연구 데이터의 출판과 인용에 관한 연구

A Study on Publication and Citation of Research Data

이 상 호
한국과학기술정보연구원

Lee sang-ho
Korea Institute of Science and Technology Information

요약

최근 오픈 사이언스 운동과 함께 정부 부처, 연구비 지원기관 등에서는 공적 기금으로 연구를 수행하는 경우 생성된 각종 연구 데이터를 관리하고 공개하도록 의무화하려는 움직임이 있다. 데이터에 DOI와 같은 식별자를 부여하여 데이터 리파지토리를 통해 출판하면 이해당사자들에게 많은 이익을 가져다 줄 수 있으며, 데이터의 인용을 활성화하기 위해 주제별 또는 기관별 리파지토리나 데이터센터에서 표준적인 인용 방법과 인용 요소들을 발표하고 있다. 앞으로 과학연구의 공개, 개방화가 더욱 추진되면 더욱 많은 연구데이터의 공유 활동이 일어날 것으로 예상되며 분야별 또는 유형별로 국제적인 데이터 리파지토리들이 출현하여 학술 논문의 근거가 되는 데이터 저장소로서의 역할을 수행할 것으로 생각된다.

1. 서론

오픈 사이언스 운동에 따라 정부 부처, 연구기관, 연구비 관리기관 등이 연구성과물의 공개를 요구하는 경향이 증가하고 있다. 또한 학회, 대형출판사 등은 논문 게재시 관련 데이터의 제출을 함께 요청하고 있으며 각종 학술 단체에서는 연구자들에게 연구 관련 데이터를 공개하거나, 인용과 출판을 권고하고 있다.

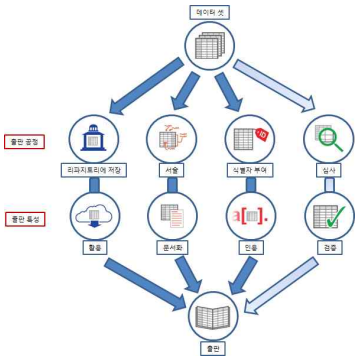
2. 출판의 목표

데이터 출판의 목표는 과학적 연구가 재생산되고 데이터가 재이용되도록 하기 위한 것으로서, 연구데이터를 널리 출판한다는 것은 논문의 부정행위뿐만 아니라 정직한 오류를 공개할 수가 있으며, 출판된 학술 논문의 근거가 되는 데이터의 투명성으로 인해 과학적 재생산성을 향상시킬 수가 있다.

또한, 데이터를 적절하게 재이용하게 되면 연구비용을 낮추고 연구를 촉진시킬 수 있는데, 데이터의 문서화, 출판과 저장에는 시간과 비용이 소요되나 데이터를 반복해서 생산하는 비용에 비하면 훨씬 저렴하며 시간 경과에 따라 측정한 기후 데이터나 특정한 시점에서 관찰한 우주 현상 데이터는 돈으로 대체할 수 없는 귀한 가치를 가지고 있다.

3. 데이터 출판의 특성

학술 단체에서 동의하고 있는 데이터 출판의 특성으로서 ① 활용성, ② 문서화, ③ 인용, ④ 검증을 4가지 요소를 들 수 있다(그림 1)[1].



▶▶ 그림 1. 데이터 출판을 위한 리파지토리 저장, 문서화, 식별자 부여와 심사

데이터를 출판한다는 것은 종래의 저널을 출판하는 것과 유사하다고 할 수 있는데, 일반적으로 출판계획 수립, 출판 신청, 데이터와 미디어 파일 제출, 품질 심사, 편집과 표준 기입, 리파지토리 저장과 공개를 통한 출판의 순서로 이루어진다. 저자들은 이와 같은 단계별로 데이터를 출판함으로써 데이터 출판에 기여하게 된다[2].

4. DOI

DOI는 디지털 학술 원문에 부여되는 영구식별자이다. DOI명은 데이터 셋의 내용이 적절하게 작성되고 장기적으로 접근할 수 있는 안정된 저장소에서 관리되는 데이

터 셋에 한해서 공인된 기관들이 DOI명을 부여하는 것이 좋다. 이는 DOI명을 가진다는 것은 해당 데이터 셋이 적절하게 관리되고 장기적으로 활용할 수 있도록 안정적으로 콘텐츠에 접근할 수 있다는 의미이며, 데이터 셋에 DOI를 부여할 경우의 장점으로서는 인터넷에 존재하는 연구 데이터에 대한 접근성이 향상되고, 과학 논문에서도 연구 데이터를 합법적으로 인용할 수 있는 콘텐츠로 인정하게 되며, 연구 결과를 검증하고 미래의 연구에 다시 재활용할 수 있다.

5. 데이터 인용

데이터 인용이란 연구자들이 저널 논문이나 기술 보고서, 컨퍼런스 논문과 같은 연구 성과물에 참고 문헌으로 서지 정보를 관례적으로 인용하는 것처럼 데이터를 인용하는 것을 말한다. 데이터를 인용하게 되면 해당 출판물의 인용 빈도를 향상시킨다는 보고가 있으며[3], 데이터가 1급 연구 성과물로서 인정받을 수 있게 되고, 본인의 이력서나 연구실적에 데이터의 출판과 인용 내용도 기재할 수 있다. 또한 데이터에 DOI를 부여하게 되면 그 데이터의 영향을 측정할 수 있는 추적과 계량화가 가능하게 된다.

6. 데이터 인용 원리

데이터 인용 원리[4]는 인용의 목적, 역할, 기여에 대한 것으로서 특정 분야를 위한 권고사항이 포함되어 있지 않으며 커뮤니티를 지원하기 위해 개발되었다. 이러한 원리들은 중요성으로 판단하기보다 쉽게 이해할 수 있도록 분류되었는데 그 내용은 ① 중요성, ② 신뢰와 기여, ③ 증거, ④ 고유 식별, ⑤ 접근, ⑥ 영속성, ⑦ 특정과 검증, ⑧ 상호 운용성과 유연성으로 나누어 볼 수 있다.

7. 데이터 인용 방법과 요소

출판물과 직접 관련이 있는 데이터를 인용하는 방법은 일반적으로 데이터에 대한 접근 방법을 알려주는 것인데, 예를 들면 공개된 데이터에 대해서는 어떤 리파지토리에서 무엇을 이용할 수 있는지, URL, 식별자, 또는 데이터에 접근할 수 있는 접근 코드 등을 알려주어야 하며, 접근이 제한된 데이터에 대해서는 접근을 제한하는 법적 또는 도덕적 이유를 설명하고 접근 조건을 설명하는 문서에 영구적으로 연결이 되도록 해야 한다.

데이터 인용도 다른 인용과 동일한 요소를 포함하고 있으며, 주제별 또는 기관별 리파지토리나 데이터센터에서는 일반적으로 데이터 인용을 위한 지침을 발표하고 있다. 다만 이러한 지침이 없을 경우에는 서적을 인용할 때와 같은 형식을 따르는 것이 좋으며 데이터 인용을 위해 고려해야 하는 인용 요소들은 ① 저자(author), ② 출판 일시(publication date), ③ 제목(title), ④ 편집(edition), ⑤ 버전(version), ⑥ 자원 유형(resource type), ⑦ 출판자(publisher), ⑧ 식별자(identifier), ⑨ 소재지(location)가 있는데 이러한 인용 요소 가운데에서 모든 데이터 인용에 공통으로 사용되는 중요한 요소로서는 저자, 제목, 출판 일시, 출판자와 소재지로서 이러한 정보들은 이용자들로 하여금 데이터 셋의 관련성을 판단하고 데이터 셋에 접근할 수 있도록 하며 데이터의 품질 혹은 지속성에 대한 확신을 줄 수가 있다.

이 밖에도 데이터의 인용에서 고려해야 할 사항으로서 데이터의 입도, 비공개 데이터의 인용, 샘플, 물질과 같은 물리적 데이터의 인용, 시간의 경과에 따라 변화하는 데이터의 인용 등이 있다.

■ 참 고 문 헌 ■

[1] F1000Research 2014, 3:94 Last updated: 09 MAR 2016, doi:10.12688/f1000research.3979.3
[2] http://opencontext.org/about/publishing, (2016. 10.)
[3] http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308, (2016. 9.)
[4] https://www.force11.org/datacitation, (2016. 9.)