Materials Research and Data Science Conference



Materials Database based on MGI and Big Data in China

Yanjing SU, <u>Haiqing YIN</u>, Chaofang DONG, Xue JIANG University of Science and Technology Beijing, China

yjsu@ustb.edu.cn, hqyin@ustb.edu.cn

Sept 26, 2017, Rockville, MD

www.wps.c

Beijing Institute of Iron	Beijing University of University of Science	General	
Founded in 1952	Tron and Steeland Technology BeijingTechnology 19601988	Schools National Science Center	14
A key national university sponsored by the Chinese Ministry of Education	School of Materials Science and Engineering School of Metallurgical and Ecological Engineering	State Key Lab National Engineering Research Center	2
Renowned for its study of	School of Civil & Resources EngineeringSchool of Mechanical Engineering	Faculty & Staff	1741/
materials science and metallurgy	School of Automation and Electrical EngineeringSchool of Computer and Communication Engineering	Members of CAS	1643 4
batch of "211 project".	School of Mathematics and Physics School of Chemistry and Biological Engineering	Professors Associate Professors	489 773
2006, Included in "985 innovation platform for	Donlinks School of Economics and Management	Students	
superiority subject" project 2014, Included in "2011	School of Humanities and Social Sciences School of Marxism	Full-time Students 2 Undergraduate Students 1	14,444 13,615
plan project" as a lead university.	School of Foreign Studies School of Energy and Environmental Engineering	Postgraduate Students Doctoral Students International Students	6,915 3,022 892
		international Students	072

國來統

Profile of USTB





Their research covers met	als, nonmetals and polymers
Material Optimization Design Laboratory	Magnetic Materials and Devices Laboratory
Materials Phase Balance and New Materials Design Laboratory	Nanocomposites Materials and Green Catalysis Laboratory
Material Science Basis and Material Anisotropy Laboratory	Materials Surface and Interface Laboratory
High Temperature Material and Applications Laboratory	Crystal Defects and Materials Irradiation Damage Laboratory
Magnetic Material Laboratory	Advanced micro and Nano Energy and Sensor Systems Laboratory
Biomedical Material Laboratory	Nano Device Design and Characterization Laboratory
Clean Energy Industry Material Laboratory	Advanced Functional Film Materials and Applications Laboratory
Laboratory of Advanced Processing Technology of Material	Polymer Display and Energy Materials Laboratory
Material Forming Process Control and Simulation Laboratory	New Photoelectric Functional Materials Laboratory
Material Forming Theory and Quality Control Laboratory	Optical Functional Materials and Devices Laboratory
Material Forming Technology and Mold Research Laboratory	New Magnetic Functional Materials Laboratory
Material Solidification Forming Laboratory	Materials Microstructure and Performance Laboratory
Material Advanced Welding and Connection Technology Laboratory	Laboratory of Refractory Materials
Material Design and Processing Integration Technology Laboratory	Laboratory of Multifunctional Inorganic Materials
Functional Structural Material Design and Application Laboratory	Laboratory of Novel Energy Materials
Metal Micro-Nano Structured Material Laboratory	Laboratory of Functional Ceramics and Devices
Composite Material Structural Engineering Center	Laboratory of Advanced Ceramics and Nuclear Materials
Powder Metallurgy and Advanced Ceramics Laboratory	

比京钾枝大學 University of Science and Technology Beijing 3.94

Materials Education

	6 Undergraduate Programs	3 Postgraduate Programs	71.29% 65.97% ^{66.80%} 66.59%
	Metal Materials Materials Processing and Contro Engineering	 Materials Science Materials Physics and Chemistry 	63.10%
	Inorganic Nonmetallic Materials Materials Physics Materials Chemistry	Materials Processing and Control Engineering	2012 2013 2014 2015 2016 Rate of advanced study (%)
•	Nanomaterials & Nanotechnology	~400 master students and ~100 Ph. D candidates get	23.23% 22.34% 21.46%
•	 ~400 undergraduate students get enrolled each year Enable students to begin specializing and getting practical experience very early in their program 	 candidates get enrolled each year They are heavily focused on research work mixed with coursework. 	2012 2013 2014 2015 2016
	5		Rate of overseas study (%)

Profile of materials science discipline







The status quo of materials database in China

- Materials Scientific Data Sharing Network
- China gateway to corrosion and protection

Progress in MGI databases and big data in China

I. Database and Big Data Technology of Material Genome Engineering (MGE)

II.Data acquisition and database fusion technology on structure -property based

on high-throughput experiments and calculations

History of Material Databases in China



Three stages

- <u>1980s</u>: around 23 materials databases were built by universities and research institutes, with the financial support from the national government.
- Small scale (small volume, less users), little update.



National Scientific Data Sharing Project of China was launched in 2000.



<u>2015</u>: MGI related database and data technology

Support from Central and Local Government

Data Resources of diverse disciplines



- 2 National projects for Materials data (data centers located in USTB):
- National Materials scientific data sharing network
 - --- distributed, covering nearly all the materials, mainly for material selection.
- China gateway to corrosion and protection
 - --- centralized, specially focusing on corrosion.





The status quo of materials database in China

- Materials Scientific Data Sharing Network
- National Environmental Corrosion Platform

Progress in MGI databases and big data

- I. Database and Big Data Technology of Material Genome Engineering (MGE)
- II. Data acquisition and database fusion technology on material microstructure -property

based on high-throughput experiments and calculations

Materials Scientific Data sharing Network



Totally 18 universities & institutes contribute to data collection.



Source of data

Handbooks, patents, papers, testing data, calculation data, products dealers, Industry statistics. Materials Data Category

- Material foundamental
- Metals & alloys
- Ceramics
- Organic Materials
- Composites
- Biomaterials
- Energy Materials
- Information Materials
- Natural Materials
- Construction Materials
- **Transportation Materials**

Materials Scientific Data Sharing Network





• Multiple sources

feature

- Heterogeneous structure
- Distributed storage

Amount of Data

Around 690,000 entries of 22,000 kinds (brand name) of materials, with 236 metadata.

Materials Scientific Data Sharing Network





Materials Scientific Data Sharing Network



Standardized Doi Registration System (standard of encoding for Material data) – Intelligent property issues

head						
'depositor name			Internation/Facility/Harrison	03/49		
*depositor email_address	4		此处道可们构中操改有关论。	永信意约 句子信稿10以		
*registrant		_	此些描写注册者, 资料注册	1月的机构的信息 长虎	不直(21209/17	
••••••						
body database						
* contributor_name			此处填写责任者坦宾			
* contributor_role	(author)/E@t	•	化合成型面包有有面包			
* contributor_sequence	(first)第一责任者	٠	此处城写奏任袁绍序			
organization, name			此论编写责任者所在机构可	8		
organization_role	(author)行者	٠	此地域写真任常所在的构象的	5 নাড		
organization_sequence	(first)第一类任者	•	此后期期期任何利益地的	safe the	重要活动	EDP#
* title			此些描写数据库名称			
subtitle			******	彩旗		
description	1		此处境可数据车(数据产品)的	通送信息 可选		
* publisher_name		_	没有机构名称 每个publishe		+77	
* publisher_place			NATIONS BOARD		475	
* del		_	2,03	ac:####################################	ol.data1.7-@@li	1255个课册
	时:DOI编码规范PD	F文件 D	01可用编码查问			
Smestamp	20161022200721					
resource			HOOMEMBURL - HEAD	2県本的いたい3社。 毎1	hdoi,data-t, 29	\$E21255个中符
body dataset				BØł		B.8 35/6
"contributor_name			计处境局由任有任务			
"contributor_role	(author)作者	٠	此此城可度任會角色			
*contributor_sequence	(first)第一类任者	•	的位于中国中国的中国中国中国中国中国中国中国中国中国中国中国中国中国中国中国中国中国			
*organization_name			此处城阳景任南州在杭构 羽	8		
"organization_role	(author)作者	•	此功值可要任者所在机构条件	8 双烟		
		2	11-11-10-11-11-11-11-11-11-11-11-11-11-1	10.00 27.0	重要16.10	Http:
*organization_sequence	(first)第一责任者	•	here a set of an other set	Dents - 2 Mar		
*organization_sequence	(first)第一责任者		NUMBERS			

Standardized data collection format ---Data quality of experimental / computational / industrial data

有色金属材料及特种合金 黑色金属材料 复合材料	材料名称: 计算软件:(BaGa4S7 CASTEP:5.	5					
有机高分子材料 无机非会属材料	计算结束 化学师37							
- 信息核科		成分包称			ACCORD.			
 ■ 能源料料 ● 生物医学材料 ● 大然材料及用品 		Ba			8.3%			
		Ga			33.3%			
建筑材料		s		58.3%				
過的交通材料	Read							
	#140.509	40	00021					
	晶体尺寸 (m	m=mm)	Pures 1					
	品把参数>	(nm)	1.4755					
	品名を約5	(nm)	0.6228					
计算数据	最终争取:	(nm.)	0.5929					
第一性原理	載肥伊数 (*)		90					
分子动力学	Ameas(*)		90					
计算法力学	ARP\$3(*)		90					
1150 A) / J 77 77	2		2					
family and a	講題(IR(nm ¹) 0.5448							
性能機能	THEAR							
加工工艺模型	性影樂明	128042.65	性影響	12854402	用成设备与型号	检测机构名称		
0.00111104	4.918	ACTED IN	3.31	ev				
随助工具 ① 1. 18元素第1月元表現	元学性期	5974	d31=-6.21;d32= -4.86;d33=16.66	pm√V				
△ 2、未要检索功能	光穿性和	所知年	delta(n)=0.063					
◎ 3. #568abe 河 4. 98月1月	THERE AREA							
ed.	THERE							
	计网络入约计							
	余件类的		##X8	*83		##12		
	初始条件		296/29682	LDA				
	10k9/1		885	norm-consi	erving			
	4700余年		平衡波频止皖属	850		ev		







The foundation of materials database in China

I. Materials Property Data - Materials Scientific Data Sharing Network

II. China gateway to corrosion and protection

New advances in Materials Genome Initiative database and big data

III.Corrosion Platform

- I. Material Genome Engineering Database and Big Data Technology
- II. Data acquisition and database fusion technology of material structure and property based

on high-throughput experiments and calculations



- □ Corrosion is classified as microscopic corrosion and macroscopic one.
- Microscopic corrosion science, based on basic data, consist of the analysis of corrosion phenomena, the establishment of corrosion theory, the development of anti-corrosion technology.
- \checkmark Macroscopic corrosion is the destruction process of all structures.

Academician Jimei Xiao established the master degree of corrosion discipline.



China gateway to corrosion and protection

 出京等故大学 University of Science and Technology Beijing

- 378 experts on corrosion across China contribute to the work
- Nonprofit, open free of charge,
- Provides services for more than **400 institutions**, from industries such as aerospace, petrochemical, electric power, electronics, machinery, iron and steel, ships and offshore structures, transportation and so on.



R&D Center and Field Research Stations

One R&D center, 30 stations across China, including **15 atmospheric stations**, **7 water stations** and **8 soil stations**.

Around 270 million RMB was invested to built over **1,600 corrosion analysis facilities.**



Atmospheric Corrosion Testing Stations







Materials Corrosion "Big Data"

- The corrosion database was built from the long-term corrosion tests of 654 kinds of materials (metals, polymers, construction materials and coatings), a total of 140,000 specimens under various environments.
- Award of National Science and Technology Progress in 2009 and 2016 were obtained. The research results were published in Nature.



Asian Materials Data Symposium (AMDS)

- USTB called on and successfully established Asian Materials Data Committee (AMDC) in 2010, between Korea, Japan and China, and then Vietnam, Russia, Indonesia.
- AMDS 2018
- Asian Materials Data Symposium (AMDS) is held every 2 years.

The 6th Asian Materials Data Symposium (AMDS) Nov, 2018, Beijing, China.



Jeju, Korea



Tokyo, Japan



Sanya, China



Jeju, Korea MRDC @Rockville, MD 9/25-27,2017





The foundation of materials database in China

I. Materials Property Data - Materials Scientific Data Sharing Network

II. Materials Corrosion Data - National Environmental Corrosion Platform

New advances in Materials Genome Initiative database and big data

I. Material Genome Engineering Database and Big Data Technology

II. Data acquisition and database fusion technology of material structure and property based

on high-throughput experiments and calculations











High throughput calculation and experiment

- 1) Atomic potentials database construction
- 2) Typical material thermodynamics and dynamics database construction
- 3) High throughput DFT driven engine construction
- 4) Capture and analysis of high throughput testing data

Database Construction for Material design

- 5) First-principles calculation database
- 6) Composition-microstructure-property database

Data mining and application technology

- 7) Intelligent software platform for data mining
- 8) 3D microstructure reconstruction technology
- 9) Application research of materials data





Materials Database:

Energy materials / Rare earth materials / Catalytic materials / Biomaterials / Alloys

1 Data processing technology of high throughput material calculation

□ Automatic data archiving technique and a driven engine for high throughput DFT.

University of Science and Techn



Data capture and integration technology for high-throughput heterogeneous experimental data, and image processing technology.



Integrated analysis of heterogeneous data



microstructure image processing and quantitative characterization



Synchrotron radiation / X-ray diffraction data processing



Experimental data processing

3 MGI database infrastructure, standards and the database



- □ MGI database based on distributed NoSQL database architecture.
- Material data ontology technology.
- Standards & specification of Materials data .





Build the analysis algorithms and models of materials data mining.
 Develop an intelligent software platform for data mining.





Large-scale 3D reconstruction & application of material microstructure.
 Interactive, multi-scale and multi-objective experiment and simulation.



MGI Database Platform Architecture



- Template based data schema design and data submission to satisfy personalized data archiving.
- Cloud service environment based on Hadoop, no-schema data storage based on MongoDB, flexible extension based on Json.



MGI Database Platform Architecture



Template design

牌号		文本型	
	数据类型 文本型	数值型	
密度		范围型	
	数据类型 数値型 数据单位 g/cm3	文件型	
7.随率		多值数据	
	数据类型 数值型 数据单位 %	表格型	
		混合型	
制备工艺	5 双据		
界面混合	旬制备工艺 合型数据	Basic	ι
	齐国厚度 数据类型 数值型 数据单位 nm	Assemb	ole
	外面厚度 数据类型 数值型 数据单位 0m 界面热处理 数据类型 文本型	Assemb	5]6
	外面修葺 数据类型 数值型 数据类型 约届单位 nm 界面熱处理 炎本型 界面制备工艺参数	Assemb	5]e
	外面均在 放面少正 数据类型 数据型 外面為处理 数据类型 数据类型 文本型 外面制备工艺参数 数据类型 数据类型 文本型	Assemble templates	5]e
基代	外面均度 数据类型 数值型 数据类型 約個型 のm 界面熱处理 次本型 弊面制备工艺参数 公本型	Assemble templates hrough dragging and tropping basic unit	
基体温含		Assemble templates hrough dragging and lropping basic unit	d
		Assemble templates hrough dragging and lropping basic unit	ole

Data sheet

陶瓷基复合材料示例模板 • 生成	法单 生成模板文件
· 提交数据文件 选择 ▶ 提迟	数据文件
牌号	
牌号	
密度	
密度	g/cm3
孔脉率	
孔旗率	%
制备工艺	
界面制备工艺	
界面厚度	
界面厚度	nm
界面热处理	
界面热处理	
界面制备工艺参数	
界面制备工艺参数	
基体制备工艺 D	ata sheet generated by
基体成分 as	sembling
52.00 (PH)	
基体热分	
基体的 公理 基体特处理	
基体机会学 基体机会理 基体机备工艺参数	



Data Conversion Interface

- **Conversion interface:** parse user data into **JSON** format
- **Data export:** transform **JSON** data into template file
- □ Application program interface: provide API for accessing database and database applications(big data analysis, data mining, etc.)





MGI Database Platform/ Atomic potential database

	Flome	Semicon	ductors, Ionic C	Crystals, Interface, Metal Allo	ys and C	Others		
	Lieme	Grouped by the num	Potentials #1 - description 1	for kcl			edit delete	
Ν	1 <mark>1</mark> 1 H	Semiconduc	Scope	Potentials #2 - description for kcl				edit dele
This	2 3 4 LI B	Ionic Crystal	Ionic Crystals	Scope				
to fe	3 11 1	RbCI	Function Formula	Ionic Crystals				
Lei	Na M	Interface	$(1) \phi_{1,1}(r_{e}) = \frac{q_{eff}^2}{2},$	Function Formula	Function Pa	rams		
	4 K C	NI/AI2O3	$(2) \phi (r_{e}) = D \begin{cases} e \\ e \end{cases}$	$\phi^{SR}(r) = Dexp\left[-\gamma\left(\frac{r}{R}\right) - 1\right]$			Cla	sses
	5 37 3 Rb S	Metal Allovs	-200	- 117 - I	#	Params	3	•-
	6 55 5	Re-Re	- 2 exp		1	D	1.7	149
	Cs B	0.1	(a) $\phi(r) = D$ ex		1	R	2.3	383
_	7 Fr R	Others	() ++=0 <i>y</i> /-2+=0		1	Ŷ	6.4	557
		Function		Function Formula	Function Pa	rams		
- 12 F		Grouped by the num		$\phi^{SR}(r) = D\left\{ exp\left[-\gamma\left(\frac{r}{R}-1\right)\right] - 2\left[-\frac{\gamma}{2}\left(\frac{r}{R}-1\right)\right]\right\}$			Cla	sses
_	1	F			#	Params	5	
_		Function Typ	Reference		1	D	0.1	177
		morse - morse functi	Bloom, B. S., & Sosniak, I		1	R	3.7	066
	l	Rep-Exp - descriptio	Other Properties		1	Y	8.8	093
			No additional properties.	Reference				



The foundation of materials database in China

I. Materials Property Data - Materials Scientific Data Sharing Network

II. Materials Corrosion Data - National Environmental Corrosion Platform

New advances in Materials Genome Initiative database and big data

I. Material Genome Engineering Database and Big Data Technology

II.Data acquisition and database fusion technology of microstructure and

property based on high-throughput experiments and calculations

Data acquisition and database fusion technology of material structure and rechnology being property based on high-throughput experiments and calculations

13th five-year plan: National Key R&D Plan Program 2017.07-2020.12



1 High-throughput experiments, calculations, and knowledge extraction at micro and nano scale



theory of reliable access and knowledge extraction for heterogeneous, dynamic and mass data
 Methods for prediction & verification of mechanical properties.



2 Materials Interface Electrochemistry high-throughput experiments, calculation and analysis techniques



In-situ high-throughput collection of interface electrochemical information
 Establish automatic analysis system for high throughput electrochemical data



- The experiment is invariant while the data quantity increases more than 10 times
- Simultaneous acquisition of electrochemical and image data

Electrochemical high throughput computational simulation



Propagation dynamics of protons and ions in microchannels and its interface electron transfer properties

3 Image acquisition and recognition technology of high precision multiscale surface/interface of material structure



- □ Preprocessing and registration rules of image data, establish models for multilevel fusion of images
- □ Material surface/interface image retrieval recognition technology
- **D** Build a dynamic simulation system for material image data



Surface/interface image retrieval and simulation





3D refactoring and dynamic simulation of microstructure MRDC @Rockville, MD 9/25-27,2017





- □ Distributed analysis, deep mining technology, automatic storage
- □ Microstructure performance data acquisition and data fusion support system

MGI Big Data All-in-One (AiO) Machine



- Machine learning methods such as convolutional neural networks, Bayesian networks and Support Vector Machine
- Methods of material evaluation by combining model evaluation with experimental verification

Closure





Complexity, flexibility, and scalability



Acknowledgement

- State Key Project of China (2016YFB0700503)
- National High Technology Research and Development Program of China (2015AA03420)
- Beijing Science and Technology Plan (D16110300240000)

THANKS FOR YOUR ATTENTION!







