ONTOLOGIES AND RICH METADATA FOR MATERIALS SCIENTIFIC DATA ANALYSIS

Emily C. LeBlanc, Marcello Balduccini, William C. Regli

MATERIALS SCIENCE AND SOCIETY

- Key to mitigating some of society's greatest challenges:
  – Poverty
  – Human health and welfare
  – Clean energy

GETTING A NEW MATERIAL TO MARKET

- Currently 20+ years from discovery of a new material to commercial deployment
- Materials Genome Initiative (U.S., 2011)

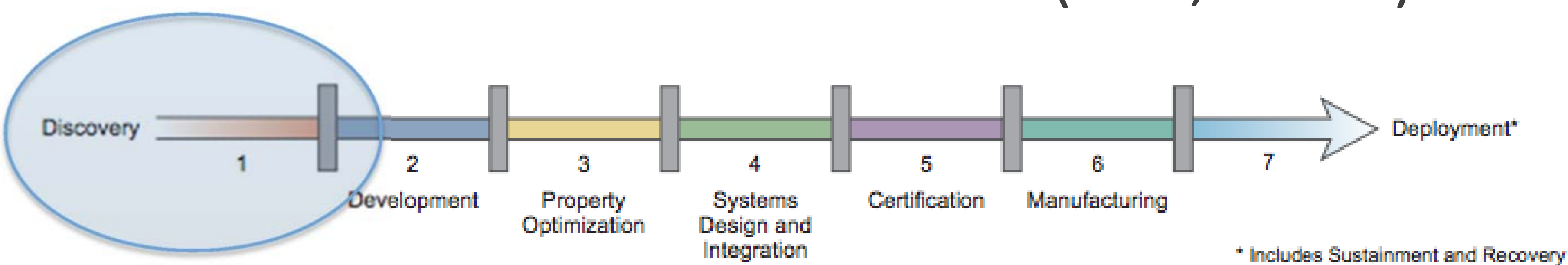


Figure 1: Materials development continuum

EXPEDITING DISCOVERY

- Data Transparency and Integration
  - Scientists require access to the largest possible datasets on which to base their models

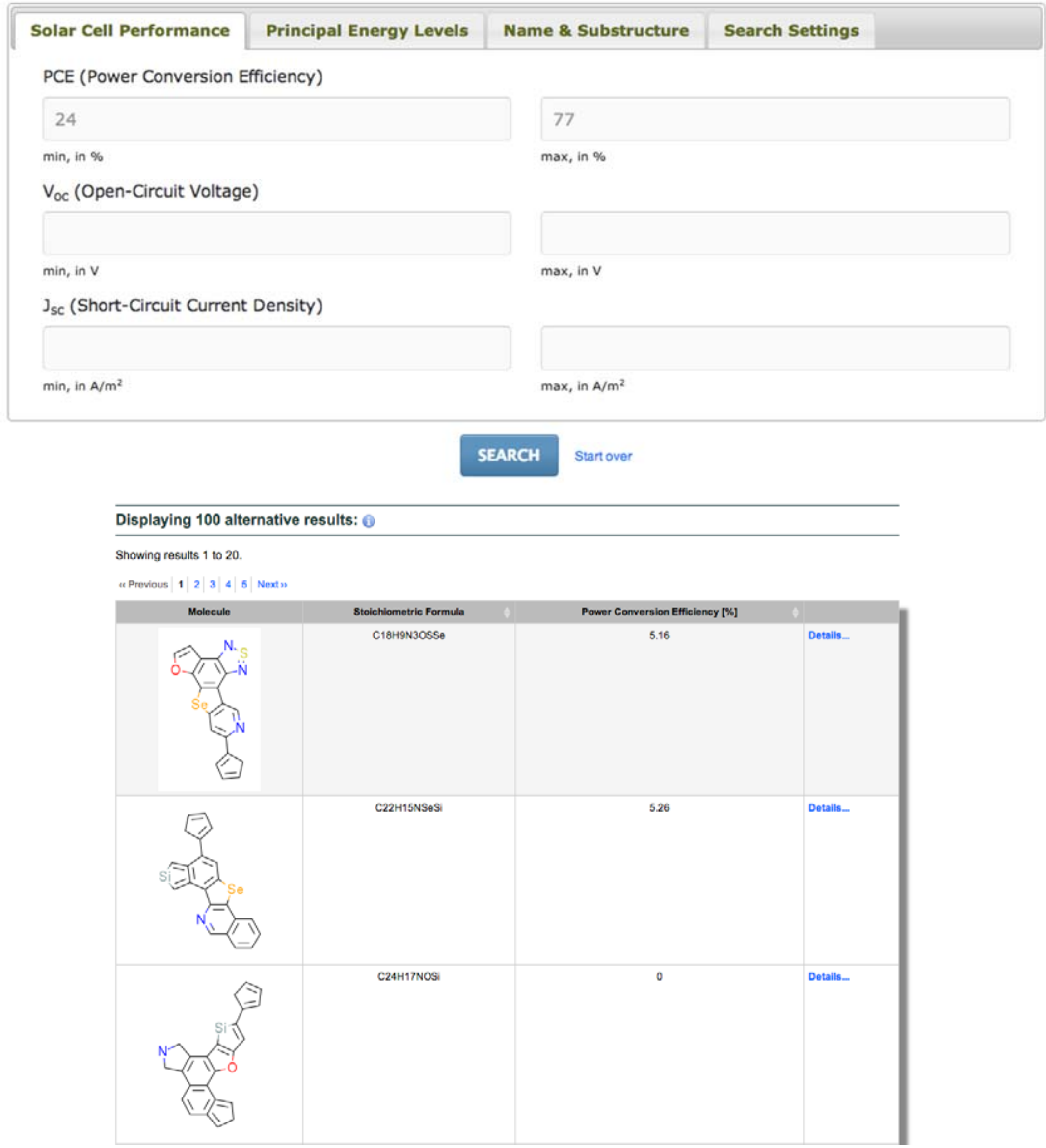
- Many materials science databases…and many unique formats for queries and results

HARVARD UNIVERSITY: CLEAN ENERGY PROJECT

Very large database of organic chemicals for researching solar energy

Search by:

- Parameters describing performance of solar cells or energy levels
- Name and/or substructure of molecules



UNIVERSITY OF TOKYO: MATNAVI

Another very large materials database, broken up into categories. Each searched individually. 3

Search by:

- Parameters describing material or chemical characteristics
- Navigate directory tree

SUPERCON (Numerical database for superconducting materials)

OXIDE & METALLIC (inorganic superconductors: Metals, Alloys, Compounds,Ceramics , etc)

ORGANIC (Organic superconductors)

info data (Knowledge data for materials researchers)

Summary Data (Data views obtained using [SUPERCON])

Home | Oxide & Metallic Menu | Organic Menu | Help

▪ Organic Search System

Structure : BETS

Anion :

Property : ALL

Search Reset

Home | Oxide & Metallic Menu | Organic Menu | Help

▪ Organic Search Result

Structure : BETS
Property : ALL

Results 1 - 38 of 38

num	fullname	tc	pcrit	pmax	fig1	fig2	refno
22	Lamda-(BETS)2GaCl4	5.6	0				P066134503
124	ramda-(BETS)2GaCl4	4.7					JPS0720369
125	ramda-(BETS)2Fe0.4Ga0.6Cl4	3.7					JPS0720369
126	ramda-(BETS)2FeCl4	5.5					JPS0720369
150	kappa-(BETS)2FeBr4	1.1	0		OTC150		ACS1230306
152	kappa-(BETS)2FeBr4	1.4	0		OTC152-1	OTC152-2	P070094514
153	ramda-(BETS)2FeCl4	5.6	.14		OTC153		P070092508
154	kappa-(BETS)2FeBr4						POL0222307
155	kappa-(BETS)2FeCl4						POL0222307
157	ramda-(BETS)2GaCl4	5.46	0				POL0222307
160	kappa-(BETS)2GaCl4	.16	0				POL0222307
196	(BETS)2(Cl2TCNQ)	1.3	.35	.35			CL19990333
197	ramda-(BETS)2GaCl4	8	0				CL19931559
203	ramda-(BETS)2FeCl4	1.8	.3	.3			AC12111243
204	kappa-(BETS)2FeBr4	1	0				ACS1215581
233	ramda-(BETS)2Fe0.45Ga0.55Cl4	6	0				P071104525
235	ramda-(BETS)2FeCl4				OPH235		P072184505
236	ramda-(BETS)2FeCl3.6Br0.4				OPH236		P072184505
237	ramda-(BETS)2FeCl3.5Br0.5				OPH237		P072184505
238	ramda-(BETS)2FeCl3.3Br0.7				OPH238		P072184505
255	ramda-(BETS)2GaCl4	4.8	0				SYN1330219
256	kappa-(BETS)2FeBr4	1.1	0		OTC256		ACS1230306
257	ramda-(BETS)2Fe0.55Ga0.45Cl4	5.4	0				CS11912392
258	ramda-(BETS)2GaBr1.5Cl2.5	9.7	.1	.3	OTC258		PR05608526
261	ramda-(BETS)2GaCl4	4.8	0		OHC261		PHC4400017
262	ramda-(BETS)2GaCl4	5	0		OHC262		JP01308325
263	ramda-(BETS)2GaCl4	5	0		OHC263		SYN1200771
273	kappa-(BETS)2FeCl4	.1	0				CEM1590407

Slide 6

3 no federation within their own database system.
Emily C. LeBlanc, 7/25/2014

FEDERATED DATABASE SYSTEM

- Defines architecture and interconnects databases
- Standard model for information exchange among materials science databases
- Simple software solution can help to expedite research process…we can do more.

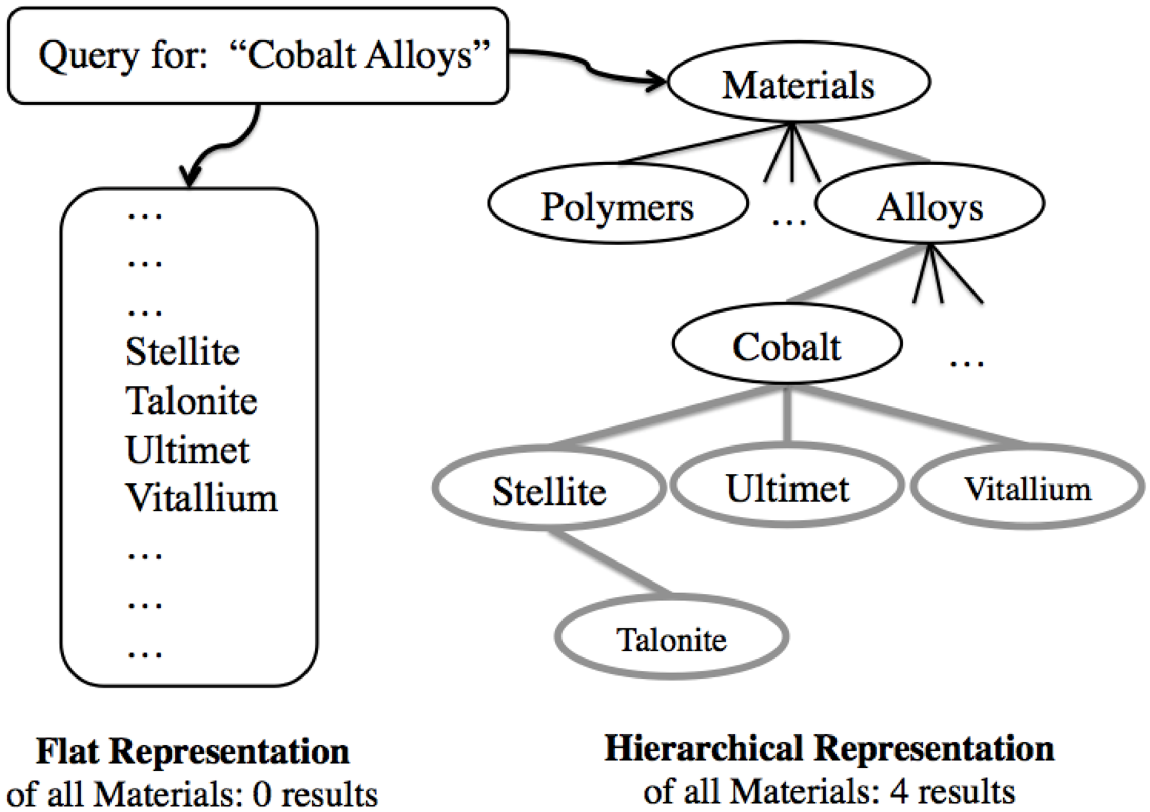
CONTENT-BASED MATERIALS SCIENCE DISCOVERY NETWORK

- Sophisticated ontological framework overlaying material science databases
  - Single query accessing heterogeneous data types
- A content-based networking context allows for intelligent routing of scientific content

A MATERIALS SCIENCE ONTOLOGY

- A materials science ontology provides terms, relationships, and structure necessary to model the domain.
- Metadata is associated with all content in the databases
- Well defined ontological structure aids in discovery of relationship between content

ONTOLOGICAL STRUCTURE

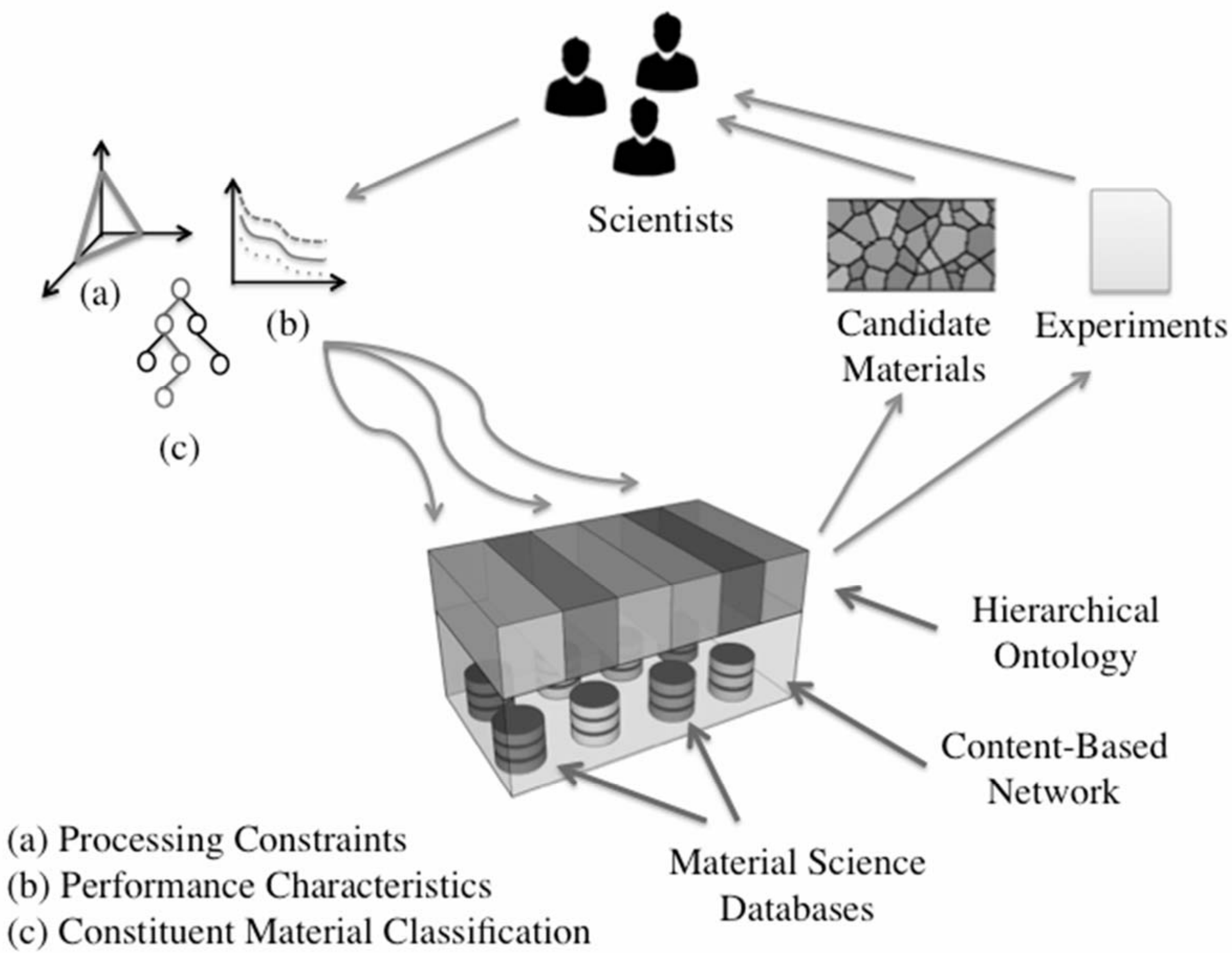


CONTENT-BASED NETWORKING

- Content is addressed instead of host machines
- Receiver declares interest in a file type
- Content providers inject files without knowing the destination
- Network is responsible for routing content

CONTENT-BASED DISCOVERY

- Employ the ontology in a content-based network federating material science databases
- Scientists receive relevant and up to date information about interests (materials, experiments, etc.) via ad hoc queries and subscriptions


Scientists
(a)
(b)
(c)
Candidate
Materials
Experiments
Hierarchical
Ontology
Content-Based
Network
Material Science
Databases
(a) Processing Constraints
(b) Performance Characteristics
(c) Constituent Material Classification

REINFORCING BIG DATA TECHNOLOGIES

- Reduce reaction times to incoming data and the resulting contextual information
- Sidestep some of the high cost of running extensive tests on potentially meaningless correlations
- Provide analytical tools with the newest and most fitting data

QUESTIONS?