NIST and The Materials Genome Initiative

James A Warren

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

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

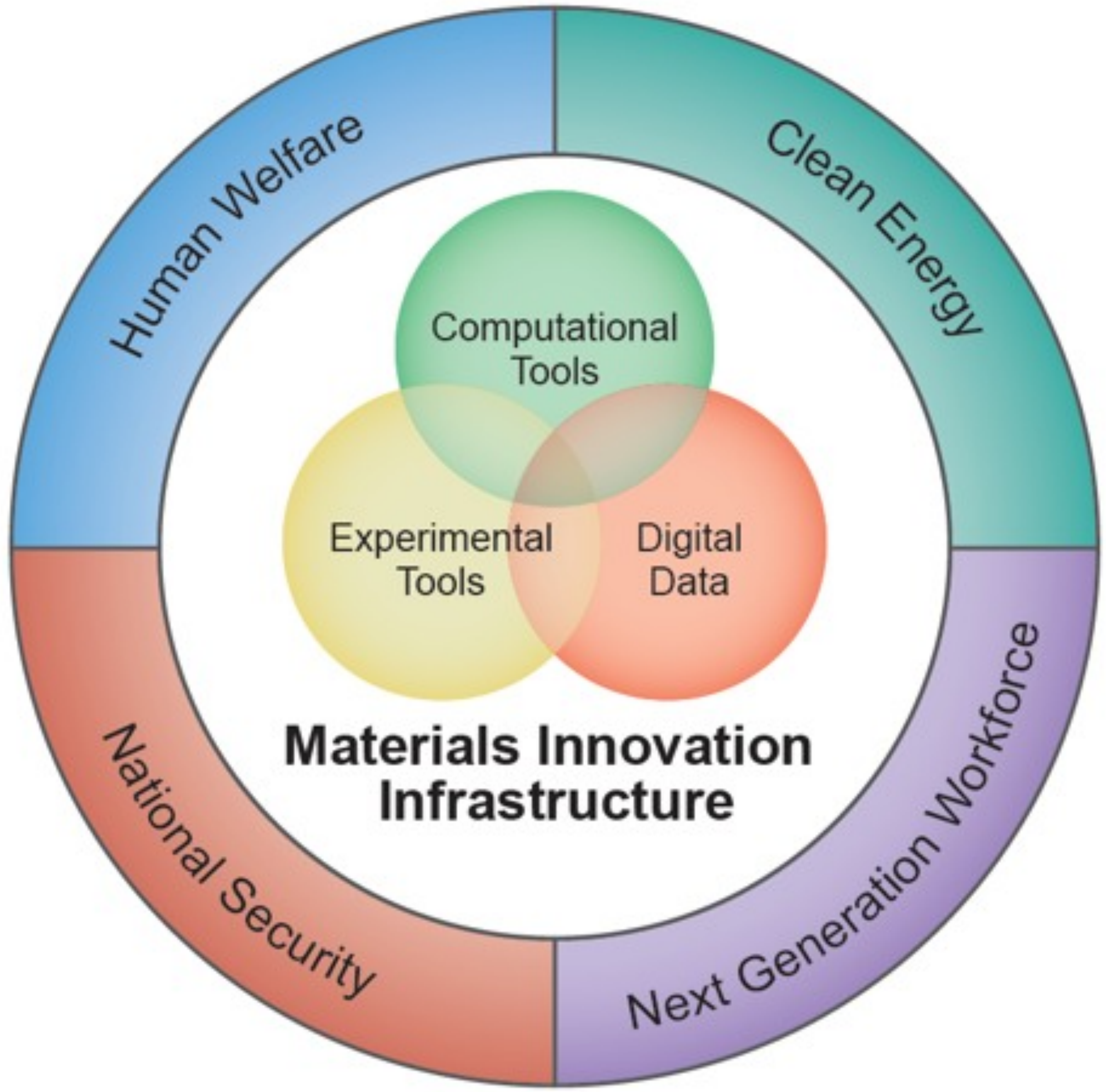
Carnegie Mellon University, June 2011

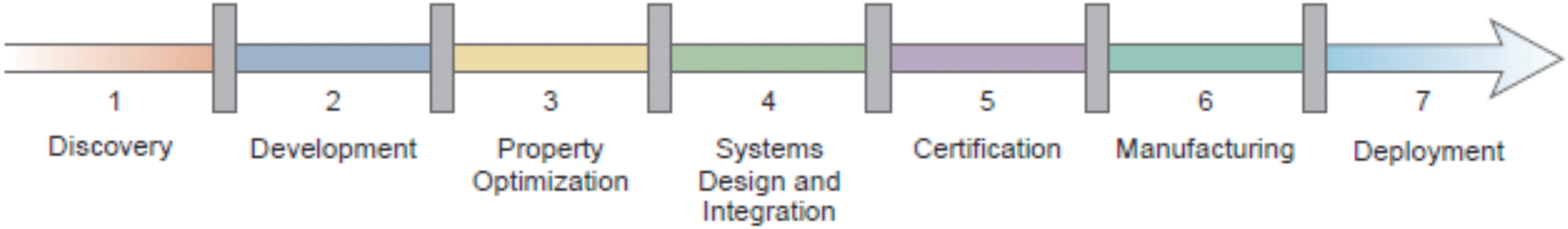
TWO YEAR ANNIVERSARY COMING!

The Materials Genome Initiative (MGI)

Goal:
To decrease the time-to-market and cost for new materials by >50%.

The Materials Innovation Infrastructure





What is the End State?

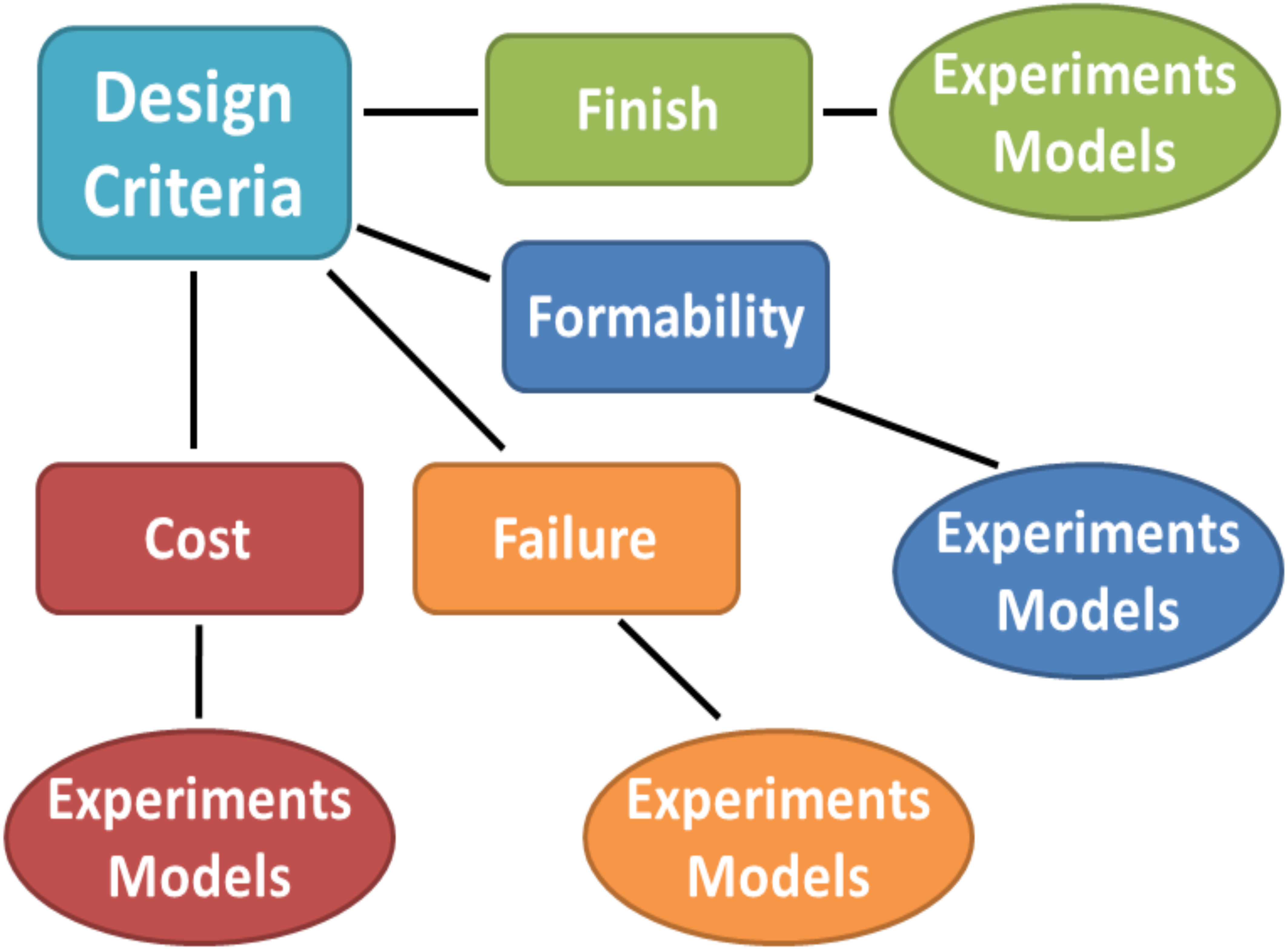
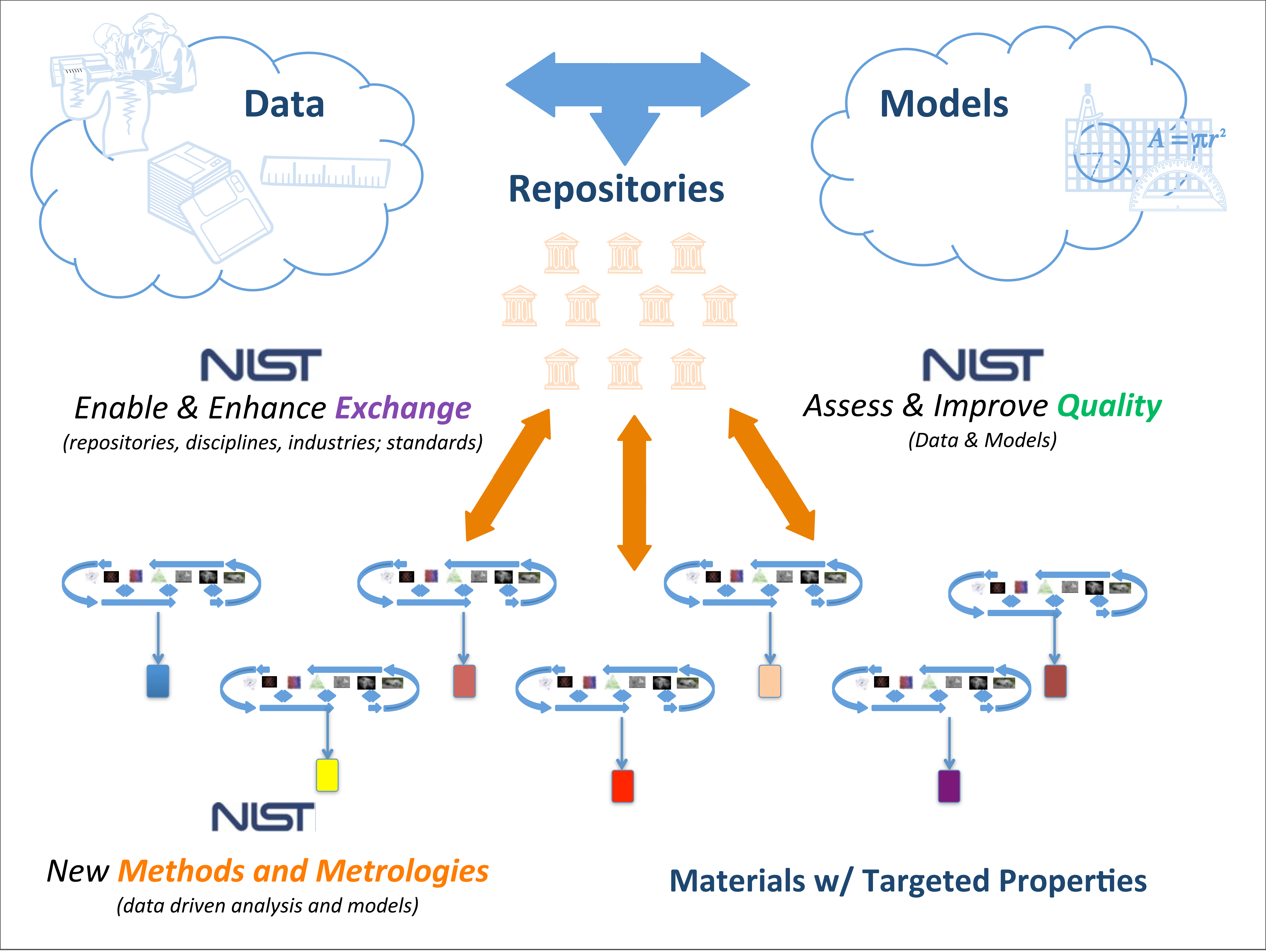


Figure 1. Schematic of how the design criteria for a given material dictate the needed material properties and thus define the needed experiments, models and data.


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

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

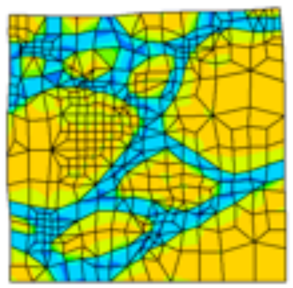
Goal 3: NIST establishes *new methods, metrologies and capabilities* necessary for accelerated materials development.

Structural Alloys Pilot

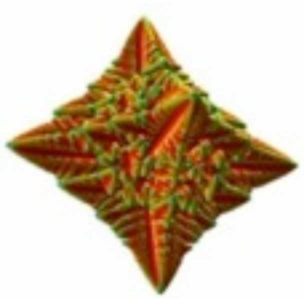
Macroscopic Materials Properties

MATERIAL PROPERTIES

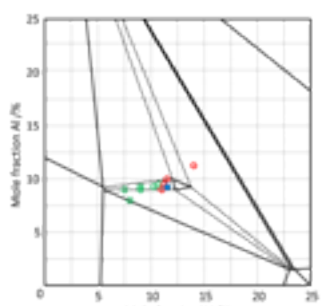
Structure-Property Relationships
Computation, Theory, Expt

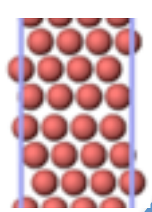


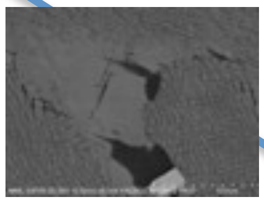
Processing-Structure Relationships
Phase-field, FE, etc.



Composition Dependent Databases
CALPHAD, Diffusion Mobility







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

Experimental Inputs
Crystal Structure, Diffusion, Thermochemical Data

Atomic Structure and Properties

Collaborations
Universities
Industry
MML (3 Divisions)
ITL

Crosscuts
Repositories
Workflow Tools
Standards
Protocols
Uncertainty
Dissemination

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

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

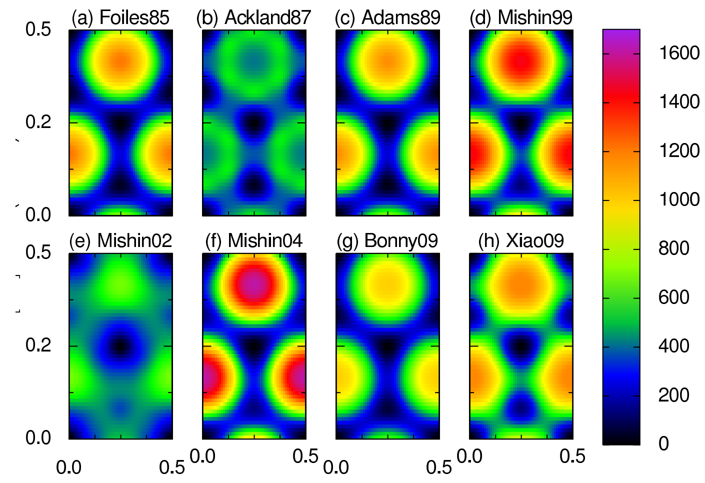
- Cloud
- Statistical Learning
- Big Data
- Knowledge Discovery

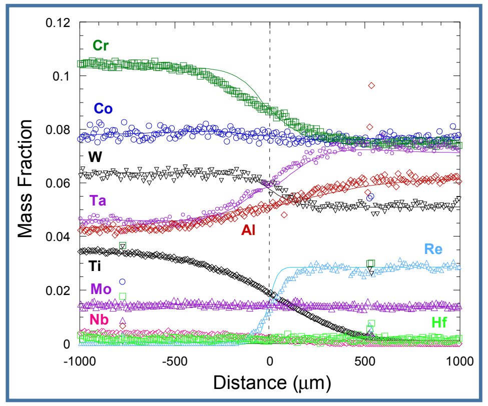
- very fast growing
- *many* new jobs
- new degrees/depts

DATA TOOLS AND INFORMATICS FOR PHASE-BASED MATERIALS DATA

Need

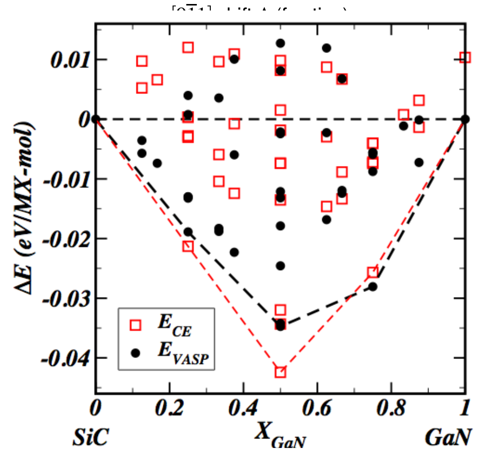
- Improved efficiency and reproducibility of thermodynamic and kinetic simulations (e.g. CALPHAD, first principles, atomistic methods).
- Accessible phase-based data described as functions of composition, temperature and pressure.

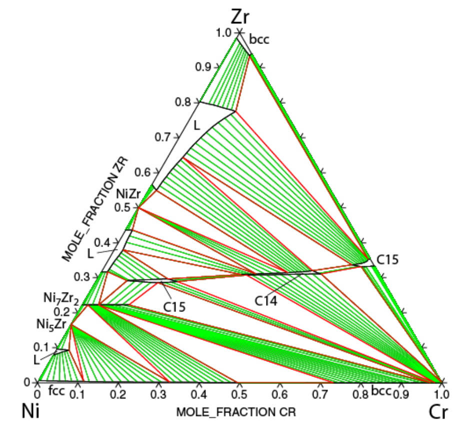




Objectives

- Develop file repositories that enable links to data files and the ability to identify key metadata.
- Develop informatics tools to enable data capture and retrieval of phase-based data.
- Develop tools that are available for the community to contribute towards and use.





Achievements and Impact

- NIST D-Space File Repository established with Kent State (http://nist.matdl.org/)
- NIST Interatomic Potentials Repository (www.ctcms.nist.gov/potentials) : *100+ down-loadable referenced interatomic potentials*
- Materials Data Curator system is under development

Customers and Partners

CALPHAD Approach

- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

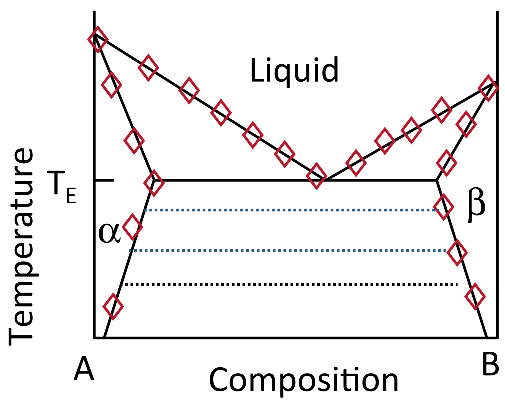
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

Binaries → Ternaries → Quaternaries → n^{th} Order Systems

CALPHAD Approach

- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data



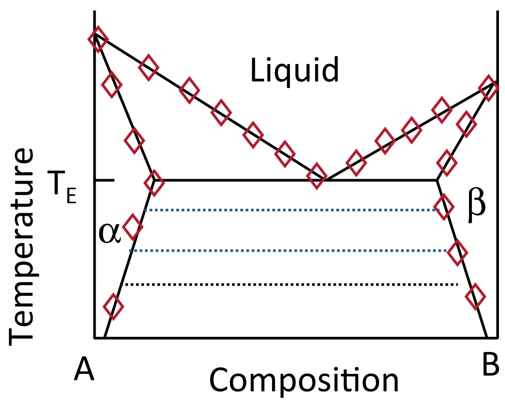
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

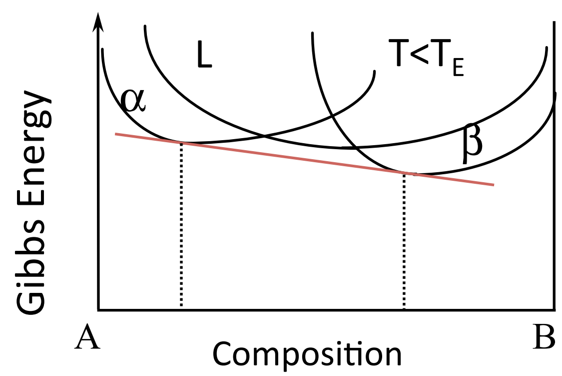
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

CALPHAD Approach

- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → →





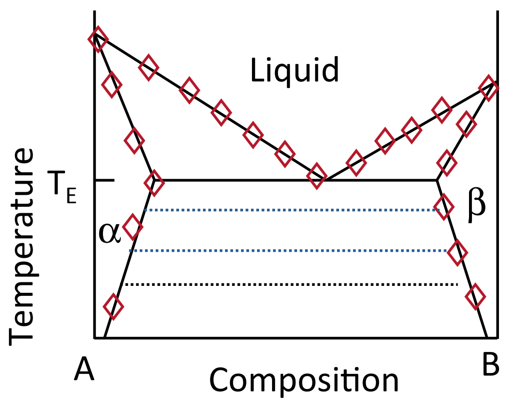
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

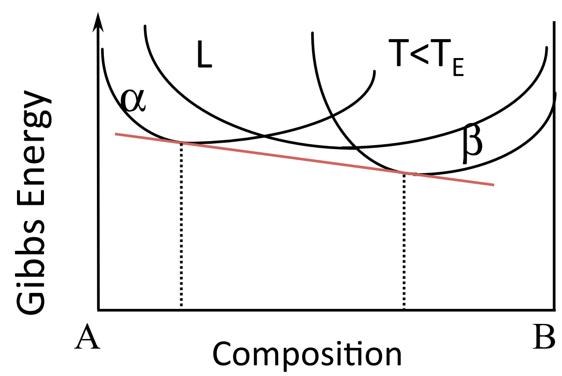
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

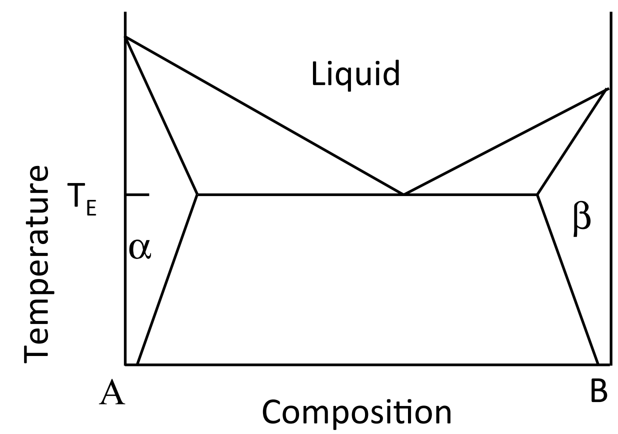
CALPHAD Approach

- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → Calculated phase diagram →







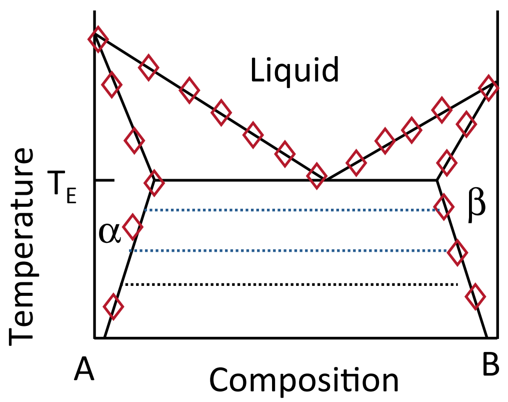
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

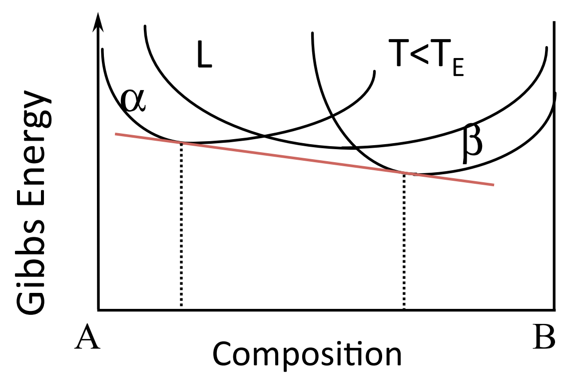
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

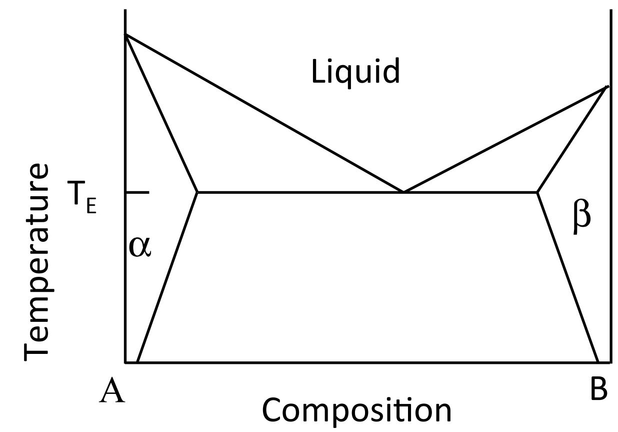
CALPHAD Approach

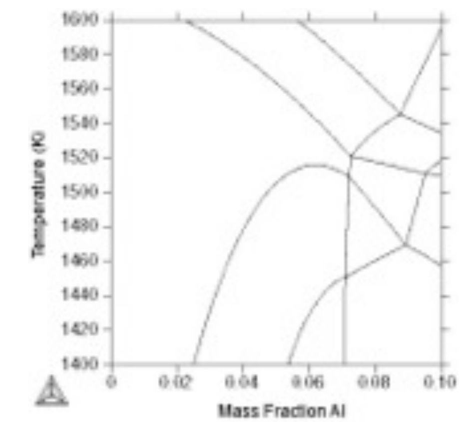
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f\,(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

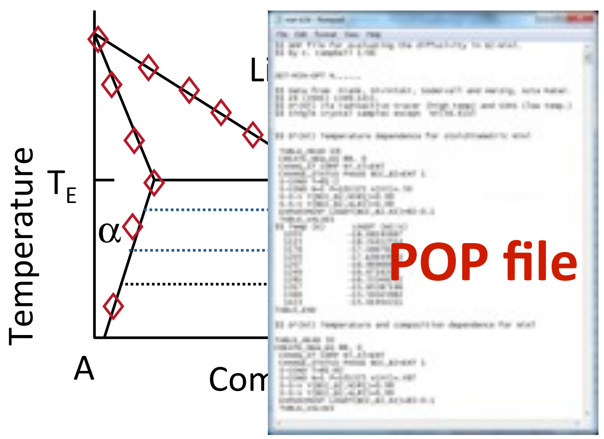
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

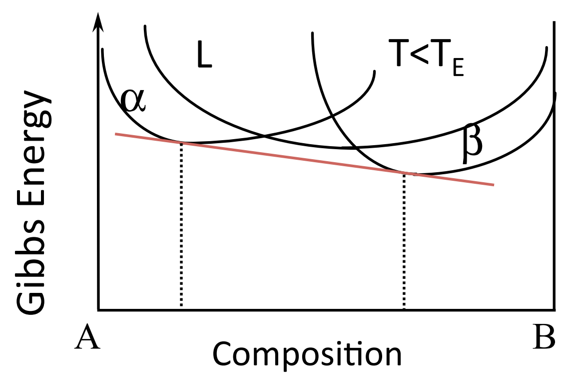
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

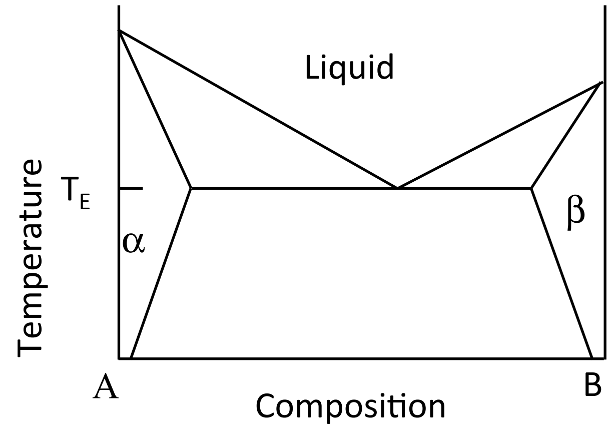
CALPHAD Approach

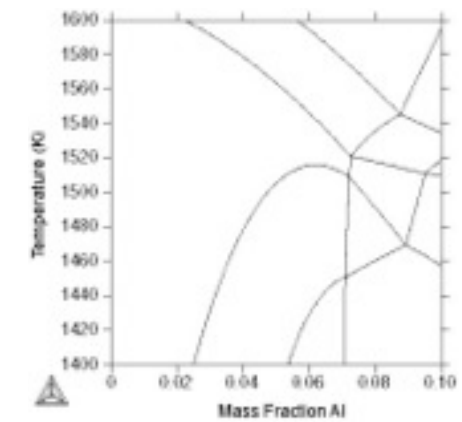
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

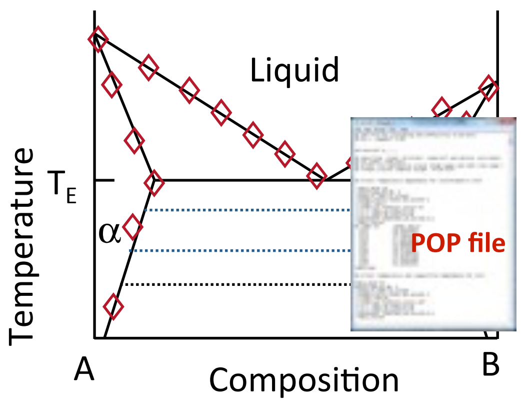
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

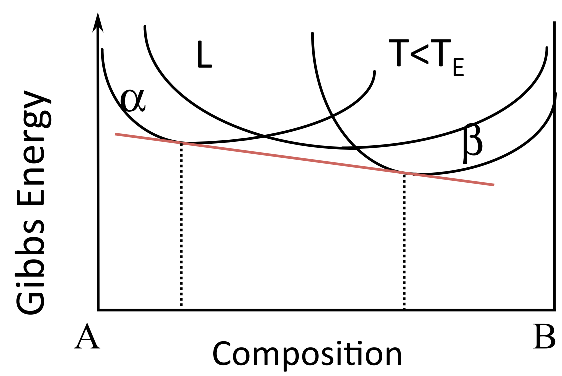
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

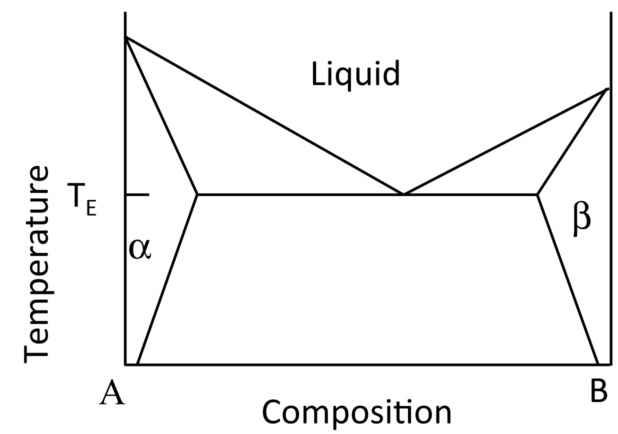
CALPHAD Approach

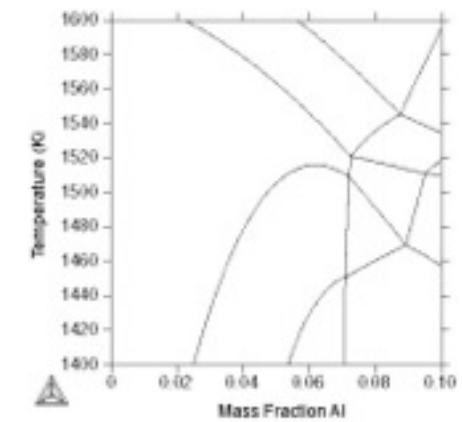
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x, T, P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

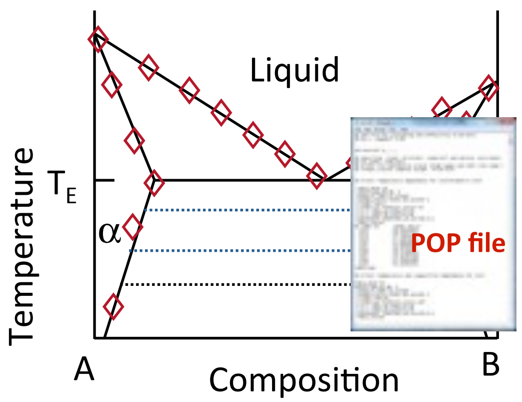
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

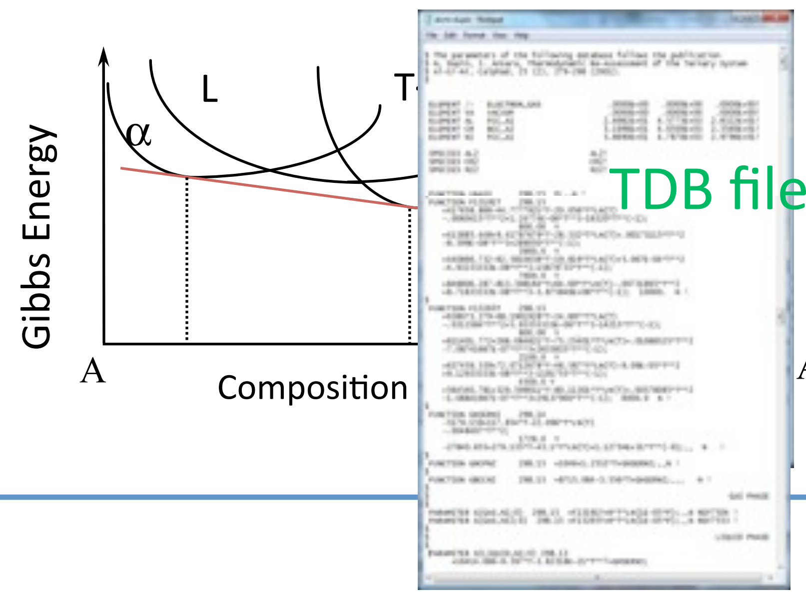
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

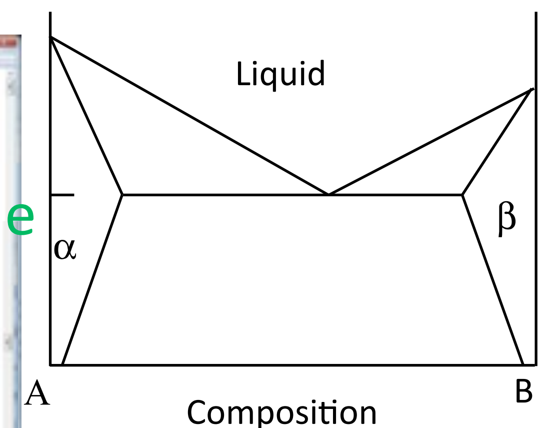
CALPHAD Approach

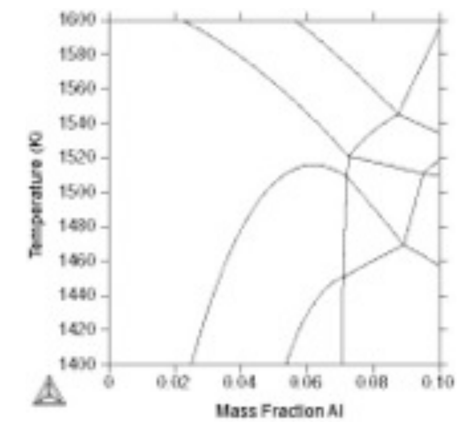
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

Binaries → Ternaries → Quaternaries → n^{th} Order Systems

CALPHAD Approach

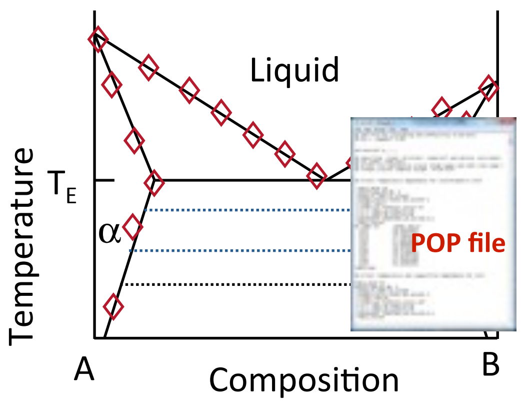
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

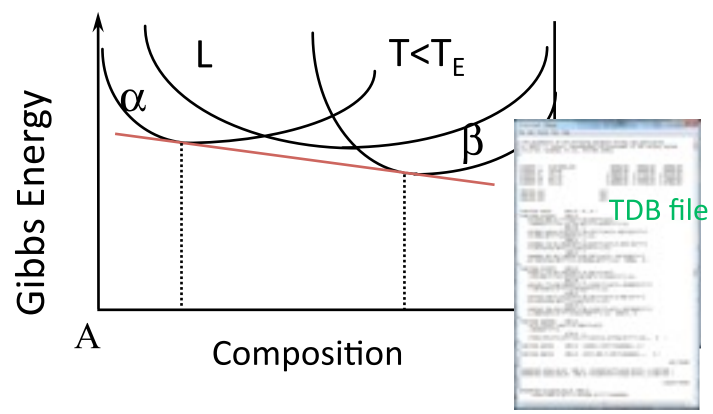
Experimental phase diagram and thermochemical data

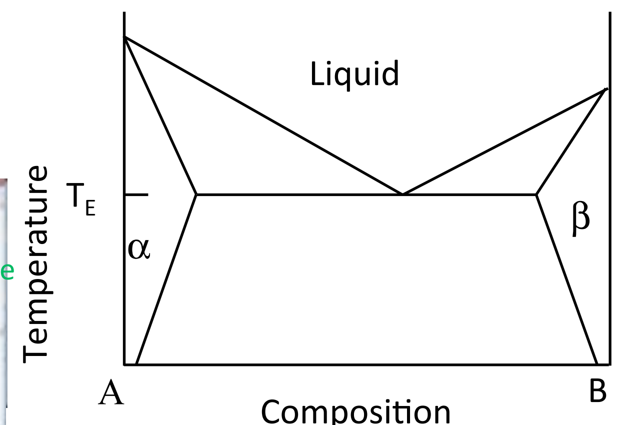
Determine Gibbs energy functions for each phase: $G = f(x,T,P)$

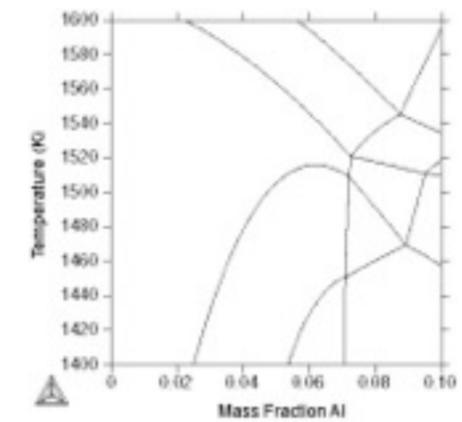
Calculated phase diagram

Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

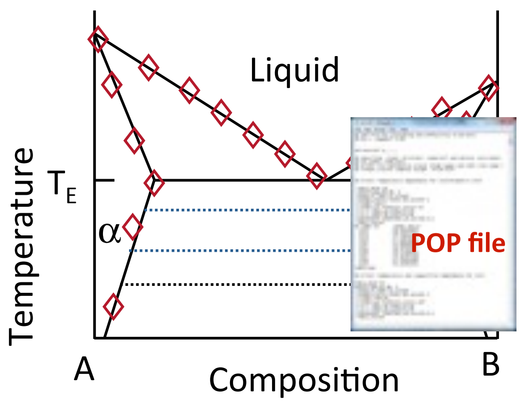
$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

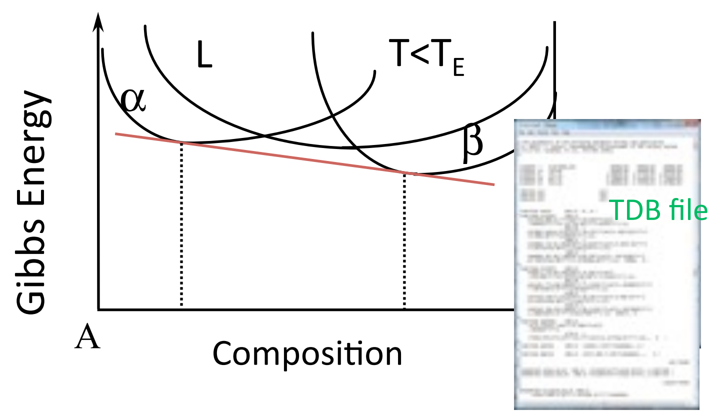
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

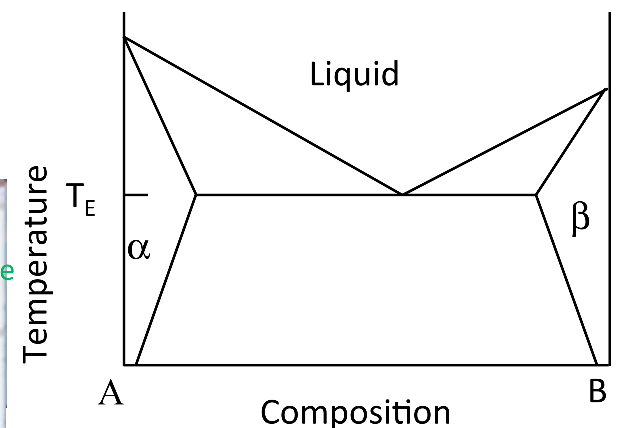
CALPHAD Approach

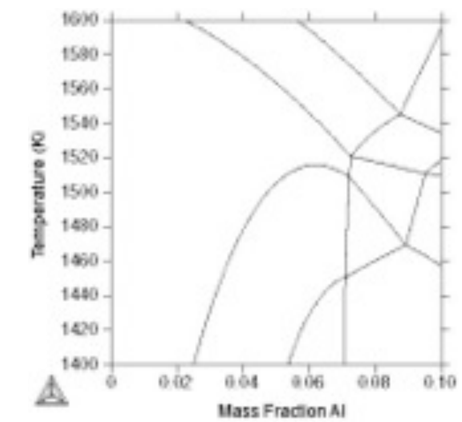
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f(x, T, P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

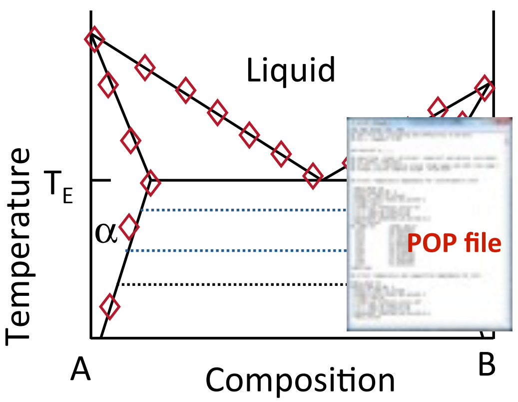
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

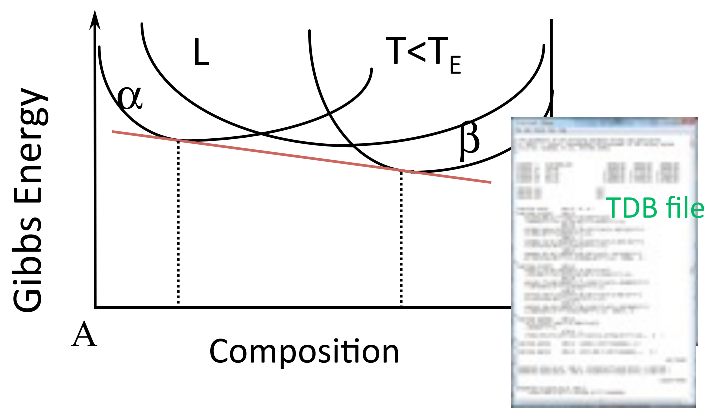
➡ **True quaternary compounds are rare in metallic systems**

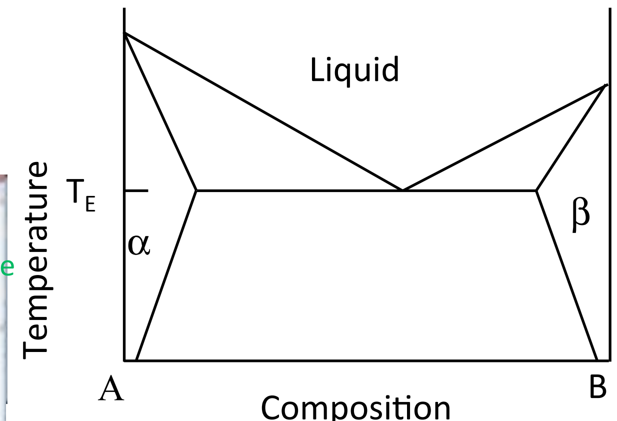
CALPHAD Approach

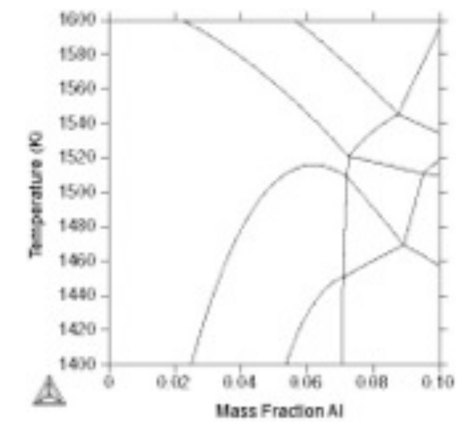
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f\,(x, T, P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

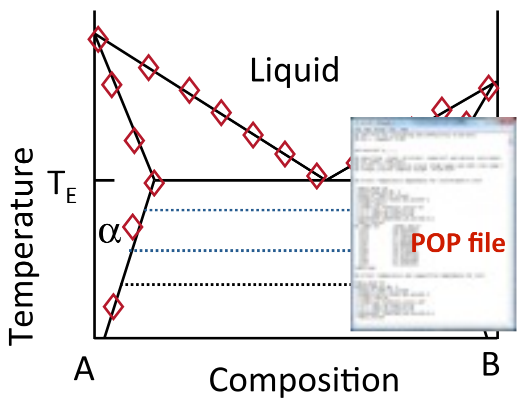
Binaries → Ternaries → Quaternaries → n^{th} Order Systems

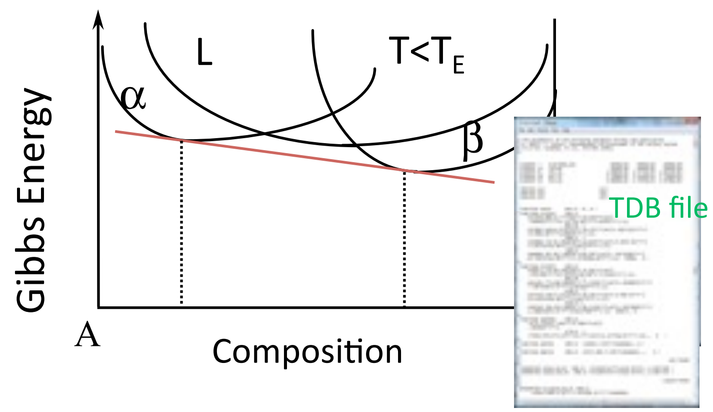
- **True quaternary compounds are rare in metallic systems**
- **Assessment of ternary systems is usually sufficient for the description of a multicomponent system**

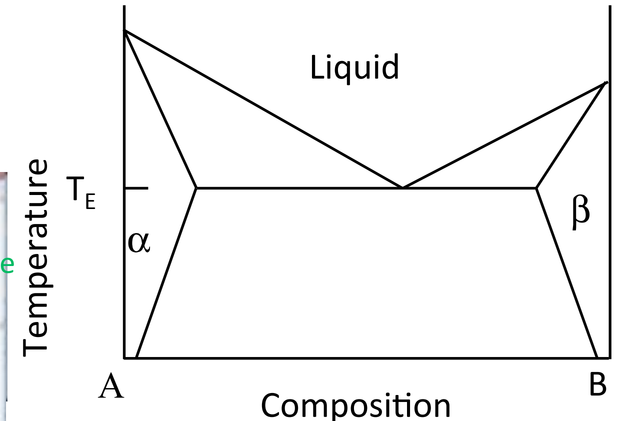
CALPHAD Approach

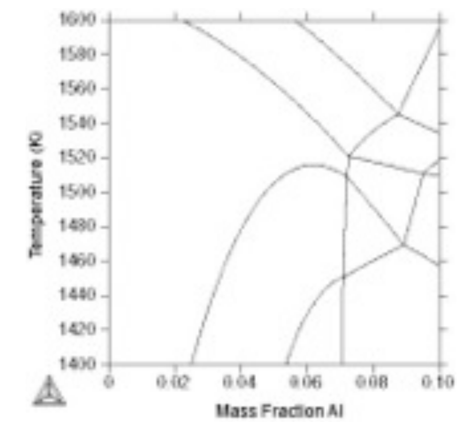
- Collected experimental and computational data are used to fit functions.
- Functions are used to calculate phase equilibria, including phase diagrams.

Experimental phase diagram and thermochemical data → Determine Gibbs energy functions for each phase: $G = f\,(x,T,P)$ → Calculated phase diagram → Combine binaries and ternaries to predict multi-component systems









Al isopleth for 8 component system

$$G^{\phi} = G^{0} + G^{ideal} + G^{excess}$$

Binaries → Ternaries → Quaternaries → n^{th} Order Systems

- True quaternary compounds are rare in metallic systems
- Assessment of ternary systems is usually sufficient for the description of a multicomponent system
- Same methodology can be applied to the description of other property data

CURRENT EFFORTS TOWARD A GENERAL MATERIALS DATA REPOSITORY

NIST priority under goals of Materials Genome Initiative with initial focus is on phase-based data (used by CALPHAD-based assessment)

- Developing needed universal identifiers
  - Crystal Structure
  - Phase Names
  - Materials (including processing history and composition)
- Developing ontologies to describe materials data with initial emphasis on
  - Phase equilibria
  - Thermochemistry
  - Diffusion
- Investigating a variety of data-interchange and representation formats (e.g. JSON, BSON, XML)
- Developing tools to transform data for use in other modeling tools
- Constructing Wiki to define terminology

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

Questions

- What metadata, and what kinds of metadata management, are needed to enable re-use of data, both across domains and across silos within domains?
  - See above presentation
- How can we incentivize researchers and providers to curate their data, organize it with useful metadata, and make it publicly available?
  - Make Transparent the ROI
- Maximum impact of data occurs when analytics make use of all available relevant data; how can analytics developers be challenged to make this standard practice?
  - Make the data available and discoverable
- What are the data ownership and personal identifiable information issues (obstacles/solutions) that can be addressed in this context?
- What are the top two data/metadata problems you would like to solve?
  - Collaboration Culture 2.0