

# **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**

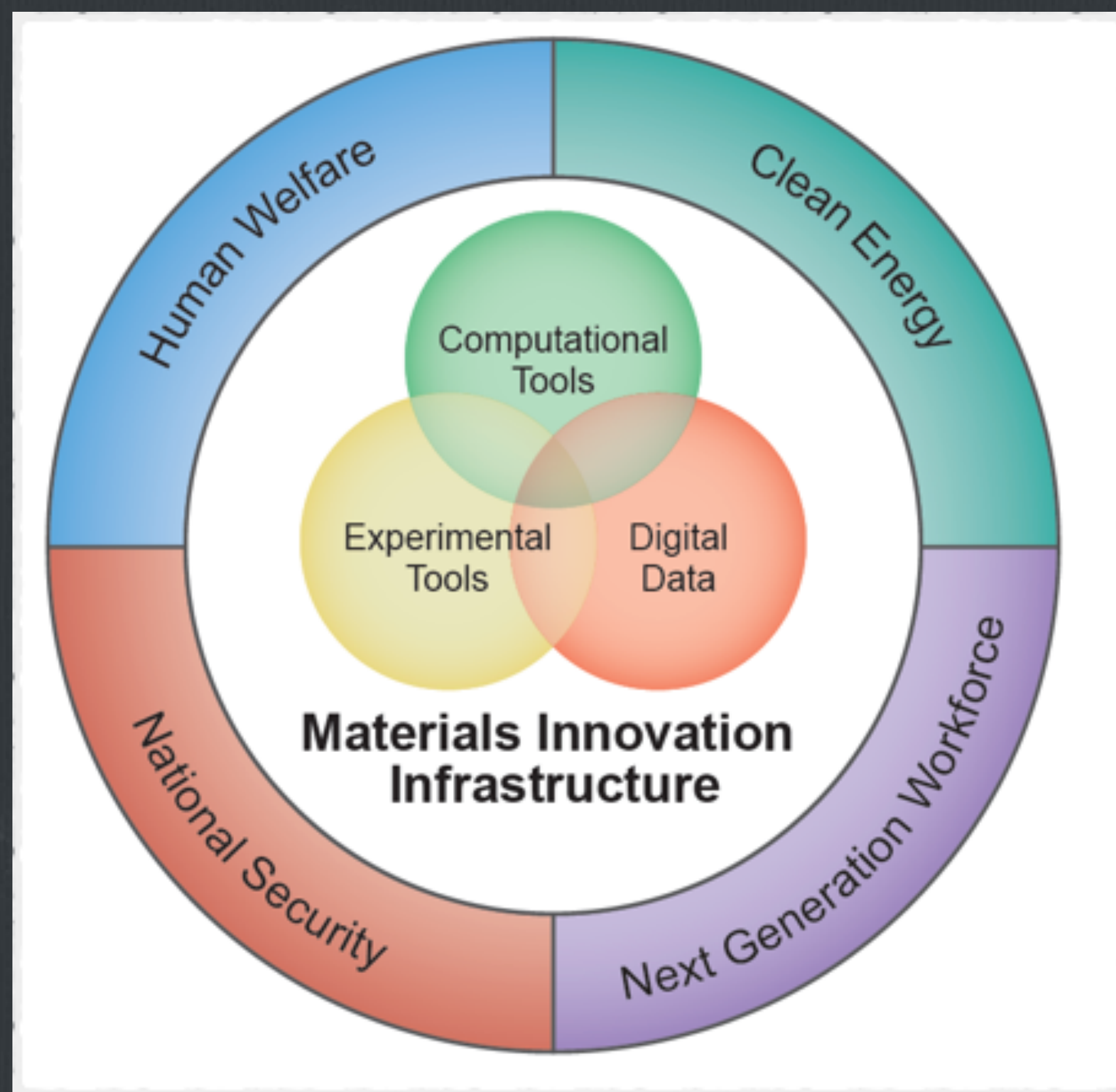
**DOE, NSF, NIST, DOD, NASA, FDA, NIH...**

**3 years into a new administration**



# To decrease time-to-market by 50% while

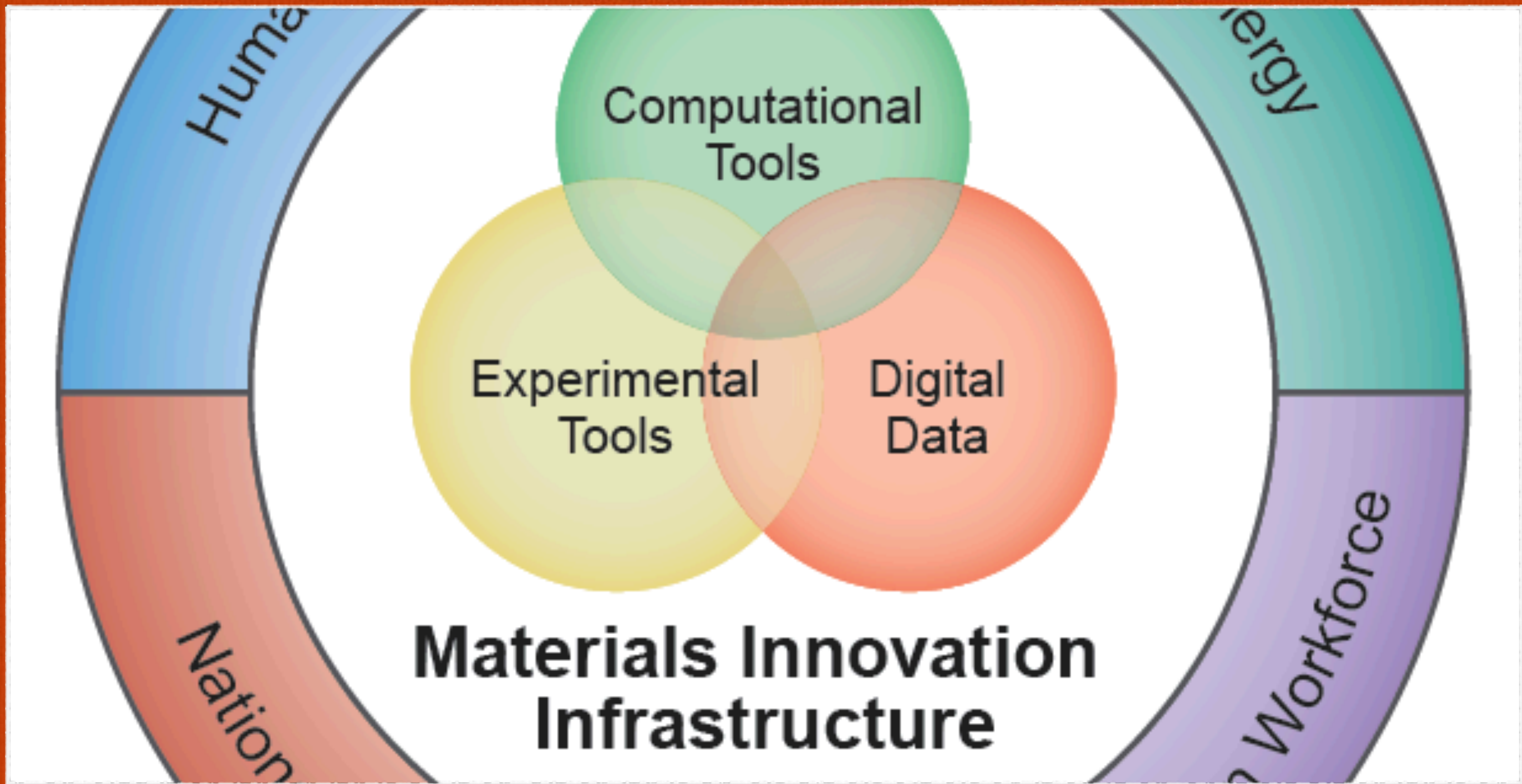
<\$\$



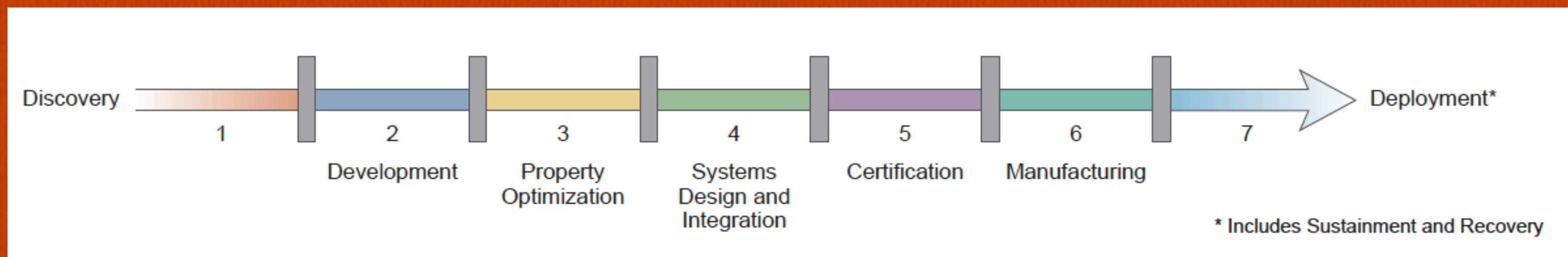
*Materials Genome Initiative for Global Competitiveness*

- 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**

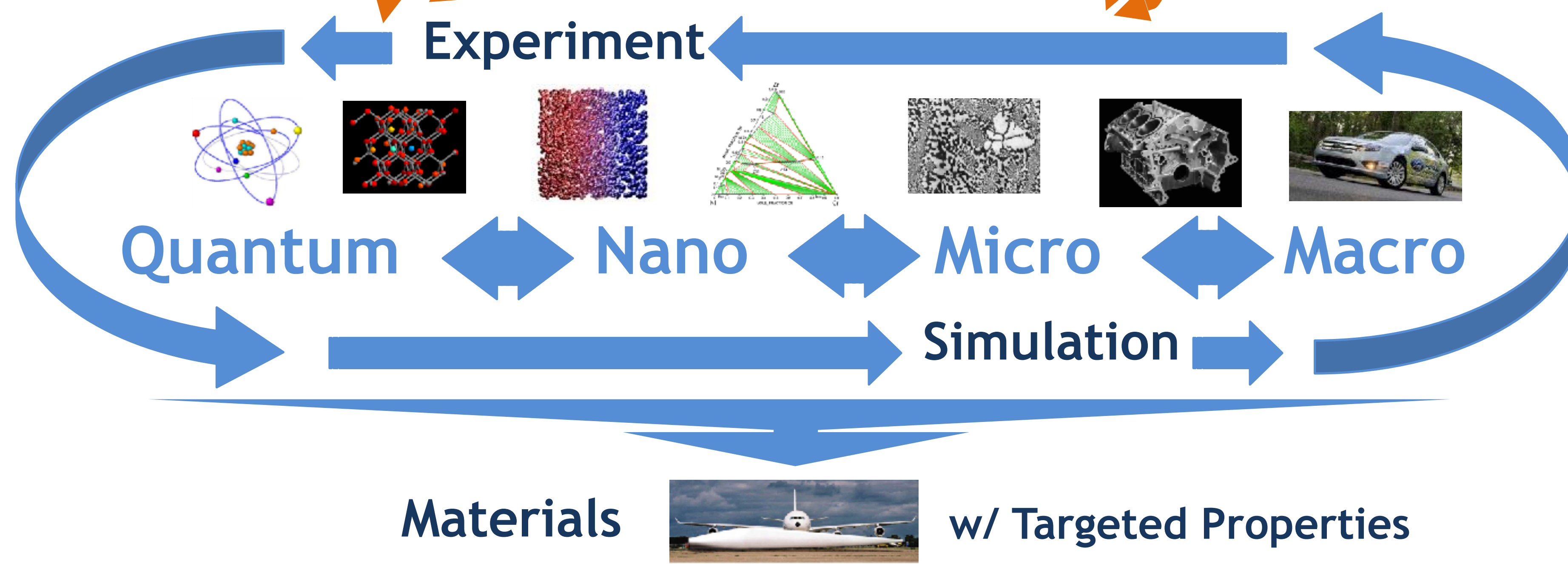
**Metaphors**





### The MGI Approach

# Creating and Capturing Knowledge of Materials



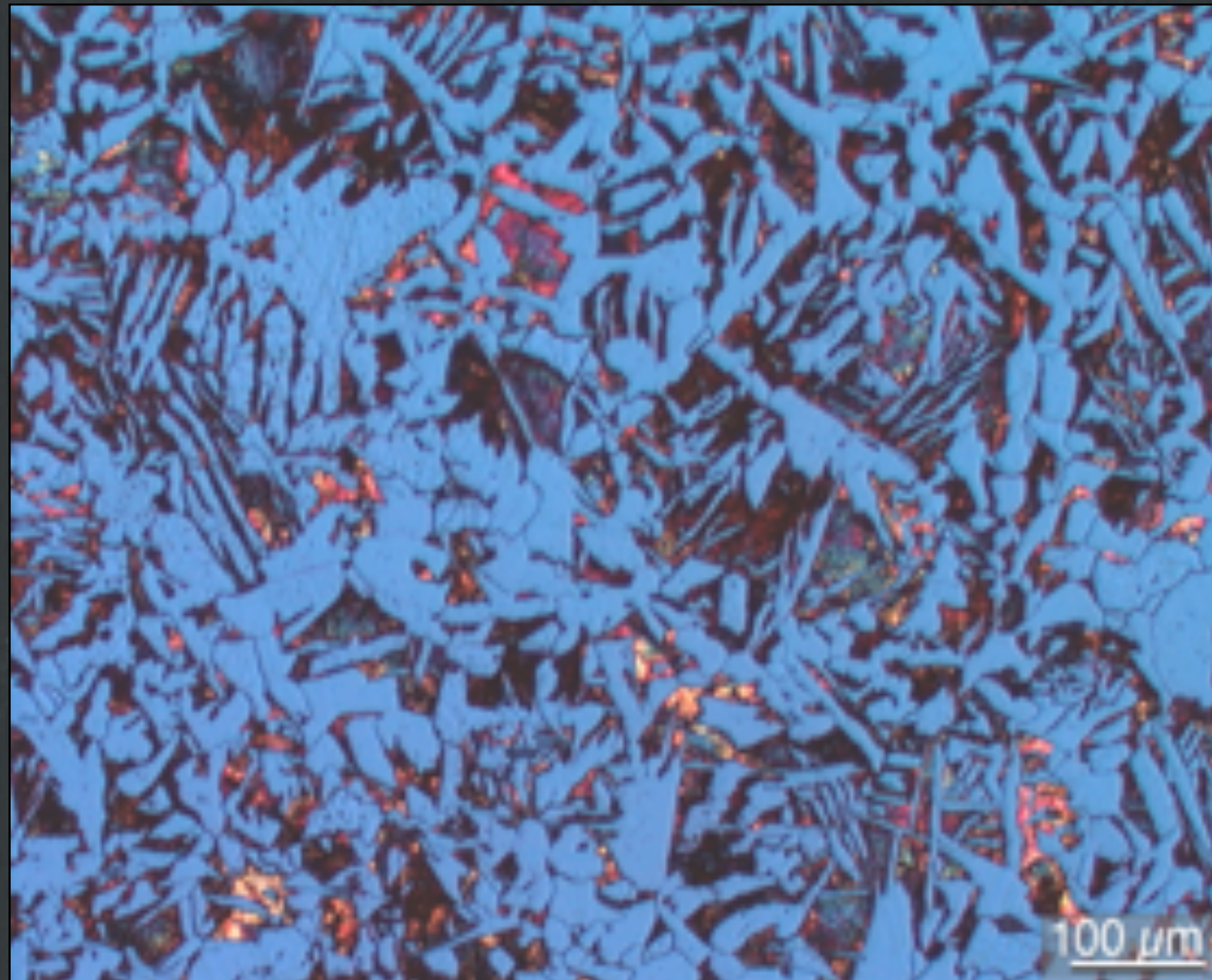


**M in MGI**

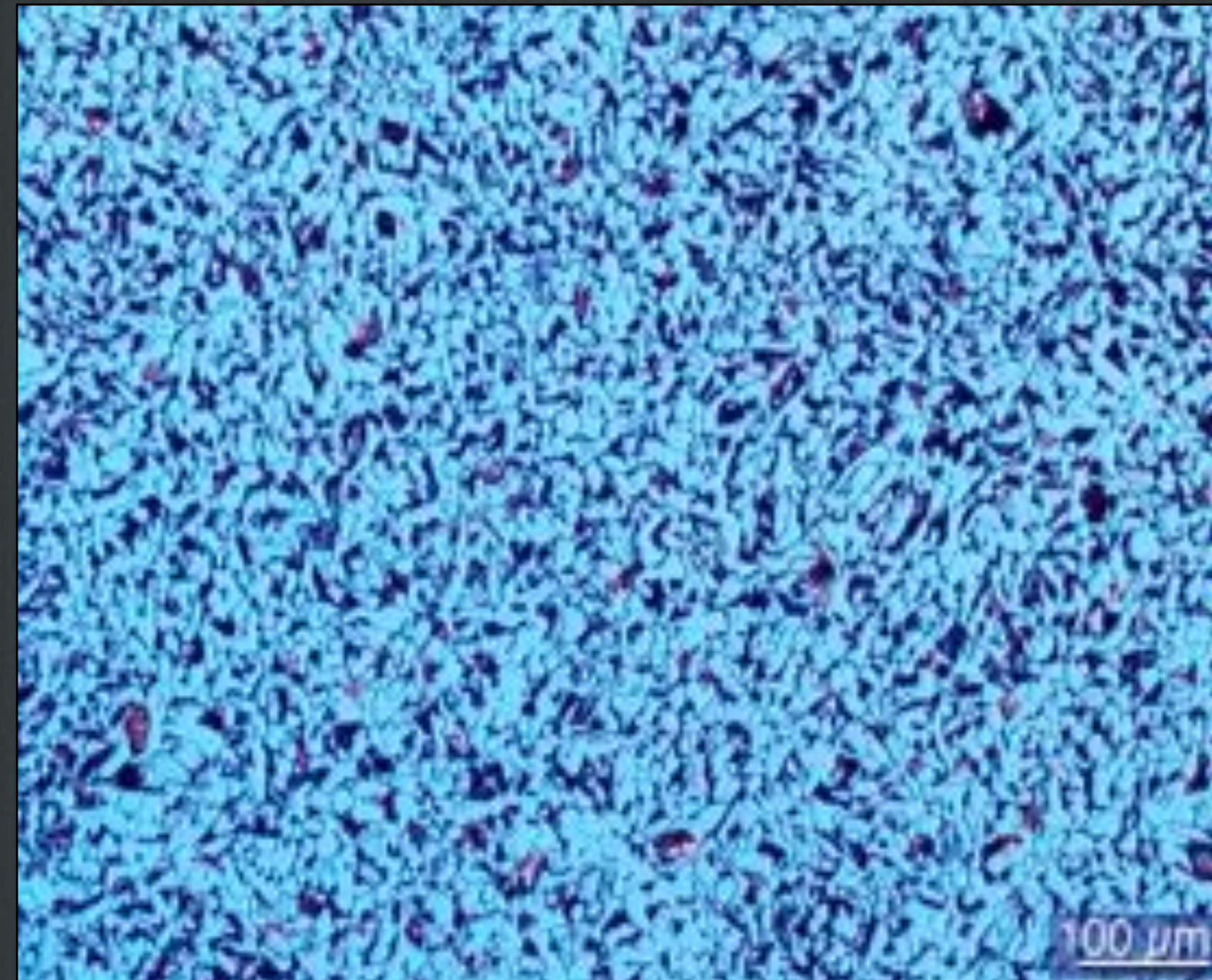


# Materials Are Complicated Systems Modeling is a Challenge

Alloy cooled from  
300 °C



Alloy cooled from  
800 °C





# The Decade of MGI?

These ideas are not new





# Apple watch

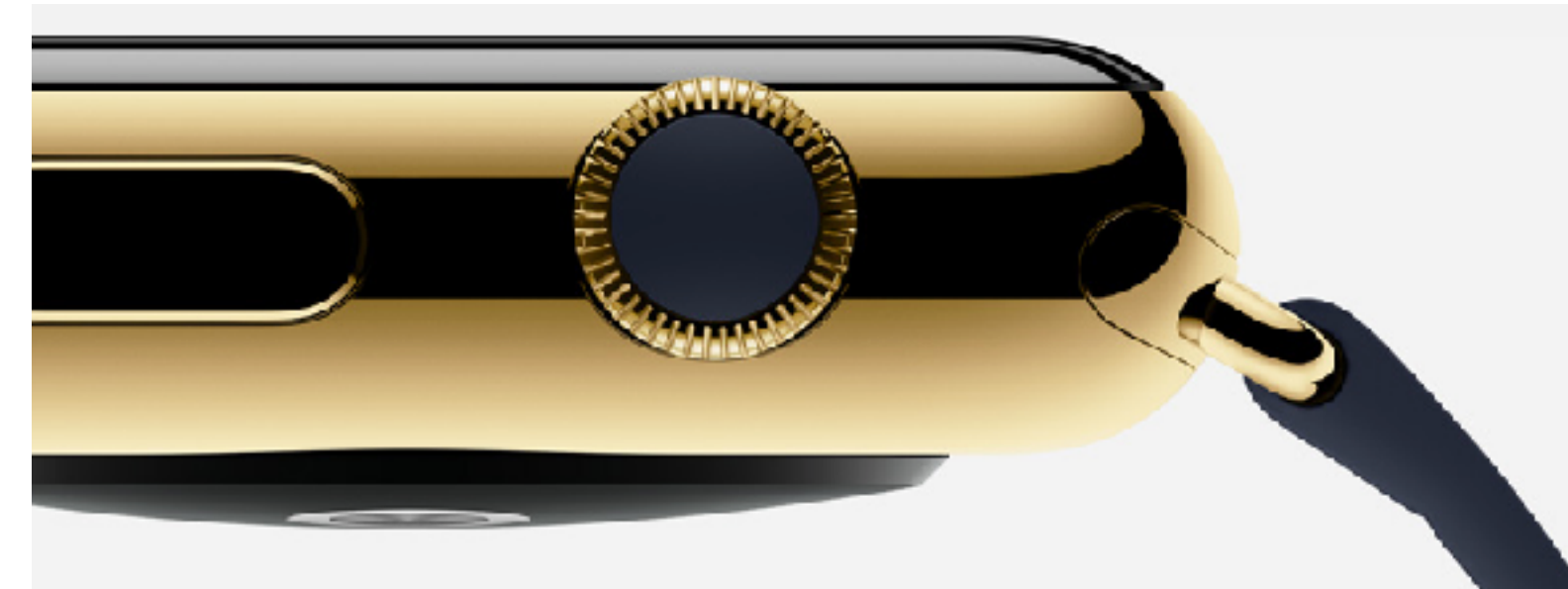
-Announced September 2014

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

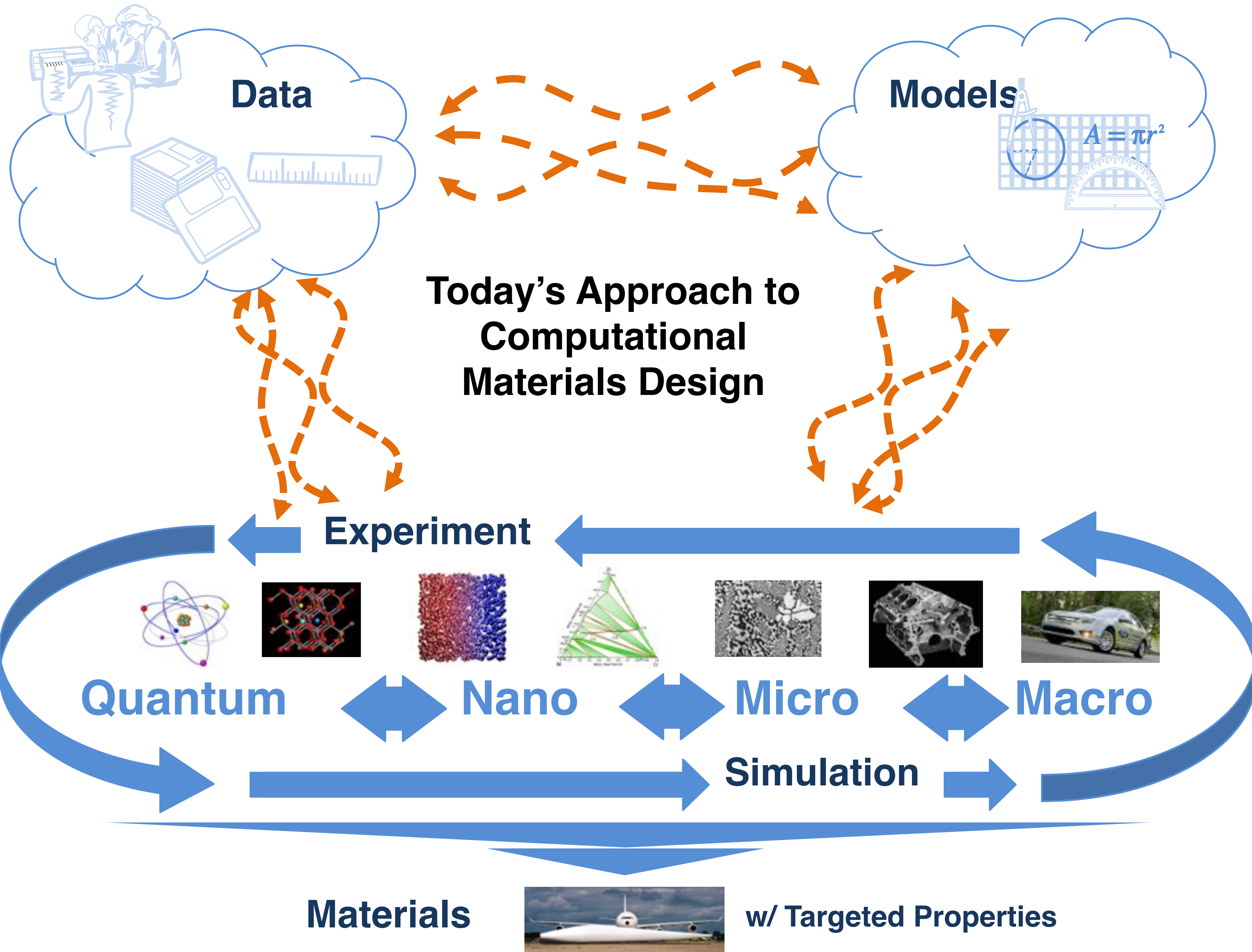
---

- 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**

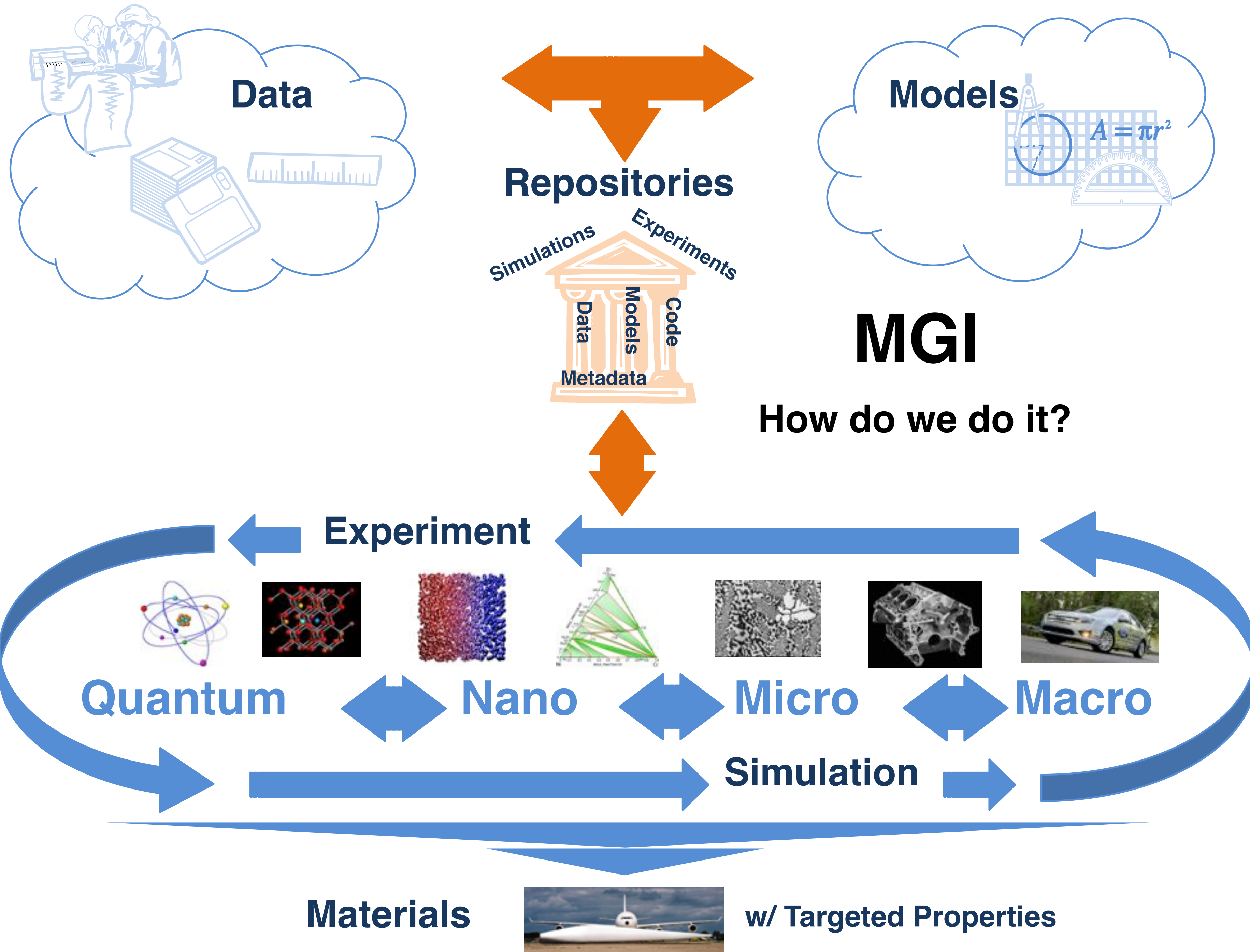


# NIST Role ~2013

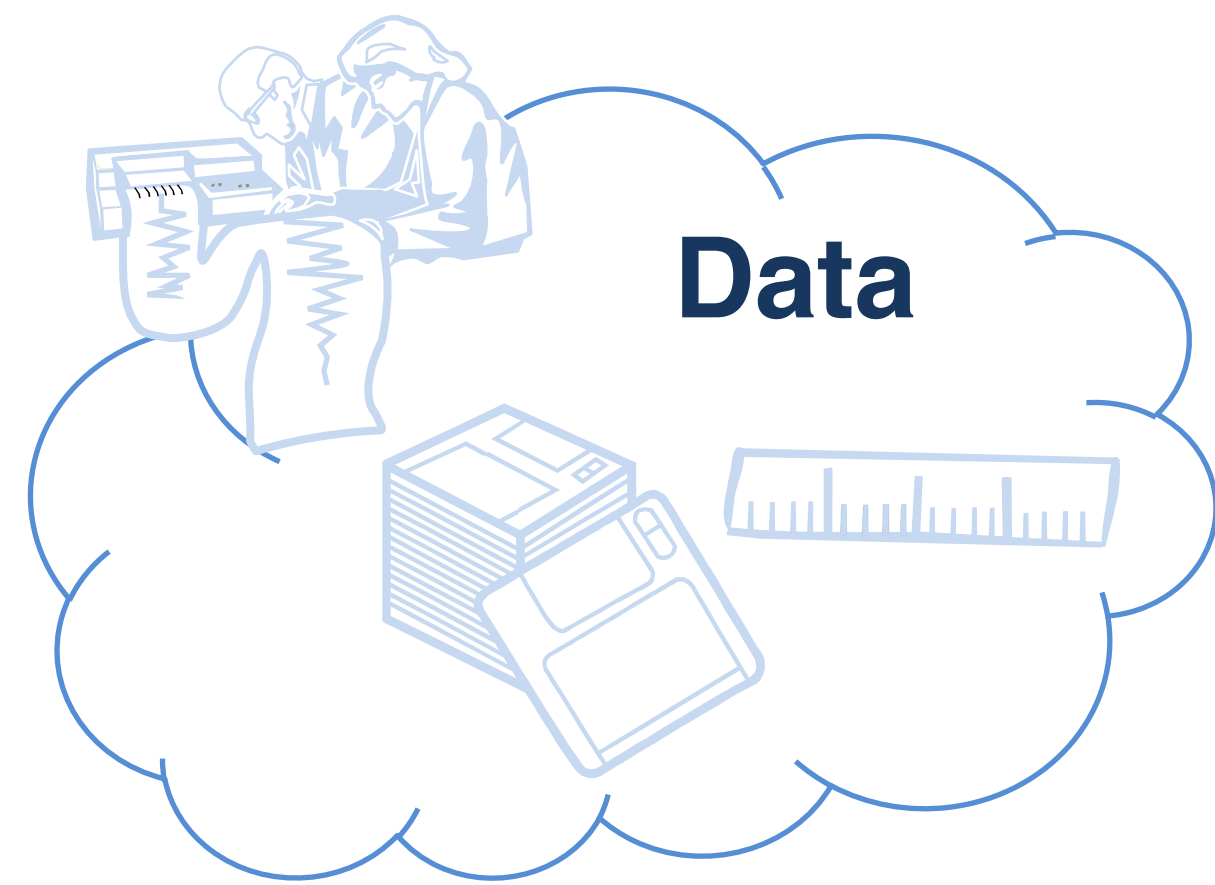




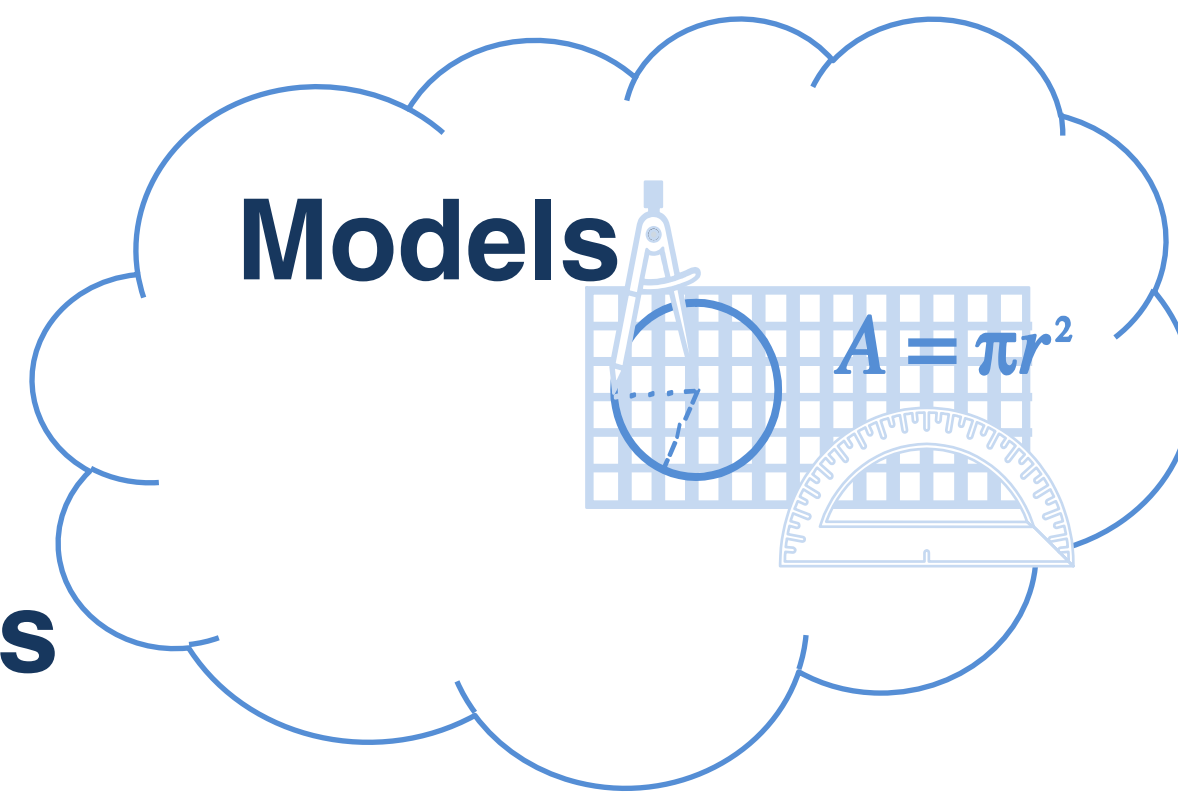




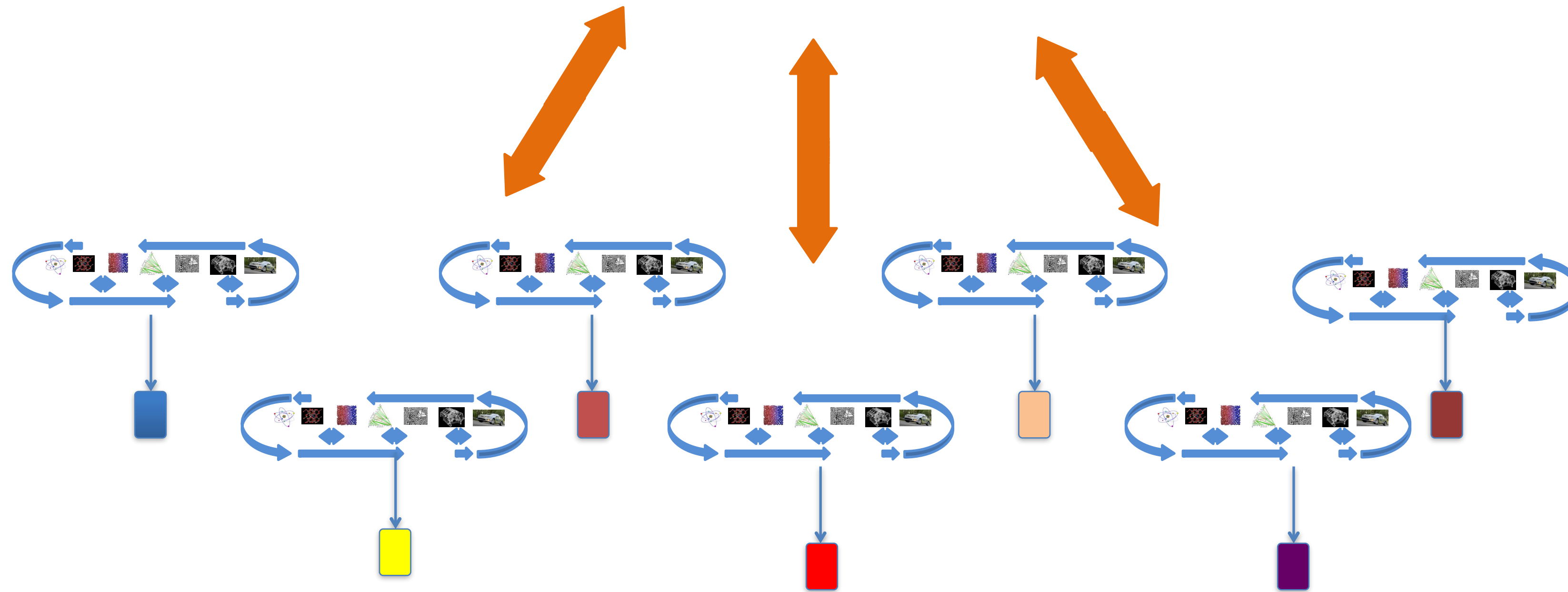




**Community-based  
Curated Repositories**

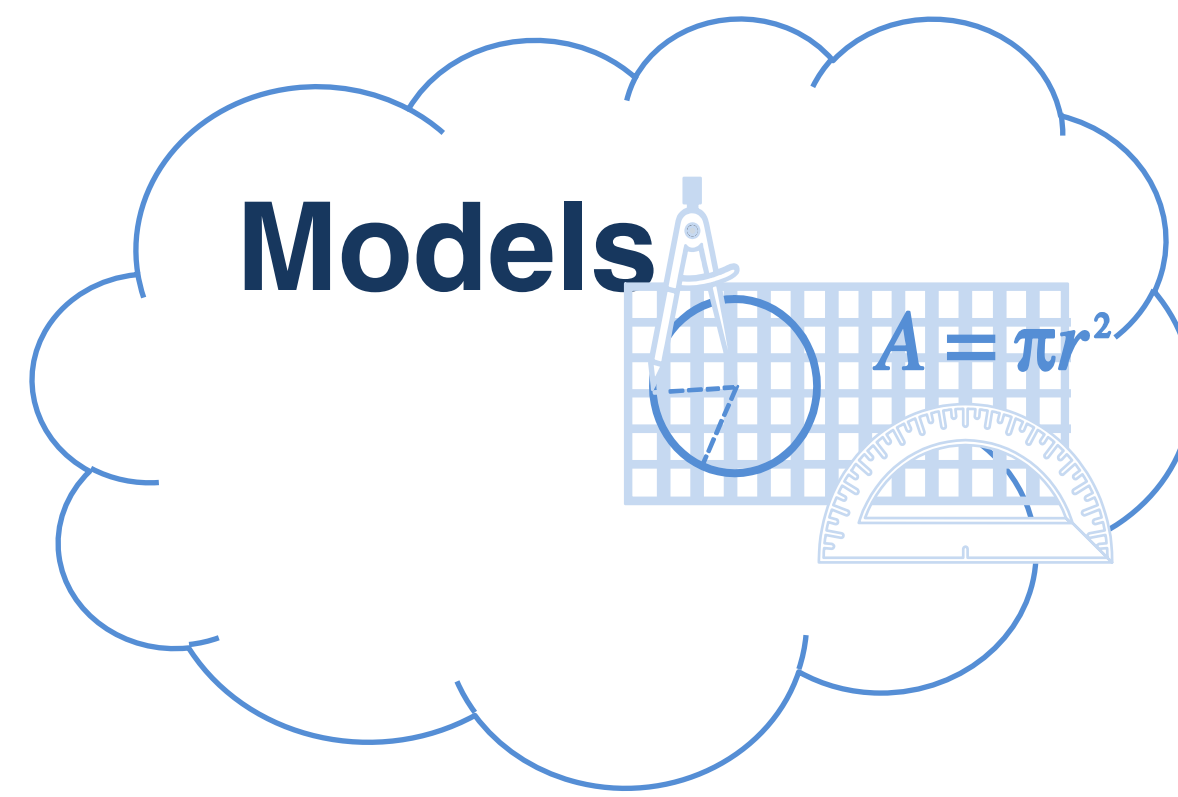
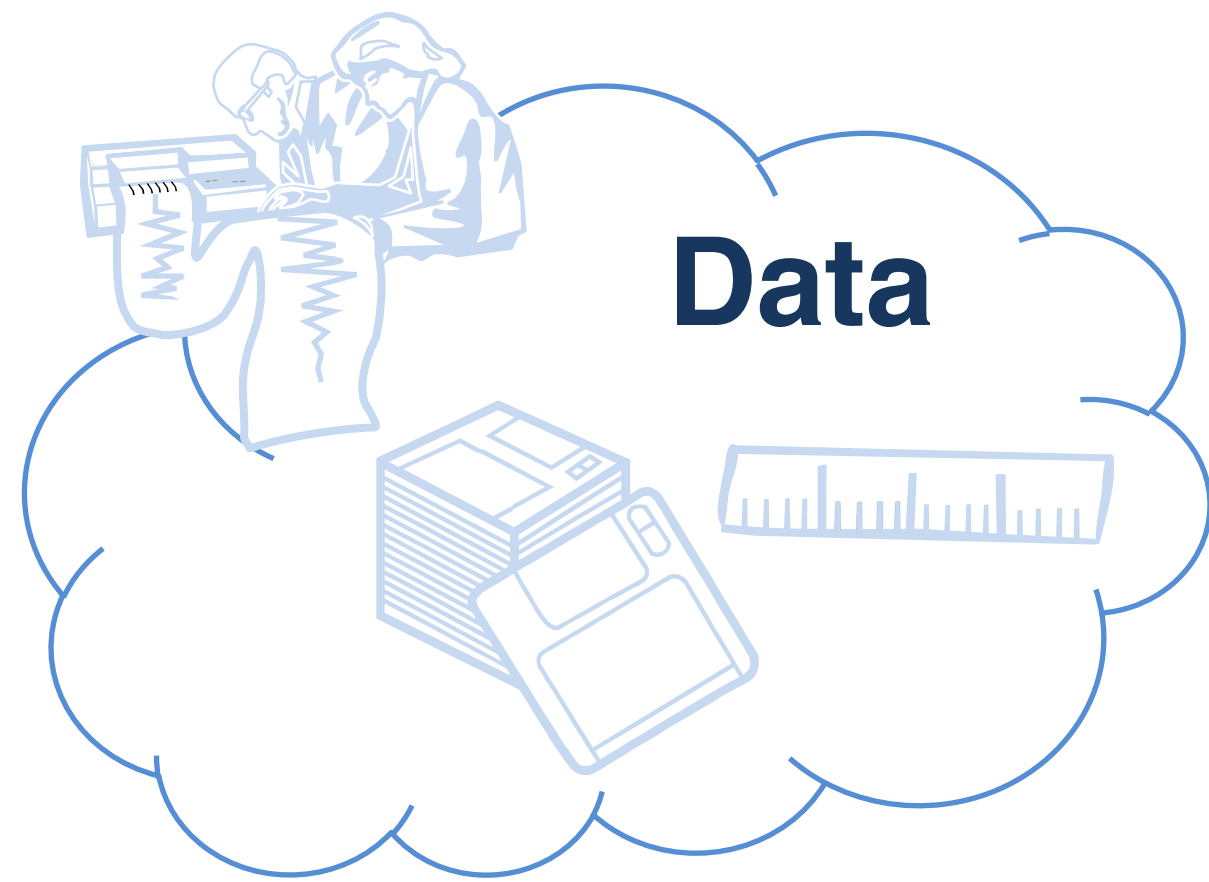


**MGI  
Ecosystems**



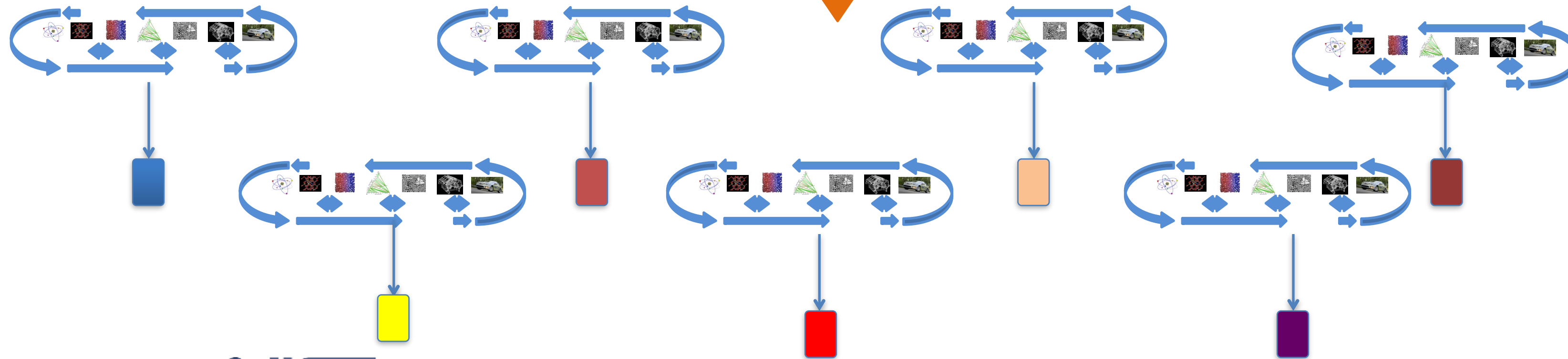
**Materials w/ Targeted Properties**





**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**





# 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

The screenshot shows a web browser window with the URL `www.usvao.org/index.html%3Fpage_id=344.html`. The page header features the VAO logo and the text "US Virtual Astronomical Observatory". A navigation menu includes "Home", "Science Tools & Services", "About the VAO", "VO News", "Support & Community", and "Contact & Connect".

## Data Discovery Tool

[Launch](#)  
(version 1.7)

Find datasets from thousands of astronomical collections known to the VO and over wide areas of the sky. This includes today's the most important collections from archives around the world such as:

- Chandra X-ray Observatory
- Hubble Space Telescope (HST) and other collections from the Multimission Archive at STScI (MAST)
- Multi-band collections from NASA HEASARC
- Sloan Digital Sky Survey (SDSS)
- Spitzer Space Telescope
- Two Micron All Sky Survey (2MASS)
- CDS Vizier Catalogs
- plus many more ...

**Features**

Retrieve astronomical data about a given position or object in the sky.

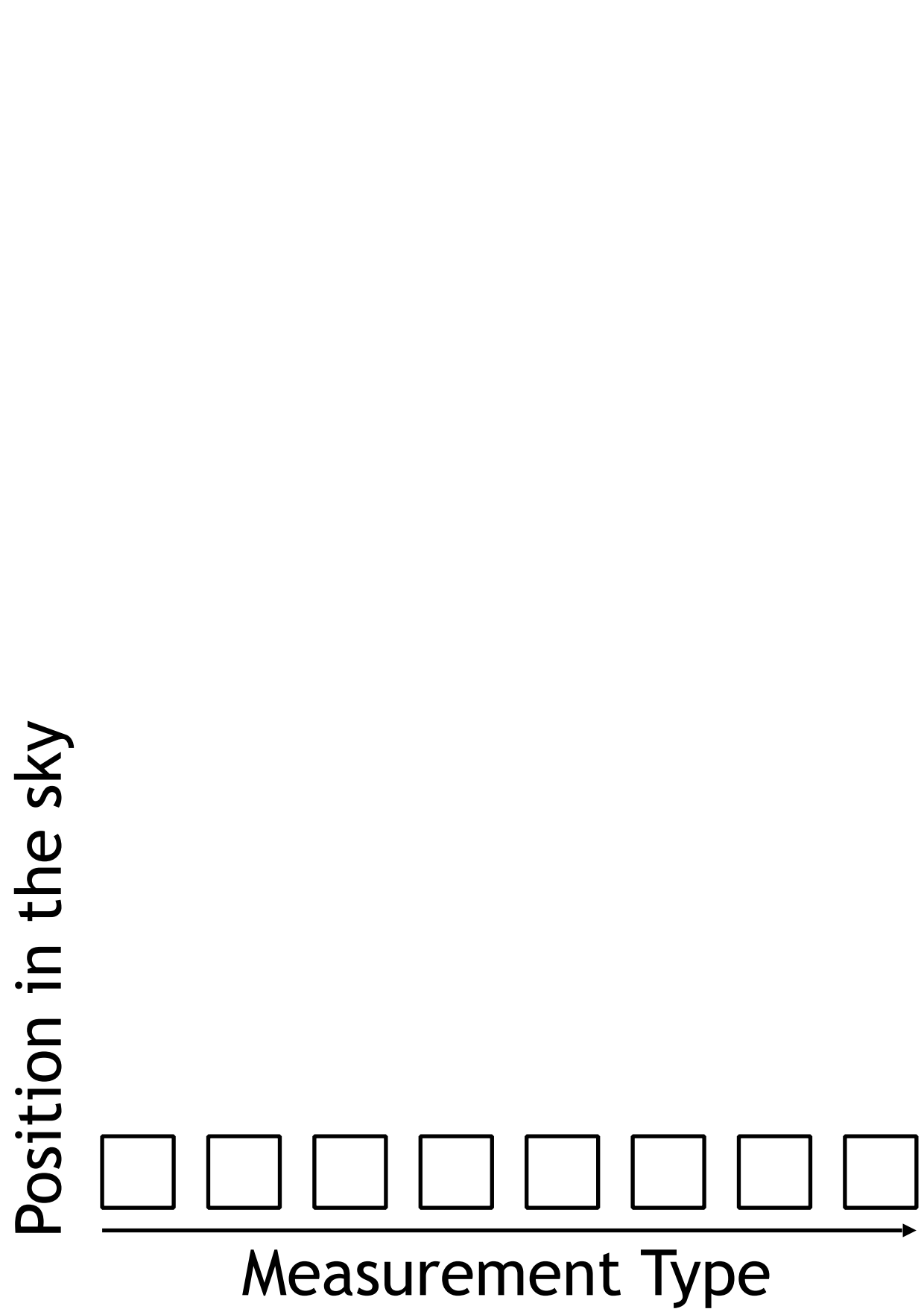
[More News from the VO](#)

- ▶ VAO Software Release: Data Discovery Tool (version 1.7)
- ▶ Summary Report of the VAO Close-Out Review
- ▶ Montage BSD 3-Clause License
- ▶ IVOA Newsletter 012 – May 2014
- ▶ VO Client Release – Access the VO from Your Desktop
- ▶ Aladin v8
- ▶ Comet VOEvent Implementation Update
- ▶ CASSIS Spectrum Analyzer Update

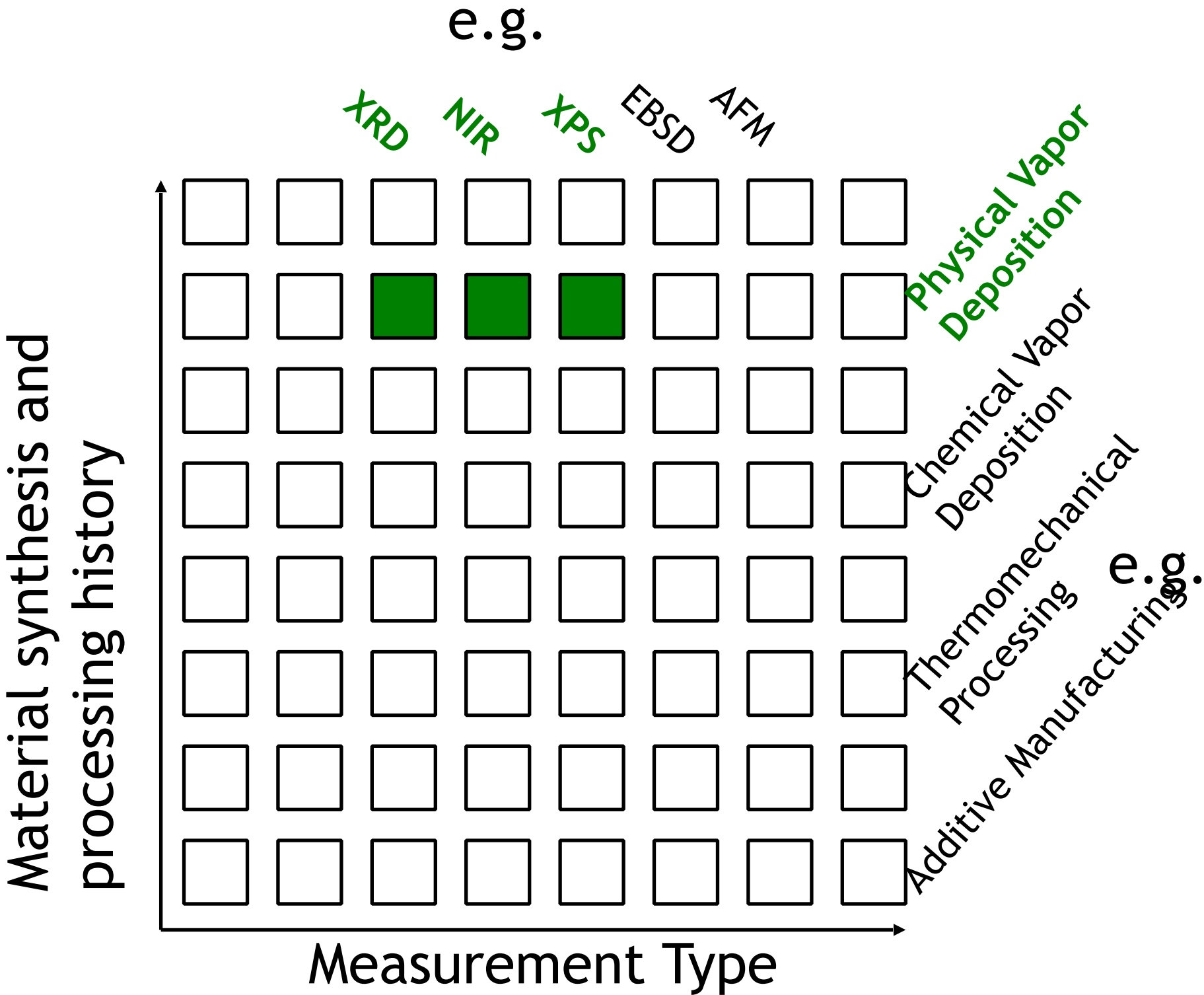
## Value Multiplier



# Astronomy vs. Material Measurement



Singular data models

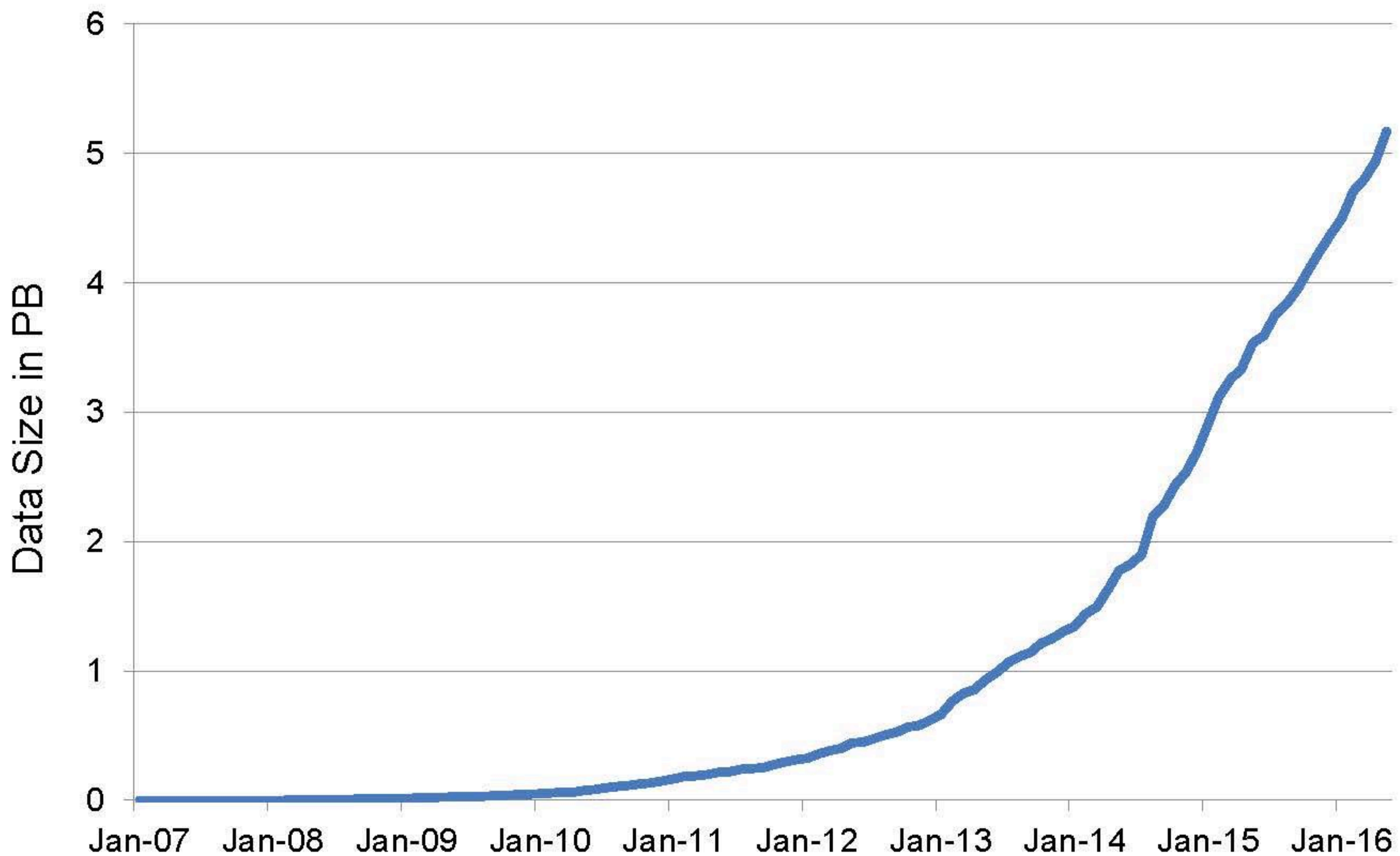


Modular data models

**WHAT'S OUR TELESCOPE?**



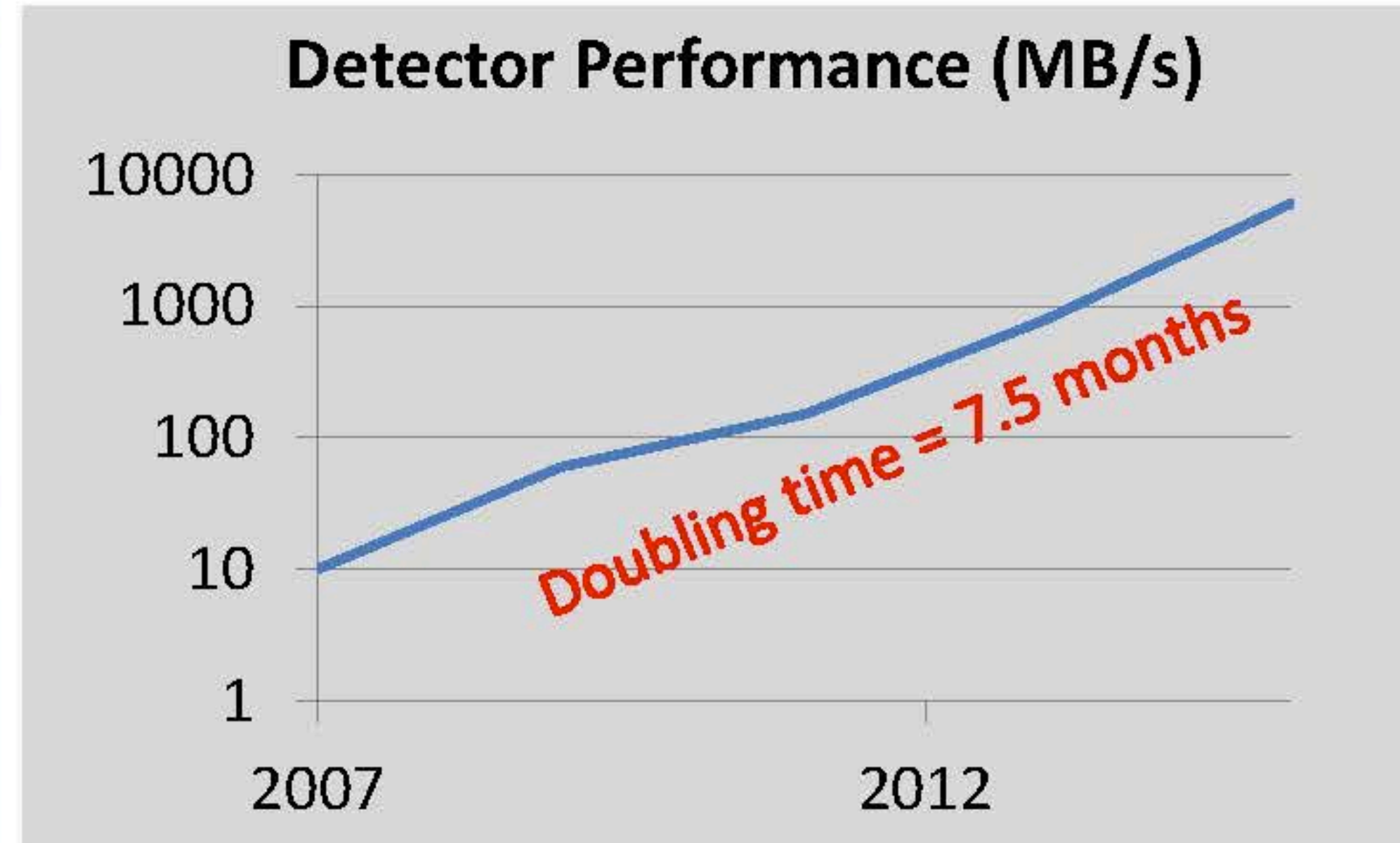
# Cumulative Amount of Data Generated By Diamond







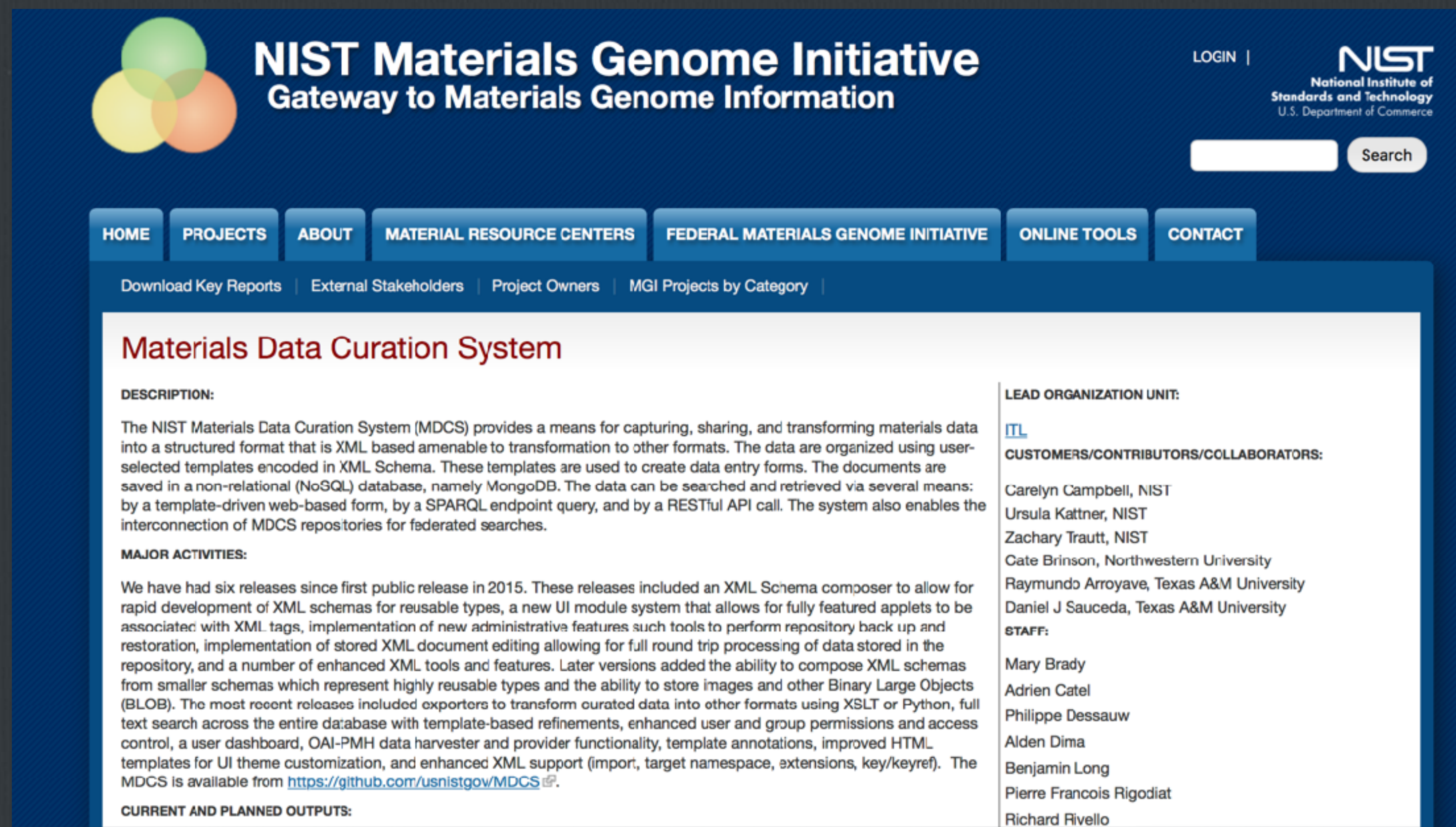
# Data Rates



- 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



# Better living through curation and schemas



**NIST Materials Genome Initiative**  
Gateway to Materials Genome Information

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

Search

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

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

## Materials Data Curation System

**DESCRIPTION:**

The NIST Materials Data Curation System (MDCS) provides a means for capturing, sharing, and transforming materials data into a structured format that is XML based amenable to transformation to other formats. The data are organized using user-selected templates encoded in XML Schema. These templates are used to create data entry forms. The documents are saved in a non-relational (NoSQL) database, namely MongoDB. The data can be searched and retrieved via several means: by a template-driven web-based form, by a SPARQL endpoint query, and by a RESTful API call. The system also enables the interconnection of MDCS repositories for federated searches.

**MAJOR ACTIVITIES:**

We have had six releases since first public release in 2015. These releases included an XML Schema composer to allow for rapid development of XML schemas for reusable types, a new UI module system that allows for fully featured applets to be associated with XML tags, implementation of new administrative features such tools to perform repository back up and restoration, implementation of stored XML document editing allowing for full round trip processing of data stored in the repository, and a number of enhanced XML tools and features. Later versions added the ability to compose XML schemas from smaller schemas which represent highly reusable types and the ability to store images and other Binary Large Objects (BLOB). The most recent releases included exporters to transform curated data into other formats using XSLT or Python, full text search across the entire database with template-based refinements, enhanced user and group permissions and access control, a user dashboard, OAI-PMH data harvester and provider functionality, template annotations, improved HTML templates for UI theme customization, and enhanced XML support (import, target namespace, extensions, key/keyref). The MDCS is available from <https://github.com/usnistgov/MDCS>.

**CURRENT AND PLANNED OUTPUTS:**

**LEAD ORGANIZATION UNIT:**

[ITL](#)

**CUSTOMERS/CONTRIBUTORS/COLLABORATORS:**

Carelyn Campbell, NIST  
Ursula Kattner, NIST  
Zachary Trautt, NIST  
Gate Brinson, Northwestern University  
Raymundo Arroyave, Texas A&M University  
Daniel J Saucedo, Texas A&M University

**STAFF:**

Mary Brady  
Adrien Catel  
Philippe Dessauw  
Alden Dima  
Benjamin Long  
Pierre Francois Rigodiat  
Richard Rivello

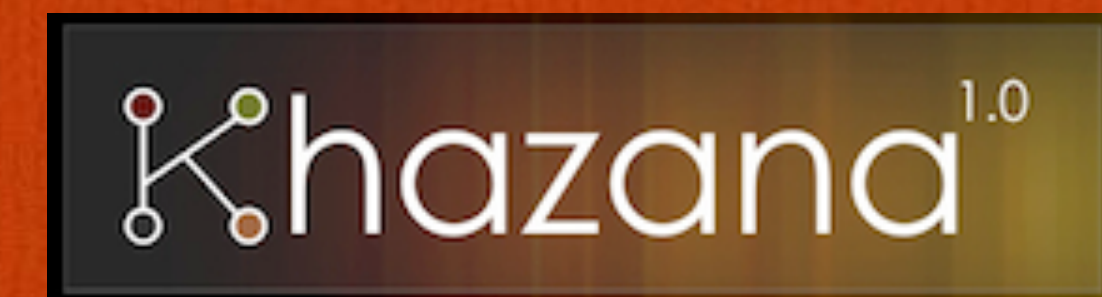
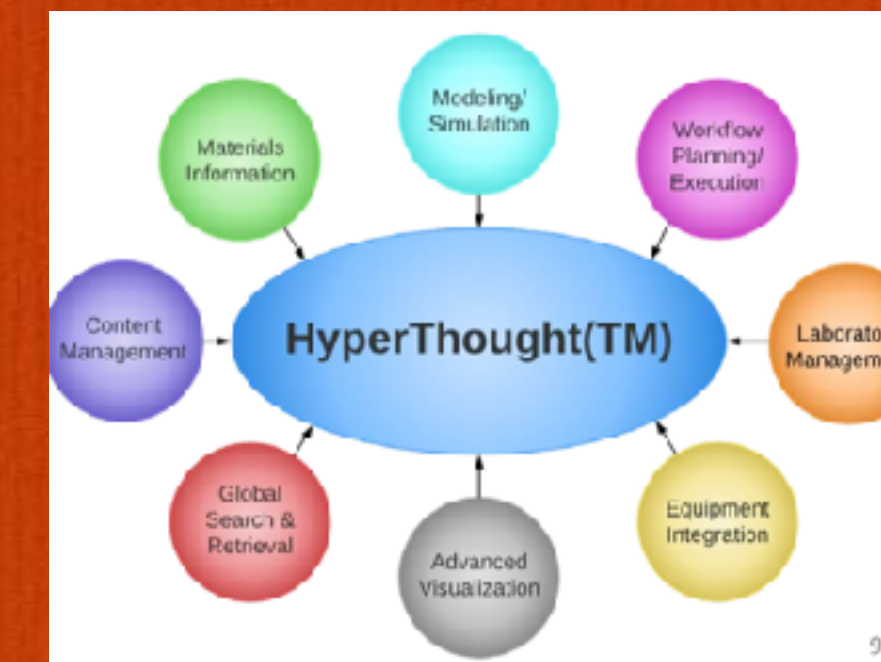
<https://github.com/usnistgov/MDCS>



# Infrastructures



Materials Innovation  
Platforms



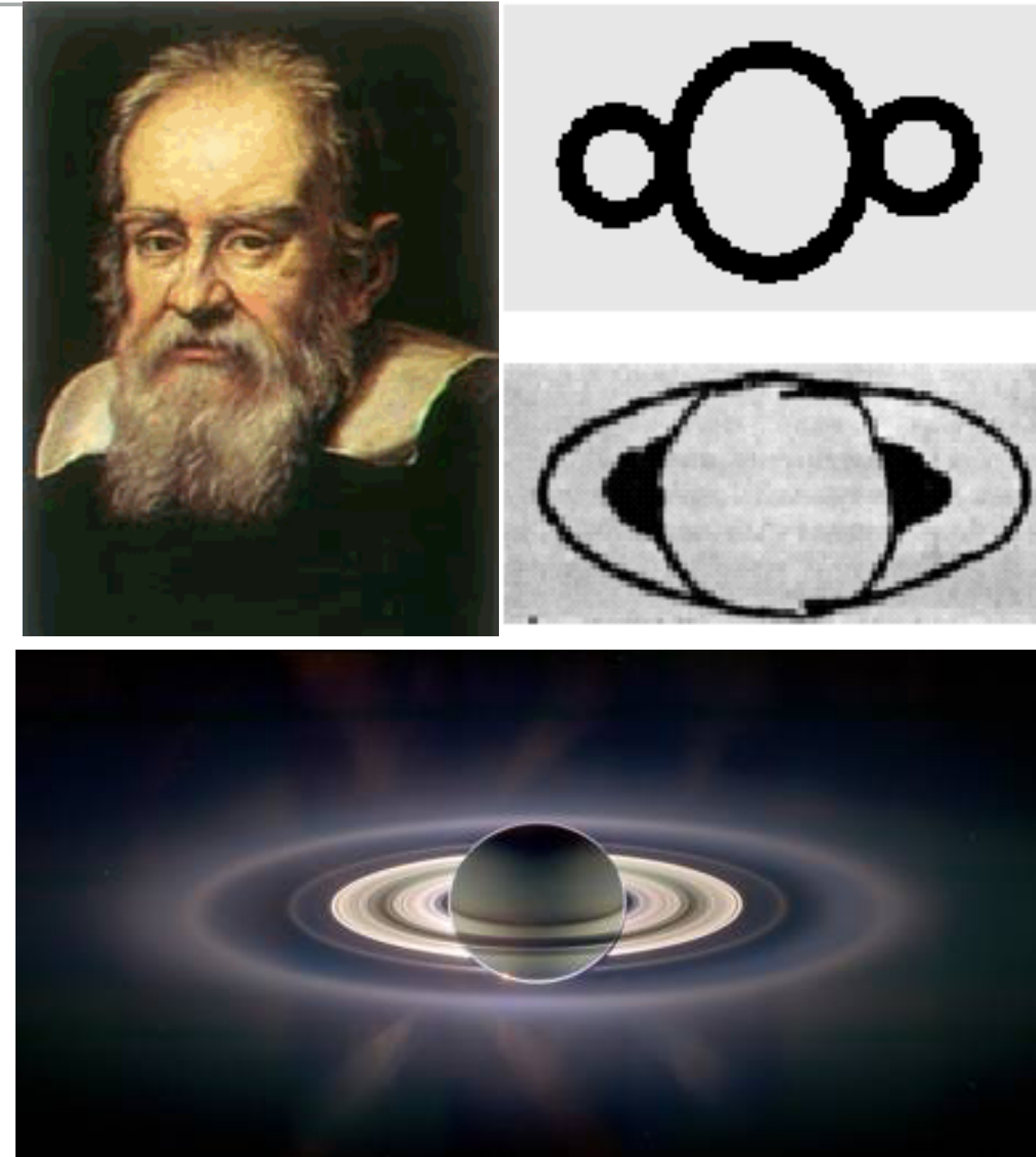


**Professional**



## 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 know 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**



**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



## 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**



2020

MATERIALS GENOME  
INITIATIVE  
STRATEGIC PLAN

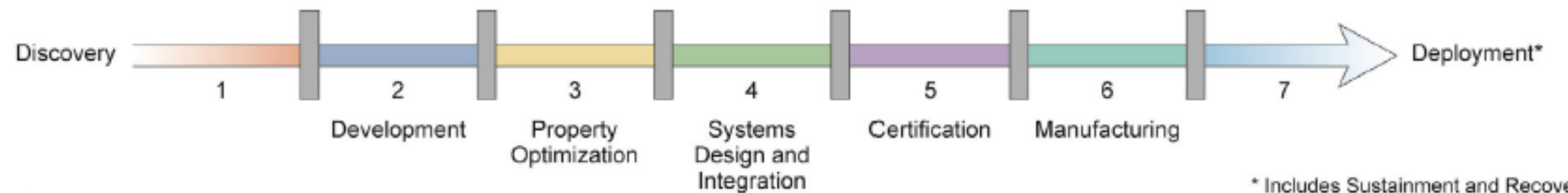
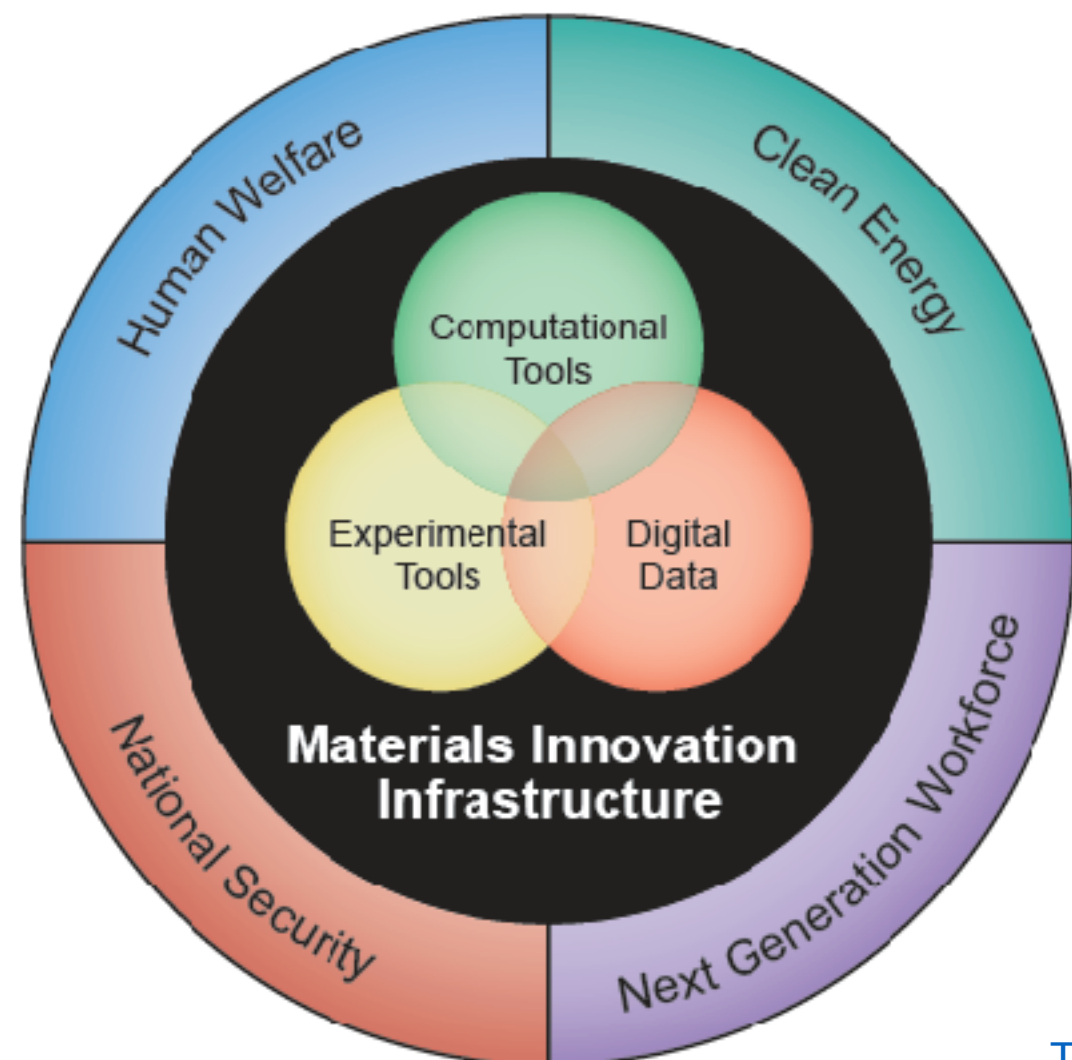
Materials Genome Initiative  
National Science and Technology Council  
Committee on Technology  
Subcommittee on the Materials Genome Initiative

DECEMBER 2014



# MGI 2.0

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



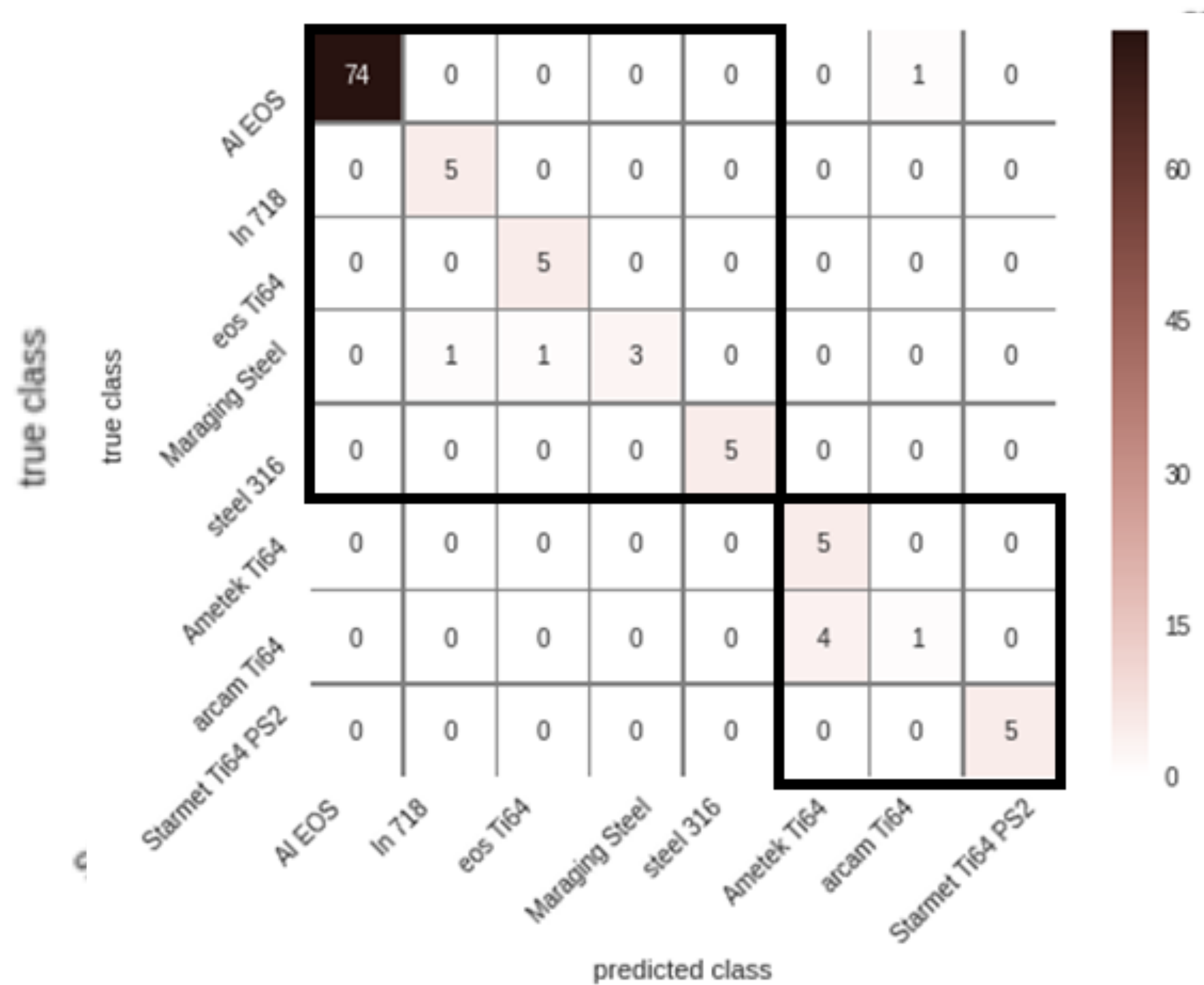


**AI/ML is a true enabler and  
driver for MGI**



# SEM powder classification

independent samples





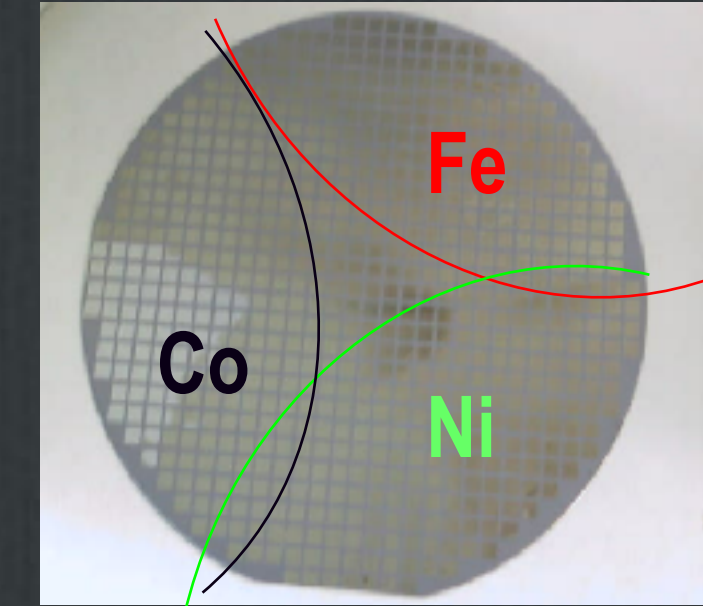
# ML and Autonomous Materials Science

A. Gilad Kusne and Collaborators

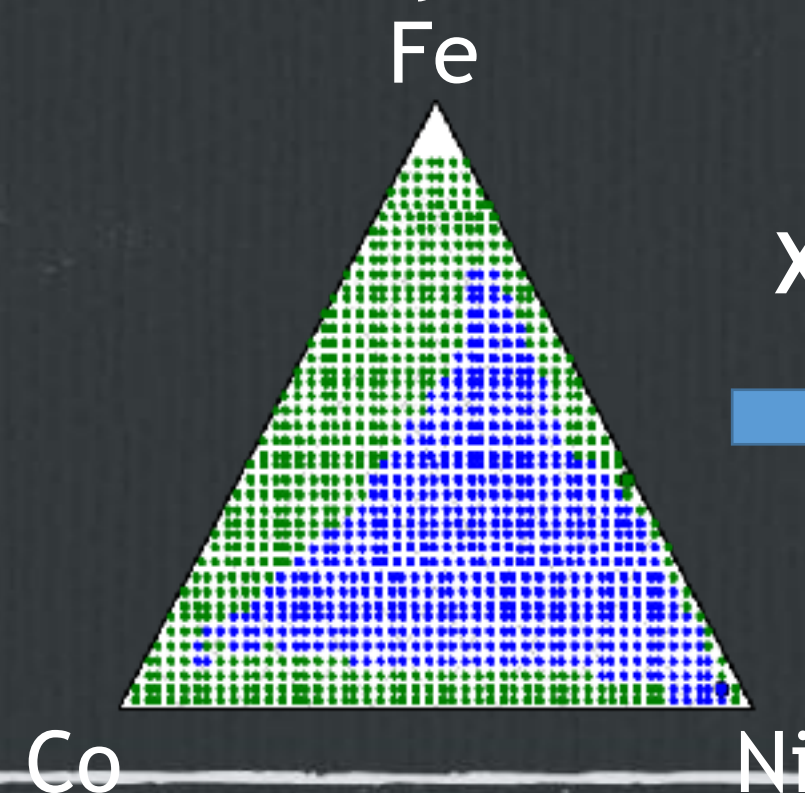


# Phase Mapping: High-Throughput Approach (APL Materials 2016)

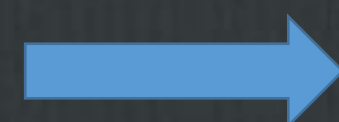
- Fabricate hundreds-thousands of samples -> HiTp Synthesis
- Measure all samples -> HiTp Characterization
- Rapid phase mapping -> Machine Learning



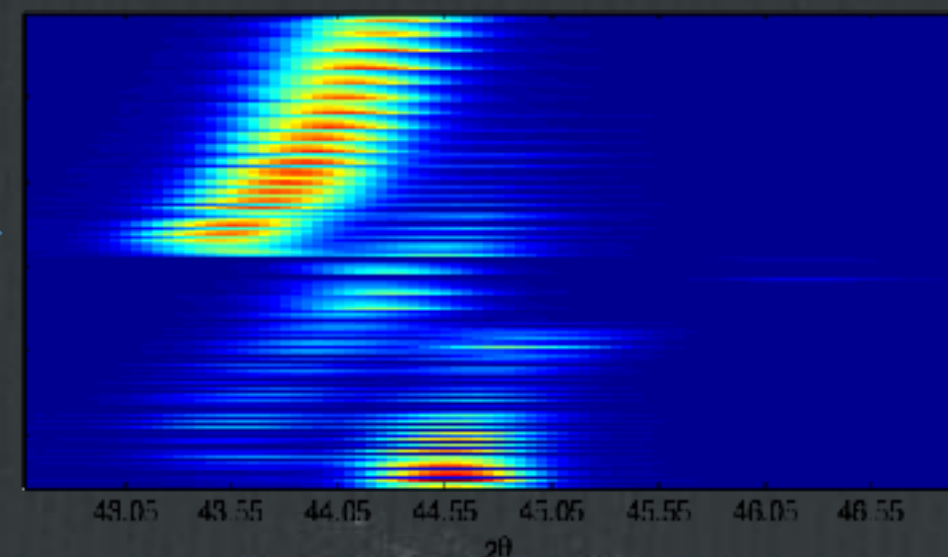
Combi Library for Ternary Spread



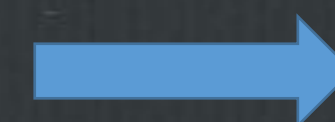
XRD



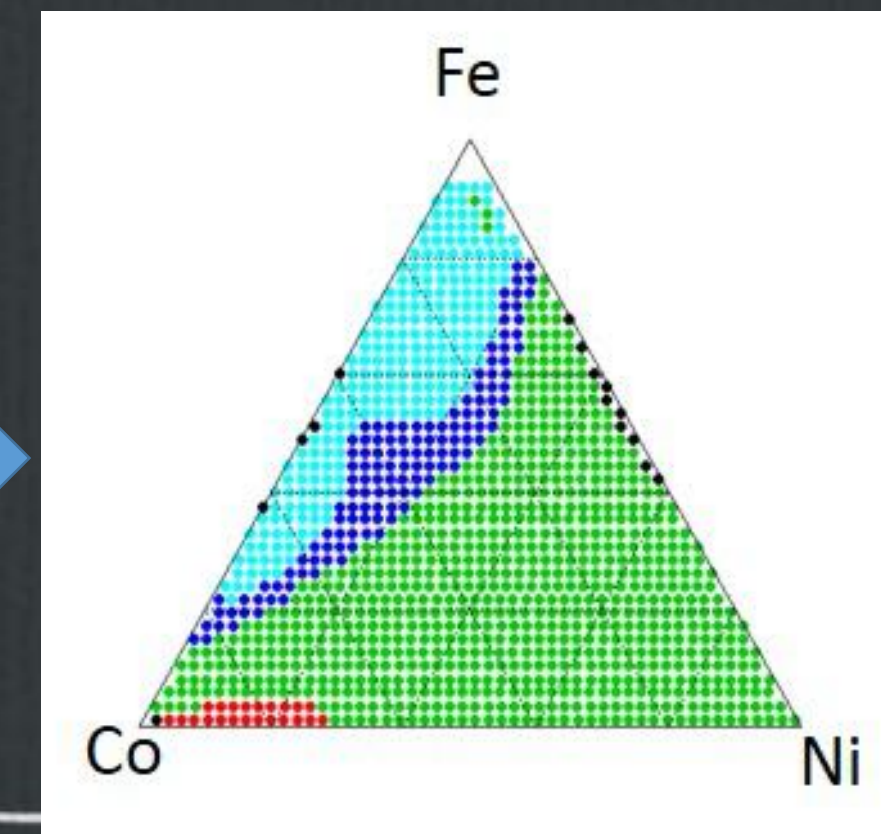
Diffraction Patterns



Cluster Analysis



Estimated Phase Map





# Phase Mapping: High-Throughput Approach

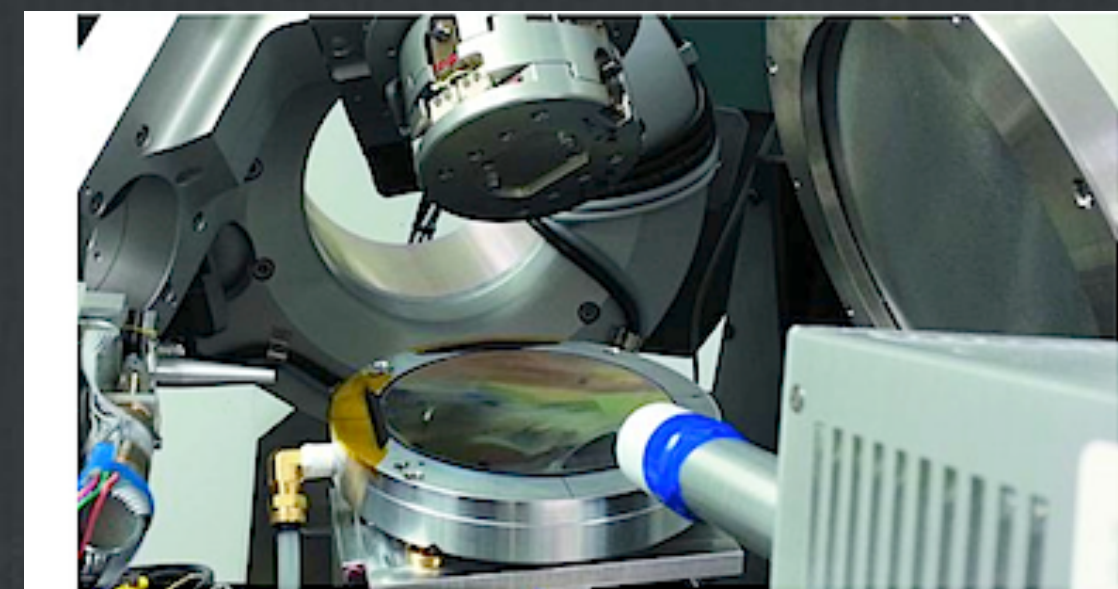
- Measurement is a time / resource sink
- For wafer of 500+ samples:
- In Lab: Takes weeks-months
- Synchrotron: Takes 5+ Hours (Every second counts)



Mn-Ni-Ge library  
535 samples



Bruker D8  
30 Minutes per  
sample  
2 weeks!

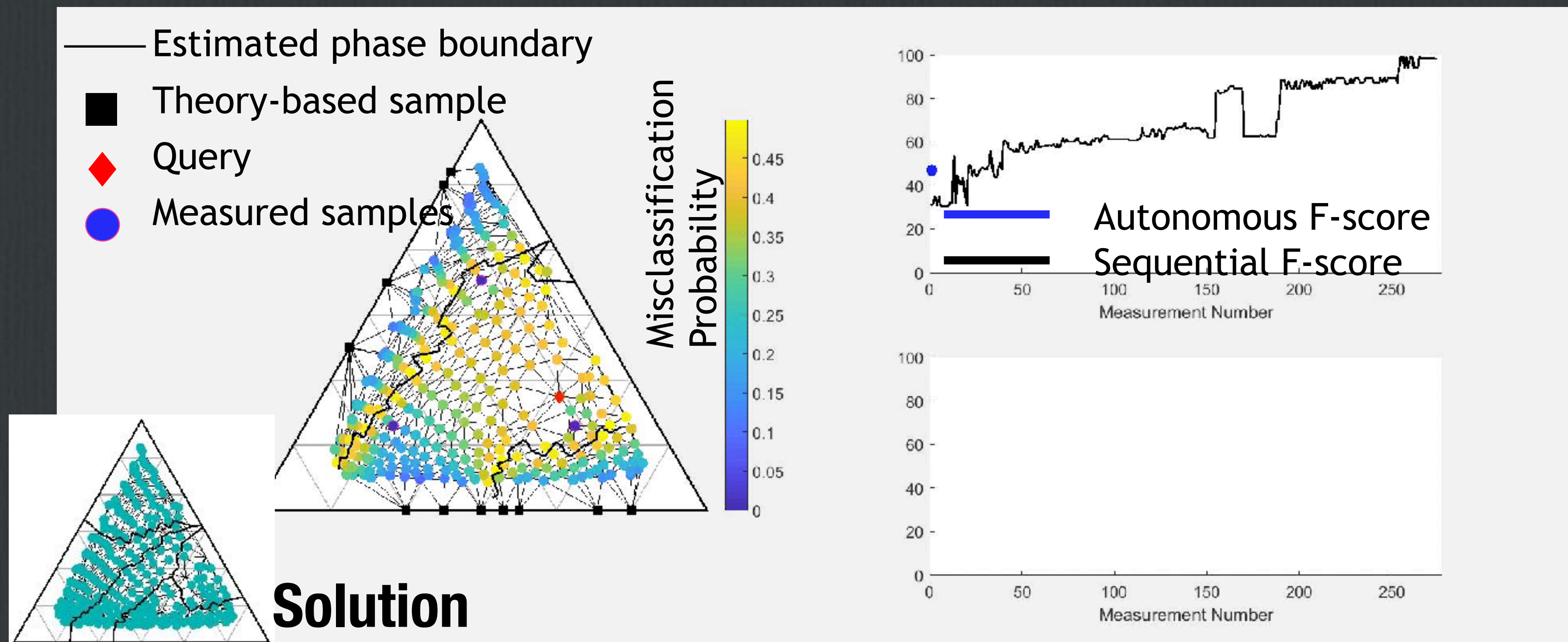


Stanford  
Synchrotron  
Radiation  
Lightsource  
30 seconds per  
sample  
4.5 hours



# Autonomous Phase Mapping

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



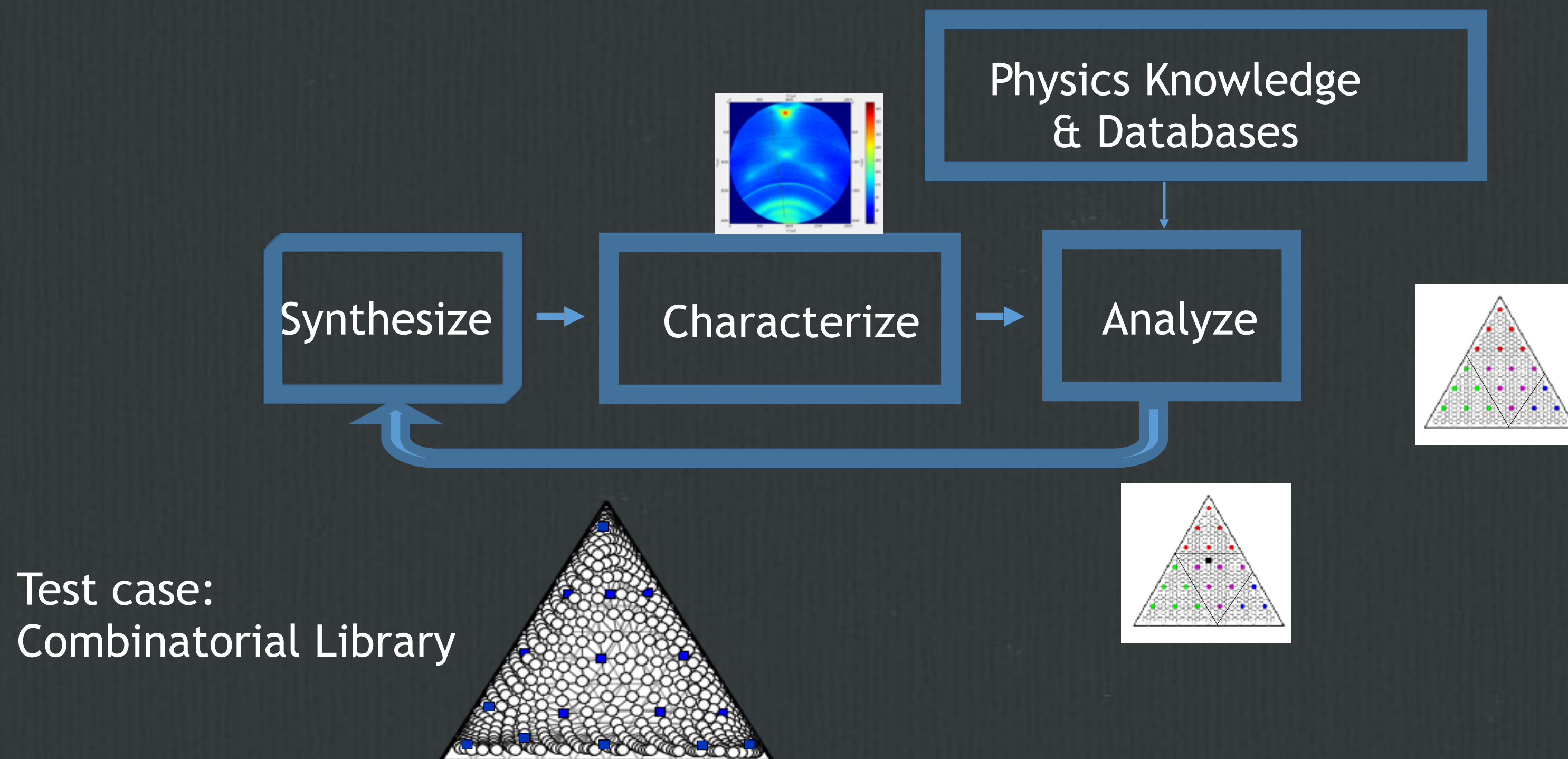
AI is controlling X-ray diffraction systems at SLAC & in the lab!





# Automated Phase Mapping

- Now: Place AI in control of Synthesis.







Questions?