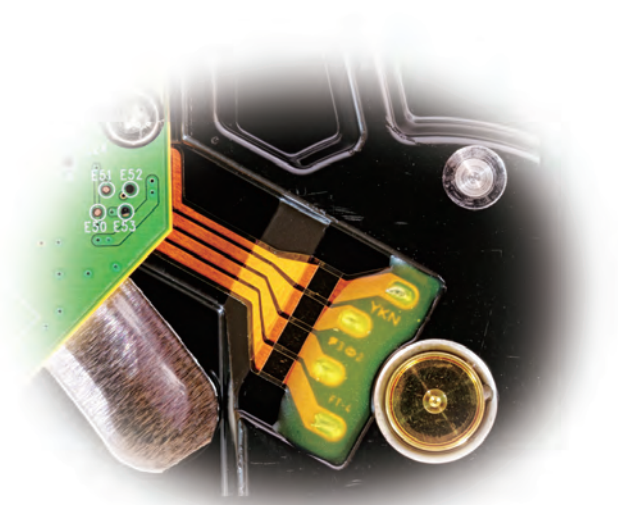


# 소재기술백서2017



The background features a complex network of molecular structures, including hexagonal and pentagonal rings, and several spherical clusters of atoms or molecules, all rendered in a light gray, semi-transparent style against a white background.

# 제4장 빅데이터 이용 소재 개발

1. 빅데이터와 소재연구
2. 소재정보학 기반 소재설계기술
3. 금속소재 가상공학 플랫폼 기술



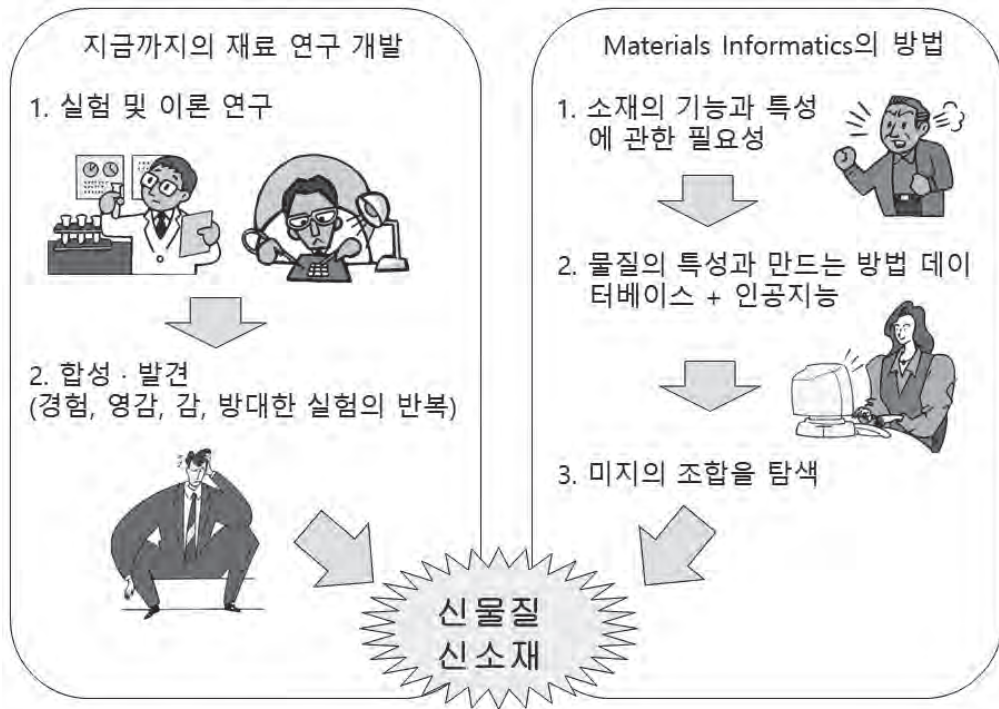
# 1. 빅데이터와 소재연구

## 1. 서 론

### 1.1 데이터 기반의 소재 연구

현대 초연결 사회를 실현한 ICT의 비약적 발전으로 인해 연구개발 데이터의 축적과 활용을 통한 연구개발이라는 새로운 장이 열리고 있다. 과학적 연구탐색의 패러다임이 실험(experiment), 이론(theory), 모사(simulation)에서 데이터 기반(data-intensive)의 연구로 전환되고 있는 것이다. 자연과학과 정보과학이 성공적으로 만난 대표적인 사례는 생물정보학(Bio Informatics)으로서, 2000년대 초반 인간 유전체인 지놈에 대한 30억 개 염기서열 해독이 완료(Human Genome Project) 되면서 이 정보를 이용한 질병의 진단, 치료 및 신약개발이 활발히 진행되고 있다.

데이터 기반의 소재 연구(소재정보학: materials Informatics)란 계산 과학에 의한 특성 예측과 이를 실증하는 고속(high throughput) 합성 및 평가, 그리고 소재물성 데이터베이스와 기계 학습 등을 통합적으로 활용하여 신소재를 탐색하고 설계하는 연구개발 활동 전반을 말한다. 실험 및 계산에 의해 얻어진 물질 정보와 데이터들을 통계적으로 분석함으로써 소재의 구조와 물성을 결정하는 핵심인자를 파악하고, 이를 통해 새로운 소재를 빠른 속도로 탐색할 수 있는 연구 방법론이다. 소재정보학의 궁극적인 목표는 이론 연구자가 정보학에 의해 파악된 핵심 인자로부터 재료 특성을 지배하는 법칙을 발견하고 재료 설계를 가능케 하는 체계적인 접근 방식을 구축하는 것이라고 할 수 있다.



〈그림 3-4-1-1〉 기존의 재료연구개발과 소재정보학에 의한 재료개발의 차이점

위 그림은 소재정보학과 기존의 소재개발 전략의 차이점을 그림으로 표현한 것이다.

컴퓨터와 계산 방법의 발전으로 인해 실험과 이론을 융합하는 계산과학은 소재 연구의 새로운 수단으로 중요한 위치를 차지하고 있다. 그러나 최근에는 이러한 실험, 이론, 계산의 모든 연구 데이터들을 활용하려는 데이터 기반 연구의 중요성이 크게 인식되고 있다. 그 배경으로는 제일원리 계산에 의한 물성 데이터베이스의 구축, 대형 시설과 첨단 실험 장비를 이용한 효율적인 합성 및 분석기술의 발전 등을 들 수 있다. 시뮬레이션과 유사한 조합실험법이나 데이터베이스와 융합시킨 ‘데이터 과학’기법을 신소재의 탐색에 적극 활용함으로써 반복되는 실험과 계산을 거쳐 결과를 얻어내는 과정을 생략할 수 있고, 따라서 신소재 개발과 활용까지의 소요시간을 크게 단축할 수 있을 것으로 기대하고 있다.

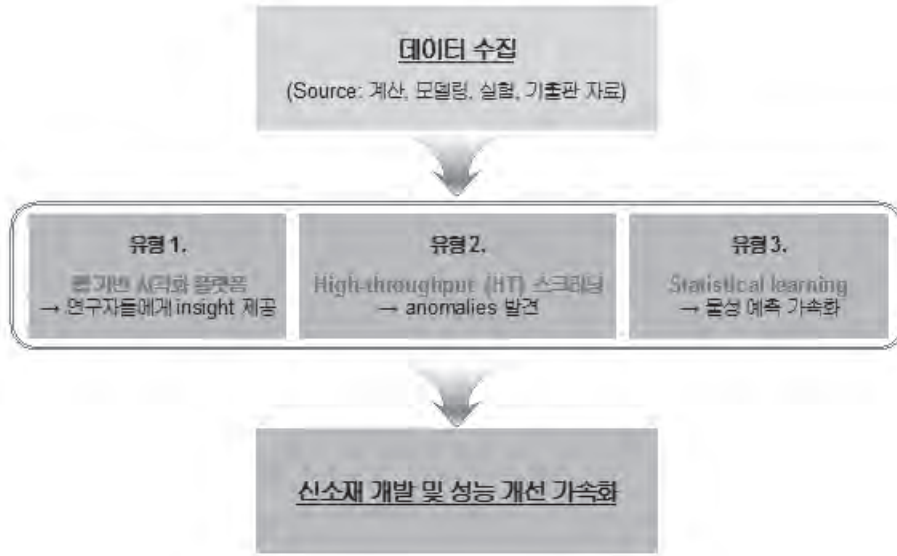
조합실험법은 10년 전에도 활발히 연구되었다가 주춤하고 있는 분야이지만, 최근 제일원리 계산에 의해 도출된 신소재 후보 물질을 조합실험법을 통해 합성하고 그것을 방사광 가속기 등

대형장비를 통해 종합적으로 평가하는 효율적 연구 과정을 구현할 것으로 기대되는 분야이다. 계산 과학에 기반을 둔 소재 정보학의 개념은 2000년경부터 제안되고 있었다. 당시에는 다룰 수 있는 원자 수와 정확도에 한계가 있었고, 대량의 데이터를 통합하고 분석하여 재료 과학의 문제를 해결하는 통계 과학적 기법도 충분치 못하여 그 적용 범위가 제한적이었다. 그러나 최근 몇 년간 컴퓨팅 기술의 발전과 대량의 데이터를 취급 할 수 있는 환경이 구축됨에 따라 이러한 상황이 극적으로 변하고 있다.

특히, 2011년 미국의 Materials Genome Initiative나 일본의 Materials Research by Information Integration Initiative 사업 그리고 우리나라의 미래소재 디스커버리 사업 등 관련된 국가적 사업이 시작됨에 따라, 계산 과학과 데이터 마이닝을 융합시킨 데이터 기반의 재료 설계 혹은 고속 스크리닝 기법이 주목을 받으며 전 세계적으로 활발한 연구가 이루어지고 있다. 이처럼 빅데이터를 활용한 정보과학 기술의 발전은 나노 기술·소재의 연구 개발 방법 자체에도 큰 영향을 주기 시작했다. 매 순간 새롭게 만들어지고 업데이트되는 대량의 실험 데이터가 축적되면, 이 데이터로부터 새로운 재료에 대한 지식의 발견이 가능해지고 원하는 특성을 가진 재료의 효율적인 설계 및 탐색이 가능해질 것이다. 이에 따라 전 세계적으로 데이터 기반 소재 연구라는 새로운 접근법에 대한 노력이 경주되고 있다.

## 1.2 데이터 기반 소재연구의 분석

빅데이터 기반 연구방식은 이미 다양한 분야에서 신소재 개발을 가속화시키고 있다. 본 글에서는 구체적인 성공 연구사례를 통해 이 분야에 대한 이해를 도모하고자 한다. <그림 3-4-1-2>와 같이 빅데이터 기반 소재연구는 크게 세 가지 유형으로 나누어 볼 수 있다. 첫째 유형은 방대한 양의 데이터를 다양한 각도로 시각화시키는 플랫폼을 개발하는 연구이다. 이와 같은 시각화 플랫폼은 언뜻 보기에는 수집된 데이터를 분류/정리하는 정도의 수준으로 보이지만, 방대한 양의 데이터가 모이면 관련 연구자들이 특정 재료들만 연구할 때 간과할 수 있는 지식과 통찰력을 제공한다. 둘째 유형은 고속 대량(high-throughput: HT) 스크리닝 연구로서 연구자들이 원하는 물성을 가지는 신소재를 발굴하는데 매우 유용하다. 방대한 양의 소재들에 대하여 관심갖는 물성을 빠르게 스크리닝하여 극소수의 원하는 재료들을 발굴해내는 접근법이다. 마지막 유형은 수집된 데이터들 간에 유용한 상관관계를 통계적으로 학습시키는 연구이다. 학습이 성공적으로 이루어질 경우, 물성 예측이 매우 가속화될 수 있기 때문에 앞서 소개된 HT 스크리닝의 한계점을 보완할 수 있다.



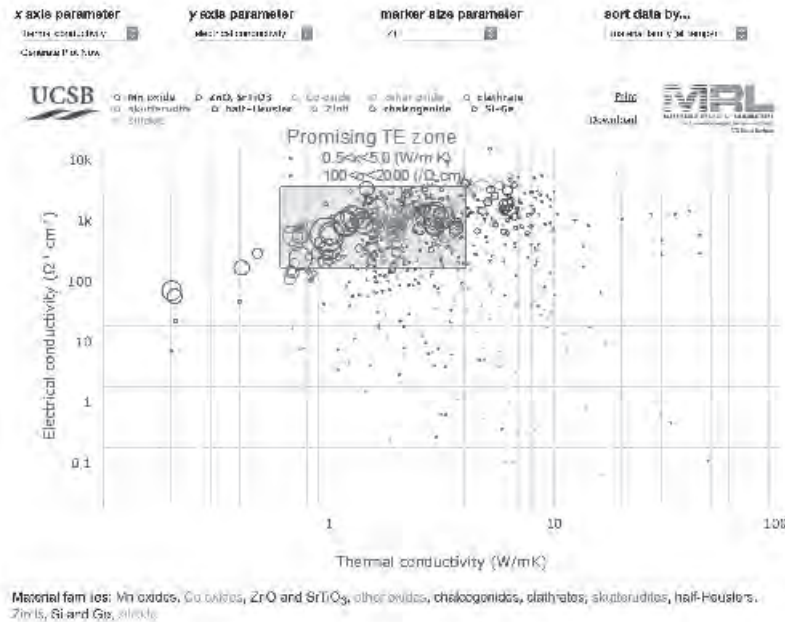
〈그림 3-4-1-2〉 빅데이터 기반 소재연구의 유형 구분 도식도

### 유형 1. 웹 기반 시각화 플랫폼

미국 캘리포니아 주립대학교(UCSB) Ram Seshadri 교수 연구실에서는 100편이 넘는 기출판된 논문들로부터 열전소자 성능변수( $zT^1$ ),  $S$ ,  $\sigma$ ,  $\kappa$ ) 데이터를 총 18,000개 이상 추출하여 데이터베이스화 시켰다.<sup>2)</sup> 더 나아가 이 방대한 양의 데이터들을 기반으로 웹 기반 시각화 플랫폼(web-based visualization platform)으로 설계함으로써, 관련분야 연구자들이 특정 재료에 대해서 단편적인 연구를 진행할 때에는 발견하기 힘든 지식 또는 통찰력을 제공해준다.

1) Dimensionless figure of merit  $zT = S^2 \sigma T / \kappa$  ( $T$  온도,  $S$  지벡상수,  $\sigma$  전기전도도,  $\kappa$  열전도도).

2) Gaultois et al, Chem. Mater, 25, 2911(2013). 해당 연구에서는 실험 데이터 값만 추출하였음.



〈그림 3-4-1-3〉 미국 UCSB Seshadri 교수 연구팀에서 개발한 열전소재 웹 시각화 플랫폼 (<http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp>).  
본 예시에서는 x축, y축, 마커 사이즈를 각각 열전도도값, 전기전도도값, zT값으로 설정하였음

위 그림은 수집된 모든 열전소재 데이터들에 대해 열전도도(x축), 전기전도도(y축), 그리고 zT(마커 사이즈)를 한 눈에 볼 수 있도록 정리해서 보여주는 그래프이다. 이 그래프에서는 특별한 기술적 도구의 도움 없이 육안으로도 마커가 큰 데이터들(즉 열전성능이 좋은 재료들)이 그래프 좌상단에 집중되어 있음을 알아낼 수 있다. 더 구체적으로는 고성능 열전재료들은 대부분 (1) 열전도도 값이 0.5~5(W/mK), (2) 전기전도도 값이 100~2000 ( $\Omega^{-1}\text{cm}^{-1}$ ) 범위에 존재한다. 이와 같이 데이터의 축적 및 시각화로부터 얻은 지식은 연구자들이 향후 연구를 기획하고 진행하는데 큰 통찰력과 방향성을 제시해준다.

#### 유형 2. HT 스크리닝

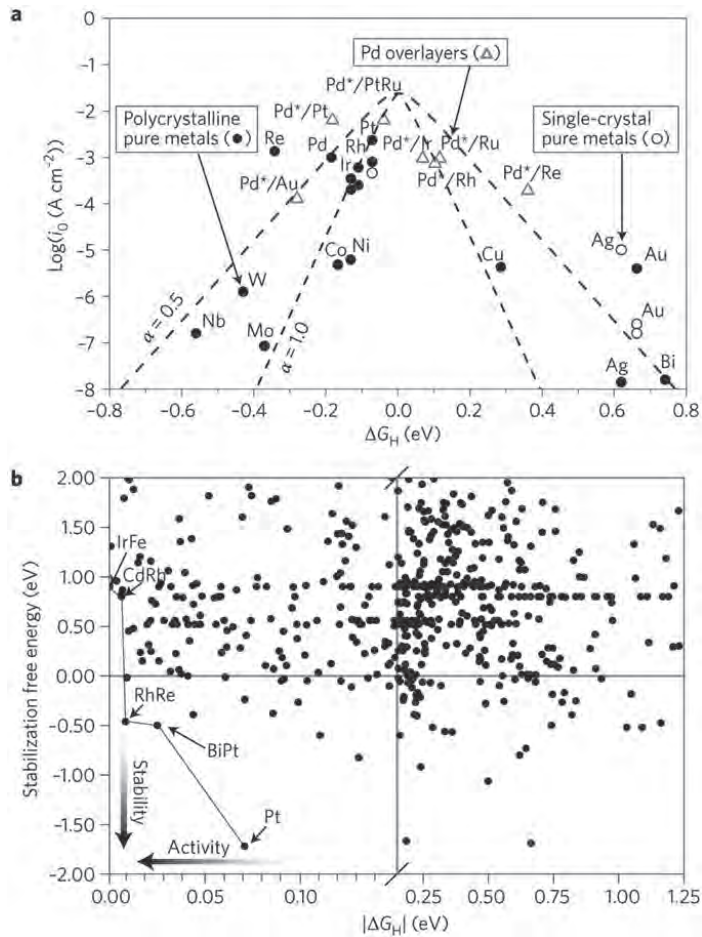
미국 듀크대 Stefano Curtarolo 교수 연구팀은 Half-Heusler<sup>3)</sup> 반도체들에 대하여 격자 열전도도(lattice thermal conductivity,  $\kappa_w$ ) 값을 HT 스크리닝하여, 유망 열전소재들을 제시하였다.<sup>4)</sup> 이 연구팀은 현존하는 최고의 열전소재들(BiTe, PbTe, CuSe 등)이 격자 열전도도가 유난히 낮다는 점에 착안하여, 조합 가능한 79,000개의 Half-Heusler 반도체들에 대하여  $\kappa_w < 5(\text{W/mK})$  인 네 가지 신재료(BiBaK, PtLaSb, RhLaTe, SbNaSr)들을 찾아낼 수 있었다.

79,000개의 재료를 “실험”에 일일이 적용해보는 것은 사실상 무한대의 시간과 노력을 필요로 한다. 해당 연구진은 이러한 한계를 극복하기 위해서 격자 열전도도를 정확하고 빠르게 예측할 수 있는 모델을 설계하는데 성공하였고, 그 결과 방대한 양의 데이터베이스를 확보할 수 있었다. 추후 HT 스크리닝을 통해 걸러진 네 가지 재료에 대해서만 집중적인 실험검증을 할 수 있어, 신소재 개발에 투입되는 시간과 돈이 매우 절약될 수 있다.

3) Half-Heusler 반도체 (XYZ): 자기성 합금의 일종 (fcc 결정구조, 조성 X:Y:Z=1:1:1)으로 열전소재로 많이 사용되는 재료군

4) Carrete et al, Phys. Rev. X 4, 011019 (2014).





〈그림 3-4-1-4〉 흡착에너지 디스크립터 이용한 촉매재료 HT 스크리닝.

- a. 수소원자 흡착에너지( $\Delta G_H$ , x축)에 따른 전기화학 촉매 특성(전류값, y축)
- b. 700여개의 이원계 합금 재료에 대한 합금 생성에너지(y축)와 수소원자 흡착에너지( $\Delta G_H$ , x축)

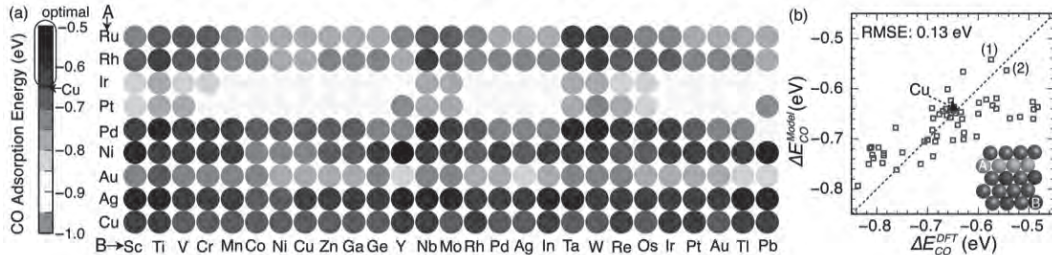
HT 스크리닝 방법론은 열전소자 분야 뿐만 아니라 촉매재료 개발 연구에도 성공적으로 적용된 바 있다. 미국 스탠포드대 Jens Nørskov 교수 연구팀은 Pt을 능가하는 수소 생성용 전기화학 촉매를 찾는 연구를 보고하였다.<sup>5)</sup> 이 연구팀은 700개가 넘는 이원계 합금 재료에 대하여 반응성 디스크립터<sup>6)</sup>로서 수소원자의 표면 흡착에너지를 제일원리 계산을 통해 추적하였다. 위 그림에서 수소 원자 흡착에너지( $\Delta G_H$ )가 0eV에 가까울수록 수소생성 반응성이 높을 것으로 예측되는데, 이런 예측에 기반하여 Pt을 넘어설 수 있는 유망 촉매재료로 BiPt, RhRe 등의 합금 등이 제시되었다. 매우 흥미롭게도 연구팀은 BiPt 촉매의 수소생성 반응성이 실제 실험에서도 순수 Pt 금속보다 더 훌륭하다는 것을 해당 연구에서 검증하였다. 본 연구는 HT 스크리닝 방법으로 신물질을 발굴해내는 훌륭한 예시로 꼽힌다. 700개가 넘는 이원계 합금의 전기화학적 촉매특성을 실험에서 일일이 살펴보는 것은 매우 도전적이고 오래 걸리는 작업이다. 본 예시에서 확인하였듯이 HT 스크리닝 방법론은 이러한 한계를 극복하여 신물질 개발을 매우 빠른 속도로 가속화시킬 수 있다.

#### 유형 3. 통계적 학습법

최근 들어서는 축적된 데이터를 통계적으로 학습시키는 기술(기계학습법)이 신소재 개발 연구에 적용되고 있다. HT 스크리닝 방법론이 신소재 개발을 가속화하고 있음은 틀림없지만, 방대한 양의 데이터를 계산 또는 실험으로 추적해내는 것은 여전히 큰 도전이다. 이를 극복하기 위한 노력의 일환으로 기축적된 데이터를 통계적으로 학습시키는 기계학습법이 소재분야에서 활발해지고 있다. 이 방법론은 기축적된 재료들의 구조 정보와 물성 정보 간에 유용한 상관관계를 찾아내어, 최종적으로는 전혀 학습되지 않은 재료에 대해서도 빠른 속도로 소재 물성을 예측하는데 목표를 두고 있다.

5) Greeley et al, Nat. Mater, 5, 909 (2006).

6) 표면 흡착에너지는 촉매 커뮤니티에서는 널리 알려진 반응성 디스크립터이다. 반응물의 흡착에너지가 적절한 값(이 값은 화학반응이나 반응물의 종류에 따라 달라짐)을 가질 때 촉매특성이 최대화된다고 알려져 있음.



〈그림 3-4-1-5〉 a. 기계학습 모델로 예측된 코어-셸 합금 표면상 CO 흡착에너지.  
b. 기계학습 모델과 DFT 계산방법론의 흡착에너지 예측값 비교

미국 버지니아 공대 Hongliang Xin 교수 연구팀은 촉매  $\text{CO}_2$  환원용 합금촉매 개발 연구에 기계학습법을 도입했다.<sup>7)</sup> 앞서 소개했듯이 촉매 표면상의 흡착에너지는 화학반응성의 디스크립터로 널리 알려져 있다. 그러나 흡착에너지 값은 다양한 반응물 및 흡착사이트 종류를 값비싼 양자계산을 기반으로 계산하여야 얻을 수 때문에 매우 많은 컴퓨팅 자원과 시간을 요구한다. 이 연구팀은 이원계 합금 재료에 대하여, 전기음성도 및 몇 가지의 d-밴드 성질들과 반응물(본 연구에서는 CO) 흡착에너지의 상관관계를 인공신경망 알고리즘으로 학습시켰다. 그 결과 개발된 기계학습 모델을 기반으로 전혀 학습된 적이 없는 촉매재료들의 흡착에너지를 매우 정확하고 빠르게 예측하였다(제곱 평균 에러 0.13eV). 촉매 분야는 촉매재료 후보군, 화학반응 종류, 화학반응 경로 등이 너무나 다양하고 복잡하기 때문에 일일이 실험이나 계산을 진행하여 신재료를 찾아내기가 매우 힘든 분야이다. 통계적 학습법의 도입은 이러한 한계를 극복시켜 신촉매 재료 개발이 더욱 가속화될 것으로 기대된다.

이러한 유형들의 빅데이터 기반 소재개발 연구는 태양전지, 물 광분해, 가스 저장/분리, 압전소자, 열전소자, 촉매, 배터리, 트랜지스터 등의 응용 분야에 이미 다양하게 적용되고 있다. 이 모든 것들을 소개하기에는 지면이 부족하므로, 본 글에서는 “열전소자”와 “촉매” 두 분야에서 이루어진 사례들을 중점적으로 설명하고, 나머지 분야들에 대해서는 〈표 3-4-1-1〉에 정리하였다.

7) Ma et al. J. Phys. Chem. Lett. 6, 3528 (2015).

〈표 3-4-1-1〉 빅데이터 기반 연구들의 추가사례들  
 \* ICSD: 무기결정구조 DB \*\* CSD: 캠브리지 구조 DB

| 응용 분야                  | 데이터 종류(계산, 실험)   | 데이터 크기   | 접근법                      | 유망 신물질   | 출처  |
|------------------------|--|--|--------------------------|--|---|
| 태양전지                   | 밴드갭, 광흡수율<br>(제일원리 계산)   | 260개<br>I-III-VI<br>chalcopyrite 화합물<br>(ICSD*)          | HT<br>스크리닝               | Cu <sub>7</sub> TiS <sub>4</sub> 포함<br>25개   | Phys. Rev. Lett.<br>108, 068701<br>(2012)   |
| Solar thermal fuels    | 이성질화 엔탈피, 에너지 밀도 (제일원리 계산)   | 430개 이성질성 분자<br>(CSD**)                                  | HT<br>스크리닝               | 물질명<br>명시되지<br>않음  | Nano Lett, 14,<br>7046 (2014)               |
| 광 물분해                  | 생성에너지, 밴드갭,<br>밴드예지 (제일원리 계산)                                      | 68개 이원계 질화물,<br>1503개 삼원계 산화물,<br>1377개 사원계산화물<br>(ICSD) | HT<br>스크리닝               | Ti <sub>3</sub> O <sub>5</sub> N <sub>2</sub> ,<br>La <sub>2</sub> TiO <sub>2</sub> N <sub>2</sub> ,<br>Li <sub>3</sub> MoO <sub>4</sub> N | Energy Environ.<br>Sci. 6, 157 (2013)       |
|                        | 생성에너지, 밴드갭,<br>밴드예지, 전하이동도<br>(제일원리 계산)                            | 5400개 페로브스카이트<br>산화물 또는 산질화물<br>(ICSD)                   | HT<br>스크리닝               | MgTaO <sub>3</sub> N<br>포함 15개   | Energy Environ.<br>Sci. 5, 5814<br>(2012)   |
| 가스 저장/분리               | 프로페인/프로필렌 분자의<br>흡착량, 분리선택도<br>(몬테카를로 계산)                          | 65,350개<br>다공성 소재<br>(ICSD, CSD)                         | HT<br>스크리닝               | 해당 없음  | J. Phys. Chem.<br>C 120, 24224<br>(2016)    |
|                        | 메탄 흡착량<br>(계산)   | 137,953개 MOF   | HT<br>스크리닝               | NOTT-107<br>포함 300개<br>이상  | Nat. Chem, 4, 83<br>(2012)                  |
| 압전소재                   | 밴드갭, 압전상수, 유전상수,<br>탄성계수<br>(제일원리 계산)                              | 987개<br>Half-Heusler 재료                                  | HT<br>스크리닝               | KMgP 포함<br>6개  | Phys. Rev. Lett.<br>109, 037602<br>(2012)   |
| 열전소재                   | 열전 성능변수 (zT, S, $\sigma$ , $\kappa$ ),<br>자원량 (기출판된 실험논문<br>100여편) | 18,000개  | 웹기반<br>시각화 기술            | 해당 없음  | Chem. Mater, 25,<br>2911 (2013)             |
|                        | 격자 열전도도 (모델링)  | 79,000개<br>Half-Heusler 반도체 재료<br>(AFLOWLIB.org)         | HT<br>스크리닝<br>+<br>기계학습법 | BiBaK 포함<br>4개   | Phys. Rev. X 4,<br>011019 (2014)            |
| 촉매                     | 수소원자 흡착에너지<br>(제일원리 계산)  | 700개 이원계 합금  | HT<br>스크리닝               | BiPt 포함<br>5개  | Nat. Mater. 5, 909<br>(2006)                |
|                        | 일산화탄소 흡착에너지,<br>합금재료 DOS<br>(제일원리 계산)                              | ~400개 이원계 합금 또는<br>코어-셸 구조 합금                            | 기계<br>학습법                | 해당 없음  | J. Phys. Chem.<br>Lett. 6, 3528<br>(2015)   |
| 배터리                    | 배터리 용량 및 전압<br>(제일원리 계산)   | 60,000개 산화물, 인화물,<br>황화물 등 (ICSD)                        | HT<br>스크리닝               | 물질명<br>명시되지<br>않음  | Mater. Res. Soc.<br>bull. 35, 693<br>(2010) |
| 트랜지스터<br>(게이트용<br>유전체) | 밴드갭, 유전상수<br>(제일원리 계산)   | 1800개 이원계 또는 삼원계<br>산화물 (ICSD)                           | HT<br>스크리닝               | c-BeO 포함<br>20개 이상   | NPG Asia<br>Materials 7, e190<br>(2015)     |

### 1.3 데이터 기반 소재연구를 위한 정책적 기술적 과제

소재 정보학의 성공을 위해서는 무엇보다도 소재 빅데이터의 구축이 선결되어야 할 과제이다. 소재 빅데이터는 공공재의 성격이 강하기 때문에 정부 정책이나 공공연구기관의 역할이 구축과 활용 단계 모두에서 매우 중요하다. 특히, 소재 관련 데이터베이스의 구축 및 새로운 알고리즘에 기반한 시뮬레이션 소프트웨어의 개발 등은 연구 성과의 실용화를 목표로 하는 과제들과는 전혀 다른 성격을 가지고 있다. 따라서, 소재정보학의 발전을 위해서는 국가적 차원에서의 정책 수립과 실행이 반드시 필요하다. 여기에는 하드웨어 인프라의 정비·공유 뿐 아니라 지적 기반(소프트웨어 인프라)의 정비와 공유도 포함된다. 즉, 소재분야의 빅데이터를 활용하기 위한 정보학의 역량도 함께 고도화되어야 한다. 또한, 다양한 소재 데이터를 연구자가 쉽게 이용할 수 있는 환경도 충분치 않기 때문에, 이를 해결할 국가적 인프라의 구축도 시급한 중요 과제라고 할 것이다.

데이터 기반 소재연구는 데이터의 상호 이용을 전제로 하고 있기 때문에 특히 정교하게 설계된 정책 추진 방안이 제시되어야 한다. 소재는 모든 산업의 기반이어서 그 데이터가 널리 공유되어야 하지만 그 수는 턱없이 부족한 상황이다. 더구나 소재 공정 데이터는 거의 전무한 실정인바 기업이 가지고 있는 공정 데이터의 공유도 필요하다. 그러나 소재 기업에서는 수십 년간 시행착오를 통해 구축한 공정 데이터를 핵심 경쟁력으로 인식하고 있어서 이는 쉽게 해결될 수 없는 문제이다. 연구개발은 기본적으로 경쟁적 성격이 강해 타인과의 데이터 공유를 쉽게 수용할 수 없는 분야이다. 따라서 데이터의 비공개와 공개의 기준, 데이터 공개의 인센티브, 연구자와 연구 성과에 대한 평가 등에 대해 모두가 수긍할 수 있는 합리적 시스템을 갖추어야 한다. 이와 함께 소재와 정보학이 융합된 새로운 분야를 이끌어 갈 융합형 소재 인력의 육성 정책 또한 시급한 과제이다.

정보학 기반 소재개발의 기술적 과제들도 산적해 있다. 제일원리계산 및 시뮬레이션 결과를 이용한 데이터 기반의 연구는 연구 대상을 유닛셀 크기의 결정체에 한정된 경우 많은 성과를 거두고 있다. 이 경우에는 컴퓨팅 자원에 따라 데이터의 수를 크게 늘릴 수 있기 때문에 높은 신뢰도 수준에서 가장 우수한 물질을 탐색해 낼 수 있다. 그러나 소재 개발이란 원재료에서 가공재료에 이르기까지 원자, 분자, 나노, 마이크로 그리고 매크로 스케일에서 복잡하게 상호작용하는 구조와 인자들을 이해하고 제어해야 하는 분야이다. 구조 재료의 개발을 좀 더 자세히 살펴보면 나노 스케일에서는 전위를 포함한 결함의 밀도와 분포, 용질

원자의 배열과 분포, 입계 및 계면 구조의 제어가 필요하다. 마이크로 스케일에서는 결정립의 크기와 모양, 응력과 변형의 분포를, 매크로 스케일에서 결함이나 용질 편석 잔류 응력 등 매우 다양한 요소들을 제어해야 한다. 이러한 요소들을 어떻게 데이터베이스화 할 것인가, 그리고 이 데이터들이 합성과 가공 공정에 민감하게 의존한다면 그러한 메타 데이터를 어떻게 부여할 것인가, 각 요소의 공간적 시간적 변화 정보를 어떻게 부여할 것인가 등 많은 문제가 검토되어야 한다. 또한, 바이오 등 타 분야의 데이터베이스와 달리 소재 데이터는 소재만의 특이점이 있다. 소재는 동일 조성에도 공정에 따라 그 특성이 모두 다르게 나타난다. 따라서 소재 데이터베이스에는 소재공정 데이터도 함께 수집되어야 한다. 어느 범위까지의 데이터베이스를 구축하는 것이 바람직할지, 그리고 이상적인 데이터베이스가 구축된다고 해도 데이터로부터 원래의 소재를 재현할 수 있는지 등에 대한 면밀한 검토가 필요하다.

소재 데이터베이스 구축에 있어서 가장 어려운 점은 어떻게 자료를 수집할 것인가 하는 점이다. 그 대책 중 하나는 실험 장치에서 나오는 데이터를 직접 가져오는 방법이다. 이것은 향후 실험 노트가 종이 기반에서 전자화되면 이를 데이터베이스에 연동시킴으로써 해결될 수 있을 것이다. 일본 NIMS의 MatNavi는 기존의 논문에서 수동으로 데이터를 추출하여 데이터베이스화하고 있지만 이미 한계에 도달하고 있다. 이는 발표 논문의 수가 최근 기하급수적으로 증가하고 있기 때문에 수작업으로 데이터를 수집하는 것이 거의 불가능한 상황이기 때문이다. 이를 극복하기 위해서는 텍스트 마이닝 기술의 적극적인 활용을 생각해 볼 수 있다.

이 경우 문제가 되는 것은 데이터의 품질이다. 텍스트 마이닝에 의한 데이터의 품질을 어떻게 보장할 것인가는 텍스트 마이닝 분야의 새로운 연구 주제이기도 하다. 텍스트 마이닝의 저작권과 법적 문제도 함께 정비되어야 한다. 데이터 수집 활동의 인센티브 부여 문제도 숙고할 필요가 있다. 데이터를 수집하고 이를 데이터베이스화하여 제공하는 활동이 현재는 많은 경우 평가 항목에서 제외되어 있다. 또한 데이터의 공유 범위를 어떻게 결정하는가 하는 문제도 정리되어야 한다. 데이터 과학자와 데이터 큐레이터의 육성 방안 역시 향후 큰 문제가 될 것으로 예상된다. 또한, 지속적인 인프라로서 데이터의 수집, 관리, 제공 기능을 유지하기 위해서는 한시적 프로젝트가 종료되어도 이를 관리하고 운영하는 새로운 거점 조직의 설치가 필요하다.



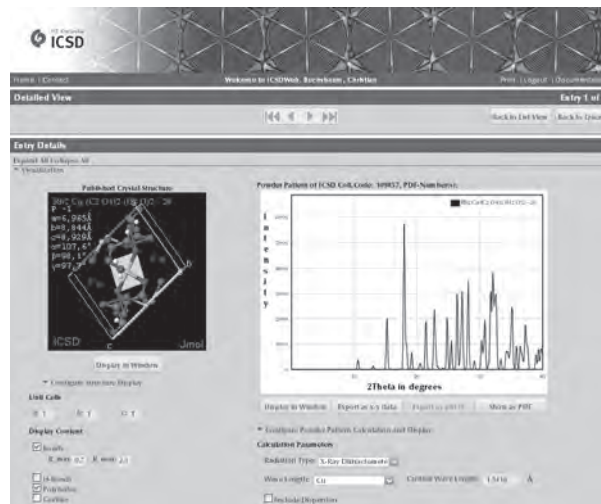
## 2. 연구개발 동향

현재 활용 가능한 전 세계 소재물성 DB 정보는 KIST 계산과학센터의 나노소재설계 플랫폼 포털인 vfab.org에서 제공하고 있으므로 참고하기 바란다.<sup>8)</sup>

### 2.1 유럽

소재정보학 연구에서 가장 우선적으로 구축되어야 하는 데이터베이스는 존재 가능한 모든 소재들의 구조정보이다. 이는 실험적으로 이미 찾았지거나 사용되고 있는 소재들은 물론이고, 가상으로 만들어질 수 있는 소재들을 모두 포함한다. 전 세계적으로 가장 큰 “소재 구조정보” 데이터베이스가 대부분 유럽기관들의 주도 하에 구축되었다. 아래에서 세 가지 사례(ICSD, CSD, Chem Spider)를 중점적으로 소개하고, 추가 사례들 중 일부는 <표 3-4-1-2>에 따로 정리하여 나타내었다.

○ ICSD(Inorganic Crystal Structure Database; 무기결정구조 데이터베이스)



자료: <https://icsd.fiz-karlsruhe.de/search/>

<그림 3-4-1-6> ICSD 공식 홈페이지

8) <http://www.vfab.org/index.php/information/>

FIZ Karlsruhe 기관의 주도 하에 수집된 세계 최대의 무기재료 결정구조 데이터베이스이다. 단원소 물질, 광물, 금속, 금속간 화합물 등 존재 가능한 모든 무기재료를 정의하고, 구조 정보 (좌표, 스페이스 그룹, Pearson 기호, Wyckoff 순서)들을 축적한다. 현재 188,000개의 소재가 등록되어 있으며, 해마다 평균 4,000개 이상의 새로운 소재들과 관련 정보들이 추가되고 있다.

○ CSD(Cambridge Structural Database; 캠브리지 구조 데이터베이스)



자료: <https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/>

<그림 3-4-1-7> CSD 공식 홈페이지

영국 Cambridge Crystallography Data Centre의 주도 하에 수집된 세계 최대의 유기물 구조 데이터베이스이다. 주로 작은 유기물 분자들이나 또는 금속-유기물 화합물에 대하여 구조 정보를 축적해 두고 있다. 현재 900,000개에 달하는 소재들이 등록되어 있으며, 해마다 출판 자료를 기반으로 평균 30,000개 이상의 새로운 소재들이 추가 등록되고 있다.

## ○ ChemSpider

Home About us Web APIs Help Sign in

**ChemSpider**  
Search and share chemistry

Search ChemSpider

Simple Structure Advanced History

### Search ChemSpider

Matches any text strings used to describe a molecule.

Search

Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID

**What is ChemSpider?**

ChemSpider is a free chemical structure database providing fast text and structure search access to over 58 million structures from hundreds of data sources.

**Search by chemical names**

- Systematic names
- Synonyms
- Trade names
- Database identifiers

**Search by chemical structure**

- Create structure-based queries
- Draw structures in the web page
- Use structure files from your computer

**Find important data**

- Literature references
- Physical properties
- Interactive spectra
- Chemical suppliers

자료: <http://www.chemspider.com/>

## 〈그림 3-4-1-8〉 ChemSpider 공식 홈페이지

Royal Society of Chemistry(RSC)를 주도 하에, 화학 구조의 이용가능한 정보원과 이와 관련된 정보를 하나의 검색 저장소에 종합하고 색인을 할 목적으로 개발되었다. 총 480여개의 데이터원으로부터 60,000,000개의 독특한 화학 물질을 저장해두고 있다.

〈표 3-4-1-2〉 유럽 기관들을 주축으로 구축된 DB의 일부

| DB 이름                              | 분류         | 수집 데이터 종류   | URL  | 유료/무료 |
|------------------------------------|------------|-------------|--|-------|
| CALPHAD databases                  | 금속, 계산     | 열역학 상태도     | <a href="http://www.thermocalc.com/products-services/databases/thermodynamic">www.thermocalc.com/products-services/databases/thermodynamic</a> | 유료    |
| Computational Materials Repository | 일반 재료, 문헌값 | 프로젝트 별로 상이함 | <a href="http://cmr.fysik.dtu.dk">cmr.fysik.dtu.dk</a>   | 무료    |
| NoMaD                              | 일반 재료, 계산  | 총에너지        | <a href="http://nomad-repository.eu">nomad-repository.eu</a>   | 무료    |

자료: Materials science with large-scale data and informatics: Unlocking new opportunities, Hill et al. MRS Bulletin 41, 20162.

## 2.2 미국

ICSD, CSD, Chem Spider 등에서 구축된 소재 구조정보 DB를 기반으로 다양한 소재 물성정보 DB(예: 열역학 성질, 전기/자기/기계적 성질 등)가 구축되어 있다. 아래에 대표적인 네 가지 사례를 소개하고, 추가 사례들은 <표 3-4-1-3>에 요약/정리하여 나타내었다.

### ○ Open Quantum Materials Database(OQMD)

**OQMD:**  
The Open Quantum Materials Database

Newsflash: OQMD v1.1 is out! (Download it here.)

**Welcome to the Open Quantum Materials Database**

The OQMD is a database of DFT-calculated thermodynamic and structural properties. This online interface is for convenient, small-scale access; for a more powerful utilization of the data, we recommend downloading the entire database and the API for interfacing with it, from the link below.

**You can...**

- Search for materials by composition,
- Create phase diagrams using the thermochemical data in OQMD,
- Determine ground state compounds at any composition,
- Visualize crystal structures, or
- Download the entire database (and the API) for your own use!

**Tweet @TheOQMD to ask what is stable at a composition, or to get a simple phase diagram!**

Chris Wolverton here is your Pb-Te-Sr-Se phase diagram!

Chris Wolverton LiFeO<sub>2</sub>: stable, dH=-1.98 eV/at E<sub>g</sub>=1.54 V=9.03A<sup>3</sup>/at

자료: <http://oqmd.org/>

<그림 3-4-1-9> OQMD 공식 홈페이지

미국 노스웨스턴 대학(Chris Wolverton 교수팀)의 주도 하에, ICSD 내의 무기재료를 대상으로 DFT 총에너지 계산데이터 470,000여개를 확보해 두었다. 해당 플랫폼 환경에서는 계산된 총에너지를 기반으로 재료의 생성에너지, 상태도와 같은 핵심적인 열역학 성질을 제공한다. 본 데이터베이스는 합금 촉매, 리튬 이온 배터리, 고효율 나노구조 열전소자 등 분야에 사용되는 재료들을 스크리닝하는데 널리 활용되고 있다.

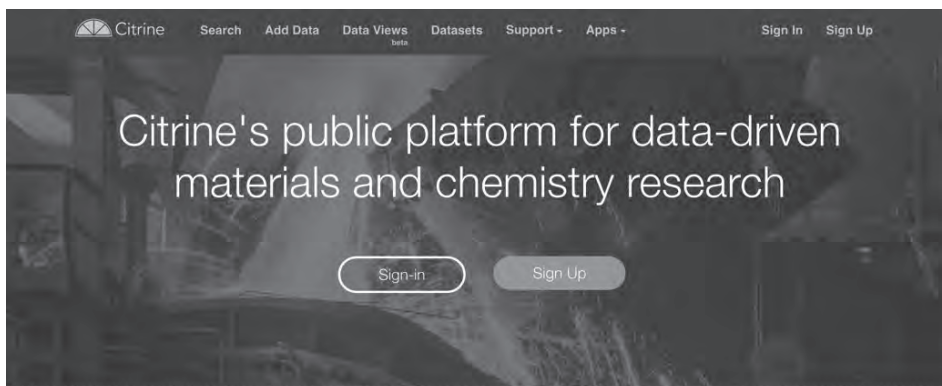
## ○ Materials Projects

자료: <https://materialsproject.org/>




〈그림 3-4-1-10〉 Materials Project 공식 홈페이지

미국 로렌스 버클리 국립연구소(Gerbrand Ceder, Kristin Persson 교수팀)의 주도 하에, ICSD 내의 무기재료(70,000여개)를 대상으로 다양한 소재 기초물성을 확보해 두었다. 앞서 소개된 OQMD와 차별적인 부분은 저장하고 있는 소재물성이 더 다양하다. OQMD의 경우 열역학 성질(예: 생성에너지)을 집중적으로 DB화 시켰지만, Materials Project는 열역학 성질은 물론이고 자기적 성질(예: 자기 모멘트 크기), 전기적 성질(예: 밴드 다이어그램, density of states(DOS)), 기계적 성질(예: 체적 탄성율, 전단 탄성율)까지 DB화하였다.

○ Citrination Materials–Data Analytics Platform



Here's what you can do with Citrination

|  |   |   |
|--|---|---|
|  <p><b>Search</b></p> <p>Explore the world's largest database of materials and chemicals information, including: polymers, alloys, semiconductors, and many more.</p> <p>Search Records    Search Datasets</p> |  <p><b>Add Data</b></p> <p>Store, share, and visualize your raw data and literature reviews.</p> <p>Upload Data</p> |  <p><b>Analyze</b></p> <p>Use your data to construct a machine learning model to fill in missing values in your data and create optimally designed experiment sets.</p> <p>Start a Model</p> |
|--|---|---|

자료: <https://citrination.com/>

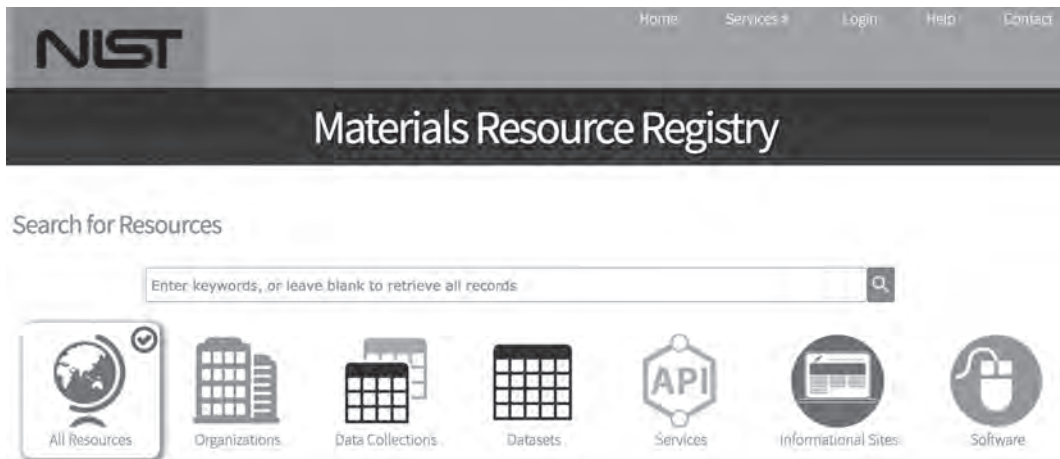
〈그림 3-4-1-11〉 Citrination Platform 공식 홈페이지

Citrine Informatics 회사에서 무료로 공개된 세계 최대의 재료 데이터 플랫폼이다. 앞서 소개된 OQMD, Materials Projects 그리고 AFLOWLIB과 같은 데이터베이스에는 ICSD에서 제공하는 구조정보를 기반으로 제일원리 계산값이 저장되어 있다. 반면, Citrination 플랫폼이 기존 DB들과 차별화되는 점은 크게 다음 두 가지이다. (1) 논문, 특히, 기존 DB 등의 많은 데이터(정형, 비정형)를 망라하고, 계산데이터 뿐만 아니라 사용가능한 실험데이터까지 모두 포함시켰다. (2) 데이터 수집에 그치지 않고, 수집된 데이터를 이용해 기계학습을 수행하는



기능을 내장시킴으로써 DB 내에 존재하지 않는 물질에 대해서도 물성 예측이 가능하게 하는 기능을 제공하고 최종적으로는 연구자들이 원하는 물성을 가지는 적절한 소재를 찾아주는 “역방향 소재설계” 기능을 추가하고 있다.

### ○ Materials Data Repository(NIST)



자료: <https://materials.registry.nist.gov/>

〈그림 3-4-1-12〉 NIST Materials Resource Registry 공식 홈페이지

미국 National Institute of Science and Technology(NIST)는 광범위한 재료 연구자 커뮤니티에서 데이터 공유 및 재사용을 촉진하기 위해 Materials Genome Initiative (MGI)와 협력하여 재료 과학 데이터 저장소(Materials Data Repository)를 만들었다. 재료와 관련된 모든 데이터를 한 곳에 저장시켜두는 역할을 한다. 앞서 소개된 OQMD, Materials Project에서는 특정 재료 물성의 계산데이터 값을 저장하였다면, 본 데이터 저장소에는 실험/계산 구분없이 모든 사용가능한 데이터(주로 문헌자료 기반)를 축적하고 있다. 데이터 제공 기관, 재료/물성/합성 및 공정법 등의 종류별로 데이터 자원 검색이 가능하며, 대부분 공공 서비스로 제공하나 일부 접근이 제한된 데이터들도 존재한다. 또한 Materials Resource Registry에서는 사용자가 직접 데이터를 추가할 수 있게 하여 데이터를 지속적으로 수집하고 있다.

〈표 3-4-1-3〉 미국 기관들을 주축으로 구축된 DB의 일부

| DB 이름  | 분류                  | 수집 데이터 종류                                    | URL                                     | 유료/무료 |
|--|---------------------|--|---|-------|
| AFLOWLIB   | 재료물성, 계산            | 172,000 재료, 기초 열역학, 전기, 기계적 물성               | aflowlib.org                            | 무료    |
| American Mineralogist Crystal Structure Database | 광물, 문헌값             | 구조 정보  | rruff.geo.arizona.edu/AMS/amcsd.php     | 무료    |
| CatApp   | 촉매, 계산              | 다양한 금속 표면 상의 에너제틱스 (반응물, 생성물 에너지, 활성화 에너지 등) | suncat.stanford.edu/catapp              | 무료    |
| CINDAS High Performance Alloys Database          | 항공우주 재료 (합금), 계산/실험 | 열안정성, 기계적 강도                                 | cindasdata.com                          | 유료    |
| DOE Hydrogen Storage Materials Database          | 수소 저장, 계산/실험        | 수소 저장 용량                                     | www.hydrogenmaterialssearch.govtools.us | 무료    |
| Harvard Clean Energy Project                     | 유기물 태양전지, 계산        | 태양전지 관련 물성                                   | cepdb.molecularspace.org                | 무료    |
| Thermoelectrics Design Lab                       | 열전소자, 문헌값           | 열전성능 변수                                      | www.tedesignlab.org                     | 무료    |

자료: Materials science with large-scale data and informatics: Unlocking new opportunities, Hill et al. MRS Bulletin 41, 20162.

## 2.3 일본

일본 정부는 2016년 결의된 제5기 과학기술기본계획에서 슈퍼 스마트 사회 ‘Society 5.0’의 실현에 공헌하는 11개 시스템 중 하나로 소위 ‘통합형 재료 개발 시스템’을 선정하였다. 이 시스템은 계산과학과 데이터과학 기법을 통해 혁신적인 기능성 재료와 구조 재료 등 신소재를 개발하는 것으로서, 소재의 개발 기간을 대폭 단축할 수 있을 것으로 기대하고 있다. 미래에는 개발기간과 비용의 절감 뿐 아니라, 인공 지능을 통해 새로운 소재를 예측하고 합성 공정을 제시하며 새로운 물리법칙을 발견하는 것도 가능할 것으로 기대하고 있다.<sup>9)</sup>

개인 연구자나 소규모 연구그룹이 한시적인 프로젝트를 통해 대규모의 소재 데이터베이스를 구축하고 지속적인 운영과 정보제공을 담당하는 것은 대단히 어렵다. 따라서 소규모 연구를 지원하는 일본의 과학 연구비 조성 사업(일명 과연비 사업)에서는 2017년 9월부터 사용된 가장 최신의 지원기술 분류표에도 소재 데이터베이스 연구는 포함되어 있지 않다. 대신 소재 데이터베이스 구축을 위한 대형 국가 프로젝트로 예산이 책정되어 집행되고 있다. NIMS(National Institute for Materials Science, 일본 국립물질·재료 연구기구)는 과학기술 진흥기구(JST) 사업인 ‘정보 통합 물질·재료 개발 이니셔티브(MI<sup>2</sup>I: Materials research by Information Integration Initiative)’의 주관기관이다. 데이터기반 소재혁신의 거점 구축을 목표로 소재 데이터베이스를 개발·정비하고, 재료 과학부터 정보 과학, 수학 분야에 이르는 산학연 협동 체제를 통한 오픈 이노베이션을 지향하고 있다. 2015년에 창설된 정보 통합 물질·재료 연구 센터(Center for Materials research by Information Integration of NIMS(CMI<sup>2</sup>))는 배터리 재료, 자성 재료, 전열 제어 및 열전 재료는 구체적인 주제의 소재개발연구를 수행하면서 동시에 데이터 중심의 연구 방법론의 개발에도 힘쓰고 있다.<sup>10)</sup> 또한, 산학연 연구자들이 연구개발의 현장에서 활용할 수 있는 정보통합형 물질탐색·재료개발 시스템을 구축하고 있다.

9) Panoramic View of the Nanotechnology/Materials Research Field (국립연구개발법인 과학기술진흥기구 연구개발전략센터, 2017)<https://www.jst.go.jp/crds/report/report02/CRDS-FY2016-FR-05.html>

10) <http://www.nims.go.jp/eng/research/MII-I/>



〈그림 3-4-1-13〉 NIMS MatNavi 데이터베이스 홈페이지

NIMS는 2003년부터 일본 최대의 오픈 데이터베이스인 MatNavi를 제공하고 있다. 이것은 고분자, 무기 재료, 금속 재료, 초전도 재료 등 11개 데이터베이스, 4개의 응용 시스템, 6종의 구조 재료 데이터 시트로 구성되어 있다. 기 출판된 방대한 학술 자료에서 유용한 논문을 추출하여 수치 데이터를 수집, 데이터베이스화하고 있기 때문에 신뢰도가 높은 데이터베이스이다. 그러나 이 데이터베이스는 빅데이터 분석 즉 데이터 기반의 학습을 전제로 구축되어 있지는 않다. 따라서 소재정보학의 관점에서 사용이 용이한 데이터베이스 구축과 함께 텍스트 마이닝을 이용하여 논문의 정보를 자동으로 데이터베이스화하는 기술의 개발도 추진되고 있다. MatNavi는 사용자 등록만 하면 무료로 이용할 수 있으며, 2017년 3월 현재 등록자 수는 12만 명을 넘어섰고 그 중 28%는 해외 사용자이다. 2017년부터 연구를 위한 기능을 담은 데이터베이스 유료 버전의 출시를 준비하고 있다.

또한 효율적 소재 데이터베이스를 구축하고 데이터베이스의 실제적인 이용을 촉진하기 위해 NIMS는 ICT 기술과 소재 기술을 융합하는 핵심 시스템인 ‘통합 재료개발 시스템’ 구축에 나섰다. 또한, 2017년 4월에 소재데이터 플랫폼의 연구개발을 위한 새로운 연구개발 부서로서 통합형재료개발·정보기반부문(Research and Services Division of Materials Data and Integrated System, MaDIS)를 발족시켰다.<sup>11)</sup> 이 부문의 하부 조직으로는 상기한 정보통합 물질·재료 연구센터와 2014년도부터 추진하는 ‘혁신적 구조 재료’ 프로젝트를 수행하는 SIP-MI 연구소(SIP-MI: Strategic Innovation Programs - Materials Integration)가 있다.<sup>12)</sup> 또한 NIMS가 산업계와의 오픈 이노베이션을 지향하는 소재 오픈 플랫폼(Materials Open Platform: MOP)에서 소재정보학을 연구하는 MOP-MI 연구소와 새로 설립된 소재 데이터 플랫폼 센터를 포괄하고 있다. 소재 데이터 플랫폼 센터(Materials data platform center, DPFC)는 통합 재료 개발 시스템을 지원하는 핵심 활동으로 세계 최대 규모의 고기능 재료 데이터 플랫폼을 구축하고 있다.<sup>13)</sup>

## 2.4 중국

미국의 MGI가 발표되고 나서 중국에서도 중국판 MGI(Chinese Materials Genome Initiative)가 시작되었다. 2014년에 상하이 시와 상하이 대학이 공동으로 상하이 Materials Genome 연구소를 설립한 바 있고, 최근 데이터 과학을 활용한 소재 개발 연구를 활발히 진행하고 있다. 2015년 5월에는 베이징 과학기술 대학 내에 MGI 연구소가 신설되었다. 또한 2016년에는 상하이 교통 대학에서도 Materials Genome 융합연구센터를 설립하는 등 거국적으로 Materials Genome 연구에 힘을 쏟기 시작하고, 선진 연구자와의 제휴를 강화하고 있다. 계산과학 및 기계학습, 그리고 데이터 저장 시스템을 구축하는 등 단기간에 눈부신 발전을 보이고 있다. 중앙 정부와 상하이 지방 정부로부터 총 1,000억 원 정도의 예산을 받아 미국에서 귀국한 중국인 연구자를 리더로 응용 연구를 가속화하고 있다. 2016년 3월에 발표된 과학 기술 혁신 제 13차 5개년 계획에서도 중국 산업의 국제 경쟁력 향상을 위한 중점 기술 중 하나인 ‘신소재 기술’ 분야 중에 새로운 재료의 개발 기간·비용을 1/2로 감소하기 위한

11) <http://www.nims.go.jp/eng/research/MaDIS/>

12) <http://www.nims.go.jp/research/SIP-MI/>

13) <http://www.nims.go.jp/research/materials-data-pf/>

‘Materials Genome 공학’을 선정하고 있다. 2016년 6월에는 중국 국가 중점 연구 프로젝트로 MGI 관련 14 과제를 채택했다.(연구비 총액: 3억 위안, 한화 약 480억 원)<sup>14)</sup>

2016년 1월 베이징 Materials Genome 공학 혁신 연맹이 구성되어 중국 과학원 물리 연구소에서 설립 기념식 및 제1회 전체 회원 대회가 개최되었다. 본 연맹은 중국 과학원 물리 연구소와 베이징 과학기술 대학에 의해 공동 설립되었다. 설립 취지는 관련 분야에서 우위를 갖는 대학, 과학 연구원(소) 및 기업 등 총 36 기관으로 구성되어있다. 주요 참가 기관은 중국 과학원 물리 연구소, 베이징 과학기술 대학, 베이징 신소재 개발센터, 닝더 시대 신에너지 과학기술 회사, 칭화 대학, 베이징 대학, 베이징 항공항천 대학, 중국 철강연구 과학기술 그룹 유한 공사, 중국 과학원 계산기 네트워크 정보센터, 베이징 공업 대학 등이다. Materials Genome에 관한 연구를 실시하는 것으로, 베이징 및 중국 전체의 신소재의 연구 개발 과정의 가속화, 개발주기 및 비용의 절반, 신소재 산업 체계의 구축·구비 제조업의 부흥, 혁신 주도 발전 전략의 수행을 지원하는 것을 목적으로 하고 있다. 이처럼 중국의 소재 빅데이터 연구활동의 특징은 국가 수준보다 상하이와 베이징 등 지방 정부 차원에서 프로젝트가 만들어지고 있는 점이다. 민간 사례의 경우 2015년 9월에는 닝보에 국제 MGI 연구소도 설치된 바 있다.

## 2.5 시사점

머신러닝과 최적 제어 이론을 결합하여, 특정한 문제의 해답을 찾아가는 인공지능에게 빅데이터는 판단의 논리를 제공하는 모수로 기능(a work of ‘calibration’)한다. 따라서, 데이터 기반 소재연구에서 가장 선행되어야 할 부분은 빅데이터 인프라의 구축이며, 소재관련 데이터를 체계적으로 수집하고 저장, 관리하는 역량은 미래 소재 경쟁력의 핵심이라고 할 수 있다.

위에 소개된 바와 같이 전 세계적으로 여러 연구기관들이 빅데이터 기반 소재개발을 가속화시키기 위해서 빅데이터 인프라 구축을 위한 노력을 다양하게 진행하고 있다. 이처럼 많은 DB가 이미 존재하고 있지만, 현재 시점에서는 다음 몇 가지 한계점들을 인지하고 극복하려는 노력이 필요하다.

14) Panoramic View of the Nanotechnology/Materials Research Field (국립연구개발법인 과학기술진흥기구 연구개발전략센터, 2017)

<https://www.jst.go.jp/crds/report/report02/CRDS-FY2016-FR-05.html>



첫째, 현재 구축되어 있는 데이터 베이스의 물성정보가 대부분 계산데이터이다. 데이터 산출방법 측면에서 분류해 보면, 실험데이터와 계산데이터로 양분이 가능하다. 계산데이터의 경우(현재는 주로 제일원리 계산) 컴퓨터 자동화 생성이 가능하고 변수 조절이 용이하기 때문에 대량으로 축적하는 것이 가능하다. 때문에 대부분의 대규모 DB는 계산데이터에 기반을 두고 있다. 하지만 사용하는 계산방법론이 산출해주는 물성만을 확보할 수 있고, 일부 물성의 경우 실험값과 동떨어진 결과들을 축적할 때가 많아 데이터의 질(quality)를 훼손시킬 수도 있다. 현재는 실험데이터 DB의 양이 상대적으로 매우 부족하다. 전 세계의 대규모 연구기관(학교, 연구소, 가속기 등)들에서 재료들의 물성 분석 결과를 체계적으로 DB화 하려는 노력이 꼭 필요한 시점이다.

두 번째는 데이터의 표준화 작업이 필요하다. 위에 나타나있듯이 이미 많은 DB가 존재 하지만 DB마다 저장 체계가 모두 제각각이다. 예를 들면 OQMD와 Materials Project는 무기 재료의 열역학 성질들을 공통적으로 저장하고 있지만, 그 저장 방식이 서로 다르다. DB 관리 기관들에서도 데이터의 표준화 필요성을 인식하고 있지만, 당사자들 간의 이해관계 때문에 합의 도출이 어려운 상황이다. 재료정보학의 발전을 위해서는 단일통합 시스템을 만드는 노력이 꼭 필요하다. Citrine Informatics 회사는 소재 데이터 표준화 작업에 앞장서서, Materials Information File(MIF, Open JSON-based file format, created by Citrine Informatics) 등을 데이터 표준 포맷으로 공식화하기 위해 노력 중이다. 재료분야 단일통합 데이터 시스템이 구축된다면 향후 기계학습과 분석을 더욱 강력하게 하여 신소재 개발을 매우 빠르게 가속화시킬 것이다.

세 번째는 인포메틱스 플랫폼의 구축이 필요하다. 인포메틱스 플랫폼을 통한 데이터 기반의 지식 창출은 데이터의 잠재 가치로부터 실질 가치를 실현하는 의미를 지닌다. 따라서 인포메틱스 플랫폼은 방대한 데이터 정보로부터 숨어있는 지식에 접근하고, 연구개발 과정에서 빅데이터를 활용할 수 있게 하는 연구 기반으로서 사용자 선호도를 반영하여 모델링, 계산, DB 구축, 기계학습, 예측/설계 등의 기능이 구현되어야 한다. 또한, 통합형 플랫폼 설계를 위해 단일화된 프로그래밍 언어, 인터넷 연결 등의 요소도 고려되어야 할 것이다. 데이터와 알고리즘을 이용해서 과학적 지식을 창출하는 R&D 정보학은, 해당 분야 전문가와 정보처리 툴을 위한 프로그래밍과 데이터 분석/처리에 익숙한 전산전공자 또는 데이터 사이언티스트들이 협력하는 융합연구를 통해 직관적이고 사용이 편리한 SW 및 알고리즘 개발이 이루어질 수 있을 것이다.

## 2.6 국내 현황

국내에서 소재 빅데이터와 데이터 기반의 소재개발은 2013년 창의소재 디스커버리사업의 기획단계에서 새로운 연구방법론의 하나로서 국가연구개발사업의 기획에 적극적으로 반영되기 시작했다.<sup>15)</sup> 그러나 소재개발에 활용될 수 있는 데이터가 부족하여 본격적인 소재정보학으로의 발전된 모습은 아직 보이지 못하고 있는 실정이다. 현재 국내에 공개되어 있는 소재 데이터는 소재종합솔루션센터<sup>16)</sup>에서 구축한 소재정보 데이터베이스가 있는데, <표 3-4-1-4>에서와 같이 분야별 소재정보은행에서 특화된 소재물성 데이터를 제공하고 있다. 소재정보은행에서 구축한 소재 물성 DB는 논문, 특허, 기업 상품설명서 등에서 수집 가공되거나, 측정에 의해 자체 생성한 물성 정보로서 각 분야의 일부 소재 및 응용분야에 집중되어 있다. 따라서 이들 DB는 한정된 범위에서 물성 검색 정도로 활용되고 있을 뿐 이를 이용하여 소재를 설계하는 수준에는 미치지 못하고 있다.

15) 창의소재 디스커버리 사업 기획보고서(2013.6. 미래창조과학부)

16) <http://www.matcenter.org>

〈표 3-4-1-4〉 국내 소재 데이터베이스 구축 현황

| 소재정보은행   | DB 건수                               | *DB 물질 종류  |
|--|-------------------------------------|--|
| 금속소재<br>(2006~2016.5. 기준)<br>한국기계연구원 부설 재료연구소<br><a href="http://www.metalsbank.com/mb/main/main.mb">http://www.metalsbank.com/mb/main/main.mb</a>                                   | 115,532건<br>(수집/가공)<br>81,592건 (생성) | 철강소재, 알루미늄합금, 마그네슘합금,<br>분말소재(철계, 초경), 구리소재 등  |
| 화학소재<br>(2007~2016.5. 기준)<br>한국화학연구원<br><a href="http://www.cmib.org/index.html">http://www.cmib.org/index.html</a>  | 529,979건<br>(수집/가공)<br>41,329건 (생성) | 플라스틱(열가소성, 열경화성),<br>고무 엘라스토머의 고분자 소재, 첨가제<br>(열안정제, UV안정제, 가소제) 등,<br>필름, 복합재료, 정밀화학소재<br>(점/접착제, 도료 등) 등 |
| 세라믹소재<br>(2007~2016.5. 기준)<br>한국세라믹연구원<br><a href="http://www.ceramicsbank.com/ceramic/body.jsp">http://www.ceramicsbank.com/ceramic/body.jsp</a>                                    | 150,144건<br>(수집/가공)<br>30,694건 (생성) | 유전/압전소재, 반도체/도전체, 에너지소재,<br>유리, 형광체, 벌크제,<br>내열소재 등  |
| 섬유소재<br>(2012~2016.5. 기준)<br>다이텍연구원<br><a href="http://210.123.142.138/miguest/index.aspx?profileKey=textile_DB">http://210.123.142.138/miguest/index.aspx?profileKey=textile_DB</a> | 36,833건<br>(수집/가공)<br>14,047건 (생성)  | 소재물성별(폴리에스테르계, 폴리아미드계,<br>폴리우레탄계),<br>용도별(수송용, 토목용, 산업용),<br>공정별(시제조, 시가공, 제직, 가공) 등                       |

한편, 국가참조표준센터에서는 신뢰성 있는 물성데이터를 수집 생산하고 있으며, 현재 금속, 물리화학, 재료 등 8개 과학기술 분야, 76개 DB, 2만 7천여 데이터를 제공하고 있다.<sup>17)</sup> 또한 산업통상자원부의 산업핵심기술개발사업을 통해 계산과 정보학의 융합플랫폼인 iBAT이 개발 공개된 바 있다.<sup>18)</sup> 최근 국내 대기업인 LG 화학, 삼성 SDI, 삼성중공업 등 산업체에서는 리튬 이차전지 전극 소재 및 전해질 첨가제 등에 소규모의 ‘Computational Materials Screening’ 방법으로 자체 데이터베이스를 만들어 활용하고 있는 것으로 파악된다. 한편, 국내 중소기업에서 소재개발에 소재 빅데이터를 활용하는 것은 거의 전무하다고 할 수 있다.

17) <http://www.srd.re.kr>

18) 에너지용 나노소재의 효율적 설계를 위한 웹기반 밀티스케일 시뮬레이션 플랫폼 개발(연구책임자: 한국과학기술연구원 이광렬)

### 3부. 기술동향분석

소재 데이터의 중요성에 대한 인식 부족과 취약한 인프라를 극복하기 위해 2016년 미래창조과학부에서는 미래소재산업의 준비를 위한 경쟁력 확보전략으로서 소재 빅데이터 플랫폼 사업을 제안하였다.<sup>19)</sup> 이에 따라 2017년 2, 3차 나노·소재기술개발 사업(연구재단)을 통해 빅데이터 플랫폼 연구가 본격적으로 추진되기 시작했다.<sup>20)</sup> 이 사업은 실험 데이터와 계산물성 데이터의 체계적인 수집·가공·분류에 의해 소재 빅데이터를 구축하고, 소재개발의 효율을 높이기 위한 빅데이터 분석 기술을 개발·활용하는 것으로 목표로 하고 있다. 또한, 축적된 빅데이터와 데이터 기반 소재설계기술을 널리 활용하기 위해 기술적, 산업적 수요에 따라 선정된 전략 소재 별로 테마형 빅데이터 플랫폼을 구축하여 공개한다. <표 3-4-1-5>는 2016년 미래소재산업 준비계획 수립을 위한 기획연구보고서<sup>21)</sup>에서 제시한 테마형 빅데이터 플랫폼 사업의 추진안이다.

<표 3-4-1-5> 테마형 소재 빅데이터 플랫폼 사업의 추진 예시

| 추진방안                        | 세부과제(예시)  |
|-----------------------------|---|
| 소재 빅데이터 구축 및 활용을 위한 요소기술 개발 | <ul style="list-style-type: none"> <li>• 소재물성 메타 데이터 설계를 위한 한-미-일-유럽-아시아의 국제공동</li> <li>• 소재 빅데이터 구축 및 관리 시스템 개발</li> <li>• 소재정보학 요소기술(데이터 저장·관리, 데이터 해석 및 가시화) 개발</li> <li>• 데이터마이닝, 머신러닝 활용기술 개발</li> </ul>   |
| 테마형 소재 빅데이터 플랫폼 구축          | <ul style="list-style-type: none"> <li>• 나노촉매 빅데이터 플랫폼</li> <li>• 에너지 하베스트 소재 빅데이터 플랫폼</li> <li>• 수소생산·이송·저장용 소재 빅데이터 플랫폼</li> <li>• 에너지 플랜트용 구조용 합금 빅데이터 플랫폼</li> <li>• 구조용 고엔트로피 합금 빅데이터 플랫폼</li> <li>• 수송기기용 탄소복합체 소재 빅데이터 플랫폼</li> <li>• 복합기능 차폐소재 빅데이터 플랫폼</li> <li>• 고내구성 표면 나노구조체 빅데이터 플랫폼</li> <li>• 환경정화용 나노소재 빅데이터 플랫폼</li> </ul> |

19) 미래소재산업 준비계획 수립을 위한 기획연구 보고서(2016.8, 미래창조과학부)

20) 소재 빅데이터 기반 고효율 열전 신소재 예측과 개발(연구책임자: 한국과학연구원 장현주)

21) 미래소재산업 준비계획 수립을 위한 기획연구 보고서(2016.8, 미래창조과학부)

소재 빅데이터의 구축과 활용을 위해서는 많은 도전과제들이 존재한다. 첫째로는 각 연구실과 실험실에 활용될 수 없는 형태로 산재되어 있는 다양한 형태의 데이터를 어떻게 수집 활용할 것인가 하는 점이다. 기존의 데이터는 수집을 포기한다고 하더라도, 향후 각 연구진이 생성하는 데이터를 축적 관리할 수 있는 국가적(혹은 기관별) 인프라의 구축은 소재기술의 미래 경쟁력 관점에서 매우 중요한 이슈이다. 또한, 이미 출판된 논문과 특허 및 기술 자료에 포함된 데이터들을 활용할 수 있는 체계 구축도 필요하다. 이들 데이터의 가공과 분석만으로도 다양한 지식의 창출이 가능하기 때문이다.<sup>22)</sup> 소재정보를 빅데이터화하기 위해서는 이러한 정보의 수집과 가공이 자동으로 일어날 수 있는 플랫폼과 거대규모 데이터의 관리 및 운용 기술 개발이 필수적이다.

둘째로는 빅데이터 기반의 소재설계 기술이 개발되어야 한다. 데이터의 상관관계를 분석하여 소재개발 전략의 수립을 지원하고, 머신러닝과 딥러닝을 통해 신소재의 물성과 구조를 최적화할 수 있는 기술의 개발은 데이터를 통한 소재개발 경쟁력의 핵심이다. 이를 위해서는 우선 소재 전문가와 데이터 전문가와의 협업을 통해 소재정보학의 핵심 역량이 확보되어야 할 것이다. 장기적으로는 소재연구자들이 데이터 과학의 기법들을 연구에 활용할 수 있도록 데이터과학 분야의 교육과 관련 기법의 개발이 필요하다. 또한, 데이터를 활용한 소재개발의 학습효과를 위해 시범적 성공 사례의 도출 또한 소재정보학의 성공을 위해 필요하다고 할 것이다.

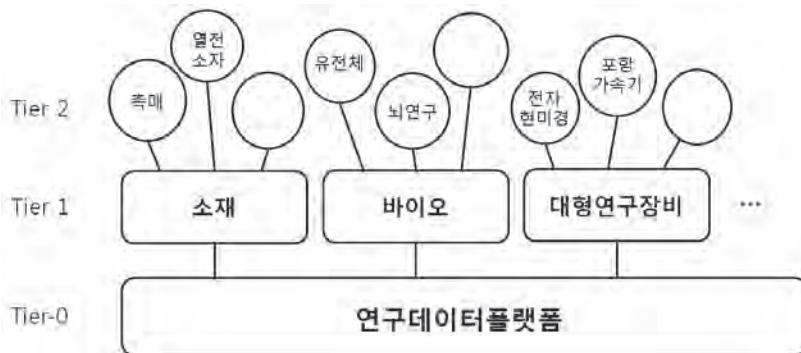
세 번째로는 데이터의 공유를 유도하기 위해 데이터 소유자의 권리를 보장하고 데이터 공유에 따르는 혜택부여 등 소재연구진이 동의할 수 있는 제도적 정비가 필요하다. 과학기술 분야에서는 데이터의 생성자체가 연구개발 행위의 목적이다. 이러한 점에서 일상생활로부터 산출되는 사회현상 데이터나 경제활동으로부터 산출되는 경제현상 데이터와는 성격을 달리한다. 따라서, 연구개발 데이터의 공유에 관한 과학기술계 내의 합의가 반드시 필요하다. 이를 위해서는 정부가 합리적인 제도의 개발에 주도적 역할을 해야 할 것이다.

22) R. Seshadri, T. D. Sparks, *APL Mater.* 4, 053206 (2016); Y. Katsura, M. Kumagai, S. Gunji, Y. Imai K. Kimura, *J. Jp. Soc. Powder Metall.* 64 (8), 467 (2017)

### 1) 과학기술정보통신부 현황

2017년 8월부터 과학기술정보통신부에서는 연구개발 과정에서 발생하는 데이터를 국가적 자산으로 관리 활용할 수 있는 기반과 정책 수립을 위한 TFT를 가동하고 있다. 연구데이터 공유·활용 체계 구축 로드맵에서는 데이터 수집 및 관리 제도, 연구데이터 수집 및 관리 지원을 위한 인프라, 공유 활용 플랫폼과 생태계 구축이 모두 고려된 전략수립을 추진하고 있다. 성공적인 생태계 구축을 위해 1) 연구커뮤니티 수요를 기반으로 추진하며, 2) 우선순위에 따라 점진적으로 확장하고, 3) 연구데이터 플랫폼을 구축 공개한다는 접근 전략의 정책을 수립하고 있다. 연구 분야 별 특성을 고려하여 공유 활용 방안을 마련할 계획이며, 2021년까지 3개 시범분야로서 소재, 바이오·의약, 그리고 대형 연구시설의 데이터 축적과 활용이 고려되고 있다.

아래 그림은 계층(Tier) 구조 방식으로 국가 데이터 공유·활용 생태계를 구축하는 연구 데이터 공유·활용 체계의 개념도이다. Tier-1, Tier-2 센터는 데이터 관리에 대한 권리·책임을 보유하는 연구 커뮤니티가 담당하며, 수집된 데이터를 관리·검증 및 공유하기 위한 큐레이팅, 데이터 활용 연구와 1차 아카이빙(Archiving)을 수행한다. Tier-0은 국가 연구데이터 플랫폼으로서 데이터의 공유·활용을 위한 하드웨어, DB, 컴퓨팅, 네트워크 및 액세스 환경과 함께 데이터 관리 및 분석 SW를 제공하는 역할을 수행한다. 따라서, Tier-1과 Tier-2 센터는 하드웨어가 없어도 Tier-0에서 제공하는 하드웨어를 통해 데이터 플랫폼 구축이 가능한 국가적 인프라를 제공하게 된다. Tier-1은 여러 Tier-2를 연계하여 소재, 바이오 등 대분류에서 1~2개씩 선정한다. Tier-1은 Tier-2 보유기관 중 한 곳이 담당하며 Tier-2의 조성계획, 운영, 데이터의 관리를 책임진다.



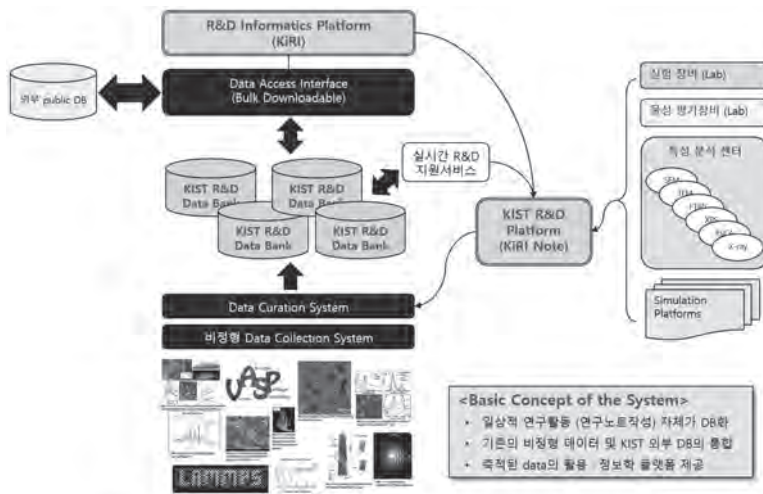
〈그림 3-4-1-14〉 연구데이터 공유·활용 체계 개념도



또한, 데이터 기반 연구 커뮤니티 형성을 촉진하기 위해 데이터 기반의 시뮬레이션 및 머신러닝을 통한 난제해결을 목표로 하는 집단연구 지원 프로그램 신설과 데이터 뱅크를 통한 정보제공 플랫폼 구축 그리고 연구커뮤니티 육성 활동을 지원하는 예산 배정을 골자로 한 기반구축 안이 수립되고 있다. 제도적인 면에서는 연구데이터 공유·활용의 법적 근거 마련, 연구데이터 소유권 명시안, 그리고 연구데이터의 관리 및 공유·활용 기본 원칙 수립을 위해 다양한 의견을 수렴하고 있다. 이 과정을 통해 성과 평가법 개정, 연구개발 사업 공동 관리 규정 개정, 연구 분야별 데이터 관리 원칙 및 계획안 등이 마련될 것으로 기대된다.

## 2) 한국과학기술연구원 (KIST) 현황

2017년부터 KIST에서는 데이터 기반의 연구개발 환경을 제공하기 위해 R&D 빅데이터 및 정보학 기술을 이용할 수 있는 연구개발 인프라 구축 사업을 진행하고 있다(그림 3-4-1-15 참조). 본 사업은 2016년부터 시작된 KIST 빅데이터 포럼을 통한 의견수렴을 바탕으로 기획되었으며 KIST 내 모든 연구 분야에 적용될 수 있는 범용의 인프라 구축을 목표로 하되, 인프라의 실효성 입증을 위한 시범 적용 분야로서 나노 촉매 분야를 선정하여 2018년 12월까지 구축 완료하는 것을 목표로 하고 있다.



〈그림 3-4-1-15〉 KIST 데이터기반 R&D 플랫폼의 구성도.  
 KIST R&D 플랫폼은 ① 일상적 연구활동의 기록 자체가 DB화 될 수 있는 시스템,  
 ② 기존의 비정형 데이터 수집 가공을 통한 다양한 R&D 지원 서비스 개발,  
 ③ 원내외 DB를 연동한 빅데이터 기반의 인포메틱스 플랫폼의 구축 제공을 지향함

이 사업의 목표는 다음 세 영역의 인프라를 구축하여 공개 활용토록 하는 것이다.

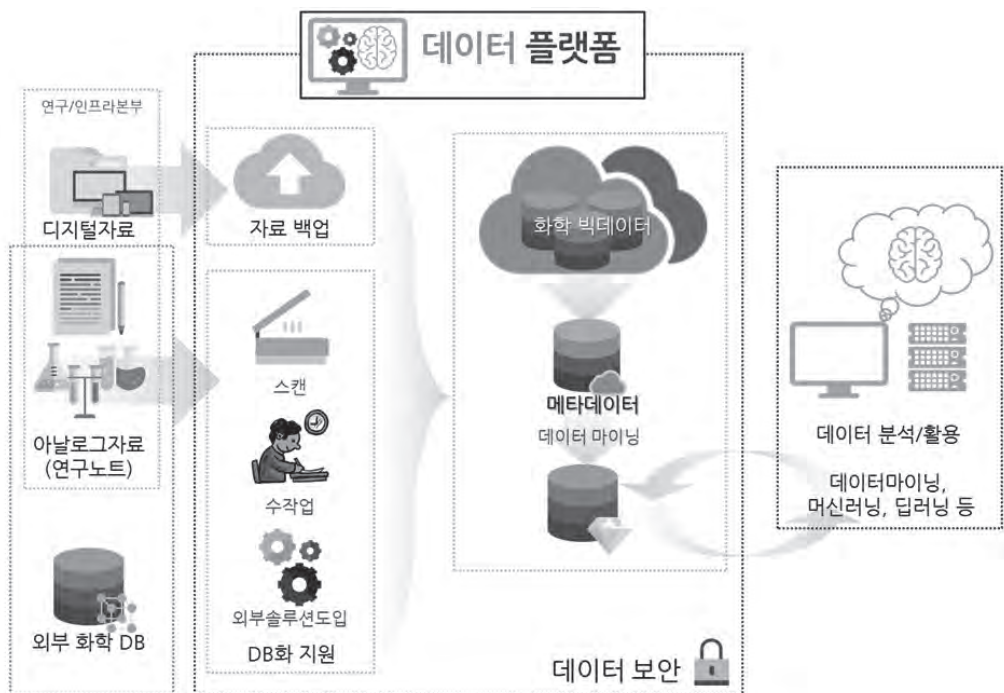
1. 특성분석센터 및 나노입자 설계 플랫폼을 중심으로 한 R&D Platform “KiRI Note v1.0” 구축
  - 연구자 개인의 통상적 연구 활동이 데이터화 될 수 있는 플랫폼으로서 KiRI Note를 구축하여 제공
    - 실험연구의 경우 실험 과정을 기록하고 생성된 시료의 분석/특성평가를 통해 data 생성
    - 특성분석센터의 장비를 통한 분석은 특성분석센터 통합관리시스템과의 연계를 통해 진행하며 data 생성
    - 계산연구의 경우 기능성 나노입자 소재설계 플랫폼과 연동
  - 모든 생성된 data는 meta data와 함께 KIST Data Bank에 비공개 조건으로 축적되며, 시스템은 이 데이터에 기초하여 work 별로 연구노트를 자동 생성한 후 연구노트 인증시스템으로 이관
  - KIST 내부 데이터뿐만 아니라 대량의 외부 글로벌 과학기술 데이터 접근을 통한 AI 기반 연구활동 지원 서비스 체계 마련
2. “KiRI Note”의 data와 외부 비정형 데이터를 수집 축적하고 활용하는 국제표준에 부합되는 KIST Data Bank 구축 및 서비스
  - KIST Data Bank 구축을 위한 국제 표준과 연계된 핵심 인프라 구축
    - 국제사실표준(DCAT) 기반 과학기술 카탈로그 확장 개발 및 상호연동
  - 글로벌 과학기술 데이터 플랫폼 기술 주도를 위한 자체 기술력 및 데이터 네트워크 인프라 확보
  - KIST 내외부 정형·비정형 데이터의 통합, 저장, 관리 및 활용
  - 대량의 외부 글로벌 과학기술 데이터 접근을 통한 실시간 연구활동 지원 서비스 기술의 개발
3. 나노촉매의 물성예측을 위한 머신러닝 알고리즘 개발 및 KIST R&D 인포메틱스 플랫폼 “KiRI v1.0” 구축
  - KIST 내/외부의 데이터, 글로벌 스케일의 데이터를 손쉽게 통합 검색 및 활용할 수 있는 데이터 기반 R&D 서비스 포털

- 연구원별 개인화 페이지, 연구원 개인 데이터 업로드, 글로벌 과학기술 데이터와의 조인, 과학기술 데이터 인지적/시각적 분석, 다양한 퍼블리싱 및 공유 기능 제공
- 촉매 소재의 여러 특성들 사이의 상관관계를 비교 분석할 수 있는 사용자 환경 구축
- 수리과학 기반 데이터 모델링과 빅데이터 분석 및 물성 예측모델 개발
  - 촉매 소재 설계 모델 구축 및 공개

데이터 기반 R&D 인프라의 구축은 기관의 미래연구 경쟁력을 크게 증가시킬 수 있을 것으로 기대하고 있다. KIST 생산 데이터의 체계적 관리를 통해 데이터를 자산화 하며, 연구노트를 통한 지적재산 보호 기능 강화, 데이터 기반 연구 활동 지원, 연구동향/연관기술 등 분석을 통한 연구정책 수립 기능 강화, 머신러닝 기반 촉매 물성 예측 알고리즘/모델 개발 등 다양한 관점에서 KIST의 주요 R&D 기반으로 위치하게 될 것이다.

### 3) 한국화학연구원 (KRICT) 현황

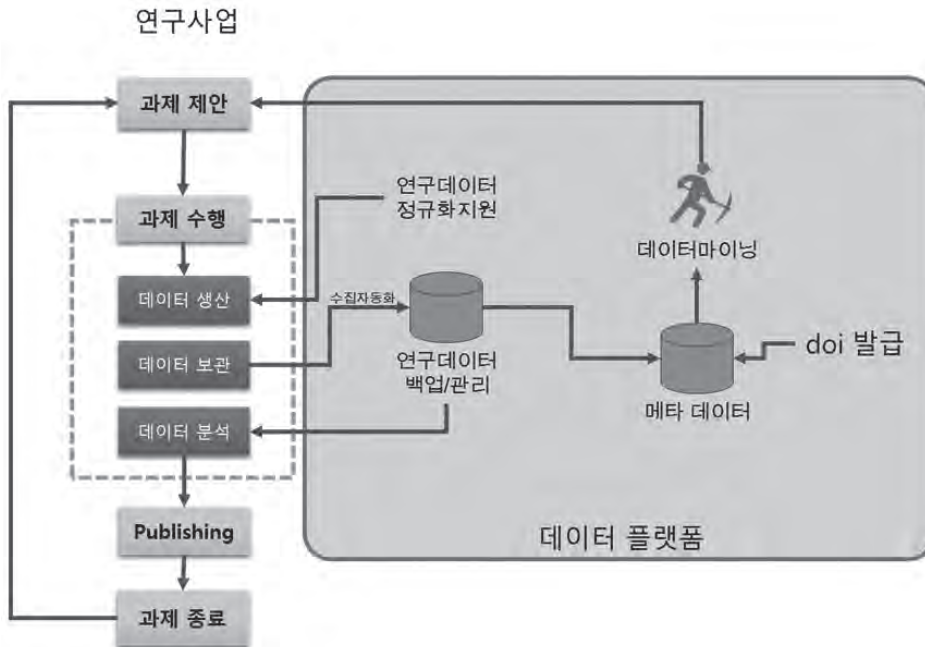
2017년부터 KRICT에서는 화학인프라본부의 화학시물레이션센터에서 KRICT에서 생산되는 연구 데이터를 수집하고 데이터베이스를 구축하여 연구데이터를 활용하기 위한 기반을 만드는 “화학데이터 플랫폼 구축”사업을 시작하였다(그림 3-4-1-16, 17 참조). 이 사업을 통하여 기존에 연구노트 및 아날로그 연구 데이터를 전산화하며 연구 데이터를 수집하고 수집된 연구 데이터를 기반으로 차후 활용 가능한 데이터베이스를 구축하고 있다. 그리고 연구 데이터 자동 수집을 위한 기반 구축을 추진하며 수집한 데이터를 바탕으로 새로운 연구 주제 창출을 위한 데이터 분석을 지원하고 데이터마닝, 머신러닝 등에 적용하기 위한 기반을 마련하는 것을 목적으로 하고 있다. 첫 번째 사업 년도인 2017년에는 KRICT에서 수행하고 있는 유기태양전지 연구 관련 데이터를 대상으로 하여 데이터 수집 및 정리를 하여 데이터베이스를 구축하고 있다.



〈그림 3-4-1-16〉 데이터 플랫폼을 활용한 데이터 수집 및 활용

본 사업을 통하여 수집된 연구데이터는 정규화 과정을 거쳐 관리되고 메타데이터를 추출하며 데이터마이닝을 거쳐 머신러닝 등 소재정보학 플랫폼에서 활용할 수 있도록 가공하며, 향후 KRICT 전체 연구데이터의 수집·활용과 함께 “화학 관련 연구데이터의 리포지토리”를 구축할 계획을 가지고 있다.

연구원 내의 소재 연구데이터 수집 및 활용을 위한 플랫폼 구축과 더불어 화학시뮬레이션 센터에서는 소재 빅데이터에 다양한 머신러닝을 적용하는 소재정보학 연구를 수행하고 있다. 현재 소재정보학 연구는 기존의 공개된 소재 빅데이터 및 KRICT의 계산 물질 데이터베이스를 활용하여 다양한 머신러닝 알고리즘 적용을 통해 소재 물성 예측에 대한 연구를 진행하고 있다. 이러한 소재정보학 연구는 향후 연구원 내에서 구축된 데이터베이스를 활용하는 연구로의 확대를 계획하고 있다. KRICT의 화학데이터 플랫폼 구축, 소재정보학 응용 연구는 기존의 화학소재정보은행과 연계하여 “소재정보학 기반 소재 개발 플랫폼”으로 구축하고 이를 다양한 소재 개발 연구에 적용할 계획이다.



〈그림 3-4-1-17〉 연구사업 지원을 위한 연구데이터의 확보 및 활용

### 3. 빅데이터 기반 소재연구 활성화를 위한 제언

머신러닝을 통해 신뢰할 만한 정보를 추출하여 이를 바탕으로 성공적인 소재의 개발을 달성하기 위해서는, 소재 빅데이터의 구축이 무엇보다도 시급하고 중요하다. 머신러닝과 최적 제어 이론을 결합하여 특정한 문제의 해답을 찾아가는 인공지능에게 빅데이터는 판단의 논리를 제공하는 모수이기 때문이다. 따라서, 데이터는 인공지능을 이용한 데이터 기반의 연구에서 가장 필요한 핵심 자원이며, 이를 체계적으로 수집하고 저장 관리하는 인프라는 미래 R&D 경쟁력의 핵심이라고 할 수 있다.

빅데이터의 구축에는 무엇보다도 제도 및 인프라 관점의 노력이 필요하다. 특히 빅데이터라 함은 기존 물성정보 DB의 규모를 훨씬 뛰어 넘기 때문에 데이터의 수집과정이 자동화

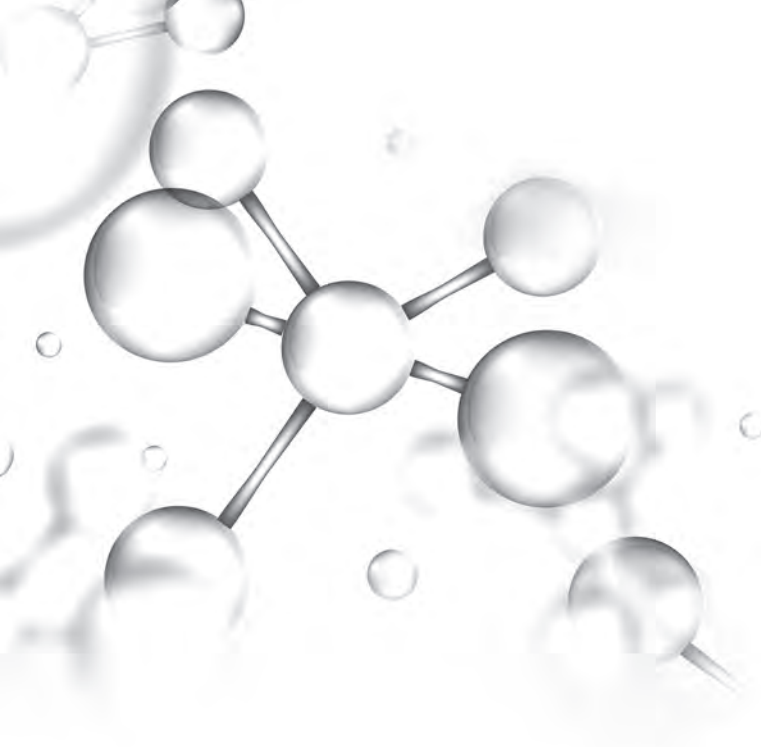
되어야 한다. 특히, 데이터의 수집을 위해 연구자에게 추가의 부담이 가해지면 빅데이터 구축은 실패할 가능성이 높다. 빅데이터 구축을 가장 효과적으로 달성할 수 있는 방법은 플랫폼화된 연구개발 환경을 구축하는 것이다. 실험의 과정을 관리하는 플랫폼, 분석 플랫폼, 계산재료과학 플랫폼 등이 제공되어 플랫폼 상에서 연구를 수행하게 되면, 그 결과들인 데이터가 메타 데이터와 함께 자동으로 축적될 수 있을 것이다. 연구개발 플랫폼의 활성화를 위해서는 플랫폼 상에서 연구개발과정을 편리하게 해주는 다양한 서비스를 개발하여 제공하여야 한다.

두 번째는 데이터의 공유에 관한 재료과학 연구자들의 인식변화가 있어야 한다. 연구자들은 데이터를 생성할 때 과제의 수행이나 논문의 작성 등 특정한 목적으로 가지고 데이터를 생성하고 관리한다. 그러나, 그 목적이 달성되고 난 뒤에는 데이터 관리에 크게 신경쓰지 않는 것이 일반적이다. 빅데이터를 통한 소재개발은 원래의 목적이 달성된 데이터들을 공유하여 빅데이터화 함으로써, 데이터 생성의 고유 목적과는 다른 관점의 가치를 창출하는 것이다. 따라서, 일정기간의 엠바고 기간을 거쳐 데이터를 공유하는 문화가 연구자들 사이에 만들어져야 한다. 또한, 데이터의 공유를 촉진하기 위해서는 데이터를 통한 연구가 어떤 가치를 창출하는가에 대해 연구자들의 구체적 학습 경험과 이를 통한 인식 제고도 필요하다.

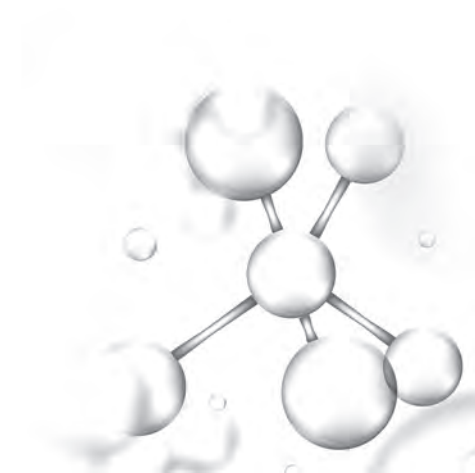
세 번째로는 글로벌 스케일의 데이터 공유 체계를 갖추어야 한다. 소재 정보학은 데이터의 규모가 매우 중요한 분야이다. 작은 규모의 데이터로는 충분히 신뢰할 만한 정보를 추출하지 못할 가능성이 높기 때문이다. 따라서, 어느 한 연구진이나 한 국가가 보유한 데이터만으로는 충분히 효과적인 소재정보학 연구가 진행되지 못한다. 따라서, 세계적인 소재 데이터 구축 노력에 적극 동참하여 데이터 공유를 위한 표준화, 공유를 위한 인터페이스 구축 등 글로벌 스케일의 데이터 공유를 염두에 둔 빅데이터 구축 노력이 진행되어야 한다.

빅데이터의 축적과 함께 소재정보학 분야의 연구개발 역량 강화에도 많은 노력이 필요하다. 거대규모의 데이터가 공유되는 시점에서 이를 최대한 활용할 수 있는 역량을 갖추고 있어야 실질적인 가치를 창출할 수 있기 때문이다. 소재 정보학 분야는 소재 분야의 전문가와 데이터 분석/처리에 익숙한 전산전공자 혹은 데이터 사이언티스트들이 협력하는 융합연구를 통해 직관적이고 사용이 편리한 SW 및 알고리즘 개발이 이루어질 수 있다. 따라서, 정부에서는 적극적으로 데이터 기반의 소재개발 융합연구를 지원하여야 한다. 또한, 정보과학과의 융합 역량 강화를 위해 재료과학 연구자들에 대한 교육이 필요하기 때문에 융합과제를 수행하는 연구진들이 소재정보학의 파급을 위한 교육프로그램을 개발하도록 유도하여야 한다.





# 집필진 및 과년도 소개



| 구분 | 내용                        | 집필자(소속)                          |
|----|---------------------------|----------------------------------|
| 3부 | 3-1. 실리콘 기반 초고집적 반도체 소재기술 | 이병훈(GIST)                        |
|    | 3-2. 인공지능 반도체 소재기술        | 박병국(서울대)                         |
|    | 3-3. 뉴로모픽 소자 기술           | 최창환(한양대)                         |
|    | 4-1. 빅데이터 기반 소재 연구        | 이광렬, Hiroshi Mizuseki, 김동훈(KIST) |
|    | 4-2. 소재정보학 기반 소재기술        | 김형준(KAIST)                       |
|    | 4-3. 금속소재 가상공학 플랫폼 기술     | 이호원(재료연), 김동규(원자력연),<br>이진우(재료연) |
| 4부 | 특집: 인공지능과 소재              | 조효진(재료연)                         |

