MatDX: A Data Warehouse Solution for the Integration and Classification of Materials Information

Jungho Shin¹, Hyunju Chang¹

¹Korea Research Institute of Chemical Technology, 141 Gajeongro, Yuseong, Daejeon 34114, Korea | jungho@krict.re.kr

ABSTRACT In the field of data-driven materials science, the emergence of useful data infrastructures based on the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles, such as NOMAD, OPTIMADE, Materials Project, AFLOW, and OQMD, has accelerated this process even further. To facilitate the utilization of these data infrastructures, the integration and classification of numerous kinds of metadata related to materials properties are required. This, in turn, enables the research community to perform more accurate predictions through the use of machine learning algorithms. MatDX (Materials Data eXplorer) has been developed with a primary focus on the integration and classification of materials ontologies, including classes and instances in material name, composition, compound, structure, property, and applications. To facilitate integration, MatDX has been developed as a kind of data warehouse solution, which serves as a useful strategy for connecting multiple databases. Material tags have been introduced through classification by using the materials ontologies, and they are used to search and view detailed information for interesting materials in an easy and quick way. Further analysis functionality to visualize statistically meaningful relationships has been served in an interactive way.

The ultimate goal of these services is to enable researchers to discover novel materials for target properties based on numerous research data. MatDX, including three sub-categories of PubDX (for published data), ExpDX (for experimental data), and CalcDX (for calculated data), is ready to be accessed through the address of http://materials.chemdx.org.





ONTOLOGY

Operator Name	^	1.00	1000 e 000	hang strategy being		_					
Experimental Name:		key	value	comment			D D Insummerskitenkorpie a	+			
SampleDescription: Excitation Wavelength: 532.04 nm Laser Current: 100 %		PVP_High.txt			← → C A D. http://waterolic.com/complexity						
Excitation Power: 10.58 mW		key		value	comment		MATERIALSDATA	4		ive a marcal	
ND Filter: 19.17 % (180/255)		excitation_wavelength		532.04 nm			MATERIALS APPICATIONS		The Deal from Nonday		
Spectrograph Center Wavelength: 2500.00 cm-1 Grating: 300 gr/mm		grating		300 gr/mm				3 compositions fr	Formula A/203	reduced from Al203	Moterial Nam Aluminium(II)
Slit width: 50 um		exposure_time		0.5 s							
Exposure Time: 0.5 s		averaging		50				- Person Long	Composition		
Averaging: 50 Readout speed: 2MHz		0.0						S M 2	Compositional arity binary	Compound class	
Gain: High Readout port: Low Noise								2qA	Structure		
CCD temperature: -70 Optical Line Shaping mode: OFF								A203	Dimensional type bulk	Crystal system cubic	Space group Pm-3m
Objective Lens: TU Plan Fluor 20x/ NA 0.45		0							AE) //W productions 1101		





{A}	Array value
{I}	Image data
{N}	Numeric value
{P}	Path of raw file
{U}	URL of original data
[Ins.]	Name of instrument

WEB SERVICE - HOME

MatDX	<u>36</u> .	VDiet			
PubDX	Q	KRICI			
ExpDX	$\overline{\mathcal{P}}_{*}$			DETA	
CalcDX		Materials D	ata Explorer	DETA	Total Compositions
MatDX provides materials informative integrated from multiple DBs as debelow: NOMAD : 563k formulas CMiB : 26.7k entries TEXplorer : 1.2k calculations	ition ecribed	MatDX provides material property KRICT and other collaboration grou experiments DB, and calculations I MatDX's material tags, found on th material information for specific re	data collected from public DBs or provid ups. The collected data is categorized int DB. is platform, allow users to quickly and ea esearch interests. Search results are disp	ed by research groups in to Public DB, asily search and filter played in a hierarchical	570,822
ExpDX DATA COLLECTIO	NS 🗸	table showing all material information and ontology used for integration and o	tion consolidated from multiple database classification.	es. We provide material	
MATERIALS APPICATION	IS 🗸				
		Search Public Data	Search Experiments Data	Search Calculations Data	Move to LitDX



Length(Å) 3.95389

3.95389

3.95389

ec 02 2019 21:17:30) complex parallel LinuxiFC

x(Å) y(Å)

1.97694 1.97694 1.97694

A unavailable

9 PAW_PBE

unavailable

Volume(Å³)

A m-3m

Pseudopotential Type

Spacegroup

GGA_C_P8E+GGA_X_P8E

5 676 eV

Computational Method

Calculation Type

Energy Cutoff

90.00

2			1					
10k								
-								
15k								
						1		
	500	1000	1500	2000	2500	3000	3500	4000
	10k 15k	10k 15k 500	10k 10k 15k 500 1000	10k 15k 500 1000 1500	10k 15k 500 1000 1500 2000	10k 15k 500 1000 1500 2000 2500	10k 15k 500 1000 1500 2000 2500 3000	10k 15k 500 1000 1500 2000 2500 3000 3500

♣ 5.9293 eV 5.9228 eV

🧈 unavailable

-

Density of States





KRICT

Materials Data Explorer BETA

MatDX provides material property data collected from public DBs or provided by research groups in KRICT and other collaboration groups. The collected data is categorized into Public DB, experiments DB, and calculations DB. MatDX's material tags, found on this platform, allow users to quickly and easily search and filter

material information for specific research interests. Search results are displayed in a hierarchical table showing all material information consolidated from multiple databases. We provide material ontology used for integration and classification.



Statistic : data visualization tools



571,143

DATA VISUALIZATION



Transparent Solvent Hole transfer layer Antiferromagnet AFM

Intermetallic Solar Cell Graphene Nano Electron transfer layer

Semiconductor | Electrode

	BA Paravalite Ball Collin		Ante-See Berla Dikrisgeeth Tapercodutte	abayab	10 N.
ngAl .	Bitary Balk Messaginal Tatrage Octoartomile: Cybic	nal (Tignal)	DOD Digarcalizes LAMID HOME Path energy Vyo Ant MIE Each alone H.F. Second Construction Units' energies Units' energies Vyo Ant MIE Each alone H.F.	EMA P	ads
8021	Reary Rick' Cobie		Liddle Digensities 406 NORE) Frail energy Fr Yes Cell energy PER Concrete Decrete Report Decret Report D	DUNA P	
Permiteitatea	Graphie Inches	Property	Sector Sector	helerma	Very Deaths
AI203	Bildy Babe	tinal every		NUMAS	and and a second se
A003	Reary Bulk Ealer	texterep		NUMAD	within the
AGR .	(control (control (control	Annen	annel constituenene Constituenen	Property.	Contra .

Details about materials

Fermilahame Dutquarter & Granter

ē		1000		1.000			Treing	· · · · · · · · · · · · · · · · · · ·		100000
		Lamada	-	-	1.44	-	-			1.0
_	1.1	1.3	6	-		446	-	*		
	1		<u></u>	- 199	-3-	- 22				
		-	100	sim.			Balk *** *****			
		-		100	100	1000	at here.		111111	
		17					-	4.10000000000	Stations.	
	-160	13	*		-	12	they are not	1.1000.000.000		
		-		-			Computation	nal Method		
							-	A		
							increase law	a second	inclusion in	1 Internet