

Scientific Data 학술지 분석을 통한 데이터 논문 현황에 관한 연구

An Investigation on Scientific Data for Data Journal and Data Paper

정은경 (EunKyung Chung)*

초 록

데이터 학술지와 데이터 논문이 오픈과학 패러다임에서 데이터 공유와 재이용이라는 학술활동이 등장하여 지속적으로 성장하고 있다. 본 논문은 영향력있는 다학제적 분야의 데이터 학술지인 Scientific Data에 게재된 총 713건의 논문을 대상으로 저자, 인용, 주제분야 측면을 분석하였다. 그 결과 저자의 주된 주제 영역은 생명공학, 물리학 등으로 나타났으며, 공저자 수는 평균 12명이다. 공저 형태를 네트워크로 살펴보면, 특정 연구자 그룹이 폐쇄적으로 공저활동을 수행하는 것으로 나타났다. 인용의 주제영역을 살펴보면, 데이터 논문 저자의 주제영역과 크게 다르지 않게 나타났으나, 방법론을 주로 다루는 학술지의 인용 비중이 높은 것은 데이터 논문의 특징으로 볼 수 있다. 데이터 논문 저자의 키워드를 사용하여 동시출현단어분석 네트워크로 살펴본 데이터 논문의 주제영역은 생물학이 중심이며, 구체적으로 해양생태, 암, 계놈, 데이터베이스, 기온 등의 세부 주제 영역을 확인할 수 있다. 이러한 결과는 다학제학문 분야를 다루는 데이터 학술지이지만, 데이터 학술지 출간에 관한 논의를 일찍부터 시작해온 생명공학 분야에 집중된 현상을 보여준다.

ABSTRACT

Data journals and data papers have grown and considered an important scholarly practice in the paradigm of open science in the context of data sharing and data reuse. This study investigates a total of 713 data papers published in Scientific Data in terms of author, citation, and subject areas. The findings of the study show that the subject areas of core authors are found as the areas of Biotechnology and Physics. An average number of co-authors is 12 and the patterns of co-authorship are recognized as several closed sub-networks. In terms of citation status, the subject areas of cited publications are highly similar to the areas of data paper authors. However, the citation analysis indicates that there are considerable citations on the journals specialized on methodology. The network with authors' keywords identifies more detailed areas such as marine ecology, cancer, genome, database, and temperature. This result indicates that biology oriented-subjects are primary areas in the journal although Scientific Data is categorized in multidisciplinary science in Web of Science database.

키워드: 데이터 학술지, 데이터 논문, Scientific Data, 계량분석, 네트워크분석, 동시출현단어분석
data journal, data paper, scientific data, bibliometric analysis, network analysis,
co-word analysis

* 이화여자대학교 사회과학대학 문헌정보학과 교수(echung@ewha.ac.kr)

■ 논문접수일자: 2019년 2월 17일 ■ 최초심사일자: 2019년 3월 19일 ■ 게재확정일자: 2019년 3월 27일
■ 정보관리학회지, 36(1), 117-135, 2019. [http://dx.doi.org/10.3743/KOSIM.2019.36.1.117]

1. 서론

오픈 과학과 데이터 공유는 최근의 인터넷과 정보기술 환경의 발전으로 인해 풍부해진 데이터로 인해 시작된 새로운 논의라기보다는 1980년대부터 연구데이터의 큐레이션과 공유에 대한 필요성으로 제기되어 왔다(Fienberg, Martin, & Straf, 1985). 데이터 보존과 공유는 최근 대량의 축적된 데이터를 통해 기후나 난치병 등의 사회 전체가 직면한 문제를 해결하는데 중심적 역할을 수행하고 있다. 이에 따라 학술 커뮤니케이션은 개방과 공유로 진화하고 있으며, 이로 인해 여러 다양한 이점을 얻을 수 있다. Tenopir, Allard, Douglass, Aydinoglu, Wu, Read, Manoff, Frame(2011)이 제시한 이점은 (1) 데이터의 재분석을 통해 연구결과 검증, (2) 다학제적 관점에서 기존 데이터에 대한 새로운 해석과 접근 제공, (3) 효율적으로 장기 보존된 데이터는 품질을 향상시키는데 도움이 됨, (4) 데이터 공유를 통해 동일하거나 유사한 데이터 수집이 지양될 수 있어 자원 최적화 가능, (5) 데이터 공유와 이용가능성은 데이터 조작과 날조를 방지함, (6) 연구와 실험 복제는 젊은 세대 연구자에게 훈련의 기회를 주는 것 등 구체적으로 살펴볼 수 있다.

그러나 Borgman(2012)이 지적한 바와 같이, 데이터 기반의 사회와 학술 커뮤니티에서 데이터 공유의 문제점은 실제로 데이터 공유라는 행위가 활발하게 일어나지 않는다는 점이다. 오픈 과학과 데이터 공유로 나아가는데 있어서 여러 장애요인은 연구자들의 연구 관행과 문화와 복잡하게 연계되어 있다. 그 중에서도 연구데이터 공유에 대한 적절한 보상체계가 없는 것이라고

지적되었다(Tenopir et al., 2011). 데이터를 공유하는 연구자 입장에서는 적절한 보상 없이 불필요한 시간과 노력을 들여서 다른 연구자가 사용하기에 편리하도록 가공하여 제출하는 행위가 활발하게 일어나기는 어려운 상황이다.

학술 커뮤니케이션 지형에서 데이터의 공유와 재이용을 촉진할 수 있는 중요한 기제로써의 변화가 데이터 학술지의 출간이다. 데이터 학술지는 논문을 수록하는 학술지와 유사한 데이터를 기술한 데이터 논문을 수록하여 발간하며, 데이터 논문은 학술적 데이터 셋 기술, 이용가능성, 인용지원, 데이터 품질, 데이터 공유와 접근을 향상시키는 역할을 한다(Candela, Castelli, Manghi, & Tani, 2015). 데이터 학술지의 출간으로 인해 크게 두 가지 측면에서 데이터 공유와 재이용을 촉진할 수 있다. 첫째는 연구데이터를 제공하여 공유시키고자하는 연구자는 인용이라는 적절한 보상을 얻게 되어 데이터 공유에 적극적으로 참여할 수 있는 동기가 부여된다. 둘째는 학술 커뮤니티에 유통되는 데이터 품질을 향상시킬 수 있다는 점이다. 데이터 논문의 심사와 평가 과정에서 동료평가가 이루어지게 되어, 자연스럽게 데이터 학술지의 데이터 논문으로 발간되는 데이터의 품질이 평가 과정을 거쳐 검증될 수 있다(Costello, Michener, Gahegan, Zhang, & Bourne, 2013; Klump, Bertelmann, Brase, Diepenbroek, Grobe, Höck, Lautenschlager, Schindler, Sens, & Wächter, 2006; Lawrence, Jones, Matthews, Pepler, & Callaghan, 2011).

데이터 학술지가 데이터의 공유와 재이용에 있어서 상당한 중요한 역할을 수행하고 있다. 무엇보다도 저명한 해외 출판사에서 데이터 학술지의

출간이 이루어지고, 활발하게 인용되어 학술지 영향력 평가에 있어서도 상당한 수준으로 자리매김하고 있다. 이 연구는 Nature Publishing Group에서 발간하는 다학제 분야의 데이터 학술지이며 학술지 영향력 평가에서 상당한 순위를 차지하고 있는 Scientific Data 학술지를 대상으로 하였다. 본 연구는 Scientific Data 학술지에 수록된 데이터 논문의 주제, 인용한 저작물, 저자 현황을 분석하여 데이터 학술지의 현황과 특성을 규명하고자 한다.

2. 관련 연구

데이터 학술지에 관련한 연구는 크게 세 그룹으로 구분하여 살펴볼 수 있다. 첫째는 오픈과학 환경에서 데이터의 공유와 재이용을 촉진하기 위하여 데이터 학술지의 중요성을 제시한 연구들이다(Borgman, 2012; Assante, Candela, Castelli, Manghi, Pagano, & Nazionale, 2015). 두 번째 연구 그룹은 데이터 학술지와 함께 병행해야 하는 시스템 체계를 다루었다. 이는 일반 논문을 게재하는 학술지와는 차이에 기반을 둔 것으로서 데이터셋이 저널 내부 혹은 외부 리포지토리 등에 저장되어 장기간 보존되어야 하며 항목 기술을 통해 데이터에 대한 이해를 높이고자 하는 논의가 이루어졌다 (Chard, Pruyne, Blaiszik, Ananthakrishnan, Tuecke, & Foster, 2015; Assante, Candela, Castelli, & Tani, 2016; Amorim, Castro, Da Silva, & Ribeiro, 2017). 세 번째 연구 그룹은 데이터 학술지 발간의 필요성, 구성요소, 특성을 규명 한 연구로써 Costello et al.(2013), Klump et al.

(2006), Lawrence et al.(2011), Candela, Castelli, Manghi, & Tani(2015) 등을 찾아볼 수 있다.

우선, 첫 번째 그룹의 연구는 오픈과학과 연구2.0이라는 새로운 학술 패러다임에서 데이터의 공유, 재이용, 데이터 학술지의 중요성을 다양한 관점에서 논의하였다. Borgman(2012)은 데이터 공유와 재이용으로 얻는 혜택은 분야마다 상이할 수 있지만 대체적으로 1) 연구의 복제나 검증, 2) 공적자원으로 구축된 연구결과물을 공적으로 이용가능하게 함, 3) 기존의 데이터에 대해서 새로운 질문을 하도록 장려함, 4) 연구과 혁신을 발전시킴 등의 네 가지로 설명하고 있다. 그러나 이러한 이점에도 불구하고 데이터의 공유와 재이용에는 여러 가지 난관이 존재하는데 이를 해결하기 위한 방법 중에서 공적자원으로 조성된 데이터를 다른 연구자가 이용할 수 있도록 하는 효과적인 방안으로 데이터 학술지를 언급하였다. 또한 Assante et al.(2015)의 연구는 정보기술이 급속도로 발전하는 연구환경의 기반구조를 분석하고 새로운 학술 커뮤니케이션 워크플로우를 제시하였다. 이러한 논의 가운데서 저자들은 연구데이터의 공유와 재이용의 저해요인은 연구자들이 데이터를 공유하거나 재이용하지 않는 학술 문화적인 요인뿐만 아니라 연구 워크플로우 상의 방법론적 문제점 규명과 해결에 대한 논의를 제시하였다. 특히 연구 워크플로우의 연구관련 저작물에는 데이터가 매우 중요하며, 데이터는 연구 논문과 유사한 방식으로 동료평가와 출판의 과정을 거쳐 다른 연구자들에 의해서 규명되고, 이용되어야 한다고 제시하였다.

두 번째 연구그룹의 논의는 데이터 학술지와 일반 논문을 수록하는 학술지와는 차이에 기반

을 둔다. 즉, 데이터 학술지의 데이터셋이 외부 리퍼지토리 등에 저장되어 장기간 보존되어야 하며, 데이터 논문과 서로 연결되어야 하므로 데이터를 위한 기반 기술체계가 필요하다. 이 그룹의 연구들은 이러한 기반 기술체계에 대하여 논의하였다. Chard et al.(2015)은 점점 규모가 커져가는 연구 데이터를 체계적으로 관리하고 접근하게 하는 것이 중요하며, Globus라는 데이터관리시스템을 제안하였다. 발전하는 정보기술환경에서 데이터관리시스템은 데이터와의 커뮤니케이션을 수행하고, 데이터를 탐색하고, 접근할 수 있는 새로운 매커니즘을 지닌 데이터관리시스템의 필요성을 주장하였다. 저자들이 제안한 데이터관리시스템인 Globus의 주요한 기능으로는 데이터의 규명, 기술, 큐레이트, 검증가능, 접근가능, 보존, 탐색, 브라우저, 검색이라고 제시하였다. 한편 Assante et al.(2016)의 연구와 Amorim et al.(2017)의 연구는 복수개의 데이터 리퍼지토리를 비교 분석하여 특성과 장/단점을 제시하였다. 우선 Amorim et al.(2017)의 연구는 데이터관리의 관점에서 15개의 데이터 리퍼지토리(CKAN, ContentDM, Dataverse, Digital Commons, DSpace, ePrints, EUDAT, Fedora, Figshare, Greenstone, Invenio, Omeka, SciELO, WEKO, Zenodo)를 비교 분석하였다. 데이터관리시스템 측면으로 비교 항목은 크게 5가지로 구분하여 아키텍처, 메타데이터, 상호운용성과 배포, 플랫폼 적용, 데이터 플랫폼이다. 이러한 비교를 통해서 개별 데이터 리퍼지토리가 지닌 특징점을 규명하여 개별기관의 규모, 데이터의 특성, 연구기금기관의 요구사항 등에 따라 선정할 수 있도록 제시하였다. 이와 유사하게 Assante et al.(2016)의

연구도 복수의 데이터 리퍼지토리를 비교 분석하였지만 데이터 출판 관점에서 수행하였다는 점에서 차이를 볼 수 있다. 3TU.Datacentrum, SCIRO DAP, Dryad, Figshare, Zenodo의 총 5개의 데이터 리퍼지토리를 선정하여 비교 분석하였다.

세 번째 그룹의 연구들은 데이터 학술지 발간의 필요성과 이를 위한 구성요소를 규명하고자 하였다. Costello et al.(2013)의 연구는 Biodiversity 분야의 데이터 공유와 재이용에 있어서 가장 큰 걸림돌은 데이터의 품질이라고 제시하였다. 저자들은 이러한 문제점을 극복하기 위한 최선의 방법이 데이터 출판이라고 주장하였다. 데이터 출판을 통해서 품질 점검, 동료평가, 편집위원회의 의사결정 과정들이 자연스럽게 포함되어 궁극적으로 데이터 품질을 향상시킬 수 있다. 향상된 데이터 품질은 데이터 재이용으로 연결되며, 데이터 재이용은 데이터 논문의 인용을 통해 저자에게 크레딧을 부여할 수 있는 장점이 있다. 이러한 환경에서 저자들은 양적으로 팽창하고 있는 연구 데이터 현황에서, 데이터 출판에 관련된 이슈로써 데이터 논문에 대한 동료 평가 과정을 수행하는 것, 오픈 액세스로 출간하는 것, 출판 비용에 관한 것이 주요 쟁점이라고 밝혔다. 특히, Biodiversity 분야 데이터와 데이터 기술에 대한 표준이 표준화된 데이터 관리와 이를 활용한 출판에 앞서 이루어져야 한다고 제시하였다. Klump et al.(2006)의 연구는 데이터 출판 환경을 규명하고 선행요건과 오픈 액세스 구현에 관한 논의를 다루었다. 데이터 출판의 선행요건으로 데이터는 지속적이고 일정한 방식으로 인용 가능해야 한다는 점을 지적했으며, 또 다른 요건은 데이터 품질이

다. 데이터 인용과 품질에 관한 선행요건이 갖추어진 경우 오픈 액세스 방식으로 데이터 학술지의 출간을 구체화시킬 수 있다고 지적하였다. Lawrence et al.(2011)의 연구는 공식적인 데이터 출판과 관련된 논의를 담고 있으며, 특히 동료평가와 데이터 인용에 관한 논의를 중심으로 하였다. 데이터 출간이 학술 커뮤니티에서 중요한 가장 근본적인 체제는 논문 저자뿐만 아니라 데이터 저자에게 인용이라는 적절한 보상체계가 주어지는 것이다. 이와 함께 동료 평가는 데이터의 품질을 보증하고 향상시키는 기제로써 작동한다. 데이터 출간이 포함해야 하는 요소로는 데이터의 인용체계 뿐만 아니라 데이터셋의 입자성, 일시성, 의미 등을 기술하는 체계를 갖추는 것이 필요하다고 제시하였다. 최근에 Candela et al.(2015)은 116건의 데이터 학술지를 조사하여 데이터셋 기술방식, 데이터 이용가능성, 데이터 인용, 데이터 품질, 오픈 액세스를 분석하였다. 저자들은 데이터 학술지가 급격하게 증가하고 있으며, WoS와 같은 인용색인DB에 색인(69.82%)되고 있다고 밝혔다. 2014년부터는 International Symposium on Biomedical Imaging과 같은 국제 학술대회에서 데이터 논문을 제출하도록 요청하는 상황이라고 밝혔다.

지금까지 살펴본 바와 같이 관련 연구들은 오픈과학 패러다임에서 연구데이터의 공유와 재이용이 중요하며 여러 학문 분야의 발전에 기여한다고 보았다. 그러나 데이터의 공유와 재이용을 장려하고 확산하는데 있어서는 현실적인 장애요인이 있으며, 그 중에 하나가 연구자가 데이터를 공유해야만 하는 동기의 부재이다. 이러한 인식에 따라서 연구자가 연구결과물에 대해

서 인정받는 형식인 “인용”이 데이터에도 적용될 필요가 있다. 따라서 데이터 논문과 데이터 학술지 출간은 연구자의 데이터 공유를 확산시키고, 동료평가를 통해 연구결과물의 품질을 보증하고 향상시킬 수 있다. 데이터 학술지의 출간 이후에 데이터 학술지를 대상으로 분석한 연구는 Candela et al.(2015)이 유일하다고 볼 수 있다. 이 연구는 데이터 학술지 출판의 전반적인 지형을 보여주었으나, 데이터 학술지의 저자사항, 데이터 논문의 주제, 데이터 논문이 인용한 저작물의 주제 등과 같이 보다 미시적인 관점의 분석은 부재한 상황이다. 따라서 본 논문의 데이터 학술지의 서지와 인용 데이터 분석을 통하여 데이터 학술지와 데이터 논문을 탐색하고 규명하고자 한다.

3. 데이터 수집과 분석

3.1 데이터 수집

데이터 학술지의 지형을 연구한 Candela et al.(2015)은 15개의 출판사에서 발간한 116건의 데이터 학술지를 대상으로 하였다. 순수하게 데이터 논문만을 다루는 학술지는 7종이며 나머지 109건은 일반 연구 논문과 데이터 논문을 혼재하여 발간하는 것으로 나타났다. Candela et al.(2015)에서 제시한 7종의 순수 데이터 학술지 중에서 Scientific Data는 여러 학문 분야를 다루며, WoS에 색인되어 영향력 있는 학술지이기 때문에 선정하였다. 2014년에 창간되어 2018년 말 현재 총 770논문이 발간되었다. 발간 논문의 수는 2014년 창간호를 시작으로 가

과른 증가 추세를 보여준다. 출판사는 Nature Publishing Group이다. Scientific Data는 JCR에 따르면 다학제 학문 분야(multidisciplinary science)로 분류되어 있으며, 오픈 액세스 형태의 학술지이다. 특히 JCR 2017년 IF에 따르면, 해당 소속 주제 분야의 전체 64종 학술지 중 8위(상위 13%, Q1)로 인용 기반 학술지 영향력에 있어서도 상당한 수준이라고 볼 수 있다. 본 연구의 데이터는 2018년 11월 20일 현재 Web of Science에서 Publication Name으로 Scientific Data로 검색하여 총 713건의 데이터 논문에 대한 서지정보와 인용정보를 수집하여 구축되었다.

3.2 데이터 분석

수집된 총 713건의 데이터에 대하여 저자, 인용, 데이터논문 주제 항목으로 분석을 수행하였다. 첫째, 저자 분석은 두 가지로 구성되었다. 저자의 소속기관/국가와 전공분야에 관한 현황을 파악하였다. 이와 함께 공저자 네트워크 분석을 수행하였다. 데이터 논문 저작활동의 빈도를 사용하여 저자들의 군집을 규명하고자 하였다. 두 번째 분석은 인용에 관한 것으로써 인용된 저작물과 인용된 저자 현황을 우선 파악하였다. 인용된 저작물과 저자의 주제 분야를 WoS의 주제분야를 활용하여 규명하였다. 이를 통하여 데이터 논문이 인용하여 활용하는 주제 분야를 규명하고자 하였다. 셋째, 데이터 논문의 주제 분석을 수행하였다. 데이터 논문에서 추출한 저자 키워드를 활용하여 동시출현

단어 네트워크를 구성하여 데이터 논문의 세부 하위 주제 분야를 규명하고자 하였다. 서지분석과 인용분석을 수행하기 위해서는 Bibexcel¹⁾을 사용하였고, 네트워크 분석과 시각화를 위해서는 이재윤의 WNET(v.0.4.1) 프로그램과 NodeXL²⁾을 사용하였다.

4. 분석결과

4.1 저자 분석

데이터 논문의 출판연도를 살펴보면 <그림 1>과 같다. 2014년 창간 이후 가파르게 증가하는 것을 볼 수 있다. 2014년에는 54건, 2015년에는 78건, 2016년에는 122건, 2017년에는 210건, 2018년에는 249³⁾건으로 각각 나타났다.

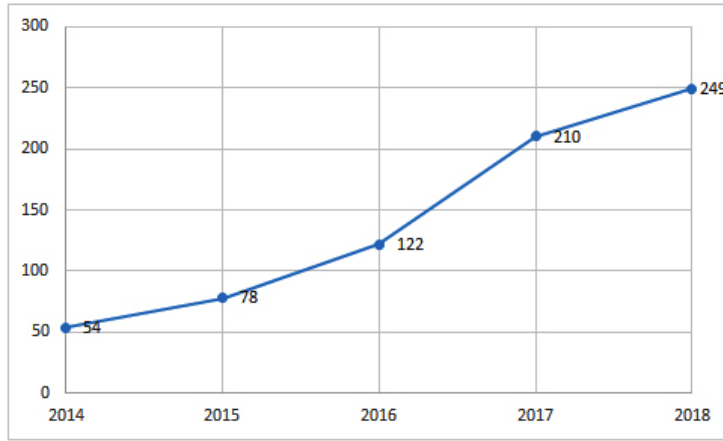
5건 이상의 데이터 논문을 상위 30명의 저자와 출현빈도는 <표 1>과 같이 살펴볼 수 있다. 9건의 논문을 발간한 2명의 저자인 Anton Barty와 Simon I Hay가 가장 활발한 것으로 나타났다. 이후에 총 6건의 논문을 발간한 7명의 저자와 5건의 논문을 발간한 21명의 저자가 상위 30명의 저자로 나타났다.

<표 1>에서 나타난 상위 30명의 저자의 국가, 소속기관, 분야와 전공을 분석하기 위해서 Scientific Data 학술지에 기재된 저자 정보, 저자 홈페이지, 구글 스칼라 연구자 페이지 등을 활용하여 규명하였다. [부록 1]에서와 같이 우선 국가별 분포를 살펴보면, 미국, 일본, 호주가 각각 6회

1) <https://homepage.univie.ac.at/juan.gorraiz/bibexcel/>

2) <https://archive.codeplex.com/?p=nodexl>

3) 2018년도 11월 20일에 검색이 이루어졌기 때문에 해당 일자 기준으로 발간된 데이터논문의 수이다.



〈그림 1〉 연도별 논문 발행 건수

〈표 1〉 논문 발행 건수 상위 30명 저자

순위	저자명	출현빈도	순위	저자명	출현빈도
1	Barty, Anton	9	3	Hayashizaki, Yoshihide	5
1	Hay, Simon I	9	3	Ingleton, Tim	5
2	Boutet, Sebastien	6	3	Kawaji, Hideya	5
2	Carninci, Piero	6	3	Lassmann, Timo	5
2	Fromme, Petra	6	3	Maia, Filipe R. N. C	5
2	Hajdu, Janos	6	3	Persson, Kristin A	5
2	Hassler, Christel	6	3	Pigott, David M	5
2	Kasukawa, Takeya	6	3	Proctor, Roger	5
2	Tatem, Andrew J	6	3	Richardson, Anthony J	5
3	Atkins, Natalia	5	3	Schlichting, Ilme	5
3	Chen, Wei	5	3	Severin, Jessica	5
3	Craddock, R. Cameron	5	3	Shah, Nigam H	5
3	Foucar, Lutz	5	3	Svenda, Martin	5
3	Golding, Nick	5	3	Swadling, Kerrie M	5
3	Hasegawa, Akira	5	3	Williams, Garth J	5

로 나타났으며, 독일이 4회, 스웨덴 3회, 영국이 3회, 스위스와 중국이 각각 1회로 나타났다. 이러한 결과는 기존의 연구 논문 생산성에 대한 북미 중심과 대조되는 것으로 유럽 국가의 강세를 찾아볼 수 있다. 또한 소속기관을 통해서 대학 보다는 연구소 소속 연구자가 집중된 것으로 나타났다. 이러한 경향은 대규모 데이터

셋을 생성할 수 있는 기반환경이 조성된 저명한 연구소에서 데이터 논문의 출간이 상대적으로 용이하기 때문이라고 볼 수 있다.

저자의 전공 현황을 Candela et al.(2015)의 연구결과에서 보여주는 바와 비교해서 살펴볼 수 있다. Candela et al.(2015)은 데이터 학술지의 주제분야 분포를 규명하였는데, 건강과학,

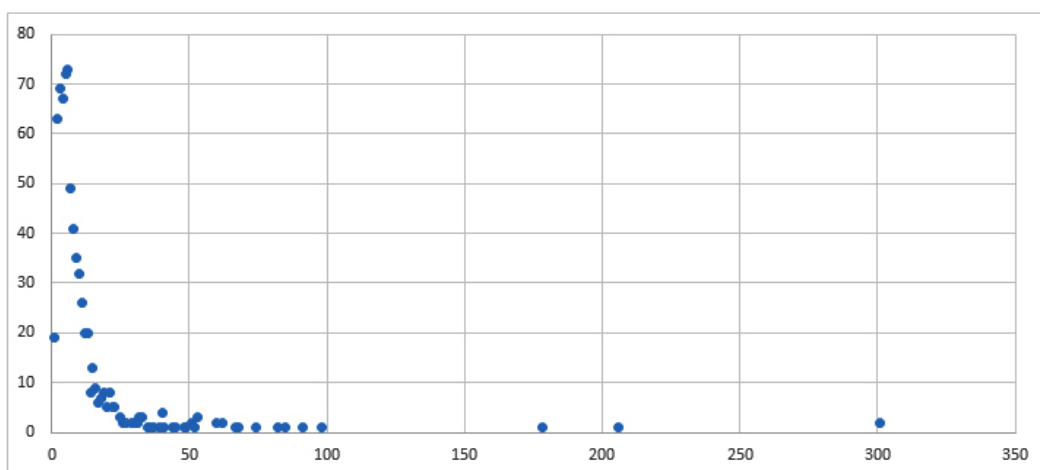
생명과학, 물리학이 41%, 36%, 17%로 주류를 이루었으며, 나머지는 인문학과 사회과학, 복합학으로 분류되었다. 이에 비추어 살펴보면, 저자들의 전공 분야 역시 생명과학과 물리학의 비중이 큰 것으로 나타났다. 데이터 학술지의 주제 분포와의 차이는 건강과학 분야의 비중이 적은 것과 기후학, 환경공학, 해양학 등의 주제 분야의 저자들을 찾아볼 수 있다는 점이다.

한편 논문당 저자 수를 살펴보면, 전체 713건의 논문에 대하여 8,032명의 저자가 참여하였으며, 이 중 고유한 저자의 수는 6,501명이다. 소수 저자가 집중하여 데이터 논문을 발간하기 보다는 다수의 저자가 참여한 것으로 볼 수 있다. 논문별 최대 저자수는 301명이며, 최소 저자수는 1명이다. 논문별 평균 저자수는 약 12명(11.26)으로 나타났다. 저자수 별 분포를 살펴보면, [부록 2]와 <그림 2>에서 제시하는 바와 같다. 가장 많은 건수는 6명이 저자로 참여하는 경우이며 73건, 5명이 저자로 참여하는 경우 72건 등으로 1명에서 20명 사이의 저자

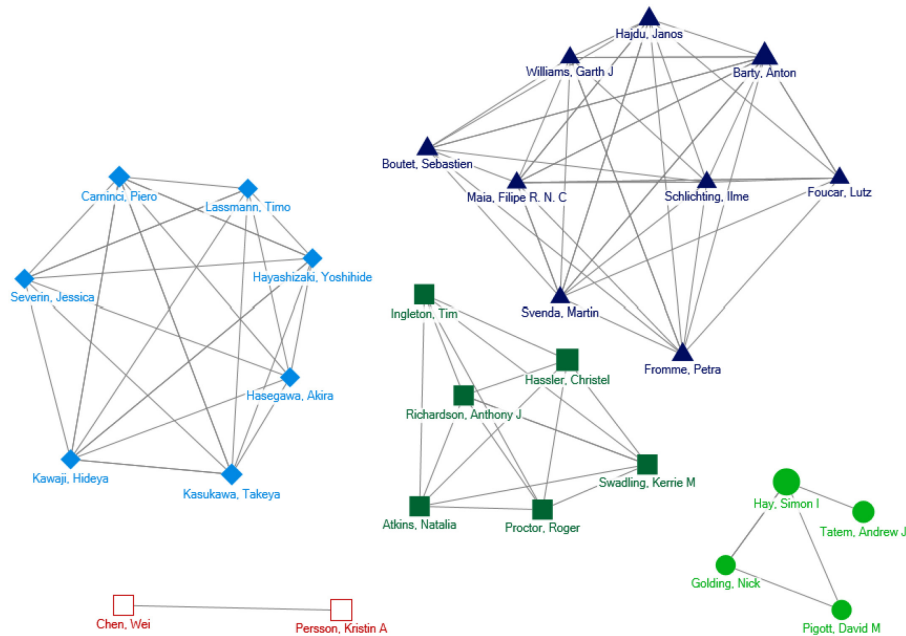
참여가 가장 빈번하게 등장하는 것으로 나타났다. 특기할 만한 사항은 단독 저자의 데이터 논문이 19건으로 [부록]의 순위상으로 13위라는 점이다.

4.2 공저자 분석

데이터 논문을 5회 이상 게재한 저자 28명에 대하여 구현된 공저자 네트워크에서 살펴볼 수 있는 바와 같이 총 5개의 하위네트워크로 구성되었다. 이러한 하위네트워크에 속한 저자들은 공저 작업을 수행했지만, 연결되지 않은 저자들과는 공저 작업을 하지 않은 결과이다. 즉, 데이터 논문의 저자들은 상당히 명확한 공저 그룹이 존재하며, 폭넓게 다양한 공저 활동을 한다고 볼 수 없다. 이러한 현상은 데이터를 생산하는 환경과 밀접하게 연관되어 있다. 데이터 논문의 공저자를 분석하여 저자들 간의 하위네트워크가 형성되었는지를 규명하기 위해서 네트워크로 구현하여 <그림 3>과 같이 나타내었다.



<그림 2> 논문별 저자수 분포



<그림 3> 데이터 논문의 공저자 네트워크(노드의 크기는 공저 빈도수 비례)

데이터 논문의 공저자 네트워크에서 5개의 서브네트워크에 각각 속한 저자들은 <표 2>에서 밝힌 전공분야와 소속기관과 대비하여 살펴볼 수 있다. 우선 가장 큰 규모의 공저 그룹은 9명의 저자가 속해 있는 서브네트워크이며, Barty, A.(물리학, Center for Free-Electron Laser Science), Boutet, S.(물리학, SLAC National Accelerator Laboratory), Foucar, L.(생명과학, Max-Planck-Institut), Fromme, P.(생명과학, Max-Planck-Institut), Hajdu, J.(생명과학, Laboratory of Molecular Biophysics), Maia, F.R.N.C.(생명과학, Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University), Schlichting, I.(생명과학, Max-Planck-Institut), Svenda, M.(생명과학, Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology,

Uppsala University), Williams, G.J.(물리학, NSLS-II, Brookhaven National Laboratory)이다. 이러한 저자들은 물리학과 생명과학 분야의 연구자들로 구성된 공저자 그룹이다. 소속기관으로 살펴보면, 미국의 저명한 물리학 연구소들과 독일의 Max-Planck-Institut, 스웨덴 대학소속 연구소 소속 연구자들의 공동저작 활동으로 볼 수 있다. 두 번째 공저자 연구그룹은 총 7명의 연구자가 속해 있으며, Carninci, P.(생명과학, RIKEN Center for Life Science Technologies, Division of Genomic Technologies), Lassmann, T.(생명과학, Telethon Kids Institute, The University of Western Australia), Hayashizaki, Y.(생명과학, Genomic Sciences Center, Genome Science Laboratory, RIKEN, Tsukuba Life Science), Hasegawa, A.(생명과학, RIKEN Center for Life Science Technologies,

Division of Genomic Technologies), Kasukawa, T.(생명과학, RIKEN Center for Life Science Technologies, Division of Genomic Technologies), Kawaji, H.(생명과학, RIKEN Center for Life Science Technologies, Division of Genomic Technologies), Severin, J.(생명과학, RIKEN Center for Life Science Technologies, Division of Genomic Technologies)이다. 이들 저자들은 전공분야는 모두 생명과학이며, 소속기관으로 살펴보면 일본의 특정기관(RIKEN Center for Life Science Technologies, Division of Genomic Technologies) 소속 5명의 연구자와 호주의 연구기관(Telethon Kids Institute, The University of Western Australia) 1명의 연구자가 지속적으로 데이터 논문을 발간한 것을 파악할 수 있다. 세 번째 공저자 연구그룹 서브네트워크는 총 6명의 연구자가 속해있다. Ingleton, T.(환경공학, Office of Environment and Heritage), Atkins, N.(해양학, Australian Ocean Data Network, Integrated Marine Observing System University of Tasmania), Proctor, R.(해양학, Australian Ocean Data Network), Richardson, A.J.(기후학, CSIRO Marine and Atmospheric Research Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, The University of Queensland), Hassler, C.(환경공학, Department F.-A. Forel for Environmental and Aquatic Sciences University of Geneva), Swadling, K.M.(해양학, Institute of Marine and Antarctic Studies, University of Tasmania)으로 나타났다. 이와 같이 6명의 저자들의 전공은 환경공학, 해양학, 기후학이고, 소속기관은 호주와 스

위스 연구기관들로 나타났다. 네 번째 공저자 그룹 서브네트워크는 4명의 저자로 구성되어 있으며, Hay, S.I.(건강과학, Health Metrics Science University of Washington), Golding, N.(동물학, Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford), Tatem, A.J.(환경공학, Geography and Environmental Science, University of Southampton), Pigott, D.M.(동물학, Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford)이다. 이 그룹의 저자들의 전공과 소속분야가 가장 다양하게 나타났다. 우선 전공 분야는 건강과학, 동물학, 환경공학이며, 소속기관은 미국과 영국의 관련 분야의 저명한 연구기관으로 나타났다. 마지막 다섯 번째 공저자 그룹 서브네트워크는 두 명의 저자로 구성되어 있으며, Chen, W.(동물학, Chengdu Zoo)와 Persson, K.A.(물리학, Department of Materials Science, University of California Berkeley)이다. 이들의 전공분야는 동물학과 물리학이며 소속기관은 중국과 미국의 대학으로 각각 나타났다.

이러한 공저자 네트워크에서 공동저작 활동을 하는 분절된 서브네트워크로 형성되어 나타난 것은 기존의 연구논문의 공저자 네트워크와는 상당한 차이라고 볼 수 있다. 연구논문의 공저자 네트워크는 대체적으로 군집형태로 구성되어 집중적으로 공동저작활동을 하는 연구자 그룹이 나타나지만, 서브네트워크와 같이 다른 그룹과는 단절된 형태는 찾아보기 어렵기 때문이다. 이러한 현상은 연구논문과 달리 데이터 논문의 생산은 대규모 연구집단, 연구장비, 연구비, 분석도구 등과 같은 요인에 영향을 받기

때문이라고 볼 수 있다.

4.3 인용 현황 분석

713건의 데이터 논문에 수록된 참고문헌의 수는 총 25,214건이며 이 중에서 고유한 참고문헌은 6,974건으로 나타났다. 가장 많은 참고문헌의 수는 총 574건을 포함하고 있으며, 가장

적은 수의 참고문헌은 1건으로 나타났다. 논문 당 평균 참고문헌의 수는 35.4건이다.

인용된 저작물 중에서 94회 이상 출현한 상위 20건을 살펴보면, <표 2>과 같다. 빈번하게 인용되는 저작물의 형태와 WoS의 Subject Category로 구분한 주제 분야를 규명하여 제시하였다. 저작물의 형태는 모두 학술지이며, 가장 많이 인용된 1위와 2위 학술지인 Nature와 Science,

<표 2> 인용된 저작물 상위 20건

순위	저작물명	빈도수	형태	WoS 주제분야
1	Nature	620	학술지	Multidisciplinary Sciences
2	Science	482	학술지	Multidisciplinary Sciences
3	Nucleic Acids Research	459	학술지	Biochemistry & Molecular Biology
4	Bioinformatics	457	학술지	Biochemical Research Methods Biotechnology & Applied Microbiology Mathematical & Computational Biology
5	Proceedings of the National Academy of Sciences of the United States of America	444	학술지	Multidisciplinary Sciences
6	Neuroimage	342	학술지	Neurosciences Neuroimaging Radiology, Nuclear Medicine & Medical Imaging
7	Nature Methods	172	학술지	Biochemical Research Methods
8	Nature Biotechnology	160	학술지	Biotechnology & Applied Microbiology
9	Genome Research	145	학술지	Biochemistry & Molecular Biology Biotechnology & Applied Microbiology Genetics & Heredity
10	Nature Genetics	138	학술지	Genetics & Heredity
11	Remote Sensing of Environment	135	학술지	Environmental Sciences Remote Sensing Imaging Science & Photographic Technology
12	Coral Reefs	116	학술지	Marine & Freshwater Biology
13	Journal of Neuroscience	115	학술지	Neurosciences
14	International Journal of Climatology	113	학술지	Meteorology & Atmospheric Sciences
15	Journal of Climate	112	학술지	Meteorology & Atmospheric Sciences
16	Marine Ecology Progress Series	101	학술지	Ecology Marine & Freshwater Biology Oceanography
17	Physical Review B	98	학술지	Physics, Condensed Matter
17	Bulletin of the American Meteorological Society	98	학술지	Meteorology & Atmospheric Sciences
19	Global Change Biology	94	학술지	Biology, Miscellaneous Ecology Environmental Sciences
19	Cell	94	학술지	Biochemistry & Molecular Biology Cell Biology

5위인 Proceedings of the National Academy of Sciences of the United States of America는 다학제과학 분야로 나타났다. 이 세 학술지를 제외한 다른 학술지는 특정 주제 분야로 분류되어 있다. 생화학, 미생물학, 수리생물학, 분자생물학, 생명공학, 신경과학, 유전학, 환경공학, 해양학, 기상학, 물리학, 생태학 등의 학문 분야의 학술지가 높은 빈도로 인용된 것을 확인할 수 있다. 이러한 분야는 인용한 데이터 논문의 분야와 크게 다르지 않다는 것을 보여준다. 그러나 한 가지 특기할 만한 사항으로 Biochemical Research Methods와 Bioinformatics와 같은 저널로 방법론을 주로 다루는 학술지가 상위에 오

른 점이다. 이러한 현상은 데이터 논문이 데이터 수집과 분석에 대한 집중적인 논의를 다루기 때문인 것으로 볼 수 있다.

한편 인용된 참고문헌에서 저자를 추출하여 빈도수에 기반으로 하여 상위에 오른 21명의 저자들을 살펴보면 <표 3>과 같다. 우선 <표 1> 저자들과 비교하면, 상당히 다른 구성임을 알 수 있다. 이를 통하여 데이터 논문의 저자가 인용되기 보다는 연구논문의 저자가 인용되는 상황을 알 수 있다. 저자들의 주제분야를 살펴보면, 데이터 논문의 저자들의 주제 분야와 크게 다르지 않게 나타났다. 저자들은 주로 생명공학이며, 물리학, 재료공학, 기후학, 건강학, 재

<표 3> 인용된 저자 상위 21명

순위	저작물명	빈도수	주제분야
1	Li, H	63	생명공학
2	Williams, S. P.	49	생화학
3	Trapnell, C	39	생명공학
4	Associacao Nacional para Difusao de Adubos	31	복합주제
4	Kresse, G	31	물리학
6	Langmead, B	30	생명공학
7	Billler, S. J.	29	생명공학
8	United Nations	26	복합주제
9	Jain, A	26	재료공학
10	Hijmans, RJ	25	기후학
11	Funk, C	23	지구과학
11	Tatem, AJ	22	건강과학
11	Anders, S	22	생명공학
14	Robinson, MD	22	생명공학
15	Gorgolewski, KJ	20	뇌과학
16	Ong, SP	19	재료공학
16	Perdew, JP	19	물리학
16	Zuo, XN	19	뇌과학
16	Wilkinson, MD	19	생명공학
20	Dobin, A	18	생명공학
20	Quinlan, AR	18	생명공학

료공학, 뇌과학 등이 소수의 주제분야로 나타났다. 기관 저자로는 Associacao Nacional para Difusao de Adubos(브라질 국립 비료 협회)와 United Nations(유엔)를 찾아볼 수 있다. 유엔의 경우는 유엔에서 발행한 인구, 보건, 기후 등에 관련된 보고서가 많이 인용된 것으로 나타났다. 따라서 이 두 기관에서 발행한 데이터와 보고서가 지속적으로 인용된 것으로 볼 수 있다. 또 한 가지 특기할 만한 사항은 16위에 오른 Wilkinson의 인용은 주로 Scientific Data 학술지에 실린 데이터 논문이다. 논문의 제목은 “The FAIR Guiding Principles for scientific data management and stewardship”으로 데이터 관리에 관한 가이드라인을 제시하였다.

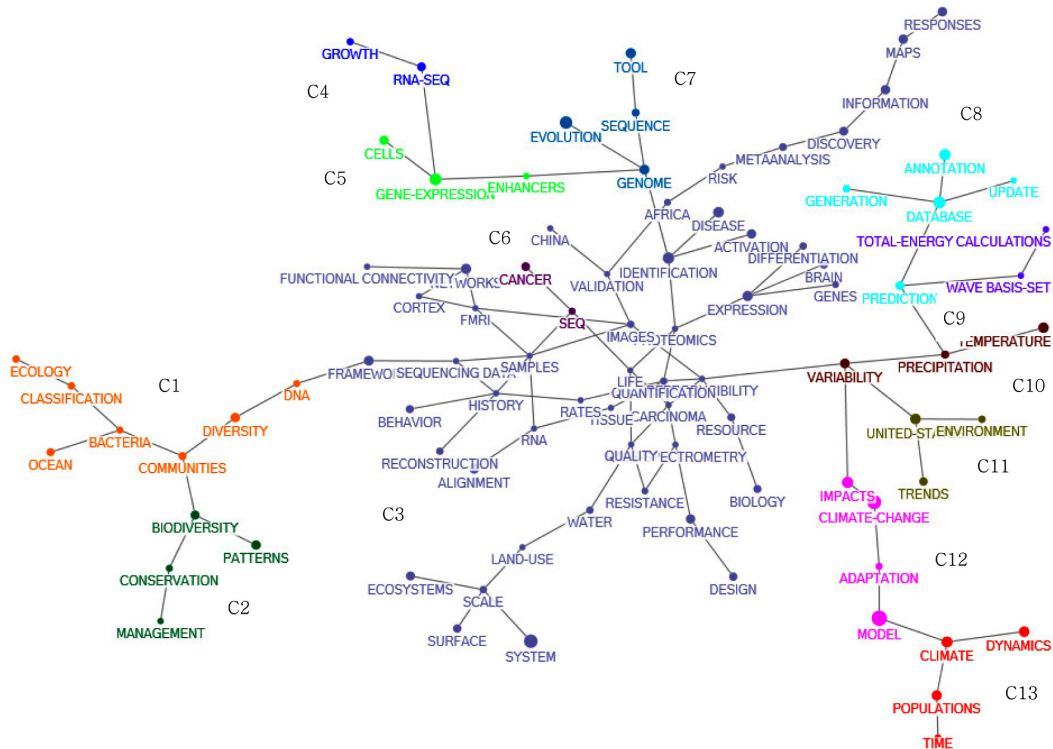
4.4 데이터 논문 주제 분석

데이터 논문의 주제를 분석하기 위하여 5회 이상 출현한 논문 저자 키워드 92개의 단어를 추출하여 네트워크로 구성하였다. 총 13개의 군집으로 형성되었으며, <표 4>로 군집과 소속 키워드를 제시하였고, <그림 4>와 같이 네트워크로 시각화하였다.

<그림 4>에서 제시하는 바와 같이 가장 큰 규모의 군집으로 C3군집을 확인할 수 있다. C3 군집은 규모가 큰 군집답게 생물학의 다양한 주제 분야가 다루어지고 있으며, 중심에 위치하고 있으며, 직접적으로 5개의 군집과 연결된 것으로 나타났다. 우선 왼쪽으로는 해양생태를 다루는 C1군집과 연결되어 있으며, 규모가 작

<표 4> 데이터 논문 저자 키워드 군집의 주제

군집	소속 키워드	중심주제
C1	BACTERIA, CLASSIFICATION, COMMUNITIES, DIVERSITY, DNA, ECOLOGY, OCEAN	해양생태
C2	BIODIVERSITY, CONSERVATION, MANAGEMENT, ATTERNS	생물다양성
C3	ACTIVATION, AFRICA, ALIGNMENT, BEHAVIOR, BIOLOGY, BRAIN, CARCINOMA, CHINA, CORTEX, DESIGN, DIFFERENTIATION, DISCOVERY, DISEASE, ECOSYSTEMS, EXPRESSION, FMRI, FRAMEWORK, FUNCTIONAL, CONNECTIVITY, GENES, HISTORY, IDENTIFICATION, IMAGES, INFORMATION, LAND-USE, LIFE, MAPS, MASS-SPECTROMETRY, METAANALYSIS, NETWORKS, PERFORMANCE, PROTEOMICS, QUALITY, QUANTIFICATION RATES, RECONSTRUCTION, REPRODUCIBILITY, RESISTANCE, RESOURCE, RESPONSES, RISK, RNA, SAMPLES, SCALE, SEQUENCING DATA, SURFACE, SYSTEM TISSUE, VALIDATION, WATER	생물학
C4	GROWTH, RNA-SEQ	RNA시퀀스
C5	CELLS, ENHANCERS, GENE-EXPRESSION	세포
C6	CANCER, SEQ	암
C7	EVOLUTION, GENOME, SEQUENCE, TOOL	계놈
C8	ANNOTATION, DATABASE, GENERATION, PREDICTION, UPDATE	데이터베이스
C9	TOTAL-ENERGY CALCULATIONS, WAVE BASIS-SET	에너지
C10	PRECIPITATION, TEMPERATURE, VARIABILITY	기온
C11	ENVIRONMENT, TRENDS, UNITED-STATES	환경
C12	ADAPTATION, CLIMATE-CHANGE, IMPACTS, MODEL	기후변화
C13	CLIMATE, DYNAMICS, POPULATIONS, TIME	기후



<그림 4> 데이터 논문 저자 키워드 네트워크와 13 군집(노드의 크기는 출현빈도 비례)

은 암을 주제 분야로 다루는 C6군집과 연결되어 있다. 또한 C7 군집 게놈 주제 분야와도 직접적으로 연결되어 있다. 오른쪽으로는 C8 데이터베이스 주제 분야와 연결되어 있으며, C10 기온 주제 분야와도 연결되어 있는 것으로 나타났다. 중심적인 주제분야인 C3군집과 직접적으로 연결되어 있지는 않지만, 외곽에 위치한 C2 주제분야 군집을 찾아볼 수 있다. 또한 C7 군집 게놈 주제 분야와 연결된 C5 군집 세포 분야와 C5와 연결된 C4 RNA시퀀스 군집 주제분야를 확인할 수 있다. C10 기온 주제분야 군집과 연결된 군집은 C11 환경 주제분야와 C12 기후변화 군집으로 나타났다. C12군집과 연결된 주제분야 군집은 유사한 주제인 C13 기

후 군집으로 확인할 수 있다. 이외에도 C9 군집인 에너지 주제분야는 C8군집인 데이터베이스 군집과 연결된 것으로 나타났다. 이러한 데이터 논문의 저자 키워드를 기반으로 한 네트워크와 군집 구현을 통해서 데이터 저널에 출간된 데이터 논문의 주제적 구성 현황을 살펴볼 수 있다.

5. 결론과 제언

데이터 공유와 재이용은 오픈 과학 패러다임에서 중요한 구성요소이다. 그럼에도 불구하고 데이터 공유와 재이용이 활성화되지 못하는 요

인 중의 하나는 연구자 스스로 연구데이터를 공유하고자 하는 동기부여가 부족하기 때문이다. 데이터 논문과 데이터 학술지 출판은 데이터의 공유에 대한 적절한 보상체계인 인용을 제공할 수 있기 때문에 연구자가 데이터를 공유하고자 하는 동기를 부여할 수 있는 기제이다. 본 논문은 이러한 상황에서 데이터 학술지와 논문에 대한 현황을 규명하고자 하였다. 이를 위하여 WoS에 색인되어 있으며, 다학제 분야의 영향력 있는 데이터 학술지인 Scientific Data를 대상으로 하여, 저자, 인용, 주제적인 측면에서 분석하였다.

우선 저자 측면에서 살펴본 결과, 생명과학, 물리학 등의 주제 분야의 저자가 주류를 이룬 것으로 나타났다. 저자의 수 측면에서 살펴보면, 최대 301명에서 최소 1명으로 나타났으며, 평균 저자 수는 약 12명(11.26)이다. 비중이 높은 저자 수는 5명에서 6명으로 확인할 수 있다. 특기할 만한 사항으로는 단독저자 논문의 비중도 상당하다는 점이다. 또한 공저자 네트워크에서 살펴보면, 관련 분야의 연구자들이 타 기관 소속 연구자들과의 공동저작 활동을 확인할 수 있다. 특히 생명공학자와 물리학자의 공저자 활동이 두드러지게 나타났다. 데이터 논문의 공저자 네트워크 상의 특색 있는 사항으로는 분절된 서브네트워크의 구성이라고 볼 수 있다. 공저 활동이 주로 폐쇄적인 소규모 연구자 그룹 사이에서 주로 일어났으며, 다른 연구 그룹과의 연결이 나타나지 않았다. 이러한 결과는 데이터 논문이 보여주는 특징이라고 할 수 있다.

두 번째로 인용 측면에서 살펴보면, 인용된 저작물과 인용된 저자의 주제 분야 분석이 수

행되었다. 인용된 저작물의 특성을 살펴보면, 우선 다학제적인 학술지의 비중이 높은 것을 확인할 수 있다. Nature, Science, Proceedings of the National Academy of Sciences of the United States of America의 3종의 학술지가 상위에 나타났다. 다학제적인 학술지 이외에 생명공학, 물리학 등 주요 저자들의 주제 분야와 유사한 형태를 보이는 것으로 나타났다. 특기할 만한 사항으로는 방법론을 주로 다루는 학술지의 인용 비중이 상당히 높은 것을 확인할 수 있다. 이러한 결과는 데이터 논문의 구성 과정에서 데이터 수집과 분석 등에 관한 기술에 있어서 인용 수요가 많은 것으로 볼 수 있다. 인용된 저자의 주제 분야를 살펴보면, 데이터 논문 저자의 주제 분야와 유사한 형태로 나타났다.

세 번째는 데이터 학술지와 논문의 주제를 세밀하게 규명할 수 있는 방식으로써 저자 키워드 동시출현분석 네트워크를 구현하였다. 키워드 동시출현분석 네트워크를 통해서 총 13개 군집의 하위 주제 영역을 규명하였다. 가장 규모가 크고 중심이 되는 주제 분야는 생물학 분야이며, 이와 직접적으로 연결되는 주제 분야 군집으로는 해양생태, 암, 계놈, 데이터베이스, 기온 주제 분야로 나타났다. 이러한 세부적인 주제 분야의 규명은 데이터 학술지와 데이터 논문의 발간을 통해 형성된 지적 구조라고 볼 수 있다. 이러한 분석결과를 통해, 다학제적 분야의 데이터 학술지와 논문을 대상으로 하였지만, 데이터 논문과 학술지 출판에 대한 논의는 생명공학 학문분야의 생물다양성 분야가 주류(Chavan & Penev, 2011)라고 지적한 바와 같이 생물학이나 생명공학 분야의 집중화 현상을

찾아볼 수 있다.

본 연구 결과는 오픈 과학의 중심 영역인 데이터 학술지와 데이터 논문의 현황을 제시하였다. 이를 통해 주요 데이터 학술지와 논문이 형성한 주제 영역, 저자 구성, 인용을 규명하였다. 이러한 분석 결과는 데이터 학술지라는 새로운 형태의 학술 생태계를 파악하는데 기여하였다.

본 연구의 후속 연구로는 보다 다양한 주제 분야의 데이터 학술지와 데이터 논문을 분석하여, 분야별 특성을 규명하는 것이 필요하다. 데이터 학술지의 발간이 이루어지고 영향력이 커지고 있는 상황에서, 이러한 연구는 주제 분야의 특성이 반영된 데이터 학술지와 논문 발간을 위한 학술 생태계를 파악하는데 도움이 될 것이다.

참 고 문 헌

이재윤. WNET v.0.4.1 소프트웨어.

Amorim, R. C., Castro, J. A., Da Silva, J. R., & Ribeiro, C. (2017). A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, 16(4), 851-862.

<https://doi.org/10.1007/s10209-016-0475-y>

Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are scientific data repositories coping with research data publishing? *Data Science Journal*, 15.

<https://doi.org/10.5334/dsj-2016-006>

Assante, M., Candela, L., Castelli, D., Manghi, P., Pagano, P., & Nazionale, C. (2015). Science 2.0 repositories: Time for a change in scholarly communication. *D-Lib Magazine*, 21(1/2), 1-14. <https://doi.org/10.1045/january2015-assante>

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.

<https://doi.org/10.1002/asi.22634>

Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747-1762.

<https://doi.org/10.1002/asi.23358>

Chard, K., Pruyne, J., Blaiszik, B., Ananthakrishnan, R., Tuecke, S., & Foster, I. (2015, August). Globus data publication as a service: Lowering barriers to reproducible science. In *e-Science (e-Science)*, 2015 IEEE 11th International Conference on (pp. 401-410). IEEE.

Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing

- in biodiversity science. *BMC Bioinformatics*, 12(15), S2.
<https://doi.org/10.1186/1471-2105-12-s15-s2>
- Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z. Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, 28(8), 454-461. <https://doi.org/10.1016/j.tree.2013.05.002>
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). *Sharing research data*. National Academy Press.
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., ..., & Wächter, J. (2006). Data publication in the open access initiative. *Data Science Journal*, 5, 79-83. <https://doi.org/10.2481/dsj.5.79>
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*, 6(2), 4-37. <https://doi.org/10.2218/ijdc.v6i2.205>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., ..., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6(6): e21101. <https://doi.org/10.1371/journal.pone.0021101>

[부록 1] 출현빈도 상위 30명의 국가, 기관, 전공/분야

번호	저자명	국가	소속기관	분야/전공
1	Barty, Anton	독일	Center for Free-Electron Laser Science	물리학
2	Hay, Simon I	미국	Health Metrics Science University of Washington	건강과학
3	Boutet, Sebastien	미국	SLAC National Accelerator Laboratory	물리학
4	Carninci, Piero	일본	RIKEN Center for Life Science Technologies, Division of Genomic Technologies	생명과학
5	Fromme, Petra	독일	Max-Planck-Institut	생명과학
6	Hajdu, Janos	스웨덴	Laboratory of Molecular Biophysics	생명과학
7	Hassler, Christel	스위스	Department F.-A. Forel for Environmental and Aquatic Sciences University of Geneva	환경공학
8	Kasukawa, Takeya	일본	RIKEN Center for Life Science Technologies, Division of Genomic Technologies	생명과학
9	Tatem, Andrew J	영국	Geography and Environmental Science, University of Southampton	환경공학
10	Atkins, Natalia	호주	Australian Ocean Data Network, Integrated Marine Observing System University of Tasmania	해양학
11	Chen, Wei	중국	Chengdu Zoo	동물학
12	Craddock, R. Cameron	미국	Center for the Developing Brain, Child Mind Institute Center for Biomedical Imaging and Neuromodulation Psychiatric Research	의학
13	Foucar, Lutz	독일	Max-Planck-Institut	생명과학
14	Golding, Nick	영국	Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford	동물학
15	Hasegawa, Akira	일본	RIKEN Center for Life Science Technologies, Division of Genomic Technologies	생명과학
16	Hayashizaki, Yoshihide	일본	Genomic Sciences Center, Genome Science Laboratory, RIKEN, Tsukuba Life Science	생명과학
17	Ingleton, Tim	호주	Office of Environment and Heritage	환경공학
18	Kawaji, Hideya	일본	RIKEN Center for Life Science Technologies, Division of Genomic Technologies	생명과학
19	Lassmann, Timo	호주	Telethon Kids Institute, The University of Western Australia	생명과학
20	Maia, Filipe R. N. C	스웨덴	Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University	생명과학
21	Persson, Kristin A	미국	Department of Materials Science, University of California Berkeley	물리학
22	Pigott, David M	영국	Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford	동물학
23	Proctor, Roger	호주	Australian Ocean Data Network	해양학
24	Richardson, Anthony J	호주	CSIRO Marine and Atmospheric Research Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, The University of Queensland	기후학
25	Schlichting, Ilme	독일	Max-Planck-Institut	물리학
26	Severin, Jessica	일본	RIKEN Center for Life Science Technologies, Division of Genomic Technologies	생명과학
27	Shah, Nigam H	미국	Stanford University Center for Bioinformatics Research	생명과학
28	Svenda, Martin	스웨덴	Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University	생명과학
29	Swadling, Kerrie M	호주	Institute of Marine and Antarctic Studies, University of Tasmania	해양학
30	Williams, Garth J	미국	NSLS-II, Brookhaven National Laboratory	물리학

[부록 2] 저자수별 논문 건수

순위	저자수	논문건수	순위	저자수	논문건수
1	6	73	29	26	2
2	5	72	29	27	2
3	3	69	29	29	2
4	4	67	29	30	2
5	2	63	29	31	2
6	7	49	29	51	2
7	8	41	29	60	2
8	9	35	29	62	2
9	10	32	29	301	2
10	11	26	38	35	1
11	12	20	38	36	1
11	13	20	38	37	1
13	1	19	38	39	1
14	15	13	38	41	1
15	16	9	38	44	1
16	14	8	38	45	1
16	19	8	38	48	1
16	21	8	38	49	1
19	18	7	38	52	1
20	17	6	38	67	1
21	20	5	38	68	1
21	22	5	38	74	1
21	23	5	38	82	1
24	40	4	38	85	1
25	25	3	38	91	1
25	32	3	38	98	1
25	33	3	38	178	1
25	53	3	38	206	1

