

Web of Science 데이터학술지 게재 데이터논문의 지적구조 규명

An Investigation of Intellectual Structure on Data Papers Published in Data Journals in Web of Science

정은경 (EunKyung Chung)*

초 록

오픈과학의 흐름에서 데이터 공유와 재이용은 중요한 연구자의 활동이 되어가고 있다. 데이터 공유와 재이용에 관한 여러 논의 중에서 데이터학술지와 데이터논문의 발간이 가시적인 결과를 보여주고 있다. 데이터학술지는 여러 학문 분야에서 발간되고 있으며, 논문의 수도 점차 증가하고 있다. 데이터논문은 데이터 자체와는 다르게 인용을 주고 받는 활동이 포함되어, 따라서 이들이 형성하는 고유한 지적구조가 생겨나게 된다. 본 연구는 데이터학술지와 데이터논문이 학술커뮤니티에서 구성하는 지적구조를 규명하고자 Web of Science에 색인된 14종의 데이터학술지와 6,086건의 데이터논문과 인용된 참고문헌 84,908건을 분석하였다. 저자사항과 함께 동시인용분석과 서지결합분석을 네트워크로 시각화하여 데이터논문이 형성한 세부 주제 분야를 규명하였다. 분석결과, 저자, 저자소속기관, 국가를 추출하여 출현빈도를 살펴보면, 전통적인 학술지 논문과 다른 양상을 보인다. 이러한 결과는 데이터의 생산이 용이한 기관과 국가에 주로 데이터논문을 출간하기 때문이라고 해석될 수 있다. 동시인용분석과 서지결합분석 모두 분석도구, 데이터베이스, 게놈구성 등이 주된 세부 주제 영역으로 나타났다. 동시인용분석결과는 9개의 군집으로 형성되었는데, 특정 주제 분야로 나타난 영역은 수질과 기후 등의 분야이다. 서지결합분석은 총 27개의 컴포넌트로 구성되었는데, 수질, 기후 이 외에도 해양, 대기 등의 세부 주제 영역이 파악되었다. 특기할만한 사항으로는 사회과학 분야의 주제 영역도 나타났다는 점이다.

ABSTRACT

In the context of open science, data sharing and reuse are becoming important researchers' activities. Among the discussions about data sharing and reuse, data journals and data papers shows visible results. Data journals are published in many academic fields, and the number of papers is increasing. Unlike the data itself, data papers contain activities that cite and receive citations, thus creating their own intellectual structures. This study analyzed 14 data journals indexed by Web of Science, 6,086 data papers and 84,908 cited references to examine the intellectual structure of data journals and data papers in academic community. Along with the author's details, the co-citation analysis and bibliographic coupling analysis were visualized in network to identify the detailed subject areas. The results of the analysis show that the frequent authors, affiliated institutions, and countries are different from that of traditional journal papers. These results can be interpreted as mainly because the authors who can easily produce data publish data papers. In both co-citation and bibliographic analysis, analytical tools, databases, and genome composition were the main subtopic areas. The co-citation analysis resulted in nine clusters, with specific subject areas being water quality and climate. The bibliographic analysis consisted of a total of 27 components, and detailed subject areas such as ocean and atmosphere were identified in addition to water quality and climate. Notably, the subject areas of the social sciences have also emerged.

키워드: 데이터학술지, 데이터논문, 인용분석, 네트워크, 동시인용분석, 서지결합분석
data journal, data paper, citation analysis, network, co-citation analysis, bibliographic coupling analysis

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

- 논문접수일자: 2020년 2월 25일 ■ 최초심사일자: 2020년 3월 5일 ■ 게재확정일자: 2020년 3월 21일
- 정보관리학회지, 37(1), 153-177, 2020. <http://dx.doi.org/10.3743/KOSIM.2020.37.1.153>

※ Copyright © 2020 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

데이터 스칼라십(Borgman, 2015)은 오픈과학의 중요한 연구자 활동으로 여겨진다. 데이터 스칼라십은 연구과정에서 발생하는 데이터가 다른 연구자에게 공유되어 재이용되는 선순환 과정에서 중요한 단계이다. McKiernan 외(2016)는 데이터와 같은 연구성과물을 공개하고 재이용하게 하는 활동이 실제로 연구자에게 유익한 활동이라는 점을 제시하였다. 예를 들면, 연구자가 연구에 사용된 데이터를 공유할 때 그렇지 않은 경우와 비교하여 논문 인용이 분야에 따라 상이하지만 9%에서 35% 정도 상승한다고 보고하였다. Belter(2014)는 적절하게 기탁되고 공유된 데이터에 대한 인용을 살펴보면, 관련 학술지 논문과 비교해 인용이 99% 상승한다고 보고하였다. 이러한 연구들이 공통으로 제시하는 사항은 데이터가 연구성과물의 부산물이 아니라 인용과 같은 학계가 인정하는 방식으로 연구성과로 포함될 필요성이다. 이러한 논의 과정에서 데이터학술지와 데이터논문이 여러 분야에서 시작되었으며, 분야도 확대되어 발행 학술지와 논문이 점차 증가하고 있다. Candela, Castelli, Manghi, Tani(2015)가 밝힌 바와 같이, 데이터학술지마다 조금씩 상이한 양상을 보이고 있지만, 데이터논문에서 공통적으로 제공하는 데이터 기술항목은 10가지 정보이다. DOI 혹은 URI와 같은 접근점을 제공하는 이용가능성, 데이터 셋 관련한 경쟁관계의 이해관계, 시공간적 데이터의 범위, 실제적인 데이터 재이용을 촉진할 수 있는 요소인 형태로써 인코딩이나 언어 등의 정보, 데이터 이용을 권장하는 정책 정보를 제공하는 라

이션스, 데이터논문의 저자 기여도 표기, 데이터 생산을 주도한 프로젝트 정보, 데이터 생산을 이끈 방법론이나 도구에 대한 정보 제공, 데이터의 품질에 대한 정보, 데이터 재이용을 촉진하는 정보로 구성되었다.

무엇보다도 Silvello(2017)가 지적한 바와 같이 데이터학술지와 데이터논문은 데이터 자체와 구분되는 고유한 특성이 존재한다. 데이터 자체는 인용을 받을 수는 있지만, 다른 논문이나 데이터를 참고문헌을 사용하여 인용을 줄 수 있는 기제가 없다. 그러나 데이터논문은 데이터 그 자체로써는 지닐 수 없는 참고문헌을 통해 다른 학술 성과물에 대하여 인용을 제공한다. 이러한 특성은 데이터학술지와 데이터논문이 형성하는 주제 영역에 관한 논의를 가능하게 한다. 본 논문은 『Scientific Data』 학술지 한 종만 분석한 정은경(2019)의 후속 연구 성격으로 Web of Science에 색인된 14종의 데이터학술지와 이에 게재된 6,086건의 데이터논문과 인용된 참고문헌 84,908건을 대상으로 분석하였다. 전통적인 학술지 논문과 같이 참고문헌을 지니는 데이터논문의 특성을 활용하여 인용을 기반으로 하는 분석인 서지결합분석과 동시인용분석을 수행하였다. 두 분석기법을 활용하여 네트워크로 시각화하여 데이터학술지와 데이터논문이 형성하는 상세 주제 영역을 제시하여 지적구조를 규명하고자 하였다.

2. 관련 연구

오픈과학의 기치 아래 데이터학술지와 데이터논문의 태생은 데이터라는 연구저작물에 대

한 공유와 재이용 필요성에 기인한다. 연구데이터의 공유와 재이용이 학술 커뮤니티에서 보편화하는 데는 여러 가지 어려움에 직면해왔다. 연구자가 데이터를 다른 연구자가 사용할 수 있도록 추가적인 시간과 노력을 들여 정제화하고 메타데이터를 작성하는 등의 일이 보편화하기 위해서는 Gorgolewski, Margulies, Milham(2013)이 지적한 바와 같이 적절한 학술적 보상 기제가 필요하다. 연구데이터를 데이터논문으로 출간하여 저자에게 그에 따른 학술적 인정인 인용이 연계되는 것이 가장 효과적인 연구자 보상 기제라고 볼 수 있다. 이러한 논의의 흐름이 이어져서 데이터논문과 데이터학술지의 출간이 지속해서 증가하는 추세이다. 본 연구에서는 데이터학술지와 데이터 논문 관련 연구를 주요한 이해관계자인 연구자와 데이터 관리 측면으로 구분하여 살펴보고자 한다.

우선 연구자의 관점으로 Gorgolewski, Margulies, Milham(2013), Huang, Hawkins, Qiao(2013), Chavan과 Penev(2011), McKiernan 외(2016), Costello 외(2013) 등의 연구를 살펴볼 수 있다. Gorgolewski, Margulies, Milham(2013)은 신경 영상 분야에서 다른 학술 연구 커뮤니티에서 수행하는 것과 같이 데이터논문 출간이 필요하다고 지적하였다. 이를 위해서 신경 영상 분야의 데이터논문이 갖추어야 하는 요소로는 최소 6가지 항목이 필요하다고 제시하였다. 우선, 전체 연구 개요, 연구 참여자 사항, 실험 설계, 표현형 평가 프로토콜, 스캔 상세 정보, 연구데이터 배포 정보 등으로 구성되어야 한다고 제시하였다. 이와 함께 데이터논문의 출간을 통해서 해당 연구자가 데이터의 품질을 심사할 수 있는 논문 심사 과정이 포함되기 때문에 질

좋은 연구데이터의 공유와 재이용에 있어서 질적 향상을 가져올 수 있다는 점을 순기능으로 지적하였다. 또한 생명다양성 분야에서 Chavan과 Penev(2011)도 연구데이터 공유와 재이용에 있어서 가장 큰 걸림돌은 연구자 스스로 추가적인 노력과 시간을 들여 다른 사람들이 이용할 수 있는 데이터로 제공하는 데 있어서 별다른 동기부여가 없다는 점을 지적하였다. Chavan과 Penev은 데이터학술지와 데이터논문이 궁극적으로 연구데이터의 공유와 재이용을 촉진하는 중요한 역할을 수행하게 된다고 주장하였다. 연구데이터를 데이터논문으로 발간할 때의 장점으로 네 가지 사항을 꼽았다. (1) 연구데이터에 대한 전통적인 학술 커뮤니티 방식의 등록이 가능하게 된다. (2) 연구데이터의 색인과 인용이 가능하게 된다. (3) 지속적인 연구데이터 이용이 인용의 방식으로 추적될 수 있다. (4) 연구데이터와 함께 출판된 데이터에 대한 메타데이터가 지속적인 접근과 설명을 제공할 수 있다. 그러나 현재 상태의 데이터논문은 고유 식별자를 부여하고, 통제 어휘를 사용하여 상호운용성을 높인 개선된 메타데이터를 사용해야 하는 등의 과제를 동시에 안고 있다고 지적하였다. 마지막으로 생물다양성 분야의 데이터학술지와 데이터논문이 주류적인 학술 활동이 되기 위해서는 연구자 평가시스템과 연구기금 기관에서 데이터논문에 대한 의무화를 추진하는 것이 필요하다고 제시하였다. Costello 외(2013) 또한 데이터학술지와 데이터논문 발간이 연구데이터의 공유와 재이용에 있는 중요한 기폭제가 된다고 주장하였다. 전통적인 학술지 논문과 같은 동료 심사제를 통해 데이터논문의 품질을 향상시키는 것이 필요하다고 제시하였다.

데이터논문의 동료 평가시에 사용할 수 있는 평가요소 중에서 몇가지 중요한 사항을 밝혔다. 가장 중요한 사항은 데이터 오류가 평가 과정에서 밝혀지고 수정되어야 한다는 점이다. 또한 연구데이터와 함께 보고되는 메타데이터는 기술적인 점검 과정과 동료 평가 과정에서 검증되어 이용 가능해야 한다고 지적하였다. 그러나 Huang, Hawkins, Qiao(2013) 역시 생물 다양성 분야의 데이터 공유를 위해서 데이터학술지와 데이터논문에 관한 논의를 연구자의 연구데이터 제공에 대한 인용과 같은 인정 효과와 연구데이터의 품질이 향상되는 점에는 동의하였다. 그러나 데이터논문 발간을 통해 연구데이터의 공유와 재이용에 있어서 함께 발생할 수 있는 문제점을 몇가지 지적했다. 우선은 연구논문 발간시에 최근의 학술지 정책이나 연구기금 정책에 따라 관련된 데이터를 리포지토리 등에 기탁하게 되어 있기 때문에 연구자 입장에서 이를 다시 데이터논문으로 출간하게 되면 중복적으로 데이터 발간이 될 여지가 있다. 이와 함께 제한된 연구자의 논문 심사자가 데이터논문까지 심사하게 되는 부담이 있다고 지적하였다.

데이터 관리와 운영에 있어서 기반 사항인 데이터 리포지토리와의 이와 관련된 기관에 대한 연구로는 Goodman(2014), Castelli, Manghi, Thanos(2013), Pampel(2013), Pampel 외(2013), Akers와 Doty(2013) 등을 찾아볼 수 있다. Goodman(2014)는 연구데이터의 공유와 재이용을 위한 10가지 규칙을 제시하였다. 그 중에서 데이터 리포지토리 이용에 대한 권고를 하였는데, 가장 이상적으로 해당 분야의 리포지토리에 기탁하는 것이다. 만약 해당 분야의 리포지토리가 없다면, 전문 사서에게 문

의하거나 소속 기관의 리포지토리를 활용하도록 권고하고 있다. Castelli, Manghi, Thanos(2013)는 현대의 학술 커뮤니케이션에 대한 통찰을 기반으로 하여 디지털 도서관과 데이터 센터의 통합을 제안하였다. 제시된 통합적 모델은 조정 단계(mediation layer), 기능 단계(enabling layer), 내용 단계(content layer), 응용 단계(application layer)로 구분하여 제시하였다. 조정 단계는 데이터와 디지털 도서관의 구성 요소들이 관련 표준 프로토콜을 사용하여 변환되는 단계로서, 기능, 내용, 응용 단계의 기저를 이루게 된다. 기능 단계는 등록기능과 사용자 로그인 과정을 담당하여, 내용 단계는 콘텐츠를 처리하고, 저장하고, 외부로 반출하는 기능을 갖추게 된다. 마지막으로 응용 단계는 출판물과 데이터의 기탁 등 여러 응용 프로그램의 기능에 대하여 설명하고 있다. Pampel 외(2013)는 최근의 연구 데이터의 공유와 재이용에 대한 요구가 높아짐에 따라 데이터 리포지토리가 다양한 분야에서 구축되어 운영되고 있으며, 이러한 현상은 전세계적으로 확산되는 추세이다. 연구 수행 시점 기준으로 약 400여 데이터 리포지토리를 크게 기관 기반, 분야 기반, 프로젝트 기반, 다학제 기반으로 구분하였다. 이러한 다양한 리포지토리가 구현되어 운영되고 있지만, 데이터 공유와 재이용으로 활발하게 연계되기 위해서는 연구자의 연구과정에 리포지토리가 자연스럽게 연계되어야 한다고 지적하였다. 이와 함께 연구자가 해당 분야의 리포지토리에 대하여 인지하지 못하는 상황도 데이터 공유의 장애물로 제시되었다. 따라서 Pampel 외는 re3data.org를 활용한 리포지토리의 등록이 데이터 공유와 재이용으로 이어

질 수 있다고 설명한다. Akers와 Doty(2013)는 미국 Emory대학 재직 중인 330명의 다양한 분야의 연구자에게 데이터 관리와 공유 관행에 관하여 설문하였다. 응답 연구자들은 인문예술 분야, 사회과학 분야, 의학 분야, 기초과학 분야로 구분하여 분석하였다. 분석결과를 살펴보면, 연구자의 분야에 따라 통계적으로 의미있는 차이를 살펴볼 수 있는데, 특히 데이터관리와 데이터관리 지원에 대한 태도와 관심 항목에 있어서 분야별로 상이하게 나타났다고 제시하였다.

3. 연구방법

데이터학술지의 데이터논문이 규명하는 지적구조를 살펴보기 위해서 본 연구는 김지현(2019)의 연구에서 밝힌 총 79종의 데이터학술

지 중에서 2019년 12월 30일 현재 Web of Science에 등재된 총 14종의 학술지(〈표 1〉)를 대상으로 데이터를 수집하였다. 데이터논문은 Candela 외(2015)가 밝힌 바와 같이 “data article”, “data descriptor”, “data in brief”, “data note”, “data original article”, “data paper”, “database article”, “database paper”, “dataset paper”, “genome database” 등 여러 학술지에서 다양한 표현으로 사용되고 있다. 현재 Web of Science는 Document Type 필드에서 일반 학술 논문(article)과 구분하여 데이터논문을 “data paper”로 색인한다. 수집 절차를 살펴보면 총 14종의 학술지에 대하여 Document Type 필드에 Data Paper로 색인된 데이터논문을 추출하였다. 〈표 1〉에서 제시된 바와 같이 14종의 데이터학술지에 수록된 총 6,086건의 데이터 논문에 대하여 서지정보와 인용정보를 수집하였다.

〈표 1〉 Web of Science 등재 데이터학술지의 데이터논문 출판 현황¹⁾

번호	데이터학술지명	창간연도	출판사명	논문건수	데이터논문건수
1	Data in Brief	2015	Elsevier	4,742	4,536
2	Geoscience Data Journal	2014	Wiley	63	28
3	GigaScience	2012	Oxford University Press	638	128
4	Biodiversity Data Journal	2015	Pensoft	492	100
5	Comparative Cytogenetics	2007	Pensoft	473	2
6	NeoBiota	2015	Pensoft	145	3
7	PhytoKeys	2011	Pensoft	684	8
8	BioInvasions Records	2012	REABIC Journals	375	1
9	Scientific Data	2014	Springer Nature	1,142	850
10	Frontiers in Marine Science	2014	Frontiers Media S.A.	2,137	8
11	Earth System Science Data	2012	Copernicus Publications	432	241
12	Ecology	1998	Wiley	7,029	63
13	Data	2017	MDPI	238	110
14	Journal of Open Archaeology Data	2015	Ubiquity Press	17	8
합 계				18,607	6,086

1) 2019년 12월 30일 현재

총 6,086건의 데이터 논문에 대하여 서지정보와 인용정보는 3단계의 분석 과정을 거쳤다. 첫 번째 분석은 데이터논문의 일반적인 현황을 살펴보고자 하였다. 이를 위해서 저자사항, 저자의 소속기관, 저자의 국가를 별도로 추출하여 Bibexcel²⁾을 사용하여 빈도분석을 수행하였다. 두 번째 단계는 데이터논문의 동시인용 분석을 수행하여 네트워크로 시각화하였다. 이를 통해서 데이터논문이 동시에 인용된 문헌의 빈도수를 기반으로 구현된 지적구조를 살펴볼 수 있다. 그러나 논문이 인용되기까지는 상당한 시간이 소요되기 때문에 최신의 현황을 제시하는 데에는 한계가 있다. 이를 극복하기 위해서 세 번째 단계의 분석으로 데이터논문의 서지결합분석을 수행하여 네트워크로 시각화하였다. 동시인용분석과 서지결합분석 네트워크는 빈도행렬에서 코사인 유사도계수로 정규화하였으며, 이재운(2006)의 WNET과 NodeXL³⁾ 도구를 사용하여 시각화하였다.

4. 분석결과

4.1 데이터논문 현황

전체적인 데이터학술지와 데이터논문의 현황은 <표 2>에서 같이 살펴볼 수 있다. <표 2>에서 제시된 바와 같이 14종의 데이터학술지에 수록된 총 6,086 건의 데이터 논문에 대하여 서지정보와 인용정보를 수집하였다. 데이터학술지의 주제분야를 WoS의 주제명으로 살펴보면,

biodiversity, biology, ecology, plant sciences, marine freshwater biology, geoscience 등의 생물학 분야의 비중이 높다. 생물학 분야 이외에는 복합학, 유전학, 고고학 등의 주제 분야로 나타났다. 데이터논문의 발행 비중으로 살펴보면, 『Data in Brief』와 『Scientific Data』 학술지의 영향으로 복합학 분야가 가장 큰 것으로 나타났다. 데이터 학술지의 데이터논문 발행 비율을 살펴보면, 『Data in Brief』와 『Scientific Data』 학술지가 각각 95.7%와 74.4%의 높은 비중을 차지한 것으로 나타났다. 『Earth System Science Data』, 『Data』, 『Journal of Open Archaeology Data』 등의 데이터학술지는 전체 발간 논문의 반 정도를 데이터논문으로 발간하는 것으로 확인할 수 있다. 한편 논문의 20% 정도를 상회하는 비율로 데이터논문을 발간하는 학술지로는 『GigaScience』와 『Biodiversity Data Journal』를 찾아볼 수 있다. 나머지 학술지는 2.1%에서 0.3% 정도의 매우 작은 비중으로 데이터논문을 발간하는 것을 확인할 수 있다.

데이터논문 6,086건에 대해서 저자의 출현 빈도 기준 12회를 기준으로 상위 11명의 저자를 살펴보면, <표 3>과 같다. 저자의 소속과 전공분야는 저자의 개인 홈페이지, 구글 스칼라, ResearchGate 사이트를 참조하여 보강하였다. <표 3>에서 살펴볼 수 있는 바와 같이 통계, 환경, 전기공학, 농학, 해양학 등의 분야의 저자이며, 소속기관은 주로 나이지리아, 이란, 한국, 일본 등의 대학으로 나타났다. 이러한 경향은 일반적인 논문 학술지의 주요 저자와 소속기관 분포와는 다른 양상을 보여준다.

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

3) <https://www.smrfoundation.org/nodexl/>

〈표 2〉 데이터학술지와 데이터논문 발간 현황

번호	데이터 학술지명	WoS 주제명	출판사명	창간연도	논문건수	데이터 논문건수	데이터 논문비율
1	Data in Brief	multidisciplinary sciences	Elsevier	2015	4,742	4,536	95.7%
2	Geoscience Data Journal	geoscience multidisciplinary/meteorology atmospheric sciences	Wiley	2014	63	28	44.4%
3	GigaScience	biology/multidisciplinary science	Oxford University Press	2012	638	128	20.1%
4	Biodiversity Data Journal	biodiversity conservation	Pensoft	2015	492	100	20.3%
5	Comparative Cytogenetics	genetics heredity/plant sciences/zoology	Pensoft	2007	473	2	0.4%
6	NeoBiota	biodiversity conservation/ecology	Pensoft	2015	145	3	2.1%
7	PhytoKeys	plant sciences	Pensoft	2011	684	8	1.2%
8	BioInvasions Records	biodiversity conservation	REABIC Journals	2012	375	1	0.3%
9	Scientific Data	multidisciplinary sciences	Springer Nature	2014	1,142	850	74.4%
10	Frontiers in Marine Science	environmental sciences/marine freshwater biology	Frontiers Media S.A.	2014	2,137	8	0.4%
11	Earth System Science Data	geoscience multidisciplinary/meteorology atmospheric sciences	Copernicus Publications	2012	432	241	55.8%
12	Ecology	ecology	Wiley	1998	7,029	63	0.9%
13	Data	computer science information systems/multidisciplinary sciences	MDPI	2017	238	110	46.2%
14	Journal of open archaeology data	archaeology	Ubiquity Press	2015	17	8	47.1%
합 계					18,607	6,086	32.7%

〈표 3〉 출현빈도 기준 상위 10명의 저자

순위	저자명	횟수	소속기관	전공분야
1	Okagbue, Hilary I	24	Covenant University Ota Ogun State, Nigeria	Statistical Data analysis Mathematical statistics statistical patterns in numbers
2	Sagane, Yoshimasa	21	Tokyo University, Japan	Agriculture
3	Ramavandi, Bahman	19	Bushehr University of Medical Sciences, Iran	Water treatment Wastewater treatment
4	Radfard, Majid	16	Environmental Health Engineering, Shiraz University of Medical Science, Iran	Environmental Epidemiology Risk Assessment Water Quality Environmental Modelling
5	Mahvi, Amir Hossein	16	Tehran University of Medical Sciences, Iran	Environment
6	Atayero, Aderemi A	16	Covenant University Ota Ogun State, Nigeria	Electrical and Information Engineering
6	Oguntunde, Pelumi E	14	Covenant University Ota Ogun State, Nigeria	Robotic Planning
8	Watanabe, Toshihiro	14	Tokyo University, Japan	Agriculture
9	Park, Chan-Il	13	Gyeongsang Natl Univ, South Korea	Marine Science
10	Yousefi, Mahmood	13	Tehran University of Medical Sciences, Iran	Environment
10	Kim, Joondong	12	Incheon National University, South Korea	Electronics Engineering

데이터논문 저자들의 소속 기관을 별도로 추출하여 65회 출현빈도를 기준으로 상위 기관을 살펴보면 <표 4>와 같다. 1위는 총 293회 출현한 나이지리아의 Covenant University이며, 출현 빈도가 높은 저자의 소속 기관에서도 살펴볼 수 있다. 2위는 중국의 Chinese Academy of Sciences로 나타났으며, 러시아의 Russia Academy of Sciences, 미국의 National Oceanic and Atmospheric Administration (NOAA)와 함께 해당 분야의 전문적인 연구기관으로 데이터를 생산하고 이를 데이터논문으로 학술 커뮤니티에 발표하는 것으로 파악될 수 있다. 이러한 현상은 분야의 특성이 반영된 결과라고 볼 수 있다.

한편, 데이터논문의 저자들의 국가를 별도로 추출하여 살펴보면, <표 5>와 같다. 1위는 미국으로 총 1,530회 출현했으며, 이어서 중국과 독일이 각각 615회, 607회 나타났다. 국가 출현 빈도에서 저자와 소속 기관 출현 양상과 달리 통상적인 주요 국가의 상위 등장을 찾아볼 수 있는데 이는 데이터논문의 특성상 많은 수의

공저자와 주제 분야의 특징이 반영되었기 때문이라고 볼 수 있다.

<표 5> 저자의 국가 출현빈도 상위 20개 국가

순위	국가명	출현횟수
1	미국	1,530
2	중국	615
3	독일	607
4	영국	595
5	이탈리아	443
6	인도	333
7	프랑스	329
8	일본	319
9	캐나다	316
10	호주	303
11	스페인	295
12	이란	257
13	나이지리아	246
14	브라질	229
15	네델란드	217
16	러시아	197
17	한국	188
18	스위스	186
19	스웨덴	165
20	덴마크	137

<표 4> 저자 소속기관 출현빈도 기준 상위 10개 기관

순위	기관명	국가	출현횟수
1	Covenant Univ	나이지리아	293
2	Chinese Academy of Sciences	중국	127
3	Tehran University of Medical Sciences	이란	119
4	Russian Academy of Sciences	러시아	83
5	Univ Colorado	미국	74
6	National Oceanic and Atmospheric Administration(NOAA)	미국	71
6	Bushehr University of Medical Sciences	이란	71
8	Stanford Univ	미국	68
10	Univ Washington	미국	65
10	Univ Calif Los Angeles	미국	65

4.2 데이터논문의 세부 주제 영역

4.2.1 동시인용분석 기반 세부 주제 영역

6,086건의 데이터논문에서 인용된 참고문헌 총 84,908건을 추출하였다. 논문당 평균 약 14건의 참고문헌이 사용되었다. 데이터논문의 참고문헌에

서 출현빈도 상위 10건의 논문을 살펴보면 <표 6>과 같다. 상위 10위 이내에 『Bioinformatics』 학술지는 1위, 2위, 3위, 7위, 8회에 각각 등장하여 총 5회 나타났다. 『Bioinformatics』 학술지는 WoS 주제 영역으로 Biochemical Research Methods, Biotechnology & Applied Microbiology,

<표 6> 데이터논문 인용문헌 출현빈도 상위 10건

순위	인용문헌	학술지 주제(WoS)	논문의 저자키워드/ 키워드플러스	출현빈도
1	Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows - Wheeler transform. Bioinformatics , 25(14), 1754-1760.	Biochemical Research Methods/Biotechnology & Applied Microbiology/Mathematical & Computational Biology	genome: oligonucleotides: sequences: program: space: dna	119
2	Simão, F. A., & et al. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics , 31(19), 3210-3212.	Biochemical Research Methods/Biotechnology & Applied Microbiology/Mathematical & Computational Biology	quality: tool	118
3	Bolger, A. M., & et al. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics , 30(15), 2114-2120.	Biochemical Research Methods/Biotechnology & Applied Microbiology/Mathematical & Computational Biology	read alignment: algorithms	87
4	Grabherr, M. G., & et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology , 29(7), 644.	Biotechnology & Applied Microbiology	schizosaccharomyces-pombe: eukaryotic transcriptome: messenger-rna: alignment: gene: resource: reveals: protein: yeast: tool	75
5	Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods , 9(4), 357.	Biochemical Research Methods	burrows-wheeler transform: human genome: ultrafast	64
6	Luo, R., & et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience , 1(1), 2047-217X.	Multidisciplinary Sciences	genome, assembly, contig, scaffold, error correction, gap-filling	57
7	Dobin, A., & et al. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics , 29(1), 15-21.	Biochemical Research Methods/Biotechnology & Applied Microbiology/Mathematical & Computational Biology	splice junctions: alignment: algorithms: sequence: encode	54
8	Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics , 27(6), 764-770.	Biochemical Research Methods/Biotechnology & Applied Microbiology/Mathematical & Computational Biology	large genomes: sequence	51
8	Vizcaino, J. A., & et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature biotechnology , 32(3), 223-226.	Biotechnology & Applied Microbiology	identifications database: phosphoproteomics data: peptideatlas: pride	51
8	Walker, B. J., & et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS one , 9(11).	Multidisciplinary Sciences	mycobacterium-tuberculosis: draft assemblies: read alignment: protein-a: sequence: epidemiology	51

Mathematical & Computational Biology으로 구분되며, 주로 방법론을 다루는 학술지임을 알 수 있다. 이 외에도 5위로 나타난 『Nature Methods』 학술지도 방법론을 주로 다루는 학술지이다. 또한 『Gigascience』와 『PloS one』 학술지는 복합학 분야의 학술지로 각각 6위와 8위로 나타났다. 데이터논문의 주제를 살펴보기 위해서 논문의 저자키워드 혹은 키워드플러스를 추출해서 살펴보면, 대체적으로 genome, dna, algorithms, tools, database 등을 주된 키워드로 확인할 수 있다. 이러한 결과로 유추해 볼 수 있는 사항은 가장 빈번하게 인용되는 문헌이 방법론을 주로 다루는 학술지와 데이터적인 성격을 지닌 학술지 논문이라는 것이다.

데이터논문으로 제시된 지적구조를 살펴보기 위해서 총 84,908건의 참고문헌 중에서 20회 이상 인용된 83건의 논문을 대상으로 동시인용분석을 수행하였다. 83건의 논문의 목록은 [부록 1]에서 제시되었다. 20회 이상 인용된 83건의 논문을 대상으로 동시인용분석을 수행하였으며, 동시인용 횟수 31회를 기준으로 상위

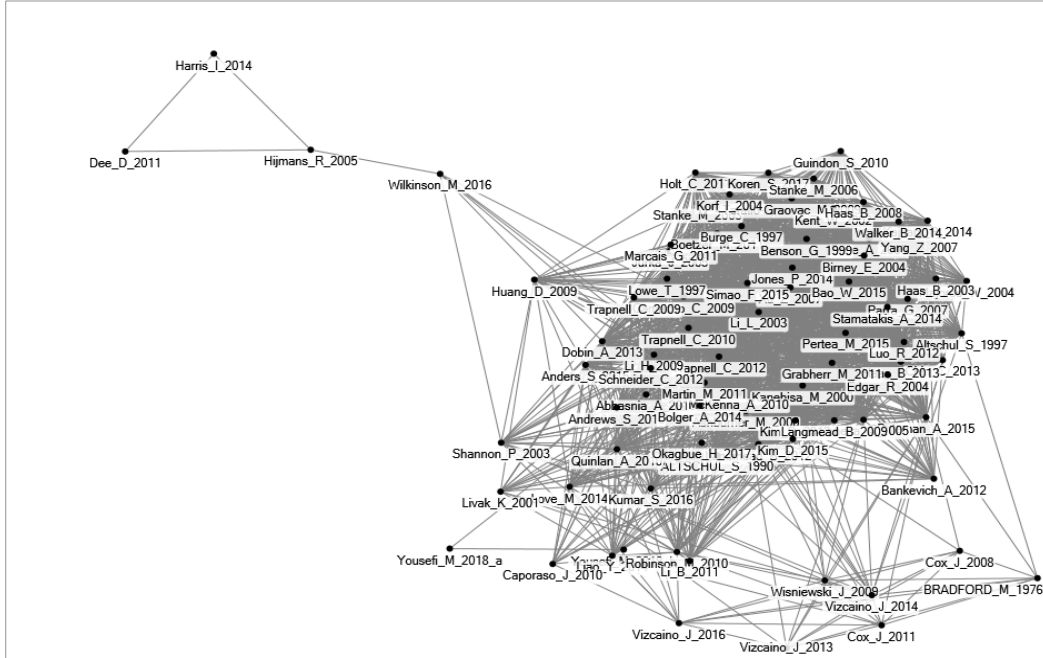
11건을 살펴보면 <표 7>과 같다. 가장 인용 횟수가 많은 두 논문 중 Li_H_2009은 genome, oligonucleotides, sequences, program, space, dna 등의 키워드를 갖고 있으며, Simao_F_2015은 quality, tool 등의 키워드를 지니고 있다. 개념과 분석 툴이 함께 인용된 논문의 비중이 높은 결과로 볼 수 있다. 또한 분석 도구를 기술한 논문인 Simao_F_2015는 상위 출현 쌍에서도 9번 등장하여 가장 많이 동시인용된 것으로 나타났다.

보다 명확하게 동시인용 네트워크를 구현하기 위해서 이재운(2006)의 WNET 소프트웨어 패키지의 패스파인더와 PNNC 군집화 기법을 사용하여 시각화하여 <그림 2>와 같이 제시하였다.

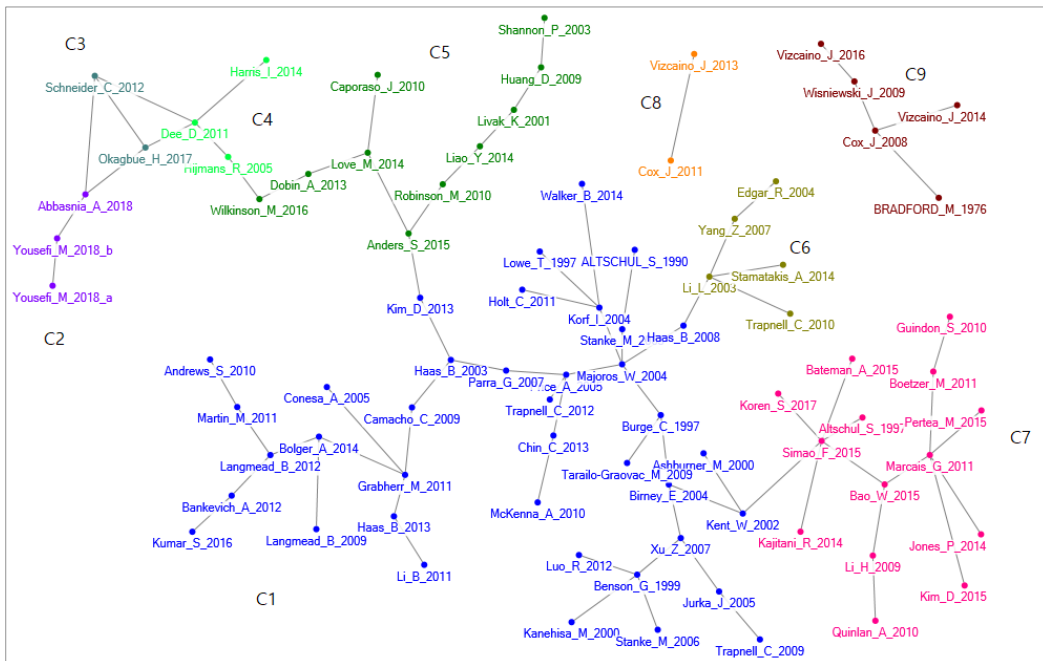
총 83건의 논문을 대상으로 동시인용 네트워크를 구현하여 <그림 1>과 같이 제시하였다. 네트워크 시각화에서 살펴볼 수 있는 바와 같이 상당히 밀도가 높게(0.52) 나타났다. 또한 상당히 밀접하게 연결되어 있는 대부분의 논문 그룹과 구분된 소수의 논문 그룹이 형성되었으

<표 7> 동시인용빈도 기준 상위 11건 논문

순위	인용문헌	인용문헌	인용횟수
1	Li_H_2009	Simao_F_2015	45
2	Grabherr_M_2011	Simao_F_2015	41
3	Birney_E_2004	Simao_F_2015	39
3	Luo_R_2012	Simao_F_2015	39
5	Simao_F_2015	Walker_B_2014	38
5	Benson_G_1999	Simao_F_2015	38
7	Marcais_G_2011	Simao_F_2015	36
8	Bolger_A_2014	Simao_F_2015	33
9	Li_L_2003	Simao_F_2015	32
10	Li_H_2009	Walker_B_2014	31
10	Benson_G_1999	Birney_E_2004	31



<그림 1> 동시인용분석 전체 네트워크(83건 논문)



<그림 2> 동시인용분석 패스파인더 네트워크와 9개 군집

며, Wilkinson_M_2016 논문으로 매개된 구성을 살펴볼 수 있다. 우선 세 논문 Dee_D_2011, Hijmans_R_2005, Harris_I_2014 은 대부분의 논문 그룹과는 별개로 그룹을 형성하여 동시에 인용되는 현상을 살펴볼 수 있다. [부록 1]의 전체 서지사항을 통해서 살펴보면, 세 논문은 기후 데이터와 관련된 주제임을 알 수 있다. 이 세 논문은 다른 논문과는 전혀 연결되지 않은 채 독자적으로 연결된 양상을 보여준다. 이 세 논문과 큰 규모의 논문 그룹과의 연결하는 역할을 하는 논문은 FAIR(Findable, Accessible, Interoperable, Reusable)를 지향하는 과학 데이터의 관리와 스투어드십에 관한 주제이다.

총 3개의 컴포넌트와 9개 군집으로 구성되었다. 우선 3개의 분절된 컴포넌트를 살펴보면, 대부분의 논문을 포함하는 큰 규모의 컴포넌트(군

집 1에서 군집 7), 군집 8인 두 논문으로 구성된 컴포넌트, 군집 9인 5개의 논문으로 구성된 컴포넌트로 형성되었다. [부록 1]의 각 논문의 서지사항과 저자키워드 등을 종합적으로 살펴볼 수 있다. 두 논문으로 구성되어 분절된 컴포넌트는 검색도구와 데이터베이스를 다루는 논문이다. 5개의 논문으로 구성된 또 다른 분절된 컴포넌트 역시 데이터베이스, 데이터처리 기법, 방법론, 도구를 주된 주제로 하는 논문으로 나타났다. 9개로 형성된 군집을 <표 8>과 같이 살펴보면, 군집 2와 군집 4를 제외하고는 대체로 분석, 방법론, 도구 등을 다룬 주제가 주를 이루는 것을 확인할 수 있다. 특히 군집4는 전체 네트워크에서 독립되어 구성된 세 논문(Dee_D_2011, Harris_I_2014, Hijmans_R_2005)으로 구성된 것을 확인할 수 있다.

<표 8> 동시인용분석 네트워크의 9개 군집과 주제

군집	소속 논문	주제
군집1	ALTSCHUL_S_1990, Andrews_S_2010, Ashburner_M_2000 Bankevich_A_2012, Benson_G_1999, Birney_E_2004 Bolger_A_2014, Burge_C_1997, Camacho_C_2009, Chin_C_2013, Conesa_A_2005, Grabherr_M_2011, Haas_B_2003, Haas_B_2008, Haas_B_2013, Holt_C_2011 Jurka_J_2005, Kanehisa_M_2000, Kent_W_2002, Kim_D_2013, Korf_I_2004, Kumar_S_2016, Langmead_B_2009, Langmead_B_2012, Li_B_2011, Lowe_T_1997, Luo_R_2012, Majoros_W_2004, Martin_M_2011, McKenna_A_2010, Parra_G_2007, Price_A_2005, Stanke_M_2003, Stanke_M_2006, Tarailo-Graovac_M_2009, Trapnell_C_2009, Trapnell_C_2012, Walker_B_2014, Xu_Z_2007	도구, 소프트웨어 프로그램, 분석
군집2	Abbasnia_A_2018, Yousefi_M_2018_a, Yousefi_M_2018_b	수질, 부식, 지하수
군집3	Okagbue_H_2017, Schneider_C_2012	이미지 분석, 랜덤 넘버
군집4	Dee_D_2011, Harris_I_2014, Hijmans_R_2005	기후, 데이터
군집5	Anders_S_2015, Caporaso_J_2010, Dobin_A_2013, Huang_D_2009, Liao_Y_2014, Livak_K_2001, Love_M_2014, Robinson_M_2010, Shannon_P_2003, Wilkinson_M_2016	분석 도구
군집6	Edgar_R_2004, Li_L_2003, Stamatakis_A_2014, Trapnell_C_2010, Yang_Z_2007	분석 도구
군집7	Altschul_S_1997, Bao_W_2015, Bateman_A_2015, Boetzer_M_2011, Guindon_S_2010, Jones_P_2014, Kajitani_R_2014, Kim_D_2015, Koren_S_2017, Li_H_2009, Marçais_G_2011, Pertea_M_2015, Quinlan_A_2010, Simao_F_2015	데이터베이스 검색도구, 분석도구
군집8	Cox_J_2011, Vizcaino_J_2013	검색도구와 데이터베이스
군집9	BRADFORD_M_1976, Cox_J_2008, Vizcaino_J_2014, Vizcaino_J_2016, Wisniewski_J_2009	데이터베이스, 데이터처리 기법, 방법론, 도구

군집 1은 가장 규모가 크며, 총 39건의 논문이 이에 해당하며, 주된 주제로 도구, 분석, 소프트웨어 프로그램 등이 파악된다. 군집 2는 수질, 부식에 관련된 주제로 나타났으며, 군집 3은 이미지 분석에 관한 주제이다. 군집 4는 기후에 관련된 주제로 구성되었으며, 군집 5와 군집 6은 분석도구를 다루고 있는 것으로 나타났다. 반면에 군집 7, 8, 9는 데이터베이스, 분석방법론, 도구, 검색도구 등이 주요한 주제로 나타났다.

4.2.2 서지결합분석 기반의 세부 주제 영역

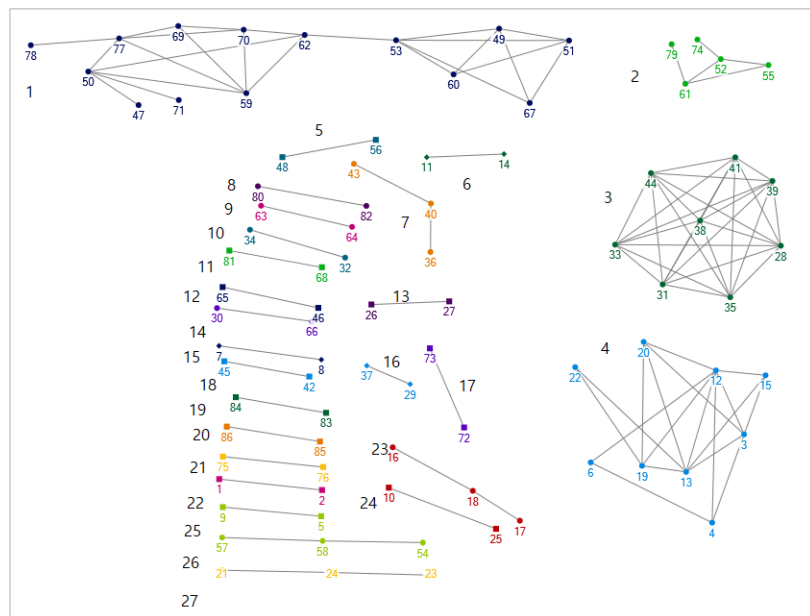
데이터논문에 대하여 주제적으로 관련이 높은 논문이 서로 연결되어 세부적 주제를 규명하기 위해서 서지결합분석을 수행하였다. <표 9>에서 보여주는 것과 같이 서지결합의 강도가 높은 순으로 살펴볼 수 있다. 가장 빈도가 높게 나타난 상위 10위 논문의 짝은 Global Carbon

Budget을 연도별로 발표한 데이터논문으로 나타났다.

<표 9> 데이터논문의 서지결합 상위 10위

순위	논문번호	논문번호	동시출현빈도
1	28	31	147
2	31	33	118
3	28	33	117
4	35	39	110
5	38	39	107
6	33	35	98
7	35	38	94
8	28	35	89
8	38	41	89
10	33	39	87

<그림 3>과 같이 상위 86건의 논문에 대해서 서지결합분석하여 네트워크로 시각화한 결과 27개의 분절된 컴포넌트로 구성되었다. 86건의



<그림 3> 서지결합분석 네트워크와 27개 컴포넌트

논문이 27개의 컴포넌트로 구성된 것은 최신의 데이터논문이 대규모의 연구영역을 형성하고 있지는 않기 때문이라고 볼 수 있다. 연속적으로 발간되는 데이터나 여러 국가가 참여해서 지속적으로 발간하는 데이터 중심으로 분절된 컴포넌트를 이루고 있는 것이다. [부록 2]의 개별 논문의 서지사항과 논문의 저자 키워드를 통해서 주제적 결합을 살펴볼 수 있다. 각 컴포

넌트에 구성된 논문과 세부주제 영역을 살펴보면 <표 10>과 같다. 앞서 살펴본, 서지결합 쌍의 상위를 차지한 탄소배출 보고서인 Global Carbon Budget 연간 보고서가 하나의 컴포넌트로 구성되었다. 전체적으로 계층 구성에 관한 5개의 컴포넌트를 1, 2, 5, 9, 26을 찾아볼 수 있다. 수질에 관련된 컴포넌트는 4와 27로 나타났으며, 나이지리아 지역의 사회조사 설문

<표 10> 서지결합분석 27개 컴포넌트와 세부주제

컴포넌트	소속 논문	세부주제
1	78, 50, 47, 77, 69, 71, 59, 70, 62, 53, 49, 60, 51, 67	계층 구성
2	79, 74, 52, 61, 55	계층 구성
3	44, 33, 38, 31, 35, 28, 39, 41	탄소 배출 데이터
4	22, 20, 6, 19, 13, 12, 15, 3, 4	이란 지역 수질
5	48, 56	계층 구성
6	11, 14	나이지리아 지역 설문 데이터
7	43, 40, 36	해양 표면 CO2
8	80, 82	세포 분석 결과
9	63, 64	계층 구성
10	34, 32	대기 가스
11	81, 68	생물기후
12	65, 46	산호 포라이트
13	26, 27	단백질
14	30, 66	해양표면
15	7, 8	보험회사 운용
16	37, 29	해양데이터
17	73, 72	척수 손상
18	45, 42	기후변화
19	84, 83	물기둥
20	86, 85	고분자
21	75, 76	뇌 데이터
22	1, 2	도시 녹지
23	16, 18, 17	저널리즘 학술지 분석
24	10, 25	나이지리아 학생 설문 데이터
25	9, 5	refractory high entropy alloy
26	57, 58, 54	계층 구성(어류)
27	21, 24, 23	수질

데이터도 별도의 컴포넌트인 6과 24로 구성되었음을 확인할 수 있다. 해양 관련 컴포넌트로는 7, 14, 16으로 나타났다. 한가지 특기할 만한 사항은 전반적으로 자연과학 데이터논문이 주류를 이루는 데 반해, 컴포넌트 6, 15, 23, 24와 같이 사회과학의 데이터논문 그룹을 찾아볼 수 있다. 이 외에도, 세포 분석, 대기 가스, 생물기후, 산호 포라이트, 단백질, 척수 손상, 기후변화, 물기둥, 고분자, 뇌 데이터, 도시 녹지 등의 다양한 자연과학 분야의 데이터논문 컴포넌트들을 찾아볼 수 있다. 서지결합분석 네트워크가 소수로 구성된 27개의 컴포넌트로 분절된 것은 데이터논문이 구성하는 영향력 있는 큰 규모의 주제가 아직은 형성되지 않았기 때문이라고 추정할 수 있다. 다양한 여러 세부 주제 볼 수 있다. 특히, 동시인용분석에서는 찾아볼 수 없는 사회과학 분야의 세부 주제를 확인할 수 있는데, 구체적으로는 6, 15, 23, 24 컴포넌트이다. 6번과 24번 컴포넌트는 나이지리아 지역의 학생들의 행동, ICT 수용 태도 등을 규명하는 데이터를 반영하였다. 15번 컴포넌트는 보험회사의 시장 실적을 분석한 데이터이며, 23번 컴포넌트는 저널리즘 학술지의 편집진을 분석한 데이터이다.

5. 논의 및 결론

최근 여러 학술커뮤니티들은 연구자들이 생산한 데이터를 공유하고 재이용하고자 하는 노력을 지속하고 있다. 연구데이터의 공유와 재이용을 촉진하고자 하는 다양한 시도들 중에는 데이터학술지와 데이터논문의 발간이 주목할

만한 성장을 이루고 있다. 데이터논문은 연구자에게 인용이라는 학계의 인정을 받은 기제가 작동해서 자발적으로 데이터를 공유하는 역할을 한다. 또한 데이터학술지는 전통적인 학술지와 마찬가지로 데이터논문에 대한 동료평가를 수행하기 때문에 공유되는 데이터에 대한 품질을 보증할 수 있다. 이러한 장점으로 인해 다양한 학술 커뮤니티에서 데이터학술지를 출간하고 많은 연구자들은 데이터를 데이터논문으로 발간하는 상황이다. 따라서 전통적인 학술지와 달리 새롭게 등장하여 성장하고 있는 데이터학술지와 데이터논문이 형성하는 지식의 구조를 살펴볼 필요가 있다. 이를 위해 본 연구는 WoS에 색인된 14종의 데이터학술지에서 총 6,086건의 데이터논문을 대상으로 동시인용분석과 서지결합분석을 수행하여 네트워크로 시각화하였다.

데이터학술지와 데이터논문의 주제영역, 저자, 저자 소속기관, 국가 등의 전반적인 현황을 살펴보면 전통적인 학술커뮤니케이션의 양상과는 다르다는 것을 확인할 수 있다. 우선 데이터학술지의 주제영역을 살펴보면, 생물학 분야의 비중이 높은 것으로 나타났으며, 데이터논문의 발행 비중이 높은 『Data in Brief』와 『Scientific Data』 학술지는 복합학 주제 분야로 설정되었다. 그러나 정은경(2019) 연구에서 밝힌 바와 같이, 복합학 분야의 학술지인 『Scientific Data』에 발행된 데이터논문도 생물학 분야의 비중이 크다. 따라서 자연과학 분야의 여러 학술 커뮤니티에서 데이터학술지와 데이터논문에 대한 관심과 논의가 있지만, 가장 활발하게 데이터논문을 발간하고 있는 분야는 생물학 관련 분야라고 볼 수 있다. 저자와 소속기관, 국

가의 관점에서 살펴보면, 전통적인 학술커뮤니케이션에서 빈번하게 등장하는 저명한 기관이나 선진국의 비중이 상당히 낮은 것으로 나타났다. 데이터학술지와 데이터논문이 주된 학술커뮤니케이션의 장이라기 보다는 아직은 미미한 영향력과 주변적인 성과로 인식되기 때문이라고 볼 수 있다. 현재는 데이터논문이 데이터생산이 많은 주제 영역에서 해당 기관에 한정되어 기관 소속 연구자들에 의해서 주로 발행된다고 해석할 수 있다. 동시인용분석을 통해 살펴본 세부 주제 영역은 9개 영역으로 구분된다. 세부 영역 중에서 도구, 분석도구, 데이터베이스 등을 다루는 군집이 주를 이루는 것으로 나타났다. 반면에 특정 주제 분야로 파악된 군집은 수질, 기후 등의 영역으로 나타났다. 데이터논문이 구성하는 세부 주제영역을 인용지체 현상 없이 살펴보기 위해서 실행한 서지결합분석 결과를 살펴보면, 총 27개의 분절된 컴포넌트로 나타났다. 우선 주된 주제 영역으로 여러 생명 종의 계층 구성에 관한 것으로 나타

났다. 이 외에도 기후 영역, 해양, 수질, 대기 등의 주제 영역을 확인할 수 있다. 한가지 특기할 만한 사항은 주된 주제 분야는 자연과학 분야의 세부 주제 영역이지만, 사회과학 분야의 주제 영역도 나타났다는 점이다. 사회과학 분야가 서지결합분석에서 나타난 현상은 최근의 데이터논문이 생물학 중심의 자연과학에서 사회과학으로 확정되는 것으로 해석할 수 있다. 본 연구에서 이러한 결과를 통해 데이터논문이라는 새로운 학술 커뮤니케이션의 장이 구성하는 지적 구조를 살펴보았다. 그러나 데이터학술지에 발행되는 모든 논문이 데이터논문은 아니기 때문에 본 연구에서 분석한 14종의 데이터학술지가 제시하는 지적구조라 이해하기에는 제한점이 있다. 향후 연구에서는 데이터논문과 기존의 전통적인 학술지 논문으로 각각 구성되는 학술분야의 비교 분석을 통해 데이터논문이 형성해 가는 영역에 대한 보다 다각적인 이해가 필요하다.

참 고 문 헌

- 김지현. (2019). 데이터 저널 현황 조사 및 발간정책 수립. 한국과학기술정보연구원.
- 이재윤. (2006). WNET 소프트웨어 패키지.
- 정은경. (2019). Scientific Data 학술지 분석을 통한 데이터 논문 현황에 관한 연구. 정보관리학회지, 36(1), 117-135. <https://doi.org/10.3743/KOSIM.2019.36.1.117>.
- Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5-26. <https://doi.org/10.2218/ijdc.v8i2.263>
- Belter, C. W. (2014). Measuring the value of research data: A citation analysis of oceanographic

- data sets. *PLoS One*, 9(3), e92590. <https://doi.org/10.1371/journal.pone.0092590>
- Borgman, C. L. (2015). *Big data, little data, no data*. Cambridge, MA: MIT Press.
- Borgman, C. L. (2016). Data citation as a bibliometric oxymoron. In Sugimoto, C. (Eds), *Theories of informetrics and scholarly communication* (pp. 93-116). Berlin: De Gruyter Mouton.
- 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>
- Castelli, D., Manghi, P., & Thanos, C. (2013). A vision towards scientific communication infrastructures. *International Journal on Digital Libraries*, 13(3-4), 155-169. <https://doi.org/10.1007/s00799-013-0106-7>
- 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. doi.org/10.1016/j.tree.2013.05.002
- Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., ... & Hogg, D. W. (2014). Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*, 10(4). <https://doi.org/10.1371/journal.pcbi.1003542>
- Gorgolewski, K., Margulies, D. S., & Milham, M. P. (2013). Making data sharing count: a publication-based solution. *Frontiers in neuroscience*, 7, 9. <https://doi.org/10.3389/fnins.2013.00009>
- Huang, X., Hawkins, B. A., & Qiao, G. (2013). Biodiversity data sharing: Will peer-reviewed data papers work?. *BioScience*, 63(1), 5-6. <https://doi.org/10.1525/bio.2013.63.1.2>
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Spies, J. R. (2016). Point of view: How open science helps researchers succeed. *Elife*, 5, e16800. <https://doi.org/10.7554/eLife.16800>
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., ... & Dierolf, U. (2013). Making research data repositories visible: the re3data.org registry. *PloS one*, 8(11), e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Silvello, G. (2018). Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, 69(1), 6-20. <https://doi.org/10.1002/asi.23917>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Kim, Ji Hyun (2019). A study on the current status of data journals and the establishment of data journal policies. Korea Institute of Science and Technology Information.

Lee, Jae Yun (2006). WNET Software Package.

Chung, EunKyung (2019). An investigation on scientific data for data journal and data paper. Journal of the Korean Society for Information Management, 36(1), 117-135.
<https://doi.org/10.3743/KOSIM.2019.36.1.117>

[부록 1] 동시인용분석 네트워크 논문목록

번호	논문	논문약칭
1	Abbasnia, A., & et al. (2018). Assessment of groundwater quality and evaluation of scaling and corrosiveness potential of drinking water samples in villages of Chabahr city, Sistan and Baluchistan province in Iran. <i>Data in brief</i> , 16, 182-192.	Abbasnia_A_2018
2	Altschul, S. F., & et al. (1990). Lipman DJ. Basic local alignment search tool, 21, 403-410.	ALTSCHUL_S_1990
3	Altschul, S. F. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. <i>Nucleic Acids Res.</i> 25: 3389.A79	Altschul_S_1997
4	Anders, S., & et al. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. <i>Bioinformatics</i> , 31(2), 166-169.	Anders_S_2015
5	Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.	Andrews_S_2010
6	Ashburner, M., & et al. (2000). The Gene Ontology Consortium Gene ontology: tool for the unification of biology. <i>Nat genet.</i> 25(1), 25-29.	Ashburner_M_2000
7	Bankevich, A., & et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. <i>Journal of computational biology</i> , 19(5), 455-477.	Bankevich_A_2012
8	Bao, W., & et al. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. <i>Mobile Dna</i> , 6(1), 11.	Bao_W_2015
9	Bateman A, et al. (2015). UniProt: a hub for protein information, <i>Nucleic Acids Research</i>	Bateman_A_2015
10	Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. <i>Nucleic acids research</i> , 27(2), 573-580.	Benson_G_1999
11	Birney, E., & et al. (2004). An overview of Ensembl. <i>Genome research</i> , 14(5), 925-928.	Birney_E_2004
12	Boetzer, M., & et al. (2011). Scaffolding pre-assembled contigs using SSPACE. <i>Bioinformatics</i> , 27(4), 578-579.	Boetzer_M_2011
13	Bolger, A. M., & et al. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. <i>Bioinformatics</i> , 30(15), 2114-2120.	Bolger_A_2014
14	Bradford, N. A. (1976). A rapid and sensitive method for the quantitation microgram quantities of a protein isolated from red cell membranes. <i>Anal Biochem</i> , 72, 248-254.	BRADFORD_M_1976
15	Burge, C. B. (1997). Identification of Genes in Human Genomic DNA (Doctoral dissertation, Stanford University).	Burge_C_1997
16	Camacho, C., & et al. (2009). BLAST+: architecture and applications. <i>BMC bioinformatics</i> , 10(1), 421.	Camacho_C_2009
17	Caporaso, J. G., & et al. (2010). QIIME allows analysis of high-throughput community sequencing data. <i>Nature methods</i> , 7(5), 335.	Caporaso_J_2010
18	Chin, C. S., & et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data <i>Nat Methods</i> , 10, 563-56910.	Chin_C_2013
19	Conesa, A., & et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. <i>Bioinformatics</i> , 21(18), 3674-3676.	Conesa_A_2005
20	Cox, J., & et al. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. <i>Nature biotechnology</i> , 26(12), 1367-1372.	Cox_J_2008
21	Cox, J., & et al. (2011). 619 Andromeda: A peptide search engine integrated into the MaxQuant environment. <i>J.</i> 620, 1794-1805.	Cox_J_2011
22	Dee, D. P., & et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. <i>Quarterly Journal of the Royal Meteorological Society</i> , 137, 553 - 597.	Dee_D_2011
23	Dobin, A., & et al. (2013). STAR: ultrafast universal RNA-seq aligner. <i>Bioinformatics</i> , 29(1), 15-21.	Dobin_A_2013
24	Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. <i>Nucleic acids research</i> , 32(5), 1792-1797.	Edgar_R_2004
25	Grabherr, M. G., & et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. <i>Nature biotechnology</i> , 29(7), 644.	Grabherr_M_2011
26	Guindon, S., & et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. <i>Systematic biology</i> , 59(3), 307-321.	Guindon_S_2010
27	Haas, B. J., & et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. <i>Nucleic acids research</i> , 31(19), 5654-5666.	Haas_B_2003

번호	논문	논문약칭
28	Haas, B. J., & et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. <i>Genome biology</i> , 9(1), R7.	Haas_B_2008
29	Haas, B. J., & et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. <i>Nature protocols</i> , 8(8), 1494.	Haas_B_2013
30	Harris, I. (2014). Jones P d., Osborn T j., Lister D h. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. <i>Int. J. Climatol.</i> , 34, 623-642.	Harris_I_2014
31	Hijmans, R. J., & et al. (2005). Very high resolution interpolated global terrestrial climate surfaces. <i>International Journal of Climatology</i> , 25(15), 1965-1978.	Hijmans_R_2005
32	Holt, C., & et al. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. <i>BMC bioinformatics</i> , 12(1), 491.	Holt_C_2011
33	Huang, T. T., & et al. (2013). Epigenetic deregulation of the anaplastic lymphoma kinase gene modulates mesenchymal characteristics of oral squamous cell carcinomas. <i>Carcinogenesis</i> , 34(8), 1717-1727.	Huang_D_2009
34	Jones, P., & et al. (2014). InterProScan 5: genome-scale protein function classification. <i>Bioinformatics</i> .	Jones_P_2014
35	Jurka, J., & et al. (2005). Repbase Update, a database of eukaryotic repetitive elements. <i>Cytogenetic and genome research</i> , 110(1-4), 462-467.	Jurka_J_2005
36	Kajitani, R., & et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. <i>Genome research</i> , 24(8), 1384-1395.	Kajitani_R_2014
37	Kanehisa, M., & et al.(2000). KEGG: kyoto encyclopedia of genes and genomes. <i>Nucleic acids research</i> , 28(1), 27-30.	Kanehisa_M_2000
38	Kent, W. J., & et al. (2002). The human genome browser at UCSC. <i>Genome research</i> , 12(6), 996-1006.	Kent_W_2002
39	Kim, D., & et al. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. <i>Genome biology</i> , 14(4), R36.	Kim_D_2013
40	Kim, D., & et al. (2015). HISAT: a fast spliced aligner with low memory requirements. <i>Nature methods</i> , 12(4), 357-360.	Kim_D_2015
41	Koren, S., & et al. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. <i>Genome research</i> , 27(5), 722-736.	Koren_S_2017
42	Korf, I. (2004). Gene finding in novel genomes. <i>BMC bioinformatics</i> , 5(1), 59.	Korf_I_2004
43	Kumar, S., & et al. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. <i>Molecular biology and evolution</i> , 33(7), 1870-1874.	Kumar_S_2016
44	Langmead, B., & et al. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. <i>Genome biology</i> , 10(3), R25.	Langmead_B_2009
45	Langmead, B., & et al. (2012). Fast gapped-read alignment with Bowtie 2. <i>Nature methods</i> , 9(4), 357.	Langmead_B_2012
46	Li, B., & et al. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. <i>BMC bioinformatics</i> , 12(1), 323.	Li_B_2011
47	Li, H., & et al. (2009). Fast and accurate short read alignment with Burrows - Wheeler transform. <i>Bioinformatics</i> , 25(14), 1754-1760.	Li_H_2009
48	Li, L., & et al. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. <i>Genome research</i> , 13(9), 2178-2189.	Li_L_2003
49	Liao, Y., & et al. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. <i>Bioinformatics</i> , 30(7), 923-930.	Liao_Y_2014
50	Livak, K. J., & et al. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2 ⁻ ΔΔCT method. <i>methods</i> , 25(4), 402-408.	Livak_K_2001
51	Love, M., & et al. (2014). Differential gene expression analysis based on the negative binomial distribution. <i>Genome Biol.</i> , 15, 550.	Love_M_2014
52	Lowe, T. M., & et al. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. <i>Nucleic acids research</i> , 25(5), 955-964.	Lowe_T_1997
53	Luo, R., & et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. <i>Gigascience</i> , 1(1), 2047-217X.	Luo_R_2012
54	Majoros, W. H., & et al. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. <i>Bioinformatics</i> , 20(16), 2878-2879.	Majoros_W_2004
55	Marçais, G., & et al. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. <i>Bioinformatics</i> , 27(6), 764-770.	Marçais_G_2011

번호	논문	논문약칭
56	Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. <i>EMBnet journal</i> , 17(1), 10-12.	Martin_M_2011
57	McKenna, A., & et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. <i>Genome research</i> , 20(9), 1297-1303.	McKenna_A_2010
58	Okagbue, H. I., & et al. (2017). Random number datasets generated from statistical analysis of randomly sampled GSM recharge cards. <i>Data in brief</i> , 10, 269-276.	Okagbue_H_2017
59	Parra, G., & et al. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. <i>Bioinformatics</i> , 23(9), 1061-1067.	Parra_G_2007
60	Pertea, M., & et al. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. <i>Nature biotechnology</i> , 33(3), 290.	Pertea_M_2015
61	Price, A. L., & et al. (2005). De novo identification of repeat families in large genomes. <i>Bioinformatics</i> , 21(suppl_1), i351-i358.	Price_A_2005
62	Quinlan, A. R., & et al. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. <i>Bioinformatics</i> , 26(6), 841-842.	Quinlan_A_2010
63	Robinson, M. D., & et al. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. <i>Bioinformatics</i> , 26(1), 139-140.	Robinson_M_2010
64	Schneider, C. A., & et al. (2012). NIH image to ImageJ: 25 years of image analysis. <i>Nat Meth</i> 9(7): 671 - 675.	Schneider_C_2012
65	Shannon, P., & et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. <i>Genome research</i> , 13(11), 2498-2504.	Shannon_P_2003
66	Simão, F. A., & et al. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. <i>Bioinformatics</i> , 31(19), 3210-3212.	Simao_F_2015
67	Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. <i>Bioinformatics</i> , 30(9), 1312-1313.	Stamatakis_A_2014
68	Stanke, M., & et al. (2003). Gene prediction with a hidden Markov model and a new intron submodel. <i>Bioinformatics</i> 19.	Stanke_M_2003
69	Stanke, M., & et al. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. <i>Nucleic acids research</i> , 34(suppl_2), W435-W439.	Stanke_M_2006
70	Tarailo-Graovac, M., & et al. (2009). <i>Curr Protoc Bioinformatics</i> .	Tarailo-Graovac_M_2009
71	Trapnell, C., & et al. (2009). TopHat: discovering splice junctions with RNA-Seq. <i>Bioinformatics</i> , 25(9), 1105-1111.	Trapnell_C_2009
72	Trapnell, C., & et al. (2010). Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. <i>Nature biotechnology</i> , 28(5), 511.	Trapnell_C_2010
73	Trapnell, C., & et al. (2012). Differential gene and transcript expression analysis of 550 RNA-seq experiments with TopHat and Cufflinks. <i>Nature Protocols</i> , 7, 562-578.	Trapnell_C_2012
74	Vizcaino, J. A., & et al. (2013). The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. <i>Nucleic acids research</i> , 41(D1), D1063-D1069.	Vizcaino_J_2013
75	Vizcaino, J. A., & et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. <i>Nature biotechnology</i> , 32(3), 223-226.	Vizcaino_J_2014
76	Vizcaino, J. A., & et al. (2016). 2016 update of the PRIDE database and its related tools. <i>Nucleic acids research</i> , 44(D1), D447-D456.	Vizcaino_J_2016
77	Walker, B. J., & et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. <i>PLoS one</i> , 9(11).	Walker_B_2014
78	Wilkinson, M. D. (2016). The FAIR guiding principles for scientific data management and stewardship (2016) <i>Scientific Data</i> 3.	Wilkinson_M_2016
79	Wiśniewski, J. R., & et al. (2009). Universal sample preparation method for proteome analysis. <i>Nature methods</i> , 6(5), 359-362.	Wisniewski_J_2009
80	Xu, Z., & et al. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. <i>Nucleic acids research</i> , 35(suppl_2), W265-W268.	Xu_Z_2007
81	Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. <i>Molecular biology and evolution</i> , 24(8), 1586-1591.	Yang_Z_2007
82	Yousefi, M., & et al. (2018). Data on corrosion and scaling potential of drinking water resources using stability indices in Jolfa, East Azerbaijan, Iran. <i>Data in brief</i> , 16, 724-731.	Yousefi_M_2018_a
83	Yousefi, M., & et al. (2018). Health risk assessment to fluoride in drinking water of rural residents living in the Poldasht city, Northwest of Iran. <i>Ecotoxicology and environmental safety</i> , 148, 426-430.	Yousefi_M_2018_b

[부록 2] 서지결합분석 네트워크 논문 목록

논문번호	데이터논문 제목	학술지명	연도	저자
1	Data on Peer-Reviewed Papers about Green Infrastructure, Urban Nature, and City Liveability	DATA	2018	Simpson GD외
2	Data for an Importance-Performance Analysis (IPA) of a Public Green Infrastructure and Urban Nature Space in Perth, Western Australia	DATA	2018	Simpson GD외
3	Dataset on assessment of physical and chemical quality of groundwater in rural drinking water, west Azerbaijan Province in Iran	DATA IN BRIEF	2018	Radfard M외
4	Data on health risk assessment of fluoride in drinking water in the Khash city of Sistan and Baluchistan province, Iran	DATA IN BRIEF	2018	Azhdarpoor A외
5	Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys	DATA IN BRIEF	2018	Couzinie JP외
6	Data on health risk assessment to the nitrate in drinking water of rural areas in the Khash city, Iran	DATA IN BRIEF	2018	Radfard M외
7	Resilience and learning from insurance firms: Dataset on British long-term insurance market performance	DATA IN BRIEF	2018	Mafimisebi OP외
8	Risky business: Data on trading results for UK general insurance firms during and after the global financial crisis	DATA IN BRIEF	2018	Mafimisebi OP
9	Database on the mechanical properties of high entropy alloys and complex concentrated alloys	DATA IN BRIEF	2018	Gorsse S외
10	Survey dataset on the types, prevalence and causes of deviant behavior among secondary school adolescents in some selected schools in Benin City, Edo State, Nigeria	DATA IN BRIEF	2018	Bishop SA외
11	Survey dataset on factors that influence satisfaction of clients with architectural services in Lagos State, Nigeria	DATA IN BRIEF	2018	Oluwatayo A외
12	Data on drinking water quality using water quality index (WQI) and assessment of groundwater quality for irrigation purposes in Qorveh&Dehgolan, Kurdistan, Iran	DATA IN BRIEF	2018	Soleimani H외
13	Data on Nitrate-Nitrite pollution in the groundwater resources a Sonqor plain in Iran	DATA IN BRIEF	2018	Jalili D외
14	Data on the awareness and adoption of ICT in town planning firms in Lagos state, Nigeria	DATA IN BRIEF	2018	Akinola AO외
15	Data on investigating the nitrate concentration levels and quality of bottled water in Torbat-e Heydarieh, Khorasan razavi province, Iran	DATA IN BRIEF	2018	Akbari H외
16	Analysis of dataset on editorial board composition of Dove Medical Press by continent	DATA IN BRIEF	2018	Okagbue HI외
17	Dataset and analysis of editorial board composition of 165 Hindawi journals indexed and abstracted in PubMed based on affiliations	DATA IN BRIEF	2018	Okagbue HI외
18	Exploration of editorial board composition, Citescore and percentiles of Hindawi journals indexed in Scopus	DATA IN BRIEF	2018	Okagbue HI외
19	Data on quality indices of groundwater resource for agricultural use in the Jolfa, East Azerbaijan, Iran	DATA IN BRIEF	2018	Heidarinejad Z외
20	Data on groundwater quality, scaling potential and corrosiveness of water samples in Torbat-e-Heydariyeh rural drinking water resources, Khorasan-e-Razavi province, Iran	DATA IN BRIEF	2018	Babaeia AA외

논문번호	데이터논문 제목	학술지명	연도	저자
21	Data for factor analysis of hydro-geochemical characteristics of groundwater resources in Iranshahr	DATA IN BRIEF	2018	Biglari H외
22	Groundwater quality assessment for drinking and agriculture purposes in Abhar city, Iran	DATA IN BRIEF	2018	Jafari K외
23	Data on performance of air stripping tower-PAC integrated system for removing of odor, taste, dye and organic materials from drinking water-A case study in Saqqez, Iran	DATA IN BRIEF	2018	Firsaheb M외
24	Data for distribution of various species of fecal coliforms in urban, rural and private drinking water sources in ten years period - A case study: Kermanshah, Iran	DATA IN BRIEF	2018	Davoodi R외
25	Survey datasets on the externalizing behaviors of primary school pupils and secondary school students in some selected schools in Ogun State, Nigeria	DATA IN BRIEF	2017	Bishop SA외
26	Active site specificity profiling datasets of matrix metalloproteinases (MMPs) 1, 2, 3, 7, 8, 9, 12, 13 and 14	DATA IN BRIEF	2016	Eckhard U외
27	TAILS N-terminomic and proteomic datasets of healthy human dental pulp	DATA IN BRIEF	2015	Eckhard U외
28	Global Carbon Budget 2019	EARTH SYSTEM SCIENCE DATA	2019	Friedlingstein P외
29	GLODAPv2.2019-an update of GLODAPv2	EARTH SYSTEM SCIENCE DATA	2019	Olsen A외
30	A machine-learning-based global sea-surface iodide distribution	EARTH SYSTEM SCIENCE DATA	2019	Sherwen T외
31	Global Carbon Budget 2018	EARTH SYSTEM SCIENCE DATA	2018	Le Quere C외
32	History of chemically and radiatively important atmospheric gases from the Advanced Global Atmospheric Gases Experiment (AGAGE)	EARTH SYSTEM SCIENCE DATA	2018	Prinn RG외
33	Global Carbon Budget 2017	EARTH SYSTEM SCIENCE DATA	2018	Le Quere C외
34	The global methane budget 2000-2012	EARTH SYSTEM SCIENCE DATA	2016	Saunois M외
35	Global Carbon Budget 2016	EARTH SYSTEM SCIENCE DATA	2016	Le Quere C외
36	A multi-decade record of high-quality fCO(2) data in version 3 of the Surface Ocean CO2 Atlas (SOCAT)	EARTH SYSTEM SCIENCE DATA	2016	Bakker DCE외
37	The Global Ocean Data Analysis Project version 2 (GLODAPv2) - an internally consistent data product for the world ocean	EARTH SYSTEM SCIENCE DATA	2016	Olsen A외
38	Global carbon budget 2014	EARTH SYSTEM SCIENCE DATA	2015	Le Quere C외
39	Global Carbon Budget 2015	EARTH SYSTEM SCIENCE DATA	2015	Le Quere C외
40	An update to the Surface Ocean CO2 Atlas (SOCAT version 2)	EARTH SYSTEM SCIENCE DATA	2014	Bakker DCE외
41	Global carbon budget 2013	EARTH SYSTEM SCIENCE DATA	2014	Le Quere C외
42	Future Flows Hydrology: an ensemble of daily river flow and monthly groundwater levels for use for climate change impact assessment across Great Britain	EARTH SYSTEM SCIENCE DATA	2013	Prudhomme C외

논문번호	데이터논문 제목	학술지명	연도	저자
43	A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT)	EARTH SYSTEM SCIENCE DATA	2013	Pfeil B외
44	The global carbon budget 1959-2011	EARTH SYSTEM SCIENCE DATA	2013	Le Quere C외
45	Future Flows Climate: an ensemble of 1-km climate change projections for hydrological application in Great Britain	EARTH SYSTEM SCIENCE DATA	2012	Prudhomme C외
46	Development of a Novel Reference Transcriptome for Scleractinian Coral <i>Porites lutea</i> Using Single-Molecule Long-Read Isoform Sequencing (Iso-Seq)	FRONTIERS IN MARINE SCIENCE	2018	Pootakham W외
47	Trochodendron aralioides, the first chromosome-level draft genome in Trochodendrales and a valuable resource for basal eudicot research	GIGASCIENCE	2019	Strijk JS외
48	An integrated chromosome-scale genome assembly of the Masai giraffe (<i>Giraffa camelopardalis tippelskirchi</i>)	GIGASCIENCE	2019	Farre M외
49	A chromosome-level draft genome of the grain caphid <i>Sitobion miscanthi</i>	GIGASCIENCE	2019	Jiang X외
50	A high-quality genome assembly for the endangered golden snub-nosed monkey (<i>Rhinopithecus roxellana</i>)	GIGASCIENCE	2019	Wang L외
51	Chromosomal-level assembly of the blood clam, <i>Scapharca (Anadara) broughtonii</i> , using long sequence reads and Hi-C	GIGASCIENCE	2019	Bai CM외
52	De novo genome assembly of the endangered <i>Acer yangbiense</i> , a plant species with extremely small populations endemic to Yunnan Province, China	GIGASCIENCE	2019	Yang J외
53	Pseudomolecule-level assembly of the Chinese oil tree yellowhorn (<i>Xanthoceras sorbifolium</i>) genome	GIGASCIENCE	2019	Bi QX외
54	Genome sequence of the barred knifejaw <i>Oplegnathus fasciatus</i> (Temminck & Schlegel, 1844): the first chromosome-level draft genome in the family Oplegnathidae	GIGASCIENCE	2019	Xiao YS외
55	Genome sequence of <i>Malaria oleifera</i> , a tree with great value for nervonic acid production	GIGASCIENCE	2019	Xu CQ외
56	A near-chromosome-scale genome assembly of the gemsbok (<i>Oryx gazella</i>): an iconic antelope of the Kalahari desert	GIGASCIENCE	2019	Farre M외
57	Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis	GIGASCIENCE	2018	Gong GR외
58	A draft genome assembly of the Chinese sillago (<i>Sillago sinica</i>), the first reference genome for Sillaginidae fishes	GIGASCIENCE	2018	Xu SY외
59	Draft genome of <i>Glyptosternon maculatum</i> , an endemic fish from Tibet Plateau	GIGASCIENCE	2018	Liu HP외
60	Long-read sequencing and de novo genome assembly of <i>Ammopiptanthus nanus</i> , a desert shrub	GIGASCIENCE	2018	Gao F외
61	High-quality assembly of the reference genome for scarlet sage, <i>Salvia splendens</i> , an economically important ornamental plant	GIGASCIENCE	2018	Dong AX외
62	Draft genome of the Peruvian scallop <i>Argopecten purpuratus</i>	GIGASCIENCE	2018	Li C외
63	Draft genome of the milu (<i>Elaphurus davidianus</i>)	GIGASCIENCE	2017	Zhang CZ외
64	Draft genome of the reindeer (<i>Rangifer tarandus</i>)	GIGASCIENCE	2017	Li ZP외
65	Novel transcriptome resources for three scleractinian coral species from the Indo-Pacific	GIGASCIENCE	2017	Kenkel CD외
66	Global sea-surface iodide observations, 1967-2018	SCIENTIFIC DATA	2019	Chance RJ외
67	A de novo genome assembly of the dwarfing pear rootstock Zhongai 1	SCIENTIFIC DATA	2019	Ou CQ외

논문번호	데이터논문 제목	학술지명	연도	저자
68	Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset	SCIENTIFIC DATA	2019	Syednasrollah B외
69	The sequencing and de novo assembly of the <i>Larimichthys crocea</i> genome using PacBio and Hi-C technologies	SCIENTIFIC DATA	2019	Chen BH외
70	The sequence and de novo assembly of <i>Takifugu bimaculatus</i> genome using PacBio and Hi-C technologies	SCIENTIFIC DATA	2019	Zhou ZX외
71	A draft genome for <i>Spatholobus suberectus</i>	SCIENTIFIC DATA	2019	Qin SS외
72	Activity/exercise-induced changes in the liver transcriptome after chronic spinal cord injury	SCIENTIFIC DATA	2019	Chariker JH외
73	Transcriptome of dorsal root ganglia caudal to a spinal cord injury with modulated behavioral activity	SCIENTIFIC DATA	2019	Chariker JH외
74	Draft genomic and transcriptome resources for marine chelicerate <i>Tachypleus tridentatus</i>	SCIENTIFIC DATA	2019	Liao YY외
75	A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults	SCIENTIFIC DATA	2019	Babayan A외
76	A functional connectome phenotyping dataset including cognitive state and personality measures	SCIENTIFIC DATA	2019	Mendes N외
77	The sequence and de novo assembly of <i>Oxygymnocypris stewartii</i> genome	SCIENTIFIC DATA	2019	Liu HP외
78	The sequence and de novo assembly of hog deer genome	SCIENTIFIC DATA	2019	Wang W외
79	Building an octaploid genome and transcriptome of the medicinal plant <i>Pogostemon cablin</i> from Lamiales	SCIENTIFIC DATA	2018	He Y외
80	Data Descriptor: Transcriptomes of cochlear inner and outer hair cells from adult mice	SCIENTIFIC DATA	2018	Li Y외
81	Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery	SCIENTIFIC DATA	2018	Richardson AD외
82	Data Descriptor: RNA-seq transcriptomic analysis of adult zebrafish inner ear hair cells	SCIENTIFIC DATA	2018	Barta CL외
83	A compendium of multi-omic sequence information from the Saanich Inlet water column	SCIENTIFIC DATA	2017	Hawley AK외
84	A compendium of geochemical information from the Saanich Inlet water column	SCIENTIFIC DATA	2017	Torres-Beltran M외
85	Data Descriptor: A hybrid organicoorganic perovskite dataset	SCIENTIFIC DATA	2017	Kim C외
86	A polymer dataset for accelerated property prediction and design	SCIENTIFIC DATA	2016	Huan TD외

