

연구 데이터의 효율적 관리·활용 방안

미국 국립아카데미에서 발간한 「디지털 시대 연구 데이터의 완전성, 접근성, 관리성 확보 방안」*의 내용을 보고드립니다

* Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age, 2009.10

※ 본 내용은 KISTEP(정책기획실)에서 정리·분석한 내용임

1. 개요

□ 미국 국립아카데미는 디지털 시대에 있어 연구 데이터(**Research data**)*의 중요성을 강조하고, 데이터의 관리·활용을 위한 3대 원칙과 11개 권고안을 제시

* 연구 수행 중 생성되거나 작성된 모든 정보로서 최종결과물뿐만 아니라 중간생성물도 포함

○ 디지털 기술의 진보는 방대한 데이터를 생산하여 새로운 연구영역을 개척하는 원동력을 제공함과 동시에, 다양한 문제점을 야기

- 방대한 데이터가 생성되어 전파되는 과정에서 잘못 가공되고 조작되어지면서 데이터의 정확성과 신뢰성을 보장하지 못하게 됨

- 또한, 디지털 기술로 인한 혁신의 가속화, 표준의 부재, 프라이버시 침해, 국가보안 침입과 같은 문제점들이 데이터 공유를 저해하고, 방대한 데이터의 장기적 보관의 문제가 발생

○ 본 보고서는 연구 데이터에 영향을 미치는 요소를 크게 세 가지로 구분하고 그에 수반되는 원칙*과 권고안을 제시

* 연구 데이터의 3대 관리 원칙 : ① 완전성(Integrity), ② 접근성(Accessibility), ③ 관리성(Stewardship)

2. 데이터의 완전성(Integrity)* 확보

* 완전성은 신뢰성과 정확성을 의미하며, 데이터의 보존과 공개의 의미도 내포

- 데이터의 완전성을 확보하는데 가장 중요한 방법은 높은 기준의 개방성과 투명성을 확보하는 것
 - 과학, 공학, 의학 등의 분야는 물리적, 생물학적, 사회적 현상을 포괄적으로 다루고, 분야별로 특이성이 존재하기 때문에 일반화하기 어려운 특징 보유
 - 방대한 양의 데이터가 수집·분석되고 새로운 데이터가 생성되어 공유하는 과정에서 데이터가 잘못 가공되거나 조작되어 결과물의 오류를 초래
 - 따라서 데이터의 수집, 분석, 해석에 있어서 오류의 발견과 정정을 위해서는 데이터를 생성하고 가공하는데 사용된 방법과 도구들이 전문가들에 의해 검증되는 절차가 필수

〈 원칙 1 : 데이터의 완전성(Integrity) 확보 〉

데이터의 완전성 확보는 다양한 학문지식의 진보와 연구 시스템에 대한 국민적 신뢰를 유지하기 위한 핵심과제이며, 이를 확보하는 궁극적 책임자는 연구원임을 명심해야 함

- ① 권고안 1 : 연구원들은 각 연구 분야별 전문적 기준에 따라 연구 데이터의 완전성을 확보할 수 있도록 과제를 구상하고 관리
 - 연구자들은 데이터 조작이나 연구결과의 불법 표절방지와 같이 모든 분야에 적용되고 연구의 근간이 되는 보편기준과 특정 분야에 적용되는 기준 모두를 명백히 이해하고 준수
- ② 권고안 2 : 연구기관들은 보편적인 기준과 특정분야 적용 기준 모두와 관련하여 신뢰성있는 연구 수행을 위한 훈련을 받도록 함. 일부 연구지원기관들은 훈련 프로그램을 개발하거나 훈련비용을 지원

- 데이터의 완전성 확보를 위한 환경 조성에는 연구공동체와 이해관계자(연구원, 연구기관, 연구지원기관, 전문학회, 학술지 등) 모두의 노력이 필요
- ③ 권고안 3 : 연구공동체와 이해관계자들은 데이터의 완전성 확보를 위한 전문 기준을 개발하여 전파하고, 이 기준을 충실히 따르도록 명시
- 분야별 전문기준들은 명확히 정의되고 설명되어야 하며, 기술 분야와 연구환경의 변화에 따라 데이터 관리지침들을 재검토하고 개정하는 것이 필수
- ④ 권고안 4 : 연구기관과 전문학회, 학술지는 연구에 기여한 데이터 전문가들의 업적을 적절히 명시하고 연구지원기관은 데이터 전문가에 대한 재정지원이 확대되고 있음을 인지할 필요

3. 데이터의 접근성(Accessibility)* 확보

* 접근성(Accessibility)에는 접근과 함께 공유(Sharing)의 의미를 내포

□ 데이터의 접근용이성 확보와 공유는 그 데이터에 대한 정확성을 검증하고, 연구 결론을 검증할 수 있는 최적의 방법

- 경우에 따라, 데이터의 접근이 불가능*하거나 공개를 연기하는 합법적인 부분**도 존재

* 상업적 물품 또는 서비스를 생성하는 단계인 경우, 국가 안보를 위한 기밀인 경우, 연구 데이터와 생성된 자료의 관리권한 및 시기가 연구 관련 고위 감독관과 직속동료들에 제한된 경우

** 상업적으로 응용될 가능성이 있는 분야의 특허출원, 계약상의 제한, 기술적 제약 등은 데이터 접근을 연기하거나 제한하는 요소로 작용

- 그러나 기본 원칙은 생성된 결과물에 사용된 연구 데이터와 방법* 및 관련 주요 정보의 공개

* 데이터의 수집, 생성, 분석에 사용되는 기법과 절차, 도구 포함

〈 원칙 2 : 데이터 접근성(Accessibility) 확보 〉

출판된 연구결과와 이와 관련된 연구 데이터와 방법 및 그외 주요 정보는 공공적 이용이 가능해야 하고, 이를 위해 연구원들이 데이터 공유와 개방성을 증진시킬 수 있는 환경을 이루도록 노력해야 함

- ⑤ 권고안 5 : 모든 연구원들은 연구결과의 검증을 위해 연구 결과를 모두가 이용 가능하도록 연구 데이터와 방법 및 기타 주요정보를 생성. 단, 비공개로 할 수 밖에 없는 경우는 예외
- ※ 정부지원금을 통한 연구뿐 아니라 사기업의 지원을 받은 연구 또한 공개하여 공유함으로써 공익에 이바지할 수 있도록 해야 함
- 데이터 접근성의 기준은 일반적으로 연구분야별로 확립된 기준*을 가지고 있으며, 데이터의 접근성을 증진하도록 하는 임무를 가진 기관이 존재할 경우 더욱 용이
- * 접근성 기준은 분야별 학자들의 의사소통 기준에 따라 차이점이 있으며 유동적임
- ⑥ 권고안 6 : 연구 데이터 공유를 위한 기준이 설립되지 않은 분야는 공유기준의 개발이 필요하므로, 관련 이해관계자 모두가 참여한 객관적 프로세스를 통해 각 분야에 적합한 기준 마련
- ⑦ 권고안 7 : 연구이해관계자들은 데이터 공유를 위한 공공인식을 제고시키고, 공개정책(Publication policies) 수립을 통해 데이터 공유를 촉진
- ⑧ 권고안 8 : 연구기관은 연구 데이터의 관리와 접근에 관련된 명확한 정책을 수립하고, 이 정책을 필히 연구원들에게 명시하여야 함
- ※ 연구기관이 수립한 정책은 타기관이나 개인이 데이터를 요청하였을 경우, 연구원과 기관이 공동으로 이에 대응할 책임을 진다는 사항을 포함

4. 데이터의 관리성(Stewardship)* 확보

* 관리성은 연구 데이터의 관리(management), 보존(preservation)의 의미를 내포

- 연구 데이터의 가치 판단과 장기적 보존을 위해서는 데이터 보존의 필요성 확립과 데이터 보존계획 수립이 필수
 - 연구 데이터는 연구 실험실, 시설물 등의 인프라 및 네트워크의 본질적 역할을 담당하는 중요한 도구로서 장기적으로 보관될 가치가 존재
 - 가치있는 데이터를 관리하기 위해서는 보존해야 할 데이터와 버려야 할 데이터에 대한 판단과 방대한 관리 비용의 분담에 대한 합의가 부재

〈 원칙 3 : 데이터 관리성(Stewardship) 확보 〉

보관가치가 있는 데이터는 장기간 보존되어야 하며, 사용자가 언제든지, 쉽게 찾아, 올바르게 사용할 수 있도록 문서화, 색인화, 인덱스화하여 작업되어 보존되어야 함

- ⑨ 권고안 9 : 연구원들은 데이터 보존 계획을 포함하여 데이터 관리 계획을 연구과제 수행 초기에 수립
 - 생성과정 및 정보를 수반하지 않는 연구 데이터는 가치가 없으므로, 데이터에 주석을 달아 장기 보존 가치를 유지하는 것이 데이터 보존계획 수립에서 가장 중요한 과제
 - 연구원들은 데이터 주석과 보존, 장기간 접근을 위한 정책과 기준 개발에 참여할 필요
- ⑩ 권고안 10 : 연구 데이터 관리를 위한 기준 개발의 일환으로 각 연구분야에서 생산되는 데이터를 평가하고, 어떠한 데이터를 보존할 것인지에 대한 기준을 수립하는 지침서를 개발

- 연구원들이 연구 데이터 관리에 대한 책임을 다하기 위해서는 제도적인 지원이 필수적

11 권고안 11 : 연구기관 및 연구지원기관은 연구원들의 데이터 보존 수요를 파악하기 위해 연구원, 데이터 전문가들과의 공동작업을 통해 데이터 관리계획을 수립하고 시행

- 쉽게 해결하기 어려운 연구 데이터의 장기 보존과 비용 부담 문제에 대한 해결책을 장기간에 걸쳐 논의하고 모색할 필요

5. 정책적 시사점

□ 국가R&D사업의 투자 효율성 제고와 연구성과의 활용을 위한 연구 데이터의 효과적이고 효율적인 관리방안이 필요

- 국가R&D 성과의 체계적인 관리·활용 촉진을 위하여 연구 데이터 관리 기준 및 지침을 수립할 필요

- 연구 데이터 관리지침과 연구윤리지침을 연계하여 전반적인 연구의 진실성 및 신뢰성 제고 검토하고, 연구 데이터 저장·보관을 위한 보존가치 판단기준 및 보존기간 산정기준 마련
- 수립된 기준 및 지침에 대해 대학·출연(연)의 연구원과 연구 데이터 관리자를 대상으로 주기적인 교육·훈련 실시

- 연구 데이터 관리 시스템을 구축하고 활용을 활성화할 필요

- 연구기관별로 표준화된 종합 연구 데이터 관리시스템 구축을 추진하고 기관 간 시스템을 연계

□ 본 보고서에서와 같이 미국에서 논의되고 있는 사항과 관련하여 국내 NTIS(국가과학기술지식포털, www.ntis.go.kr)도 국가R&D사업의 연구 데이터를 관리하는데 있어 개선 사항을 검토할 필요

- 현재의 NTIS에 등록되는 정보들은 연구의 최종결과물 및 성과이며, 연구 기획단계부터 종료까지 생성되는 모든 연구 데이터에 대한 취합은 이루어지지 않음

- 모든 연구로부터 생성되는 방대한 연구 데이터의 관리는 매우 어려운 문제이므로 다음과 같은 사항들을 고려
 - 최종 연구결과물 및 성과는 **NTIS**에서 관리하고, 연구 중간 생성물은 각 연구기관별로 표준화된 관리시스템을 통해 관리하며 중간생성물에 대한 수집·접근 및 공유 기준 마련
 - 연구 데이터에 대한 검증절차를 강화하고 NTIS뿐만 아니라 각 연구기관에서 관리하는 연구 데이터의 가치판단을 통한 장기적인 보존계획 수립이 필요
 - 연구성과물에 사용된 연구데이터 및 방법 등에 대한 정보 관리 필요성을 검토하여 ① **NTIS 표준항목**을 확대 및 ② **참조표준 (Reference Standards*)**과 연계 등을 통한 연구데이터에 대한 접근성 및 관리성을 확보하기 위한 방안 마련이 필요한 것으로 보임
- * **참조 표준(Reference Standards, 參照標準)** : 신뢰할 수 있는 수치 데이터를 말하는데, 데이터와 정보의 정확도와 신뢰도를 공인하기 위한 자료로 사용된다. 주로 국제적으로 저명한 과학 기술 분야의 각종 저널, 논문과 핸드북, 편람 등을 통해 발표된 신뢰성 높은 각종 과학 기술 물성값, 실험측정 데이터, 수치 및 상수 데이터 등으로 구성된다. 데이터와 정보의 정확도와 신뢰도를 분석 심사하여 참조 표준으로 설정하고 이를 공인해 주는 **표준 참조자료(SRD: Standard Reference Data)**를 대표적인 참조 표준으로 볼 수 있다.