



05

2021 May | Vol. 7

융합연구리뷰

Convergence Research Review

AI 로봇이 스스로 소재를 개발한다고?

한상수(한국과학기술연구원 책임연구원)

김동훈(한국과학기술연구원 선임연구원)

소재연구를 위한 빅데이터 및 플랫폼 활용

안재평(한국과학기술연구원 책임연구원)

김홍규(한국과학기술연구원 선임연구원)

CONTENTS

- 01 편집자 주
- 03 AI 로봇이 스스로 소재를 개발한다고?
- 39 소재연구를 위한 빅데이터 및 플랫폼 활용



융합연구리뷰 | Convergence Research Review
2021 May vol.7 no.5

발행일 2021년 5월 10일

발행인 김현우

발행처 한국과학기술연구원 융합연구정책센터

02792 서울특별시 성북구 화랑로 14길 5

Tel. 02-958-4977 | <http://crpc.kist.re.kr>

펴낸곳 디사플래닝 Tel. 02-6315-4600



● AI 로봇이 스스로 소재를 개발한다고?

작년 9월부터 LG사이언스파크는 캐나다의 토론토대학교와 맥마스터대학교, 프랑스 에너지·석유회사 토탈(Total)과 함께 'AI 기반 소재 개발 컨소시엄(A3MD, The Alliance for AI-Accelerated Materials Discovery)'을 결성하고 친환경 촉매, 차세대 소재 등 화학소재 개발을 목표로 공동연구를 시작했다. AI 기반 소재 개발 컨소시엄의 특징은 글로벌 최고 수준의 학계와 산업계가 참여하는 세계 최초 AI연합이라는 사실 외에도 화학소재 분야에 인공지능 및 로봇기술을 활용한다는 점이다.

오늘날 인공지능 및 로봇은 다양한 분야에 접목되고 있다. 소재분야에서도 예외가 아니다. 신소재를 찾기 위해서는 힘든 실험을 거쳐 소재를 합성하고 물성을 평가하는 과정을 끝없이 반복해야 하는데 인공지능을 활용함으로써 소재 개발 비용 및 시간을 대폭 단축할 수 있게 되었다. 뿐만 아니라 인공지능은 연구자의 도움 없이 연구자가 원하는 물성의 소재들을 스스로 합성함으로써 신소재 연구개발에 도움을 주고 있다.

기존의 성능을 뛰어넘는 소재에 대한 수요가 급증하고 있는 가운데, 복잡한 데이터를 활용하여 인간이 예상하지 못한 결과를 도출하는 인공지능의 적극적 활용으로 소재분야에서 우수한 경쟁력을 갖게 되기를 기대하며, 1부에서는 소재 자율실험실을 소개한다.

● 소재연구를 위한 빅데이터 및 플랫폼 활용

2020년 10월 4일자 헬로디디 뉴스기사에 따르면, 일본 신에너지 산업기술종합개발기구는 플렉시블 투명필름 개발에 인공지능을 활용한 결과 실험횟수를 기존대비 25분의 1 이하로 줄이는데 성공했다는 내용을 발표했다고 한다. 이는 숙련된 연구원이 25차례 실험을 거듭해도 발견하지 못한 특성을 인공지능은 1차례의 실험을 통해 얻어낸 것이라고 한다. 또한 개발된 플렉시블 투명필름을 살펴본 결과, 연구원이 제안한 25가지 후보보다 인공지능이 제안한 3가지 재료 후보 특성이 우수한 값을 보였다고 한다. 이를 가능하도록 한 것은 충분한 데이터가 있었기 때문이었다. 인공지능은 사전에 27종의 플렉시블 투명필름의 구조와 특성 관계를 기계적으로 학습했다고 한다.

현재 인공지능의 활용은 여러 분야에서 큰 혁신을 일으키고 있다. 그런데 데이터가 없는 상황에서의 인공지능은 무용지물이다. 인공지능은 소재분야에서 소재의 물성 예측, 공정의 에러 방지 등 넓은 범위에서 활용이 되고 있는데 그 중에서도 신소재 개발 시 기존에 있는 소재의 결합으로 새로운 물성을 찾아내기 때문에 소재데이터 및 관리는 중요할 수밖에 없다.

과학기술정보통신부는 2020년 '데이터 기반 소재연구 혁신허브 구축·활용 방안'을 발표하며 인공지능을 신소재 R&D에 활용하기 위해 2021년까지 학습데이터 420만 건을 수집할 계획이라고 한다. 2부에서는 미래의 중요한 자산인 소재데이터에 대한 내용을 다룬다.



융합연구리뷰

Convergence Research Review 2021 May vol.7 no.5



01

AI 로봇이 스스로 소재를 개발한다고?

한상수(한국과학기술연구원 책임연구원)
김동훈(한국과학기술연구원 선임연구원)

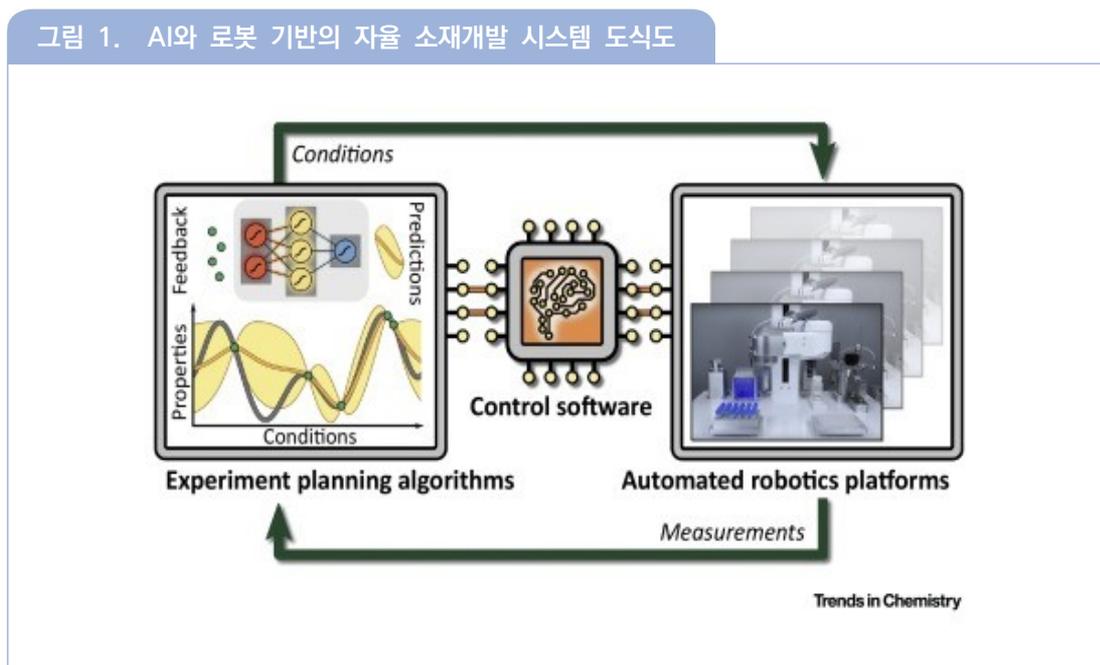
I 서론

최근 계산과학을 이용한 소재개발 성공사례가 발표되면서 계산과학을 이용한 소재개발 방식은 ‘뜨거운’ 연구분야로 여겨지고 있다. 한편, 2015년쯤부터 인공지능(Artificial Intelligence, 이하 AI) 기술이 소재설계 분야에서 많이 활용되고 있다. AI를 위해서는 데이터베이스(Database, 이하 DB)가 필수적인데, 2015년경부터 전 세계적으로 계산과학 기반의 소재 DB를 구축하기 시작하면서 AI가 소재분야에도 본격적으로 도입되었기 때문이다. 이러한 이유로 현재 세계적으로 소재 DB에 대한 관심이 높아지고 있고, 우리나라에서도 2020년에 과학기술정보통신부 지원으로 국가소재연구데이터센터가 설립되었다.

AI 기술이 소재분야에 빠르게 접목되는 이유는 무엇일까? 여러 이유가 있겠지만 크게 두 가지로 요약할 수 있다. 첫 번째 이유는 기존 계산과학 기술에 비해 소재의 물성을 훨씬 빠르게 예측할 수 있다. 계산과학으로는 몇 달씩 걸리던 계산을 AI를 이용하면 몇 분만에 결과를 얻을 수도 있다. 또 다른 이유는 AI를 이용하면 기존 계산과학 기술로는 수행하기 어려웠던 연구가 가능하다. 우리는 이 두 번째 이유에 더 큰 의미를 부여하고 싶다. 최근 컴퓨터 기술이 발달하면서 과거에 비해 계산과 실험 사이의 간극이 많이 좁혀졌으나 여전히 많은 경우에서 간극이 존재한다. 대표적인 예로, 계산과학으로 예측된 소재가 실험에서는 합성이 되지 않는 경우와 계산과학으로 예측된 소재 물성이 실제와 차이나는 경우 등이 있다. 계산-실험 간의 이러한 간극이 발생하는 가장 큰 이유는 계산과학으로 예측된 물성은 주로 소재 본연(intrinsic)의 특성만을 보기 때문이다. 하지만 실제 측정되는 소재 물성은 소재 본연의 특성뿐만 아니라 외부환경(extrinsic)에 의해서도 많은 영향을 받는다. 기존 계산과학(제일원리계산 등)은 소재의 ‘구조-물성’ 간의 상관관계에 주로 역점을 두었으나 소재-실험 간의 간극은 소재의 ‘구조-물성’ 연구만으로는 해소하기 힘들다는 것을 알 수 있다. 이 간극을 없애기 위해서는 합성(공정)에 대한 이해, 즉, ‘구조-물성-공정’간 상관관계 이해로의 연구 확장이 필요하고, 이를 위해서 AI 기술이 중요한 역할을 한다.

“그러면 AI 기술로 어떻게 ‘구조-물성-공정’상관관계를 이용해서 신소재 개발을 가속화 할 수 있을까?” 라는 의문이 생긴다. 이에 대한 예로, AI 기술로 소재 합성·공정 조건을 예측하고, 이를 기반으로 로봇을 통해 소재합성 및 성능측정을 자동화하는 방식이 있다. 이 방식에는 단순히 지정된 후보물질에 대한 소재합성·성능

측정의 자동화에 그치는 것이 아니라, 분석결과를 기반으로 새롭게 다음 실험을 계획하고, 이에 따른 후보물질을 자율적으로 샘플링하고, 다시 합성·측정을 반복함으로써 자율적으로 소재를 개발할 수 있는 Closed-loop 신소재 개발 전략이 사용되며, 이는 최근에 개발되고 있다(그림 1).



출처 : Häse et al.(2019)

기존에도 신소재개발 가속화를 위한 새로운 연구방법으로 조합실험(한 번에 여러 조건에서 신소재합성 혹은 물성을 평가하는 기술)에 로봇자동화를 접목하는 기술이 활용되기도 했었다. 그러나 이 기술은 단순 반복적인 작업의 효율성을 높이지만, 기본적으로 신소재개발 접근방식이 시행착오적 연구방식에서 탈피하지 못하는 단점이 있다. Closed-loop 신소재 개발 전략은 로봇에 AI 기능을 접목하여(지능을 갖도록 하여), AI 로봇이 스스로 소재 설계부터 합성·분석까지 진행하며 신소재를 개발할 수 있도록 한다. 이런 점에서 이 기술을 소재 자율실험실 (self-driving lab)이라 부른다. 이는 자율주행 자동차와 개념적으로 유사하다. 완전 자율주행자동차가 실현되면 인간이 원하는 목적지만 지정해주면 자동차가 스스로 목적지까지 데려다 줄 것이다. 자율실험실도 인간이 원하는 소재 혹은 물성만 입력해주면 AI 로봇이 스스로 해당 소재를 개발해주게 된다. 소재 연구자들에게 이런 꿈만 같은 일들에 대해 최근 해외 선진국을 중심으로 연구가 시작되고 있다.

흔히 4차 산업혁명 시대가 도래함에 따라 소재개발 분야에도 큰 변화가 있을 것이라고 한다. 하지만, 정작 어떻게 변화를 가져다 줄 것인지를 구체적으로 설명할 수 있는 연구자는 많지 않다. 우리는 소재 자율실험실 (self-driving lab)이 그 한 예가 될 것이라 확신한다. 머지않아 소재 연구에서 자율실험실(self-driving lab)을 흔하게 보게 될 것이고, 이는 소재개발의 새로운 패러다임으로 이어질 것으로 예견된다. 이런 시점에서, 소재 분야의 자율실험실 연구동향 및 사례를 살펴보고, 향후 발전방안과 대응 방안에 대해 알아보는 것은 매우 중요하다. 융합연구리뷰는 관련 연구동향 분석, 향후 발전방안 및 맺음말로 구성되어 있다.

II 소재 자율실험실 연구동향

1. 자동화(Automation)와 자율화(Autonomy) 차이점 이해

본격적인 연구동향을 파악하기에 앞서 용어 정리를 할 필요가 있다. 자동화 프로세스(automated process)와 자율화 프로세스(autonomous process)의 차이를 이해해야 한다. 이 두 용어는 혼동해서 사용하는 경우가 많은데 이를 혼동하면 소재 자율화실험실의 개념을 완전히 이해하지 못하는 것이므로 용어들을 명확히 구분해서 이해하는 것이 매우 중요하다. 아래 설명과 <그림2>를 함께 확인하는 것이 차이점 이해에 도움이 될 것이다.

먼저, 자동화(automation)와 자율화(autonomy)는 제어 시스템과 정보기술을 조화롭게 사용하여 산업 기계류와 공정을 제어함으로써 사람이 관여할 필요를 줄이는 접근방식이라는 점에서 유사해 보인다. 그러나 자동화 프로세스(automated process)는 인간의 명령대로 움직이는 프로세스를 의미한다. 즉, 로봇이 자동화 실험을 한다는 것은 인간(연구자)이 미리 프로그램에 어떤 실험들을 수행해야 하는지 입력을 해두면 로봇은 입력값에 따라서 충실히 실험을 수행하는 것이다. 기본적으로 로봇 기계의 움직임을 비롯한 모든 제어통제권이 인간(연구자)에게 있는 것이다. 반면에 자율화 프로세스(autonomous process)는 어떠한 실험을 수행할지를 인간이 아니라 AI 로봇이 자체적으로 결정한다. 즉, 실험 순서나 조건들을 AI 로봇이 스스로 판단하여 수정해 나가면서 실험을 진행하는데, 이를 가능하게 하는 것은 바로 AI 로봇에 탑재된 빅데이터와 인공지능 소프트웨어 때문이다. 자동화 시스템에서는 인간이 실험조건을 일일이 입력값으로 제공해주어야 하지만, 자율화 시스템에서는 그럴 필요가 없다. 오로지 얻고자 하는 최종물성(target property)만 알려주면 그 물성을 가지는 소재를 합성해주는 모든 프로세스는 AI 로봇이 자율적으로 판단한다.

자율화(autonomy)의 개념을 이해하는데 가장 쉬운 유사예시는 바로 자율주행차(self-driving car 또는 autonomous vehicle)이다. 자율주행차는 단순한 자동화 차량이 아니다. 자율주행차도 자율화의 정도에 따라서 레벨1부터 레벨5까지 다양하게 정의되지만, 궁극적 목표인 레벨5 자율주행차의 경우는 인간이 자동차의 움직임에 대한 능동적 통제권을 더 이상 가지지 않는다. 물론 심각한 안전문제가 발생하였을 때는 인간(운전자)이 개입할 수는 있겠으나, 평소에는 자동차의 움직임에 세세한 명령을 내리지 않는다. 이러한 것이 가능한 이유는

자율주행차에는 수많은 사진들을 바탕으로 한 빅데이터와 이를 훈련시킨 인공지능 모델이 탑재되어 있기 때문이다. 덕분에 탑승자는 최종 목적지만 알려주면, 자율주행차는 어떠한 길을 선택할지, 보행자를 어떻게 피해갈지, 교통 신호를 어떻게 지킬지 등에 대한 자율적 판단이 가능하며 안전하게 목적지에 도착할 수 있게 되는 것이다. 자율주행차의 실험실 버전을 자율실험실로 이해하면 쉽다.

융합연구리뷰에서는 주로 자율화실험실 개발의 연구동향을 다루지만, 독자들의 이해를 돕기 위해서 그보다 수년, 길게는 10년 먼저 구축된 자동화 실험실의 2가지 예시들을 먼저 소개한다. 그 다음에 지난 3년간 미국과 영국을 중심으로 활발하게 발전되고 있는 자율화실험실에 대한 상세한 예시들을 소개한다.

그림 2. 자동화(automation)와 자율화(autonomy)의 차이점

AUTONOMY FROM START TO FINISH			
LEVEL	CONCEPT	DEFINITION	WHO'S IN CONTROL
0	Human Operation	The operator controls the machine at all times.	
1	Automation (Function-specific)	The operator has overall control of the machine and is responsible for its safe operation, but can transfer limited control over a specific function (like moving a bucket or blade) to the machine.	
2	Semi-autonomous	The machine accomplishes a subset of its defined tasks without operator interaction. The operator performs the remaining tasks.	
3	Autonomous	The machine accomplishes all its defined tasks without operator interaction and is responsible for all safety-critical earthmoving functions.	

출처 : Caterpillar 홈페이지

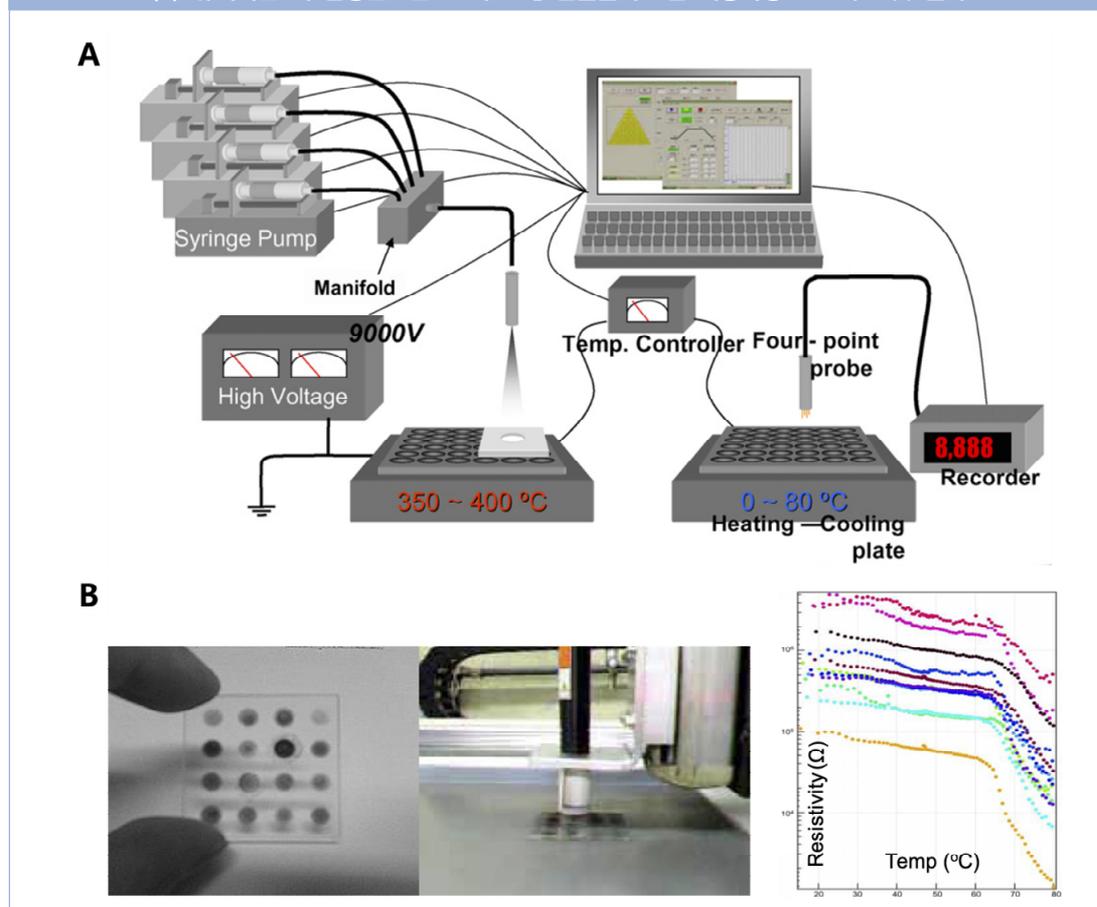
2. 소재개발 자동화 프로세스(Automation)

실험실의 루틴(routine)한 과정을 자동화시키고자 하는 욕구는 대학교 연구실 또는 기업체 연구소에서 오래도록 존재했다. 자동화에 성공하면 인간이 해야 하는 일들을 현저하게 줄일 수 있기 때문이다. 많은 연구 사례들 가운데, 2008년 일본 도쿄대학교 후지모토(Fujimoto) 교수팀의 연구보고를 가장 먼저 소개한다(Fujimoto et al., 2008). 후지모토(Fujimoto) 교수팀은 대표적인 열색성 재료(thermochromic material)인 바나듐 산화물(VO_2) 박막 합성과 이들의 물성평가 과정을 자동화하기 위한 자동화 플랫폼을 만들었다(Fujimoto et al., 2008). 자동화 플랫폼의 이름은 “M-ist Combi”이다. M-ist Combi는 자동으로 샘플들을 준비하고 물성을 평가한다. 연구팀은 바나듐 산화물로 재료를 선택하였는데, 이 재료는 열색성 화합물로서 68°C 에서 금속-반도체 상 변이가 일어나는 소재이다. 68°C 이상에서는 사각형의 금홍석(tetragonal rutile)구조를 가지면서 금속성 성질을 가지지만, 68°C 이하에서는 단사정(monoclinic) 구조로 바뀌면서 반도체성 성질을 가진다.

후지모토(Fujimoto) 교수팀은 <그림3A>에 소개된 것처럼 첨가제(dopant)의 양을 바꿔가면서 서로 다른 바나듐 산화물(VO_2) 박막을 자동으로 합성하였다. 주사기 펌프(syringe pump)에서 토출되는 첨가제의 용량을 조절함으로써 서로 다른 박막재료를 연속적으로 합성하는 것이 가능하다. 바나듐 산화물(VO_2) 박막과 같은 무기물을 합성하는 데에는 통상적으로 고가의 진공 장비(sputtering 장비 등)를 사용하지만, 본 연구에서는 하드웨어 구축의 편의성을 위하여 솔루션(solution) 상태에서 합성 가능한 합성법을 채택하였다. 합성 완료된 박막의 두께는 약 300nm, 평균 50-100nm 직경의 입자들로 구성되었다.

이어 교수팀은 물성(이 경우 저항 resistivity)을 자동으로 평가하는 시스템을 개발하여 자동합성 파트와 연동시켰다. 해당 시스템에서 사용한 측정 장비는 4점 저항측정장비(four-point resistance method)이다. <그림 3B>에서 보이는 바와 같이 유리기판 위에 16개의 바나듐 산화물(VO_2) 박막이 분리되어 합성이 되어있고, 이 박막들 위로 4점 저항측정장비가 위치를 옮겨가서 저항값을 측정하였다. 교수팀은 M-ist Combi 플랫폼을 이용하여 총 16개의 샘플에 대해 재료를 합성 및 물성평가를 하는데 총 2시간이 채 걸리지 않았다고 보고하였다. 본 플랫폼은 인간이 루틴(routine)하게 하는 실험 합성 및 평가과정을 기계적으로 자동화하는데 그치지만, 자동화를 위한 하드웨어 구축에 성공했다는 점에서 의의가 있다.

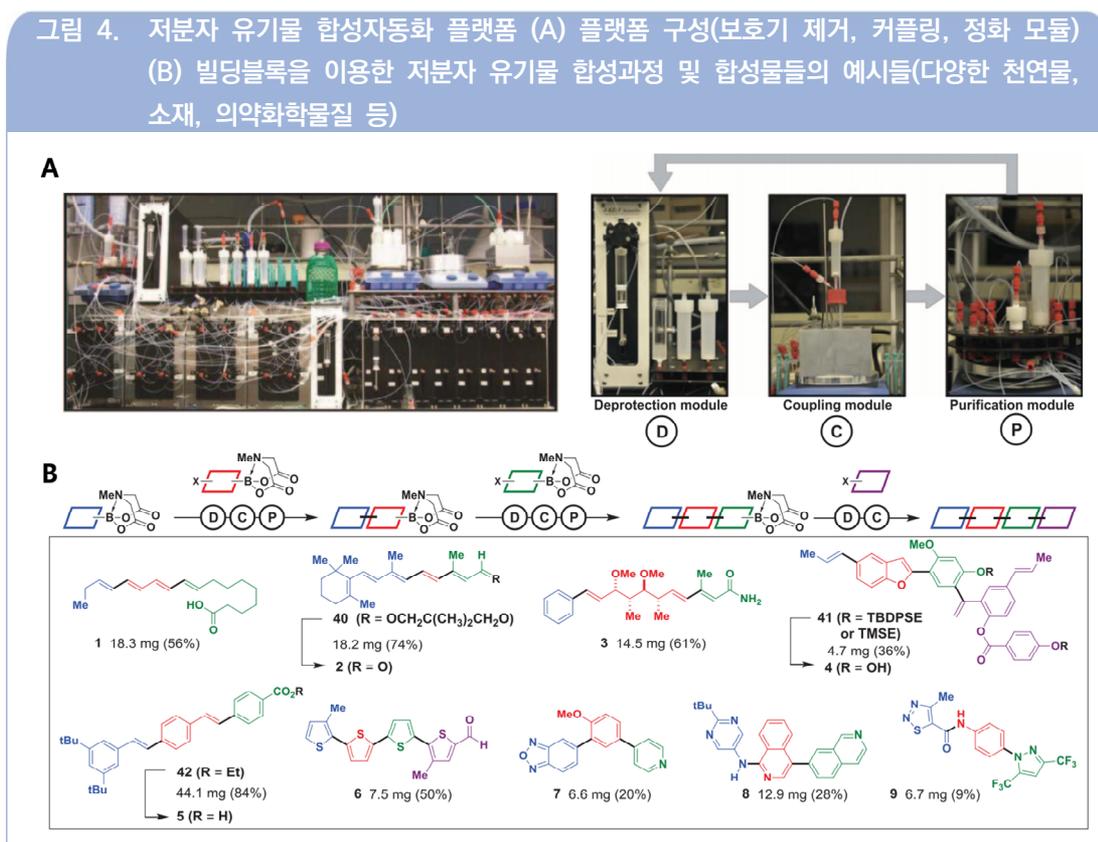
그림 3. 열색성 VO₂ 박막 자동합성 및 저항측정 자동화 플랫폼 (A) 자동화 플랫폼 구축 도식도 (B) 유리기판 위 합성완료된 16개 VO₂ 샘플들과 4점 저항측정 프로세스 및 결과



출처 : Fujimoto et al.(2008)

다음으로 소개할 자동화 플랫폼 사례는 미국 일리노이 공대의 버크(Burke) 교수팀의 연구사례이다. 버크(Burke) 교수팀은 유기물 합성 자동화 플랫폼을 2015년에 Science지를 통해서 공개한 바 있다(Li et al., 2015). 연구팀은 저분자 유기물(small organic molecules)의 합성 자동화에 집중하였는데, 저분자 유기물은 의학과 공학에서 매우 가치있게 사용되는 물질의 한 종류이다. 대표적으로 유기 발광 다이오드(OLED, Organic Light Emitting Diodes)에 발광체로 많이 사용되며, 의학에서도 암세포 추적용 발광체 물질로써 많이 활용된다. 저분자 유기물들의 물성을 최적화하기 위해서는 실험연구자들이 수많은 실험을 통하여 다양한 종류의 저분자 유기물들을 합성하여야 하는데 이는 시간과 에너지 소모가 심한 작업이다.

버크(Burke) 교수팀은 이를 해결하기 위하여 유기물 합성용 자동화 플랫폼을 만들었다(그림4). 플랫폼 내의 합성 시스템은 크게 세 가지 모듈인 보호기 제거 모듈(deprotection module), 커플링 모듈(coupling module), 그리고 정화 모듈(purification module)로 구성이 되어있다. 이 세 가지 과정(보호기 제거, 커플링, 정화 과정)은 저분자 유기물을 합성하는데 필수적인 과정이므로 각각의 과정을 모듈화하여 단일화된 플랫폼으로 연결하였다.



출처 : Li et al.(2015)

이 시스템을 활용하는 경우 총 14개의 서로 다른 종류(분자의 구조나 성질에 따른 분류)의 저분자 유기물들이 생성 가능하다는 것을 데모(demo)로써 보여주었다. 합성 과정은 기본적으로 빌딩블록(building block) 기반의 합성 방식을 채택하였는데, 빌딩블록은 수천 개 이상이 이미 상업적으로 구입 가능하므로 이 플랫폼으로 엄청난 수의 저분자 유기물 합성이 가능할 것으로 전망된다.

본 연구는 저분자 유기물 합성의 복잡한 프로세스를 자동화시킨 점에서 큰 주목을 받았으며, 산업체에도 큰 파급력을 끼쳤다. 그러나 본 연구는 합성 자동화에만 초점이 맞추어져 있으며, 생성된 유기물들에 대한 물성 평가 작업과 관련된 플랫폼은 존재하지 않는다. 물질들을 대량으로 생산하는 데에는 특화되어 있으나, 플랫폼 내에 연구자들이 평가하고자 하는 물성측정 파트가 존재하지 않으면 소재 합성의 자율화를 이룰 수 없다.

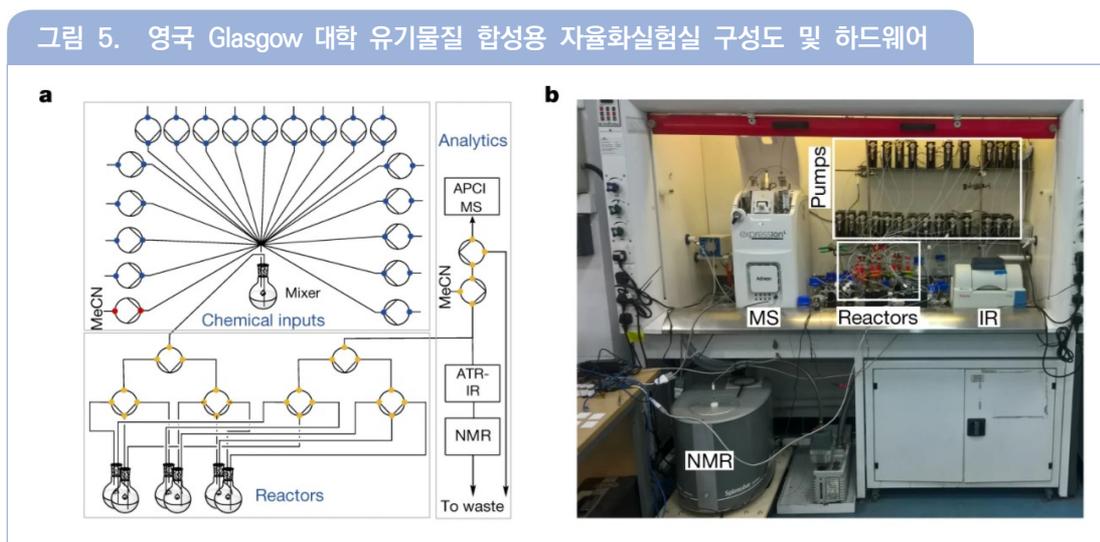
3. 소재개발 자율화 프로세스(Autonomy)

앞의 2.2장에서 소개한 사례 이외에도 하드웨어적인 부분에서 자동화 플랫폼 구축에 성공한 사례는 무수히 많고, 산업체에서도 비용 감소를 위해 공정 자동화 플랫폼들을 자체적으로 구축하고 있다. 그러나 융합연구 리뷰에서는 앞의 2건 이외의 '자동화' 사례들에 대한 소개는 생략하고, 소재개발 '자율화' 프로세스(autonomous process)에 초점을 두고 관련 연구사례들을 소개한다.

인터넷의 발달과 더불어 사물인터넷(IoT, Internet of Things) 시대의 등장, 그로 인해 축적되는 빅데이터, 그리고 컴퓨터 하드웨어 중 하나인 그래픽 처리 장치(GPU, Graphics Processing Unit)의 발전으로 인공지능 분야에서 딥러닝(deep learning)의 개발 속도가 비약적으로 증가하면서, 재료과학 연구자들은 자동화되어 있는 실험실에 인공지능을 도입하기 시작하였다. 인공지능이 도입되기 전에는 연구자들의 사전지식을 기반으로 실험들이 이루어졌다. 반면에 인공지능 모델을 탑재한 자율화실험실은 합성-물성 간의 상관관계에 대한 연구자들의 사전지식 없이도 연구자들이 원하는 물성의 소재들을 스스로 합성한다. 이와 같은 자율화 시스템이 고도화됨에 따라 재료과학 분야에서 신소재 개발 및 새로운 공정에 대한 연구개발은 점점 가속화되고 있다. 최근 들어 대규모 국가주도 연구사업이 진행되고 있는 유럽과 미국을 중심으로 Closed-loop 신소재 개발법을 적용한 무인실험실에 대한 많은 연구보고가 이어지고 있다. 무인실험실 특성상, 데이터·AI·로봇·소재분석 등의 기술 융합 및 시스템 통합을 위해 다양한 분야의 전문가가 필요하나 대규모 국가주도 연구사업 없이 다학제 연구팀 구성이 어렵기 때문에 연구개발 성과가 미국, 유럽에 집중되는 것으로 판단된다. 특히 지난 2-3년간 집중적으로 성과가 도출되고 있는데, 2018년 영국 글래스고우(Glasgow) 대학을 비롯하여 영국 리버풀대, 미국 MIT 공대, 하버드대, 캐나다 토론토대 등이 앞장서고 있는 반면, 반면 아시아권(중국, 한국, 일본)에서는 괄목할만한 성과가 보고되지 않았다. 아래에서는 자율화실험실이 발전되어 온 순서와 양상을 보여주기 위해서, 문헌보고가 된 시간 순으로 연구 성과들을 소개한다. 그리고 각 사례에서 어떤 요소들이 새롭게 추가되고 있고, 그러한 요소들이 신소재 개발을 어떻게 촉진시키며 학문적 또는 산업적 의의를 가지고 있는지 덧붙인다.

3.1. 지능형 소재개발 자율화실험실의 첫 걸음(영국 글래스고 대학 Cronin 교수팀)

영국 글래스고 대학 크로닌(Cronin) 교수팀은 유기물 합성 및 분석용 실험실에 인공지능 개념을 처음으로 탑재하여 2018년 7월 Nature지에 자율화실험실을 보고하였다(Granda et al., 2018). 하드웨어 관점에서는 기존의 유기물 합성 자동화 플랫폼과 상당히 유사하다. <그림 5>에 나타나 있듯이, 하드웨어는 크게 용액 주사 펌프, 반응기, 분석기, IR(Infrared spectroscopy), NMR(Nuclear Magnetic Resonance)로 구성된다.



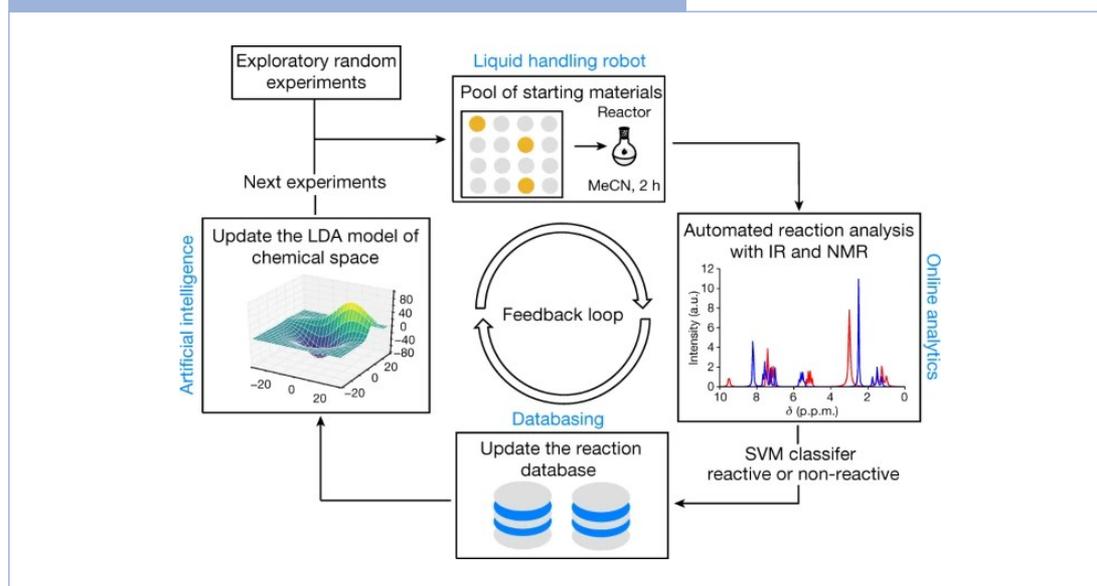
출처 : Granda et al.(2018)

본 연구의 목표는 유기물 합성 및 분석의 자동화 시스템과 인공지능 플랫폼을 접목시켜, 반응성이 있는 새로운 화학반응들을 찾아내는 것이다. 크로닌(Cronin) 교수팀은 많은 유기물 합성반응 종류 가운데 Suzuki-Miyara 화학반응을 주로 탐구하였다. 이 화학반응에서는 반응물, 리간드, 베이스, 용매 등의 다양한 물질들이 사용되는데, 이 물질들의 종류와 용량을 다르게 하였을 때의 반응성(반응유무 또는 수율)을 예측하기 위한 머신러닝 모델을 개발하였다. 이를 바탕으로 새로운 화학반응 합성법을 개발하는 방식이다.

자율화실험실에서는 실험조건을 인간이 아니라 인공지능 모델이 스스로 결정하기 때문에 환류체계(feedback loop)가 중요하다. <그림 6>에 정리된 본 연구에서의 환류체계는 다음의 1~4 단계를 계속적으로 반복하며 최적의 반응조건이 나오면 looping을 멈춘다. 1) 먼저 Suzuki-Miyara 화학반응에 사용되는 다양한 물질들을 랜덤으로 선택하여 유기물 합성반응을 진행한다. 2) 합성 완료된 소재에 대하여 IR 및 NMR 장비로 물성을

평가한다. 3) 머신러닝 모델(IR 및 NMR 스펙트럼을 인풋(input)으로 사용하는 support vector machine 분류모델)을 활용하여 반응성 유무를 판단한다. 4) 예측된 반응성 정보를 바탕으로 합성 데이터베이스와 인공지능 모델을 수정하고, 다음 실험조건을 결정한다.

그림 6. 유기물 합성을 위한 자율화실험실의 환류체계



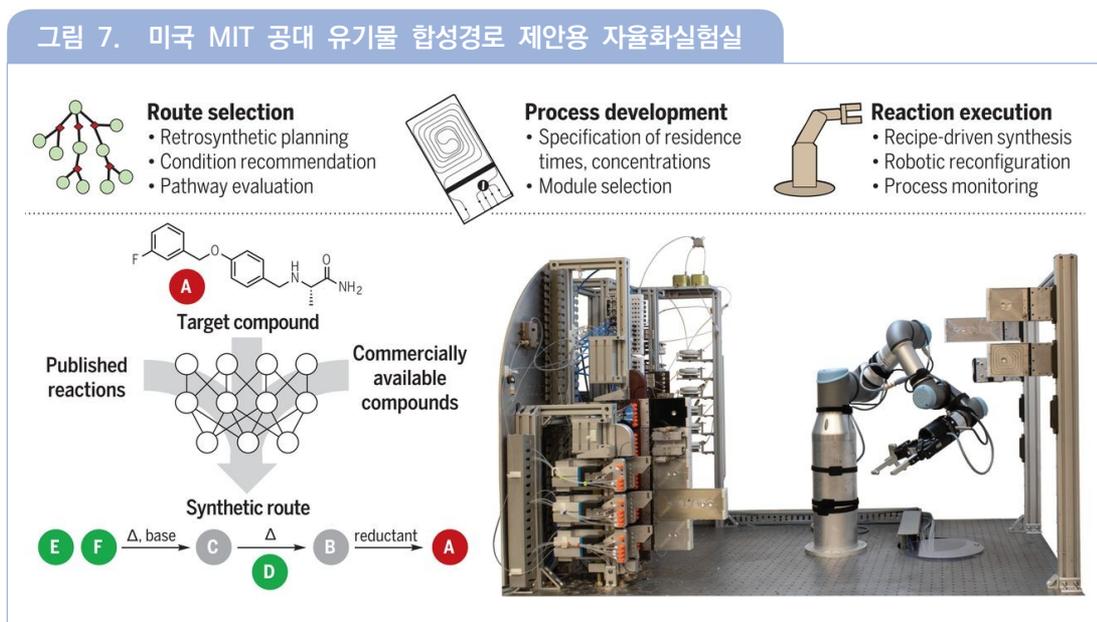
출처 : Granda et al.(2018)

해당 연구는 물질합성 영역에서 자율화실험실의 개념을 최초로 도입한 연구라는 점에서 큰 의미가 있으며, 재료과학에서 인공지능과 자율화실험실이 성공적으로 융합될 수 있는 가능성을 시사했다는 점에서 학계에서 높은 평가를 받았다.

3.2. 협동 로봇을 도입한 유기물 합성(미국 MIT 공대 Jensen 교수팀)

미국 MIT 공대 젠슨(Jensen) 교수팀은 연구자가 원하는 유기물 소재를 합성해주는 지능형 자율화실험실 개발에 성공하였고, 2019년 8월 Science지에 발표하였다^[4]. 본 연구에서는 연구자가 원하는 유기물 소재의 화학구조를 입력값으로 제공하면 자율화실험실이 그 소재를 합성하는 것을 목표로 한다(그림7). 화학구조가 입력되면 인공지능 모델이 해당 물질의 구조를 파악한 후 합성의 성공을 위해서 어떤 시약이 필요한지, 또는 어떤 시약의 조합이 경제적인지 등을 판단하여 초기 반응물 시약들을 선정한다. 그 후, 합성 공정을 인공지능이

제안하면, 협동 로봇(robotic arm)이 해당 시약이 담겨있는 모듈들을 직접 옮기면서 실제 합성을 진행한다. 이 과정대로 원하는 유기 물질이 나올 때까지 연구자의 개입 없이 실험을 반복한다.



출처 : Coley et al.(2019)

본 연구에서는 협동 로봇(robotic arm)을 도입하여 합성과 분석을 진행하였다는 점에 주목한다. 로봇공학의 발전으로 이미 시중에 연구용 또는 산업용으로 판매되고 있는 로봇들의 움직임은 유연하고 정교하다. 화학 합성반응이 복잡해질수록(즉, 시약의 종류가 다양해질수록), 또는 합성, 분석, 물성평가 모듈들이 물리적으로 멀리 떨어져 있을수록 협동 로봇의 도움을 받을 수밖에 없는데, 본 연구에서 협동 로봇의 도입을 통해 효율적으로 실험을 진행할 수 있음을 보여주었다.

합성을 희망하는 유기물의 구조만 넣었을 때 화학 반응의 경로 전체를 예측해낸다는 것은 유기화학자들에게는 진보적인 성과임이 분명할 것이다. 이를 가능하게 하기 위하여 연구팀은 인공지능 모델을 Reaxys 데이터베이스(expert-curated chemistry database)와 미국 특허청(U.S. Patent and Trademark Office)에 있는 접근가능한 모든 유기물 화학반응 경로들을 학습하여 생성하였다. 이후 역합성 분석(Retrosynthesis analysis, 작용기를 변화 혹은 단계적으로 작은 분자로 분해하여 합성법을 설계하는 수단) 기반의 AI 실험계획법을 적용하여 Closed-loop의 설계가 가능하도록 개발하였다. 연구팀은 특허로부터 얻어진 5만 종의 화학반응을 기반으로

개발된 역합성 모델과 협동 로봇을 이용하여 합성법이 매우 복잡한 15종의 의약품 분자를 연구자 개입 없이 자동으로 합성하는데 성공하였다.

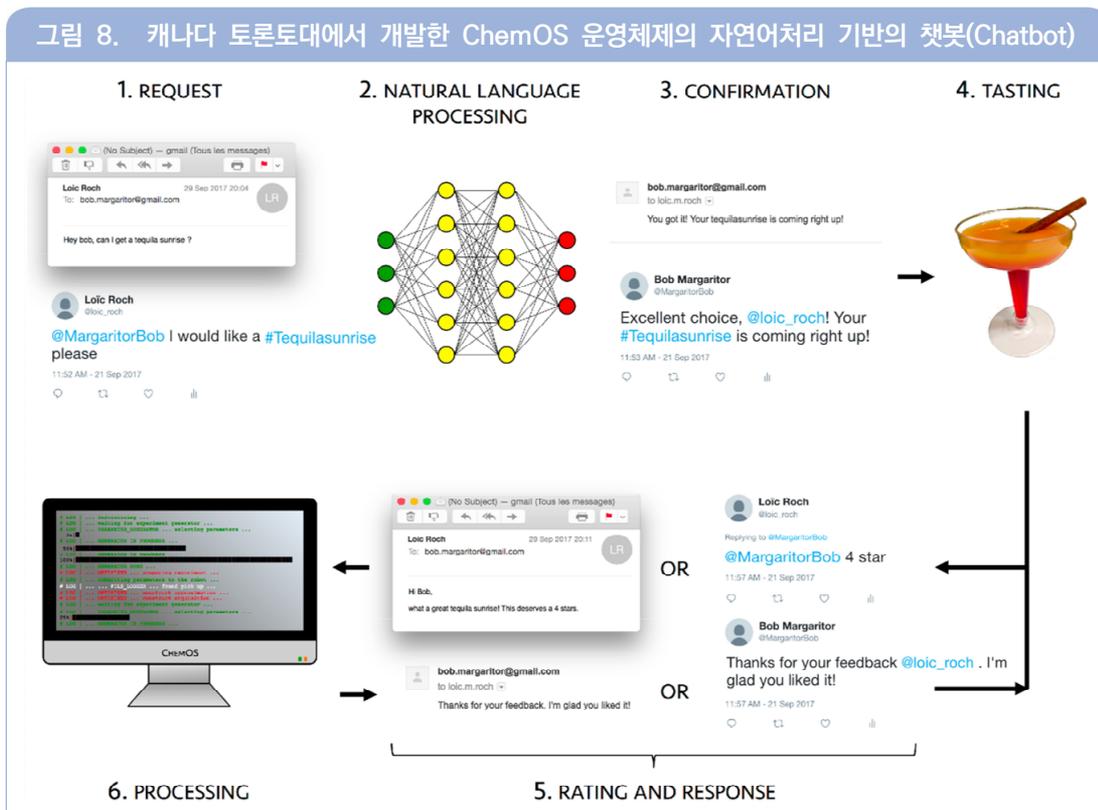
이 연구는 협동 로봇을 소재개발용 자율화실험실에 도입한 첫 연구라는 점, 그리고 플라스크나 바이알에서 일어나는 회분식(batch) 합성이 아니라 용액의 흐름을 이용한 flow synthesis를 도입했다는 점에서 주목할 만하다. 추가적으로 합성 순서를 정할 때, 경제성까지 고려하여 대상 물질 제작에 필요한 시약 선택을 하고, 합성과정을 결정하기 때문에 개발된 인공지능 모델은 상업화 관점에서도 본 연구는 큰 의미가 있다. 다만, 실험실 안전 측면에서 연구자가 부재중일 때 안전사고 발생 시 실험 중단, 책임자 통보 등이 가능한 센서와 제어시스템이 필요하다. 또한 유동화학 기반 합성 시스템과 역합성 모델은 무기소재 개발에 적용하기에는 무리가 있어 이에 대한 확장성은 아직 미흡하다.

3.3. 연구자-플랫폼 소통 챗봇(chatbot) 소프트웨어 개발(캐나다 토론토대 Aspuru-Guzik 교수팀)

캐나다 토론토대 아스푸루-구직(Aspuru-Guzik) 교수팀은 앞으로 개발될 다양한 형태의 소재 자율화실험실에 범용적으로 적용될 수 있는 인공지능 소프트웨어 개발의 필요성을 느끼고 인공지능 소프트웨어(ChemOS)를 개발 및 공개하였다[18]. 자율화실험실의 효율 극대화는 내부에 탑재되어있는 인공지능 모델이 얼마나 효율적으로 다음 실험조건을 제안해주는가에 달려있다고 해도 과언이 아니다. 모델이 최적 조건에 가까운 실험조건을 빠르게 제안해준다면 정답에 금방 수렴할 수 있을 것이고, 궁극적으로 소재개발에 필요한 비용과 시간이 절약될 수 있다. ChemOS는 기본적으로 베이지안 최적화 기법이 탑재되어있는 운영체제이다. ChemOS는 데이터 샘플링 방식 및 머신러닝 학습법에 따라 Phoenix, SMAC, Sperrmint, Uniform의 네 가지 기능을 제공한다.

ChemOS의 흥미로운 기능 중 하나는 연구자와 플랫폼이 소통할 수 있는 자연어 처리 기반의 챗봇(chatbot) 기능이다. <그림 8>의 예시에서는 이해를 돕기 위해서 데킬라(Tequila) 제조를 예시로 보여주는데, 화학합성도 여러 시약들을 솔루션 형태로 섞어서 제조한다는 측면에서 데킬라 제조와 본질적으로 동일하다. 연구자가 원하는 물성조건(request)을 입력하면(예를 들면, Hey bob, can I get a tequila sunrise?), 자연어처리 기법을 이용하여 ChemOS 소프트웨어는 요구조건을 재확인(Excellent choice! Your tequila sunrise is coming right up!) 후 공정 플랫폼에 실제 제조를 명령하게 된다. 추후, 데킬라 맛에 대한 평가(예를 들면, 4 out of 5 stars)를 하면 이를 바탕으로 머신러닝 학습을 통해 새로운 제조 조건을 제안하게 된다. 궁극적으로 자율화실험실은 음성이나 텍스트 기반으로 연구자와 플랫폼이 소통할 수 있는 창구가 마련될 것인데, 이를

위해 자연어처리 기법과 음성인식 등의 고유 소프트웨어 기술들이 도입될 것이다. 아스푸루-구직(Aspuru-Guzik) 교수팀의 예시는 매우 초기 버전이라 아직 몇 가지 예시들에 한정되지만 이들의 시도는 주목할 만하다.



출처 : Roch et al.(2020)

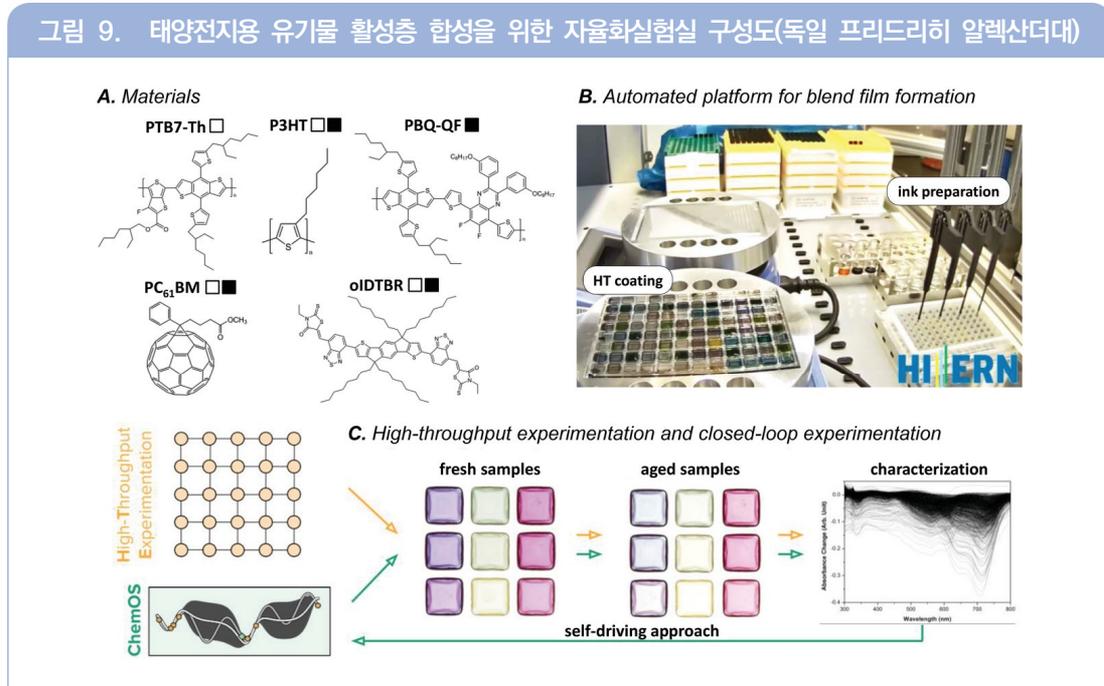
3.4. 적정 기계를 도입하여 태양전지용 활성층 용액 합성(독일 프리드리히 알렉산더대 Brabec 교수팀, 캐나다 토론토대 Aspuru-Guzik 교수팀)

독일 프리드리히 알렉산더대 브라벡(Brabec) 교수팀은 유기물 기반 태양전지(OPV, Organic PhotoVoltaics)를 오랜 기간 연구해 온 연구팀이다. 브라벡(Brabec) 교수팀은 유기물 기반 태양전지 내의 소재를 개발하는 과정을 자율화시키기 위해서 노력해왔으며, 2020년 Advanced Materials에 첫 성과를 보고하였다[11].

본 연구에서의 초점은 태양전지에 활성층으로 사용되는 유기물이다. 유기물은 공기 중에 오랜 시간 노출(aging) 되면 빛을 흡수하는 능력이 빠르게 퇴보한다. 따라서 오랜 시간이 지나도 빛 흡수 능력이 처음과 같이 유지되는

물질을 태양전지의 활성층 재료로 사용하는 것이 바람직하다. 본 연구의 목적은 빛 흡수 능력을 유지할 수 있는 활성층 소재를 찾아내는 것, 더 정확하게는 활성층 내의 최적의 성분 조합비율을 찾아내는 것에 있다.

하드웨어 측면에서는 적정 기계를 도입하였다는 사실이 주목할 만하다. 이전의 연구들에서는 용액 주사 펌프를 사용하였으나, 본 연구에서는 high-throughput 실험이 가능하도록 피페팅(pipetting) 기계장비를 사용하여 용액을 직접 유리기판 위에 적정하여 실험을 진행하였다. 소프트웨어 관점에서는 베어지안 최적화 기법을 탑재한 ChemOS 운영체제를 도입하여 활성층을 구성하는 최적의 용액 조합 비율을 찾아내었다. 4개의 서로 다른 유기물 용액을 섞어가며 실험을 진행하는데, 유기물 이름은 각각 IDTBR, PCBM, P3HT, PTB7-Th이며 이들의 화학구조는 <그림 9>에 나타나있다. 최적화 결과, IDTBR과 PCBM, 이 두가지 시약을 첨가제로 넣었을 때 다른 물질에 비해서 빛 안정성이 떨어지는 것을 확인하였다. 반면, P3HT와 PTB7-Th 이 두가지 시약이 전체 용액의 대부분을 차지할 때 빛 안정성이 가장 높음을 보여주었다.



출처 : Langner et al.(2020)

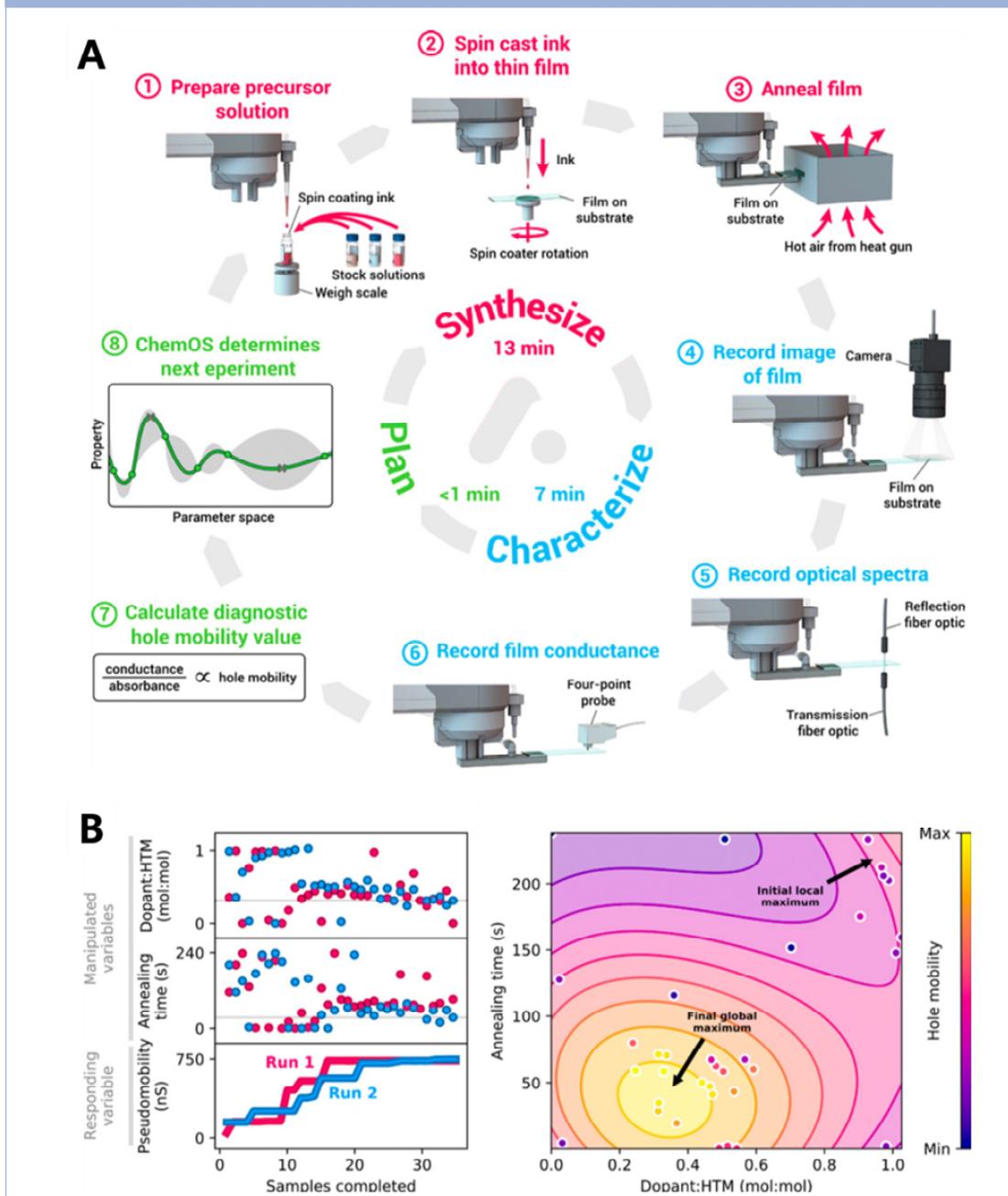
이 연구는 연구자가 변수를 일일이 조정하면서 오랜 시간 걸려 실험을 하지 않아도, 머신러닝을 도입한 자율화실험실이 실험적인 비용(재료비, 실험 시간 등)을 줄여준다는 점을 특히 강조한다. 기존의 high-throughput

접근법을 사용할 경우 2,000번의 실험을 진행하는데 약 100mg 수준의 시약이 필요한 반면, 자율화실험실을 사용할 경우 같은 실험들을 수행하는데 1mg 이내 수준의 시약이 필요하다. 즉, 사용하는 재료의 양이 100배 이상 절감되는 효과가 있다.

캐나다 토론토 대학의 아스푸루-구직(Aspuru-Guzik) 교수팀은 위 사례와 매우 비슷한 연구결과를 동일년도(2020년 5월)에 Science Advances 저널에 보고하였다[15]. 본 연구에서는 페로브스카이트 태양전지(perovskite photovoltaics)의 정공수송층(hole transporting layer)에 사용될 소재를 탐구하였다. 정공수송층은 정공(hole)을 전극으로 이동시키는 것을 도와주는 층인데, 이때 정공의 이동도(hole mobility)가 높을수록 태양전지 성능이 향상된다. 본 연구에서는 대표적인 정공수송층 소재인 spiro-OMeTAD 소재를 기반으로 첨가제(dopant)의 비율과 가열시간(annealing time)을 조절하여 정공 이동도를 최적화시키는 것을 목표로 한다. 합성, 분석, 물성평가의 환류체계는 <그림 10A>에 도시되어 있다.

하드웨어 측면에서는 스피코팅 기계와 가열(annealing) 기계들을 도입하여 양질의 박막 형성에 성공하였다는 점이 주목할 만하다. 박막이 완성되면 4점 저항측정장비(four-point resistivity probe)를 사용하여 연구팀이 최적화하고자 하는 물성인 정공 이동도를 측정할 수 있다. 소프트웨어는 상기 소개된 베어지안 최적화 기법이 도입된 ChemOS 운영체제를 사용하였다. <그림 10B>에는 spiro-OMeTAD에 첨가된 첨가제(dopant)의 양과 가열시간 조건이 수정되는 과정을 보여주는데, 초반의 local maximum에 빠지지 않고 약 20-30회 실험을 거쳐서 global maximum으로 최적화에 성공하는 것을 확인할 수 있다. 기존의 high-throughput 방식을 사용하였다면, 최소 수백 번에서 수천 번의 반복실험이 필요하겠지만 자율화실험실에서는 최적화 조건 찾기가 불과 20회 이내에 완료된다. 특히 조절 변수들의 개수가 많으면 많아질수록 자율화실험실의 효율성이 훨씬 더 극대화되는 것은 당연하다.

그림 10. 태양전지용 유기물 정공수송층 소재 합성을 위한 자율화실험실 도식도 및 결과. (A) 합성, 분석, 물성평가, 다음단계 예측의 환류체계. (B) 첨가제의 양과 가열시간의 최적화과정 및 결과



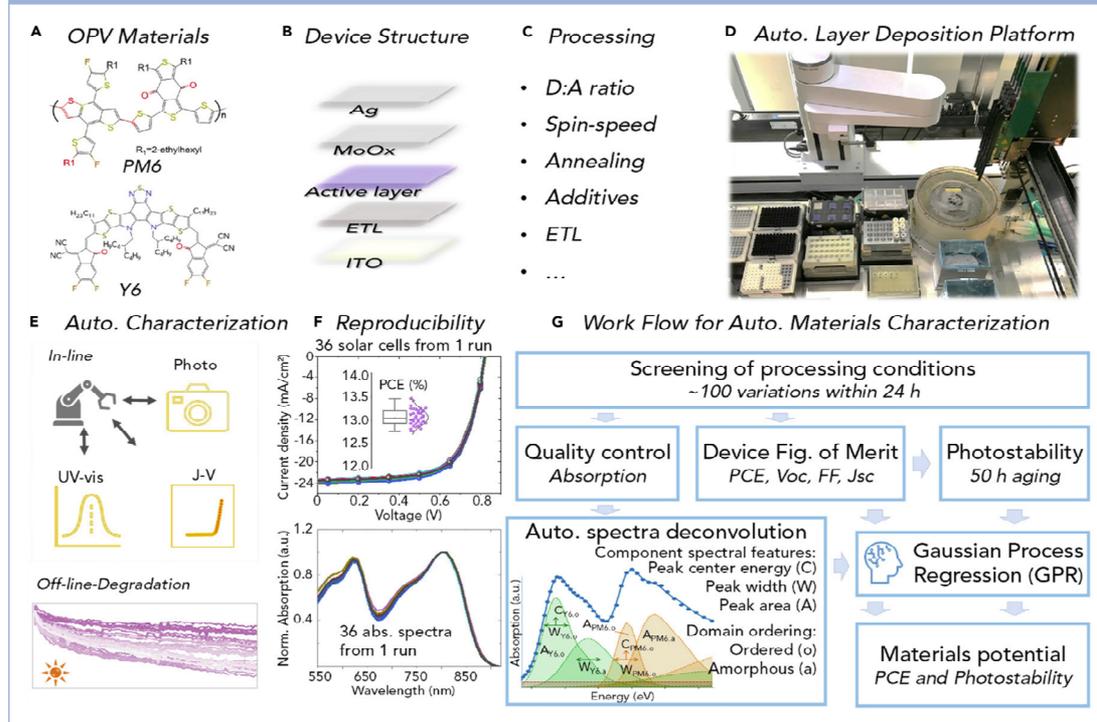
출처 : MacLeod et al.(2020)

3.5. 소자(device) 제작 및 성능평가 사례(독일 프리드리히 알렉산더대 Brabec 교수팀)

독일 프리드리히 알렉산더대 브라벡(Brabec) 교수팀은 태양전지의 소자 제작(device fabrication)으로 까지 확장하여 태양전지 공정 프로세스 전체를 자율화실험실 플랫폼 상에서 완성하였다[5]. 이 연구는 2021년 2월 Joule지에 등재되었다. 이 연구는 소자 제작을 완성했다는 점에서 앞선 연구들보다 훨씬 진보적이다. 앞서 소개되었던 연구들은 소자 내의 특정층에 사용되는 소재만을 최적화하는 연구였다. 반면, 본 연구는 소자 전체를 제작하는데 성공하였는데 이 과정이 얼마나 복잡한지 설명하기 위해 태양전지를 예로 든다. 태양전지는 많은 소재들, 예를 들면 ITO(Indium Tin Oxide), 전자수송층(ETL, Electron Transporting Layer), 광활성층(Active Layer), 정공수송층(HTL, Hole Transporting Layer), Ag(전극) 등을 적층시켜야 한다. 이러한 다양한 소재들을 적층시키기 위해서는 다양한 장비들이 필요하고, 만일 탠덤 태양전지로 제작하면 이보다 두배, 세배 많은 재료들을 적층시킬 필요가 있다. 하드웨어적 측면에서 기존 연구사례들보다 훨씬 더 난이도가 높은 것이다. 이러한 복잡한 과정을 오차나 실수없이 셀 제작까지 완료했다는 점에서 큰 의미가 있는 연구 사례이다.

〈그림 11〉에 정리되어 있듯이 본 연구에서는 유기물 태양전지에서 빛 흡수 및 전자-정공쌍 생성을 담당하는 활성층에 PM6:Y6(donor=PM6, acceptor=Y6)을 사용하였다. PM6:Y6 물질을 바탕으로 bulk heterojunction 구조를 가지는 유기태양전지를 제작하였으며, 조절 변수들은 donor:acceptor 비율, 활성층 스펀코팅 속도, 활성층 가열온도, 전자수송층 가열온도 총 4가지를 사용하였다. 최종목표는 광안정성(photostability, 소자가 빛을 계속 받았을 때 소자성능이 시간이 지남에 따라 안정적으로 유지되는 특성)을 극대화시킬 수 있는 실험조건들을 찾아내는 것이다. 연구팀은 자율화실험실을 활용하여 총 100번의 소자제작을 70시간 이내에 완료했고, 최대 14%의 에너지 전환효율(power conversion efficiency)을 획득할 수 있었다고 보고하였다.

그림 11. 유기물 태양전지의 소자(device) 제작과 소자성능 평가를 위한 자율화실험실 도식도 그리고 워크플로우



출처 : Du et al.(2021)

앞서 소개된 연구들은 태양전지의 특정층(활성층 또는 정공수송층)만을 최적화시키는 연구였다면, 본 연구는 유기물 태양전지 소자 전체를 제작하고 소자의 성능을 최적화시켰다는 점에서 차별된다. 비록 본 연구는 태양전지 소자에 국한되어 있지만, 다른 연구 분야에서도 결국 최종적인 산물은 소자(device)인 경우가 많다. 배터리도 소자이고 발광 다이오드(LED, Light Emitting Diode)도 소자이다. 연구실이나 산업체에서는 소자(device)가 최종산물이 될 것이기 때문에, 앞으로는 소자 제작을 플랫폼 상에서 완성시키는 연구보고들이 이어질 것으로 예견된다.

3.6. 연구실 주행로봇(mobile robot) 도입(영국 리버풀대 Cooper 교수팀)

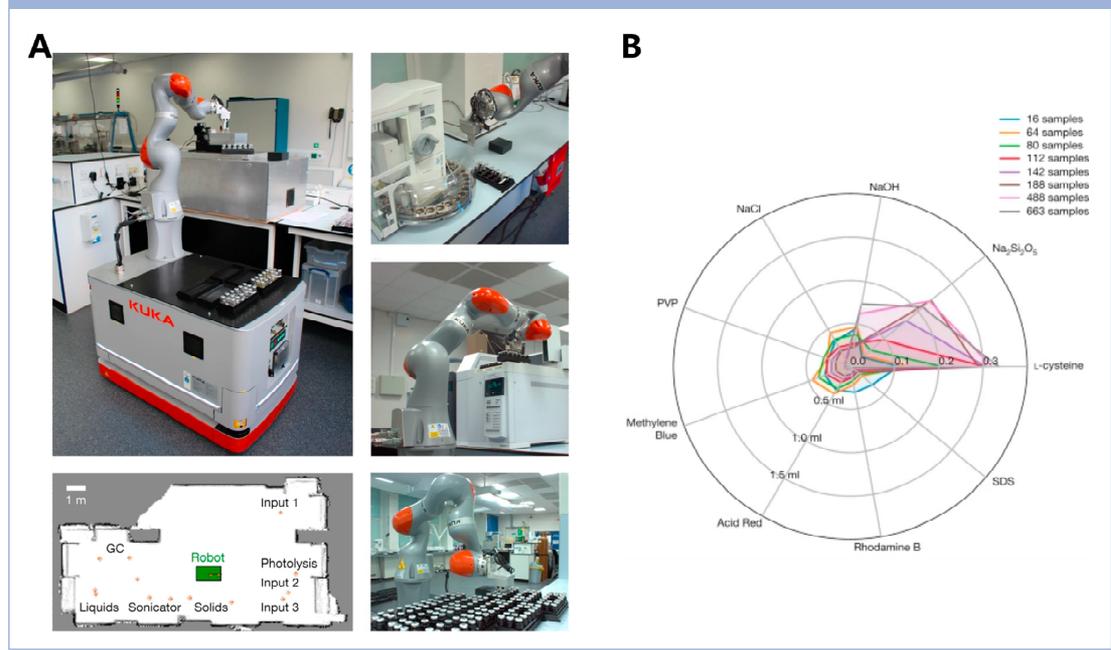
영국 리버풀대 쿠퍼(Cooper) 교수팀은 자율화실험실에 주행로봇(mobile robot)을 도입하였다. 쿠퍼(Cooper) 교수팀은 이 시스템을 mobile robotic chemist라 명명하고 2020년 Nature지에 발표하였다[3].

AI 로봇이 연구실 전체를 휘젓고 다니기 시작한 것으로, 매우 바람직한 방향이며 큰 진보로 판단된다. 앞으로는 협동로봇(robotic arm)만으로는 한계가 있을 것이다. 왜냐하면 소재의 합성, 분석, 물성평가를 위한 장비들은 물리적으로 많이 떨어질 수밖에 없는데, 주행이 불가능한 협동로봇은 물리적으로 접근 가능한 범위가 제한적이기 때문이다. 주행로봇의 도입은 꼭 필요하다. 본 연구에서는 mobile robotic chemist(이동 로봇 화학자, 실험실 조수 로봇을 의미)를 사용하여 수소 생성 반응인 수소발생반응(HER, Hydrogen Evolution Reaction)에 사용되는 P10 광촉매를 최적화시키는 것이 목적이다. P10 광촉매를 대상으로 최고의 효율을 가지는 촉매 반응조건(첨가제 종류/용량, pH 등)을 찾고자 한다. 물론 HER 촉매반응은 그들의 주행 AI 로봇이 얼마나 훌륭한 일을 해낼 수 있는지 보여주는 하나의 예시일 뿐, 다양한 연구분야로 확장되는 것은 시간문제이다.

〈그림 12〉에 연구실 하드웨어 구성이 소개되어 있다. 바이알 저장소, 고체 시약 분사기(solid dispensing machine), 액체 시약 분사기(liquid dispensing machine), 소결 처리기(sonication), 불활성 처리기(inerization machine), 광분해 기계(photolysis), 기체 크로마토그래피(GC, gas chromatography) 등으로 촉매의 합성, 소재 분석, 그리고 촉매특성 평가에 필요한 장비들이 연구실 전체에 널려있다. 이처럼 연구실 전체에 다양한 장비들이 널려있기 때문에 주행로봇을 통해 접근할 필요가 있다. 이때 주행 로봇은 laser scanning system을 탑재하고 있기 때문에, 실험실의 구조가 바뀌어도 합성 공간과 분석 공간을 정확히 인식하고 움직일 수 있다는 장점을 가지고 있다.

광촉매는 촉매 주위의 조건(pH, 빛 흡수 염료, 이온 전하량, 수소 결합세기, 등등)에 따라 촉매특성이 크게 변할 수 있다. 본 연구에서는 이러한 조건들을 조절변수들을 베어지안 기법을 통해서 최적화시켰다. 〈그림12B〉는 실험을 반복함에 따라 변수들이 최적화되어 가는 과정을 보여주는데, 궁극적으로 광촉매 특성에 도움이 되지 않는 시약들 SDS, Rhodamine B, Acid Red, Methylene Blue 등의 농도는 실험을 진행할수록 점차 줄어드는 반면, 성능에 긍정적 영향을 주는 L-cysteine, NaOH 등의 농도는 특정값으로 수렴하는 것을 확인할 수 있다.

그림 12. 영국 리버풀대 움직이는 로봇을 도입한 자율실험실 사진과 HER 광촉매 최적화 결과



출처 : Burger et al.(2020)

본 연구 이전의 모든 자율화실험실들은 정적인 공간 안에서 실험들이 진행되었다. 그래서 공간의 제약이 많이 발생하고 때문에 다양한 분석 방법 도입이 제한되었다. 공간의 제약은 결국 연구자가 실험 중간중간에 개입을 하도록 만들기 때문에 완전한 의미의 자율화실험실에는 도달하지 못하였다. 하지만 본 연구에서 주행로봇을 도입함으로써 연구실 사용의 자유도를 훨씬 제고시켰으며, 하드웨어적 측면에서 이는 큰 진보적 성과이다. 다만, 아직도 연구자(사람)와 로봇이 연구실에서 함께 작업을 하였을 때 생길 수 있는 변수들에 대한 대처능력은 미흡하므로 이런 점들을 개선할 필요가 있다.

3.7. 기타 사례들

융합연구리뷰에서는 꼭 짚어볼 필요가 있는 기술적 요소들을 가지고 있는 연구들만 선별하여 소개하였다. 그러나 앞서 소개된 사례들 이외에도 대규모 국가주도 연구사업이 진행되고 있는 유럽과 북아메리카(미국, 캐나다)를 중심으로 최근 3년간 신소재 개발용 무인실험실에 대한 연구가 많이 이어지고 있다. 아래 <표 1>은 기타 사례들을 추가로 정리하여 보여준다. 대부분의 사례들이 아직은 유기소재의 합성과 평가에 집중되어 있으나, 전 세계적으로 리튬이온전지 전해질, 유무기 태양전지 소재 등 다양한 소재군으로의 확장이 시도되고 있다.

표 1. 신소재 개발용 자율화실험실 추가 사례들

실험로봇명	국가	기관명	소재군
Adam [10]	영국	애버리스트위스 대학 (Aberystwyth University)	유기소재
Continuous-flow chemical synthesis system [2]	미국	MIT	유기소재
Chemobot [19]	영국	글래스고 대학 (University of Glasgow)	유기소재
High-throughput battery evaluation system [16]	일본	국가수리과학연구소 (NIMS, National Institute for Mathematical Sciences)	리튬이온전지 전해질
Robot-Accelerated Perovskite Investigation and Discovery (RAPID) [14]	미국	로런스 버클리 국립연구소 (LBNL, Lawrence Berkeley National Laboratory)	유무기 태양전지 소재
RoboRXN [21]	미국	IBM	유기소재
Scientific Autonomous Reasoning Agent(SARA) [1]	미국	코넬 대학 (Cornell University)	무기소재(산화물)
Materials acceleration operating system in cloud [13]	중국	홍콩 중문 대학 (Chinese Univ. of Hongkong)	유무기 소재
Continuous flow synthesis system of pharmaceutical agents [20]	영국	캠브리지 대학 (University of Cambridge)	유기소재

출처 : 저자 작성

① AI로봇이 스스로 소재를 개발한다고?

② 소재연구를 위한 빅데이터 및 플랫폼 활용

III 소재 자율실험실의 발전방향

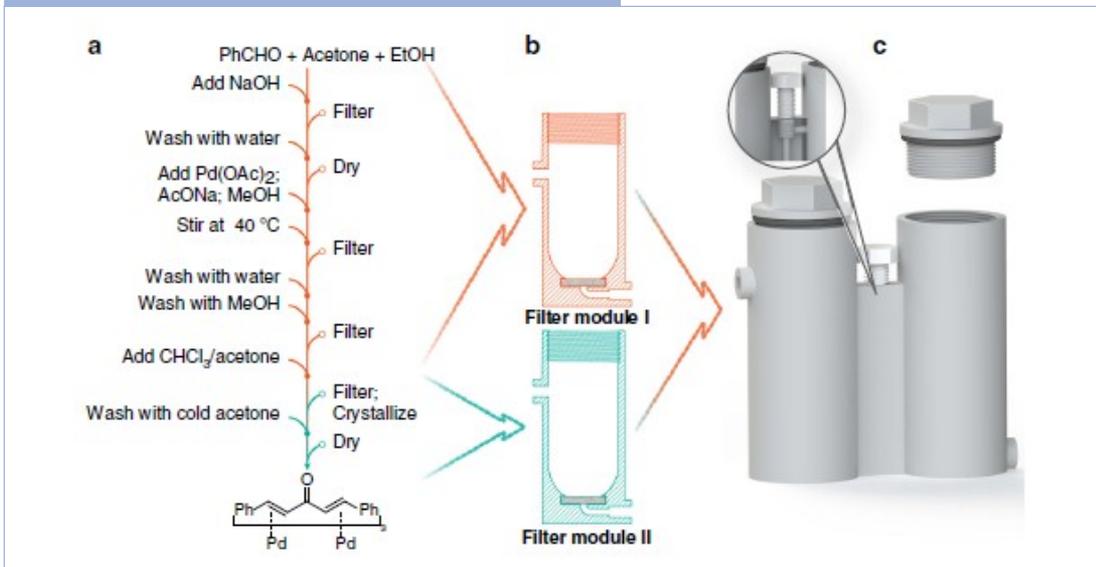
2장에서 살펴본 자율실험실 연구동향을 보면 연구의 대부분이 최근(2~3년 이내)에 발표되었다는 점에 주목할 필요가 있다. 소재분야에 AI 기술은 2015년경부터 본격적으로 도입되기 시작하였으며, 이는 타 분야에 비해 늦게 도입되었다. 그럼에도 불구하고 소재 자율실험실은 이미 우리 곁에 가까이 있고 앞으로 어떤 모습으로 발전될지 기대된다. 이런 점에서 앞으로 소재 분야에서 자율실험실이 어떤 방향으로 연구되고 발전할 지를 AI 기술, 소재데이터, 실험자동화장치 및 로봇 등의 하드웨어적 요소로 나누어서 생각해보는 것도 중요하다.

1. 소재합성·평가 자동화장치 및 기술 개발

2장의 연구동향을 살펴보면 특이점을 발견할 수 있다. 현재까지 개발된 자율실험실은 대부분 유기소재 분야에 집중되어 있다. 이는 무기소재에 비해 상대적으로 합성절차가 간단하고, 고가의 합성장비를 요구하지 않기 때문이다. 예를 들어, 반도체 박막소재를 개발하기 위해서는 고진공 상태에서 금속 등을 증착할 수 있는 장비가 필요하고, 이런 장비는 일반적으로 고가이고 유기소재 합성장비에 비해 규모 면에서도 훨씬 크다. 또한 증착공정 절차도 여러모로 복잡하다. 자율실험실에서는 사람 대신에 로봇이 합성을 해야 하기에 현재는 합성공정이 상대적으로 간단한 분야에 우선 적용되고 있다. 이는 현재 상용화 중인 협동로봇은 인간의 손만큼 자유자재로 움직이거나 정교하지는 않기 때문이다. 따라서, 향후 무기소재 등의 다양한 분야로 확장하기 위해서는 (사람에 의한) 신합성공정 기술 개발이 선행되어야 한다. 이는 합성공정을 단순화하는 기술 개발과 합성장치를 간소화, 자동화하는 기술 개발 등도 포함한다. 또한, 새로운 합성공정을 위한 자동화장치를 개발하기 위해 연구자의 연구환경에 맞게 장치를 제작할 수 있는 3D 프린팅의 역할이 더욱 증대될 것이다.

<그림 13>은 3D 프린터로 제작한 유기합성용 반응기 사례를 보여주고 있다. 반응기 제작을 위해서는 우선 소재합성 절차를 명확히 이해하고, 이 중에서 중요한(key) 요소를 확인하는 절차가 선행되어야 한다(그림 13a). 공정이 크게 2개의 그룹으로 나뉠 수 있어, 각각의 합성공정을 위한 모듈화 장치를 설계한다(그림 13b). 마지막으로 이렇게 2개로 모듈화된 장치를 하나로 연결하도록 설계하고, 이를 3D 프린팅으로 제작하여 연구자 맞춤형 반응기를 제작하였다.

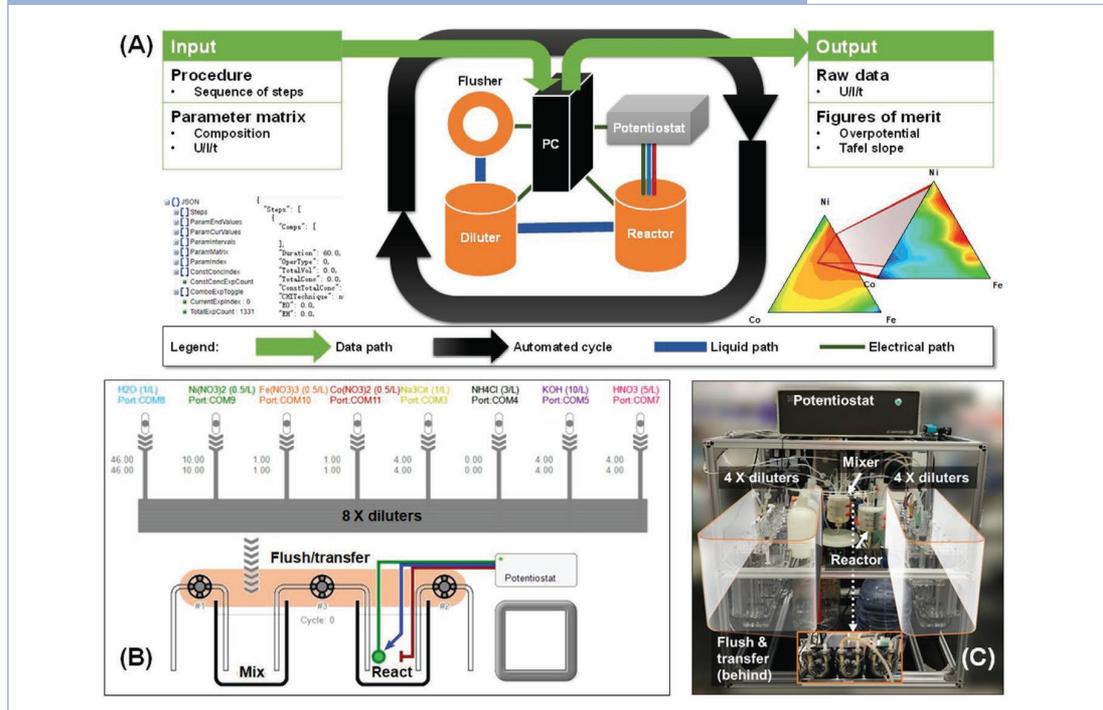
그림 13. 3D 프린터로 제작한 유기합성용 반응기



출처 : Zaleskiy et al.(2019)

합성공정 측면 뿐만 아니라 소재 분석기술도 자율실험실 개발에 매우 중요하다. 로봇이 소재를 합성하더라도 소재물성이 평가되지 않으면 소재개발은 이루어지지 않는다. 그리고 자율실험실의 응용분야의 다양성을 위해서도 물성평가 및 분석기술이 뒷받침되어야 한다. 기본적으로 AI 로봇이 물성평가 및 분석을 하기 위해서는 관련 자동화기술 및 장치 개발이 되어야 한다. 더불어 관련 장치들을 디지털화하여 컴퓨터에 의해 제어될 수 있도록 해야 한다. 이는 소재합성 자동화장치에서도 마찬가지이다. 이러한 이유로 현재의 소재분석 장치에는 많은 제약이 있을 수 있다. 수소(H₂)와 산소(O₂) 가스로부터 과산화수소(H₂O₂)를 합성하는 촉매성능 평가로 예를 들면, 일반적으로 관련 촉매성능을 평가하기 위해서는 반응기에 수용액을 담고 이 수용액에 촉매를 넣은 후 수소와 산소 가스를 불어 넣으면서 일정 시간에 따라 생성되는 과산화수소의 양을 측정한다. 그리고 새로운 촉매의 성능을 평가하기 위해 이전 실험에서 수행한 반응기는 모두 비우고 새로 수용액과 촉매를 넣는다. 이 과정을 로봇이 한다고 생각해보면, 아마 가장 문제가 되는 과정은 새로운 촉매성능 평가를 위해 이전 촉매성능 평가에 사용한 반응기를 비우고 새로 그 반응기에 수용액과 촉매를 넣는 과정일 것이다. 자율실험실에서는 이를 디지털화 및 자동화해야하기에 사람이 실험할 때 사용하는 반응기와는 다른 형태의 반응기를 제작하거나 또 다른 방법을 개발해야한다. 이런 실험자동화의 노력으로 최근에 중국에서 Ni-Co-Fe 삼원계 합금을 합성하고 전기화학적 특성을 자동으로 평가할 수 있는 통합 자동화장치를 개발한 바 있다(그림 14).

그림 14. Ni-Co-Fe 합금제조 및 전기화학특성 평가 자동화장치



출처 : Yu et al.(2021)

2. 소재 합성·공정 DB 구축의 중요성

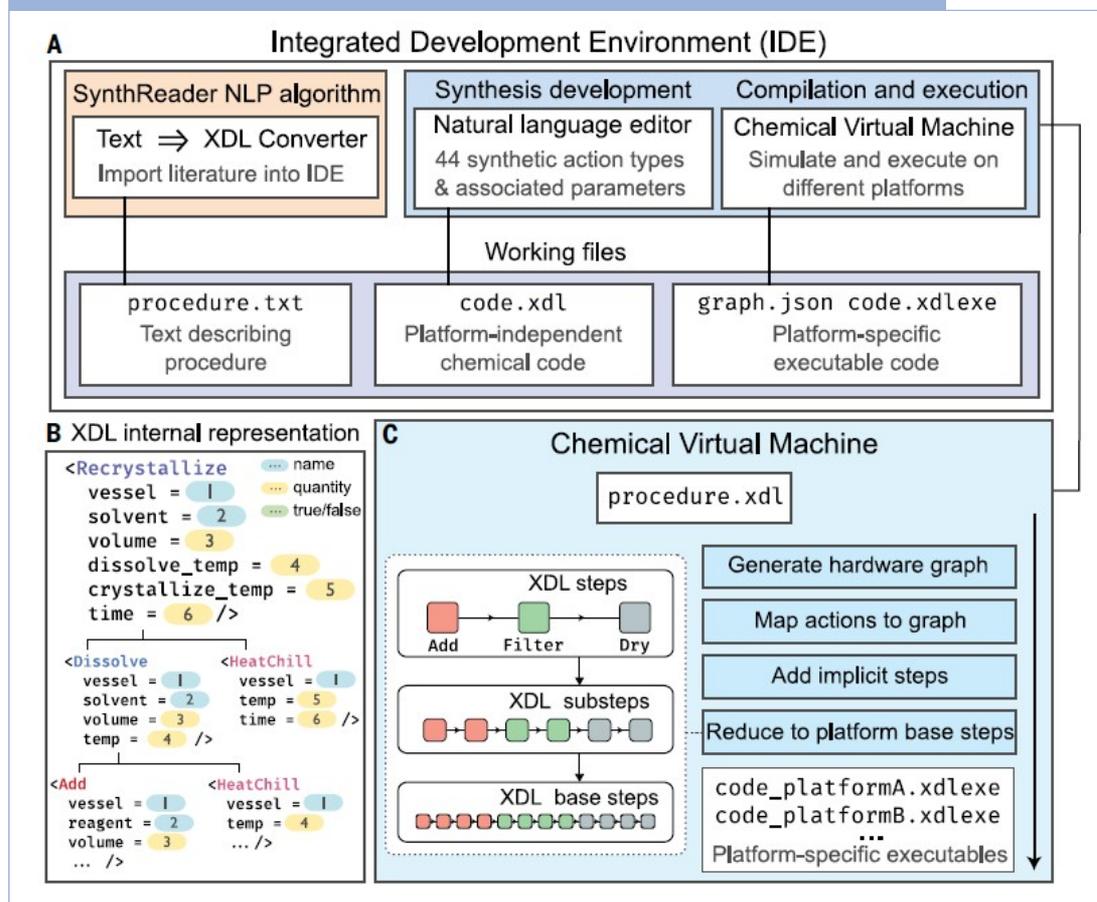
자율실험실을 통한 소재개발 가속화를 위해서는 무엇보다 AI가 최적의 소재 및 공정조건을 정확하고, 빠르게 찾아주어야 한다. 이를 위한 핵심요소가 소재 DB이다. 물론 자율실험실에서는 소재 DB 없이 AI 로봇이 스스로 실험을 수행, 소재 데이터를 생산하며 DB를 축적해갈 수도 있다. 하지만 처음부터 DB가 있는 상태와 비교했을 때, 소재개발 속도가 느릴 수밖에 없다.

현재 세계적으로 많이 알려진 소재 DB는 CSD, ICSD와 같은 결정구조 DB와 Materials Project, NoMAD, OQMD와 같은 계산 DB가 있다. 반면 실험데이터 DB의 경우, 연구실 차원에서 구축한 사례는 있지만 체계적으로 구축하여 공개한 사례는 아직까지는 존재하지 않는다. 계산데이터는 컴퓨터 시뮬레이션으로 얻어지는 경우가 많고, 이 경우 인풋(input)/아웃풋(output) 정보를 체계적으로 수집·관리하는 것이 상대적으로 용이하지만 실험데이터는 그렇지 않기 때문이다. 한편, 실험데이터는 다시 공정데이터, 분석데이터, 특성평가데이터로

크게 나눌 수 있다. 그 중 공정데이터란 소재를 합성하거나 시스템을 구축(예, 전극제작, 디바이스 제작 등) 할 때의 각종 실험조건을 일컬으며 보통 메타데이터로 통칭하고 있다. 공정데이터의 예로는 원료(precursor)의 종류 및 투입량, 열처리 온도 및 시간, 합성 시 압력 등이 있고, 응용분야의 종류에 따라 구체적 예시는 달라질 수 있다. 공정데이터는 계산 및 타 실험데이터와 달리 ‘공정순서(sequence)’ 정보도 매우 중요하다. 예로써, 두 가지(A, B) 원료를 같이 넣고 합성하는 경우와 A 원료를 넣고, 어느 정도 시간이 지난 후 B 원료를 넣고 합성하는 경우에 다른 소재가 합성될 수 있다. 이런 점에서 공정데이터는 분석데이터 및 특성평가데이터에 비해 복잡한 구조를 갖는다. 이런 이유로 인해 공정데이터를 체계적으로 구축하는 일은 쉽지만은 않다. 그러나 분석데이터와 특성평가데이터의 상관관계 파악으로는 실제 실험에서 합성 가능한 최적의 실험조건과 경로를 제시할 수가 없다. 따라서 실험 공정데이터 DB를 구축하는 것은 소재 자율실험실 개발을 위해 반드시 선행되어야 한다. 공정데이터 DB를 체계적으로 구축하기 위해서는 관련 용어(vocabulary)와 데이터 구조(hierarchy) 등을 표준화하는 노력이 필요하고, 이를 위해서는 세계적으로 소재데이터 표준화협의체 구성 및 협력연구가 뒤따라야 한다.

공정데이터 DB를 구축하기 위해서는 자연어처리(Natural Language Process) 기술이 중요하게 활용된다. 여기서 자연어처리 기술은 크게 두 가지 관점에서 활용될 수 있다. 공정데이터 DB를 구축하기 위해 새롭게 실험을 할 수도 있지만 이 경우 많은 시간과 비용을 필요로 한다. 따라서 이전에 발표된 문헌에서 자연어처리 기술을 활용하여 공정데이터를 추출하고 이를 DB화하는 것이 더욱 효율적이다. 하지만 현재의 자연어처리 기술로는 이러한 공정데이터를 추출하는데 정확성이 높지 않아, 관련 연구가 활발하게 진행될 필요가 있다. 또한, 실험 자동화장치에 의해 소재를 합성하고 이 정보를 DB화하더라도, 그 DB를 바로 AI에 적용하기 힘들 수 있다. 자동화장치에서 생성된 데이터를 DB화할 때 AI가 바로 활용할 수 있도록 디지털화할 필요가 있고, 이때 자연어처리 기술이 중요하게 활용될 수 있다. 따라서 자율실험실 구축을 위해서는 자연어처리 기술 또한 핵심기술이 될 수 있다. <그림 15>는 영국 글래스고(Glasgow) 대학의 크로닌(Cronin) 교수팀이 최근에 자연어처리 기술을 이용하여 공정데이터를 XML(Extensible Markup Language) 형태로 DB를 구축한 사례를 보여준다. 소재 자율실험실 구축 시 이와 유사한 연구가 반드시 진행되어야 한다.

그림 15. 영국 글래스고 대학에서 구축한 XML 형태의 공정데이터 DB 구조

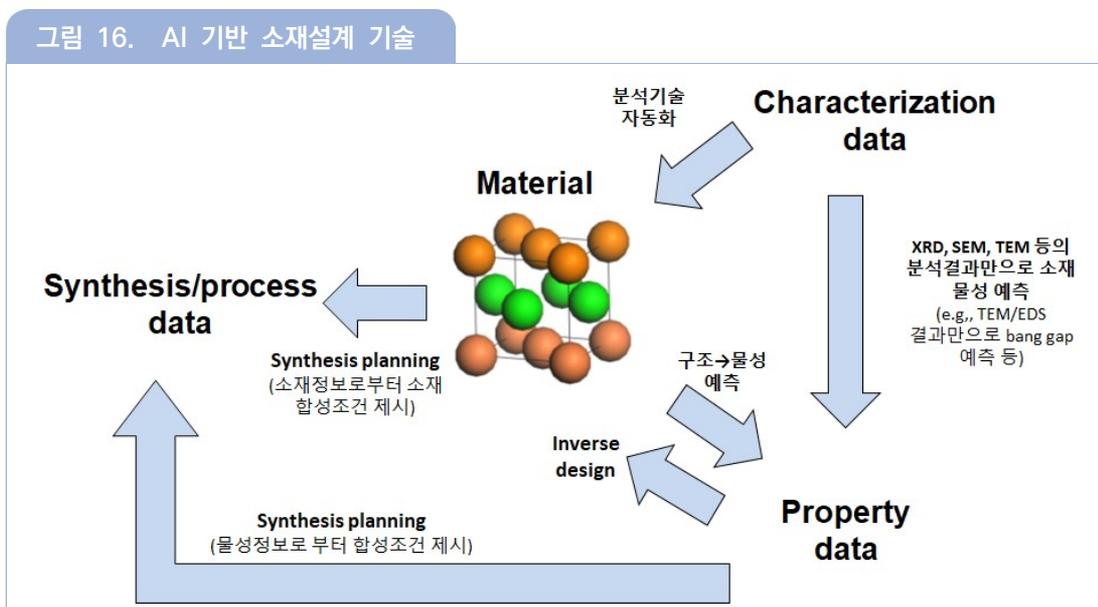


출처 : Mehr et al.(2020)

3. 공정최적화 및 맞춤형 소재설계 AI 기술

서두에서 언급한 것처럼 현재 AI 기술은 소재개발에 활발히 활용되고 있다. 특히 현재의 계산과학기술 및 소재분석 기술의 한계점을 극복한 기술개발 분야에서 활발히 연구되고 있다. 예로써, 기존 제일원리계산의 느린 계산 속도를 개선시키는 AI 기술 및 전자현미경에서 얻은 회절패턴 정보로부터 결정구조를 자동적으로 분류해주는 AI 기술 등이 있다. <그림 16>은 한국과학기술연구원(KIST) 계산과학연구센터에서 계획 중인 소재개발을 위해 필요한 AI 기술을 나타낸다. <그림 16>에서 Material ⇌ Property 예측은 기존 계산과학의 한계극복형 AI 기술을 나타내고, Characterization → Material 예측은 소재분석기술의 한계극복형 AI를

의미한다. 일반적으로(실험적으로) 소재를 연구할 때 합성공정을 통해 소재가 합성되고, 합성된 소재의 구조 및 조성 등을 파악하기 위해 분석장비(XRD, SEM, TEM 등)를 통해 소재분석(characterization) 작업이 이루어지며, 궁극적으로 합성소재가 원하는 물성(기계적, 화학적, 물리적 등)을 갖는지 평가하는 작업이 수행된다. 그러나 AI 기반 소재설계 기술에는 위에서 설명한 소재연구 프로세스의 반대방향으로 소재정보를 예측하는 기술 개발이 필요하다. 소재 자율실험실에서는 '구조-물성' 간 상관관계를 뛰어넘어 '구조-물성-공정' 간의 상관관계를 이해할 수 있는 AI 기술 개발이 필요하다. 특히, 연구자가 원하는 물성 및 소재정보를 입력값으로 하여 그 소재의 합성/공정조건을 예측할 수 있는 AI 기술(synthesis planning 기술) 개발이 반드시 필요하다.



출처 : KIST 계산과학연구센터

일반적으로 신소재를 개발한다고 하면 어떤 한 성능(물성)을 극대화할 수 있는 소재를 개발하는 것으로 이해한다. 계산과학 혹은 AI를 이용한 소재 개발연구에서는 이러한 연구방식이 활발하게 진행되고 있다. 하지만 산업적 관점에서는 어느 한 성능만 극대화하기 보다는, 비록 최고성능은 아닐지라도 여러 물성의 기준을 충족하는 소재를 개발하는 것이 더 중요할 수 있다. 촉매 개발을 예로 들면 촉매성능은 크게 활성도, 선택성, 안정성으로 나눌 수 있다. 통상적으로 새로운 촉매를 개발한다고 하면 활성도 혹은 선택성이 우수한 촉매를 개발하고자 한다. 하지만 산업적으로는 활성과 선택성이 어느 정도 이상의 성능을 가지면서 안정성 또한

보장되는 소재를 더욱 원하게 된다. 산업적 관점에서는 수요자가 요구하는 여러 물성을 충족하도록 하는 맞춤형 소재개발이 필요하다. 4차 산업혁명 시대를 맞이하여 개인맞춤 시장(개별고객의 주문에 맞춰 생산이 가능한 시장)의 부상으로 제품 다양화와 그에 따른 맞춤형 소재의 신속한 개발에 대한 수요가 더욱 커질 것으로 전망된다. 기존의 최대성능을 갖는 소재를 설계하는 방식에서는 타겟(target) 물성이 하나이지만, 맞춤형 소재설계 방식에서는 타겟 물성이 여러 개이다. 따라서 소재설계 방식에 기존과 다른 기술이 필요하고, 여기에 AI 기술이 중요하게 활용될 수 있다.

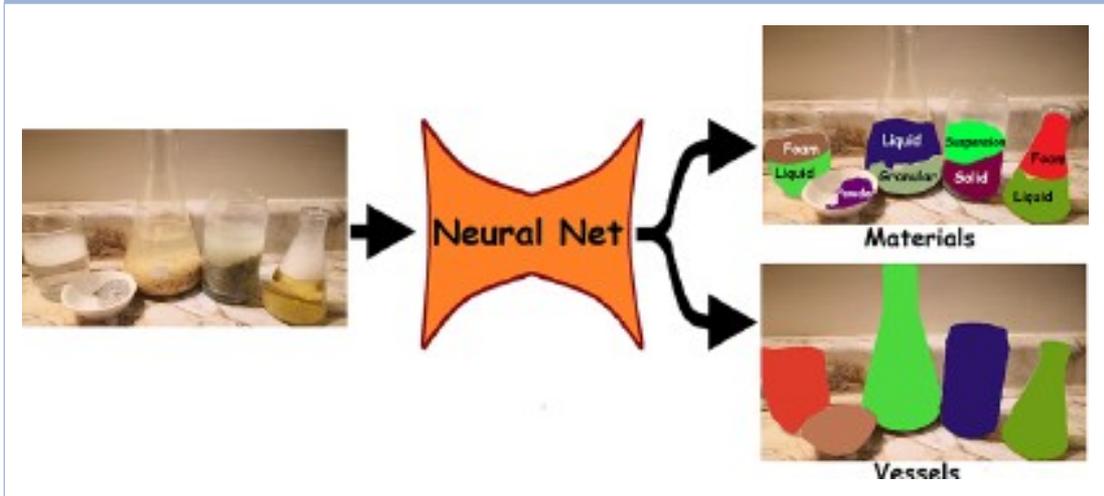
4. 완전무인화를 위한 AI 로봇 기반기술 개발

소재 자율실험실에서는 사람 대신에 로봇이 실험을 수행한다. 따라서 사람이 실험할 때는 손쉽게 할 수 있는 일도 현재의 로봇에게는 힘든 일이 될 수 있다. 예를 들면, 시약병을 사람이 잡을 때는 큰 어려움이 없지만, 로봇에게는 시약병의 무게 및 크기/모양에 따라 제약이 있을 수 있다. 대부분의 시약병은 유리로 되어 있어서 로봇이 조금만 힘을 가하면 깨질 가능성도 존재한다. 따라서 로봇이 유리 등으로 된 작고 약한 실험용품을 안전하게 파지하기 위한 로봇손 하드웨어 및 제어기술 개발 연구가 반드시 필요하다. 실험용품의 물체자세 및 깊이값을 입력으로 하는 실험실 전용 로봇손 파지 모션 제어 학습 알고리즘 개발 및 학습을 위한 시뮬레이션 환경 개발 등도 같이 필요하다.

장기적으로는 AI 로봇에 자율주행 기능을 탑재하여 사람이 실험하는 것처럼 실험실을 자유롭게 이동하며 실험할 수 있도록 해야 한다. 여기에 AI가 예측한 소재합성 조건에 따라 로봇이 실험에 앞서 스스로 합성에 필요한 시약을 찾는 행위도 포함된다. 이를 위한 다양한 기술이 필요하며, 대표적으로 실험실 환경 내 연구용품을 인식할 수 있는 기술과 로봇의 자율주행을 위한 실험실 환경 인식 기술 등이 개발될 필요가 있다. 실험실 자율주행 로봇의 안전한 주행을 위해서는 실험실의 3차원 공간 정보를 취득하는 것이 필수적이다. 기존의 깊이 센서 및 깊이 지각 알고리즘을 활용할 경우, 무반사 불투명 표면에서만 정확한 깊이값 획득이 가능하나 소재 실험실 환경의 많은 배경과 도구들이 금속 혹은 유리 재질로 되어 있어 이에 대응하기 위한 기술개발이 반드시 필요하다. 또한 로봇이 실험도구를 가지고 이동하여 원하는 곳에 정밀하게 위치시키는 기술도 매우 중요하다. 특히 소재합성의 경우 크기가 작은 바이얼(vial) 등의 실험도구를 많이 사용하기 때문에 정밀하게 위치시키지 못하면 다른 형태의 안전사고도 생길 수 있다. 이 밖에도 화학실험 등에 쓰이는 실험도구 및 시약병은 크기, 모양, 색깔 및 재질이 유사한 경우가 많아 AI 기반의 컴퓨터비전(Computer vision) 및 센서를

활용한 기술개발 등도 필요하다. 최근에 캐나다 토론토 대학에서는 시약병의 종류 및 시약병 내 함유물의 상(고체, 액체, 분말, 거품 등)을 구별하는 AI 컴퓨터비전 기술을 개발한 바 있다(그림 17). 아직은 시약병 내의 함유물이 구체적으로 어떤 화학물질인지까지 판별하는 기술이 개발되지는 않았지만, 이를 위해 시약병의 인식라벨 등을 통해 시약병 내의 화학물질을 구별하는 기술이 개발될 필요가 있다.

그림 17. 캐나다 토론토 대학에서 개발한 시약병 종류 및 함유물질 식별을 위한 AI 기반 컴퓨터비전 기술



출처 : Eppel et al.(2020)

IV 결론

최근 국내 저출산 및 고령화에 따른 노동인구 감소로 국가경쟁력이 크게 위협받고 있다. 세계무역기구(WTO)의 보고에 따르면 국내의 노동인구 감소는 국내총생산에 영향을 끼쳐 우리나라의 잠재적인 성장률은 2040년까지 65%로 세계 평균(80%)에도 못 미칠 것으로 전망된다. 이런 상황에서 소재 자율실험실은 우리나라의 산업경쟁력에 이바지 할 수 있을 것으로 기대한다. 향후 AI 로봇의 투입으로 국내 제조업 분야의 노동력 감소 문제를 일부 해결할 수도 있을 것으로 낙관한다. AI 로봇은 저출산 고령화 현상을 극복할 수 있는 기술적 열쇠가 될 수 있을 뿐만 아니라 지속적인 경제성장을 위한 국가의 유망산업이기도 하다.

코로나-19로 시공간 제약에서 자유로운 유비쿼터스 연구기반 조성의 필요성이 대두되고 있다. 즉, 재난·재해 등의 국가 위기 시에도 중단 없는 연구수행이 보장될 필요가 있다. 소재 분야의 비대면 연구를 위해서는 디지털 전환뿐만 아니라 사람 없이도 실험연구를 수행할 수 있는 환경이 필요하며, 이를 위해 AI 로봇을 통한 자율실험실 개발이 필요하다. 또한, 소재 개발 시에는 위험한 물질을 다루는 경우가 많아 실험실 안전사고가 종종 발생하고 있어 미래에는 실험실 무인화가 활발히 진행될 것으로 기대된다. 이런 점에서 앞으로 AI 로봇을 통한 자율실험실이 소재분야 뿐만 아니라 타분야에도 대중화될 것이며, 특히 위험한 물질을 다루는 실험에 우선적으로 적용될 가능성이 높다.

또한, 4차 산업혁명 시대에서 국가의 소재경쟁력은 데이터가 좌우한다 해도 과언이 아니다. AI 기술을 통해 신소재를 빠르게 설계 및 개발하기 위해서는 양질의 소재데이터가 뒷받침되어야 한다. 현재 세계적으로 소재 데이터를 체계적으로 수집·관리하려는 노력이 진행되고 있다. 여기서 우리가 생각해 볼 이슈가 있다. 소재 DB 구축 시 각 연구그룹에서 창출되는 데이터는 모두 양질이라 하더라도 연구그룹마다 실험조건 및 장비 등이 달라서, 같은 소재에 대해서도 데이터를 모아서 보면 데이터의 양은 늘었지만 DB 관점에서는 양질이라 부르기 힘들 수 있다. 따라서 양질의 데이터를 체계적으로 생성하기 위해서는 동일장비에서 데이터를 대량·신속(high-throughput)하게 생산할 수 있는 플랫폼이 필요하다. 이런 점에서도 소재 자율실험실은 중요한 역할을 할 수 있다. 자율실험실의 로봇을 통한 실험자동화로 소재공정 및 물성데이터를 대량·신속하게 생산할 수 있기 때문이다. 또한 자율실험실이 실현되면 낮에는 사람이 실험을 하고, 밤에는 로봇이 실험을 이어서 수행할 수 있어 소재개발이 더욱 가속화될 수 있다.

위에서도 언급한 바와 같이 소재 자율실험실은 사람이 아닌 AI 로봇이 실험을 하기 때문에 사람이 다칠 염려가 없어 안전할 수 있지만, 한편으로는 사람이 없기 때문에 인명피해 외에 또 다른 안전사고가 발생할 수 있다. 이를 방지하기 위한 기술적인 노력으로 센서 등의 개발연구도 필요하지만, 안전한 자율실험실을 위한 정책 수립에도 신경을 쓸 필요가 있다. 이는 자율주행자동차가 완전자율화 되기까지 기술적으로 단계를 수립하고 이에 따른 안전정책을 마련하는 것과 유사하다.

아마 많은 연구자들이 AI 로봇을 통한 소재개발 연구가 활성화되면 인간의 일자리를 빼앗지 않을까하고 염려할 수 있다. 저자는 이에 대해 두 가지 답변을 하고 싶다. 첫 번째는 소재 자율실험실은 4차 산업혁명 시대에서 거스를 수 없는 현실이라는 점이다. 즉, 우리가 하지 않으면 다른 국가에서 개발할 것이 분명하다. 이미 미국, 유럽 등에서는 이에 대해 국가적으로 많은 투자가 이루어지고 있다. 따라서, 거스를 수 없는 현실이라면 우리나라가 이 분야를 주도하는 것이 낫지 않을까라고 조심스럽게 얘기하고 싶다. 다행스럽게도 현재 이 분야는 세계적으로도 아직 태동기로 우리나라에서 지금 시작해도 늦지 않았다고 판단된다. 두 번째로 너무 걱정하지 않아도 된다는 답을 하고 싶다. 기본적으로 AI 기술은 학습된 데이터 환경에서 최적화된 결과를 얻을 수 있다. 즉, AI 로봇 스스로는 DB로 학습되지 않은 창의적인 일을 할 수가 없다. 나노촉매 개발용으로 학습된 AI 로봇이 새로운 배터리 소재를 개발할 수는 없는 것과 같다. AI 로봇이 스스로 소재를 개발할 수 있어도 사람처럼 창의적일 수는 없다.

저자_ **한상수**(Sang Soo Han)

• 학력

한국과학기술원 신소재공학 박사
 한양대학교 재료공학 석사
 한양대학교 재료금속공학 학사

• 경력

現) 한국과학기술연구원 계산과학연구센터 센터장
 現) 한국과학기술연구원 책임연구원
 前) 한국표준과학연구원 선임연구원

저자_ **김동훈**(Donghun Kim)

• 학력

MIT Materials Science and Engineering 박사
 포항공과대학교 신소재공학 학사

• 경력

現) 한국과학기술연구원 계산과학연구센터 선임연구원

참고문헌

〈국외문헌 : 알파벳순〉

- 1) Ament S., Amsler M., Sutherland D. R., Chang M., Guevarra D., Connolly A. B., Gregoire J. M., Thompson M. O., Gomes C. P., Bruce van Dover R., (2021). Autonomous synthesis of metastable materials. *Materials Science*, arXiv:2101.07385
- 2) Bédard A., Adamo A., Aroh K. C., Russell M. G., Bedermann A. A., Torosian J., Yue B., Jensen K. F., Jamison T. F. (2018). Reconfigurable system for automated optimization of diverse chemical reactions. *Science*, Vol. 361, Issue. 6408, pp. 1220
- 3) Burguer B., Maffettone, P. M., Gusev, V. V., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., Rankin, N., Harris, B., Sprick, R. S., Cooper, A. I. (2020) *Nature* 583, 237–241.
- 4) Coley, C. W., Thomas III, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., Hicklin, R. W., Plehiers, P. P., Byington, J., Piotti, J. S., Green, W. H., Hart, A. J., Jamison, T. F., Jensen, K. F. (2019) *Science* 365, eaax1566.
- 5) Du, X., Luer, L., Heumueller, T., Wagner, J., Berger, C., Osterrieder, T., Wortmann, J., Langner, S., Vongsaysy, U., Bertrand, M., Li, N., Stubhan, T., Hauch, J., Brabec, C. J. (2021) *Joule* 5, 1–12.
- 6) Eppel, S. Xu, H., Bismuth, M., & Aspuru-Guzik, A. (2020). Computer Vision for Recognition of Materials and Vessels in Chemistry Lab Settings and the Vector-LabPics Data Set, *ACS Central Science*, 6, 1743–1752.
- 7) Fujimoto, K., Onoda, K., Sato, M., Matsuo, H., Yamaguchi, T., Ito, S. (2008). High-throughput synthesis and evaluation of thermochromic materials by a combinatorial approach, *Materials Science and Engineering A* 475, 52–56.
- 8) Granda, J. M., Donina, L., Dragone, V., Long, D.-L., Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity, *Nature* 559, 377–381.
- 9) Häse, F., Roch, L. M., & Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories, *Trends in Chemistry*, 1, 282–291.
- 10) King, R. D., Rowland J., Oliver S. G., Young M., Aubrey W., Byrne E., Liakata M., Markham M., Pir P., Soldatova L. N., Sparkes A., Whelan K. E., Clare A. (2009). The Automation of Science. *Science*, Vol. 324, Issue 5923, pp. 85
- 11) Langer S., Häse, F., Perea, J. D., Stubhan, T., Hauch, J., Roch, L. M., Heumueller, T., Aspuru-Guzik, A., Brabec, C. J. (2020) *Advanced Materials* 32, 1907801.
- 12) Li, J., Ballmer, S. G., Gillis, E. P., Fujii, S., Schmidt, M. J., Palazzolo, A. M. E., Lehmann, J. W., Morehouse, G. F., Burke, M. D. (2015). Synthesis of many different types of organic small molecules using one automated process, *Science* 347, 1221–1226.

- 13)** Li J., Li J., Liu R., Tu Y., Li Y., Cheng J., He T., Zhu X., (2020). Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nature Communications*, 11, Article number: 2046
- 14)** Li. Z., Najeeb M. A., Alves L., Sherman A. Z., Shekar V., Parrilla P. C., Pendleton I. M., Wang W., Nega P W., Zeller M., Schrier J., Norquist A. J., Chan E. M. (2020). Robot-Accelerated Perovskite Investigation and Discovery. *Chemistry of Materials*, Vol. 32, Issue. 13, pp. 5650
- 15)** MacLeod, B. P., Parlane, F. G. L., Morrissey, T. D., Häse, F., Roch, L. M., Dettelbach, K. E., Moreira, R., Yunker, L. P. E., Rooney, M. B., Deeth, J. R., Lai, V., Ng, G. J., Situ, H., Zhang, R. H., Elliott, M. S., Haley, T. H., Dvorak, D. J., Aspuru-Guzik, A., Hein, J. E., Berlinguette, C. P. (2020) *Science Advances* eaaz8867.
- 16)** Matsuda S., Nishioka K., & Nakanishi S. (2019). High-throughput combinatorial screening of multi-component electrolyte additives to improve the performance of Li metal secondary batteries. *Scientific Reports*, Vol. 9, Article number: 6211.
- 17)** Mehr, S. H. M., Craven, M., Leonov, A. I., Keenan, G., & Cronin, L. (2020). A universal system for digitization and automatic execution of the chemical synthesis literature. *Science*, 370, 101-108.
- 18)** Roch, L. M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., Yunker, L. P. E., Hein, J. E., Aspuru-Guzik, A. (2020). ChemOS: An orchestration software to democratize autonomous discovery, *PLOS ONE* 15: e0229862
- 19)** Steiner S., Wolf J., Glatzel S., Andreou A., Granda J. M., Keenan G., Hinkley T., Aragon-Camarasa G., Kitson P. J., Angelone D., Cronin L. (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. *Science*, Vol. 363, Issue 6423, eaav2211
- 20)** Ley S., (2018). Chemistry Freed from Space and Time – Automated optimization and synthesis of pharmaceuticals in the cloud. *Angewandte Chemie International Edition*, 57, 15128-15132
- 21)** Vaucher A. C., Zipoli F., Gelyukens J., Nair V. H., Schwaller P. & Laino T. (2020). Automated extraction of chemical synthesis actions from experimental procedures. *Nature Communications*, Vol. 11, Article number: 3601
- 22)** Yu, C., Xiong, Q., Yang, K., Chen, H., & Pan, F. (2021). A Programmable and Automated Platform for Integrated Synthesis and Evaluation of Water Electrolysis Catalysts, *Advanced Materials Technologies*, 6, 2001036(1)-(7).
- 23)** Zaleskiy, S. S., Kitson, P. J., Frei, P., Bubliauskas, A., & Cronin, L. (2019). 3D designed and printed chemical generators for on demand reagent synthesis, *Nature Communications*, 10, 5496 (1)-(8).

<기타문헌 : 마지막순서(홈페이지주소 등)>

- 24)** Caterpillar 홈페이지. https://www.cat.com/en_ZA/articles/ci-articles/automation-autonomy-what-s-the-difference



융합연구리뷰

Convergence Research Review 2021 May vol.7 no.5



02

소재연구를 위한 빅데이터 및 플랫폼 활용

안재평(한국과학기술연구원 책임연구원)

김홍규(한국과학기술연구원 선임연구원)

*도움주신 분 : 이병현(한국과학기술연구원, 데이터 해석 전문가)

I 인공지능 시대, 소재 데이터란?

1. 소재 데이터 분야 R&D 환경 변화

신소재는 환경·에너지, 정보통신 및 의료 등 모든 산업의 원천으로 과거 수십 년간 각국은 소재 연구개발에 막대한 자원을 투입해 왔으나, 신물질 발견에서 실용화 과정을 거쳐 시장 진입까지 여전히 10~30년에 달하는 긴 기간이 소요된다. 최근에는 소재 특성 측정·분석 기술의 발달과 계산과학을 기반으로 한 신소재 탐색·설계 연구의 활성화에 따라 실험과 계산에 의한 소재 연구데이터가 대폭 증가하여, 디지털화된 소재 연구데이터를 활용한 소재 R&D 환경을 조성하려는 움직임이 있다. 이런 변화는 2011년 미국의 소재게놈이니셔티브(MGI, Materials Genome Initiative) 및 2012년 딥러닝(Deep Learning)을 계기로 일어난 제3차 인공지능 붐에 의해 시작되었으며, 이에 따라 실험, 이론, 계산과학에 이어 제4의 과학으로 불리는 데이터 과학을 활용한 소재 연구개발이 전 세계적으로 확산되는 추세이다.

소재 데이터를 이용하여 재료과학의 제반 문제를 해결하는 소재정보학(Materials Informatics)의 활용으로 소재 연구개발의 효율화를 추구하는 데이터 구동형 소재 개발 방법론이 소재 R&D 패러다임으로 새롭게 대두되고 있다. 소재 연구개발에 활용되는 데이터 과학은 복잡하고 방대한 데이터로부터 특징적인 패턴을 발견하고 유용한 지식을 추출하는 기술로 데이터 마이닝, 기계학습(머신러닝)이 대표적인 방법이다. 소재정보학의 도입으로 대량의 복잡한 데이터로부터 규칙을 추출하는 귀납적(데이터 구동형) 방법론과 이론·모델에 기초한 연역적(원리 구동형) 방법론이 시너지 효과를 창출하여 향후 소재 연구에 큰 변화를 이룰 것으로 전망된다. 데이터 기반 소재 연구개발에는 다양한 분야의 전문가 및 연구그룹의 참여가 필수적이며 소재정보학을 위한 오픈플랫폼을 구축하고 대학 및 산업체의 모든 연구자들이 데이터를 쉽게 활용할 수 있는 총체적인 데이터 생태계의 조성이 필요하다. 데이터 기반 소재정보학을 활용하는 신소재 연구개발은 아래와 같은 주제들을 다룬다.

- ① 다양한 종류·형태의 대량 데이터 분석에 의한 구조-물성 관계 및 기구 해명
- ② 대량의 데이터로부터 구조 또는 물성 예측

- ③ 최적화 방법에 의한 물질(구조) 탐색
- ④ 멀티스케일 모델링
- ⑤ 데이터를 이용한 고차원 수리모델
- ⑥ 물질공간 또는 해석 데이터 가시화

1.1. 연구 환경 및 패러다임의 변화

과학기술 역사를 살펴보면, 근대과학은 1600년대 갈릴레오, 뉴턴 등 과학자들로부터 시작되어 1900년대 초반까지 이어졌다. 이 시기에 과학적 탐색방법은 반복적인 실험을 통해 이론을 정립하는 것이었다. 이후 컴퓨터의 등장과 이론이 접목되며 실험할 수 없거나 증명할 수 없는 문제들에 대한 예측과 해석이 계산과학을 통해 가능해졌다. 이런 과정에서 인간이 감당할 수 없을 정도로 대량으로 발생하는 빅데이터를 이용한 해석과 예측의 시대가 열리며 연구의 패러다임이 급격하게 변하고 있다. 최근에는 정보기술(IT) 발달 및 연구 장비의 현대화로 방대한 양의 연구자료 생산·수집·활용을 통한 연구개발의 패러다임 변화가 진행 중이고 세계적으로 실험·이론·계산과학 중심의 연구 패러다임이 빅데이터 중심의 데이터 기반 연구 패러다임으로 변화하고 있다.



출처 : 한국과학기술연구원(2018)

빅데이터가 활용되기 전의 기존 재료과학의 연구 패러다임으로는 새로운 소재를 찾아 상업화하기까지 10~20년이라는 긴 기간이 소요되나, 재료정보학이 발달하면 반복되는 실험과 계산 단계를 거쳐 결과를 얻어내는 과정이 생략됨으로써 신소재 개발에서 활용까지의 소요기간이 크게 단축될 수 있을 것으로 전망되고 있다. 소재 연구 분야에서는 1980년대부터 재료과학과 정보과학(데이터 마이닝과 머신러닝)의 융합에 대한 필요성이 대두되었고 이는 소재정보학의 탄생 및 소재 개발의 새로운 시대를 여는데 밑거름이 되었다.

빅데이터 연구는 거대 자료를 활용하는 연구로써, 연구의 중심이 기존의 논리성을 기반으로 한 연역적인 과학에서 '합리성에 기반한 귀납적'인 과학으로 이동하는데 가장 중요한 역할을 하고 있다(이영의, 2015). 최근 빅데이터 중심의 연구 성공사례로는 생물정보학(bioinformatics)에서 생성된 빅데이터로부터 생물학적 이해를 도출하고 새로운 형태의 의미를 구성해 인간 유전체인 게놈(genome)에 담긴 30억 개의 염기서열을 해독한 사례가 있다. 빅데이터는 바이오 및 헬스케어 분야의 유전체 정보로부터 질병을 진단하여 치료하고 신약개발에도 활발하게 이용되고 있을 뿐만 아니라 마케팅, 의료, 교통 분야 등에서는 이미 적극 활용되고 있다.

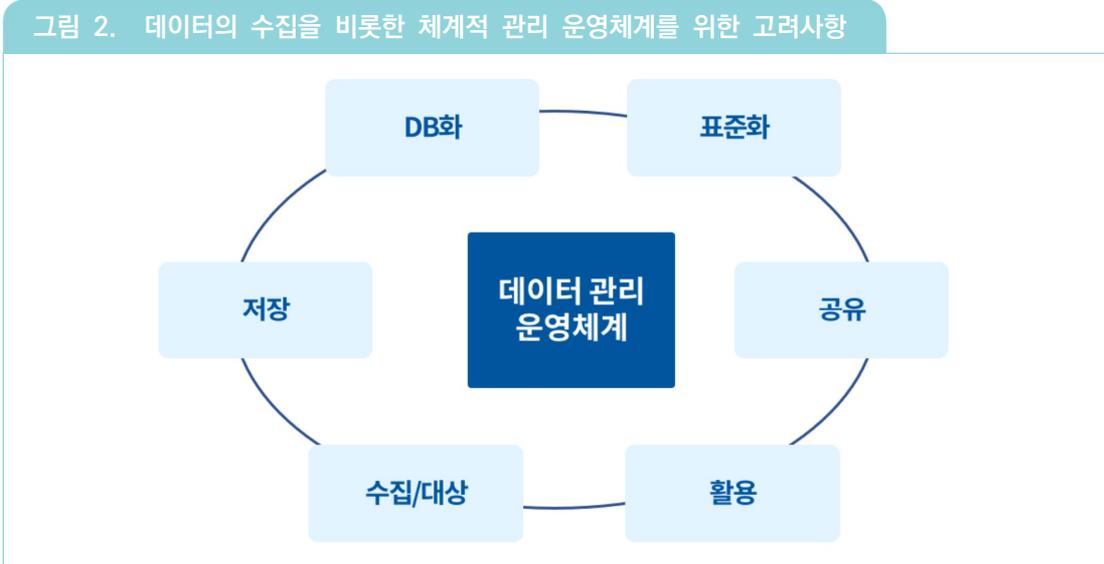
1.2. 소재 연구데이터 수집·활용의 필요성

4차 산업혁명 시대의 도래에 따라 빅데이터 및 이를 활용하는 인공지능(머신러닝 등) 기반의 연구 환경 변화가 다양한 분야에서 이루어지고 있다. 이러한 환경에서는 활용 목적별로 충분한 양의 데이터가 확보되지 않으면 데이터 활용 자체가 불가능하거나, 머신러닝 등을 통해 얻어진 모델의 신뢰성에 악영향을 끼치기 때문에 충분한 양의 연구데이터 확보가 필수적이다. 현재 인공지능을 구현하기 위한 기술, 수학적 방법론, 컴퓨터 언어 등은 이미 높은 수준에 이르렀으나, 이를 작동시키기 위한 소재 관련 연구데이터가 부족한 상황이다. 주로 제일원리에 기초한 계산데이터의 경우 컴퓨터 자동화 생성이 가능하고 변수 조절이 용이하여 대량으로 데이터를 축적하는 것이 가능하다. 현재 구축되어 있는 데이터베이스의 물성정보는 대부분 계산데이터를 사용하는데, 실험데이터의 데이터 양이 상대적으로 매우 부족하다. 이러한 이유로, 대부분의 대규모 데이터베이스(Database, 이하 DB)는 계산방법론이 산출해주는 물성 데이터만 확보하고 있어 일부 물성의 경우 실험값과 동떨어진 결과들이 축적될 때가 많아 데이터의 질(quality)이 훼손되는 사례도 빈번히 발생한다. 따라서 해외 연구기관들에서는 소재 관련 실험 및 물성 분석 데이터를 중앙 집중화하고 체계적으로 데이터베이스화 하려고 시도하고 있다.

앞으로는 소재 및 물성과 관련된 빅데이터를 기반으로 한 기계학습 등을 통합적으로 활용함으로써 신소재를 탐색하고 설계하는 소재정보학이 점차 중요해질 전망이다. 소재정보학을 활용하는 이유는 실험 및 계산으로 도출된 물질 정보와 데이터들을 통계적으로 분석하여 소재 구조와 물성을 결정하는 핵심 인자를 파악하기 위함이다. 빅데이터를 기반으로 하는 소재정보학의 목표는 정보학에 의해 파악된 핵심 인자로부터 소재 특성을 지배하는 법칙을 발견하고, 미지의 소재 설계가 가능한 체계적인 접근 방식을 구축하는 것이다.

1.3. 소재 연구데이터의 관리체계

소재 데이터를 체계적으로 수집, 저장, 관리하는 인프라 구축은 미래 R&D 경쟁력의 핵심이며 소재정보학이 데이터를 기반으로 하기 때문에 데이터의 수집부터 활용까지 전주기에 걸친 체계적 관리 및 운영체계를 구축하는 것이 무엇보다 중요하다.



출처 : 한국과학기술연구원(2018)

연구 환경에는 많은 데이터와 인프라가 존재하지만, 연구데이터의 저장 체계가 모두 다르다. 예를 들면 소재 데이터베이스 서비스인 오픈소스 소재 데이터베이스(OQMD, Open Quantum Materials Database)와 Materials Project는 공통적으로 무기 재료의 열역학 성질의 데이터를 저장하고 있지만, 서로 다른 저장 방식을 채택하고 있다. 연구데이터 활용을 위한 관리가 미흡하고 대부분 단순 저장 및 관리에 그치며 기관·정책 차원의 지원이 부족하다. 소재 개발의 발전을 위해서는 데이터 분류, 수집대상 선정 및 데이터베이스화, 저장방식의 체계화, 디지털 큐레이션 등을 통한 중앙 집중적인 단일 통합 시스템을 통한 관리가 필요하다.

연구데이터를 관리하기 위해서는 시스템적인 관리체계를 구축하는 것과 함께 데이터에 따른 분류체계를 갖추는 것이 중요하다. 연구데이터는 일반적으로 정형성 또는 산출방법 기준으로 정형(structured data)/비정형(unstructured data) 또는 실험/분석계산데이터로 구분된다. 정형데이터는 데이터베이스의 고정된 양식에 저장되는 데이터로 처리·분석이 용이한 반면 비정형데이터는 고정된 양식에 저장되지 못하는 데이터로 의미 있는 정보 추출이 어렵다. 비정형데이터로는 전통적으로 사람이 직접 정보를 추출하는 이미지, PDF 파일, 동영상, 오디오, 위치정보 등이 있으며 데이터 관리·활용·분석을 위한 별도의 처리방법이 필요하다. 한편 실험데이터는 실험환경, 실험장비, 실험방법 등 다양한 실험인자로 인해서 구조화·체계화 하는 것이 어렵다. 반면에 현재 주로 제일원리계산을 통해 얻어지는 계산데이터는 실험데이터보다 데이터 관리가 용이하나 사실과의 차이로 인해 검증의 필요성이 있다. 이와 같은 다양한 종류의 데이터를 수집·활용하기 위해서는 데이터의

수집대상을 정하고 저장·데이터베이스화·디지털 큐레이션·관리체계 등에 대한 이해가 필요하다. 또한 연구자들에게 지적 산물들의 공유 및 확산과 관련된 적절한 예시를 제공함으로써 소재 연구데이터의 재활용 및 공유가 소재 개발의 효율성을 증대시킨다는 공감대가 형성될 수 있도록 해야 하며 산재된 양질의 연구정보를 중앙 집중화해 연구의 수월성과 효율성을 제고할 필요가 있다.

표 1. 데이터의 분류 기준에 따른 정의 및 특징

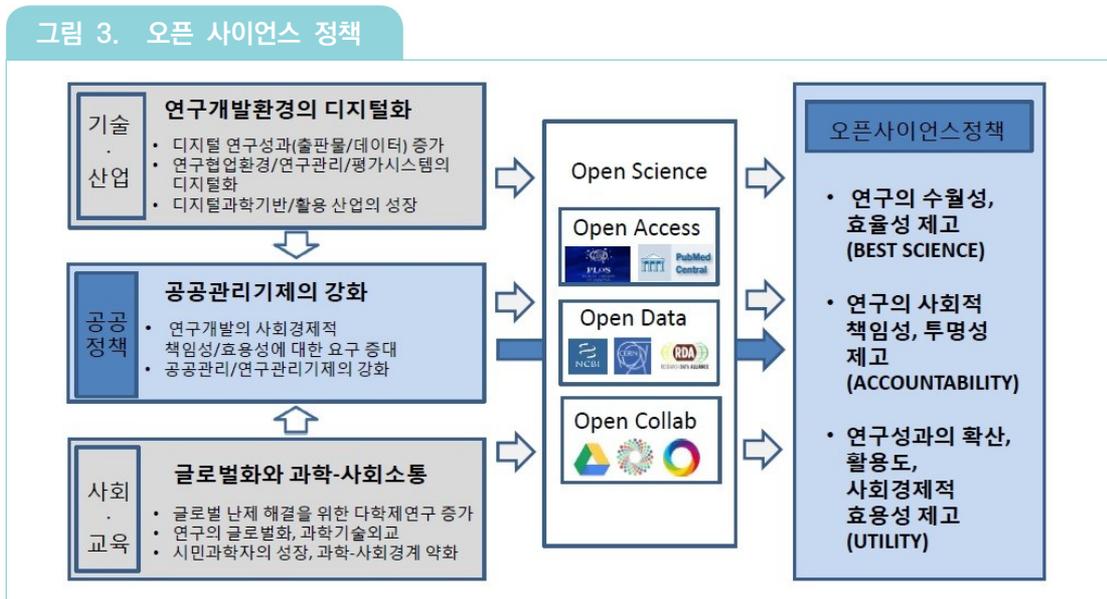
분류기준	데이터 type	정의 및 특징
정형성	정형	• 고정된 필드에 저장되는 데이터
	비정형	• 이미지, PDF파일, 동영상 등 - 구조화/체계화 어려움
산출방법	실험데이터	• 연구노트, 실험을 통해 얻어진 데이터 - 실험인자의 다양성 때문에 구조화/체계화 어려움
	계산데이터	• 계산을 통해 획득한 데이터 - 데이터 관리가 용이함

출처 : 한국과학기술연구원(2018)

1.4. 소재 연구데이터 공유의 필요성

오픈 사이언스 정책은 공공연구의 성과인 연구데이터를 디지털 포맷으로 공개, 확산시켜 사회경제적 편익을 제고하려는 공공정책이다. 오픈 사이언스의 발전사는 최초의 온라인 저널을 기반으로 1991년 개설된 arXiv와 같은 웹사이트가 과학 출판물에 대한 오픈 액세스를 제공하면서 시작된 것으로 볼 수 있다. 연구개발 결과물이 온라인으로 제공되면서 물리적 도서관의 역할이 감소하고 저널 구독 비용이 낮아졌다. 한편, 문헌이 아닌 데이터의 수집 및 저장과 관련해서는 조금 더 일찍 역사가 시작되었다. 1950년대부터 이미 과학적 글로벌 데이터 센터를 설립하기 위해 데이터를 장기적으로 저장하기 위한 노력을 해왔고 이 시기부터 많은 맞춤형 과학 데이터베이스가 확산되어 전 세계 거의 모든 분야에 데이터 저장소를 제공하여 현재는 수천 개의 데이터 센터가 존재한다. 이후 과학적 데이터의 공유에 대한 요구 증가로 2004년 대부분의 선진국들은 공적 예산을 지원하기로 결정했으며 관련 저장소를 검색하고 그 안에 있는 데이터를 활용하려는 노력들이 진행되었다. 이에 발맞추어 유관 기관들은 데이터 관리 계획과 함께 데이터 저장, 보안 및 보호 방법에 대한 오픈 사이언스 정책 관련 규칙의 필요성을 주장하였으며, 그 결과 데이터 관리 개발을 지원하고 통일된 지침을 널리 제공하기 위하여 FAIR(Findable, Accessible, Interoperable, Reusable) 데이터 원칙이 제정되었다.

오픈 사이언스의 실천전략인 기관 리포지토리(Institutional Repository, 정보자원의 공유 및 저장 역할을 하는 저장소) 구축을 통해, 전 세계 누구나 자유롭게 연구데이터 정보에 접근해 연구와 이용자 간 데이터를 공유함으로써 연구자들은 연구 성과를 가시화 및 확산시키고 활용하며 기관은 위상을 제고해 연구경쟁력을 향상시키는 것이 세계적 흐름이다. 연구데이터의 양과 종류는 기하급수적으로 증가하고 있고, 앞으로 더욱 급속하게 증가할 것으로 예측된다. 또한 하나의 시스템에서 다른 시스템으로 메타데이터를 전송하고 전송받을 수 있는 프로토콜이 개발되는 등 다양한 종류의 디지털 자원에 대한 개방성 및 접근성을 향상시키려는 움직임이 있다. 최근에는 연구자가 기관 리포지토리에 직접 연구데이터를 올리고 저장하는 셀프 아카이빙(self archiving) 시스템을 통해 기관 리포지토리는 연구데이터를 전 세계에 신속하고 효과적으로 배포하고 공유한다. 이처럼 연구커뮤니티를 중심으로 글로벌 연구협력이 일상화되고 있다.



출처 : 한국과학기술정보연구원 이상환(2019)

1.5. 소재 연구데이터 플랫폼의 필요성

2007년부터 소재 물성 데이터베이스가 구축된 소재정보은행(2016년 기준으로 108만 건의 소재 데이터 구축) 중심으로 산업용 소재 물성 데이터 검색 및 분류 서비스가 제공되고 있으나 일부 소재 및 응용분야에 집중되어 있어 한정된 범위의 물성 검색만 제공되고 있다. 2013년 과학기술정보통신부의 창의소재 디스커버리

사업을 통해 소재 빅데이터와 데이터 기반 소재 개발 사업이 시작되었으나 데이터 부족으로 소재 개발에 활용할 수 있는 소재정보학으로 발전하기에는 미흡한 수준이다. 이성과 같이 현재 국내에서는 데이터베이스화된 소재 R&D 데이터를 적용할 수 있는 플랫폼이 미흡하며 연구데이터 기반의 소재 개발을 위해 데이터 수집·공유·활용을 위한 연구데이터 인프라를 구축할 수 있는 자체 기술을 확보할 필요가 있다. 선진국들은 소재 개발 R&D 전주기 동안 생성되는 연구데이터를 체계적으로 수집·공유·활용하고자 데이터 기반의 인프라를 강화하고 있다. 2000년대 들어 미국은 소재계놈이니셔티브(MGI), 일본은 MatNavi를 통해 데이터 수집·공유·활용 기반을 마련하고 있다.

입자물리, 바이오, 천문학 분야 등 타 분야의 데이터 공유 상황을 보면 거대규모 실험 장치에서 생산되는 데이터 양이 워낙 방대하여 데이터 공유가 필수이다. 이에 반해서 소재분야 데이터 연구는 전통적으로 소규모 데이터 생성과 분석 접근에 익숙하여 데이터 재활용 및 공유에 대한 연구자들의 인식이 부정적이다. 소재 연구데이터 공유 및 빅데이터를 활용한 기계학습과 분석정밀도 증가는 우수한 연구결과로 귀결될 것으로 예견된다. 따라서 소재 응용분야별 소재 데이터 수집 및 공유 플랫폼 구축이 필요하고 범국가적 연구데이터 플랫폼과 연계·활용 가능한 네트워크를 구성해야 한다.

R&D 환경의 경우, 연구 주제 및 방법, 활용 목적에 따라 수집되는 데이터의 형태가 다르고, 연구자별 데이터 저장 및 분류 방식이 상이하여 정형화된 메타데이터를 통한 빅데이터 수집은 한계가 있고, 정형화된 메타데이터는 정형화를 위한 연구자(End-User)의 추가적인 작업이 수반되어야 한다. 추가적인 작업은 연구자의 불편과 업무 효율성 저하를 가져올 수밖에 없고, 이로 인해 연구자들이 수집 플랫폼을 사용하지 않는 악순환이 발생할 가능성이 매우 높다. 불편함에서 오는 한계를 극복할 수 있는 하나의 방법으로 연구자 행동기반 데이터가 있다. 이는 연구노트(KRI, Korean Research Informatics) 및 분석 신청 등 연구자의 행동이 실험, 시뮬레이션 등에서 도출되는 데이터의 메타데이터의 역할을 하는 것으로, 빅데이터 수집을 위해 연구자에게 추가적인 작업을 요구하지 않아도 되며, 비정형 메타데이터를 통해 유연한 데이터 운용이 가능할 것으로 전망된다. 기술적으로 연구자 행동기반 데이터는 실험 장비 및 시뮬레이션으로부터 출력되는 데이터와 연구자 행동 기록 데이터로 이루어지는데, 데이터의 수집을 위해서는 연구노트 작성, 합성, 분석, 시뮬레이션 등 연구자의 행동을 데이터화 하고, 실험 장비 및 시뮬레이션 데이터를 처리할 수 있는 일관된 데이터 처리·수집 환경이 필요하다.

주제별 또는 분야별로 구축되는 데이터 수집 플랫폼의 표준화를 통해 데이터와 사용자 환경의 일관성을 확보하고 개발 절차 등의 단순화를 위한 표준 개발 프레임워크의 구축이 시급한 상황이다. 소재 연구데이터

플랫폼 구축에 프레임워크 개념을 도입하면 몇 가지 중요한 장점이 있다. 프로그램 개발 중 사용되는 재사용이 가능한 수많은 코드를 통합 및 모듈화를 통하여, 모듈들을 단순히 조립하여 표준화된 소재 세부분야에 따라 특화 플랫폼을 생성하는 것이 가능하다. 연구데이터 플랫폼 개발에 필요한 데이터베이스 연동, 템플릿 형태의 표준 그래픽 사용자 환경(GUI, Graphical User Interface), 세션 관리 등의 기능을 모듈화하여 프레임워크로 제공함으로써 복잡한 웹 서비스 및 웹 애플리케이션을 신속하게 작성하는 것이 가능하다. 예로써 정부 주도로 '전자 정부 표준프레임워크'를 구축 및 공개하였으며, 2018년 3월 756개의 공공/민간 기관의 정보화 사업에 활용 중이다.

2. 소재 연구데이터 구축·관리·활용의 전략적 변화

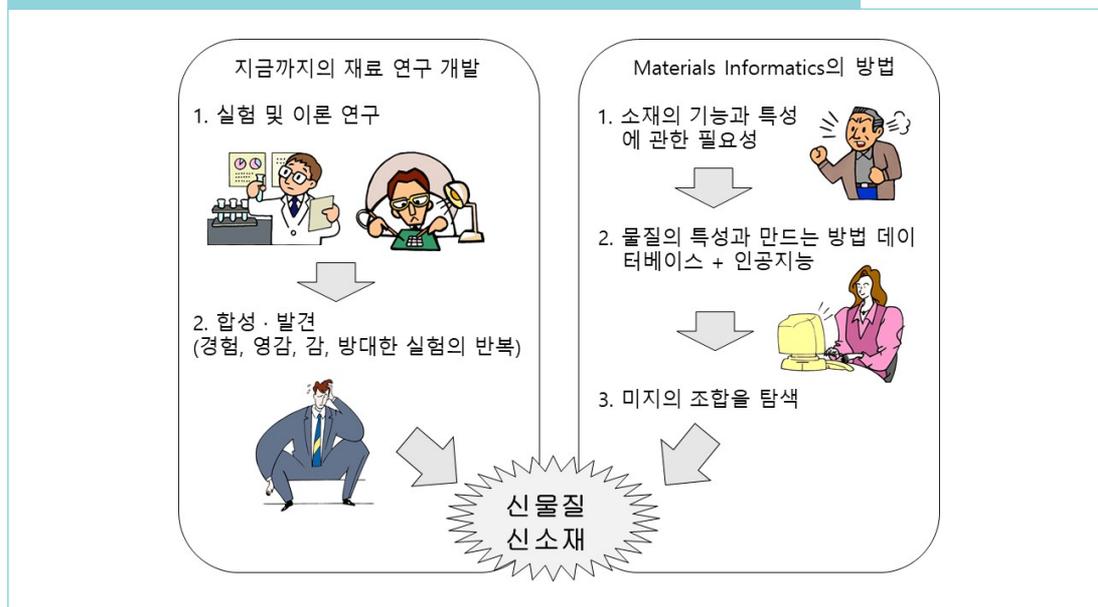
2.1. 소재 연구데이터의 국가적 전략 및 정책

연구데이터의 정보학을 시작하기 위해 미국에서는, 국가가 주도적으로 소재게놈이니셔티브(MGI) 프로젝트를 시작한 후, 대학과 연구소를 중심으로 소재 연구데이터 인프라가 구축되면서 가시적인 성과가 도출되고 있다. 예로써, 브리티시 컬럼비아대, 로런스 버클리 국립연구소(LBNL, Lawrence Berkeley National Laboratory), 텍사스 A&M 대학교(D3EM, Data-Enabled Discovery and Design of Energy Materials), 듀크대학(소재게놈센터) 등이 설립되었다. 국내는 R&D 과정에서 발생하는 연구데이터를 국가적 자산으로 관리·활용할 수 있는 기반과 정책이 선진국에 비해서 늦게 진행되었고 소재정보학은 이제 막 시작하는 단계로 볼 수 있다. 소재연구를 통해 생성된 데이터를 연구자만이 소유하고 활용하기에는 투입된 국가 R&D 비용에 비해 활용도가 낮다는 평가다. 따라서 데이터 신뢰도를 높이고 공유함으로써 후발 연구자와 산업체가 공공목적으로 용이하게 활용하도록 국가적 유인책이 필요하다. 우선 다양한 소재 데이터를 연구자가 쉽게 저장·관리·활용할 수 있는 환경을 제공할 수 있는 국가적 인프라 플랫폼을 시급히 구축하는 것이 필요하다. 이를 위해 과학기술정보통신부는 2020년 9월에 '소재연구데이터 플랫폼사업단'을 향후 7년간 진행하는 프로젝트를 시작하였다. 사업단은 연구데이터 수집, 공유, 활용에 대한 로드맵과 연구데이터 수집 및 관리 지원 등의 데이터 생태계의 활성화를 위한 공유 활용 플랫폼을 구축할 예정이다. 또한 국제 수준의 소재 연구데이터의 혁신적인 표준화를 통해 범세계적으로 소재 데이터 표준화에 대한 주도권을 확보하여 한국 중심의 국제 소재 연방 네트워크(Federated Materials Registry Network)를 구성함으로써 소재 데이터 국제표준을 선점하는 것이 중요하다.

2.2. 소재 연구데이터를 활용한 소재정보학

시트린 인포매틱스(Citrine Informatics)사는 소재 데이터 표준화 작업선도, Materials Information File(MIF, Open JSON-based file format, created by Citrine Informatics) 등을 데이터 표준 포맷으로 공식화하기 위해 노력 중이다. 재료분야에서 다양한 소재에 대한 통합 데이터 시스템이 구축된다면 향후 더욱 강화된 기계학습과 분석을 통해 신소재 개발이 가속화될 것이다. 데이터 정보학을 구현하고 연구데이터 활용·촉진을 위해서는 소재의 합성방법 및 특성평가 방법에 따른 소재 데이터의 표준화가 필수적이다. 데이터를 표준화한다는 것은 재활용·공유·분석 데이터들의 통일된 저장방식을 구현하고, 소재 연구데이터 수집·활용의 편리성을 제공하며 측정 장비의 신뢰성(측정소급성)을 확보할 수 있는 장점이 있다.

그림 4. 기존의 재료연구개발과 소재정보학에 의한 재료개발의 차이점



출처 : 소재기술백서(2017)

- 소재정보학 : 계산 과학에 의한 특성 예측과 이를 실증하는 고속(high throughput) 합성 및 평가, 소재물성 데이터베이스와 기계 학습 등을 통합적으로 활용하여 신소재를 탐색하고 설계하는 연구개발 활동
- 소재정보학 목표 : 정보학에 의해 파악된 핵심 인자로부터 재료 특성을 지배하는 법칙을 발견하고, 미지의 재료 설계가 가능한 체계적인 접근 방식 구축

- 소재정보학 방법론 : 실험 및 계산으로 도출된 물질 정보와 데이터들을 통계적으로 분석, 소재의 구조와 물성을 결정하는 핵심인자를 파악하고, 새로운 소재를 빠른 속도로 탐색하는 연구 방법론
- 소재정보학 기대효과 : 일반적으로 소재 개발은 ①발견, ②개발, ③물성최적화, ④시스템 디자인, ⑤통합, ⑥인증, ⑦제조의 7단계를 거치며, 하나의 소재를 개발하기 위해 통상적으로 10-20년이 소요되며, 빅데이터 활용을 통한 물성 예측 및 고속 실증으로 신소재 탐색, 설계, 제조까지 혁신적 기간 단축 가능
- 소재정보학 활용을 위한 요구사항 : 소재 실험/분석/계산 데이터베이스 구축, 소재 연구데이터의 효율적 활용을 위해 소재분야에 적용 가능한 데이터 마이닝 기술과 머신러닝, 인공지능 기술개발 필요, 선진국과 비교하여 소재 연구 데이터베이스 구축의 후발 주자이기에 데이터베이스 구축과 동시에 소재 데이터베이스에 대한 융합적 활용방법론(머신러닝 방법 등)을 개발하는 Two-Track적 플랫폼 구축 필요

소재 연구데이터의 수집부터 활용까지 고려할 때 몇 가지 핵심적인 개념이 필요하다. 먼저 소재 연구데이터의 표준화를 통하여 소재 응용분야 전반에 활용할 수 있는 유연하게 수집·관리·활용 가능한 소재 연구데이터 수집 및 공유 플랫폼을 구축하고 범국가적 연구데이터 플랫폼과 연계·활용 가능한 네트워크를 구성할 필요가 있다. 또한 범국가적 소재연방 네트워크를 형성하여 전 세계 규모의 빅데이터 수집 및 연구자 간의 교류를 활성화함으로써 국내의 소재 연구데이터와 함께 글로벌한 데이터의 수집이 필요하다. 궁극적으로는 데이터베이스 구축과 동시에 소재 데이터베이스에 대한 융합적 활용방법론(머신러닝 방법 등)을 개발하는 계층적(Tier) 플랫폼의 구축이 필요하다. 개발 플랫폼을 이용함으로써 축적된 빅데이터 기반 머신러닝 알고리즘을 실제 소재 응용분야에 적용할 수 있을 것이다.

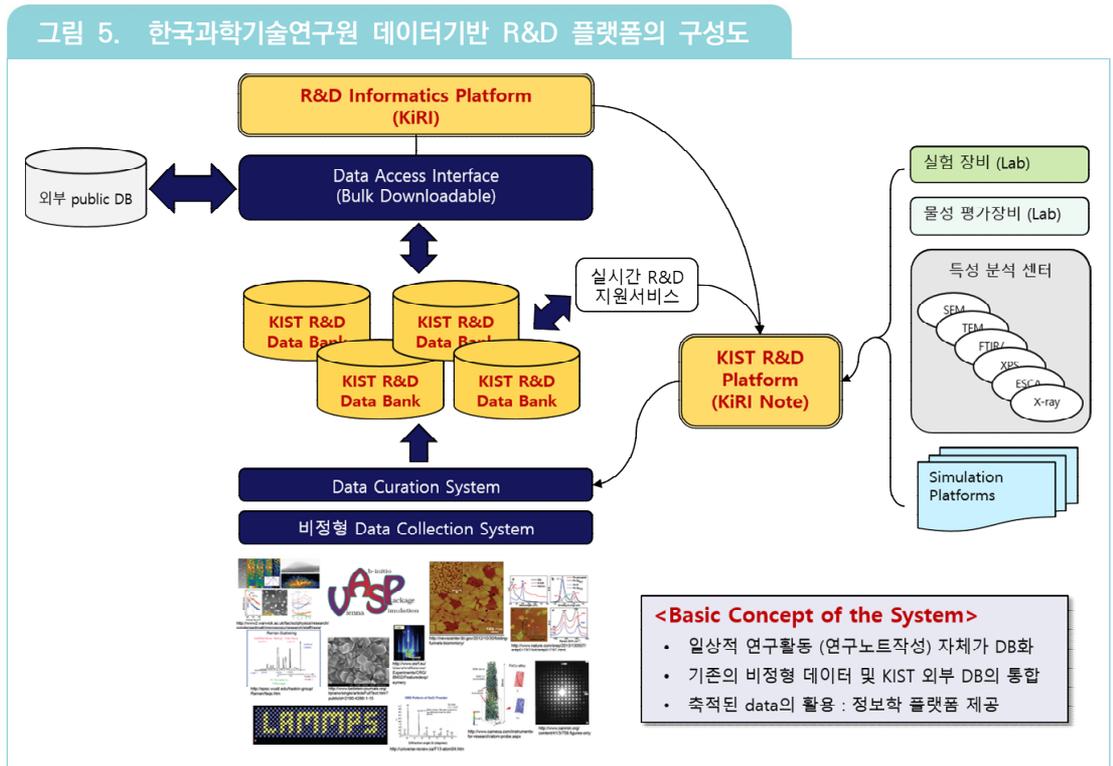
II 국내외 소재데이터 현황 및 규정

1. 국내현황

2020년 10월부터 과학기술정보통신부는 소재 연구데이터를 체계적·안정적으로 수집·관리할 수 있는 인프라를 구축하고, 소재 연구개발에 활용하는 것에 주안점을 두고 데이터 기반 소재연구 혁신허브 구축·활용 방안을 추진하고 있다. 특히, 데이터 입력 표준 템플릿을 2021년까지 확립하고, 인공지능 학습을 위한 양질의 메타데이터 400만 건 이상을 확보할 계획이다. 수집대상은 연구개발로 생성되는 모든 데이터로, 과학기술정보통신부 및 타 부처 과제, 기초연구 등으로 연구 과제를 확대하며, 논문·특허 추출, 기존 데이터 재규격화, 신규 생성 등 수집방식을 다각화하는 것을 목표로 한다. 또한, 대용량 데이터 관리를 위한 전용 초고성능컴퓨팅 환경을 2026년까지 구축하고, 수집한 데이터 및 인공지능(AI, Artificial Intelligence)을 활용해 신소재 탐색·설계, 공정개발, 양산에 이르는 기간을 줄이기 위한 서비스도 제공한다. 이외에도 ‘국가 소재 연구데이터 센터’를 운영해 데이터를 체계적으로 관리하고 에너지·환경, 스마트·정보통신(IT), 구조(안전) 등 소재 응용분야별 특화센터를 설립하며, 정부 과제에 대한 데이터 등록 의무화 등을 담은 가이드라인도 마련한다. 또한 축적된 대용량 데이터를 안정적으로 관리하기 위해 2026년까지 3,000개 연산코어를 갖춘 슈퍼컴퓨터를 도입할 예정이다. 단기 시범과제와 중장기 대형과제를 병행하며 대학생, 대학원생, 소재 R&D 종사자, 기업 재직자 등을 대상으로 대상별 맞춤형 전문교육을 통해 인공지능 기반 소재 R&D를 수행할 전문인력 양성도 지원한다.

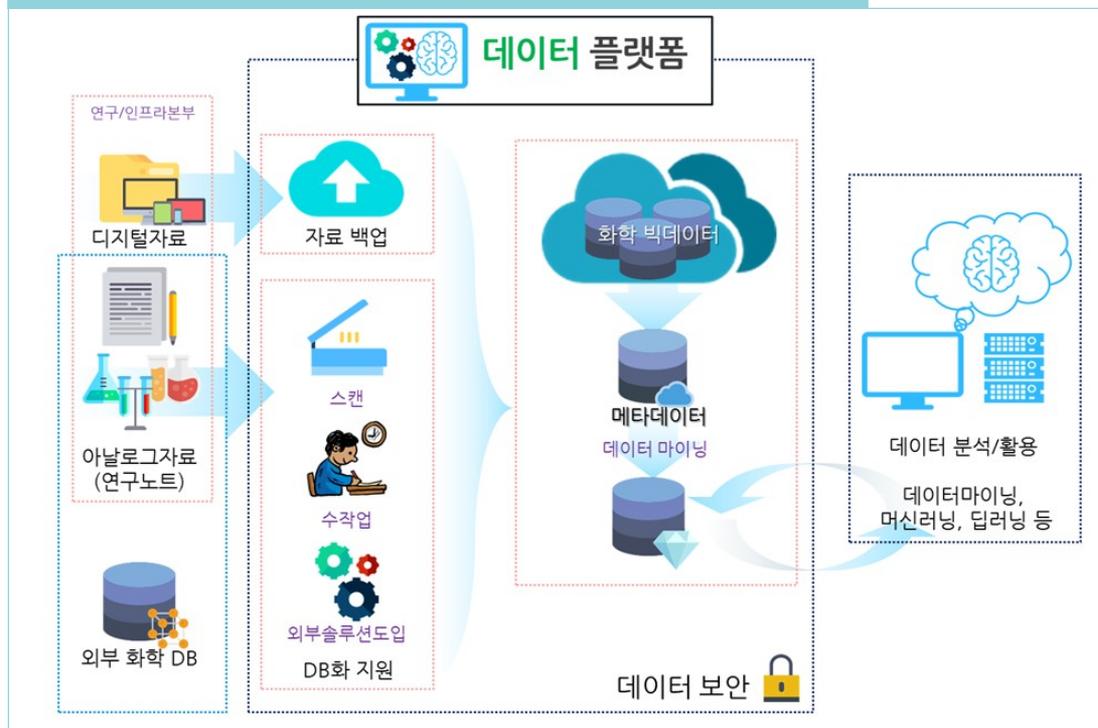
소재 연구데이터를 수집하고 활용하기 위해 이미 진행되고 있는 과제들로는 미래소재디스커버리사업, 나노·소재기술개발사업 등이 있다. 미래소재디스커버리사업은 계산과학, 조합화학법, 복합측정 등 정보통신기술(ICT) 기반의 新연구기법을 활용하여 새로운 물성의 소재 개발사업(2015년~, 16개 연구단)을 수행 중이며 나노·소재기술개발사업은 웹기반 계산과학 플랫폼 구축을 위한 시범사업(2016년~), 소재 빅데이터 플랫폼 시범구축(2017년~), 도전형 소재기술개발(2017년~) 등을 추진 중이다. 국내 연구 결과물인 논문과 특허에 대해서는 국가과학기술지식정보서비스(NTIS)에서 수집하여 관리하고 있으며 국가 R&D 사업 전주기 정보 이력과 온라인 성과평가 시스템 구축, 기술이전·사업화 관련 정보를 통합 제공한다.

한국과학기술연구원(KIST)은 2017년부터 촉매 및 전지 소재에 대해 기계학습을 통한 신소재 설계를 위한 기초 사업을 수행 중이다. 특히 한국과학기술연구원(KIST) 내 모든 연구 분야에 적용될 수 있는 범용의 전자연구노트를 구축 중이다. 또한 한국화학연구원(KRICT)은 화학데이터의 수집 플랫폼 구축과 열전 및 유기태양전지 데이터 수집을 통해 소재 예측과 개발에 필요한 기초 사업을 수행 중이다.



출처 : 한국과학기술연구원(2018)

그림 6. 한국화학연구원 데이터 플랫폼을 활용한 데이터 수집 및 활용



출처 : 한국화학연구원(2018)

산업통상자원부는 2021년 3월 디지털 소재혁신 강화 실행계획을 발표하고, 소재 개발에 필요한 데이터 축적과 활용을 확산하기 위한 과제를 지원한다. 프로젝트별로 데이터의 항목·구조·방식 등 표준입력 템플릿을 확보한다는 계획이다. 금속, 화학, 세라믹, 섬유 등 4대 소재의 개발 단계별로 공통 구성항목을 도출하고 구체화한다. 2022년까지 10만 건 이상의 양질의 데이터 확보를 목표로 소재 개발에 필요한 데이터 규모·수준을 확보할 계획이다. 기존 데이터는 해당 프로젝트와 관련된 본래의 소재정보은행 데이터를 재규격화하고, 문헌 자료 및 추가 실험 등을 통해 보완해 나가며, 실험데이터는 프로젝트별 컨소시엄이 직접 실험을 통해 고품질 기계학습이 가능한 대량의 신규 데이터를 생성한다. 축적된 데이터로는 소재 개발 인공지능 예측모델 개발, 활용목적별 서비스 제공, 소재의 부품화 가능성 검증, 전문인력 양성 등을 추진한다. 또 기업, 연구자 등에게 데이터를 제공해 자체모델 개발이 어려운 중소기업에 인공지능 표준모델 소스 코드를 개방한다. 이와 함께 데이터 표준화·축적·활용을 선순환 구조로 확산해 나가기 위한 표준 제정, 인센티브 부여, 관리 시스템 구축과 같은 제도·체계 등도 확충한다.

소재정보은행의 경우 2007년부터 금속, 화학, 세라믹, 섬유 분야에 대한 약 108만 건의 소재물성 데이터베이스를 수집 중이며 이외에 기술 및 시장동향, 특히, 각종 보고서 등을 생성·수집하고 있다. 또한 소재정보은행은 축적된 정보를 기반으로 부품화에 필요한 물성정보와 기술적 솔루션을 부품업체에 제공한다.

국가참조표준센터는 측정 데이터 및 정보의 정확도와 신뢰도를 과학적으로 분석·평가하여 국가가 공인한 고품질의 데이터를 수집·생산, 평가하고 있으며 참조표준(안) 제정을 통해 신뢰성 있는 물성 데이터를 제공한다. 2006년부터 금속, 재료 등 8개 분야의 물성 데이터(10개 분야 65개 DB, 3만 8천개 데이터)를 제공하고 있다.

표 2. 소재정보은행 물성 DB 구축 현황(2016년 12월말 기준)

분야	주관기관	주요 내용	구축 건수		
			수집·가공	생성	합계
금속	재료연구소	철강, 알루미늄합금, 마그네슘합금, 분말, 구리소재 등	121,129	90,561	211,690
화학	한국화학연구원	플라스틱, 고무, 첨가제, 필름, 접착제, 도료, 복합재료 등	558,352	43,182	601,534
세라믹	세라믹기술원	반도체소재, 센서소재, 기계구조재료, 생체소재 등	159,503	35,145	194,648
섬유	다이텍연구원	PET, 나일론, 아라미드섬유, 탄소섬유, 토목용 섬유 등	56,402	17,560	73,962
합 계			895,386	186,448	1,081,834

출처 : 미래소재 연구데이터 플랫폼 구축사업 보고서(2019)

2. 해외현황

우리 정부는 오픈 사이언스, 빅데이터 및 인공지능을 포괄하는 새로운 소재정보학을 위해 데이터 기반 연구개발을 추진하는데 초점을 맞춘 전략적 연구 기회를 제공하고 있다. 그러나, 아직 소재·부품·장비 응용 프로그램이나 산업 응용 프로그램을 찾고 개발하는데 제한적이다.

선진국의 경우 데이터 수집·관리의 중요성을 인지하고 국가적 차원의 지원과 육성 정책을 활발히 추진하고 있으며 데이터 수집·활용 및 플랫폼 구축을 위해 미국, 일본, 유럽 등은 자국의 연구역량 강화를 추진하고 있다. 미국 또는 유럽연합의 경우 레지스트리 연합(Registry Federation)을 통해 상호 기록 교환 방식 네트워크를 구축하는 개념을 도입 중인데 레지스트리는 데이터 기반 연구에 유용한 데이터 저장소, 아카이브, 웹사이트 및 서비스 같은 리소스에 대한 상위 수준의 메타데이터를 수집, 기록을 공유하고 통일성 있는 시스템을 구축하여 빅데이터의 한계를 극복하려는 노력을 하고 있다. 글로벌 레지스트리 연합 프레임워크(Federated Registry Framework)의 연결 수준에서 보면 마스터 레지스트리를 만들지 않고 여러 개의 레지스트리들이 함께 구동할

수 있는 네트워크 방식으로 구축한다. 레지스트리 연합은 개방적이며, 모든 조직은 그 조직이 보유하고 있는 자원을 세계에 광고하는 수단으로 레지스트리를 호스팅(Hosting)한다. 레지스트리 제공자는 특정 자원의 하위 세트에 대한 설명 레코드를 작성하고 큐레이션 데이터를 만든다.

계층 재료 설계 센터(CHiMaD, Center for Hierarchical Materials Design)는 특정 하위 커뮤니티에 대한 레코드 관리 및 사용자 정의된 API(Application Programming Interface, 애플리케이션 프로그래밍 인터페이스, 응용 프로그램에서 사용할 수 있도록 운영 체제나 프로그래밍 언어가 제공하는 기능을 제어할 수 있게 만든 인터페이스)를 통해 센터 회원 조직이 제공하는 데이터 공유가 가능하고 국립표준기술연구소(NIST, National Institute of Standards and Technology)는 일반적으로 그래픽 사용자 인터페이스(GUI, Graphical User Interface)를 통해 대규모의 커뮤니티와 데이터 공유가 가능하며, 작은 단위의 데이터베이스 저장이 가능하다.

소재개념이니셔티브(MGI)를 통해 데이터 기반 재료과학 분야를 발전시키기 위해 상당한 정부 자금을 제공한 미국을 제외하고, 우리 정부를 포함한 대부분의 정부 지원 자금 계획은 보다 일반적이다. 유럽연합에서는 두 개의 성공적인 프로젝트(MaX 및 NOMAD CoE)에 자금이 지원되었다. 그러나 두 센터 모두 유럽연합의 Horizon 2020 고성능 컴퓨팅 보조금에 의해 촉진된 것으로 데이터 기반 재료과학에 초점이 맞춰져 있지는 않다. 새로운 프레임워크 프로그램(FP, Framework. Program)에서 변경될 수 있지만 유럽연합에서 아직까지 데이터 기반 재료과학의 탐색 및 활용에 대한 요청은 발행되지 않았다.

국가적 차이가 존재하지만, 각국은 소재 데이터를 수집하고 활용하는 저장 인프라를 제공하고 있으며 제공하는 데이터에 따라서 아래 표와 같이 정리된다.

표 3. 각국에서 운영 중인 소재 데이터 인프라 및 서비스

	Open access	Comp. data	Exp. data	Data upload (DOIs)	Workflow management tools	Web API	Data analysis tools
AFLOW	✓	✓			✓	✓	✓
Computational Materials Repository	✓	✓			✓		✓
Crystallography Open Database	✓	✓	✓	✓			
HTEM	✓		✓	✓		✓	✓
Khazana	✓	✓	✓				✓
MARVEL NCCR	✓	✓		✓	✓		✓
Materials Data Facility(MDF)	✓	✓	✓	✓(DOI)		✓	
Materials Project	✓	✓			✓	✓	✓
MatNavi/NIMS	✓	✓	✓				✓
NOMAD CoE	✓	✓		✓(DOI)		✓	✓
Organic Materials Database	✓	✓					✓
Open Quantum Materials Database	✓	✓					✓
Open Materials Database	✓	✓		✓	✓	✓	✓
SUNCAT	✓	✓				✓	✓
Citrine Informatics	✓	✓	✓	✓		✓	✓
Exabyte.io						✓	✓
Granta Design		✓	✓				✓
Materials Design		✓	✓				✓
Materials Platform for Data Science	✓	✓	✓			✓	✓
Materials Zone			✓				✓
Springer Materials			✓				✓

* Open access : 무료 액세스 제공/ Comp. data : 소프트웨어 시뮬레이션에서 생성된 데이터/ Exp. data : 실험에서 생성된 데이터/ Data upload : DOI(Digital Object Identifier) 발급 가능성과 함께 자체 데이터의 업로드 가능/ Workflow management tools: 오픈 소스 소프트웨어 도구를 제공하거나 개발하는데 협력/ Web API: 자동화된 스크립트를 사용하여 데이터에 원격으로 액세스/ Data analysis tools: 기계학습을 포함한 온라인 또는 오프라인 데이터 분석 도구 제공
출처 : Himanen(2019)

미국은 국가적인 연구데이터 혁신으로 소재연구 생태계의 연결 고리를 강화하는 프로젝트를 가장 먼저 시작하였다. 1960년대부터 데이터의 중요성을 인식하였고, 2011년 소재계놈이니셔티브(MGI)를 발표하며 빅데이터의 소재분야 연구 패러다임 변화를 주도하고 있다. 1968년 신뢰성이 검증된 표준참고자료(SRD, Standard Reference Data) 데이터베이스 구축사업을 시작으로 디지털 데이터를 생산·활용하는 연구를 집중 지원했으며, 2005년 디지털 데이터의 중요성과 파급효과를 인식하고 데이터의 체계적 수집 및 관련 교육 시스템의 구축을 권고한 바 있다. 2008년 소재 개발 효율성 확대를 위해 통합 전산 재료 공학(ICME, Integrated Computational Materials Engineering)을 활성화하였으며 2011년 소재계놈이니셔티브(MGI) 수립을 계기로 단순 참고자료 수준이 아닌 연구데이터의 체계적인 수집·공유·유통 및 활용으로 이어지는 확산적 연구를 활성화하였다.

현재 소재 전 분야 연구데이터의 수집·활용을 통한 빅데이터화 및 소재 개발 효율성 향상을 위한 다양한 연구가 추진 중이다. 국립표준기술연구소(NIST)를 중심으로 외부 연구진 간의 공동 연구를 통해 데이터 수집·검색 서비스 제공, 데이터를 활용한 소재 개발 및 빅데이터 인프라 구축 등을 추진하고 있다. 국립표준기술연구소(NIST) 역할은 체계적으로 데이터를 광범위하게 수집할 수 있는 데이터 관리를 양식 프로그램(MDCS, Materials Data Curation System)을 개발하여 필요로 하는 글로벌 연구자에게 무료로 제공하는 것이다. 국립표준기술연구소(NIST) 주관 사업 이외에도 특정소재 및 주제에 대한 데이터베이스 구축·보급 및 활용사업이 활발하게 진행되고 있다. 구조용 금속소재 실험 및 계산과학 연구 그룹(PRISMS, PRedictive Integrated Structural Materials Science)과 계산·시뮬레이션 연구를 위한 실험 정보 수집·공유 그룹(Materials Common) 간 협력 연구가 진행 중이다. 또한 소재·화학분야 연구데이터 정보의 공유·분석이 가능한 플랫폼 구축 사업화도 추진 중이다.

국립표준기술연구소(NIST)는 데이터 수집을 위한 메타데이터의 기록·검색을 위한 포맷을 결정하기 위해 MDCS(Material Data Curation System)를 개발하여 2015년 첫 공개한 이후 지속적으로 업그레이드 중이다. MDCS의 무상 배포를 통해 관련 연구데이터 수집 단계에서부터 체계적 관리를 유도하며 향후 빅데이터 형태로의 전환에 대비하여 범국가적 연구데이터 관리체계의 확립을 추진 중이다. MDCS는 소재 데이터의 수집·공유·가공을 일정한 포맷 내에서 관리할 수 있는 XML(Extensible Markup Language, 확장성 생성 언어) 기반의 시스템을 활용하여 데이터를 일정한 포맷에 맞추어 입력·검색·출력을 가능하게 한다. XML은 HTML의 한계를 극복하기 위해 만든 언어로, 웹상에서 데이터를 표현하고 교환하기 위한 표준화된 텍스트 형식을 의미한다. 데이터 포맷은 국립표준기술연구소(NIST) 내 전문가들에 의해 결정되며, GitHub(<http://github.com/usnistgov/MDCS>)에서 다운로드가 가능하다.

일본은 소재 연구개발 시 소재 데이터베이스 기반 구축으로 연구데이터의 활용을 촉진하려는 프로그램을 국립 재료 과학 연구소(NIMS, National Institute for Materials Science)를 통해서 구축하였다. 국립 재료 과학 연구소(NIMS)는 운영·관리하는 소재 빅데이터 통합사이트인 MatNavi를 통해 물질·재료 데이터베이스 서비스를 제공한다. MatNavi(<http://mits.nims.go.jp>)는 공표된 학술문헌, 계산과학으로부터 연구데이터를 집적화하여 고분자, 무기재료, 초전도재료 등으로 분류한 정보(현재 51만 건 이상 수집)를 제공 중이다.

국립 재료 과학 연구소(NIMS)는 2001년 Materials Information Technology Station(MITS)을 구성하여 사용 가능한 데이터 인프라를 구축하여 일찍이 2003년부터 고분자 DB, 결정구조 DB, 원자력물질 DB, 전자구조 DB 등 11개 연구 분야로 서비스를 시작하였다. 2004년에는 검색 엔진(Search Engine)을 탑재한 'MatNavi 사이트'를 공개하고 격년으로 일본 내에서 'Symposium on Materials Database' 개최를 통해 MatNavi 서비스의 활용성을 향상시키고 있다. 2010년 이전에는 주로 문헌이나 자료, 실험 등을 통한 소재 특성의 데이터베이스를 구축하였으나 2013년부터 제일원리 계산과학 기반의 소재 물성 데이터베이스를 구축하고 있다. 2015년에는 소재 개발에 빅데이터를 활용한 플랫폼(Materials Information Integration Initiative, MI2I)을 구축하고 이론·실험·계산과학 및 데이터 과학을 연계한 혁신적 소재 개발 연구를 확대하고 있다. 소재정보학 등 새로운 연구방법론을 활용하여 소재 개발 기간과 비용을 획기적으로 감축하기 위한 연구개발에 집중하는 추세이다. MatNavi의 주요 소재 데이터는 발표된 학술 논문을 통해 수집·구축되며 대부분의 물성 값들은 검색 결과로서만 데이터 값이 제공되며 디지털 데이터로는 제공되지 않아 활용하는데 한계가 존재한다. 또한 산업용 DB와 연구용 DB 제공 및 웹 플랫폼을 통일하여 4개의 그룹으로 운영하며 기본 특성(폴리머, 무기물, 상평형, 전자재료 DB 등), NIMS Structural Materials Datasheet(인장강도, 부식 등), 산업용 DB, 응용분야별 DB 등을 제공하고 있다.

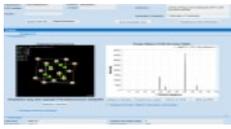
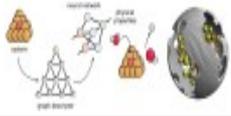
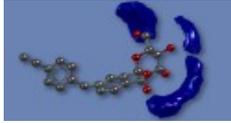
표 4. MatNavi 데이터베이스 내용

명칭	분야 & 특성	데이터베이스 크기
Polymer Database (PoLyInfo)	<ul style="list-style-type: none"> 고분자의 구조 및 물성 문헌상의 정보 및 연구자 제공 데이터 NMR 데이터(154 samples) 	333,583개
Inorganic Material Database(AtomWork)	<ul style="list-style-type: none"> 문헌상에 보고된 무기물, 금속 소재의 결정구조, 물성, 상태도, XRD, 화합물 물성별(기계적 특성, 광학특성) search 기능 	28,000 결정구조, 55,000 재료 특성, 15,000 상태도
Computational Electronic Structure Database (CompES-X)	<ul style="list-style-type: none"> Inorganic Materials Database(AtomWork)의 구조에 따른 제일원리 전자구조 계산 물성 DB 제일원리 계산의 입력값 및 최적화된 구조, DOS, 밴드구조, 전하분포 등 문헌에 보고된 계산 값의 데이터베이스 VASP 계산 결과 데이터베이스 	80,000가지 화합물

출처 : 국가연구데이터 공유활용 체제 구축 사업 기획보고서(2019)

유럽은 과학 데이터 기반의 가치 창출을 위한 오픈데이터 전략을 추진하면서 국제적인 리더십을 이끌고 있다. 소재 연구데이터 개방을 통해 접근성 및 활용성을 극대화하며 2007년 제7차 프레임워크 프로그램(FP7)을 시작으로 계산과학 등을 기반으로 한 모델링 연구를 활성화 하였다. 영국은 빅데이터 및 신소재 연구, 우주, 에너지, 고령화 사회 연구에 약 10조 원(2016~2021년)의 예산을 집중 투자하고 있다(Our plan for growth : science and innovation, 2014). 특히 계산과학을 기반으로 하는 모델링과 정보학 연구가 활발히 진행되고 있다. Horizon 2020(2014~2020년)의 NMP 프로그램(Nanoscience, Nanotechnology, Materials and New Production technologies), 유럽 재료 모델링 위원회(European Materials Modeling Council, 2014년 설립) 등을 중심으로 소재 모델링 분야를 연구하고 있다. 2014년부터 Horizon 2020의 지원을 받은 연구 프로젝트로부터 생산된 연구 성과물 공개를 의무화하고 있다. 유럽의 소재 연구데이터 인프라로는 계산재료과학의 주요 소프트웨어 입출력 파일을 수집·관리하는 NoMAD(Novel Materials Discovery) 프로젝트가 대표적이다. NoMAD는 저장소(Repository), 아카이브(Archive), 소재 데이터베이스(Encyclopedia), 빅데이터 분석 도구(Big-Data Analytics), 첨단 그래픽 도구(Advanced Graphics), 고성능 컴퓨팅 서비스 및 인프라(HPC (High-Performance Computing) Services and Infrastructures) 등으로 구성(440여만 건 데이터 구축)되어 있다. 또한 과학데이터 저널을 중심으로 데이터를 활용한 논문 발표가 확산되고 있다. 예로써 Nature Publishing Group(영국)의 저널인 Scientific Data Journal(2014년~)과 Elsevier(네덜란드)의 Data in Brief(2014년~)는 오픈 액세스(Open Access, 법적, 경제적, 기술적 장벽 없이 인터넷을 통해 누구나 무료로 정보에 접근, 활용할 수 있도록 저작물 생산자와 이용자가 정보를 공유) 저널 운영 및 데이터 논문 게재를 중점적으로 지원하고 있다.

표 5. 주요 연구 그룹 및 데이터베이스 현황

명칭	분야 & 특성	데이터베이스 크기
ICSD (Inorganic Crystal Structure Database)	<ul style="list-style-type: none"> • 실험에 기반 한 inorganic compounds 결정구조(187,000개 이상) 보유 • DFT기반 제일원리계산 접목 가능 	 [독일] FIZ, https://www.fiz-karlsruhe.de/en/leistungen/kristallographie/icsd.html
NoMAD COE (Novel Materials Discovery)	<ul style="list-style-type: none"> • 소재물성 데이터, 빅데이터 분석, 가시화 등 다양한 기능 • 계산연구그룹(8개)과 고성능컴퓨팅센터(4개)가 참여 중 	 [EU] CoE, http://nomad-coe.eu
CSD (Cambridge Structural Centre)	<ul style="list-style-type: none"> • 실험적으로 결정된 inorganic, metal-inorganic 결정 구조의 repository • 약 80만개 이상 데이터 보유 	 [영국] Cambridge Crystallography Data Centre, https://www.ccdc.cam.ac.uk

출처 : 국가연구데이터 공유활용 체제 구축 사업 기획보고서(2019)

중국은 데이터 수집·공유 강화 및 과학자원 시스템 통합을 추진 중이다. 1986년부터 중국과학원의 과학 데이터베이스(SDB, Science Database)를 통해 데이터 수집·관리를 시작하고 중국 과학원을 중심으로 연구데이터를 활용한 소재 개발 연구를 추진하고 있다. 재밌는 사실은 미국의 소재개놈이니셔티브(MGI) 발표에 영향을 받아 중국 과학원과 공학원이 2014년 중국판 소재개놈이니셔티브(MGI) 정책 및 소재정보학을 추진하고 있다는 점이다. 국가자연과학기금을 통해 열전재료의 개발에 소재 연구데이터 적용을 위한 조합실험법을 이용하여 재료개발 가속화를 추진하고, 중국과학기술부는 소재 데이터베이스, 재료개발 툴(Tool) 구축 지원을 위해 '중국제조 2025'를 통해 'Materials Genome 공정기술 개발'을 중점적으로 지원하고 있다. 특히, 중국의 각 대학기관은 미국의 버클리·연구기관에서 경험이 있는 중국 과학자들을 적극 유치하여 상해대학 MGI, 닝보 MGI 연구소 등 소재정보학 연구소를 설립하여 운영 중이다.

3. 데이터 수집 및 활용을 위한 규정

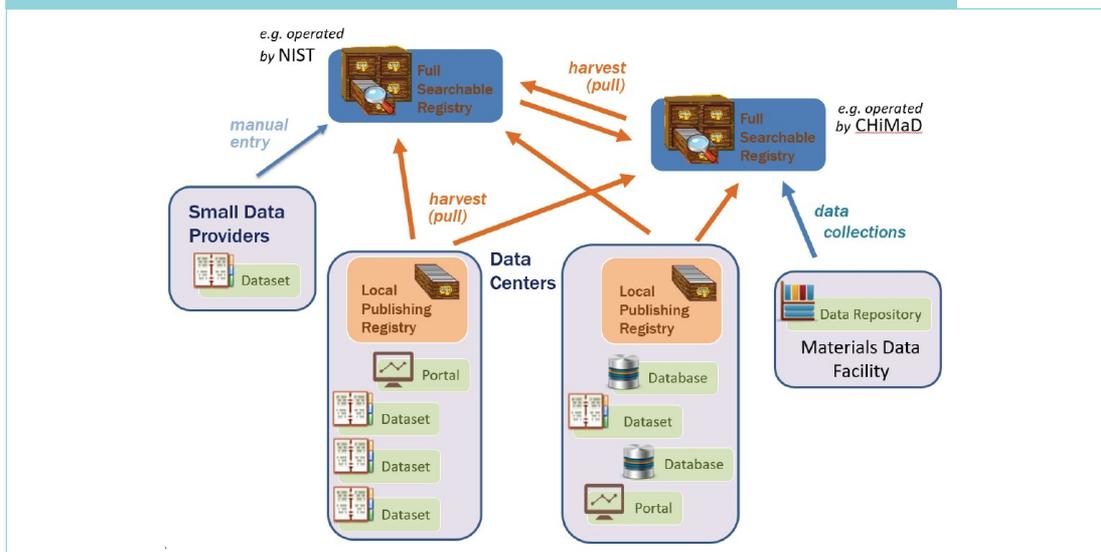
3.1. 소재 연구데이터 표준 프로토콜

앞서 살펴본 바와 같이 전 세계적으로 여러 기관에서 소재 연구데이터의 수집·활용 플랫폼 인프라 구축이 진행되고 있다. 그러나 각 나라별로, 각 기관별로 구축되고 있는 소재 연구데이터 플랫폼 데이터베이스 간의 연동이 불가능하다면, 진정한 의미에서의 빅데이터라고 할 수 있을까? 이러한 의문에서 출발하여, 플랫폼

인프라 데이터베이스 간의 연동성 및 확장성을 담보하기 위하여 범세계적으로 다양한 시도가 이루어지고 있다. 다양한 정부조직, 국제기구, 대학 등이 참여한 커뮤니티 형태의 비정부 기구가 조직되어 연구데이터 인프라, 정책, 규격 개발 등을 주도하고 있다. 특히, 2013년 연구데이터 연맹(Research Data Alliance, 이하 RDA)이 발족되어, 137개 국가, 11,732명의 회원이 참여하여 연구데이터 공유·교환을 촉진하고 데이터 기반 R&D 혁신을 가속화하기 위한 표준 프로토콜 개발 및 커뮤니티 활동을 수행하고 있다.

RDA 기구에서는 각 이슈별 표준 프로토콜 권고안을 개발하는 다양한 작업그룹(Working Group)을 운영하고 있으며, 현재 약 95여 개의 작업그룹이 활성화되어 있다. 각 작업그룹에서는 작업그룹에 참여한 회원들 간의 의사교환을 통하여 목표로 하는 표준 프로토콜 이슈의 권고안을 보고하고 있다. 데이터 종류의 저장을 위한 영구적인 식별자 표준 프로토콜부터 연구데이터 수집 과정의 API 표준화까지 연구데이터 수집·활용과 관련된 다방면의 표준 프로토콜을 개발하여 권고하고 있다. 그 중 International Materials Resource Registries (IMRR) Working Group은 소재 연구데이터 플랫폼 간의 연동성 및 확장성을 담보한 레지스트리 연합 프레임워크 (Federated Registry Framework)를 제안하였다. 이 프레임워크는 소재 연구데이터 플랫폼 간의 데이터를 상호 검색·교환을 수행할 수 있는 그림을 제시하였으며, 제안한 프레임워크를 실현할 수 있는 소재 데이터 저장소 오픈소스 소프트웨어를 개발하여 제공하고 있다.

그림 7. IMRR Working Group이 권고한 Federated Registry Framework



출처 : IMRR WG Final Report and Recommendations

3.2. 연구데이터 수집·활용 규정 및 규약

소재 연구데이터를 수집·활용하기 위한 인프라, 표준규격 등을 개발 및 구축하는 각국의 노력과 별개로, 개별 연구자의 데이터를 수집·활용하는 과정에서 발생할 수 있는 지적재산권의 보호와 동시에 오픈사이언스 생태계 형성을 위한 데이터 공개 장려 정책 제정 또한 다방면으로 이루어지고 있다. 경제협력개발기구(OECD)는 2007년 연구데이터 오픈 액세스(Open Access) 원칙 및 가이드라인을 공표한 이후, G7 장관회의를 비롯한 여러 형태의 회의를 통하여 오픈사이언스 관련 성명 발표, 합의기구 설치 등 다방면으로 오픈사이언스 생태계 구축에 노력을 기울이고 있다. 유럽은 Horizon 2020 프로젝트의 일환으로 FAIR(Findable, Accessible, Interoperable, Reusable) 데이터 관리 가이드라인을 공표하여 연구데이터 관련 규약을 제시한 바 있다. 또한, Horizon 2020 프로젝트의 지원에 의하여 발표한 논문 및 유관 연구데이터의 오픈 액세스(Open Access)를 의무화함으로써 오픈사이언스 클라우드를 구축하고 있다. 미국의 경우, 미국국립과학재단(NSF, National Science Foundation)의 지원을 받는 모든 연구과제에 대하여 연구기획 단계에서 데이터관리계획(DMP, Data Management Plan) 제출을 의무화하였으며 연구종료 후 12개월 이후 연구성과 데이터를 공개하도록 정책화하였다.

우리나라도 여러 연구데이터 수집·관리 체계 인프라 구축개발이 현실화됨에 따라 이에 발맞추어 과학기술정보통신부를 중심으로 연구데이터 공유·활용 전략안을 제시하였다. 또한, 연구데이터의 체계적인 관리를 위하여 '국가연구개발사업의 관리 등에 관한 규정'을 개정함으로써 연구개발계획서 작성 시 데이터관리계획(DMP)을 수립 및 제시를 의무화하였다. 과학기술정보통신부 산하 기관(한국연구재단, 정보통신기획평가원)을 중심으로 일부 과제에 데이터관리계획(DMP)을 시범 적용하고 있다. 더 나아가, 최근 소재 연구데이터 플랫폼 구축방안을 수립, 발표하여 과학기술정보통신부 소재 R&D를 수행하는 모든 과제에 대하여 데이터관리계획(DMP)의 제출 의무를 확대·적용하는 한편 현재 구축되고 있는 '국가 소재 연구데이터 플랫폼'에 연구성과물 등록을 의무화할 것을 공표하였다.

표 6. 각국의 오픈사이언스 및 연구데이터 정책 현황

국가	주요정책
미국	<ul style="list-style-type: none"> 정보자유법(1966), 미국경쟁력 강화법(2010), 백악관 '연구데이터 관리 및 공유에 대한 지침'(2013, OSTP)으로 공공데이터의 오픈 액세스 원칙 제시 공적자금이 투입된 연구데이터는 연구기관별(1억달러 이상) 관리·공개*, 연구기획단계에서 데이터관리계획(DMP) 제출 의무화
영국	<ul style="list-style-type: none"> 정보자유법에서 공개정보의 범위에 데이터셋을 포함시켜 연구데이터를 포함한 광의의 데이터를 공개 대상으로 규정
독일	<ul style="list-style-type: none"> 2010년 출판법(Copyright Act, UrhG)을 개정하여, 공적자금으로 지원된 연구성과에 대해 연구자가 출판사에 출판권을 이전한 경우에도 일정기간의 유예기간 경과 후 비상업적·교육적 용도로 2차 출판하는 것을 인정
프랑스	<ul style="list-style-type: none"> '16년 디지털공화국법(Law No. 2016-1321 on a Digital Republic)을 통해, 공공연구에서 생산된 논문을 비상업적 용도로 배포 권리를 부여
호주	<ul style="list-style-type: none"> 호주 국가연구위원회(ARC, Australian Research Council)의 재정지원을 받는 연구기관에 대하여 의무적으로 데이터 관리계획 제출 규정(2014)
일본	<ul style="list-style-type: none"> 과학기술진흥기구(JST, Japan Science and Technology Agency)는 '13년 '공공자금으로 지원된 연구의 성과에 관한 오픈액세스 방침'을 공표, '17년 공공연구로 생산된 연구데이터에 대한 데이터관리계획(DMP) 수립을 권고하는 오픈사이언스정책 발표

출처 : 미래소재 연구데이터 플랫폼 구축사업 보고서(2019)

III 소재 연구데이터 수집·활용 전략

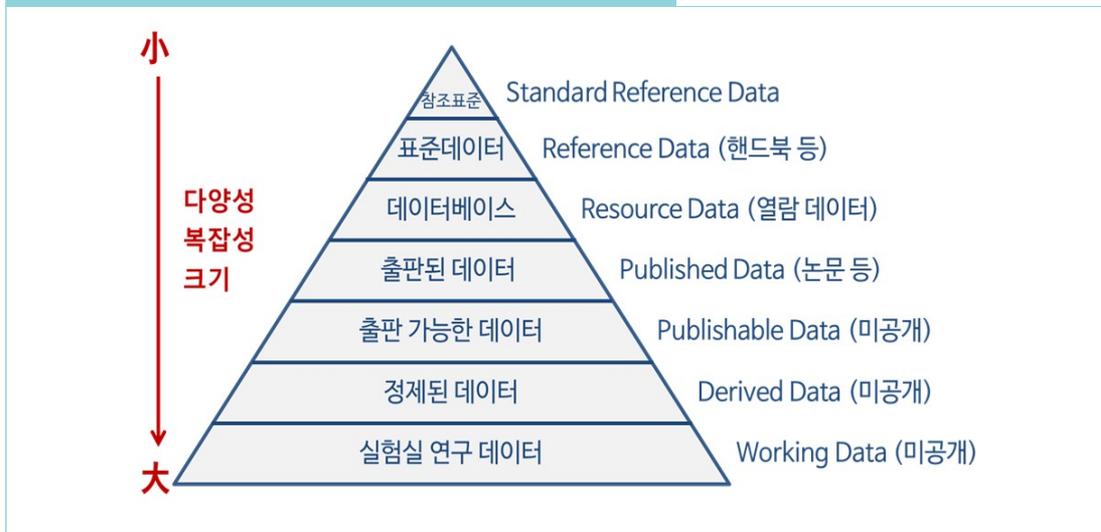
1. 소재 연구데이터 정의 및 특징

앞서 살펴본 바와 같이, 4차 산업혁명 시대 진입에 따른 연구개발 패러다임의 변화로 연구데이터 수집·활용 역량이 과학기술 경쟁력의 핵심으로 부각되고 있다. 이러한 시대적 요구에 따라 연구데이터의 재사용을 향상과 불필요한 연구를 최소화하여 연구개발 사업 효율성과 경쟁력을 향상시키기 위하여 세계 각국은 다양한 정책을 펼치고 있다. 그런데 각국들이 모으고자 하는 연구데이터는 모두 동일한 데이터를 지칭하고 있는 것일까? 세계 각국은 공통적으로 연구데이터를 ‘연구개발 과정 혹은 그 결과로 생산되는 모든 종류의 데이터’로 정의하고 있다. 그러나 이를 수집하는 관점에서 국가별로 시각차가 존재한다. 거시적으로 볼 때, ‘모든’ 연구데이터를 수집할 것인지 또는 ‘결과성산물’의 연구데이터만을 수집할 것인지에 따라 차이가 있다. 다시 말해, 연구데이터 수집·활용 인프라 구축 및 유관 정책 개발의 초기 단계에서 각국은 연구개발 결과를 검증하기 위한 ‘성공한 데이터만’을 수집할 것인지, 또는 ‘실패한 데이터를 포함한 연구개발 과정의 모든 데이터’를 수집할 것인지를 선택지에 놓여 있는 것이다. 데이터관리계획(DMP) 권고 혹은 정책을 고려할 때, 대부분의 국가에서는 현재, 전자의 수집전략을 기반으로 한 데이터 기반 R&D 생태계 형성에 주력하고 있으나, 후자의 수집전략을 최종목표로 하고 있다. 우리나라 또한, 각 기관에서 개발되고 있는 연구데이터 플랫폼 인프라 구축 현황을 볼 때, 대체적으로 실패한 데이터를 포함한 모든 종류의 데이터 수집 목표로 하고 있다. 이러한 관점에서 소재 연구데이터를 조금 더 세분화하여 다음과 같이 정의할 수 있다.

- 소재 연구개발 전주기 과정에서 실험과 계산으로 생산되는 모든 정형·비정형 데이터, 문헌 및 인터넷을 통하여 공개된 소재 데이터를 모두 포함한다.
- 소재 연구데이터는 합성·제조 과정에서의 합성방법, 조건 등의 데이터, 측정·분석에 의하여 생산된 스펙트럼, 이미지, 동영상 등의 데이터, 계산에 의하여 생산된 물리·화학·기계적 물성을 모두 포함한다.
- 연구개발 과정에서 목표물성에 도달하지 못했거나, 예측된 물성을 가진 소재 합성 및 제조에 성공하지 못한, 실패한 데이터도 연구데이터로 포함한다.

연구결과 성과물의 연구데이터(정제된 혹은 출판에 활용된 데이터)로부터 실험실에서 생산되는 모든 종류의 데이터로 확장하면, 수집해야할 데이터의 다양성, 복잡성 및 크기가 상대적으로 커지게 된다. 반대로 데이터를 정제하면 할수록 그것의 양이 줄어드는 것은 당연한 결과이다. 또한, 소재 연구개발 과정에서 측정결과데이터와 계산물성데이터를 동시에 고려하여 정제된 단일의 해석결과를 도출하는 과정을 상기할 때, 데이터 정제에 따른 다양성 및 복잡성이 줄어드는 것도 이해할 수 있게 된다. 수집대상 연구데이터 정의를 연구개발 전주기 과정의 모든 데이터로 확장·적용하였을 경우 다양성·복잡성·크기가 증가하게 된다. 따라서, 활용 가능한 소재 연구데이터 수집을 위해서는 연구데이터의 다양성, 정확성 등에 대한 깊은 고찰이 요구된다.

그림 8. 연구데이터 정의의 범위에 따른 데이터 특징



출처 : 국가연구데이터 공유활용 체제 구축 사업 기획보고서(2019)

2. 소재 연구데이터 종류

관점에 따라 소재 연구데이터의 종류를 분류하는 기준은 천차만별이다. 예를 들어, 연구데이터의 출처가 수집에 있어서 주요 관심요소인 경우 각 데이터의 생산 출처를 기준으로 연구데이터의 종류를 분류한다. 한국과학기술정보연구원(KISTI) 데이터센터는 연구데이터의 종류를 실험데이터, 관측데이터, 계산데이터뿐만 아니라 참조데이터, 조사데이터 등 다양한 출처에 대하여 분류 기준을 제시한 바 있다. 특이한 점은, 가속기·화학 실험데이터와 전자현미경데이터를 서로 다른 종류로 분류하고, 텍스트마이닝에 의하여 추출한 물성데이터와

계산 혹은 평가에 의한 물성데이터를 서로 다르게 분류한다는 점이다. 원천출처에 기반한 분류기준이 적용되었기 때문에 데이터 자체의 의미가 크게 고려되지 않았지만, 데이터의 큐레이션 및 관리 관점에서 출처에 기반한 연구데이터 종류의 분류 기준은 중요한 의미를 갖는다.

소재 연구과정을 기반으로 한 연구데이터 종류의 분류도 가능하다. 소재 연구개발 과정을 상기해보면, ①소재 합성·제조, ②합성한 소재의 측정·분석, ③계산과학 기반 구조·물성 계산, ④앞서 기술한 각 데이터의 재가공을 통한 해석의 4가지로 분류가 가능하다. 이러한 연구데이터 종류의 분류 방법은 앞서 살펴보았던 출처에 기반한 분류와는 다른 활용 가능성·방향성을 내포하고 있다. 예를 들어, 소재 물성 향상을 위하여 소재 합성·공정의 최적화를 이루고 싶다면, 물성과 관련된 측정·분석·계산데이터와 실험·공정데이터 간의 상관관계를 살펴봄으로써 그러한 목표에 달성할 수 있을 것이다. 즉, 공정(Process)-구조(Structure)-특성(Property)-성능(Performance) 간의 상관관계 도출을 위한 적합한 분류 기준이다. 또는 측정·분석데이터와 이론을 기반으로 한 계산데이터 간의 상관관계를 고찰함으로써, 측정·분석의 정확도 향상을 위한 측정·분석조건 탐색을 하거나, 계산의 정확도 향상을 위한 다양한 계산 상의 변수 탐색을 할 수도 있다. 앞서 기술한 데이터 출처에 기반한 연구데이터 종류의 분류가 데이터 관리 등의 IT 관점에 가까운 분류 방식이라면, 이러한 소재 연구과정에 기반한 분류 방식은 조금 더 소재 연구의 관점에 가까운 방식으로 볼 수 있다. 하지만, 서로 간의 관점이 다르다고 해서 하나만의 분류 방식으로 데이터 종류를 분류할 필요는 없다. 오히려 매트릭스 구조 형태로 연구데이터 종류를 다방면으로 분류한다면, IT 관점에서 데이터 관리의 효율성을 확보함과 동시에 소재 연구 관점에서 활용의 가치를 높일 수 있게 된다.

표 7. 소재 연구데이터 종류 분류의 예

종 류	내 용
실험·공정데이터	소재를 합성·제조하는 과정의 모든 데이터 (예) 합성 원물질, 합성법, 합성조건 등
측정·분석데이터	소재 성능·물성 측정 및 분석과정의 모든 데이터 (예) 응용분야별 물성측정, 구조·화학분석 등
계산데이터	계산과학 기반 시뮬레이션 통해 생산되는 데이터 (예) 계산 입·출력데이터, 계산결과 소재물성데이터 등
해석데이터	각 데이터의 재가공을 통하여 도출된 해석결과 데이터 (예) 현미경 관찰을 통해 도출된 입자크기 값, 분포 등

출처 : 미래소재 연구데이터 플랫폼 구축사업 보고서(2019)

3. 소재 연구데이터의 표준화

수집대상 연구데이터의 다양성 및 복잡성은 데이터 과학 관점에서 빅데이터화 되었을 때 높은 가치 창출로 연결될 수 있는 긍정적인 속성이지만, 수집기술 개발의 난이도가 과도하게 높아진다는 부정적 요소가 숨어있다. 다양하고 복잡한 데이터를 체계적으로 수집하기 위해서는 앞서 다양한 기준으로 분류한 데이터의 종류에 따라 특화하여 수집전략을 세울 필요가 있다. 당연히 데이터가 다양할수록 분류체계가 복잡해질 수밖에 없으며 각 분류에 맞는 수집기술 개발에 어려움이 존재할 수밖에 없는 것이다. 소재 연구데이터 수집기술 개발의 어려움은 비단 앞서 분류한 데이터의 다양한 종류에만 있지 않다. 조금 더 구체적으로 들어가, 같은 종류의 데이터임에도 불구하고 데이터의 출처, 장비, 형식 등의 상이성에서 기인한 수집기술 개발의 난이도 상승이 존재한다. 예를 들어, 똑같은 XRD 데이터로 분류됨에도 불구하고 서로 다른 제조사의 XRD 장비에서 생산된 데이터 간에는 측정·분석조건, 데이터 파일 형식 등의 차이가 존재한다. 데이터의 의미 관점에서 동일한 데이터로 수집해야함에도 불구하고, 장비의 종류별·제조사별로 수집기술을 개발해야하기 때문에 소재 연구데이터의 수집기술 개발에 어려움이 증가하게 된다. 서로 다른 장비 혹은 소프트웨어에서 생산되었다고 하더라도 동일한 의미를 갖는 데이터라면 통일된 내용·형식으로 수집될 수 있도록 데이터의 전처리 표준화 기술 개발이 요구된다. 이 표준화 과정은 데이터의 출처에 따라 의미상으로 동일한, 하지만 형식상으로 서로 다른 메타정보를 통일시키는 의미로의 표준화와 데이터 형식의 상이성(예를 들어, JPG와 PNG의 관계)을 통일시키는 의미로의 표준화 모두를 포함한다.

3.1. 실험·공정데이터 표준화

소재 합성·제조 과정에서 생산된 모든 연구데이터를 지칭하는 실험·공정데이터는 합성 시료, 합성법, 합성조건 등의 데이터가 주를 이루고 있다. 또한, 실험 과정에서 발생하는 사건 혹은 관찰되는 현상들은 특정한 틀에 정형화할 수 없는 경우가 상당 부분 존재한다. 그래서 보통 수기로 혹은 최근에는 전자기기를 활용하여 연구노트에 이런 다양한 비정형의 실험·공정 정보들을 입력한다. 광범위한 소재 응용분야와 소재 그 자체, 그리고 그 소재들을 합성·제조하는 과정의 다양성을 고려할 때, 합성·제조하는 과정에서 발생하는 비정형 연구데이터를 모두 정형화시키는 것은 불가능에 가깝다. 연구데이터를 수집하는 과정에서 데이터의 과도한 정형화는 수집과정의 불편함이 증폭되어 오히려 연구데이터 수집이 이루어지지 않을 가능성이 높아질 뿐이다. 따라서, 소재 합성·제조 과정의 데이터 중 정형화시켰을 때 얻을 수 있는 이점이 높은 데이터들을 정형화하여 수집하며, 나머지 비정형의 정보들은 수기 또는 전자기기로 연구노트에 자유롭게 입력하여 추후 문자인식 및 자연어처리 기술이 발달함에

따라 이러한 비정형데이터를 활용할 수 있는 여지를 남겨두는 것이 좋다. 한국과학기술연구원(KIST)에서 개발·제공되고 있는 KiRI 전자연구노트는 이러한 소재 실험·공정데이터의 비정형데이터를 수집하는, 좋은 예시의 수집전략 기술이다. 특히 KiRI 전자연구노트는 사람의 손으로 기록한 수기데이터를 광학식 문자 판독(OCR, Optical Character Recognition) 등의 기술로 문자인식하여 자동으로 디지털 텍스트로 변환하는 기능도 제공하고 있다. 문자인식기술의 고도화와 더불어 자연어처리 기술이 발전하면, 이러한 필기 기록들로부터 정형화된 데이터들을 추출하게 될 날이 곧 올 것이라 예견된다.

한편, 소재 실험·공정데이터의 정형성을 극대화하기 위해서, 소재 합성·제조와 관련된 어떤 정보들을 정형화 혹은 표준화시킬 수 있으며, 해당 데이터들을 어떤 방법으로 수집할 수 있을까? 실험·공정데이터의 표준화는 바로 이러한 부분에 초점을 맞추고 있다. 데이터는 소재 응용분야별로 다양할 뿐만 아니라, 같은 응용분야의 동일한 소재라고 하더라도 다양한 합성시료·합성법·조건 등이 있을 수 있다. 그러므로 연구자·연구그룹별로 서로 상이할 수 있는 다양한 합성순서·조건에 대응할 수 있는 유연한 구조의 표준화 방식이 필요하다. 소재 합성·제조 과정을 상기해보면, 어떠한 순차적인 혹은 병렬적인 과정을 통하여 소재를 합성 및 제조하고 있음을 알 수 있다. 이런 순차·병렬 구조가 복잡하게 얽혀있는 흐름도를 표준화하고 유연하게 설계할 수 있는 시스템은 이미 우리가 잘 알고 있는 것이 하나 있다. 바로 제어시스템 설계 등에 활용되는 그래픽 기반 프로그래밍 툴인 랩뷰(LabVIEW)이다. 랩뷰와 같이, 소재의 합성·제조 과정에서 자주 활용되는 합성공정들을 표준화하고 이를 그래픽 함수화하면 각 연구자별, 연구그룹별로 각자가 소재 합성·제조 과정을 회로화할 수 있게 되고, 이를 기반으로 정형화된 실험·공정데이터를 수집한다면 소재를 합성·제조하는 과정의 많은 영역의 데이터들을 정형화하여 수집할 수 있게 될 것이다.

3.2. 측정·분석데이터 표준화

합성한 소재가 원하는 물성을 발현하는지 또는 어떤 구조를 갖는지를 확인하기 위하여 측정·분석을 수행한다. 장비 등을 통하여 소재를 측정·분석하는 과정에서 발생하는 모든 종류의 데이터를 측정·분석데이터라 칭한다면, 소재 실험·공정데이터와는 달리 다행히도 측정·분석데이터는 거의 정형화되어 있는 데이터 밖에 없다. 그도 당연한 것이, 측정·분석하기 위하여 장치 혹은 장비로 입력하는 값들은 정형화되어 있을 수밖에 없으며, 측정·분석의 결과물 또한 마찬가지이다. 따라서 측정·분석데이터는 실험·공정데이터에 비해 상대적으로 쉽게 수집할 수 있을 것으로 보인다. 그러나 측정·분석데이터의 경우에는 동일한 종류의 측정·분석데이터를 얻기 위하여 서로 다른 제조사의 측정·분석 장치, 장비를 활용하는 경우가 있어 이에 대한 표준화 기술 실현이 필요하다.

한국과학기술연구원(KIST)에서 개발한 KiRI 플랫폼 시스템과 개발 중인 소재 연구데이터 수집·관리·활용 플랫폼 내에서 측정·분석의뢰를 간편하게 수행하고, 측정·분석에 대한 결과물 데이터를 효율적으로 수집할 수 있도록 한국과학기술연구원(KIST)에서 보유 중인 65종, 76개 분석의 장비에 대한 전수조사를 통해 측정·분석 의뢰과정의 메타정보를 표준화한 바 있다. 측정·분석 의뢰과정의 표준화는 측정·분석 샘플의 소재 정보(상, 원소, 구조 등)를 측정·분석 조건 및 결과와 연결시켜줌으로써 측정·분석결과를 소재 합성·제조의 실험·공정 데이터와 연결 구조를 만들어준다.

그림 9. XRD 분석의뢰 메타정보 표준화 결과물

의뢰자 작성란

1. 시료 정보 (Sample Information)

1-1. 상태 (Phase) 박막 (두께:) 벌크 파우더 고분자막 기타 1-2.

시료 조성 원소 (Composition)

1-3. 시료 구성 예상 화합물

1-4. 시료 환수 여부 : 네 아니오

1-5. 구조 및 두께 (Structure of sample) (파일 첨부, 직접 그림 가능)

ex) Ti 10nm / SiO₂ 100nm / Si substrate ex) pattern 있을 경우 SEM 사진 등.

2. 분석 모드 (Analysis mode)

2-1. 일반분석 Theta/2Theta scan : 파우더, 벌크
 2theta scan (incident angle: () deg) : 박막

2-2. 특수분석
 Rietveld Refinement 5~140deg, Step scan
 이 실험은 원자간 거리 결정격자 미지구조분석 기타

표준 XRD
 (온도 구간: () ~ () °C, 온도유지시간 () min, 승온속도 () °C/min)
 (기종: Cu, Ni, Ag, Pt, Mo, Fe, Co, Cr, V, Mn, Zn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn, Pb, Bi, Sb, Sn, Te, Se, As, S, P, Si, Ge, Ga, In, Sn,

것이 매우 어렵다. 가끔 각 장비 제조사별로 원시데이터의 독자포맷을 공개하는 경우가 존재하여 공개된 정보에 따라 측정·분석 조건정보 및 실험데이터 등을 추출할 수 있다. 그러나 대부분의 경우에는 각 제조사의 독자규격을 공개하지 않고 있으며, 이러한 경우에는 바이너리로부터 개별 주소값의 데이터가 어떤 의미를 갖는지에 대한 맵핑을 일일이 해줄 필요가 있다. 각 장치·장비 활용 전문가만이 역공학(reverse engineering) 과정을 통하여 제조사의 독자규격 바이너리 원시데이터로부터 측정·분석 조건정보 및 실험데이터 등을 추출할 수 있기 때문에 많은 인력과 시간이 필요하다. TEM 데이터의 가장 대표적인 원시데이터 파일인 DM3, DM4 등은 세계의 많은 연구자들의 노력에 의하여 핵심요소 정보들을 추출할 수 있도록 각 정보별 바이너리 파일 주소값에 대한 맵핑정보가 공개되어있다. 이런 방식으로 각 장치·장비로부터 생산된 원시데이터로부터 측정·분석 조건정보 및 실험데이터 등을 직접 추출할 수 있게 된다면, 데이터들의 수집 과정에서 각 메타정보를 사람이 입력할 필요 없이 자동으로 수집할 수 있게 되며, 또한, 앞서 정의한 장치·장비 제조사별 다양한 독자규격들을 표준화된 양식으로 변환시키기가 쉬워지므로 소재 연구데이터 수집 과정에서 매우 중요한 부분을 차지한다.

그림 10. TEM 원시데이터로부터 메타정보 등을 직접 추출할 수 있는 맵핑정보

```

Example DM4 file

00 00 00 04 00 00 00 00 01 09 58 7e 00 00 00 01
01 00 00 00 00 00 00 00 0f
15 00 11
41 70 70 6c 69 63 61 74 69 6f 6e 42 6f 75 6e 64 73
00 00 00 00 00 00 00 84 25 25 25 25
.....
00 00 00 00 00 00 00 00

Header

00 00 00 04          i4be  DM version = 4
00 00 00 00 01 09 58 7e i8be  file length - 24
00 00 00 01          i4be  byte order, 1 = little endian (PC) order

Root tag directory

01          i1    1 = sorted (normally = 1)
00          i1    0 = closed, 1 = open (normally = 0)
00 00 00 00 00 00 0f i8be  number of tags in root directory (0fh = 15)

Tag in root tag directory

15 00 11 41 70 70 6c ... Tag directories and tags, see below

End of file

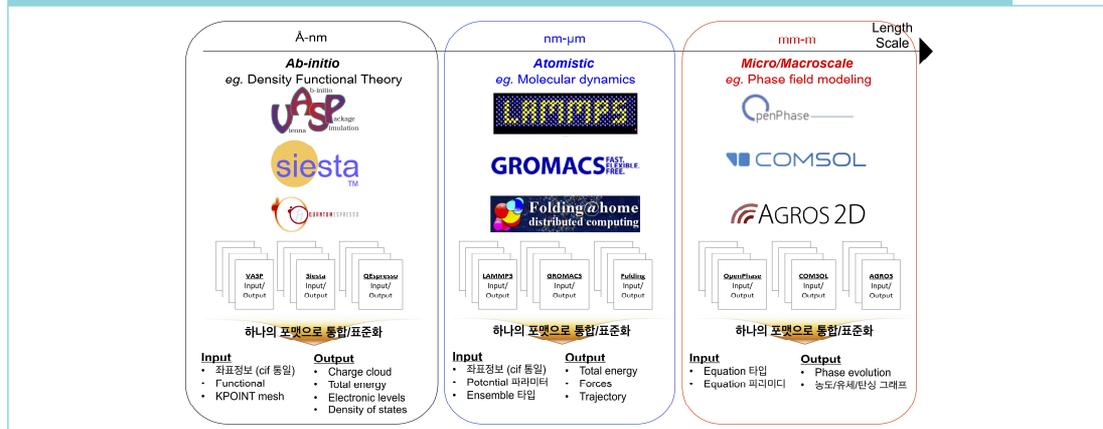
00 00 00 00 00 00 00 00      End of file appears to be marked with 8 nulls
    
```

출처 : Boothroyd 교수 홈페이지

3.3. 계산데이터 표준화

계산과학에 의하여 생산된 데이터는 전 세계적으로 가장 많이 그리고 체계적으로 구축되어 있다. 미국의 Materials Project 및 오픈소스 소재 데이터베이스(OQMD), 유럽의 NOMAD 등 계산과학 데이터베이스는 이미 플랫폼화 되어있어 전 세계의 계산과학 데이터베이스 포털로 활용되고 있다. 계산과학 데이터의 수집은 기 구축되어 있는 데이터베이스에서 이미 밀도범함수이론(DFT, Density Functional Theory)에 기반한 각종 소프트웨어(VASP, Quantum Espresso, Gaussian 등)에 대한 표준화가 이루어져 데이터베이스화 되어 있어 이를 최대한 활용하는 것이 중요하다. 각 계산과학 소프트웨어에서 생산되는 데이터는 입력·출력데이터로 이루어지며, 입력정보(원자 좌표, 사용한 포테셜 정보 등)가 서로 동일한 의미를 지니더라도 계산과학 소프트웨어 간의 형식이 상이하며 결과정보 또한 마찬가지로이다. 이렇게 서로 다른 표현에 의하여 묘사된 정보들을 체계적으로 수집하기 위해서는 각 정보들에 대한 표준화가 이루어져야 체계적인 데이터 수집이 가능할 것이다. 다양한 계산과학 데이터 전·후처리 소프트웨어들이 개발됨에 따라 동일한 의미를 갖는 데이터의 통일된 명칭을 부여하여 표준규격을 제시하려는 시도들이 많이 이루어지고 있다. 예를 들어, 파이썬(Python)으로 작성된 Atomic Simulation Environment 패키지는 다양한 밀도범함수이론(DFT) 계산의 입·출력에 대한 데이터처리를 통하여 각 정보를 표준 규격화하여 계산 소프트웨어의 종류와 무관하게 활용할 수 있도록 다양한 래퍼 함수(wrapping function)를 제공하고 있다. 다만, 기 구축되어 있는 다양한 계산 데이터베이스 관련 기관과의 공조를 통한 표준화 연구개발이 필요하다.

그림 11. 다양한 스케일 및 방법론으로 생성된 계산데이터들의 통합 표준화 모식도



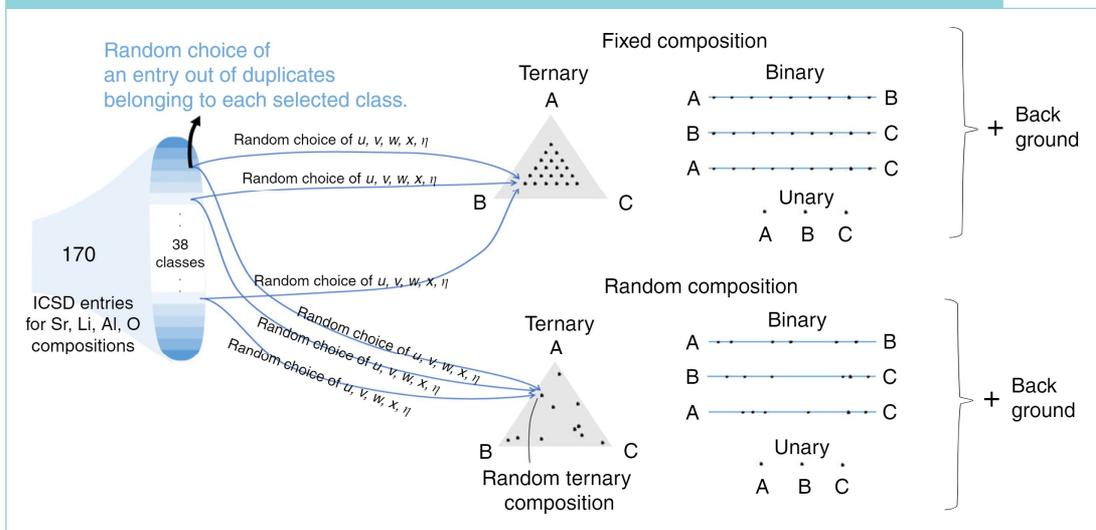
출처 : 한국과학기술연구원(2018)

3.4. 해석데이터 표준화

소재 연구개발 과정에서 소재 물성과 매우 밀접하게 연관되어 있음에도 불구하고 실험·공정, 측정·분석, 계산데이터에서 직접적으로 추출할 수 없는 데이터가 있다. 예를 들면, 합성한 소재의 입자크기, 분포 등은 일반적으로 숫자로 표현되지만 이 숫자들은 실험·공정, 측정·분석데이터에서 직접적으로 도출되지 않아 사람이 샘플의 미세조직 사진으로부터 입자의 크기 및 분포를 계산하는 경우가 대부분이다. 다시 말해, 이는 측정·분석 데이터를 해석한 결과값인 것이다. 이러한 데이터들을 해석데이터라 하며, 주로 측정·분석 및 계산데이터의 결과물을 사람이 가공하거나 측정하여 추출하는 경우가 많다. 특히, 스펙트럼 혹은 이미지로 표현된 측정·분석데이터로부터 소재 연구자가 원하는 해석값을 추출하는 것은 사람의 손을 거치지 않고서는 어려운 일이다. 따라서, 소재 연구데이터의 빅데이터 구축을 위해서는 각 측정·분석의 결과물로부터 사람의 손을 거치지 않고 해석데이터를 추출하기 위한 새로운 기술이 필요하다. 예를 들면, TEM 이미지로부터 입자형태로 나타나는 상(phase)의 크기를 자동으로 계산하여 평균적인 입자크기 값을 추출하거나, XPS 스펙트럼으로부터 각 원소 조성의 분율을 자동으로 추출할 수 있다면, 측정·분석 후 사람의 손을 거치지 않고 해석데이터를 쉽게 수집할 수 있어 소재 연구데이터의 빅데이터화를 가속화 할 수 있을 것이다.

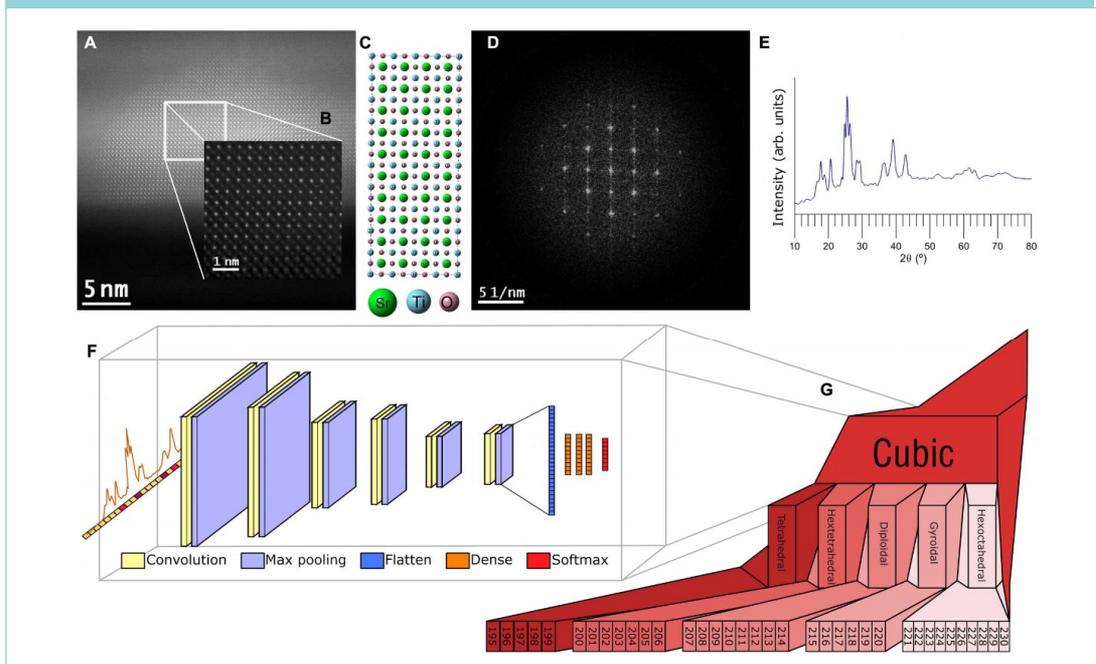
최근 인공지능 신경망의 발달로, 측정·분석 연구분야에서도 인공지능 기술을 활용하여, 사람의 손을 거치지 않고 해석데이터를 추출·생산할 수 있는 새로운 기술에 대한 연구개발이 진행 중이다. 세종대학교의 손기선 교수 연구그룹은 실험에 의하여 추출된 소재 구조정보 데이터베이스인 Inorganic Crystal Structure Database(ICSD)를 활용하여 XRD 스펙트럼을 생산하였고, 이를 합성곱 신경망으로 학습하여 순물질 소재의 공간군을 자동으로 판별할 수 있는 모델을 개발하였다. 또한 최근에는 4개의 상(phase)이 다양한 분율로 존재하는 시스템을 전산모사하여, 복합상이 존재하는 XRD 스펙트럼으로부터 각 상의 분율을 자동으로 계산할 수 있는 모델을 개발함으로써 특정 응용분야(LED 소재)에서 다양한 상분율과 소재 물성 간의 상관관계를 데이터에 기반하여 연구할 수 있는 가능성을 제시하였다. 인공지능 기반 해석데이터 추출 기술 개발은 전자현미경 분야에서도 활발하며, 미국 아이다호 국립연구소의 밀러(Miller) 박사 연구그룹은 TEM 회절패턴 이미지로부터 변환한 스펙트럼을 학습하여 각 순물질 소재의 결정계 및 공간군을 판별할 수 있는 모델을 제시한 바 있다. 이러한 다양한 시도들은 측정·분석데이터의 자동 해석데이터 생산기술로 분류할 수 있는 시도들이며, 인공지능(AI) 기반 해석데이터의 표준화·수집 및 생산기술 개발은 소재 연구데이터의 빅데이터화의 초석이 될 것으로 예견된다.

그림 12. XRD 복합상 스펙트럼에 각 상분율을 자동 계산하는 합성곱 신경망 도식



출처 : Lee(2020)

그림 13. TEM 회절패턴으로부터 결정계 및 공간군을 자동 추출 합성곱 신경망 모델 도식



출처 : Aguiar(2019)

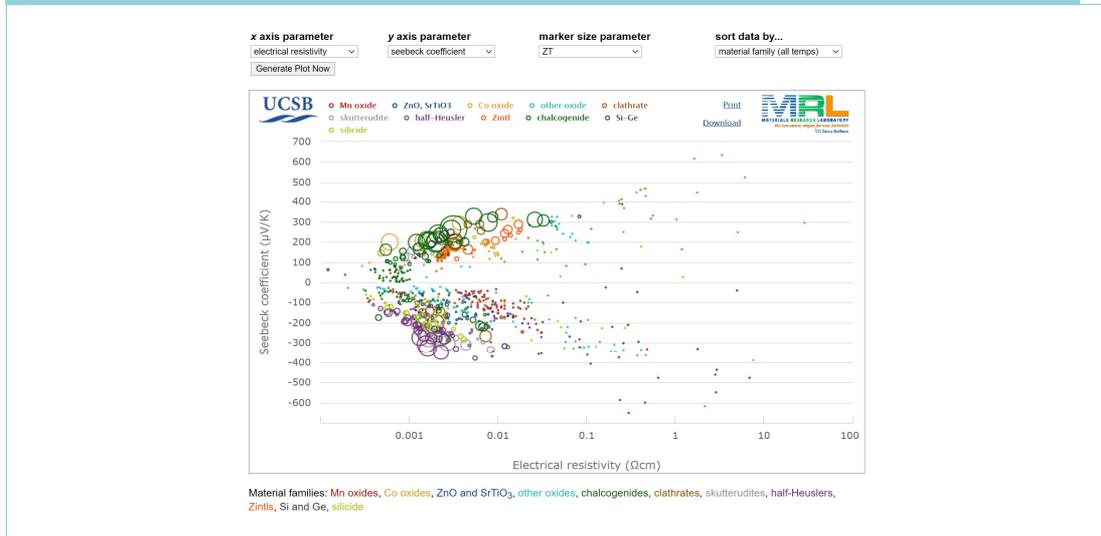
4. 소재 연구데이터 관리·활용

소재 실험·공정, 측정·분석, 계산, 해석데이터를 표준화하여 수집하더라도, 데이터들을 효율적이고 체계적으로 관리 및 활용하지 못한다면 수집하는 의미가 전혀 없을 것이다. 물론 데이터베이스 관리, 분산처리시스템, 하드웨어 등의 IT 관점에서의 데이터 관리 및 활용 관련 문제들도 있지만, 융합연구리뷰에서는 소재 연구 관점에서 연구데이터의 관리 및 활용에 대하여 기술하고자 한다.

연구데이터의 관리는 크게 두 가지로 나눌 수 있다. 연구데이터를 수집하는 과정에서 데이터를 자동으로 분류하고 데이터 간의 연결 구조를 만들어주는 디지털 큐레이션 관점의 관리와 수집한 연구데이터의 검색, 삭제, 통계화 등 개별 연구자, 연구그룹, 기관 등에서 활용할 수 있는 측면에서의 관리가 있다. 수집대상 소재 연구데이터는 데이터 간의 연결관계가 존재하는 복잡한 구조의 데이터이기 때문에, 데이터를 검색할 때 연관 데이터까지 결과를 같이 보여줄 필요가 있다. 뿐만 아니라 이러한 체계적인 데이터 관리 기능은 연구자별, 연구그룹별 소재 연구데이터에 대한 접근성을 높여주어 데이터 관리 과정에서 소재 연구개발의 새로운 단초를 제공할 수 있다. 최근, 전 세계적으로 산업계에서 목표로 삼고 있는 디지털 트랜스포메이션(Digital Transformation)이 실험실·연구실에서도 일어나고 있는 것이다.

또한, 데이터가 수집되어 체계적인 관리시스템 하에 놓이게 되면 이를 다각도에서 활용하는 방법이 개발될 수 있다. 캘리포니아 대학교 샌타바버라(UC Santa Barbara)에서는 열전소재와 관련하여 수집된 각종 물성 정보를 다양한 방식으로 시각화할 수 있게 하여, 각 소재의 물성값 혹은 구조와의 상관관계를 한눈에 확인할 수 있도록 기능을 제공하고 있다. 다양한 시각화를 통한 수집된 연구데이터를 활용하는 것은 시각화 기술개발 과정이 상대적으로 단순할지 몰라도 소재 연구자에게 강한 영감을 줄 수 있는 강력한 툴(Tool)이 될 수 있다. 뿐만 아니라, 한국과학기술연구원(KIST) 내 구현되어있는 KiRI 플랫폼은 촉매 소재와 관련된 데이터를 수집할 수 있는 시스템으로 개발되었으며, 축적되어 있는 데이터베이스의 데이터들을 활용하여, 플랫폼을 사용하는 연구자 스스로 플랫폼 내에서 데이터 간의 상관관계를 도출할 수 있는 기계학습 모델 개발을 수행할 수 있도록 인공지능(AI) 관련 기능을 제공하고 있다. 이렇듯 다양한 데이터 활용기술이 개발되고 있으며, 연구자들이 소재 연구의 새로운 영감을 받을 수 있는 하나의 통로로 사용될 수 있을 것이다.

그림 14. 열전소재의 물성값 간의 상관관계를 한눈에 확인할 수 있는 데이터 시각화의 예



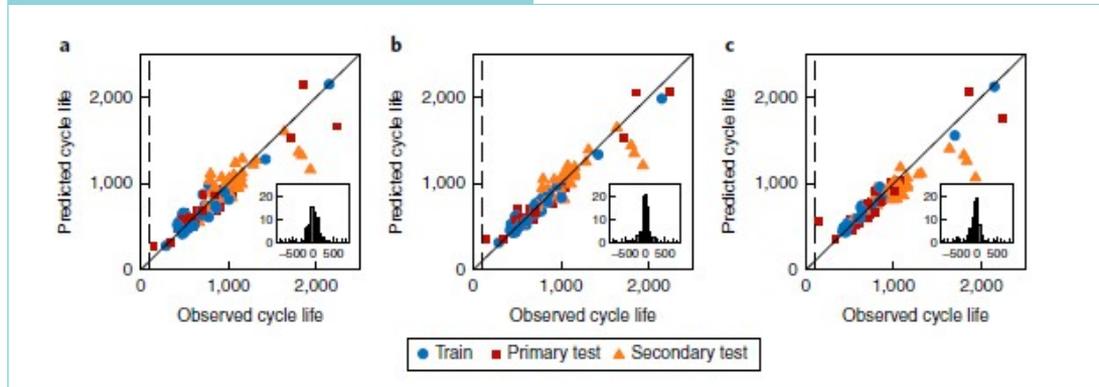
출처 : UCSB MRL 홈페이지

IV 데이터 기반 소재 설계 및 물성 예측

지금까지 기술한 내용을 간략하게 정리하면 다음과 같다. 소재 연구데이터는 실험·공정, 측정·분석, 계산, 그리고 해석데이터로 분류할 수 있으며, 각 데이터의 종류별로 표준화를 수행하여 데이터 전·후처리를 통해 수집하게 된다. 표준화 설계 단계에서부터 데이터 간의 연결구조를 고려하였기 때문에, 데이터는 자연스럽게 디지털 큐레이션을 거쳐 데이터 간의 연결관계를 유지하여 수집된다. 이후 소재 연구자는 데이터의 체계적인 관리와 시각화·기계학습모델 개발을 통하여 이를 활용할 수 있다. 융합연구리뷰에서는 모든 응용분야에 적용할 수 있는 소재 연구데이터의 전반적인 수집·관리·활용 과정에 대해 기술했기 때문에 모든 것이 반영된 실제 사례는 많이 다루지 않았다. 그러나 특정 응용분야의 특정 소재에 데이터 수집·관리·활용기술 개발이 이루어져 활용된 몇 가지 사례들을 소개한다.

최근 전기차 중심으로 에너지저장소재 수요가 급속히 증가함에 따라, 이차전지 등 저장소재의 교체시기를 판단할 수 있는 잔존 수명, 잔존 에너지·용량 등 건전성(State of Health) 판별에 대한 수요가 증가하고 있으며, 데이터를 기반으로 한 에너지저장소재의 건전성을 판단하는 인공지능 모델 개발 연구가 있다. 미국 MIT의 브라츠(Braatz) 교수 연구그룹은 A123 시스템사의 LFP/Graphite 셀을 대상으로 충전 속도에 따른 배터리의 수명을 실험적으로 수집하였다. 다양한 실험조건에 따라 데이터를 수집하였으며, 수집된 데이터 간의 상관관계를 LASSO(Least Absolute Shrinkage and Selection Operator) 및 ElasticNet을 활용하여 선형예측, 제시한 바 있다. 또, 미국 미시건대학교의 스테파노폴루(Stefanopoulou) 교수 연구그룹은 LFP/Graphite와 LMC/Graphite 소재의 수명 예측 모델을 Ridge 회귀분석을 통해 개발한 바 있다. 특히, 이 연구그룹은 소재의 에이징에 따른 수명 변화를 예측하였으며, 인자추출(Feature Extraction)을 통하여 방전용량은 수명과 상관성이 상대적으로 적음을 밝혀낸 바 있다. 마지막으로, 영국 옥스포드 대학교의 Lee 교수 연구그룹은 LCO/Graphite 셀을 대상으로 Electrochemical Impedance Spectroscopy(EIS) 전기계측을 활용하여 배터리 내부의 열화 상태를 평가, 이 측정·분석데이터를 수집하여 배터리 잔존 수명을 예측할 수 있는 기계학습 모델을 개발하였다.

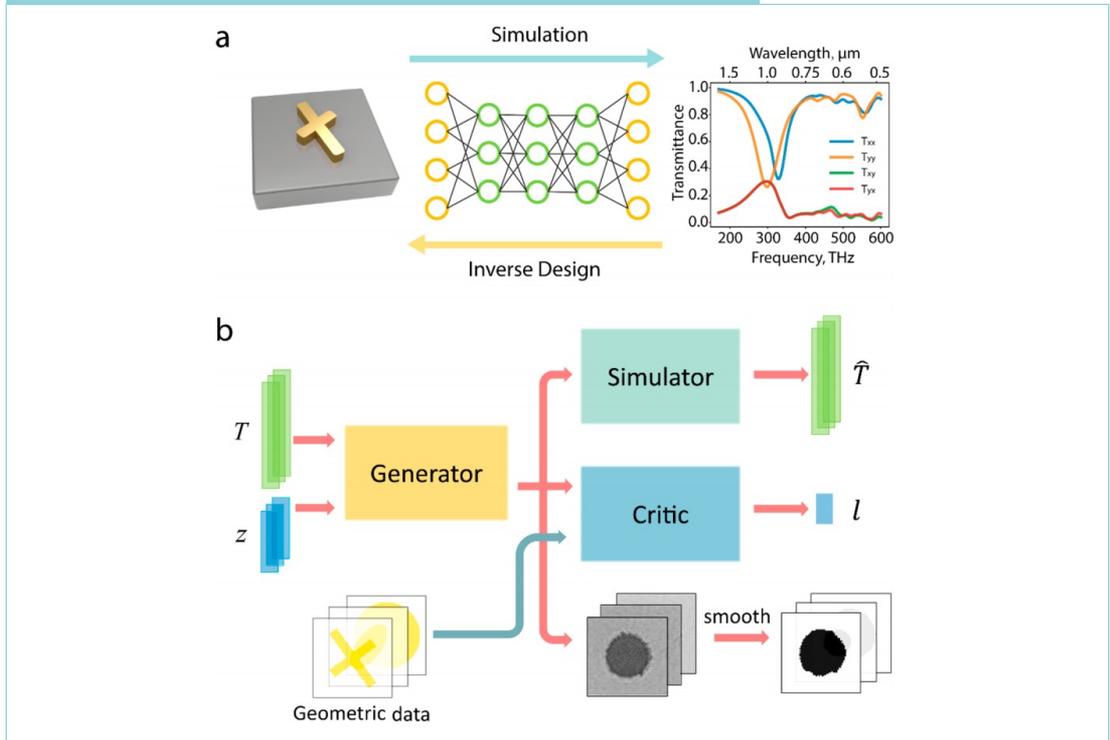
그림 15. 배터리 수명 예측 모델의 성능



출처 : Severson(2019)

수집된 데이터를 활용하여 소재의 물성 예측뿐만 아니라, 원하는 물성값을 갖는 신소재를 설계할 수도 있다. 예를 들어, 삼성종합기술원은 계산데이터를 활용하여 DNN 및 RNN 기반 요구물성값을 갖는 유기소재를 탐색할 수 있는 모델을 개발하여 보고한 바 있다. 또한, 미국 조지아공대 카이(Cai) 교수 연구그룹은 소재 형상에 따라 발생할 수 있는 투과율 등 광학 물성의 변화를 전산모사하여 데이터를 구축한 후, 이를 기반으로 요구물성값을 갖는 소재 형상을 역설계할 수 있는 기계학습 모델을 개발하여 발표하였다. 특히, 이 기술은 기계학습 모델이 단순히 소재의 물성 혹은 원소에 대한 설계뿐만 아니라 가짜 이미지 생성기술로 유명한 생성적 적대 신경망(GAN, Generative Adversarial Network) 기술을 활용하여 소재의 형상 또한 데이터를 기반으로 설계할 수 있음을 보여준 좋은 예이다. 이렇듯 최근 다양한 응용분야에서 데이터 기반 소재를 설계하거나 물성을 예측하는 기술이 개발되어 보고되고 있다. 지난 2019년 융합연구리뷰의 “인공지능, 소재 개발에 침투하다” 편에서도 촉매소재 관련 다양한 연구의 예가 소개된 적 있다. 하지만, 모든 사례들에서 알 수 있듯이, 거의 모든 데이터 기반 소재 탐색·설계 및 물성 예측 기계학습 모델 개발 연구는 기 구축되어 있는, 또는 구축하기 쉬운 계산과학 데이터베이스에 기반하고 있다. 이는 아직 소재의 실험·공정, 측정·분석데이터가 충분히 축적되지 않았기 때문이며, 향후 실험데이터가 충분히 축적되어 소재를 설계하거나 물성을 예측하는 데에 활용되기를 기대한다.

그림 16. 소재 형상을 역설계할 수 있는 기계학습 모델 도식



출처 : Liu(2018)

1 AI로봇이 스스로 소재를 개발한다고?

2 소재연구를 위한 빅데이터 및 플랫폼 활용

V 소재연구를 위한 빅데이터 및 플랫폼

우리는 인공지능으로 대변되는 4차 산업혁명의 격변기에 있으며 빅데이터의 수집·활용에 초점을 맞춘 데이터 오픈 사이언스는 그 기반이 되고 있다. 융합연구리뷰를 통해 소재 빅데이터, 데이터 수집·관리, 오픈사이언스 및 기계학습과의 연관성에 대한 중요성을 살펴보았다. 최근 구글(Google), 아마존(Amazon) 및 페이스북(Facebook)과 같은 거대 IT 기업들은 데이터 인프라, 알고리즘 및 분석 도구 개발을 통해 이 분야를 주도하고 있다. 이러한 초대형 IT 기업과 소셜 미디어 등과 비교할 때 재료과학 분야에서 다뤄지고 있는 소재 데이터는 아직 시작 단계에 있다.

재료분야 과학자들은 소재 연구데이터 기반의 효율적인 소재 설계, 머신러닝 알고리즘 개발을 통하여 소재 개발에 필요한 새로운 방법론을 개발하고 차세대 소재 개발을 가속화시킬 것이다. 그러나 소재 데이터는 여전히 비정형화 되어 있고 흩어져 있으며 수집·활용을 위한 생태계가 부족하다. 소재정보학 분야는 오래 전부터 발전해 왔지만 관련성, 완전성, 표준화, 수용 및 수명과 같은 여전히 해결해야 하는 주요 과제를 갖고 있다. 소재분야의 온톨로지를 통한 재료 데이터의 표준화는 소재 데이터의 인공지능 기반 분석을 공유·통합·채택 하는데 크게 도움이 될 것이다. 또한 오류 추정을 위해 실험 데이터와 계산 데이터 사이에 피드백 매커니즘을 만들면 소재 데이터의 정확성 문제를 해결하는데 도움이 될 것으로 예견된다.

소재 연구데이터 수집·관리·공유를 위한 데이터 인프라 환경이 구축되고 있으며 소재·부품·장비 주력산업 및 혁신성장 산업에 요구되는 소재에 대해 데이터 기반 신소재 설계방법론이 개발되고 있다. 이를 위해서는 체계적인 소재 연구데이터 수집·공유·활용 플랫폼 구축 및 소재 분야에 특화된 데이터 분석 및 기계학습 모델 개발, 그리고 데이터 관련 국제기구 RDA 소재분야 워킹그룹 참여를 통한 표준화 활동, 빅데이터화를 위한 소재 데이터 공유 협력 및 인프라 요소 기술을 개발하는 것이 필요하다. 국제적 신뢰성이 확보된 분석 플랫폼을 구축하여 소재 분야 표준화를 선도하고 국제적 신뢰성 및 표준화 평가의 중심점 역할을 수행한다면 소재산업 경쟁력 상승에 기여할 것이다.

기계학습은 혁신적인 연구 방법이지만 소재 연구데이터를 이용한 기계학습 모델 개발을 위해서는 선제적으로 연구데이터 큐레이션 및 표준화에 대한 노력이 필요하다. 학문적 발전과 산업적 관심 사이의 시너지는 여전히

중요한 과제이지만 소재 데이터 및 전문 지식을 위한 지속 가능한 생태계를 만드는 것이 중요하다. 이를 통해 신규 소재군을 탐색·물성 예측하고 소재(태양광, 연료전지, 촉매 등) 개발을 위해 연구개발에 소요되는 시간, 비용, 위험요인 등을 획기적으로 감소시킬 수 있다. 연구개발 전 과정의 연구데이터(실험 과정, 측정 결과, 계산 결과 등)를 수집·활용·제공 및 플랫폼 기반의 신소재 연구 가시화로 소재 R&D의 방향을 명확하게 제시함으로써 다가올 미래를 전략적으로 선점하고 글로벌 기술경쟁력을 향상시키게 될 것이다.

저자_ 안재평(Jae-Pyoung Ahn)

• 학력

고려대학교 금속공학 박사
고려대학교 금속공학 석사
고려대학교 금속공학 학사

• 경력

現) 한국과학기술연구원 연구자원·데이터지원본부 본부장
現) 한국과학기술연구원 책임연구원
前) 한국과학기술연구원 특성분석센터 센터장

저자_ 김홍규(Hong-Kyu Kim)

• 학력

고려대학교 신소재공학 박사
고려대학교 신소재공학 석사
고려대학교 신소재공학 학사

• 경력

現) 한국과학기술연구원 특성분석센터 선임연구원

참고문헌

〈국내문헌〉

- 1) [한국과학기술연구원(2018)] 한국과학기술연구원, 『소재 연구데이터 수집 관리 및 활용 플랫폼 개발』, 2018. 08.
- 2) [소재기술백서(2017)] 한국재료연구원, 『소재기술백서2017』, 2017. 12
- 3) [한국화학연구원(2018)] 한국화학연구원, 『열전 소재 빅데이터 구축 및 데이터기반 소재 설계 플랫폼 개발』, 2018. 04.
- 4) [미래소재 연구데이터 플랫폼 구축사업 기획보고서(2019)] 한국재료연구원, 『미래소재 연구데이터 플랫폼 구축사업 상세기획 연구』, 2019. 02.
- 5) [국가연구데이터 공유활용 체제 구축 사업 기획보고서(2019)] 과학기술정보통신부, 『국가연구데이터 공유·활용 체제 구축』, 2019. 02.
- 6) [이영의(2015)] 이영의, 베이즈주의, 한국연구재단 저술총서, 2015

〈국외문헌〉

- 7) [Aguiar(2019)] Aguiar et al., "Decoding crystallography from high-resolution electron imaging and diffraction datasets with deep learning", *Sci. Adv.* 5, eaaw1949 (2019)
- 8) [Himanen(2019)] Himanen et al., "Data-Driven Materials Science: Status, Challenges, and Perspectives", *Adv. Sci.* 6, 21, 1900808 (2019)
- 9) [Lee(2020)] Lee et al., "A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns", *Nat. Commun.* 11, 86, 1-11 (2020)
- 10) [Liu(2018)] Liu et al., "Generative Model for the Inverse Design of Metasurfaces", *Nano Lett.* 18, 10, 6570-6576 (2018)
- 11) [Severson(2019)] Severson et al., "Data-Driven prediction of battery cycle life before capacity degradation", *Nat. Energy* 4, 383 (2019)

〈기타문헌〉

- 12) [IMRR WG Final Report and Recommendations] https://www.rd-alliance.org/system/files/documents/RDA_Working_Group_Report_Final_0.pdf
- 13) [Boothroyd 교수 홈페이지] <https://personal.ntu.edu.sg/cbb/info/dmformat/index.html#dm4>
- 14) [UCSB MRL 홈페이지] <http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp>

융합연구리뷰

Convergence Research Review 2021 May vol.7 no.5

이 보고서는 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 사업임

(No. NRF-2012M3C1A1050726)