

NIST Materials Data Informatics Efforts

C. E. Campbell, U.R. Kattner

Material Measurement Laboratory, Material Science and Engineering Division

A.A. Dima and S. Youssef

Information Technology Laboratory, Software and Systems Division

L. Bartolo

Kent State University

April 29, 2014



NIST Collaborators

Information Systems Group

Mary Brady, Alden Dima, Y.-S. Li-Baboud, Benjamin Long, , Philippe Dessauw, Pierre Savonitto , Sharief Youssef

Thermodynamics and Kinetics Group

Carrie Campbell, Ursula Kattner, Ben Burton, Chandler Becker, Francesca Tavazza, Zach. Trautt , Dan Wheeler, Andrew. Reid. Shenyen Li

Thermodynamic Research Group

Stephanie Hooker, Robert. Chirico, Michel Frenkel, Vladimir Diky, Chirs Munzy.

Cell Systems Science Group TN Bhat and John Elliott

Kent State, Laura Bartolo



NIST Data Efforts

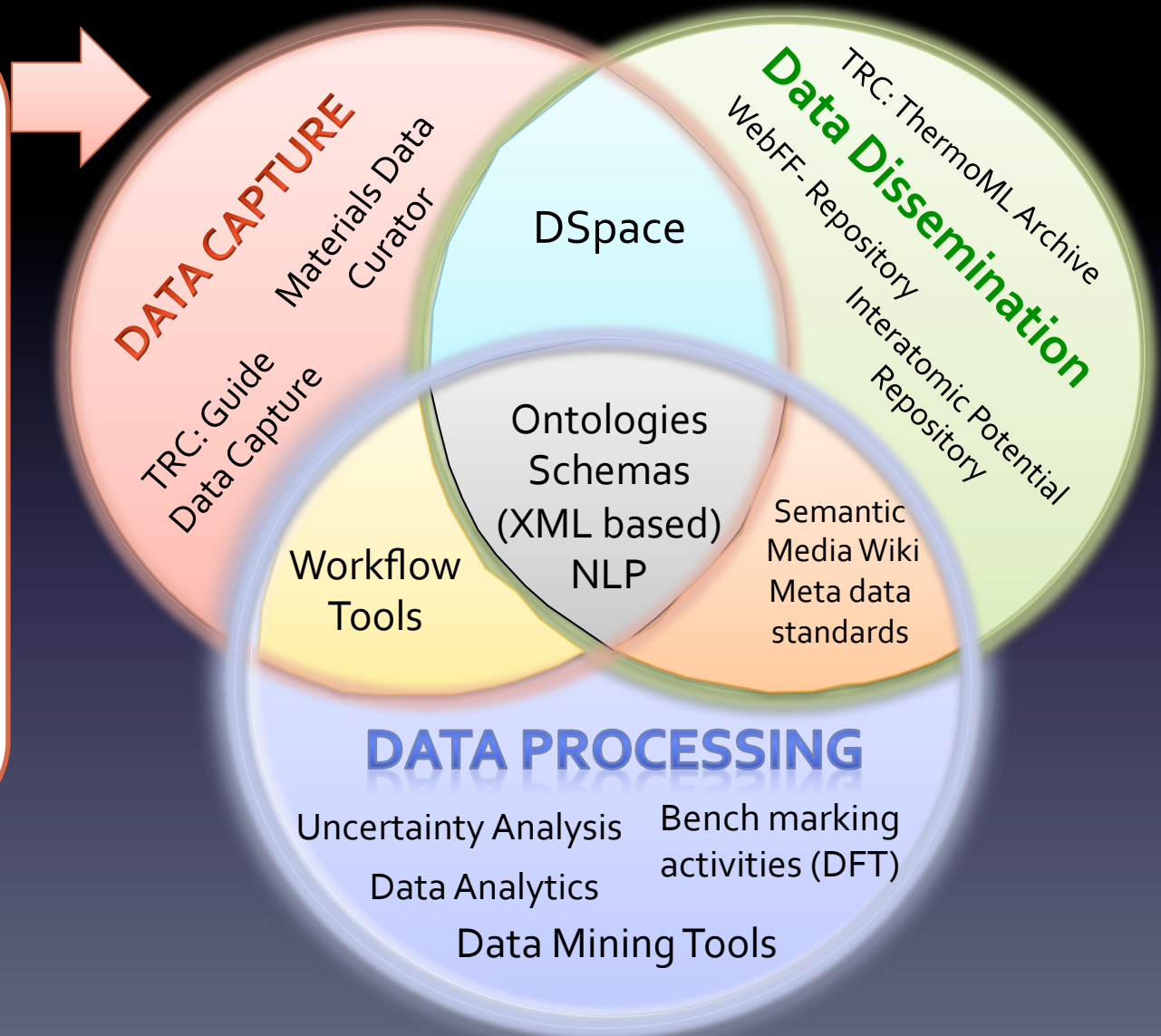
Collaborations

ASM International:
Structural Data
Demonstration Project

DOE/EERE Kinetics of
Cast Mg Alloys

Journals collaboration

- IMMI
- *Others under discussion*



Phase-Based Property Database

➤ Material Property Database Exist




Everything
Material.



AFLOWLIB.ORG

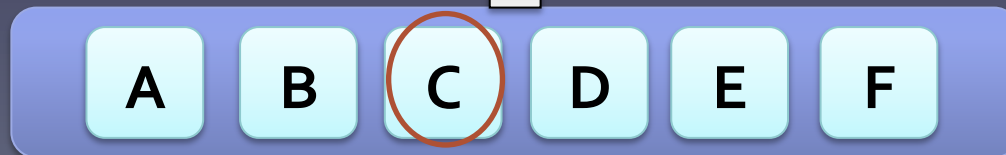
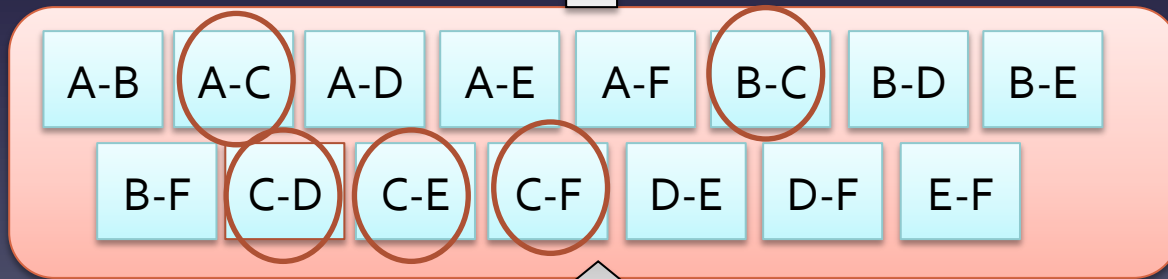
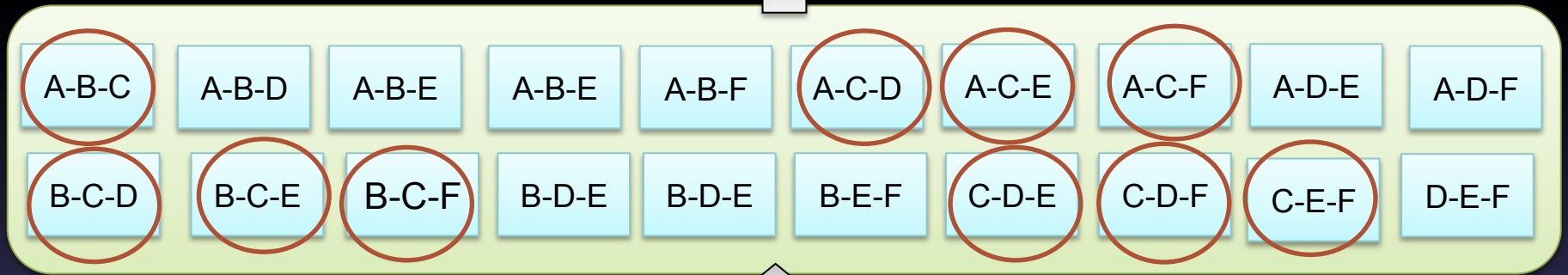
OQMD:
An Open Quantum Materials Database

Generally, focused on engineering/design specs or first-principle calculations results.

- 
- Focus on phase-based properties that are needed to describe the composition, temperature, and pressure functions of a phase.
 - Unary, binary and ternary data are primary focus.
 - Multicomponent data are needed for validation

Data Dependencies

A-B-C-D-E-F
(ABCDEF.TDB)



Examples of CALPHAD Data

For each assessment: Evaluated data file (e.g. POP, DOP)

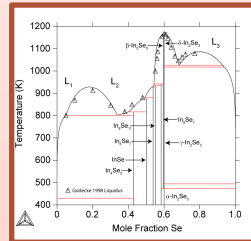
Functional descriptions for phase quantity (e.g. TDB)

- Emphasis on binary and ternary data to predict multicomponent properties
- Data can be experimental or computational.

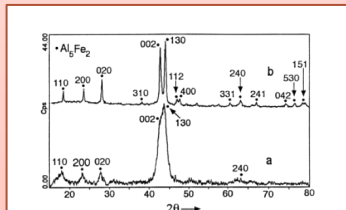
1-D (Points)

Melting Temperatures

Critical Temperatures
(Phase Changes)



Lattice Parameters



Heat of Formations

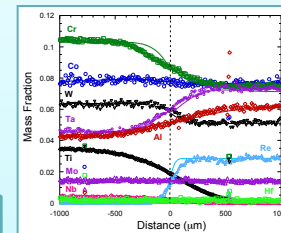
Phase fractions
and compositions

Tracer
Diffusivities

Activation energies

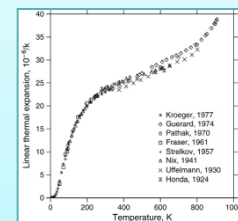
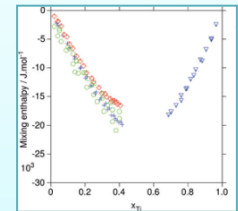
2-D (Lines)

Composition
Profiles



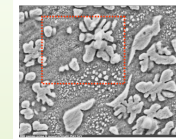
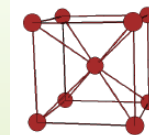
Heat Capacities

Enthalpies of mixing



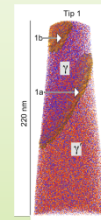
3-D

Crystal structures



Micrographs/Morphologies

3-D Atom probe Tomography



Current Assessment Process

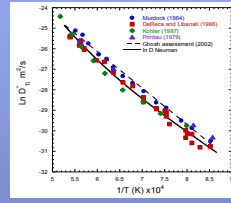


Search data

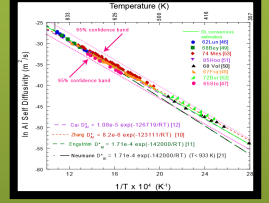
Digitize Data



Plot data



Reference values



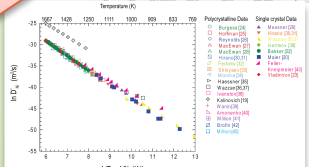
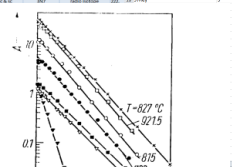
Future

Computational and Experimental Databases and Repositories



Search infrastructure

Element	Phase	Diff. Coeff. (m²/s)	Diff. Coeff. (cm²/s)	Temp. (K)	Ref.	Material	Experimental method/Measurement	Address
Ni	Ni	5.27	2793	0.40	1490-1553	33	Diff. Exp.	Wahlman, Wilson, Ward & Mital, 1976
Ni	Ni	8.4	286.9	0.34	1575-1688	3	Diff. Exp.	Przybyl
Ni	Ni	3.36	2943	0.33	1420-1670	33	Diff. Exp.	Wahlman
Ni	Ni	5.13	2913	0.34	1510-1610	33	Diff. Exp.	Wahlman
Ni	Ni	3.3	293.4	0.30	1633-1635	4	Diff. Exp.	Wahlman
Ni	Ni	1.7	288.7	0.32	1575-1575	3	Diff. Exp.	Wahlman
Ni	Ni	3.00	296.2	0.40	1480-1576	3	Diff. Exp.	Wahlman
Ni	Ni	1.9207	284.7	0.40	1480-1576	3	Diff. Exp.	Wahlman

CALPHAD user needs data for A-B-C system,

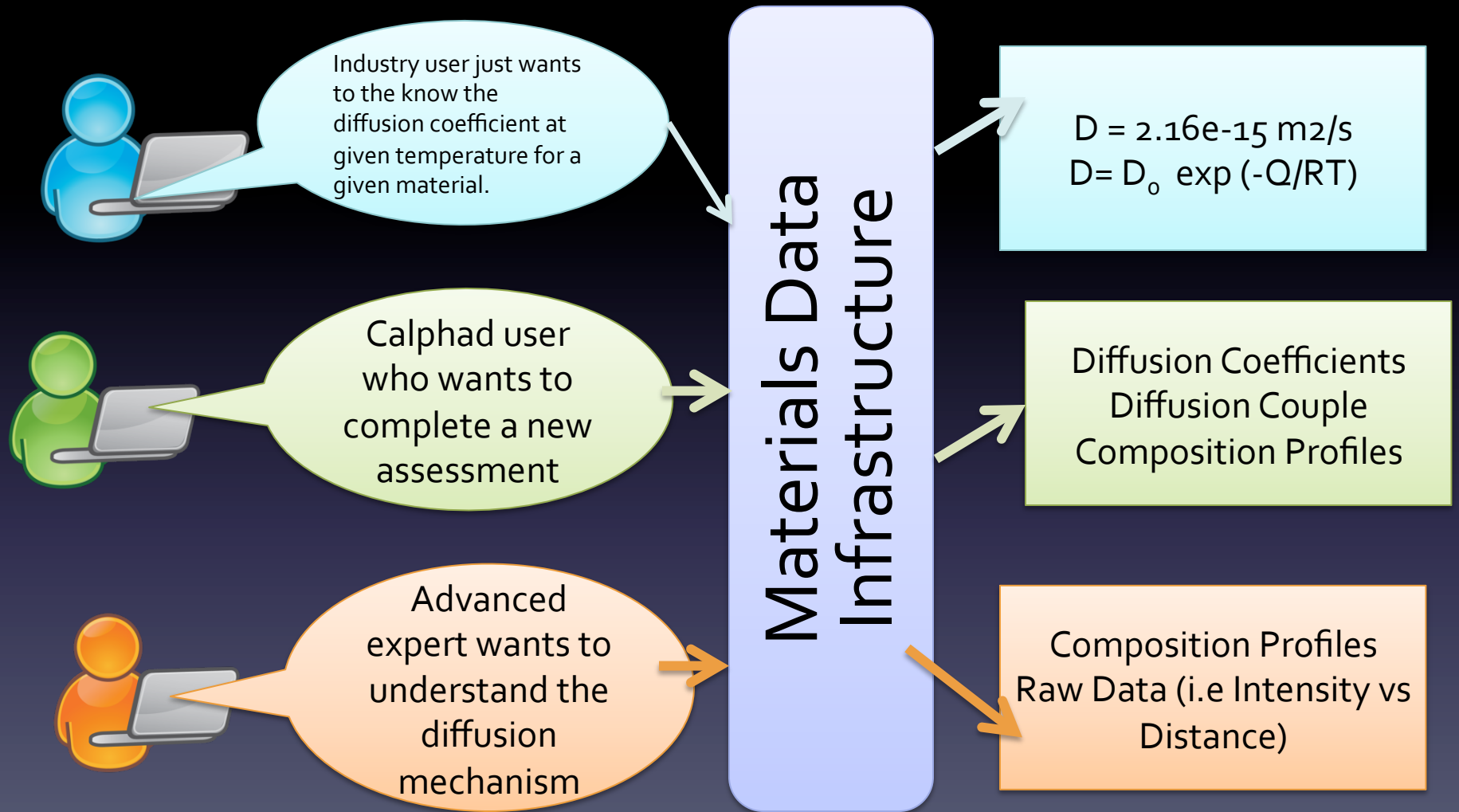
Interfaces for input of computational and experiment data

Federated System

Search returns available data And points out missing data

Fig. 1. Gaussian penetration plots for self-diffusion in nickel single crystals. Value of one division on the abscissa: □ 18, × 5.5, ○ 4.3, ● 3.7, ■ 3.4, ▽ 0.056, ▼ 0.19 × 10⁻¹¹ cm²

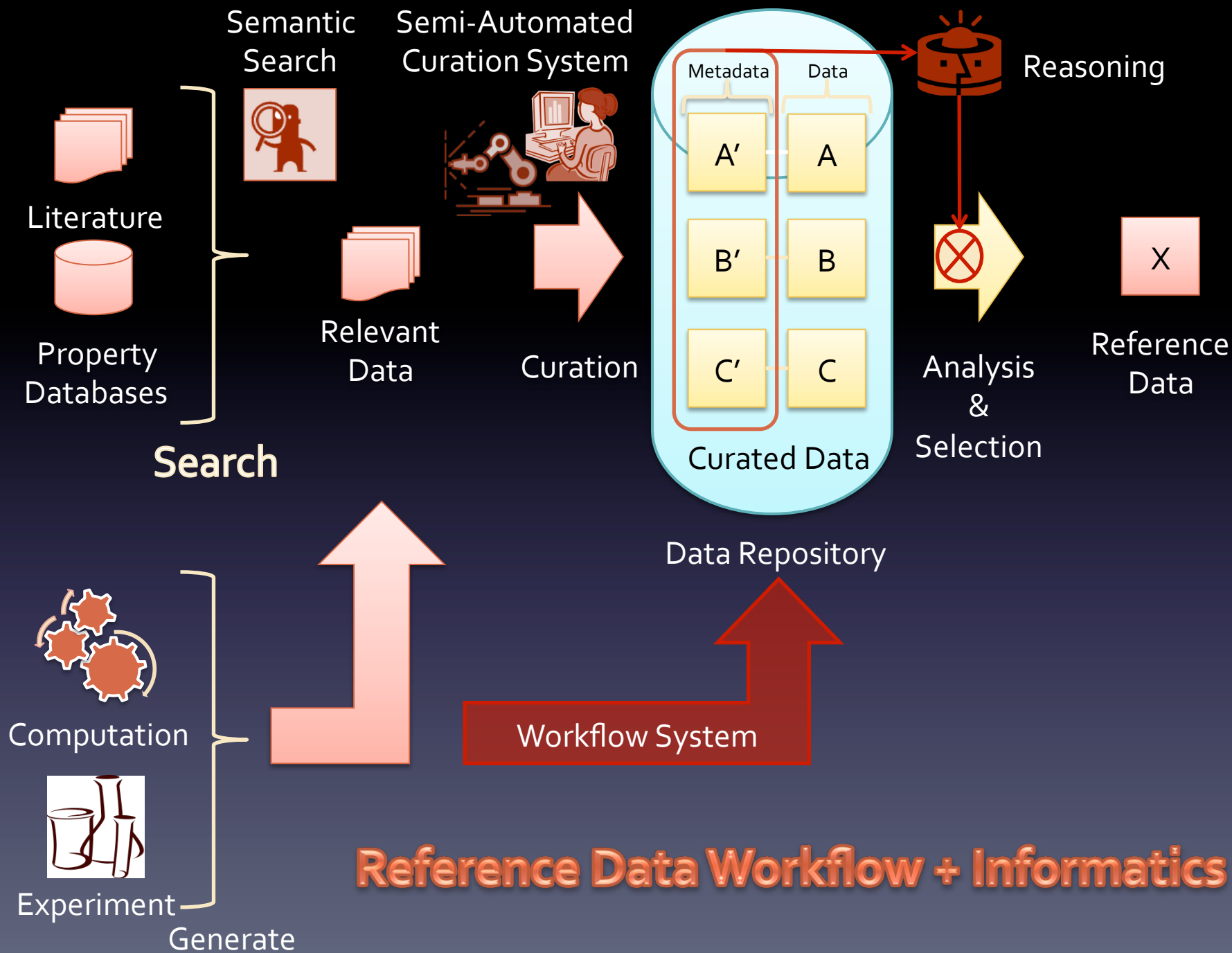
EXAMPLE OF DIFFERENT TYPES OF DATA USERS: DIFFUSION DATA



Data are diverse

Data are semi-structured

Need complete data sets

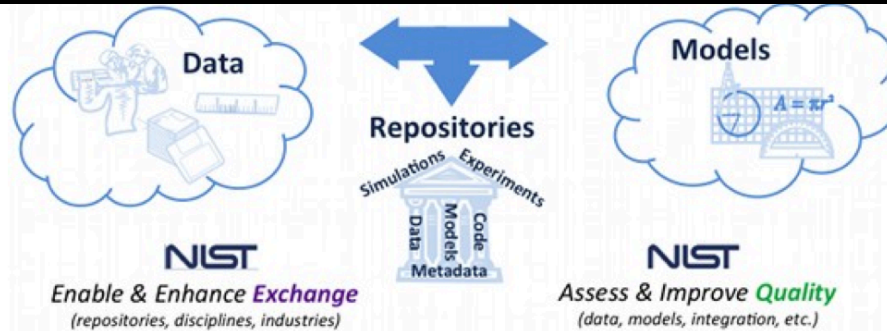


Two Biggest Problems

- **Need to collect data and organize**
 - Data are diverse
 - Data are semi-structured
 - Data are incomplete
- **Need to make the data useful to a wide variety of users.**

D-SPACE CUSTOMIZED FILE REPOSITORY

nist.matdl.org



[Login](#)

[NIST Repositories](#) → [Community List](#)

NIST Repositories

The National Institute of Standards and Technology is establishing essential data exchange protocols and mechanisms for widespread adoption to ensure quality materials data and models and to foster data sharing and reuse.

- **[NIST Computational File Repository](#)**
 - [CALPHAD Assessments](#)
 - [First Principles Phase Stability \(FPPS\) Files](#)
 - [Interatomic Potentials](#)
- **[NIST Experimental Data Repository](#)**
 - [Diffusion Data](#)
 - [Phase Equilibria](#)
 - [Texture Data](#)
- **[NIST Structural Materials Data Demonstration Project ASM](#)**
 - [6061 Aluminum Alloys: Diffusion Data](#)
 - [6061 Aluminum Alloys: Handbooks, Journal Articles, and Technical Papers](#)
 - [6061 Aluminum Alloys: Mechanical Properties](#)
 - [6061 Aluminum Alloys: Phases, Phase Diagrams](#)
- **[TMS Springer Integrating Materials and Manufacturing Innovation \(IMMI\)](#)**
 - [General Articles](#)
 - [Thematic Series on Three-Dimensional Materials Science](#)

Search NIST Repositories

Computational
CALPHAD
First Principles

Browse

All of NIST Repositories

Experimental
Phase Equilibria
Diffusion

[Login](#)
[Register](#)

Sample Entry

(repositories, disciplines, industries) (data, models, integration, etc.)
NIST File Repositories → NIST Data File Repositories → CALPHAD Assessments

CALPHAD Assessments

Search D Space

Browse by

- By Issue Date
- Authors
- Titles
- Subjects

Search within this collection

Submit a new item to this collection

Recent Submissions

[Al-Cr-Ni Diffusion Mobilities](#)
Campbell, C.E. (2013-02-11)
This work presents the assessment of the diffusion mobilities in the gamma prime and B2 phases in the Ni-Al-Cr system. Available experimental data are compared with the CALPHAD model results.

[Ni-Al-Cr system Thermodynamic](#)
Dupin, N.; Ansara, I.; Sundman, B. (2010-11-11)
A re-assessment of the thermodynamic energy function for the gamma prime experimental liquidus temperature.

[Ag-Al Functional Description](#)
Du, Zeting; Jing, Zhan-Peng; Li, J. (2010-11-11)
The energy expressions for gamma prime phases, are established by comparing experimental energy and the ...

NIST File Repositories → NIST Data File Repositories → CALPHAD Assessments → View Item

Data Citation:
Al-Cr-Ni Diffusion Mobilities in Gamma Prime and B2
Campbell, C.E.
<http://hdl.handle.net/11115/51> → **Digital Identifier**
Affiliation: Metallurgy Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-6555, USA
Contact Email: carelyn.campbell@nist.gov

Publication Citation:
Campbell, C.E. "Assessment of the diffusion mobilities in the gamma prime and B2 phases in the Ni-Al-Cr system," *Acta Mater.* 2008;56:4277.
<http://dx.doi.org/10.1016/j.actamat.2008.04.051> → **Related Work**

Related Work:
Dupin, N., Ansara, I., Sundman, B. "Thermodynamic Re-Assessment of the Ternary System Al-Cr-Ni," *CALPHAD* 2001;25:279. Publication: [http://dx.doi.org/10.1016/S0364-5916\(01\)00049-9](http://dx.doi.org/10.1016/S0364-5916(01)00049-9)
<http://hdl.handle.net/11115/10088> → **Similar Work**

Similar Work:
Zhang, L., Du, Y., Chen, Q., Steinbach, I. "Atomic mobilities and diffusivities in the fcc, L12 and B2 phases of the Ni-Al system," *International Journal of Materials Research*, 2010;146:1. <http://dx.doi.org/10.1464/110428>

Abstract:
This work presents the assessment of the diffusion mobilities in both the γ' (Ni₃Al-L12) and B2 phases in the Ni-Al-Cr system utilizing the phenomenological model developed by Helander and Available experimental tracer diffusivity, interdiffusion coefficients and activation energies evaluated and then used to optimize the composition- and temperature-dependent diffusion model. For both the B2 and γ' phases, the assessed diffusion mobility descriptions reproduce the Arrhenius temperature dependence for the Ni, Al and Cr tracer diffusivities and interdiffusion coefficient assessment reproduces the strong composition dependence of the diffusivities in the B2 observed experimentally. The measured composition dependences of the diffusivities in the γ' phase also replicated by the present mobility descriptions. The assessed mobility descriptions are validated by comparing calculated and measured composition profiles for a variety of Ni-Al and Ni-Al-Cr diffusion couples, including B2/B2, γ (fcc)/ γ' and γ /B2 couples.

Files in this item


	Name: exp-b2.zip Size: 9.374Kb Format: application/zip Description: Experimental data for NiAl B2 phase	View/Open
	Name: exp-ni3al.zip Size: 9.619Kb Format: application/zip Description: Experimental diffusion data files for NiAl	View/Open
	Name: alcrni-mob-NIST-0 ... Size: 57.23Kb Format: application/tdb Description: Diffusion mobility description for Ni-Al-Cr using N. Dupin thermodynamics (CALPHAD 2001)	View/Open
	Name: Re-assessment of the ... Size: 237.1Kb Format: PDF Description: Explanation of revised diffusion mobility descriptions	View/Open

The following license files are available:

- [Creative Commons](#) → **Offer licenses with attribution 3.0**

This item appears in the following collections:

- [CALPHAD Assessments](#)



https://materialsdata.nist.gov/dspace/xmlui/



NIST Repositories → Community List

NIST Repositories

The National Institute of Standards and Technology is establishing essential data exchange protocols and mechanisms for widespread adoption to ensure quality materials data and models and to foster data sharing and reuse.

- **Computational File Repository**
 - [Atomistics Simulations](#)
 - [CALPHAD Assessments](#)
 - [First Principles Simulations](#)
 - [Other Computational Methods](#)
- **Experimental Data Repository**
 - [Diffusion Data](#)
 - [Mechanical Properties](#)
 - [Other Experimental Data](#)
 - [Phase Equilibria and Thermodynamic Data](#)
- **NIST/DOE-EERE Advanced Automotive Cast Magnesium Alloys**
 - [A systematic multiscale modeling and experimental approach to protect grain boundaries in magnesium alloys from corrosion](#)
 - [Corrosivity and Passivity of Metastable Mg Alloys](#)
 - [Dealloying, Microstructure and the Corrosion/Protection of Cast Magnesium Alloys](#)
 - [High-Throughput Study of Diffusion and Phase Transformation Kinetics of Mg-Based Systems](#)
 - [In-situ Investigation of Microstructural Evolution During Solidification and Heat-Treatment in a Die-Cast Magnesium Alloy](#)
 - [Phase Transformation Kinetics and Alloy Microsegregation in High Pressure Die Cast Magnesium Alloys](#)
- **NIST Thermodynamics and Kinetics Test Space**

Search NIST Repositories

[Advanced Search](#)

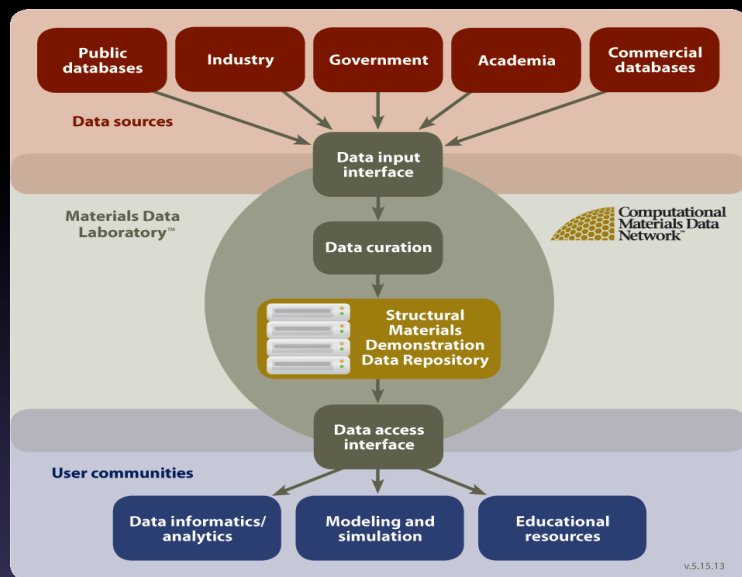
Browse

All of NIST Repositories
[Communities & Collections](#)
[Subjects](#)
[Titles](#)
[Authors](#)

My Account

[Login](#)

To be added soon:
Huesler First Principle Simulations (DFT Magnetic Properties)



March 2014: Phase 1 release.
June 2014: Phase 2 release.
Dec 2014: Project Completion

Goal: Establish well-pedigreed and curated demonstration datasets for non-proprietary metallic structural materials data over all length scales.

NIST's role

- Provide data schemas and meta-data formats for diffusion and phase equilibria data.
- Provide sample diffusion and phase equilibria data for the Al-Mg-Si system.
- Use expanded TRC Guided Data Capture program with available binary and ternary phase equilibria literature
- Expand use and implementation of DSpace Repository
- Link with developing ontology and semantic web tools



Start of process

1. Experiment Planning (Article Authors)

2. Article Preparation and Submission (Article Authors)



3. Journals (Editors)

End

5. Decision

End

ThermoLit

Journal Support Websites

NIST Literature Report

NIST A

4. Traditional Peer Review

Approve (not "Accept")

6a. In-House Data Capture (Student Associates)

NIST Data Report

6b. Guided Data Capture

6c. ThermoData Engine

NIST B

NIST/TRC SOURCE Database

7. Journals (Editors)

7a. Revisions (Authors)

End

8. Final Decision

Accept Publish

After publication

9. ThermoML Archive of published experimental data

NIST C

10. Data Users

End of process

Thermodynamic Property

Property Type

Material

- Elements
- Sample

Sample Preparation

Reaction

First Principles Calculation

- Package/method
- Potential
- Approximation
- K-points

Approach

- First Principles Calculation
- Experiment

Experiment

- Method
- Calibration/Reference Material
- Operating Conditions
- Ambient Conditions
- Characterizing Method

Identification Method

(D) Sample Description

1	Sample series Composition Composition unit Composition error	<input checked="" type="checkbox"/> Series <input type="checkbox"/> Single 0.95Al-xMg-(0.05-x)Si <input checked="" type="checkbox"/> Mass fraction <input type="checkbox"/> Mole fraction not reported	xMg in Table
2	Form Size/Dimension	<input checked="" type="checkbox"/> Chunk <input type="checkbox"/> Foil <input type="checkbox"/> Powder 0.3 g	
3	Sample preparation	<input checked="" type="checkbox"/> Casting <input type="checkbox"/> Sintering <input type="checkbox"/> Mechanical Alloying <input type="checkbox"/> Rapid cooling	
4	Mechanical history	<input type="checkbox"/> Rolling <input type="checkbox"/> Extrusion <input type="checkbox"/> Milling <input checked="" type="checkbox"/> Forming	
5	Thermal history	<input checked="" type="checkbox"/> Annealing	
6	Sample analysis	<input type="checkbox"/> EPMA <input type="checkbox"/> EDS <input type="checkbox"/> WDS <input type="checkbox"/> AES <input type="checkbox"/> XRF <input type="checkbox"/> Chemical	

Casting

1. Melted in high frequency furnace under argon gas
2. Casted into 8 mm diameter 100 mm length copper mold

Forming

1. Hammered to 6.5 mm diameter

Annealing

Container	not reported
Atmosphere	not reported
Temperature, Time	550 °C, 5 days
Quench	not reported

Sample Preparation Description

Facilitating Materials Ontology Development via NLP and Machine Learning

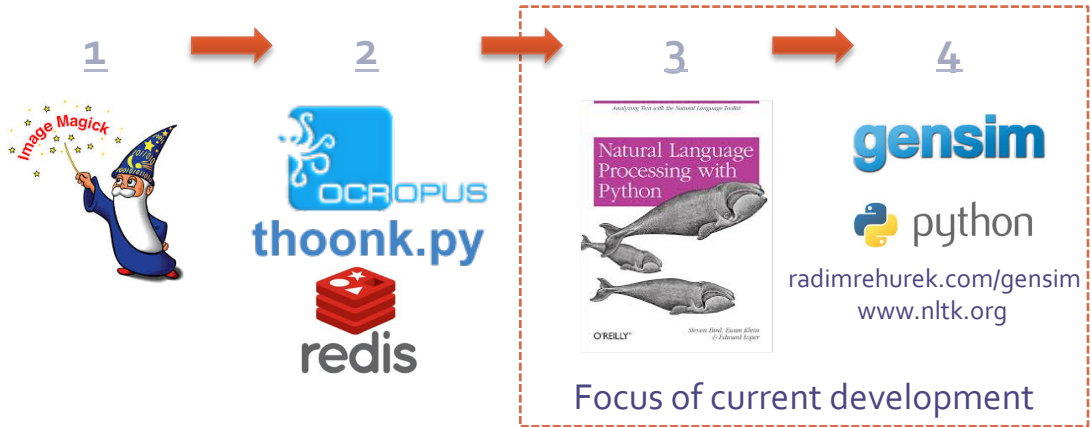
Starting with a corpus of 5893 PDFs



2009, 2010, 2011, 2012

NIST Diffusion Data Center

www.matscitech.org www.tms.org patapsco.nist.gov/diffusion



5893 PDFs - 5.8 GB

63K pages }
3.8M lines } 53GB

3.4M lines extracted

20M tokens

temperatur	69079
materi	53673
diffus	49743
process	48594
result	43713
surfac	33899

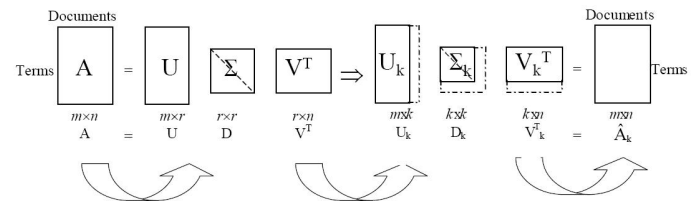
Identify key stems by frequency³

Extract text^{1,2}

Generate concordances for key stems³

tivity method. **Lattice diffusion coefficients** and grain boundary diffusion coefficients were measured. Plots of the data for all the alloys, the **diffusion coefficients** at temperatures 1150 'C and higher, were only accepted. From these **coefficients**, the parameters of temperature of which were also determined at those alloys the **self-diffusion coefficients** of which were also determined at **intercrystallite diffusion coefficients**, the following relation coefficient place. **Values of bulk diffusion coefficients** at temperatures below 1150 'C

Use Latent Semantic Indexing to group similar concordance entries⁴

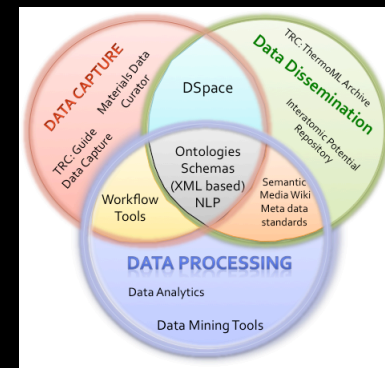


<http://liqianguo.wordpress.com/2011/06/09/latent-semantic-analysis/>

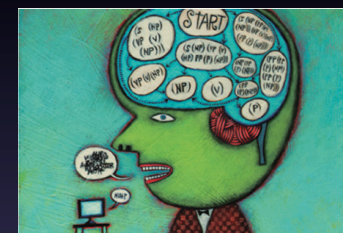
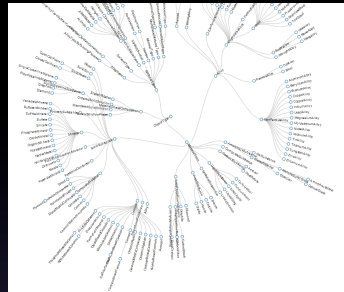
STOP

**Go to Alden and Youssef
Presentations**

Integration Tools: How to Gain Knowledge from Unstructured Data



- Ontologies
- Natural Language Processing
- Semantic Web/ Semantic Media Wiki



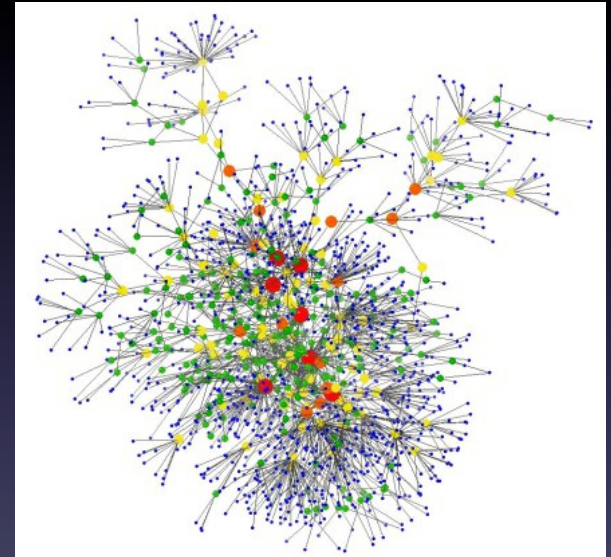
What is an Ontology and Why Build One?

- An ontology defines concepts and the relationships between concepts for a given community.

Purpose:

- To share a common understanding of the structure of information among people or software tools.
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

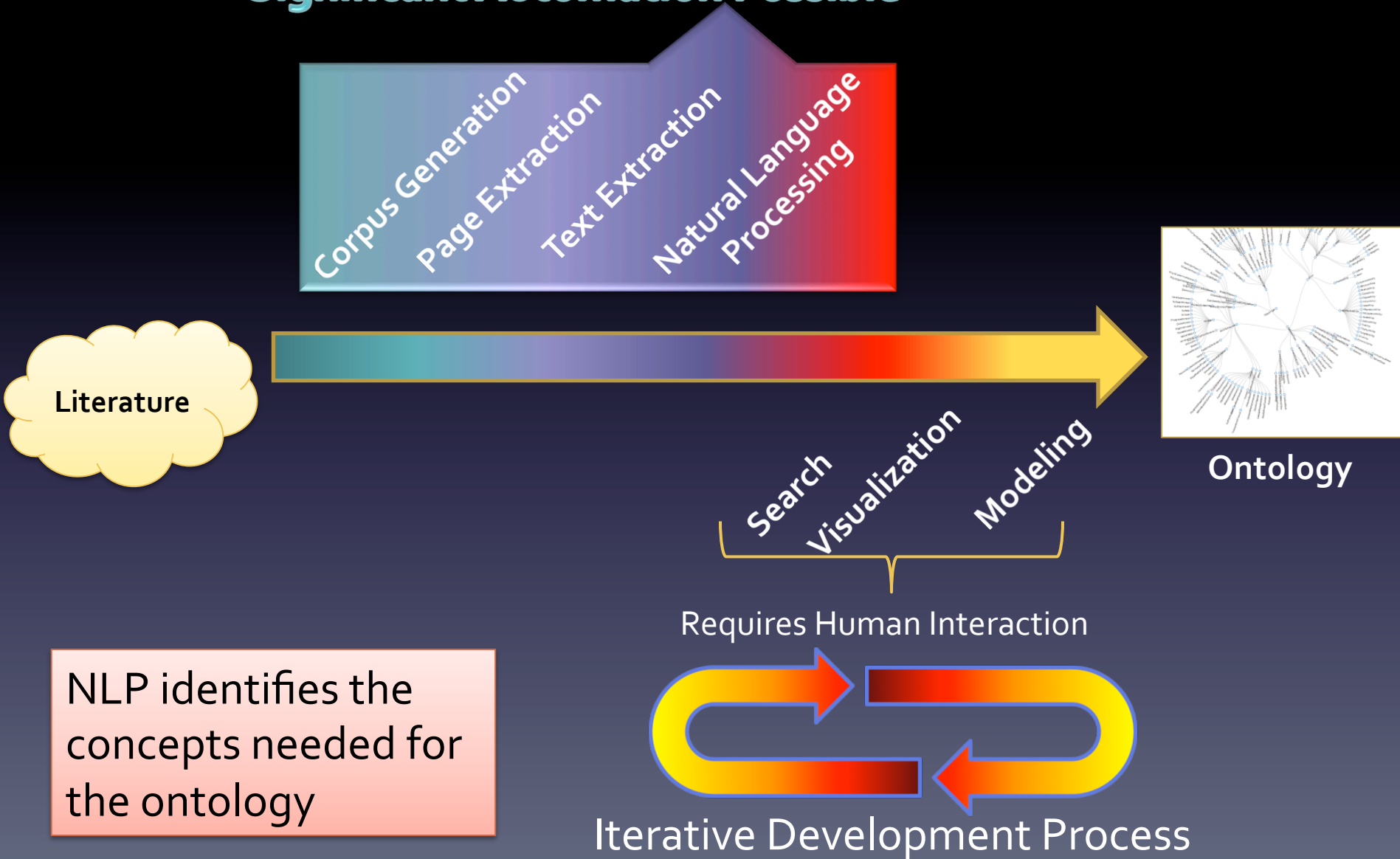
<http://www.ksl.stanford.edu/>Ontology 101



An ontology model is dynamic. The domain knowledge changes.

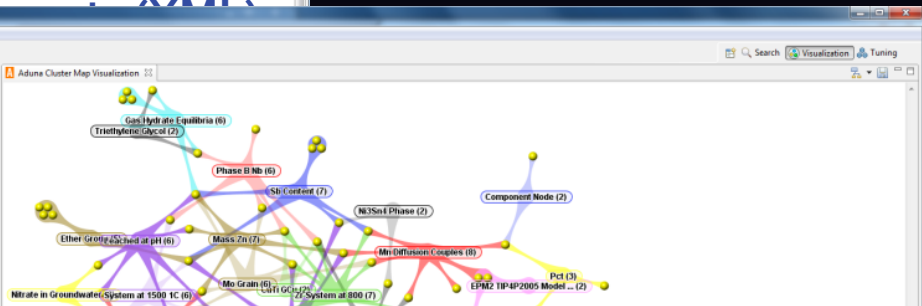
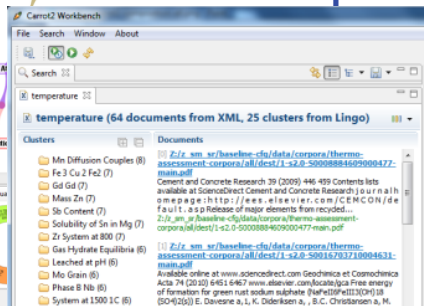
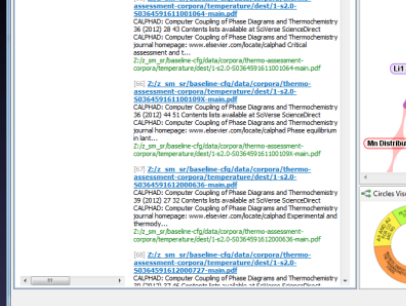
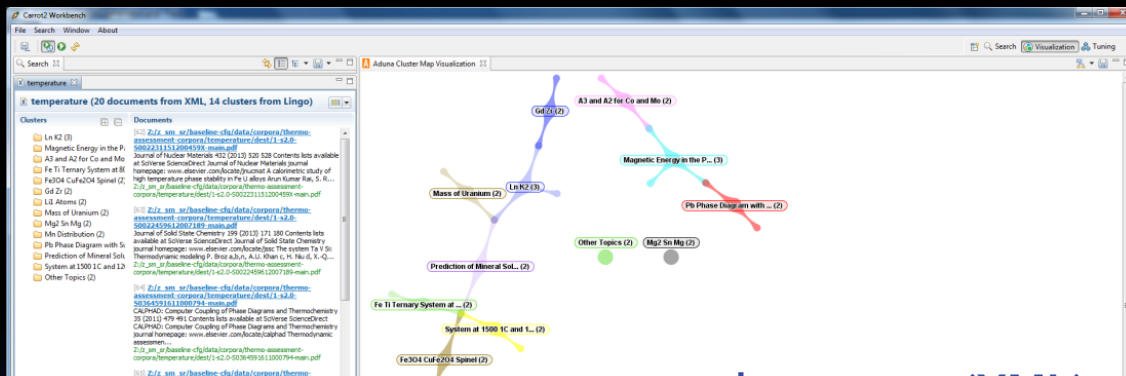
Informatics for Ontology Acceleration

Significant Automation Possible



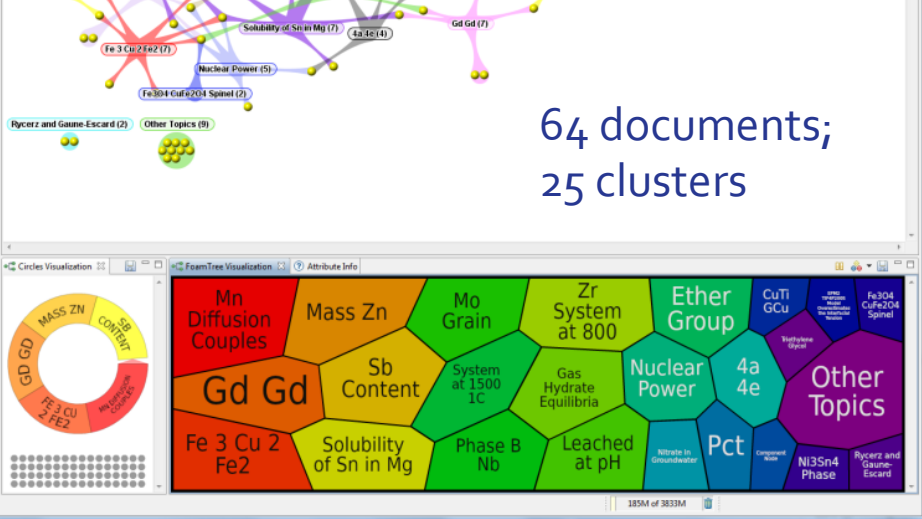
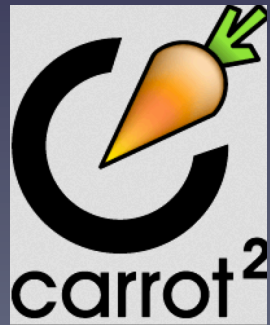
Ontology Acceleration – Clustering

Search term:
"Temperature"



64 documents;
25 clusters

Carrot²



Informatics Task (research): Research Ag Alloys for CALPHAD Assessment

Informatics Task

Research Ag Alloys
+ CALPHAD
Assessment

Traditional
Approach

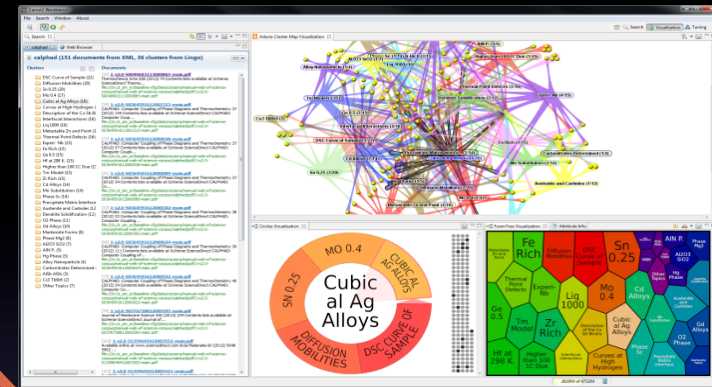
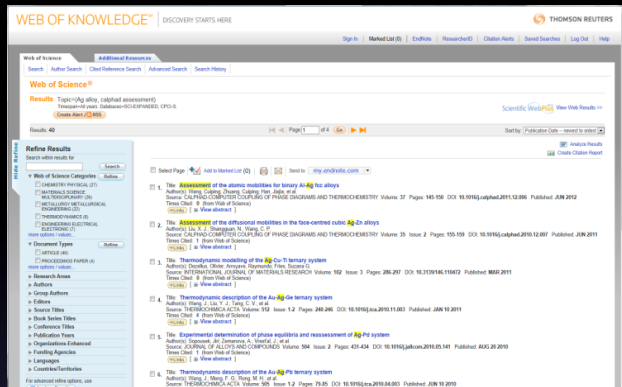
Web of Science Search
(Ag alloy, CALPHAD
assessment)

- Term-only criteria
- Relies on pre-processed, structured corpora

Our
Approach

Carrot2 Search
(Ag alloy, CALPHAD
assessment)

- Domain-specific representation
- Tunable clustering
- Can process less structured, noisy corpora
- Thematic visualization & browsing



Semantic Medline (SM)

1. What is it?

1. SM is a semantic layer on top of PubMed
2. SM supports semantic applications of semantic search, semantic visualization, and automatic summarization



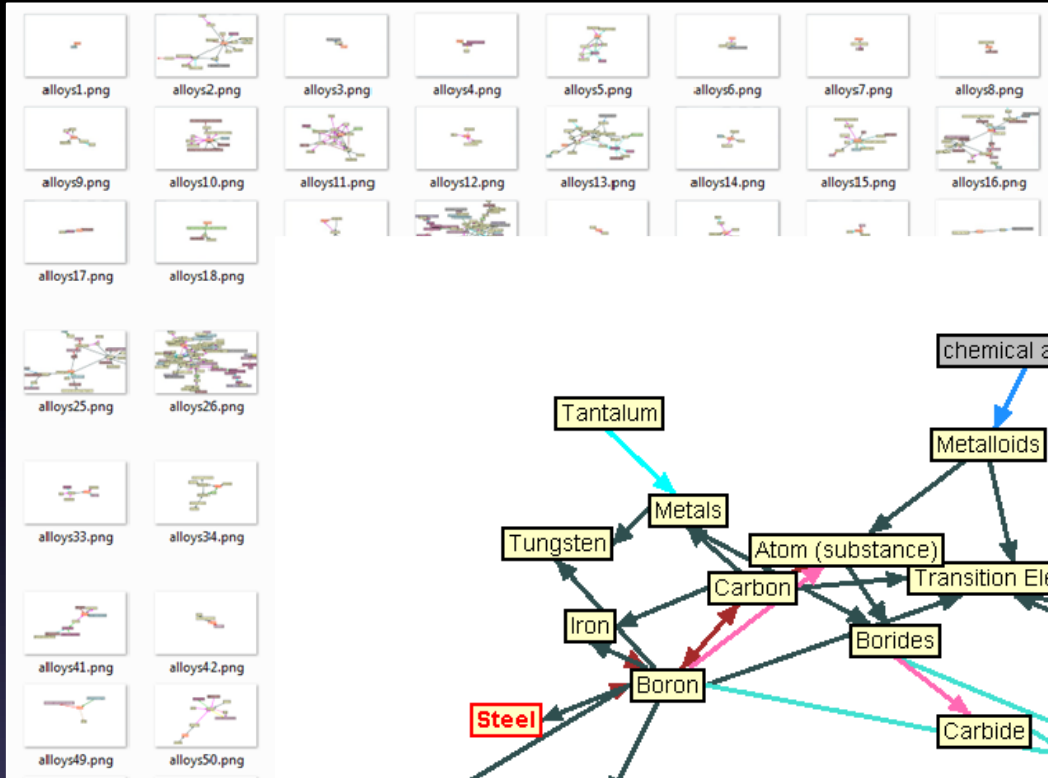
2. Collaborators: NIH/NLM

3. How does it work?

1. Supported by DB of corpus-based semantic information
2. Uses search to access semantic-info which supports semantic application operations

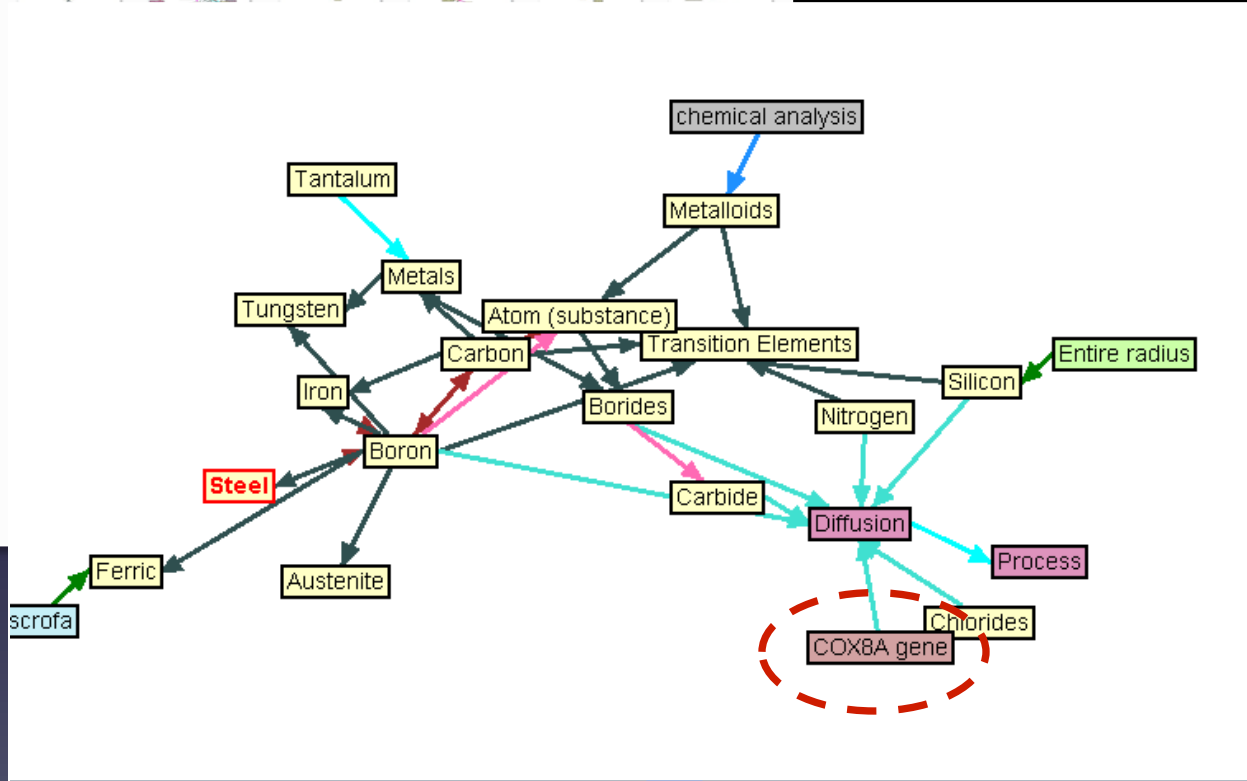
4. **GOAL:** *Retarget* SM from **biomedical** domain to **MS** domain “Semantic Matline”

Ontology Acceleration – Semantic Medline

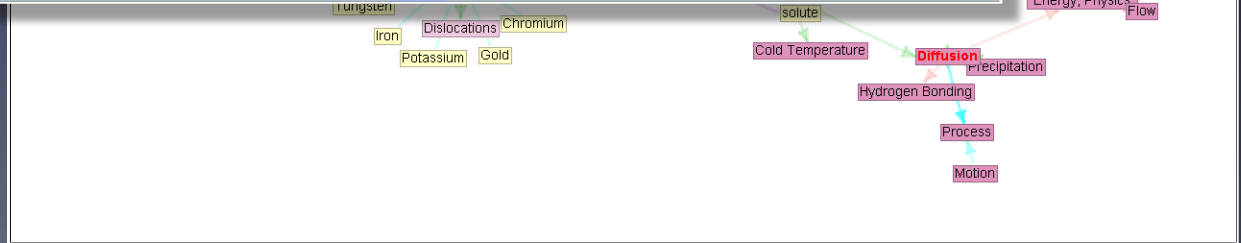


Iterative Process for Ontology Evolution

1. Start with Existing Ontology

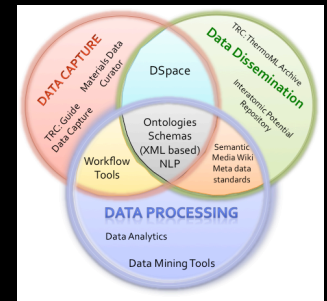


pus
y Coverage





Semantic Web



- Creates a “Web of Data” that can be machined processed
- Provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.
- Elements of the Semantic Web
 - XML: elemental syntax for content structure within documents
 - Examples (ThermoML, MatML, UnitsML)
 - XML Based- Schemas : a language for providing and restricting the structure and content of elements contained within XML documents.
 - RDF: a simple language for expressing data models



Linking Semantic Media Wiki to D-Space: Better Searching



Set \$wgLogo to the URL path to your own logo image.

Main page
Community portal
Current events
Recent changes
Help

Tools
Experiment entering
Ontology visualizer

Page Discussion

Thermodynamics

Gibbs energy
Enthalpy
Entropy
Heat capacity
Phase equilibrium
Phase transition
Melting Temperature

Beginning an effort to define critical concepts using controlled natural language

Diffusion

Tracer Diffusivity
Intrinsic Diffusivity
Chemical Diffusivity

Page Discussion

Search

Tracer Diffusivity

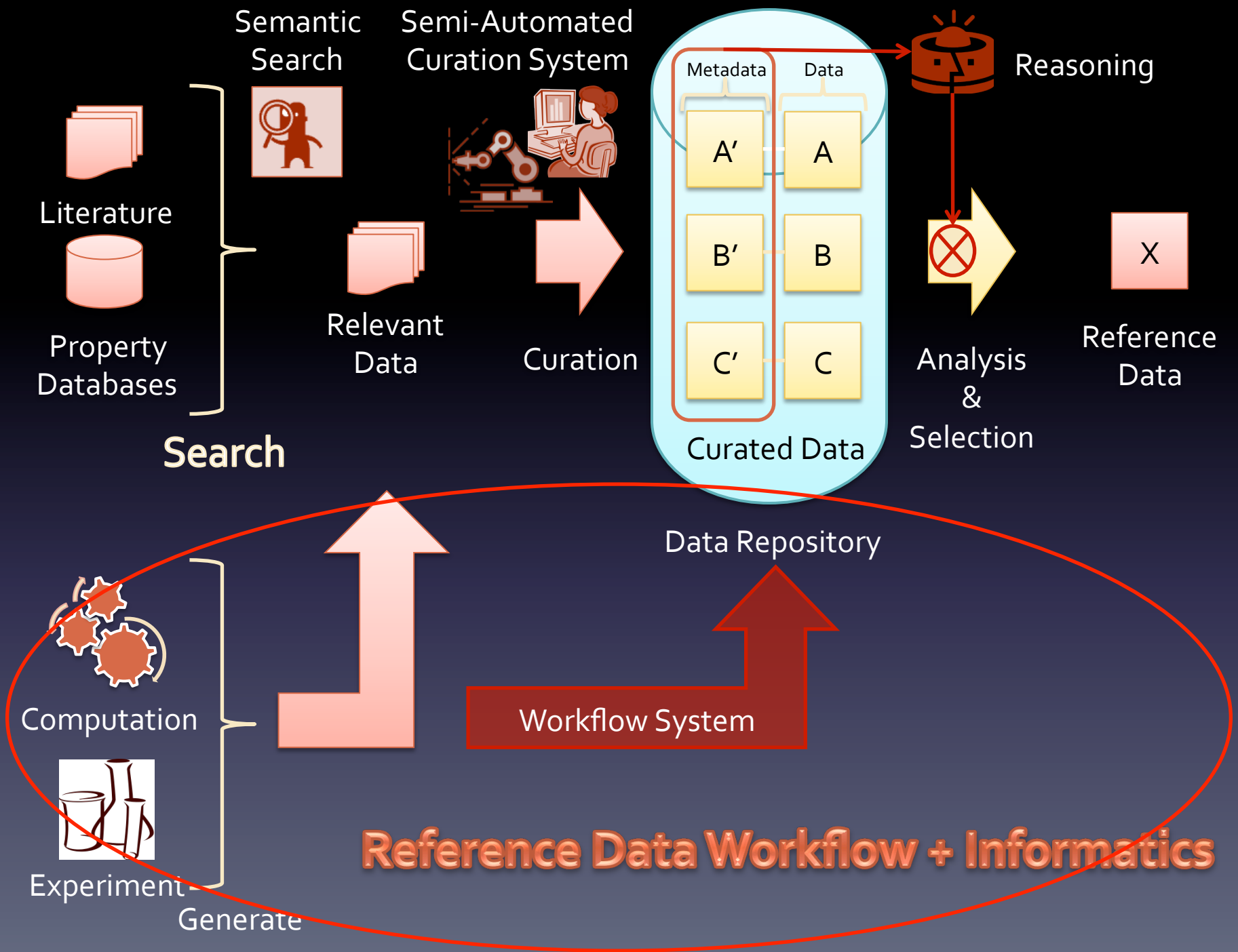
Tracer diffusivity is the migration of a tagged atom (i) through a material of which i is a component.

Some tracer diffusivity is measured by using migration of a dilute concentration of a radioactive isotope of the tracer atom into a homogeneous material.

Tracer diffusivity equals the self-diffusivity multiplied by a correlation factor.

Uncorrelated jump imply the correlation factor equals 1.

Reference: Kizilyalli, M., et al. (1999). "Definitions of Terms for Diffusion in the Solid State (IUPAC Recommendations 1999)." Pure Appl.Chem. 71 (7): 1307-1325.



Data Curation

Materials Data Curation System

Part of the Materials Genome Initiative

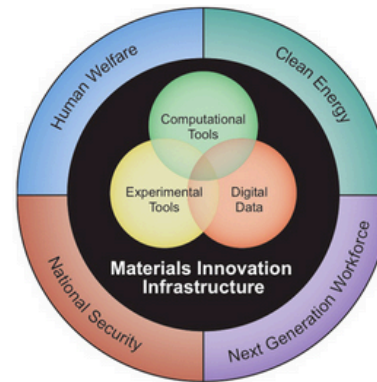
Login | My Profile | Help

Home | **Data Curation** | Data Exploration

Materials Data Curator

This system allows for the curation of Material Data in a repository using predefined templates and a prototype ontology.

This is being developed at the National Institute of Standards and Technology and is made available to solicit comments from the Material Science community. Please do not enter any proprietary data into this system.



Available Options

[All Options »](#)



[Curate your Materials Data](#)

Click here to select a form template and then fill out the corresponding form.



[Explore the repository](#)

Most Recent Templates

[Browse All »](#)

Demo Diffusion Data v2.0 | demoDiffusionData_v2.0.xsd

Demo Light | demo...

Demo Diffusion | der...

New features:

- Ability to store templates
- Schema management tools
- API interface

- Re-written in python
- Backed by MongoDB

Data Curation: Tracer Diffusivity Test Schema

Material Genome Initiative

XML Form Editor

Contact us | F.A.Q | Site

Home Register Experiment Data Exploration

Enter Data View XML

Data Entry

In this step, you have to fill in the form. During the process
Once you have fill every field, you can view the XML.



Experiment

- ExperimentType
 - Choose **TracerDiffusivity**
 - TracerDiffusivity
 - Material
 - MaterialName
 - Phase
 - Name
 - CrystalStructure
 - SpaceGroup
 - SymbolOrNumber
 - WyckoffSequence
 - Sequence
 - Composition
 - QuantityUnit
 - Constituents
 - Element
 - Quantity
 - Purity
 - Error
 - MaterialForm
 - Choose **SingleCrystalline**
 - SingleCrystalline



© 2012 NIST - MGI - XML Form Editor 0.3f | Privacy Policy | Terms of Use



Experiment

- ExperimentType
 - Choose **TracerDiffusivity**
 - TracerDiffusivity
 - DiffusingSpecies
 - Element
 - MaterialPurity
 - ExperimentalConditions
 - MeasurementConditions
 - Time
 - Duration
 - Unit
 - Uncertainty
 - Type
 - Value
 - Temperature
 - Temperature
 - Unit
 - Uncertainty
 - Type
 - Value
 - Environment
 - Environment



© 2012 NIST - MGI - XML Form Editor 0.3f | Privacy Policy | Terms of Use

Future of Materials Data Informatics

- Most materials data are only semi-structured
- Ontologies and the Semantic web offer opportunities to link unstructured data and enable knowledge creation.

Opportunities to Get Involved: (Lots of work to do)

- Share your data and tools
 - <https://materialsdata.nist.gov/dspace/xmlui>
 - <http://nist.matdl.org>
- Curate your data
- Participate in working groups to develop ontologies and XML schemas
 - (NIST Diffusion & CALPHAD Workshop, April 28-30, 2014)