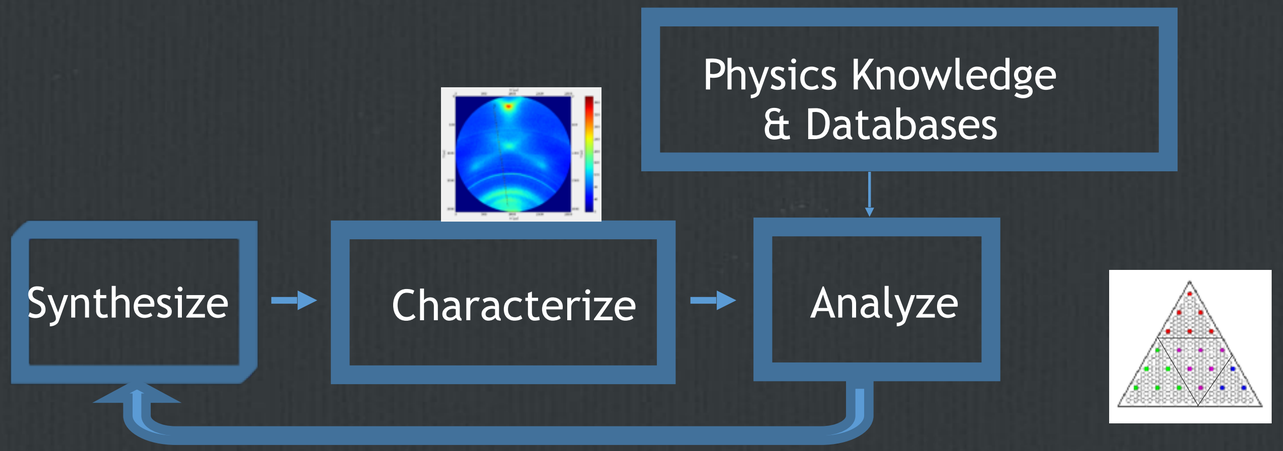


AI is controlling X-ray diffraction systems at SLAC & in the lab!

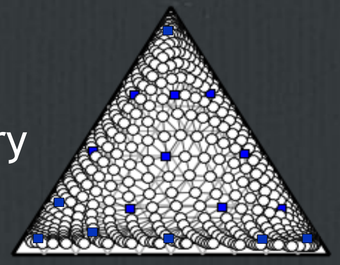


Automated Phase Mapping

- Now: Place AI in control of Synthesis.



Test case:
Combinatorial Library



Questions?