

Materials R&D Data Schema and Standard Vocabulary

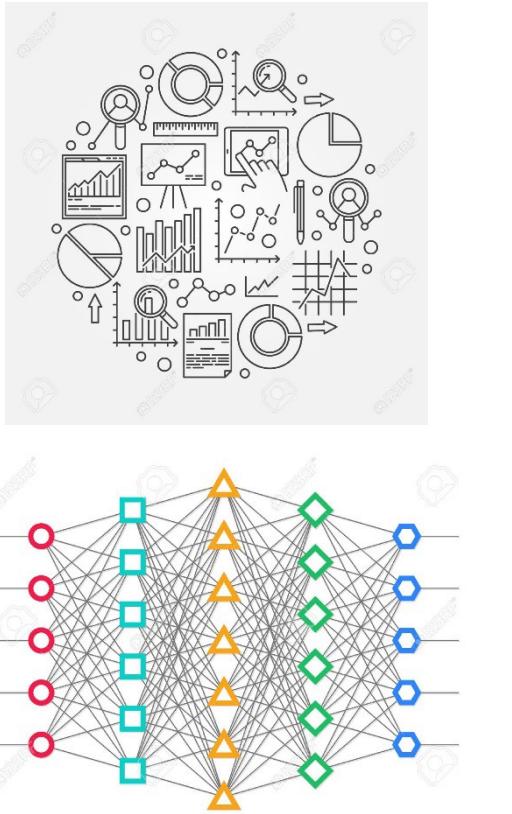
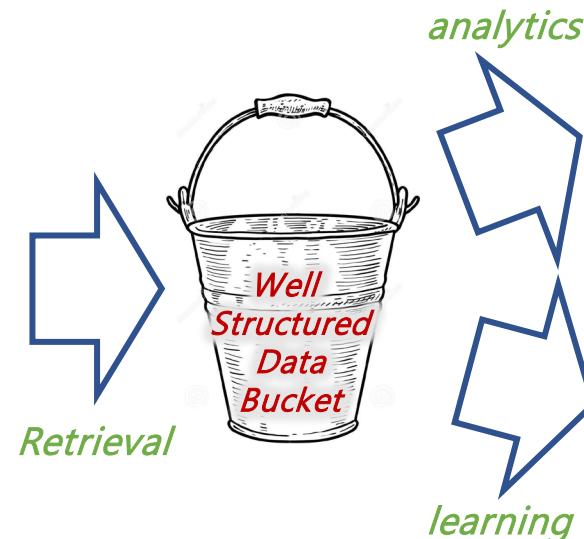
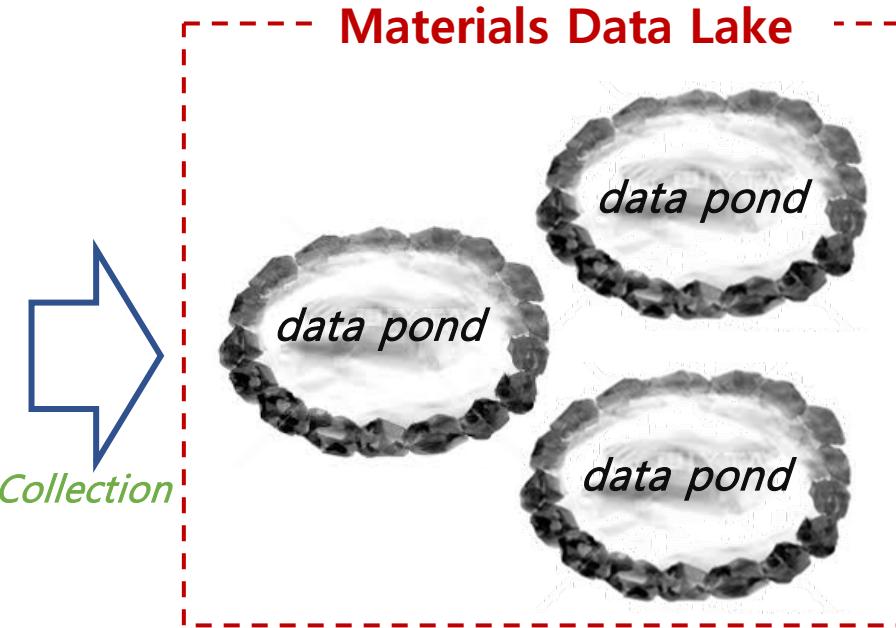
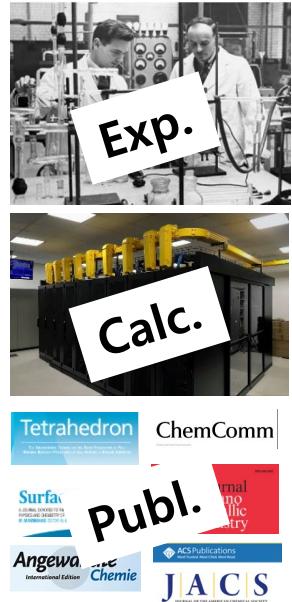
Kwang-Ryeol Lee (krlee@kist.re.kr)

Korea Institute of Science and Technology, Seoul, Korea

Committee for Standardization of Materials R&D Data

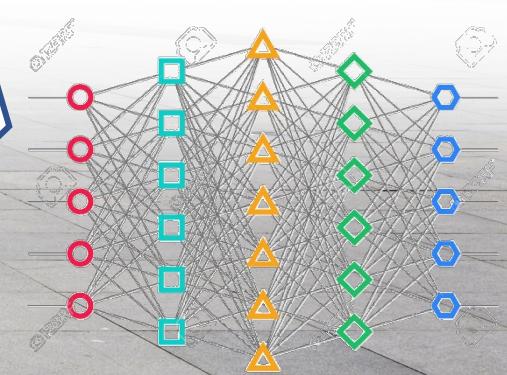
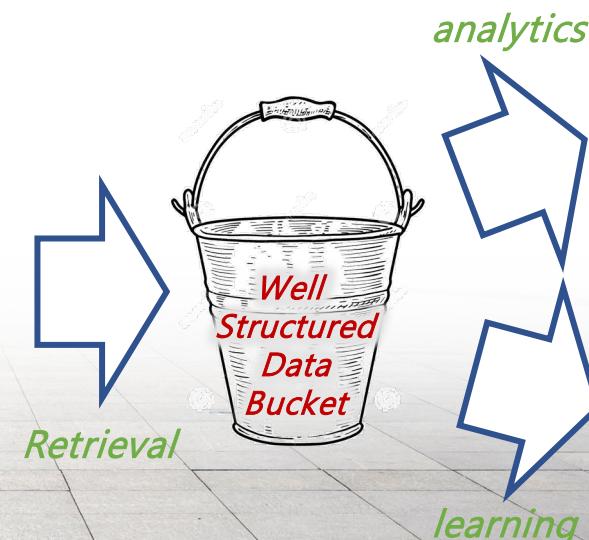
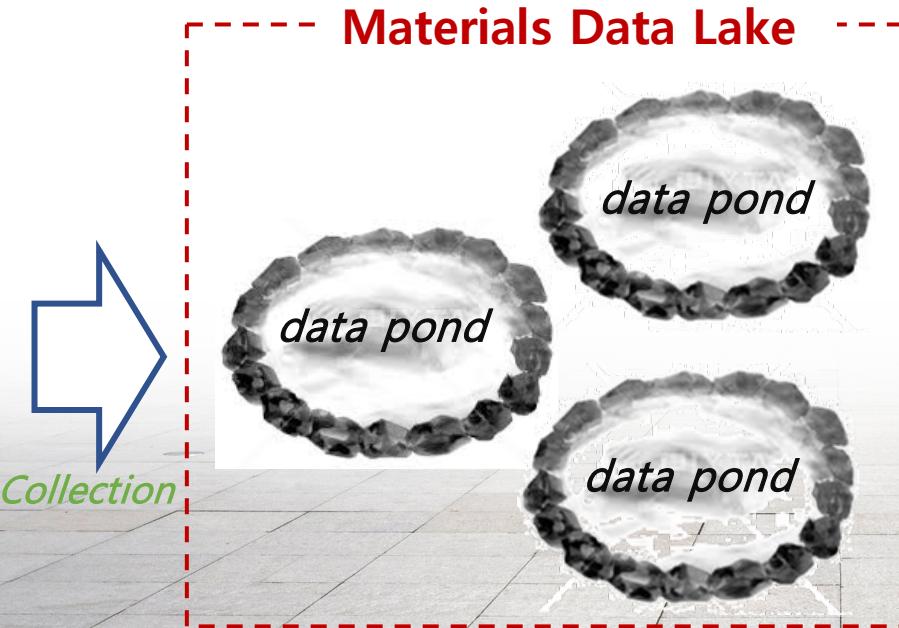
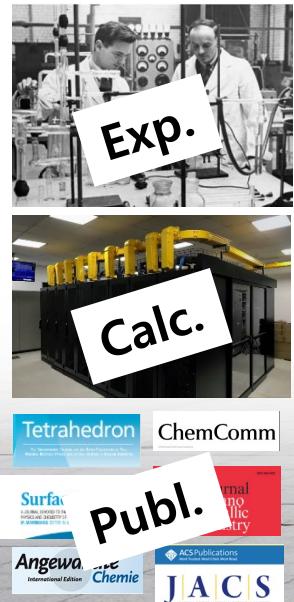
Chair	이광렬	한국과학기술연구원
Energy & Environment Materials	이정훈	한국과학기술연구원
	김지한	한국과학원
Smart IT Materials	박선화	한국표준연구원
	이동화	포항공과대학교
Structure and Safety Materials	김수현	한국재료연구원
	김정한	한밭대학교
Consulting	안재평	한국과학기술연구원
	장현주	한국화학연구원
	한승우	서울대학교
	허용학	한국표준연구원
	한상수	한국과학기술연구원
	신호선	한국표준연구원
	오창석	한국재료연구원

Data Driven Materials Design



- Collection of R&D Data from Various Sources
- Data Storage and Maintenance
- Data Driven Knowledge

Rock Base for the Data Driven Materials Design



Standard Data Scheme and Keywords

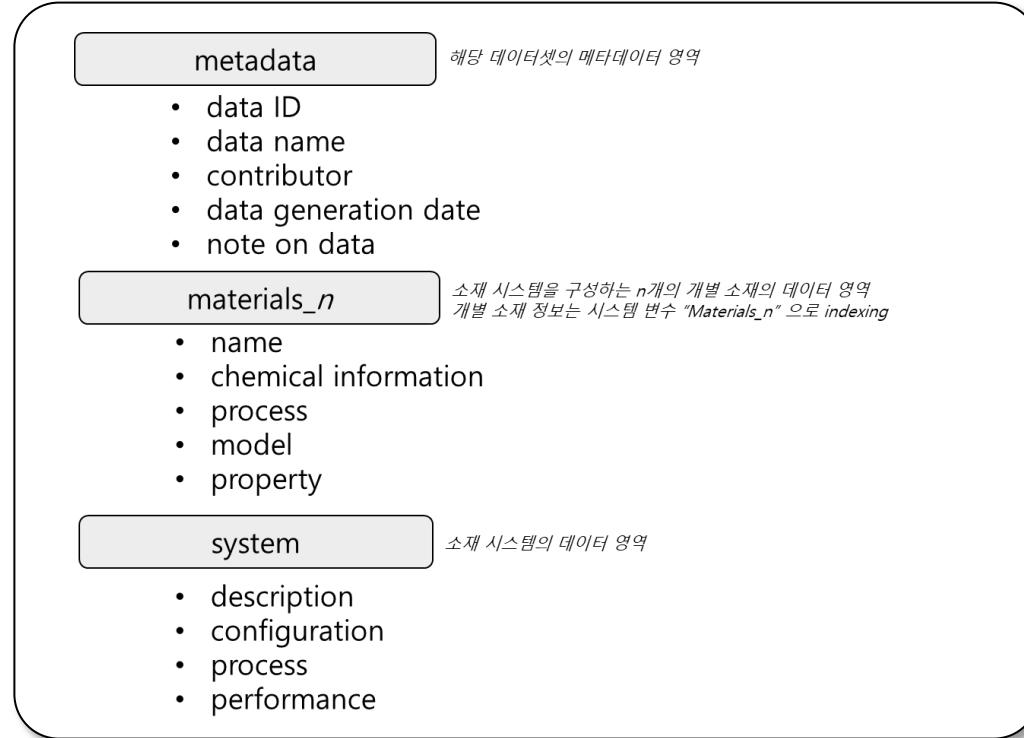
Dictionary of Vocabulary for Materials Research Data

Set /set/ **verb** (sets, setting; past and past participle set)
set¹ ▶ **verb** (**sets, setting**; past and past participle set)
1 [with obj. and usu. with adverbial] put, lay, or stand (something) in a specified place or position: Delaney set the mug of tea down | Catherine set a chair by the bed.
■ (**be set**) be situated or fixed in a specified place or position: the village was set among olive groves on a hill. ■ represent (a story, play, film, or scene) as happening at a specified time or in a specified place: a private-eye novel set in Berlin. ■ mount a precious stone in (something, typically a piece of jewellery): a bracelet set with emeralds. ■ mount (a precious stone) in something. ■ Printing arrange (type) as required. ■ Printing arrange the type for (a piece of text): article headings will be set in Times fourteen point. ■ prepare (a table) for a meal by placing cutlery, crockery, etc. on it in their proper places. ■ (**set something to**) provide (music) so that a written work can be produced in a musical form: a form of poetry which can be set to music. ■ Bell-ringing move (a bell) so that it rests in an inverted position ready for ringing. ■ cause (a hen) to sit on eggs. ■ put (a seed plant) in the ground to grow. ■ Sailing put (a sail) catch the wind.
can henceforth be set, someone apart, superiority: his ability set him apart. set thing and keep it books and rooms aside 1 save or for time, for a to set aside at remove la a legal deci back 1 de or somet research a partic you ba some stop set ■ e lo a ;

Standard Vocabulary & Grammar for Materials R&D Data

What to be Standardized?

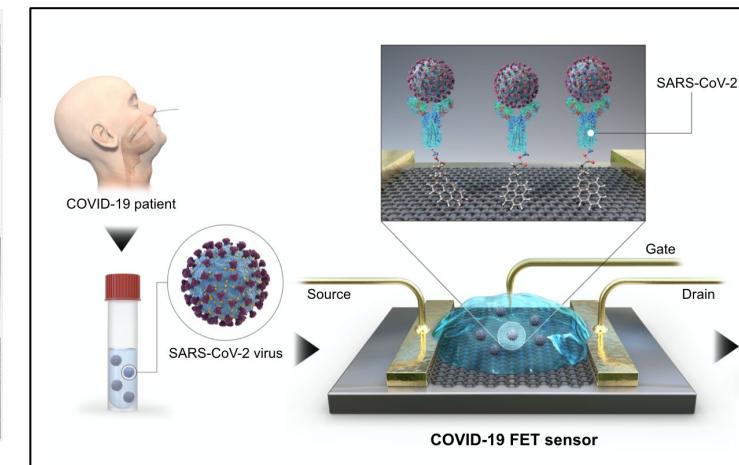
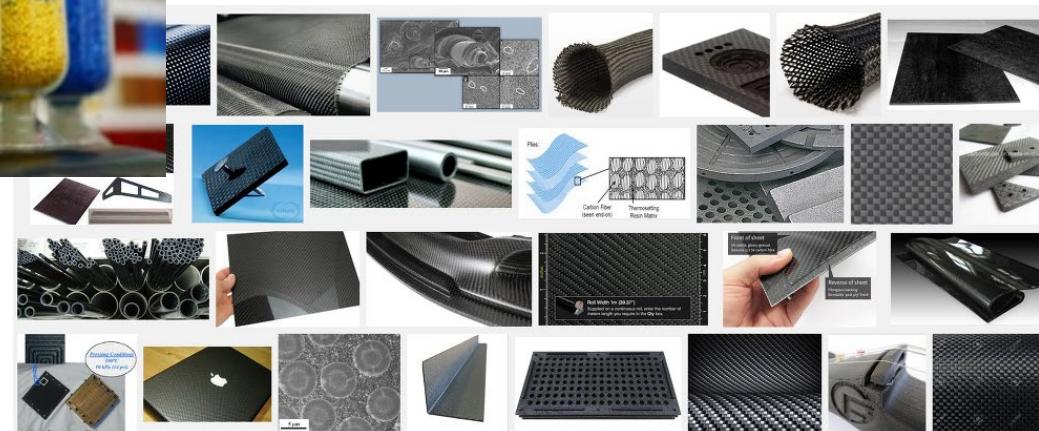
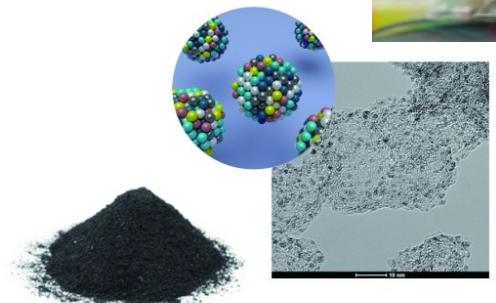
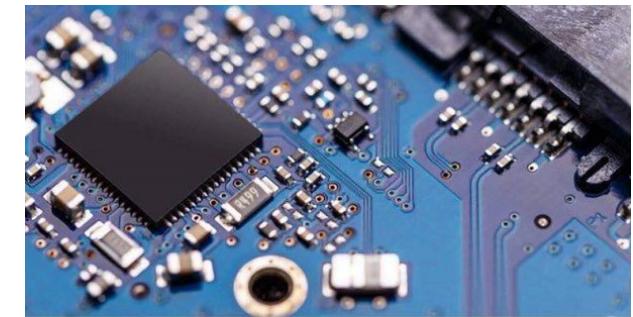
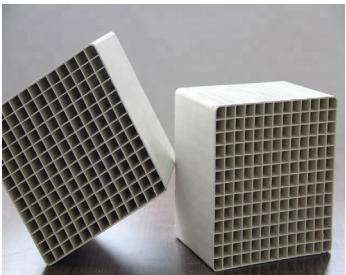
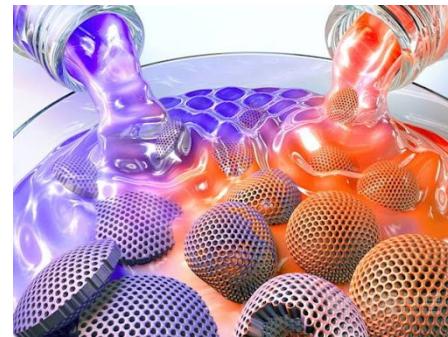
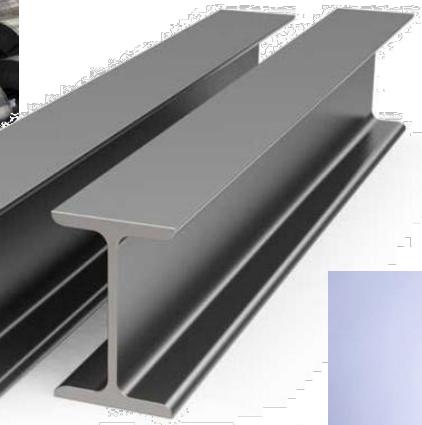
• Structure of Materials R&D Data : Grammar (FAIR strategy)



• Data Keywords in the Standard Data Structure : Vocabulary

경도 = 65 GPa {"hardness":65}

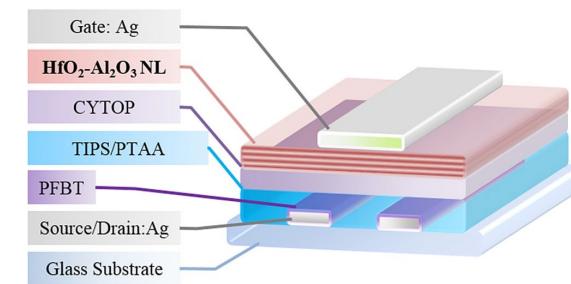
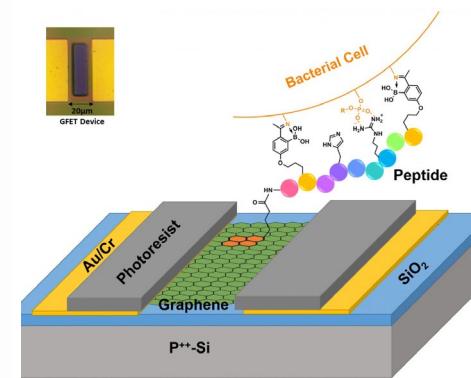
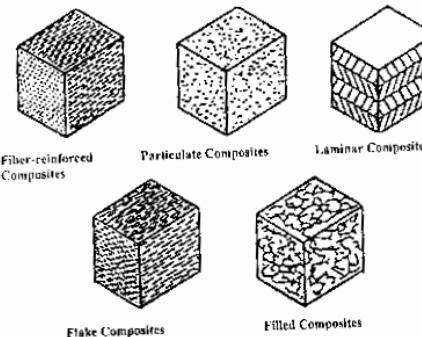
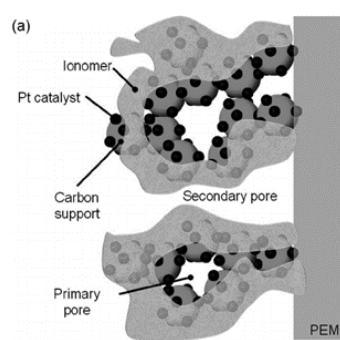
Data Structure for Wide Range of Materials



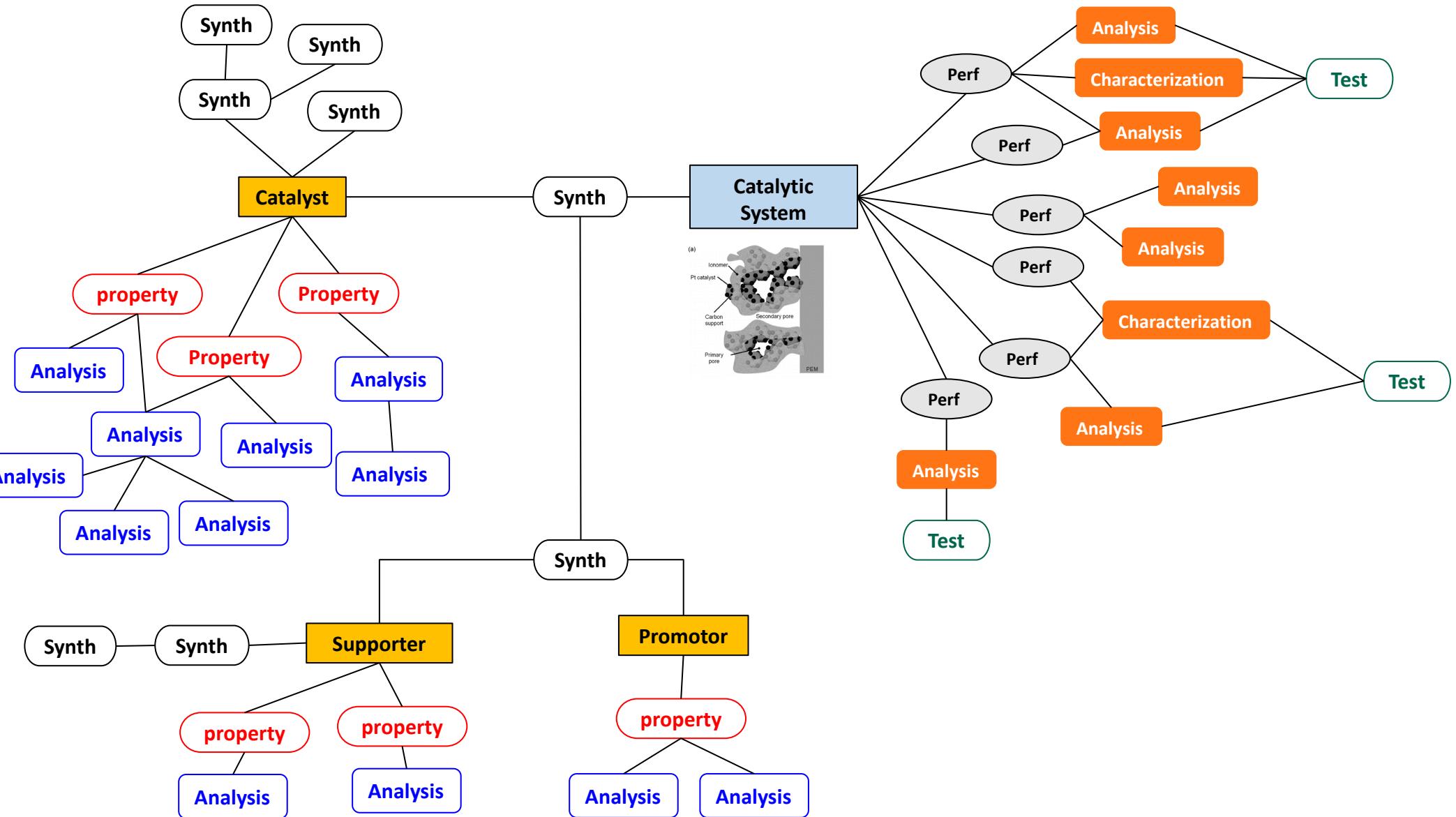
Introduction to Materials System

- System composed of one or more materials to characterize the performance of the material.

Materials System	Materials 1	Materials 2	Materials 3	Materials 4	...
Catalytic Materials	Catalyst	Supporter	Promotor		
Composite Materials	Fiber	Matrix			
Sensor Materials	Sensor	Electrode	Substrate	Protective Layer	
Cathode in Secondary Battery	Cathode 1	Cathode 2	Electrode	Electrolyte	
Chrome Coating Steel	Substrate Steel	Cr Coating Layer			
Special Steel	Special Steel				



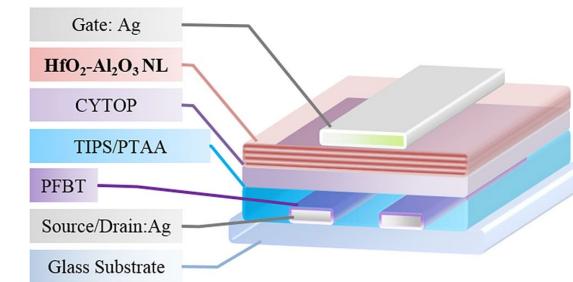
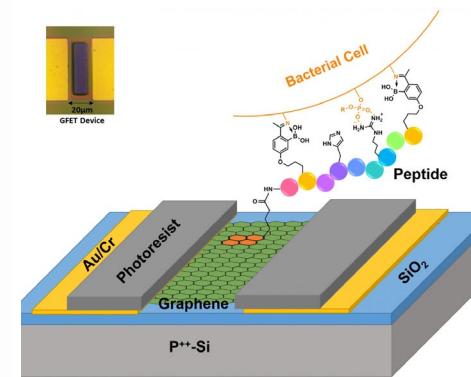
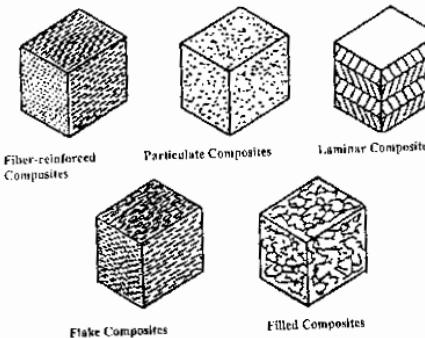
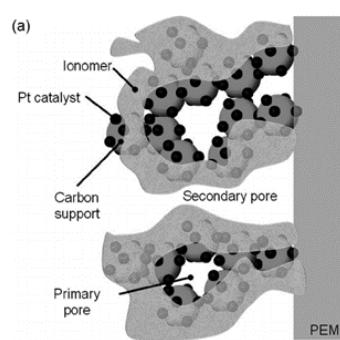
Data Structure Example : A Catalytic Materials System



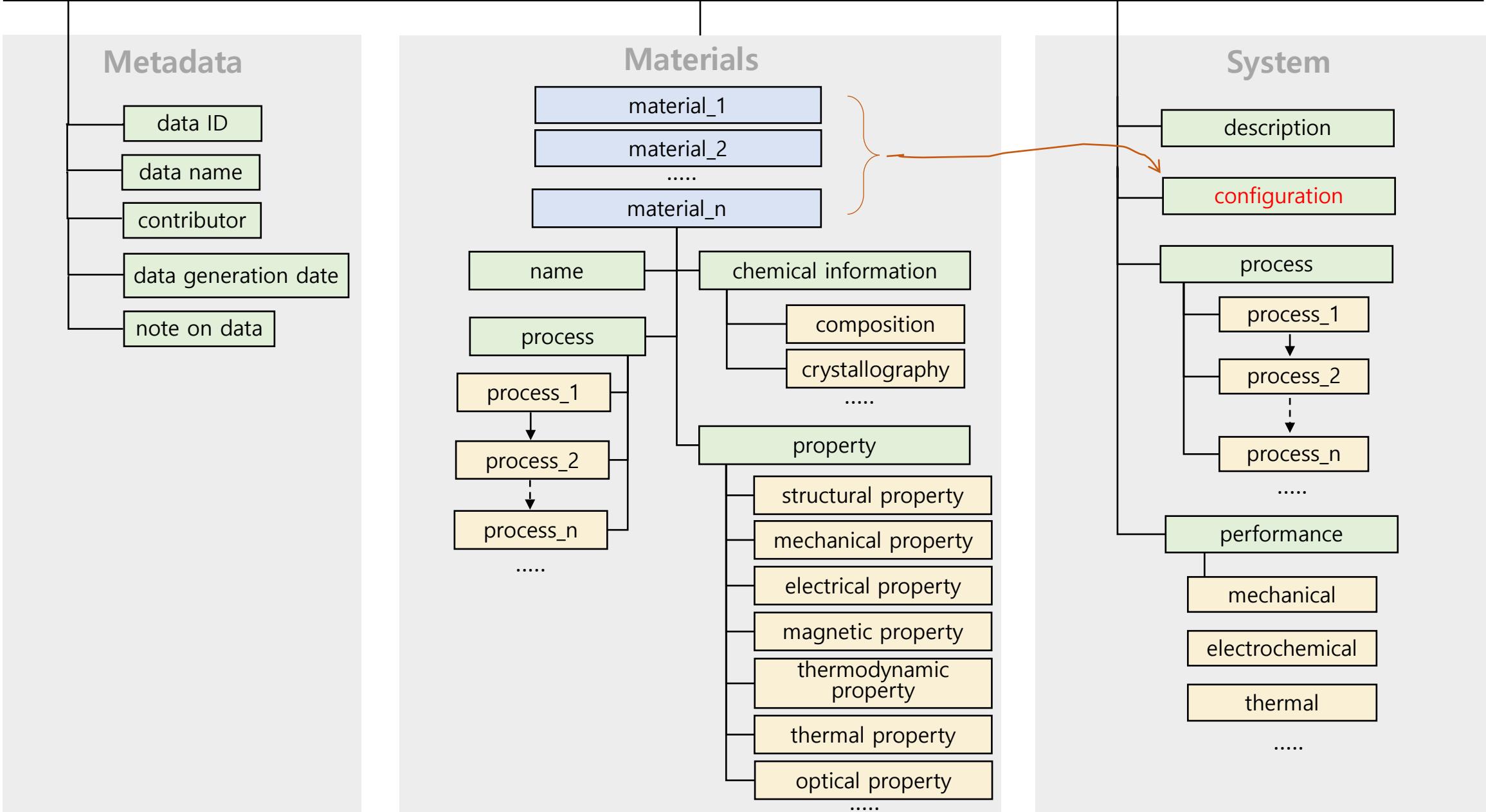
Introduction to Materials System

- System composed of one or more materials to characterize the performance of the material.

Materials System	Materials 1	Materials 2	Materials 3	Materials 4	...
Catalytic Materials	Catalyst	Supporter	Promotor		
Composite Materials	Fiber	Matrix			
Sensor Materials	Sensor	Electrode	Substrate	Protective Layer	
Cathode in Secondary Battery	Cathode 1	Cathode 2	Electrode	Electrolyte	
Chrome Coating Steel	Substrate Steel	Cr Coating Layer			
Special Steel	Special Steel				



Materials System Data



General Structure of Materials Data Set

metadata

- data ID
- data name
- contributor
- data generation date
- note on data

Meta data section of the data set

materials_n

- name
- chemical information
- process
- model
- property

*Data section of each n materials constructing the materials system
Each materials' data is indexed by the system variable "Materials_n"*

system

- description
- configuration
- process
- performance

Data section of Materials System

```
0 {4}
  meta {5}
    data ID : DOI:111.222.333.444
    data     : Pt3Ni particle on graphite support data
    name    : 1
  contributor {2}
    data generation date : 2020-04-28
    note on  : information of material_2 comes from
    data     : Aldrich catalog\n10.1021/n1401881z
  material_1 {4}
    name : Pt3Ni
    process [3]
      0 {1}
        solvothermal {8}
      1 {1}
        heat treatment {10}
      2 {1}
        centrifugation {5}
    property {1}
  material_2 {2}
    system {4}
      description : Pt3Ni/C catalyst system for CO2 reduction
      configuration {3}
        active material : Material_1
        amount of active material : 0.003
        support material : Material_2
      process [3]
      performance {1}
        electrochemical {1}
          ORR {3}
            area-specific activity {3}
              value : 49
              uncertainty : 0.1
            measurement {1}
              electrochemical activity {5}
            mass-specific activity {3}
            Faradaic efficiency {3}
```

Contents of Standard Vocabulary Dictionary

▼ object {6}
▶ metadata vocabulary {4}
▶ materials vocabulary {1}
▶ system specific vocabulary {3}
▶ common vocabulary for analysis method {18}
▶ common vocabulary for materials process {29}
▶ numeric data expression {2}

3 vocabulary groups
 2 common vocabulary groups
 1 numeric data expression rule

이휘군	category 1	category 2	category 3	electrode..1
metadata vocabulary	data name			buffer..1
	contributor	name		active layer
		affiliation		buffer..2
	data generation date			electrode..2
materials common vocabulary	note on data			(process..n)
		name		current
				endurance
		chemical information	composition	mechanism
			crystallography	operating speed
			SMILES	resistance
		process	(process..n)	retention
			elements	selectivity
			state	voltage
		model	box dimension	measurement
			periodic boundary condition	
			structure file	
			structural property	creep test
system specific vocabulary			mechanical property	DET
			electrical property	electrochemical activity
			magnetic property	empirical MD
		property	thermodynamic property	fatigue test
			thermal property	gas chromatography
			optical property	hardness test
			corrosion property	impact test
				infrared spectroscopy
		description		memristive activity
	system_catalyst		active material	nuclear magnetic resonance
			amount of active material	optical microscopy
		configuration	promotor	TEM
			amount of promotor	tensile test
			support materials	thermal activity
	process		(process..n)	thermogravimetric
			electrochemical	x-ray diffraction
		performance	thermal	
system_porous_materials	description			centrifugation
	configuration		framework..(n)	chemical mechanical polishing
			inserted molecule	chemical synthesis
	process	(process..n)		Czochralski growth method
	performance	gas adsorption		drying
system_memristi	description			electrochemical deposition
				epitaxy
				etching
				exfoliation
				thermomechanical process

Keyword format in the dictionary (in json format)

```
{"JSON_Name" : {"kor_definition":string,  
"eng_definition":string,  
"alias":string,..],  
"data_type":string (python),  
"data_unit":string (Latex),  
"data_example":string}
```

어휘 사전의 작성 규칙 (2021.10.19.)

1. 본 소재 데이터 표준어휘 사전은 json 포맷을 사용하여 소재 데이터의 구조와 소재 데이터에서 사용되는 keyword 들을 정의한다.

2. 각 데이터 keyword 들은 아래의 정보를 담고 있다.

- eng_definition : keyword의 영문 설명
- var_name : keyword의 DB 시스템 변수명
- alias : keyword와 동일한 의미로 사용되는 용어
- data_type : keyword에 해당하는 데이터의 형태
- data_unit : keyword에 해당하는 데이터의 단위
- data_example : keyword에 해당하는 데이터의 예시

(중요) 이를 명칭을 이용하여 parsing한 어휘 사전의 정보를 바탕으로 플랫폼이 구축된다.
따라서, 위 명칭들은 어휘 사전에서 오류없이 사용되어야 한다.

3. 국제표준어휘와의 연동을 위해 keyword는 모두 영문어휘를 사용한다.

4. 본 사전에 사용되는 모든 keyword는 고유명사를 제외하고 소문자로 작성한다.
예) chemical information: Czochralski growth method

5. keyword의 괄호로 표시된 영역은 실제 data로 대체된다. 같은 성격의 반복되는 키워드는
뒤에 _(n)을 붙인다.
예)

본 사전 내의 keyword	실제 데이터 예
materials_(n)	"materials_1"
precursor_(n)	"precursor_3"
(process)	"chemical synthesis"
(analysis)	"optical microscopy"
(unit)	"g/mole"

6. DB 시스템 변수명은 keyword의 단어를 대문자로 시작하는 단어로 변환하여 붙여서 구성
한다. 다만 필요할 때는 단축어를 사용할 수 있다.

예) "structural property" → StructuralProperty
"optical performance" → OptPerf

7. data_type으로는 아래의 것들이 있다.

- * string : 문자열 데이터
- * numeric : 수치 데이터

Keyword format in the dictionary (in json format)

```
{"JSON_Name": {"kor_definition": "string",
    "eng_definition": "string",
    "alias": "[string, ...]",
    "data_type": "string (python)",
    "data_unit": "string (Latex)",
    "data_example": "string"}}
```

▼ data name {4}

```
eng_definition : name of the data set, named by data provider
var_name : DataName
data_type : string
data_example : Zinc coated Fe-6Mn-3Al steel plate; Ti/Al dissimilar welded joint; Fe/C catalyst; Gd doped MOF 370; LLZO (Li7La3Zr2O12) Solid Electrolyte
```

```
{"data name": "Pt3Ni catalyst"}
```

어휘 사전의 작성 규칙 (2021.10.19.)

1. 본 소재 데이터 표준어휘 사전은 json 포맷을 사용하여 소재 데이터의 구조와 소재 데이터에서 사용되는 keyword 들을 정의한다.

2. 각 데이터 keyword 들은 아래의 정보를 담고 있다.

- eng_definition : keyword의 영문 설명
- var_name : keyword의 DB 시스템 변수명
- alias : keyword와 동일한 의미로 사용되는 용어
- data_type : keyword에 해당하는 데이터의 형태
- data_unit : keyword에 해당하는 데이터의 단위
- data_example : keyword에 해당하는 데이터의 예시

(중요) 이들 명칭을 이용하여 parsing한 어휘 사전의 정보를 바탕으로 플랫폼이 구축된다. 따라서, 위 명칭들은 어휘 사전에서 오류없이 사용되어야 한다.

3. 국제표준어휘와의 연동을 위해 keyword는 모두 영문어휘를 사용한다.

4. 본 사전에 사용되는 모든 keyword는 고유명사를 제외하고 소문자로 작성한다.
예) chemical information; Czochralski growth method

5. keyword의 괄호로 표시된 영역은 실제 data로 대체된다. 같은 성격의 반복되는 키워드는 뒤에 _(n)을 붙인다.
예)

본 사전 내의 keyword	실제 데이터 예
materials_(n)	"materials_1"
precursor_(n)	"precursor_3"
(process)	"chemical synthesis"
(analysis)	"optical microscopy"
(unit)	"g/mole"

6. DB 시스템 변수명은 keyword의 단어를 대문자로 시작하는 단어로 변환하여 붙여서 구성한다. 다만 필요할 때는 단축어를 사용할 수 있다.

예) "structural property" → StructuralProperty
"optical performance" → OptPerf

7. data_type으로는 아래의 것들이 있다.

- * string : 문자열 데이터
- * numeric : 수치 데이터

Numeric Data Examples

```
▼ thermal property {6}
    eng_definition : physical properties that a material exhibits upon the application of thermal forces
    var_name : ThermalProperty
    alias : [value]
▼ thermal conductivity {6}
    eng_definition : quantity of heat that passes in unit time through unit area of a substance
    var_name : ThermalConductivity
    alias : [value]
▼ value {5}
    eng_definition : data value
    var_name : DataValue
    data_type : numeric
    data_unit : W/(m.K) (watts per meter-kelvin)
    data_example : 12.3
▼ uncertainty {4}
    eng_definition : uncertainty of the value
    var_name : Uncertainty
    data_type : numeric
    data_unit : same as value
▼ measurement {4}
    eng_definition : mesurement method
    var_name : Measurement
    alias : characterization tools, analysis method
    ▶ (analysis) {2}
▶ thermal diffusivity {6}
▶ thermal expansion {6}
```

```
▼ thermal property {2}
    ▼ thermal conductivity {2}
        value : 12.3
        uncertainty : 0.02
    ▼ measurement {1}
        ▶ themal calorimetry {2}
```

Array Numeric Data

```
▼ array numeric data {4}
  format : Array data is composed of 'x_definition' and 'value_array' in the form of
  array of objects {x, x_uncertainty, value, and value_uncertainty}.
  Independent parameter of the data array, x, should be defined in the
  'x_definition' section in the form of array [definition, x_unit].
  'value_array' follows the 'x_definition'.
  ▼ x_definition {3}
    eng_definition : define the independent parameter of the array
    var_name : X_definition
    data_type : array string [definition, x_unit]
  ▼ value_array {4}
    eng_definition : array data value
    var_name : ArrayDataValue
    data_type : numeric array of {x, x_uncertainty, value, value_uncertainty}
    value unit : (unit)
```

x	x_uncertainty	value	value_uncertainty
130	0.5	0.8	0.001
180	0.5	1.3	0.001
240	0.5	2.2	0.001

```
▼ x_definition [2]
  0 : temperature
  1 : K
▼ value_array [3]
  ▼ 0 {4}
    x : 130
    x_uncertainty : 0.5
    value : 0.8
    value_uncertainty : 0.001
  ▶ 1 {4}
  ▶ 2 {4}
```

Process Data Construction

▼ common vocabulary for materials process {29}

▼ centrifugation {6}

eng_definition : materials separation process using centrifugal force

var_name : Centrifugation

► revolution per minute {5}

► time {5}

► temperature {5}

► additive_(n) {4}

► chemical mechanical polishing {12}

► chemical synthesis {7}

► Czochralski growth method {10}

► drying {6}

► electrochemical deposition {16}

► epitaxy {9}

► etching {6}

► exfoliation {7}

► thermomechanical process {13}

► hydrothermal growth method {8}

► ion intercalation {7}

► Kyropoulos growth method {10}

► lithogrtaphy_e-beam {12}

► lithography_photon {8}

▼ process [3]

▼ 0 {1}

► solvothermal {8}

▼ 1 {1}

► heat treatment {10}

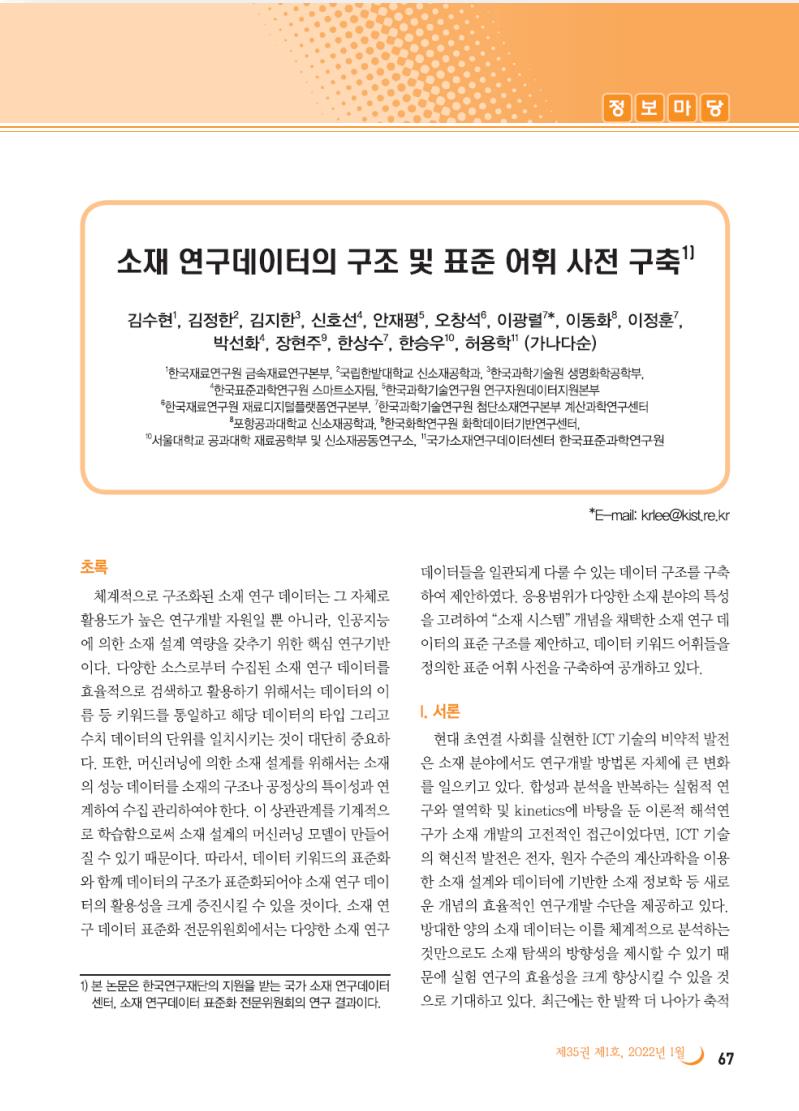
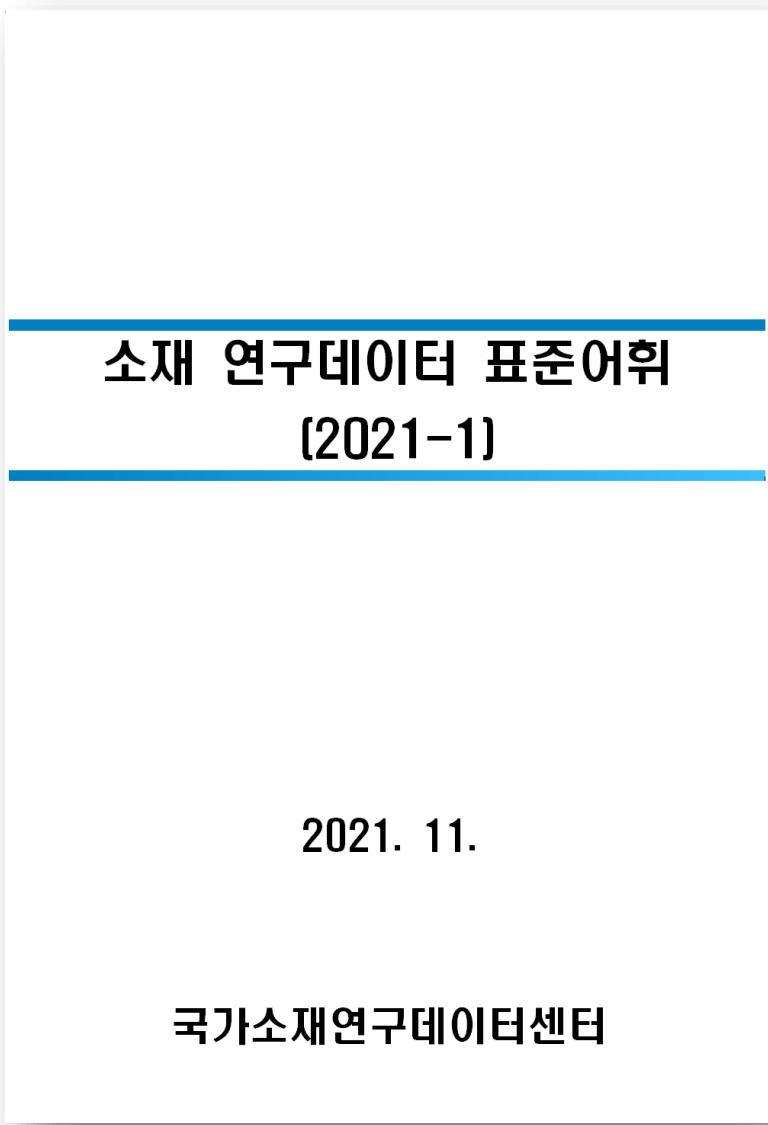
▼ 2 {1}

► centrifugation {5}

Measurement Data Construction

- ▼ common vocabulary for analysis method {26}
 - ▼ atom probe tomography {5}
 - eng_definition : Method for analyzing 3D chemical composition by atomic level
 - var_name : APT
 - instrument {4}
 - acquisition condition {7}
 - data {4}
 - creep test {5}
 - DFT {19}
 - electrochemical activity {9}
 - empirical MD {5}
 - EPMA {6}
 - fatigue test {11}
 - focussed ion beam {6}
 - gas adsorption isotherm {9}
 - gas chromatography {11}
 - hardness test {4}
 - impact test {4}
 - infrared spectroscopy {7}
 - memristive activity {9}
 - nano-indentation {6}
 - nuclear magnetic resonance {6}
 - optical microscopy {6}
 - Raman spectroscopy {6}

- ▼ crystallography {3}
 - Bravis lattice : cubic
 - ▼ lattice parameter {1}
 - a : 3.897
 - ▼ measurement [2]
 - ▼ 0 {1}
 - X-ray diffraction {2}
 - ▼ 1 {1}
 - TEM {2}



Korea Materials Data Station

The screenshot shows the K-MDS homepage with a dark blue background featuring a molecular or network structure. At the top, there are four navigation tabs: '소개' (About), '데이터' (Data), '분석' (Analysis), and '문서' (Documents). Below the tabs is a search bar with placeholder text '찾으시는 연구 데이터를 검색하세요.' (Search for research data). The main content area displays four circular statistics: '63 DataSet', '24K Data', '103 Users', and '62 Organization'. Below this is a section titled 'K-MDS Notice' with a link to 'more'.

K-MDS Notice

국가 소재 데이터 스테이션(K-MDS) 계산 자원 서비스 오픈 안내
2022-01-26

K-MDS 회원을 대상으로 계산 자원 서비스를 제공합니다. 계산 자원은 JupyterHub, Terminal, 시뮬레이션 워크벤치를 통해 활용할 수 있습니다. 1. 일정 - 2022년 4월 이후 정식 서비스 예정 - 자원 신청 방법 및 일정은 별도 공지 예정 2. K-MDS의 계산 자원 보유 현황 - Jupyter용 노드: 3개 노드 (96 코어) - 계산용 노드: 20개 노드 (640 코어) - 스토리지: 3PB

소재 연구데이터 수집·관리·공유·활용 가이드라인 및 표준 어휘 배포
2021-12-27

소재 연구데이터 수집·관리·공유·활용 가이드라인과 소재 연구데이터 표준 어휘를 배포합니다. K-MDS와 더불어 보다 나은 플랫폼 생태계 구축을 위해 첨부드리는 PDF 파일의 내용을 숙지하시면 플랫폼의 이해와 활용도 제고에 많은 도움이 될 것입니다. 감사합니다.



For Data Sharing in Large Scale

- Study the Standard Schema and Vocabulary of the counterpart
- Build Translator (API) between peer-to-peer level
- Build Translator to the global standard schema and vocabulary

