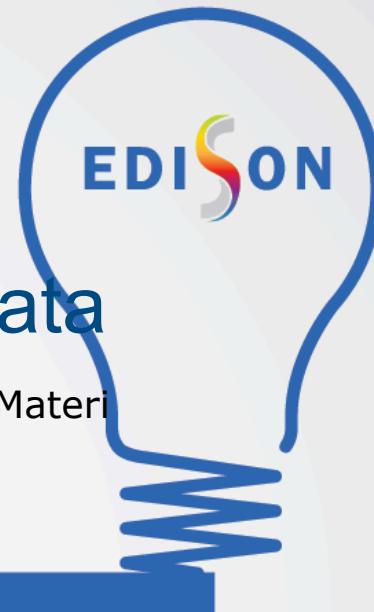


# Data Platform for Computational Science Data

(Focusing on Data Management and Utilization of Computational Science Data in Material field)



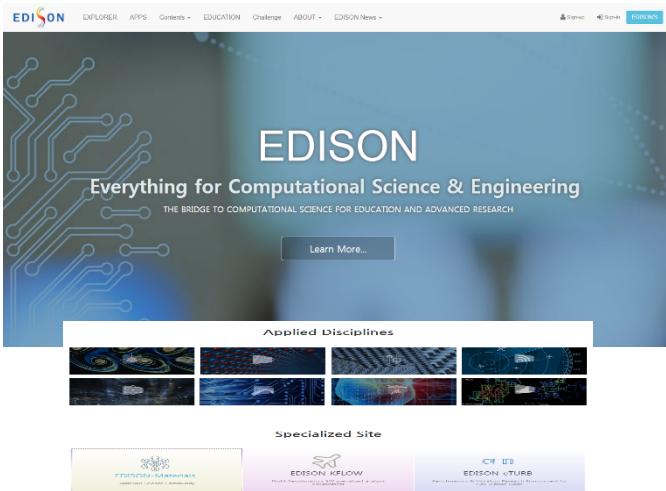
2018. 10

Korea Institute of Science and Technology Information (KISTI)

Jaesung Kim

# 1.1 Background

## 1 An Example of Existing Computational Science Platform : EDISON



- Web-based computational science software (Over 400 SWs, 8 fields)
- Support multi-disciplinary fields
  - Computational Fluid Dynamics, Computational Chemistry, Computational Structural Dynamics, Nano Physics, Computer Aided Optimal Design, Computational Electromagnetics, Computational Medicine, Urban Environment

## 2

### Paradigm Shift to Data-driven methodology



- Avoid duplication of computation/experiment that take a lot of time
- Extraction of meaningful information through analysis of accumulated data

## 3

### Issues in Computational Science Platform

- The rise of Data-driven science methodology
- The heterogeneity of computational science data
- The absence of reliability of computational science data
- The need for convergence of multi-disciplinary software

## 4

### Technical Solutions

- Computational science data platform
- Customizable data curation
- Provenance management
- Scientific workflow

## 1.2 Global Trends

 EDISON

# Open Science & Convergence of Computational Science and Data

## Materials

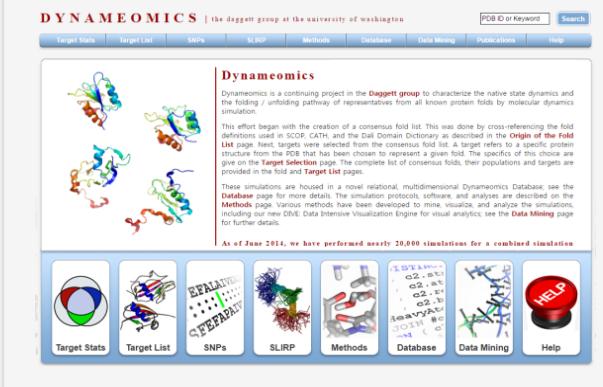
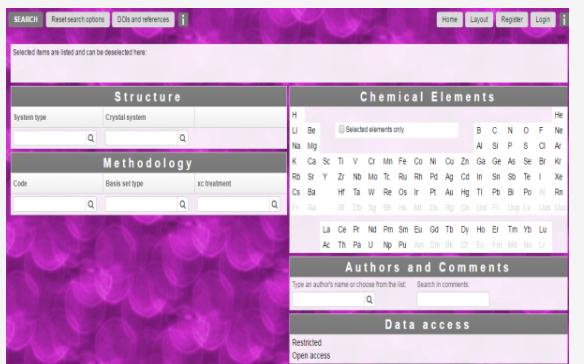
The screenshot shows the MaterialCloud interface. At the top, there's a navigation bar with links for Home, About, Apps, Documentation, API, Tutorials, and Dashboard. Below the navigation is a search bar with the placeholder "Search for materials information by chemistry, composition, or property". To the left of the search bar is a sidebar titled "Explore Materials" with a dropdown menu set to "by Elements". The search bar also includes a dropdown for "Advanced Search Syntax" and a "search" button. The main area features a periodic table where elements can be selected. A tooltip "2 # of elements: e.g., 4 or >2 & <6 excluded elements C1 Br" is shown over the table. Below the table, there are sections for "Material Tags" (including "radioactive" and "isotope") and "Band Gap (eV)" (with a slider from 0 to 10). At the bottom, there's a section for "Energy Above Hull" (with a slider from 0 to 6) and "Formation Energy" (with a slider from -4 to 4).

# Bio Informatics

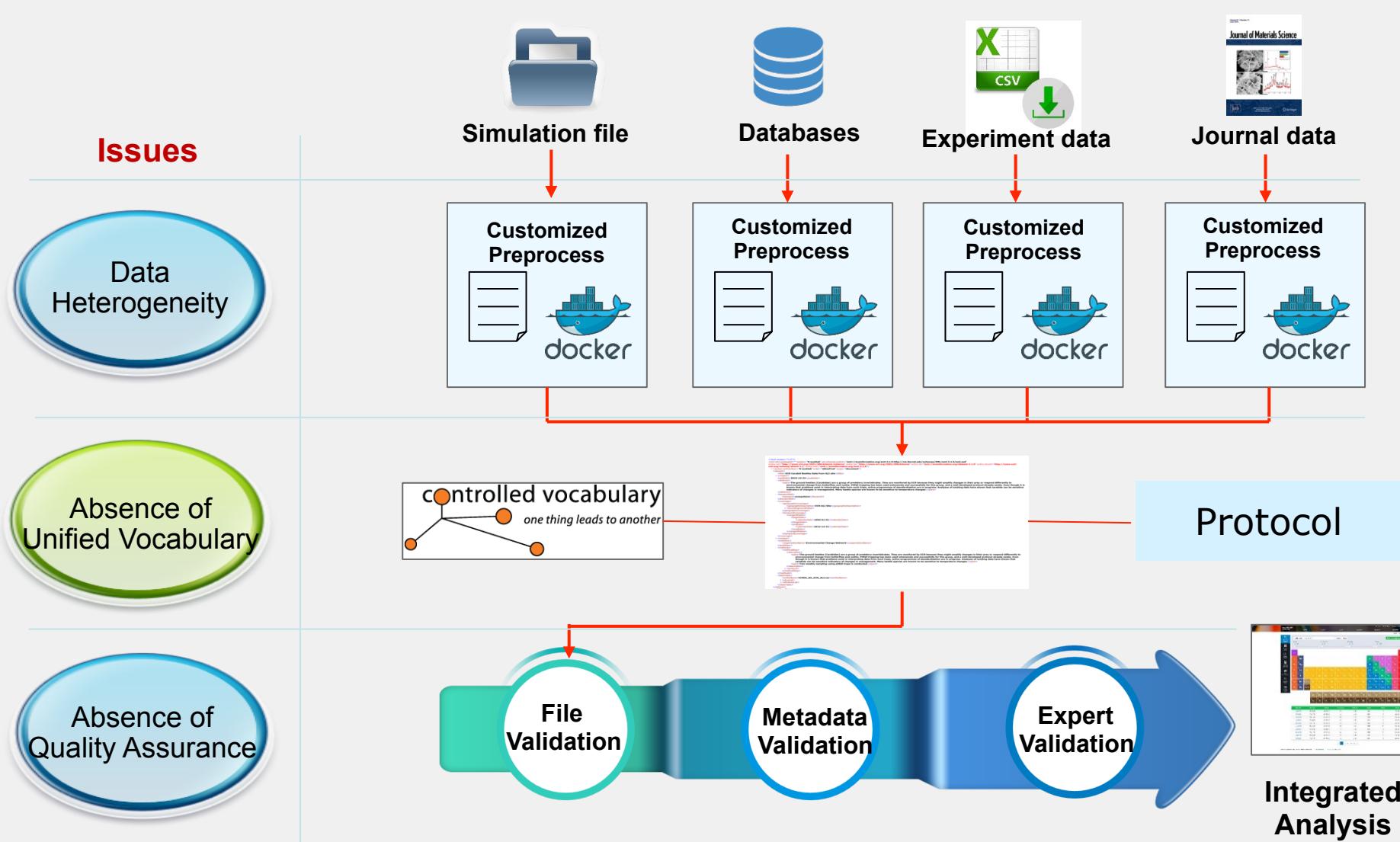
My experiments						
Local indexing of biomolecular simulation data						
Method	Name	Software	Molecules	Publisher	Date	Experiment path
X	TUTORIAL_1-DNA	AMBER	DNA	#0063160	2014-02-12 14:09	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	TUTORIAL_3	AMBER	Protein	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	IBONI1	AMBER	RNA / Protein	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	H-REND	AMBER		#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	MULTI_D	AMBER		#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	FOSFINA	GAFFTAN	CytochromeP450	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	FONT_A_DFT	GAFFTAN	CytochromeP450	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	MORDE DUNDANT	GAFFTAN	CytochromeP450	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	ACETONTRIFOL	GAMES3	CytochromeP450	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
S3		GAMES3	S <sub>3</sub>	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	SODIUM_TODT	GAMES3	No	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	NAMD-AMBER	NAMD	Protein + DNA	#0063160	2014-02-12 14:10	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	FAD	GROMACS	CytochromeP450	#0063160	2014-02-12 14:11	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	NMR1	GROMACS	Protein	#0063160	2014-02-12 14:11	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	NMR2	GROMACS	Protein	#0063160	2014-02-12 14:11	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	SPEEDEITE	GROMACS	Protein	#0063160	2014-02-12 14:11	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	WATER	GROMACS		#0063160	2014-02-12 14:11	/u/fch/jahab.edu/common/home/#0063160/simulat...
X	TUTORIAL_1_BASICS	AMBER	DNA	#0063160	2014-03-05 15:28	/u/fch/jahab.edu/common/home/#0063160/Basics/...
X	ONCE OLDER	AMBER	Nucleic acid	#0063160	2014-03-05 15:44	/u/fch/jahab.edu/common/home/#0063160/Basics/...
X	SUBFOLDERS	AMBER	Nucleic acid	#0063160	2014-03-05 16:23	/u/fch/jahab.edu/common/home/#0063160/Basics/...

# High Energy Physics

- **Search/Reuse** → Avoid duplication of calculation  
→ Efficient use of resources
  - **Analysis/Prediction** → New knowledge



## 2.1 Issues and Our Solutions



## 2.2 Data-driven Open Platform

- Developing services to make it more simple for various types of **Scientific Big Data**
  - a to be ...
    - Collected from **Various sources**
    - Validated : **Quality Assurance**
    - Curated Continuously : **Data Life-cycle Management**
    - Integrated and Analyzed Seamlessly : **Integrated Analysis**



Journal Course Material



SW



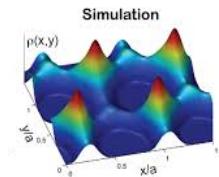
Presentation



Observation



Experiment



Simulation

## 3.1 Data-driven Open Platform (EDISON-DATA Platform)



Sign-in  

Home About Data Community Tool News Sample

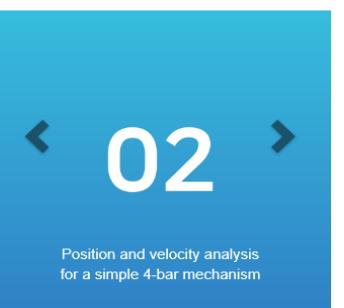


### Simulation



시뮬레이션을 수행할 수 있습니다

### Science App



02

Position and velocity analysis  
for a simple 4-bar mechanism

Everything for Computational Science & Engineering :: Bridge to Computational Science for Higher Education and Advanced Research

관련 사이트 링크



전산열유체



나노물리



계산화학



구조동역학



전산설계



전산의학

# Multi-disciplinary

공지사항

더보기

System Resource Statistics

[공지] 시뮬레이션 서비스 장애 안내 - 수정 완료

2017-05-17

[공지] 해석 시간 지연 문제 - 수정 완료

2017-05-12

[안내] EDISON 포털 교육 서비스 안내

2017-04-28

[공지] 제6회 EDISON SW 경진대회 분야별 세부 일정 안내

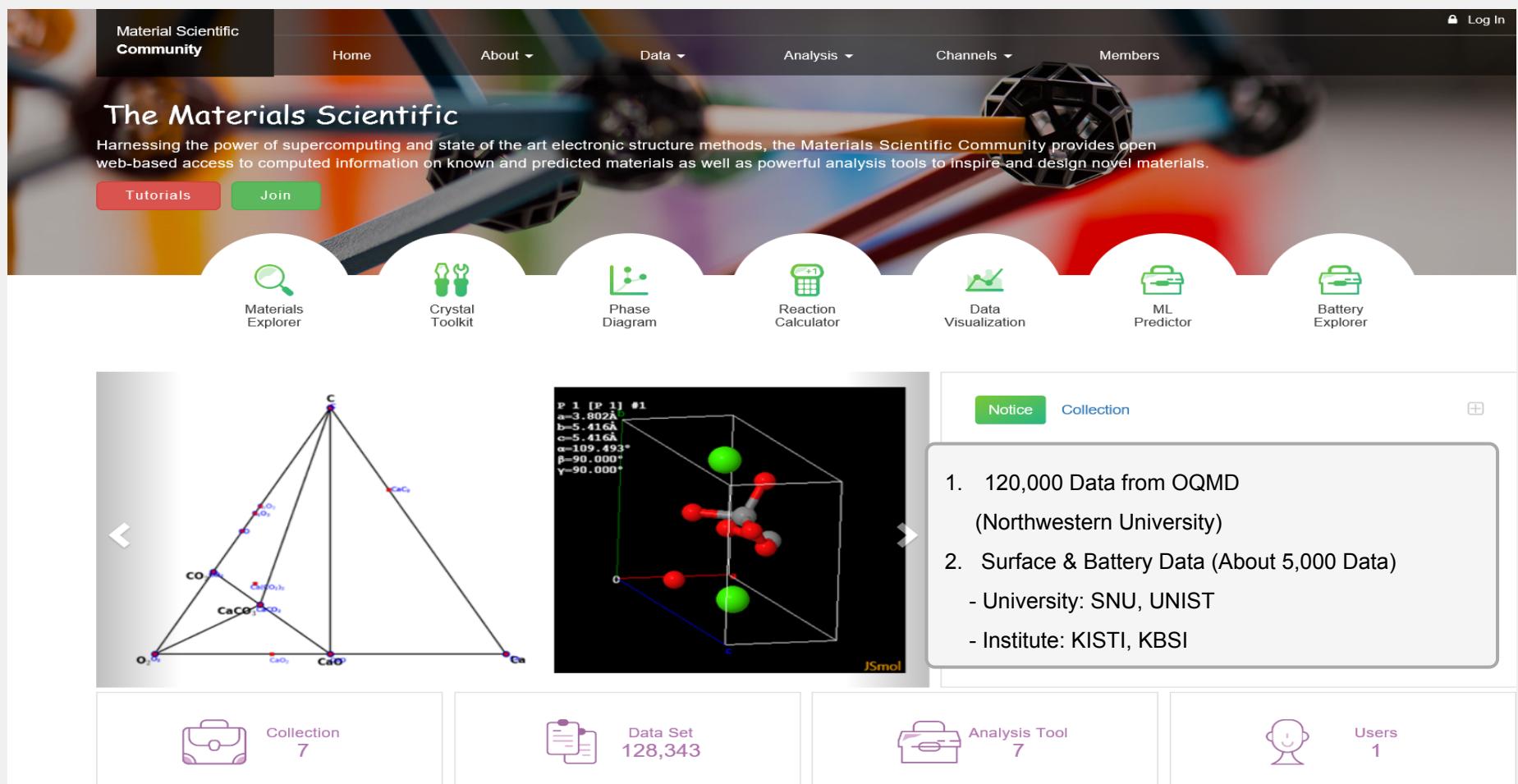
2017-03-20

[공지] 제6회 EDISON SW 경진대회 일정 안내

2017-03-16

Site	Cluster	Total	Used	Avail
EDISON	EDISON-CFD	424	143	281
	EDISON-CHEM	296	65	231
	EDISON-NANO	512	53	459
	EDISON-CSO	192	14	178
	EDISON-DESIGN	136	32	104

# 3.2 Materials Portal as a Pilot



The Materials Scientific Community provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.

**Tutorials** **Join**

**Materials Explorer** **Crystal Toolkit** **Phase Diagram** **Reaction Calculator** **Data Visualization** **ML Predictor** **Battery Explorer**

**Notice** **Collection**

- 1. 120,000 Data from OQMD  
(Northwestern University)
- 2. Surface & Battery Data (About 5,000 Data)
  - University: SNU, UNIST
  - Institute: KISTI, KBSI

**Collection 7** **Data Set 128,343** **Analysis Tool 7** **Users 1**

# 3.3 Data View

CollectionId	21912	Community	Materials
Title	Materials Design and Discovery with High-Throughput Density Functional Theory... (OQMD)		
Owner(Creator)	Sunil Ahn	Usage Right	MIT
DOI	10.5072/sdr.kisti.21912	Contributor	oqmd database
High-throughput density functional theory (HT DFT) is fast becoming a powerful tool for accelerating materials design and discovery by the amassing tens and even hundreds of thousands of DFT calculations in large databases. Complex materials problems can be approached much more efficiently and broadly through the sheer quantity of structures and chemistries available in such databases. Our HT DFT database, the Open Quantum Materials Database (OQMD), contains over 200,000 DFT calculated crystal structures and will be freely available for public use at <a href="http://oqmd.org">http://oqmd.org</a> . In this review, we describe the OQMD and its use in five materials problems, spanning a wide range of applications and materials types: (I) Li-air battery combination catalyst/electrodes, (II) Li-ion battery anodes, (III) Li-ion battery cathode coatings reactive with HF, (IV) Mg-alloy long-period stacking ordered (LPSO) strengthening precipitates, and (V) training a machine learning model to predict new stable ternary compounds.			

## Access Control

Allowed Users	
Permission	Allowed Us
Community M	Community M
Non-Community	Non-Community
Guest	Guest

**Datasearch**

- [Material Infomation](#)
- [X-Ray Diffraction](#)
- [File](#)
- [Metadata](#)
- [Comments](#)

**Material Details**

Formula: LiCoO<sub>3</sub>

Final Energy Per Atom: -5.211972

Formation Energy: -1.115486

Nsites: 10

Volume: 83.2698

Band Gap: 0

Mass: 227.744802

Density: 4.540138

**Structure**

P 1 [P 1] #1  
 $a=4.870\text{\AA}$   
 $b=4.870\text{\AA}$   
 $c=4.870\text{\AA}$   
 $\alpha=60.864^\circ$   
 $\beta=60.864^\circ$   
 $\gamma=60.864^\circ$

JSmol

**Space Group**

<b>Num</b> 167	<b>Symbol</b> R-3c
<b>Group Hall</b> -R 3 2c	<b>Point Group</b> -

**Lattice Parameters**

<b>a</b> 4.870138	<b>b</b> 4.870138	<b>c</b> 4.870139
<b>Alpha</b> 60.86405	<b>Beta</b> 60.86405	<b>Gamma</b> 60.864027

## Dataset List (120543)

datasetId	Title	Status	doi
706401	Lu	approved	10.5072/sdr.kisti.21912.706401
706405	Tm	approved	10.5072/sdr.kisti.21912.706405

# 3.4 Data Search for Material Data (1/2)

Materials Explorer

Crystal Toolkit

Phase Diagram

Reaction Calculator

3D Visualization Chart

ML Predictor

Battery Explorer

by Elements ▾ Li,O,Si   Periodic Table Close

density  # of Elements  Bandgap  Volume

1 H													2 He						
3 Li Lithium	4 Be													8 O Oxygen	9 F	10 Ne			
11 Na	12 Mg													13 Al	14 Si Silicon	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr		
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe		
55 Cs	56 Ba	57-71 La-Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn		
87 Fr	88 Ra	89-103 Ac-Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo		
57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu					
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr					

Material ID	Data Type	Reduced Formula	Spacegroup	Bandgap	Final Energy/Atom	Formation Energy	Density	Nsites	Volume
738446	oqmd	Li8SiO6	Cmc21	4.91800	-5.50956	-2.32706	2.26713	30	263.02200
738572	oqmd	Li15Si2O12	P1	.00000	-5.50223	-2.27539	2.23217	29	261.97900
738673	oqmd	Li15Si2O12	P1	5.32300	-5.48870	-2.26186	2.21514	29	263.99400

# 3.4 Data Search for Material Data (2/2)

## Materials Explorer

Materials Explorer

Crystal Toolkit

Phase Diagram

Reaction Calculator

3D Visualization Chart

ML Predictor

Battery Explorer

by Elements ▾ O,Ca,Ti,Sr,Ba

Search Reset

density 0-25 # of Elements □

0-8 Bandgap

0-18 Volume

220-620

Periodic Table Close

1 H	2 He																
3 Li	4 Be																
11 Na	12 Mg																
19 K	20 Ca Calcium	21 Sc	22 Ti Titanium	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr Strontium	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba Barium	57-71 La-Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl Thallium	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	89-103 Ac-Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu			
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr			

No data found.

# 3.5 Advanced Search

**Data Type**

oqmd x Please enter the Data Type

**Query**

elements: Li NOT elements : Si

**Result of**

finalenergyperatom x volume x Please Select Return

Search

**Example**

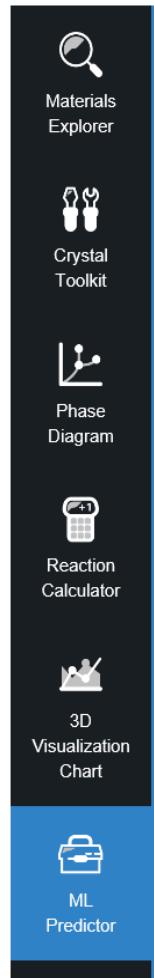
1. Equal search
  - 1-1. elements: Li
    - To search for datasets that contain elements 'Li'
  - 1-2. nsites : 3
    - To search for datasets that contain nsites '3'
2. AND operator
  - 2-1. elements: Li AND elements: O
    - To search for datasets that must contain elements 'Li' and 'O'
  - 2-2. elements: Li AND nsites : 1
    - To search for datasets that must contain elements 'Li' and nsites '1'
3. OR operator
  - 3-1. elements: Li OR elements: O
    - To search for datasets that contain elements 'Li' or 'O'
  - 3-2. elements: Li OR nsites: 2
    - To search for datasets that contain elements 'Li' or nsites '2'
4. NOT operator
  - 4-1. elements: Li NOT elements: O
    - To search for datasets that contain elements 'Li' but not 'O'
5. Range search
  - 5-1. nsites: [1 TO 5]
    - To search for datasets that nsites are between 1 to 5 (For Integer)
  - 5-2. density: [1.0 TO 5.0]
    - To search for datasets that density are between 1.0 to 5.0 (For Float)

5691 Datasets Found.

datasetId	Title	finalenergyperatom	Volume
706647	Li	-1.895582	18.0942
706668	Li	-1.893234	18.1678
707098	Li	-1.872281	16.7856

# 3.6 Material Property Prediction Service based on DL (1/2)

## ML Predictor



### Input

Type select

POSCAR Quantum Espresso

New structure  
1.0

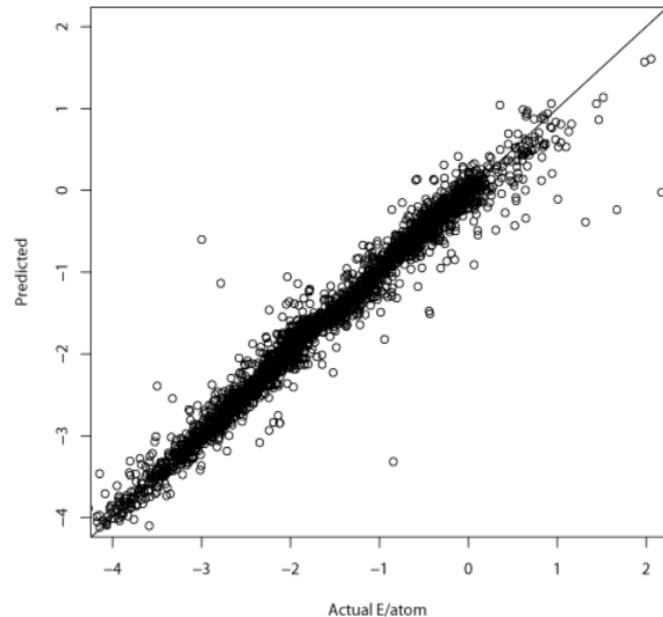
3.9429521561	0.0000000000	0.0000000000		
0.0000000000	3.9429521561	0.0000000000		
0.0000000000	0.0000000000	12.0000000000		
Ba	Sr	Ca	Ti	0
1	1	1	3	9

Direct

0.00000000	0.00000000	0.00000000
0.00000000	0.00000000	0.333332986
0.00000000	0.00000000	0.666666627
0.49999970	0.49999970	0.166666657
0.49999970	0.49999970	0.49999970
0.49999970	0.49999970	0.833333254
0.49999970	0.49999970	0.00000000
0.49999970	0.49999970	0.333333313
0.49999970	0.49999970	0.666666627
0.00000000	0.49999970	0.166666657
0.00000000	0.49999970	0.49999970
0.00000000	0.49999970	0.833333254
0.49999970	0.00000000	0.166666657
0.49999970	0.00000000	0.49999970
0.49999970	0.00000000	0.833333254

RUN PREDICTION

### Structure



Training MAE(80%): 0.05

Testing MAE(20%): 0.12

Training COD(R2): 0.987

Testing COD(R2): 0.967

# 3.6 Material Property Prediction Service based on DL (2/2)

## ML Predictor



### Input

Type select

POSCAR Quantum Espresso

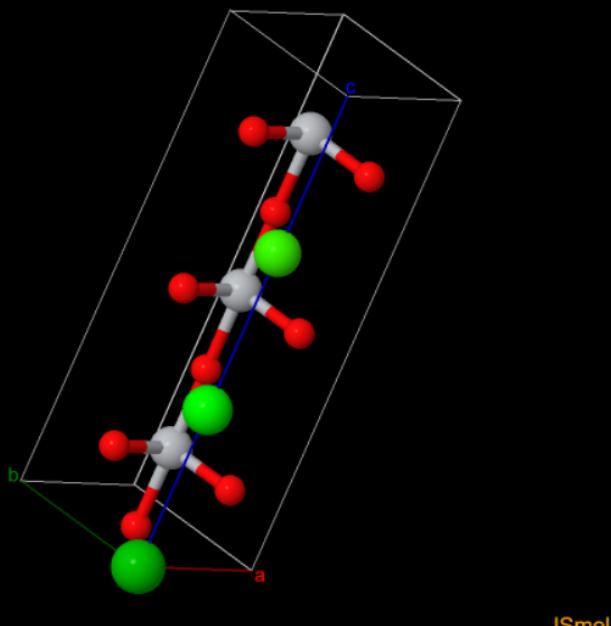
New structure

```
1.0
      3.9429521561      0.0000000000      0.0000000000
      0.0000000000      3.9429521561      0.0000000000
      0.0000000000      0.0000000000      12.0000000000
Ba   Sr   Ca   Ti   O
1     1     1     3     9
Direct
 0.0000000000      0.0000000000      0.0000000000
 0.0000000000      0.0000000000      0.333332986
 0.0000000000      0.0000000000      0.666666627
 0.49999970      0.49999970      0.166666657
 0.49999970      0.49999970      0.49999970
 0.49999970      0.49999970      0.833333254
 0.49999970      0.49999970      0.000000000
 0.49999970      0.49999970      0.333333313
 0.49999970      0.49999970      0.666666627
 0.0000000000      0.49999970      0.166666657
 0.0000000000      0.49999970      0.49999970
 0.0000000000      0.49999970      0.833333254
 0.49999970      0.0000000000      0.166666657
 0.49999970      0.0000000000      0.49999970
 0.49999970      0.0000000000      0.833333254
```

RUN PREDICTION

### Structure

P 1 [P 1] #1  
a=3.943Å  
b=3.943Å  
c=12.000Å  
α=90.000°  
β=90.000°  
γ=90.000°



JSmol

### Predictions

#### Base Properties

Formation Energy per Atom (eV)

-3.444

Final Energy per Atom (eV)

-7.975

Volume (Å³)

186.562

Density (g/cm³)

4.919

Space Group Symbol

P4mm

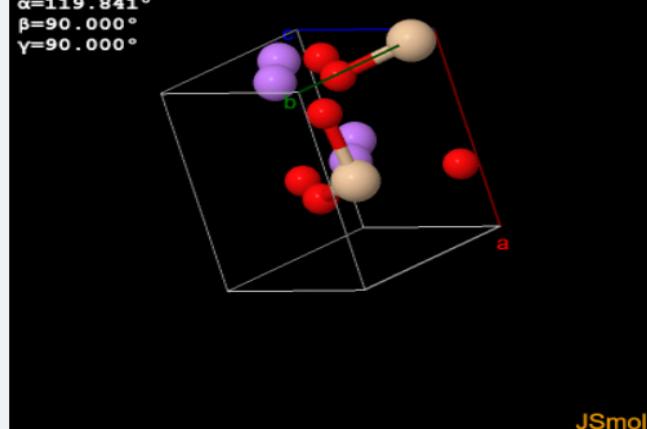
# 3.7 Crystal Toolkit

Li<sub>2</sub>SiO<sub>3</sub>

1	0	0
0	1	0
Fe		
Fm		
Fr		
Ga		
Gd		
Ge		
H		
Hf		
Hg		
Ho		
I		
In		
Ir		
K		
La		
<b>Li</b>		
Lr		
Lu		
Mg		
Mn		
Mo		
N		
Na		
Nb		
Nd		
Ni		
No		
Np		
O		
Os		

## Crystal Data

```
P 1 [P 1] #1
a=4.651Å
b=5.362Å
c=5.388Å
α=119.841°
β=90.000°
γ=90.000°
```



## Lattice Parameters

a	4.65
a	5.36
c	5.39
α	119.84
β	90.00
γ	90.00
volume	116.53

Choose File Format ▾

Li

O

Si

a	b	c
0.51	0.51	0.35
0.01	0.83	0.35
0.51	0.17	0.65
0.01	0.49	0.65

a	b	c
0.65	0.12	0.00
0.15	0.88	0.00
0.09	0.46	0.29
0.59	0.84	0.29
0.09	0.16	0.71

a	b	c
0.01	0.17	0.00
0.51	0.83	0.00

# 3.8 Phase Diagram

**Phase Diagram**

C, O, Ca

Search Reset Periodic Table Close

Materials Explorer  
Crystal Toolkit  
Phase Diagram  
Reaction Calculator  
3D Visualization Chart  
ML Predictor  
Battery Explorer

Periodic Table:

1 H	2 He																
3 Li	4 Be																
11 Na	12 Mg																
19 K	20 Ca Calcium	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	57-71 La-Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	89-103 Ac-Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu			
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr			

Ca + C + O

Mouse over dots for details

Compounds

stable (7)	Unstable (181)	
1206904	CaO	-3.03848
1205637	O <sub>2</sub>	0.00000
1205642	C	0.00000
1207978	CaO <sub>2</sub>	-2.02587
1233158	CaCO <sub>3</sub>	-2.21991
1205695	Ca	0.00000
1206668	CO <sub>2</sub>	-1.31107

Diagram showing phase relations between Carbon (C), Oxygen (O), and Calcium (Ca). Key points labeled include CO<sub>2</sub>, CaO<sub>2</sub>, CaCO<sub>3</sub>, and CaO.

# 3.9 Reaction Calculator

## Reaction Calculator

🔍  
Materials Explorer

⚖️  
Crystal Toolkit

📈  
Phase Diagram

🔢  
Reaction Calculator

🎥  
3D Visualization Chart

Reactants

Products

HΔ →

$\text{CaO} + \text{CO}_2 = \text{CaCO}_3$  >  $\Delta H_{\text{calculated}}$   
-1.09 eV (-105.11 kJ mol<sup>-1</sup>)

Formation Energies

Compound	Calculated
CaO	-6.08
CO <sub>2</sub>	-3.93
CaCO <sub>3</sub>	-11.10

# 3.10 3-D Visualization Chart

## 3D Visualization Chart

