

NIGT BRIEF

기후기술 연구자정보 구축을 위한
논문DB와 ORCID 연계 프로토콜
개발 연구: 수소기술 분야 적용

정민경 / 안세진 / 임종서 / 염성찬





하이라이트

- 기후기술법이 시행됨에 따라 기후기술 전문인력의 체계적 육성이 시급한 과제로 대두되고 있으나 현재 기후기술 인력에 대한 현황 파악이 어려워 효과적 정책수립에 한계가 존재함
※ 데이터 기반 인재정책의 중요성이 강조되고 있으나, 기후기술 분야를 포괄하는 인력정보 수집 체계는 부재한 상황
- 기후기술 분야의 다학제적 특성을 반영하여 관련 인력의 현황을 파악할 수 있는 방법론으로서 기후기술 논문 DB와 연구자 식별체계인 ORCID를 연계한 연구자 정보 구축 방법론을 수립하고 이를 수소 기술 분야에 시범 적용함
- 2019~23년 수소기술 분야의 피인용 상위 논문(5%) 저자 중 ORCID를 보유한 약 1,600여명의 데이터를 수집하고 DB를 구축하였으며, 분석 결과 국내 연구자에 대한 다학제적 정책 강화와 신진연구자 육성이 필요한 것으로 확인됨
- 기후기술 분야 인력 정책을 지원하기 위한 사람 중심 데이터베이스 구축방안을 개발·시도했다는 의의가 있으며, 이는 향후 과학기술 연구자의 경력발전과 전문성의 축적과정을 파악하는 기초자료로 활용될 수 있음

키워드

- 데이터 기반 과학기술 인력정책, 기후기술 인력정보 구축 방법, 기후기술 논문 DB, ORCID

추진배경

- 기술혁신을 통한 기후위기 대응을 위해 「제1차 기후변화대응 기술개발 기본계획(‘23~’32)」이 수립됨에 따라, 기후기술 분야의 전문인력 양성 계획과 이를 지원할 수 있는 과학기술 인력정보 수집이 시급한 과제로 부상
 - 기후기술법 제14조(전문인력 양성)은 기후기술 전문인력의 양성 및 교육 그리고 인력수급 전망 수립 및 양성계획 반영이라는 수요와 공급의 연계 측면을 강조
 - 그러나 기후기술 전분야를 포함하는 인력정보 수집 및 축적 체계가 아직 부재하며, 이는 기후기술 인력정의 및 분류체계 미비, 부처별 개별적 기후기술 인재양성 추진¹⁾ 등 한계에서 비롯함
 - 특히 인력정보는 유동성이 높고 연구자 정보와 실시간 활동이 연계되어 업데이트될 필요성이 있어 이러한 동적, 실시간적 특성을 반영한 데이터 구축방식이 필요
- 글로벌 산업구조의 변화와 신기술이 주도하는 경쟁 속에서 과학기술 인력의 중요성이 강조되고 있으며, 특히 데이터 중심의 연구자 정보수집에 대한 관심 제고

1) 에너지인력양성사업, 녹색융합기술인재양성사업, 탄소중립기술인력양성사업, 산림융복합전문인력양성사업 등

- OECD STI(Science, Technology and Industry)는 회원국을 중심으로 각국의 과학기술 혁신과 관련하여 연구인력 규모, 과학적 성과, 협력 네트워크 등의 통계지표와 분석결과를 발표하면서 국가간 비교가능한 데이터를 제공함²⁾
- 이러한 국가비교 데이터는 연구자 중심의 정보를 충분히 제공하지 못하다는 평가를 받았으며, 특히 초기경력 연구자에 대한 지원 및 모니터링을 위하여 연구자(사람) 중심의 데이터 구축의 필요성이 제기됨³⁾
- 2024년 9월 OECD-EU위원회는 연구자 경력 발전과 동향 정보를 실시간으로 제공하는 데이터 허브 구축을 목표로 RelCO(연구 혁신 경력 관측) 프로젝트 계획을 발표, 이를 통해 연구자의 ①스킬/훈련, ②직업/경력, ③이동성이라는 3가지 차원의 정보(통계, 방법론, 지표)를 수집·제공하고자 함(2025년 상반기 데이터 공개 예정)⁴⁾

[표 1] 글로벌 과학기술 인력 관련 데이터 구축 동향



- 국내 과학기술 정책에서도 데이터 기반 인력 정책이 강조되고 있으며 연구자 정보가 정책에 통합되는 흐름 등장
 - 과학기술정책 분야에서 과학적 증거를 정책 결정에 통합함으로써 경제적 또는 정치적 요인보다 과학적 기준을 우선하고자 하는 근거기반 정책이 지속적으로 강조
 - 데이터 중심 과기인력정책 전략이 2023년 12월에는 「국가전략기술 인재확보전략」, 2024년 6월에는 「데이터기반 과학기술 인재정책 고도화전략」*등에서 등장하며, 데이터확보/활용을 통한 정책지원 전략 제시
- 최근 과학기술 분야에서 인력정보를 수집하여 활용하는 실용적 사례가 증가하고 있으며, 이러한 데이터는 연구자 현황 파악, 경력경로 분석, 정책지원 및 정보제공 등 다양한 목적에 활용되고 있음
 - (KISTEP 전략기술 분야 R&D 인력 IRIS-고용 DB 연계 분석) 범부처통합연구지원시스템(IRIS) 및 고용보험 피보험자 DB간 가명정보를 연계하여 연구자 전주기 정보를 분석/활용할 수 있도록 DB 구축⁶⁾
 - (KISTEP 12대 전략기술 인력지도) 12대 국가전략기술 분야의 논문, 특허 정보를 활용하여 전세계 우수 연구자-연구기관 현황을 분석하고, 직관적으로 파악할 수 있는 지도를 구축⁷⁾

2) <https://www.oecd.org/en/topics/science-technology-and-innovation-indicators.html>

3) <https://initiative-se.eu/2022/09/25/press-release-a-manifesto-for-early-career-researchers/>

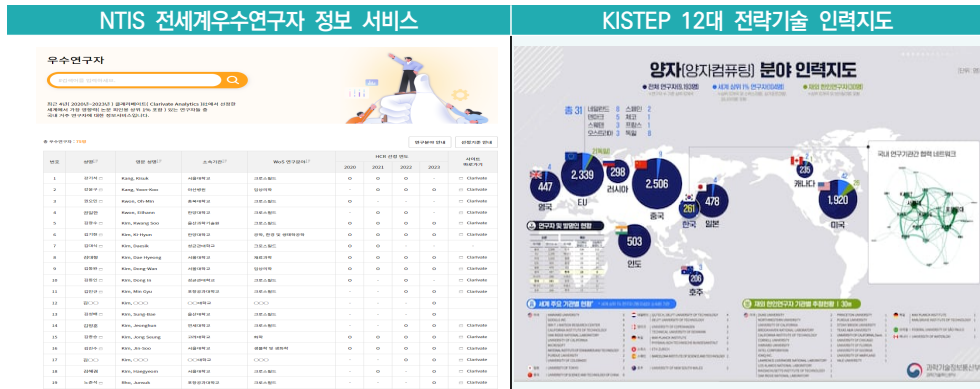
4) <https://ec.europa.eu/era-talent-platform/reico/>

5) <https://www.oecd.org/en/networks/research-and-innovation-careers-observatory.html>

6) KISTEP. 2023. 국가전략기술 핵심인력 양성 종합전략 수립 연구 최종보고서, 4장.

- (NTIS 과학기술 분야 우수연구자 지도) 최근 4년(2020~2023년) 클래리베이트(Clarivate Analytics)社에서 선정한 세계에서 가장 영향력(논문 피인용 상위 1% 포함) 있는 연구자들 중 국내 거주 연구자 정보 제공⁷⁾

[표 2] 국내 과학기술 데이터 기반 인력정보 활용 서비스



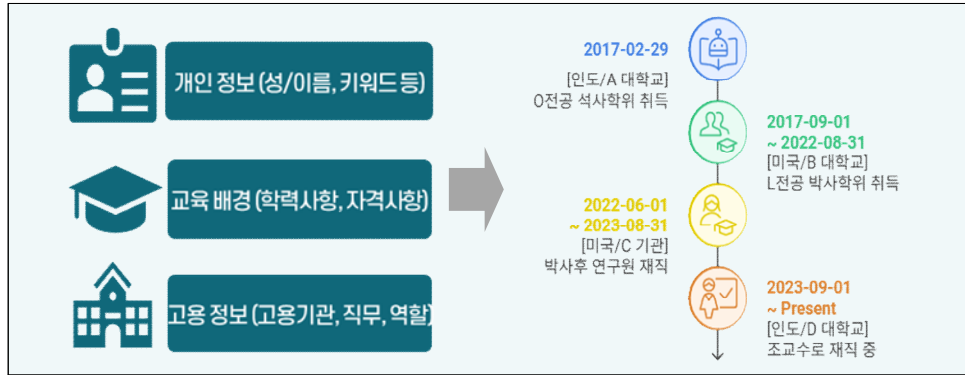
기후기술 논문 DB와 ORCID 연계 인력정보 데이터 구축: 수소 분야 시범 적용 방법

- 탄소중립 핵심기술 가운데 수소 기술 분야에서 정책 및 인력 정보 수요가 증가함에 따라, 수소기술 분야에 연구자 정보 DB 구축 방법론을 우선 적용함
 - * ‘수소 기술혁신 전략 이행단’(2022.2), 글로벌 R&D 전략지도안(2024.2)) 및 ‘글로벌 TOP 전략연구단 자원사업-청정수소저장 활용 연구단’ 사업 등에서 연구개발 인력 현황과 관련된 분석 정보 수요가 지속적으로 제기되고 있는 상황
- 논문 정보와 ORCID 정보를 연계하여 수소 기술 분야의 글로벌 연구인력 현황 데이터 구축하는 시범 연구를 진행함
 - ※ 기존 NIGT 기후변화대응 활동조사에서 기후기술 논문 DB(Web of Science)에 딥러닝 모델을 적용하여 기술 분야별로 논문을 분류하고 정량적 기준으로 추이와 성과를 분석하는 연구 수행
- 구축된 기후기술 논문 DB를 기반으로 연구자 정보를 추출하여 ORCID 연구자식별자 시스템 정보와 연계하여 기후기술 분야 연구자 데이터베이스를 구축*
 - * 시범적 연구로서 38대 기후기술 가운데 수소분야 피인용 상위 (5%) 논문을 대상으로 추진하여, 연구자 현황과 연구자의 교육 및 고용 정보 이력을 연계한 데이터베이스를 구축하고 인력정책 수립을 위한 기초자료를 생산
- 논문 서지정보-연구자 정보구축에 있어 ORCID 정보 연계 활용은 연구자 정보의 신뢰성을 제고하고 주요 경력정보를 통해 연구자의 전주기 경력에 관한 정보를 제공할 수 있다는 점에서 의미가 있음
 - 일반적으로 저자명에 기반한 연구자식별의 문제점은 동명이인, 이름변형 등 저자명 모호성 문제로 인한 신뢰성 저하문제, 연구자 경력변동 등으로 인한 추적 단절 문제(소속 기관 변경 시 동일 연구자 식별 어려움) 등이 존재
 - ※ ORCID는 세계적으로 인정받는 연구자 식별자(researcher identifier) 시스템으로, 논문저자명 식별 문제를 해결할 수 있도록 연구자에게 고유 16자리 디지털 식별자를 제공하여 연구자의 추적 및 관리 가능

7) 뉴시스. 2024. <https://news.nate.com/view/20240604n29974>

8) <https://www.ntis.go.kr/hurims/im/exRsch/selectExRschList.do> (최종 검색일:2024/10/1)

[그림 1] ORCID 연구자 정보 구조 및 정보연계 예시



• 수소 기술 분야 논문 DB와 ORCID 연계 프로토콜

- (1단계) 2019년도부터 2023년도까지 5년간 발표된 논문 대상으로 검색식 및 답러닝 분류모델을 적용
- (2단계) 기술훈 논문 피인용 상위 5%의 논문을 선별하고, 그 중에서 기술분야 연구자 정보를 추출함
- (3단계) ORCID API를 통하여 연구자의 연계정보를 수집하고 개인정보, 교육정보, 고용정보 DB 구축
- (4단계) 구축된 데이터베이스의 정제 및 표준화

[표 3] 수소기술 분야 적용 인력정보 데이터 구축 절차

절차	수행내용	결과
1. 논문 서지정보 수집	·키워드기반 Web of Science DB 논문 수집 ·답러닝 모델 적용 기술 분류 및 노이즈 제거	(논문) 15,207건
2. 논문수집 및 연구자 ORCID 추출	·피인용 상위 5% 논문 선별 ·논문 대상 저자정보(ORCID) 추출 ·추출된 ORCID 저자 비중 검토 및 데이터 정제	(논문) 769건 * ORCID 식별수 (전체) 2,624건
3. ORCID API 연동 정보 수집	·ORCID API를 통한 실시간 연구자 정보 수집 ·DB 구축, DB 통합·연계(개인/고용/교육정보 등)	-
4. 데이터 전처리 및 정제	·중복데이터 확인 및 제거, 데이터 표준화 작업	-

* 출처: 저자 작성

수소기술 분야에 대한 ORCID 인력정보 구축 방법론 시범적용 결과

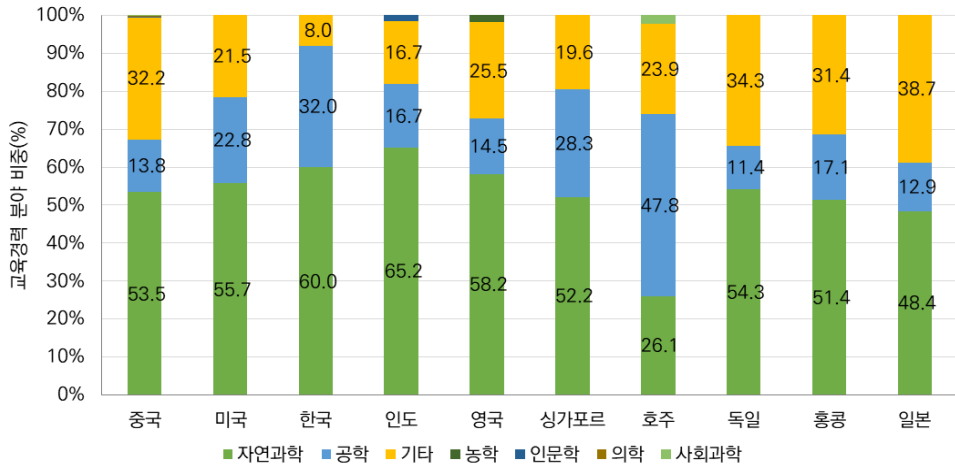
- 수소기술 분야 피인용 상위논문(총 769건)의 ORCID 식별 저자는 총 1,624명으로 확인되었으며, 연구자 정보 DB 분석 결과 개인정보 - 고용정보 - 교육정보 순서로 정보의 완결성이 높게 나타나 이를 통한 연구자의 고용, 교육정보 분석을 수행함
 - 개인정보는 기초정보 형태로 1,623개 프로필에서 정보를 보유하고 있으나, 분석을 위한 정보는 고용정보에서 가장 많은 건수를 보유하고 있으며(평균 2.3건), 교육정보는 평균 1.9건 수준으로 확인되어, 피인용 상위 5% 논문 저자들은 평균적으로 2개 이상의 경력 경험을 보유하고 있다는 것을 확인할 수 있음

[표 4] ORCID 연구자정보 결과

구분	ORCID 관련정보 건수	관련 정보 보유 ORCID 프로필 수	관련 정보 누락 ORCID 건수	ORCID당 평균 보유건수
개인정보	1623	1,623	0	1
교육정보	1,793	932	691	1.9
고용정보	2,568	1,118	505	2.3

- (국가단위 비교) 교육 경력을 위주로 살펴보면 대부분의 국가에서 자연과학 및 공학분야가 높은 비중을 차지하고 있으며, 특히 국내에서 이러한 연구자 전공 집중이 더욱 크게 나타나고 있음. 국내에서는 타 국가에 비해 기타 분야 전공 비중이 매우 낮게 나타나고 있어 다학제적 접근, 융복합 분야의 교육 정책 강화 등 인력양성 정책의 필요성이 확인됨

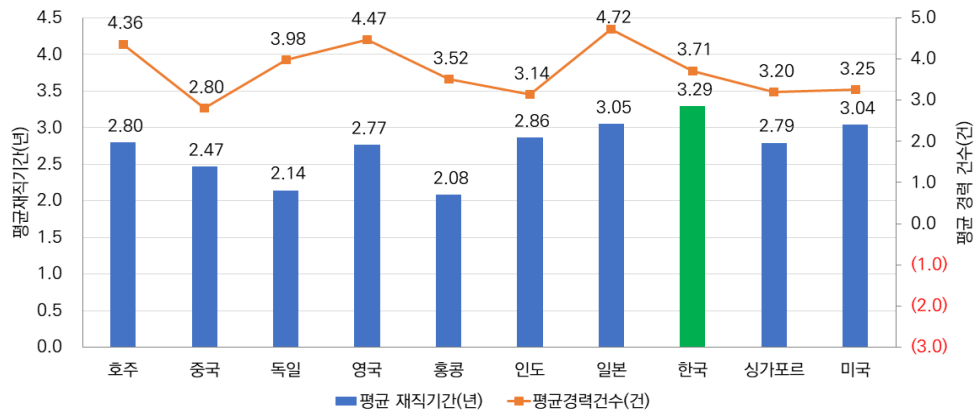
[그림 2] ORCID 연구자 교육경력(분야 및 국가별) 정보연계



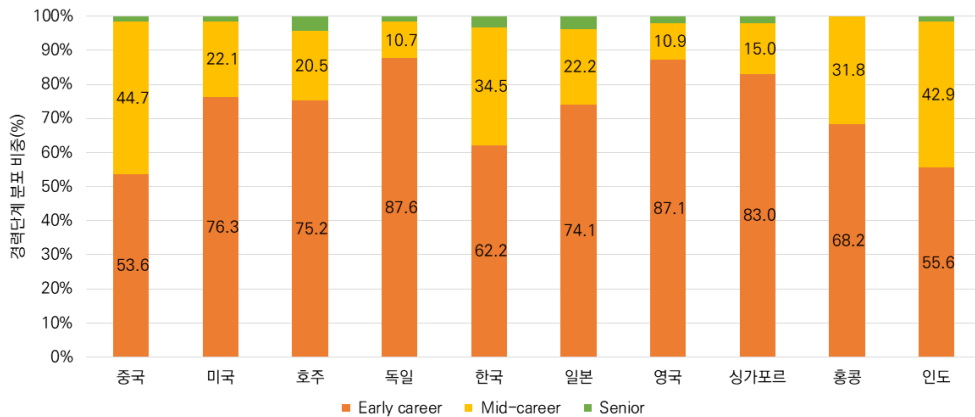
구분	자연과학	기타	공학	사회과학	의학	인문학	농학	총합계
중국	502	302	130	2	2	1	0	939
미국	127	49	52	0	0	0	0	228
한국	45	6	24	0	0	0	0	75
인도	43	11	11	0	0	1	0	66
영국	32	14	8	0	0	0	1	55
싱가포르	24	9	13	0	0	0	0	46
호주	12	11	22	1	0	0	0	46
독일	19	12	4	0	0	0	0	35
홍콩	18	11	6	0	0	0	0	35
일본	15	12	4	0	0	0	0	31
총합계	837	437	274	3	2	2	1	1,556

- 고용정보 DB 내 경력건수 및 경력당 평균 재직기간 분석 결과, 한국은 평균 3.7개의 경력경험 보유, 경력당 평균 재직기간 3.29년 수준으로 타 국가에 비해 이동성이 다소 낮게 나타나는 특징을 보임. 또한 중위경력 그룹의 비중이 다소 높은 유형으로 인도, 중국과 다소 유사한 형태로 나타남. 이는 신규 인력그룹의 성과보다는 기존 연구자 그룹에서 높은 성과를 발생시키고 있음을 시사함

[그림 3] ORCID 연구자 경력건수 및 재직기간 정보연계(국가별)



[그림 4] ORCID 연구자 경력단계별 분포 현황(국가별)



- (DB 정보구조 및 주요 변수) 연구자의 기본적인 식별을 위한 개인정보, 연구자의 학문적 배경과 전문성을 추적하기 위한 교육정보, 연구자의 경력이력을 추적하기 위한 고용정보로 구성되어 있음
 - 개인정보의 핵심변수는 ORCID 식별자, 이름, 성, 소개글, 키워드 등으로서 기본적인 식별정보들로 구성되어있고, 교육정보에서는 학위 수준, 학과, 국가, 학업시기(시작/종료시기) 등의 정보들을 제공하며, 이를 바탕으로 학위 유형, 전공 정보 표준화, 학위취득 국가명 매칭 진행하여 변수 생성 가능
 - 고용정보에서는 연구자의 고용기관, 직위, 고용국가, 고용시기(시작/종료시기) 등의 정보를 제공하고 있으며, 이를 바탕으로 고용기관 유형, 표준화된 직위명 등의 변수 생성이 가능
- (추가 변수 및 분석가능성) 교육정보와 고용정보를 연계하거나, 추가적인 변수 생성을 통하여 다양한 분석 활용 가능
 - 최종학위 이후 고용시점까지 기간을 기준으로 경력단계(초기 연구자/중견 연구자/고경력 연구자) 구분
 - 학위취득 국가와 고용기관 소속 국가 일치여부 등을 기준으로 연구자의 국가간 이동성 분석 등에 활용
 - 고용기관의 유형(학계-연구기관-민간기업) 변수를 생성하여 연구자 경력 분석에 활용

[표 5] ORCID 식별자 정보 변수 생성 예시

구분	설명	변수명	변수 설명	변수(예시)
개인정보	ORCID 시스템에 등록된 연구자의 기본 식별 정보	orcid_id	ORCID 식별자	0000-0002-1825-0097
		first_name	이름	Carberry
		last_name	성	Josiah
		biography	소개글	Josiah Carberry is a fictitious person. This account is used as a demonstration account by ORCID, CrossRef and others who wish to demonstrate the interaction of ORCID with other scholarly communication systems without having to use a real-person's account.
교육정보	연구자 교육/학위과정 관련 정보 (연구자 id별로 다수 건 연계)	organization	학교명	-
		normalized_degree	정규화된 학위 유형	'Bachelor', 'Master', 'Doctorate', 'Other'
		department_category	전공/학과	'Natural Sciences', 'Engineering and Technology', 'Medical and Health Sciences', 'Agricultural Sciences', 'Social Sciences', 'Humanities', 'Other'
		organization_country	학위 취득 국가	ISO 3166-1 alpha-2 국가 코드 (예: 'US', 'GB', 'JP')
		start_date	학업 시작 날짜	YYYY-MM-DD 형식
		end_date	학업 종료 날짜	YYYY-MM-DD 형식
고용정보	연구자 취업이력/고용 기록 정보	organization	기관명	-
		organization_type	고용 기관 유형	'University', 'Research Institute', 'Company', 'Government Agency', 'Other'
		normalized_role	정규화된 직위	Professor', 'Associate Professor', 'Assistant Professor', 'Researcher', 'Postdoctoral Researcher', 'Engineer', 'Manager', 'Other'
		is_current	현재 고용 여부	True, False
		organization_country	고용 국가	ISO 3166-1 alpha-2 국가 코드 (예: 'US', 'GB', 'JP')
		start_date	고용 시작 날짜	YYYY-MM-DD 형식
		end_date	고용 종료 날짜	YYYY-MM-DD 형식
교육-고용정보 연계	교육정보와 고용정보를 결합하여 생성한 연구자 경력 통합 정보	career_stage	경력 단계	'Early career', 'Mid-career', 'Senior'
		highest_degree	최종 학위	normalized_degree와 동일
		years_since_degree	최종 학위 취득 후 경과 연수	0 이상의 실수
		country_move	학위 취득 국가에서 현재 고용 국가로의 이동 여부	True, False
		career_trajectory	경력 궤적 유형	'Academic', 'Industry', 'Mixed', 'Other'

결론 및 기대효과

- **기후기술 분야 논문 DB와 ORCID 연계를 통해 연구자 정보를 체계적으로 구축하고 데이터기반의 기후기술 인력정책 지원기반을 마련하는 방법론을 제시**
 - 기존 기후기술 DB 의 활용도를 제고할 수 있는 글로벌 인력정보 구축 방안을 제시하였으며, 특히 ORCID 시스템 활용을 통해 동명이인, 이름변형 등의 저자명 모호성 문제를 해결하고 연구자 중심의 교육-고용 경력데이터 생산이 가능함을 확인
 - 본 방법론을 활용한 기후기술 인력에 대한 현황 정보는 기후기술 촉진법 제 7조에 따른 기후기술 인력양성 수립계획 또는 중장기 전망 등을 수립하는데 유용한 기초자료로 활용될 수 있으며, 이와 같은 방법론은 타 과학기술 분야로도 확장 적용이 가능함
- **기후기술 분야 논문 서지정보-ORCID 연계정보 향후 활용 가능성**
 - 38대 전체 기후기술 분야로의 확대 적용, 시계열적 분석을 통한 연구자 경력패턴의 도출, 국가별, 기관별 인력현황 및 이동성 분석 등을 수행하여 기술분야별 연구자 경력개발 비교 등 기후기술 인력양성 현황 파악에 용이한 자료 생산 가능
 - * (연구자 수준 데이터 활용) 연구자의 장기적인 경력경로 패턴, 연구자 성과 정보와 고용/교육정보 연계분석, 경력단계와 국가간 이동 등의 정보를 연계하여 연구자 생애주기 동향 조사 및 정책 기초자료 마련에 활용 가능
 - * (기관/국가 수준 데이터 활용) 피인용 상위 논문 주요 저자들이 소속된 기관 및 국가 추이, 국가단위 연구자 이동성과 성과, 연구자 경력단계별 국가/기관 분포 등 분석을 통하여 기후기술 분야 인력정책 수립의 근거자료 마련에 활용 가능
- **수소분야 ORCID 시범 적용 검토 결과 교육/고용정보 측면에서 데이터 누락 프로파일의 규모가 한계점으로 등장**
 - 연구자 현황 파악과 경력 분석의 정확성을 높이기 위해 데이터 수집 범위의 확대 및 누락정보 보완 연구를 지속적으로 진행할 필요가 있으며, 이를 바탕으로 향후 38대 전 분야 데이터 구축에 대한 정책적 수요를 발굴하고 후속 연구를 기획할 필요가 있음

[표 6] ORCID 활용 한계에 따른 후속 연구 시 주요 고려 사항

- ▶ (연구자 정보 포괄성) 연구자식별자(ORCID)의 등록률과 관련하여 연구자들의 등록률이 국가/분야별로 편차가 크다는 지적이 존재하며, 특히 신진연구자나 개발도상국 연구자들의 등록률이 낮을 수 있음
 - * ORCID Sandbox 타입의 테스트 데이터를 활용하여 최종 업데이트 데이터와는 다소 상이할 수 있으나, 연구에서 38대 기후 기술 논문 가운데 상위 5%를 선별하여 ORCID를 추출한 결과 식별자가 누락된 논문 비중이 전체의 10% 미만으로 확인
- ▶ (데이터 품질) 본인이 직접 입력하는 방식으로 운영됨으로써 연구자 정보 일부가 누락될 가능성이 있으며, 교육 및 고용정보의 업데이트가 불규칙적일 수 있음. 이와 같은 한계는 데이터 검증을 위한 추가 정보원 확보 등을 통해 보완할 필요가 있음
- ▶ (기술적 어려움) ORCID API의 호출제한으로 인하여 대규모의 데이터 수집에 시간이 소요되며, 38대 전 분야 데이터 구축을 위해서는 효율적 데이터 수집방안과 API 활용전략 등 연구 보완방안 필요

본 연구는 NIGT 데이터정보센터 주요사업 “F2430101 기후변화대응 기술개발 활동조사 연구” 내용을 토대로 작성되었습니다.