Project Overview

Project Timeline
- Start: Mid-Aug 2009
- Finish: October 2011
- ~2/3 complete (Feb 2011)

Budget
- Total project funding
  - DOE – $350K
  - FY09 Funding - $60K
  - FY10 Funding - $170K
  - FY11 Funding - ~$120K

Technology Gaps/Barriers
- Greater materials diversity / joining dissimilar materials
- Information Gaps in ICME Cyberinfrastructure
  - Data sufficiency
  - Incomplete structure among partners
  - Completeness and availability of repositories
- Mg Processing Model Development
  - First principles to Processing - multiscale
  - Mg-sheet forming

Partners
- Ford Motor Company (John Allison)
- Mississippi State University (Mark Horstemeyer, Tomasz Haupt)
- Canada Center for Mineral & Energy Technology (CANMET) (Kevin Boyle)
- University of Virginia (Sean Agnew)
- Metals Bank – Korea (YM Rhyim)
- ESPEI Workgroup (Liu, Wolverton, Haupt)
- Joining Materials Knowledge (Beckham)
Project Overview: Project’s place in the cyberinfrastructure

Repository
- Federated-consolidated
- Data exchange support
- ISO 10303 STEP support

Current Programs

Analysis DB
- Fills similar role to data warehouse
- Informatics storage/retrieval focused
- Analytic platform, not a re-formed repository

Informatics Effort

Informatics repository augmentation
- Property correlations
- ID feature space for complex data such as processing sequences
- When can data from distinct systems be safely fused
- Coverage and lineage anomalies
- Insufficient information for prediction (predictor diversity)
- Property definition inconsistencies
- Material definition inconsistencies
Overview: Basic Tasks & Motivation

1) As parameters are passed from scale to scale, do models adequately represent each scale and what information is lost?

2) How can the ‘model’ as a whole be verified; reliable data for verification

3) Data from different sources needed for greater variety of materials and joining

4) Does data support design?
Challenges:

• Greater diversity of materials in design problems
• Reliable information needed for new materials and model development, and verification
• Mg-alloy performance properties can conflict. Use data-driven design/informatics to address this challenge?

Objectives:

• Develop materials informatics techniques for robust model development
• Identify gaps in the fundamental information contained in the ICME knowledge infrastructure
• Augment knowledge –address gaps/barriers to developing Mg alloys
• Collaborate with current repositories on knowledge content
## Milestones / Deliverables

<table>
<thead>
<tr>
<th>Month/Year</th>
<th>Milestone / Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aug 2009</strong></td>
<td>Program start. ✓</td>
</tr>
<tr>
<td><strong>Feb 2010</strong></td>
<td>Milestone: Assess data support needs within context of workgroup research. Deliverable: Preliminary toolkit and data-storage design, assessment of data support. ✓</td>
</tr>
</tbody>
</table>
| **Sep 2010** | Milestone: Toolkit for informatics data assessment Preliminary Mg-alloy data assessment *  
Deliverables: Create ✓ and deliver toolkit for lightweight alloys. Validated tools against prototype alloy design data ✓ |
| **Jun 2011** | Milestone: Extend work to multiple repositories  
Deliverable: Apply toolkit to address global data sufficiency |
Technical Approach

Task 1
- Informatics Toolkit Design
  - Contextual assessment of data support for workgroup models development
  - Identify algorithms and storage approach for materials knowledge retention and generation

Task 2
- Informatics Toolkit Development
  - Informatics toolkit for data sufficiency/completeness, support for model development, and knowledge augmentation
  - Analytic support for cyberInfrastructure

Task 3
- Assessment of data support for model development
  - Contextual assessment of data support compatible with heterogeneous sources
  - Collaborative contact with data sources (MSSU, CanMet, MetalsBank, DataSoc)
Accomplishments & Impacts

Assessed data support for first principles development and materials repository

Data sufficiency – informatics tools to assess modeling power of finite data sets

Knowledge content of multiple information sources
  • Repository/repository
  • PI/repository
  • PI/PI

Design of prototype data schema
Knowledge retention model for workgroups

Guidance to First Principles and Cyberinfrastructure Workgroups for data completeness

Data analytics for first principles and sheet forming groups for robust models
Tools to assess repository support for information-driven design

Guidance on integrating ICME applications with repositories and cyberinfrastructure
Accomplishment #1: Analysis Database Schema

- Developed prototype schema as example for knowledge retention
- Identify key weaknesses in conversion from conventional storage
- Strongly-typed records
- Formed design constraints for information transport


Jones, DM and KF Ferris. 2010. “Generic Materials Property Data Storage and Retrieval for Alloy Material Applications,” TMS 2010
Technical Accomplishment 2: Enabling the Phase Diagram Infrastructure to MSSU Repository

First principles workgroup: Bridging first principles calculation to phase diagrams

PNNL Contribution
• Informatic effort provided Query/Response ‘glue’ for ESPEI, MSSU Cyberinfrastructure, first principles, Calphad.
• Automated procedure, retaining critical information for data
• ‘Glue’ formalism applicable to other information joining situations
Technical Accomplishment 3: Tools to Assess Data Completeness

Data Completeness

Are the existing data complete for the desired design parameters?

- Existing system
- To-be-created system

More data are required, extrapolation is too great

OK...

System Definition

Do systems have incomplete structure definitions, so one structure gives more than one property result?

Density models will be ambiguous!

Property Definition & Scale

Are properties ambiguously defined? Should they be resolved by scale or other parameters?

- Small scale
- Large scale

Here, a well-defined system has different density measurements, depending on scale.
- These must be clearly distinguished and parameterized in the database
- Different end-properties may require one, or both measurements.

Property Correlation

Are some properties correlated, so informatics can proceed without complete measures of all properties?

Correlated for a class of structures
- Not correlated

Data alone do not guarantee successful information-driven design.
Example: Predicting mechanical constants from individual data points

- Harmonic behavior
- Two different phases w/ different responses

Knowing there are two different groupings is critical to correct predictions/models/behaviors.

Not having enough information about this data may lead to –

- Linear fit
- Quadratic model for average behavior
- Averaging non-equivalent system (two phases)
- Greater property value uncertainties; validation/verification difficulties

Step-through animation sequence illustrates challenge
Property Value Deviations Resulting From Incomplete System Definition (Metals Bank)

- Illustration: % Deviation in Property Values for ‘Nominally’ Equivalent Systems
- Problem: Improperly joined information not useful for verification and validation
Tools for data assessment (Metals Bank): Contribution of system definition components to property error

**Principle:** the importance of a system definition component relates to the observed increase in error when that component is removed.

Example: for density, observed error using alloy+shape as a system definition is nearly the same as that for alloy alone.

Shape is not a crucial system definition component *for density.*
Tools to identify system definition components controlling property value deviations

- Two important signatures: “thermodynamic,” “constrained”
- Error for core mechanical properties depend on all components
Data density map for lightweight alloy section of Metals Bank
Illustrates dense and sparse data regions; information gaps
Caveat: Data by itself does not necessarily construe knowledge
Tools for data assessment:
Data coverage summary (Metals Bank/ASM)

Elongation
Strength
Hardness
Equilibrium
Corrosion
Processing/Tempering
Shape

AZ91
AZ31
AZ91D
AZ31B
AZ61*
AZ81*
AZ91
AZ91[A-C,E]
AZ92*
Other (18)
Summary of Technical Accomplishments

- Assessed data support in current data resources and current state of data capabilities (Milestone Feb 2010)
- Analysis repository schema and tested implementation. (Milestone Feb 2010)
  - Design helps resolve incomplete system or property definitions
  - Allows direct entry of "as is" data in its original form
  - Accommodates differing system and property definitions
- Translated knowledge retention schema into information transfer method for first principles workgroup (ThermoCalc-ESPEI-MSSU-First Principles)
- Repository assessment tools; assess support for information-driven validation, verification and design (Milestones Sep 2010).
  - Diagram database content; compare data population (single & combined)
  - Diagram of property value deviations as a function of system; identify system/property combinations with ambiguous property values
  - Map and diagram impact of system definition components against property deviations
  - Systematically look for associations between properties and processing variables
Collaborations

- Partners/Collaborators
  - Mississippi State University (materials repository)
  - Pennsylvania State University (ESPEI integration with repository)
  - ThermoCalc (ESPEI integration, Calphad model)
  - Materials Atlas (Iowa State)
  - CANMET (Kevin Boyle)
  - Sean Agnew (University of Virginia)
  - MetalsBank (materials repository)
  - Discussion in progress for collaborations with other repository developers (ASM)

- Technology Transfer
  - Techniques/design recommendations for software/repository integration (Pennsylvania State University collaboration)
  - Informatics tools / database design available to ICME community
  - Sharepoint site: https://spteams1.pnl.gov/sites/mat_informatics/default.aspx
Future Work

**FY2011**

- Informatics support for multiple data resources
  - Repository+Repository, Individual+Repository, Individual+Individual
  - Display ‘global’ information gaps – multi-resource
  - Physics-based data merging
  - ‘Multi-repository’ additions to system definitions

- Materials informatics Toolkit
  - Data completeness and support
  - Multiphysics deconvolution for knowledge/system definition augmentation
  - Modeling power visualization for workgroup validation and verification tasks

**Follow-on Proposal**

- Feature development to identify new or missing multiphysics
- Bridge new compositions to enable more robust design
- Granularity and multiscale hierarchy in data resources
- Robust strategies to identify/address global information gaps
Technical Back-Up Slides (limit of 5)
Technical Approach – Data Completeness

- Identify and fill gaps in ICME knowledge infrastructure required for Mg/Alloy design
  - Identify data ambiguities -- “same” measurement, different value (missing context in repository).
  - Identify data discrepancies between current repositories
  - Assure data are sufficiently complete to engage Phase II modeling – structure-processing/property relations for alloys

- Method: materials informatics techniques
  - Look for structure or processing data correlated with final properties
    - Examine impact of associating data across multiple length scales
    - Examine whether data forms must be updated to enable modeling
  - Anomaly detection
    - Property values that are not defined by specified structure-processing information
    - Ambiguities across repositories

- Collaborate with current repositories on inputs and outcomes
Technical Approach - Knowledge Augmentation

What: after identifying data completeness problems, suggest knowledge augmentation in current alloy/processing databases

- Make outcomes available for updates to repositories as needed
- Informatics tools

Why: Current data may be incomplete or ambiguous for modeling purposes

- Assess and enable overall data support for information-driven design
- Examine issues with data ambiguities
- Knowledge assessments are measured against a problem (what is to be learned)

How:

- Look for structure or processing data correlated with final properties
  - Examine impact of associating data across multiple length scales
  - Examine whether data forms must be updated to enable modeling
- Anomaly detection
  - Property values that are not defined by specified structure-processing information
  - Ambiguities across repositories
Example: Data completeness and knowledge augmentation in MSSU repository

► **Goal:** Identify compositions and processing conditions for Mg alloys consistent with improved properties
► **Data resource:** MSSU/CAVS Materials Properties repository

► **Within the repository**
  - Available: Approximately 30 Mg alloys
  - Available: Stress-strain curves for each alloy
  - *Not available*: Detailed composition or processing information
  - *Not available*: Yield strength numbers deduced from the stress-strain curve

► **In advance of detailed analysis (knowledge augmentation)**
  - Processing sequences should be added to the repository
  - (Bulk) composition should be added to the repository
  - A yield stress feature should be added to the stress-strain curves
  - Other features, based on experience and theory
  - *Try for a small set first, assessing potential association with yield stress*

**More generally, systematically identify missing information, anomalies in characterization, and associations between structure and properties…**
Example: Knowledge augmentation; where, how

It is easy to intermingle properties and systems...

Table 1. An example of alloy property data [CRC Handbook 64th edition]

<table>
<thead>
<tr>
<th>System</th>
<th>Name</th>
<th>Composition</th>
<th>Density (g/cm³)</th>
<th>Tensile Strength Kg/m.m</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alnico I</td>
<td>Fe 37-39; Al 12; Ni 20-22; Co 5</td>
<td>6.9</td>
<td>2.9</td>
<td>Cast, isotropic</td>
</tr>
<tr>
<td>2</td>
<td>Alnico II</td>
<td>Fe 52.5; Al 12; Ni 24-26; Cu 6; Co 12.5</td>
<td>7.1</td>
<td>2.1</td>
<td>Cast, isotropic</td>
</tr>
<tr>
<td>2</td>
<td>Alnico II</td>
<td>Fe 52.5; Al 12; Ni 24-26; Cu 6; Co 12.5</td>
<td>7.1</td>
<td>45.7</td>
<td>Sintered, isotropic</td>
</tr>
<tr>
<td>3</td>
<td>Alnico III</td>
<td>Fe 59-61; Al 12; Ni 24-26; Cu 3</td>
<td>6.9</td>
<td>8.5</td>
<td>Cast, isotropic</td>
</tr>
<tr>
<td>4</td>
<td>Alnico IV</td>
<td>Fe 55-56; Al 12; Ni 27-28; Co 5</td>
<td>7.0</td>
<td>6.3</td>
<td>Cast, isotropic</td>
</tr>
<tr>
<td>5</td>
<td>Hypothetical Calculation</td>
<td>Fe 40, Ni 24, Co 5</td>
<td>9.0</td>
<td>60</td>
<td>Repeated fcc unit cell</td>
</tr>
</tbody>
</table>

Inconsistent or ambiguous properties (name, form)

- Difficult to compare across systems
- Merged systems (e.g. Alnico II)
- Merged properties (e.g. Form)

*Augmentation in this case: can rectify ambiguities*
More Detailed Summary - Toolkit Accomplishments

- Repository assessment tools; assess support for information-driven validation, verification and design
  - Diagram of database content
  - Ability to survey data population and compare repository content
  - Locate information gaps; identify/prioritize maximum leverage points
  - Identify data conditions conducive to robust validation of model predictions/experimental measurements
  - Diagram of property value deviations as a function of system
  - Identify repository system/property combinations with ambiguous property values (anomalies/system and measurement uncertainties)
  - Indicate which systems/property combinations are most susceptible to large property variations
  - Map and diagram impact of system definition components against property deviations
  - Systematically look for associations between properties and processing variables