

국내 기관 소속 연구자의 데이터 가용성 진술 (Data Availability Statements) 현황 연구: PLOS ONE 학술지를 중심으로

A Study on the Data Availability Statements of Researchers Affiliated with Korean Institutions: Focusing on the PLOS ONE

안병균 (Byoung-Goon An)*

변제연 (Jeayeon Byun)**

초 록

본 연구는 국내 연구자가 저술한 논문의 데이터 가용성 진술(DAS)에 명시된 데이터 공유 메커니즘과 리포지토리를 조사함으로써 국내 연구자의 연구데이터 공유 현황과 특징을 탐구하는 것을 목적으로 한다. 이를 위하여 2014년부터 2022년까지 PLOS ONE에 게재된 국내기관 소속 연구자의 논문을 연구의 대상으로 선정하였다. 우선 논문 내 DAS 존재 현황을 파악하고 선행연구를 활용하여 데이터 공유 메커니즘의 유형을 분석하였으며, 시간의 흐름에 따른 데이터 공유 메커니즘별 변화 추이 등을 조사하였다. 그 결과, 대상 논문의 99.6%에 DAS가 작성되어 있으며 데이터 공유 메커니즘의 유형별 언급 양상은 국제적인 양상과 유사하되, 시간의 흐름에 따라 선호되는 유형이 변화하고 있음을 파악하였다. 이후 데이터 공유 메커니즘 중 리포지터리에 주목하여 DAS에 언급된 리포지터리의 횟수와 비율을 파악하고 다수 언급된 5개 리포지터리의 이용 변화 추이를 시계열적으로 분석하였다. 또한 리포지터리와 함께 언급된 데이터 접근점의 제시 방식과 유형, 유효성 등도 함께 조사하였다. 이를 통해 빈번하게 언급되는 상위 5개 리포지터리가 전체 리포지터리 언급의 60%를 차지하며 데이터 코드를 다루는 리포지터리의 이용이 증가하는 현상이 확인되었고, 리포지터리와 함께 제시된 데이터의 접근점은 대부분 유효하다는 사실을 파악할 수 있었다.

ABSTRACT

The purpose of this study is to investigate the current status and characteristics of research data sharing by domestic researchers by analyzing the data sharing mechanism and repository specified in DAS of papers authored by domestic researchers. To this end, in this study, papers of researchers belonging to domestic institutions published in PLOS ONE from 2014 to 2022 were selected as the subject of the study. First of all, the status of DAS's existence in the papers was identified, the types of data-sharing mechanisms were analyzed using precedent studies, and the trend of changes in each data-sharing mechanism over time was investigated. As a result, it was found that DAS was written in 99.6% of the target papers, and the types of data-sharing mechanisms were similar to international patterns, but preferred types were changing over time. Afterward, focusing on repositories among data sharing mechanisms, the number and ratio of repositories mentioned in DAS were identified, and the trend of changes in use of the five repositories mentioned a lot was analyzed in a time series. In addition, the presentation method, type, and validity of the data access point mentioned along with the repository were also investigated. It was confirmed that the top five frequently mentioned repositories account for 60% of all repository mentions, and the use of a repository dealing with data codes is increasing; in addition, it was found that most of the data access points presented with the repository were valid.

키워드: 연구데이터, 데이터 가용성 진술, 연구데이터 공유, 연구데이터 리포지터리, PLOS ONE
research data, Data Availability Statements(DAS), research data sharing, research data repository,
PLOS ONE

* 한국과학기술정보연구원 박사후연구원(rns1305@gmail.com) (제1저자)

** 성균관대학교 문헌정보학과 초빙교수(bjy0228@gmail.com) (교신저자)

- 논문접수일자: 2023년 2월 20일 ■ 최초심사일자: 2023년 3월 7일 ■ 게재확정일자: 2023년 3월 13일
- 정보관리학회지, 40(1), 225-258, 2023. <http://dx.doi.org/10.3743/KOSIM.2023.40.1.225>

© Copyright 2023 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

‘학술 정보에 대한 자유로운 접근 및 이용’의 개념을 지닌 오픈액세스(Open Access)는 2002년 BOAI(Budapest Open Access Initiative) 선언을 통해 처음 제시된 후, 전 세계적으로 수용되어 다양한 방식으로 실현되어 왔다. 오픈액세스에서 더 나아가, OECD(Organization for Economic Cooperation Development)에서는 2004년부터 공공 연구 성과물에 대한 접근성 제고 차원에서의 오픈 사이언스(Open Science)를 논의해 왔으며, 이는 학술지 논문의 무료 공개를 중심으로 전개된 오픈액세스를 넘어서 출판물뿐만 아니라 음성, 영상, 시료, 기록 등 연구의 수행 과정에서 산출된 데이터까지 공개하는 것에 초점을 맞추고 있다. 오늘날 디지털 기술의 급속한 발전과 확산에 따라 오픈 사이언스는 ‘디지털 기술을 통해 연구의 전 과정을 보다 개방적으로 전환하고자 하는 일련의 움직임’으로서 정의되고 있다(김소영 외, 2016; 신은정, 정원교, 2017; OECD, 2016). 한편, 데이터의 접근과 재이용을 위한 장기적 관리에 초점을 맞추고 있는 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙이 2014년 네덜란드 라이덴에서의 Jointly Designing a Data Fairport 워크숍을 통해 제안된 바 있다. 데이터의 탐색가능성(Findability), 접근가능성(Accessibility), 상호운용성(Interoperability), 재이용성(Reusability)의 제고를 위하여 DOI(Digital Object Identifier)와 같은 고유 식별자, 신뢰할 수 있는 데이터 리포지터리, 광범위한 데이터의 접근 및 이해를 지원하는 메타데이터, 데이터 라이선스 등이 요구됨을 강조

하였으며(Wilkinson et al., 2016), 이 원칙은 특정 학문 분야를 넘어선 학제 간 오픈 사이언스의 영역으로 점차 확산되고 있다(Cousijn et al., 2018; Helliwell et al., 2019; Stall et al., 2019).

이와 같은 흐름에 따라 데이터 공개 및 공유의 사례가 빠르게 증가하고 있으며, 주요 학술지 출판사에서도 연구의 신뢰성, 재현성 등의 제고를 목적으로 연구에서 생산·활용된 데이터 및 해당 소재를 공개할 것을 요청하고 있다. 그 가운데 PLOS(Public Library of Science)는 2014년 3월부터 학술지 게재 논문의 기초가 되는 데이터에 대한 공개적 접근을 촉진하기 위하여 데이터 정책을 실시한 최초의 출판사이며, 이후 Nature, Science, Elsevier 등의 출판사에서도 유사한 정책을 마련하였다. 이러한 정책은 데이터의 재이용을 가능하게 하고 연구 결과의 검증용 용이하게 하며, 저자를 위한 데이터 인용이 이루어지도록 지원하는 것을 공통된 목표로 하고 있다. 그러나 출판사에 따라 데이터 공유를 ‘권장(encourage)’하거나 ‘요구(require)’하는 등 정책의 의무화 정도에는 차이가 존재한다. 공식적인 요구 사항이나 특정 인센티브가 없는 한 연구자가 적극적으로 데이터를 공유할 가능성이 희박하므로 출판사의 ‘권장’과 ‘요구’는 실제로 얼마나 많은 데이터가 공개되어 이용될 수 있는가에 큰 영향을 미칠 수 있다(Federer et al., 2018; Jiao, Li, & Fang, 2022).

PLOS에서는 데이터 정책으로서 논문 관련 데이터의 공개·공유 여부와 접근 방법을 명시하는 ‘데이터 가용성 진술(Data Availability Statements, 이하 DAS)’의 제출을 ‘요구’하고

있다. PLOS 학술지에 원고를 제출하는 과정에서 저자는 DAS를 통해 '연구 결과의 기초가 되는 모든 데이터의 제한 없는 공개적 이용 방법' 및 '법적·윤리적 제한으로 공개 불가능한 데이터에 대한 접근 방법'을 설명해야 하고, 작성된 DAS는 출판된 논문과 함께 게시된다. 저자는 '최소한의 데이터 세트(Minimal Data Set)'로서 관련 메타데이터 및 방법론과 함께 논문의 모든 연구 결과를 재현하는데 필요한 데이터를 반드시 공유해야 하며, 정책의 권고에 따라 '데이터 리포지터리 저장', '논문 내 게시', '보충자료(Supporting Information file) 제출' 등의 방법을 통하여 공개적으로 데이터를 이용할 수 있도록 해야 한다. 또한 정책에서는 법적·윤리적 제한으로 인해 공개할 수 없는 데이터에 있어서도 DAS를 통해 해당 제한 사항을 명확히 밝혀야 한다고 규정했으며, 접근을 제한하는 제3자가 소유한 데이터, 인간을 대상으로 진행된 연구의 데이터 및 기타 민감 데이터 등에 한하여 접근 제한을 허용하고 있다. 논문 출판은 이러한 정책의 준수를 조건으로 하며, 출판 이후 데이터 정책에 위배되는 문제가 확인될 시 PLOS에서 저자 및 저자의 소속기관, 연구비 지원 기관 등에 연락을 취하거나 극단적인 경우 출판을 철회할 수 있음을 밝히고 있다. 한편, 공개적 데이터 공유 방법 중 '데이터 리포지터리 저장'은 PLOS 정책에서 강력히 권장(strongly recommended)하는 대표적 방법이다. 제출된 원고 내에 포함('논문 내 게시')되지 않은 모든 기초 데이터 및 관련 메타데이터는 적절한 공개 데이터 리포지터리에서 보관·관리되도록 하고, 리포지터리는 특정 유형의 구조화된 데이터를 취급하는 주제 분야 특정 리포지터리

(Subject-Specific Repositories) 또는 다양한 유형의 데이터를 수용하는 교차학문적 일반 리포지터리(Cross-Disciplinary Generalist Repositories)를 선택할 것을 제안하고 있다. 그 외 가능한 데이터 공개 방법으로서 '보충자료 제출'을 통한 경우에는 해당 분야에서의 표준 형식에 따른 데이터를 제출하거나 데이터의 효율적 추출이 가능한 파일 형식으로 제출함으로써 데이터에 대한 접근 및 재이용을 최대화할 것을 명시하고 있다. 뿐만 아니라 PLOS에서는 제출된 모든 '보충자료'를 논문 출판 시 FigShare에 업로드하여 FAIR 원칙 준수의 제고를 도모하고 있다(PLOS, 2019). 이를 통해 데이터 공개 및 공유 활성화를 위한 가장 효과적이고 효율적인 방법으로서 데이터 리포지터리 활용이 다시 한 번 강조되고 있음을 알 수 있다.

이처럼 PLOS의 데이터 정책은 예외적 경우를 제외한 모든 데이터의 공개 및 소재 파악을 위하여 DAS 제출을 의무화하고 있으며, 출판 논문의 일부로 게시되는 DAS는 데이터의 공유 여부 및 접근 방법에 대한 보다 상세한 정보 제공을 가능하게 한다(Jiao, Li, & Fang, 2022). 이와 같은 DAS는 연구자의 데이터 공유 행위를 조명해 볼 수 있는 실제적 정보원이라고 할 수 있으며, 따라서 DAS에서의 데이터 공유 여부와 접근 방법에 대한 진술이 어떻게 이루어지고 무슨 특징을 지니고 있는지, 또한 강력히 권장되는 데이터 공유 방법으로서의 데이터 리포지터리는 무엇이 제시되며 어떤 특징을 보이는지에 대해 살펴볼 필요가 있을 것이다. 앞서 언급한 바와 같이 DAS 정책은 2014년 초 PLOS의 주도로 시작되어 주로 국외에서 운영되고 있으며, 현재까지의 시행기간이 10년에도 미치지

못함에 따라 관련 논의와 연구가 세계적으로 매우 드물고 국내의 경우 DAS에 대한 고찰이나 정책이 미비한 실정이다. 따라서 DAS 정책을 시행하고 있는 국외 학술지에 게재된 국내 연구자 논문의 DAS 현황을 파악함으로써 국내에서 아직 생소한 개념인 DAS에 대한 이해를 증진시킬 수 있을 뿐 아니라, 국내 연구자의 데이터 공유 행위에 대한 실증적 확인을 함으로써 향후 국내의 DAS 정책 마련 및 데이터 공개·공유의 발전 방향을 모색하기 위한 기초를 다질 수 있을 것이다.

이에 본 연구는 PLOS의 학술지 게재 논문 중 국내 연구자 논문의 DAS에서 파악되는 데이터 공유 메커니즘 및 데이터 리포지터리를 분석함으로써 국내 연구자의 데이터 공유 현황과 특징을 살펴보는 것을 목적으로 한다. 이를 위해 국내 기관 소속 연구자가 PLOS ONE에 게재한 논문을 중심으로, 첫째, DAS의 존재 여부, DAS에 제시된 데이터의 공개 여부, DAS 작성 논문의 인용 양상 등 일반적 현황을 조사하였다. 둘째, DAS의 데이터 공유 메커니즘 유형에 대한 구체적인 분석으로서 국제적 차원에서의 비교, 유형별 언급 비율, 시간 흐름에 따른 변화 추이, 논문 인용 차이를 확인하였다. 셋째, DAS에 언급된 리포지터리에 주목하여 리포지터리의 종류별 비율을 파악하고 상위 리포지터리의 시계열적 변화 추이와 논문 인용 차이를 분석하였다. 마지막으로, DAS에 제시된 리포지터리 접근점의 유형별 비율과 실제 접근 가능 여부를 확인하고, 시계열적 변화 추이와 논문 인용 차이에 대한 조사를 실시하였다.

2. 선행연구

2.1 DAS에 관한 연구

학술지에서 연구데이터 공유 활성화를 위한 방안으로 DAS에 대해 논의하고 정책으로 도입한 것이 2014년 무렵이므로, DAS 자체에 주목한 연구는 국내뿐 아니라 국제적으로도 많지 않다. 현재 DAS에 대한 연구는 기본적으로 특정 학술지의 DAS 작성 내용에 대해 분석한 후 DAS 유형과 변화 양상을 파악하고 작성 소요 시간 등을 산출하거나, DAS에 기술된 리포지터리 등에 대한 세부적인 조사를 하는 방식으로 진행되고 있다.

McDonald et al.(2017)은 British Medical Journal(이하 BMJ)에 게재된 논문에서 연구데이터 공유를 위하여 어떠한 내용을 명시하고 있는지 분석하고자 2015년 1월 1일부터 2017년 8월 31일 사이에 발표된 관찰 연구 유형의 논문 237편을 연구의 대상으로 설정하였다. 데이터를 사용할 수 없다고 명시한 경우 외에도 논문 내 명시된 데이터 외 추가적인 데이터가 없는 경우, 데이터가 존재한다고 하더라도 구체적인 접근 방법이 명시되지 않은 경우를 모두 데이터가 공유되지 않은 것으로 간주하였다. 분석 결과 63%의 논문에서 연구데이터를 공유하지 않고 있다는 사실을 확인하였고, 이러한 현상은 데이터 공유를 권장하지만 특정한 요구 사항은 명시하지 않은 당시 BMJ의 데이터 정책과 연관이 있을 것이라고 추정하였다.

Grant와 Hrynaszkiewicz(2018)는 Nature 학술지의 DAS 도입이 저자와 편집자에게 어떠한 영향을 미치는지를 연구하였다. 이를 위해

여 2016년 3월부터 5월까지 2개월 간 Nature의 5개 세부 분과(Nature Neuroscience, Nature Physics, Nature Communications, Nature Cell Biology, Nature Geoscience)에 게재된 학술논문의 DAS를 분석의 대상으로 선정하고 이를 4가지 유형으로 분류하였다(from the author on request, included in the manuscript or its supplementary material, data is publicly available(repository), included with the manuscript). Grant와 Hrynaszkiewicz(2018)는 세부 분과별 DAS 작성 개수와 유형의 분포를 분석하였을 뿐만 아니라 논문에 DAS를 작성하기 위하여 소요되는 편집 시간을 측정하였으며, 이를 통해 논문당 약 15분~20분의 편집 및 제작 시간이 소요된다는 사실을 확인하였다.

Federer et al.(2018)은 PLOS ONE에서 DAS 정책이 실시된 2014년 3월부터 2016년 5월까지 게재된 학술논문을 대상으로 DAS 분석을 진행하였다. 분석 대상으로 선정한 표본의 수가 상당했으므로(논문 약 50,000편) 자동 클러스터링 방식을 기반으로 하되 자동으로 분류되지 않는 일부에 대해서는 수동으로 분류하였다. 연구에서는 DAS의 데이터 공유 메커니즘을 10가지로 유형화(access restricted, combination, in paper, in paper and SI, in SI, location not stated, N/A, other, repository, upon request)하고 그 기준과 클러스터링 방법을 명시하였다. 분석을 통해 연도 변화에 따른 공유 메커니즘별 변화 양상을 확인하였고, DAS에서 다수 언급되는 리포지터리 20개를 제시하였다.

Jiao, Li, Fang(2022)은 2014년부터 2020년까지 PLOS ONE 저널에 게재된 논문(228,505편)의 DAS를 추출하고 R 프로그래밍을 기반으

로 한 분석을 진행하였다. 이 과정에서 Federer et al.(2018)이 제시한 10가지 연구데이터 공유 메커니즘 유형을 사용하되, 분석 결과를 고려하여 단순화함으로써 6가지 메커니즘(in paper, in SI, in paper and SI, repository, Upon request, Combination)으로 유형을 구분하였다. 시간의 흐름에 따른 공유 메커니즘의 변화 양상과 주로 사용되는 연구데이터 리포지터리를 분석하였을 뿐만 아니라, 분석 대상이 된 논문을 5가지 학문분야로 구분하고 학문분야별 데이터 공유 메커니즘 선택 특징과 리포지터리 이용 변화 추이에 대한 조사도 진행하였다.

Graf et al.(2019)은 2013년부터 2019년 사이 Wiley의 176개 저널에 제출한 논문 124,000편의 DAS를 분석하고자 토픽 모델링을 진행하였다. 토픽 모델링을 기반으로 DAS의 유형을 20가지로 구분하였으며, 이 과정에서 동일한 유형에 속한다고 할지라도 구체적인 기술 내용이 다른 경우 이를 세분하였다. 이후 각 유형에서 특징적으로 언급되는 용어와 시간의 흐름에 따른 유형별 언급 빈도 추이를 분석하였으며, DAS 내 연구데이터 공유에 대한 용어의 빈도가 낮은 점을 언급하면서 실질적으로 연구데이터의 공유를 보장하는 DAS의 작성을 위해서는 연구 설계 단계에서부터 데이터를 계획하는 DMP가 중요하다고 언급하였다.

Page et al.(2022)은 PubMed, Scopus 등 다양한 출처에서 수집된 300편의 연구논문을 대상으로 DAS를 분석하였다. 연구에서는 DAS가 연구의 과정에서 생산된 데이터뿐 아니라 코드를 공유하기 위한 역할도 수행한다는 견해 하에 데이터와 코드의 이용 및 접근 가능성에 대한 언급을 종합적으로 조사하였다. 이를 기반

으로 DAS 진술 중 데이터 파일이나 코드를 공개하는 경우는 극히 적고 요청시 데이터를 제공한다는 진술도 소수에 그친다는 점을 지적하였다. 데이터 공유를 증가시키고 DAS가 일상적으로, 올바르게 작성되도록 만들기 위해서는 데이터 공유에 관한 교육 프로그램과 데이터 공유에 대한 보상 제도 등이 필요하다고 주장하였다.

Colavizza et al.(2020)은 논문 인용의 측면에서 DAS의 이점을 파악하는 연구를 진행하였다. BMC와 PLOS에서 출판된 논문을 대상으로 조사하여 데이터 공유 메커니즘을 네 가지 유형으로 분류하고(Not available, Available on request or similar, Available with the paper and its supplementary files, Available in a repository), 데이터 공유 메커니즘 유형과 인용 간의 상관관계를 분석하였다. BMC는 'Available on request or similar'이, PLOS는 'Available with the paper and its supplementary file'이 진술된 DAS의 절반 이상을 차지하면서 학술지의 정책에 따른 DAS 양상 차이가 드러났으며, 학문 분야에 따라서도 지배적인 DAS 유형이 달리 나타났다. PubMed OA 데이터 세트를 통해 수집한 H-index와 Science-Metrix를 기반으로 분석한 결과, DAS 제공 유형과 논문의 인용 사이에 상관관계가 존재하며, 논문의 인용과 가장 높은 상관관계를 보이는 것은 리포지터리 링크를 통해 데이터를 제공하는 유형이라는 사실을 밝혔다.

McGuinness와 Sheppard(2021)는 프리프린트 단계에서 작성된 DAS가 논문으로 출판되는 과정에서 변경되는지 여부를 확인하고자 2019년 6월부터 2020년 5월까지 약 1년 사이에

medRxiv 리포지터리에 게시된 프리프린트 3,938건과 해당 프리프린트의 출판 논문의 DAS에 대한 분석을 진행하였다. DAS 유형을 9가지로 세분화되, 논문 혹은 SI에 관련 데이터가 존재한다고 명시한 유형이나 온라인 리포지터리를 제시한 유형은 'Open'으로, 연구 과정에서 데이터가 생산되지 않았다고 응답한 유형은 'Not applicable'로, 그 외 모든 비공개 유형에 대해서는 'Closed'로 분류하였다. 동일한 분류 기준을 프리프린트와 출판 논문 두 가지 모두에 적용함으로써 프리프린트의 DAS와 학술논문의 DAS 사이에 차이가 있는지 확인했다. 그 결과 대부분 DAS의 변화가 없었으나 학술논문의 DAS가 프리프린트의 DAS보다 공개적으로 변화하는 경우가 그 반대의 경우보다 많다는 사실을 발견하였으며, 데이터 공개를 의무로 명시한 학술지에 투고된 논문에서 이러한 경향이 더욱 두드러진다는 사실을 밝혔다.

2.2 데이터 리포지터리에 관한 연구

국외에서는 데이터 리포지터리의 유형 및 특성을 살펴본 초기 연구를 시작으로, 리포지터리의 메타데이터 분석, 특정 리포지터리 사례 조사 등의 연구가 수행되어 왔으며, 비교적 최근에는 리포지터리의 개선방향을 모색한 연구도 확인된다.

데이터 리포지터리의 유형과 특성을 조사한 연구로, Marcial, Hemminger(2010)는 Google 검색을 통해 파악된 100개의 데이터 리포지터리를 정부, 의학/소규모, 대학, 커뮤니티 중심 생물학 리포지터리의 4개 그룹으로 제시했으며, 리포지터리의 성공·지속가능성 관련 요인

으로 보조금 및 계약을 통한 지원 여부, 다수의 후원 여부, 보유 데이터 규모, 보존 정책 존재 여부의 4가지를 확인하였다. 이후 2012년부터 데이터 리포지터리 색인화를 시작한 re3data를 기반으로 수행된 연구(Magazine, 2017; Pampel et al., 2013)에서는 이질적인 리포지터리 환경을 설명하면서 리포지터리 유형을 기관, 분야, 다학제, 프로젝트에 따라 범주화 하였으며, 리포지터리의 서비스 수준, 지원하는 언어, 준수하는 표준 등에서의 차이를 발견하였다.

데이터 리포지터리의 메타데이터를 살펴본 연구에서는 주로 세부 학문분야에 초점을 맞추고 있다. re3data의 사회과학 및 인문학 분야 데이터 리포지터리를 대상으로 메타데이터를 조사한 Gómez, Méndez, Hernández-Pérez(2016)는 사회과학-Data Documentation Initiative (DDI), 인문학-Dublin Core(DC)의 일반적인 적용 경향을 확인한 한편, 인문학 분야 리포지터리의 경우 언어학, 고고학 등 특정 분야에서의 자체 체계 사용 및 DC 일부 적용과 같은 보다 복잡하고 다양한 특징이 있음을 파악했다. Löffler et al.(2021)은 생물다양성 분야의 여러 이질적인 데이터를 취급하는 리포지터리가 학술적 정보 요구를 반영하는지를 조사하고자 FigShare, Dryad, GBIF, PANGEA 및 Zenodo의 메타데이터를 분석하였다. Dublin Core, DataCite 등 일반적 표준이 주로 적용된 것을 확인했으며, 이에 분야별 메타데이터 표준의 다양한 도입을 통한 생물다양성 분야 데이터의 보다 풍부하고 정확한 검색 및 이용을 제안하였다.

특정 데이터 리포지터리의 사례 연구로는, 정량적 관점에서 FigShare, Zenodo, Dryad가 조

사된 바 있다. FigShare의 주제 분야 내 연구자 수 및 데이터 수가 데이터의 이용 및 공유 건수와 비례하지 않는 것으로 나타나, 분야의 경계를 넘어선 리포지터리의 운영이 강조됐으며 (Thelwall & Kousha, 2016), Zenodo의 경우, 2017년 1월까지의 DOI 부여 데이터 141,777건 중 논문의 약 13%가 Web of Science Core Collection에, 데이터 세트 및 소프트웨어의 약 80%가 Data Citation Index에 색인된 것을 통해 Zenodo의 주요 정보원으로서의 역할이 확인되었다(Gorraiz et al., 2017). 한편, 진화생물학 분야 학술지 Evolution과 Heredity를 중심으로 Dryad를 살펴본 연구(Thelwall & Kousha, 2017)에서는 Dryad의 논문 이용과 별개로 이루어지는 관련 데이터 다운로드, Dryad에 데이터가 있는 논문과 없는 논문 간 미미한 인용률 차이, 데이터 다운로드와 논문 인용 횟수 간 낮은 상관관계가 확인됐으며, 다운로드는 데이터 복합성 및 논문 이해를 위한 필요 여부에 영향을 받을 수 있음을 언급하였다.

데이터 리포지터리의 개선방향을 모색한 연구로 Mannheimer et al.(2019)은 질적 데이터 공유에 있어 연구자와의 협력 방법을 ① 데이터 공유와 재이용에 대한 연구자 동의, ② 데이터의 법적·윤리적 공유 보장을 위한 데이터 관리 및 공유 계획 수립, 데이터 익명화, 메타데이터 지원, ③ 익명화 불가 데이터 접근 제한의 3가지 측면에서 제시하였다. 이 외에도 re3data의 리포지터리 대상 설문조사를 통하여, 주요 과제로서 이용자 및 인적자원 부족, 시스템 자동화 및 데이터와의 상호운용성 문제를 파악한 연구(Khan, Thelwall, & Kousha, 2022), 보다 성공적인 데이터 품질 보증을 위한 방안으로

데이터 검수자 및 리포지터리의 기여에 대한 인식, 데이터 검수의 경로의존성, 데이터 품질 보증 조치 관련 정보 부족 등의 문제가 해결되어야 함을 확인한 연구(Kindling & Strecker, 2022)가 최근 수행되었다.

국내의 경우, 데이터 리포지터리 정책을 살펴본 연구와 국내 대표 리포지터리의 개선방안을 제시한 연구가 초반부터 이루어졌으며, 리포지터리의 운영 수준 및 데이터 유형 분석, 리포지터리에 대한 국내 연구자 인식 조사 등의 연구가 근래까지 수행되어 왔다.

데이터 리포지터리의 정책에 관한 연구는 데이터 공유 및 재이용의 주요 정책 파악을 위해 국외 선진 사례를 중심으로 수행되었다. 정영미, 이상기(2010)는 CSIC의 Ranking Web of World Repositories와 JISC의 Registry of Open Access Repositories를 통해 사례를 선정했으며, 운영, 지원, 수집, 관리, 배포, 이용, 보존, 기술, 저작권 관련 정책을 분석하고 주제·기관 리포지터리 간 차이 및 국내·외 리포지터리 간 차이를 살펴보았다. 김지현(2016)은 데이터 접근 및 이용 통제 정책에 초점을 맞추어 Nature of Scientific Data 추천 기반 국외 리포지터리 37곳을 분석한 결과, 저작권 및 라이선스, 데이터 인용, 면책조항, 엠바고 적용 관련 규정이 공통적 정책 요소였고, 분야별 정책 요소의 다양성에는 차이가 있는 것으로 나타났다.

국내 데이터 리포지터리의 개선방안을 제시한 연구는 인문사회 분야의 대표적 리포지터리인 한국연구재단 기초학문자료센터(KRM)를 대상으로 하고 있다. KRM 분석, 국내외 리포지터리 사례 조사, 인문사회 분야 연구자 심층면담을 통해 국내 인문사회 데이터 아카이브의 개

념적 모델을 제안한 연구(신영란, 정연경, 2012), KRM 인문학 분야 데이터 수집 및 구축 운영자의 관점에서 실제 적용된 방법론을 소개하고 데이터의 체계적인 수집, 활용, 보존을 위한 프로세스와 전략을 제시한 연구(심원식, 안혜연, 변제연, 2015), 국외 주요 리포지터리를 조사하여 KRM의 개선점으로 데이터 관리의 중요성 인식 교육, DMP 및 메타데이터 작성 교육, 국제적 표준의 채택, 전담 인력의 확보 등을 제시한 연구(박규리, 안병근, 2017)가 있다.

데이터 리포지터리의 운영 수준 및 데이터 유형을 파악한 연구가 학문분야별로 이루어졌다. re3data의 305개 인문사회 분야 리포지터리 분석(조재인, 박종도, 2019) 및 402개 자연과학 분야 리포지터리 분석(김우중 외, 2021) 결과, 운영 수준에 따라 우수, 보편, 비공개, 열악의 4개 군집이 도출되었다. 인문사회 하위 분야의 데이터 유형은 역사 및 예술분야-이미지, 사회계열-통계데이터, 오피스문서, 언어학-오디오, 텍스트, 코드 등이 주류를 차지했으며, 자연과학 분야에서는 연구주제 및 리포지터리 방문 빈도에 따른 데이터 유형의 차이가 확인되었다.

데이터 리포지터리에 대한 인식을 조사한 연구로, 황혜경, 이지연(2016)은 설문조사 및 심층면담을 통해 물리, 수학, 생명과학 분야 국내 연구자의 기관 리포지터리 인지 경로 및 데이터 공유 현황을 파악하고, 데이터 공유에 따른 기대로 연구성과 인정, 우선권 확보, 연구역량 강화, 취업, 중복연구 가능성 축소 등을 확인하였다. 한편, 국내 연구자와 관계자의 인식 조사를 통해 AccessON의 Self Archiving과 Journal Repository의 경제적 가치를 측정한 연구(권나

현, 표순희, 2022)에서는 총 사업기간 8년간 Self Archiving은 1억 450만원, Journal Repository는 312억원, 두 서비스 통합의 경우 313.5억원으로 가치가 산출된 바 있다.

이상에서 살펴본 바와 같이, 그동안의 DAS 관련 논의와 연구는 주로 국외를 중심으로 제한적으로 이루어져 왔으며, 특히 DAS의 진술 내용에 초점을 맞추어 그 양상과 특징을 실질적으로 살펴본 연구는 극히 드물게 존재한다. 또한 국내외에서 수행된 데이터 리포지터리에 관한 연구는 리포지터리의 유형, 정책, 운영 및 관리, 특정 사례, 연구자 인식, 개선방향 모색 등 다양한 측면에서의 고찰이 이루어졌으나 여러 학문분야를 아우르는 차원에서 연구자가 사용하는 리포지터리의 종류와 특징을 실제적으로 파악한 연구는 찾아보기 어렵다. 이에 따라 국내 연구자의 DAS에서 나타나는 데이터 공유 메커니즘 및 제시되는 데이터 리포지터리에 대한 조사를 통해 데이터 공유 현황과 주요 특징을 확인하는 실증적 연구가 수행되어 향후 국내의 DAS 정책 마련 및 데이터 공개·공유를 위한 발전 방향의 모색에 활용될 필요가 있다.

3. 연구 방법

3.1 연구대상 선정

국내 기관 소속 연구자의 데이터 공유 현황을 조사하기 위하여 PLOS ONE 학술지 논문의 DAS를 연구대상으로 선정하였으며, 그 이유는 다음과 같다. 첫째, 앞서 언급한 바와 같이 PLOS는 데이터 공유 및 활용 촉진을 위해 2014

년 초부터 DAS 정책을 구현한 최초의 출판사로, 원고 제출 시 DAS를 통해 '연구 결과의 기초가 되는 모든 데이터의 제한 없는 공개적 이용 방법' 및 '법적·윤리적 제한으로 공개 불가능한 데이터에 대한 접근 방법'을 명시하도록 하고 있다. 둘째, PLOS의 학술지 중 2006년 창간된 PLOS ONE은 대표적인 오픈액세스 메가 학술지이다. 2013년에는 3만편 이상의 논문을 출판하였으며, 최근의 감소 추세에서도 매년 약 2만편의 논문을 다루는 세계적인 규모의 학술지로서 광범위한 학문분야를 포괄하는 논문의 경향을 살펴볼 수 있다. 셋째, PLOS ONE은 국내 기관 소속 연구자의 투고 빈도가 높은 학술지에 해당된다. 논문 출판 속도, 온라인 투고의 용이성 등의 이유로 국내 연구자가 PLOS ONE을 선호하고 있으며, 2006년부터 2019년까지 PLOS ONE에 게재된 국내 연구자의 논문은 약 6,500여 편으로 이는 세계 11위에 해당하는 수준이다(심원식, 2021; 심원식 외, 2020). 넷째, 선행연구(Federer et al., 2018; Jiao, Li, & Fang, 2022)를 통해 PLOS의 DAS 정책 시행 이후 출판된 대다수 PLOS ONE 논문에서 DAS의 게시가 확인된 바 있으며, 본 연구의 결과를 선행연구와 함께 살펴봄으로써 국내 기관 소속 연구자의 데이터 공유 현황을 국제적 차원에서 비교·분석할 수 있다.

3.2 데이터 수집 및 분석

국내 기관 소속 연구자의 PLOS ONE 논문을 추출하고자 대표적 인용 색인 데이터베이스인 Web of Science를 활용하였다. 데이터 수집은 2023년 1월 9일 이루어졌으며, 검색 조건

으로 ① 저널명(Publication Titles): “PLOS ONE”, ② 연구비 지원 기관(Funding Agency): “National Research Foundation Of Korea”, ③ 문서유형(Document Type): “Article”, ④ 출판 날짜(Publication Date): “2014-03-01 ~ 2022-12-31”을 적용하여 검색을 실시하였다. PLOS ONE 시스템에서의 직접적인 검색이 아닌 Web of Science를 통해 연구대상 논문을 검색한 것은 ② 연구비 지원 기관: “National Research Foundation Of Korea” 필드의 적용을 통해 ‘국내 기관 소속 연구자의 PLOS ONE 논문’ 범주에 최대한 부합하는 표본을 풍부하게 수집하고자 함에 있다. 한국연구재단은 우리나라를 대표하는 최대 규모의 연구비 지원 기관으로서 이공 분야 기초연구, 인문사회 분야 학술연구, 정책연구 등 국내의 다양한 연구 성과를 위한 지원에 주력하고 있으며, 연구기금논문의 대부분이 국외 오픈액세스 학술지에 출판되고 있다(정경희 외, 2021). 이에 따라 한국연구재단의 연구기금논문 가운데 본 연구의 표본을 수집하는 것이 적합하다고 판단하였다. ④ 출판 날짜: “2014-03-01 ~ 2022-12-31” 필드의 경우, PLOS의 DAS 정책 발효 시점인 2014년 3월 경부터 최근까지의 경향을 전반적으로 살펴보기 위한 목적으로 적용되었다. 검색된 총 2,276편 논문의 레코드 파일 다운로드를 통해 DOI 정보를 수집하였으며, DOI 랜딩페이지인 PLOS ONE 논문 페이지에서 Python을 활용하여 데이터 크롤링함으로써 분석대상 항목을 추출하였다. 논문 제목(Article Title), 저자(Authors), 제출일(Received), 게재승인일(Accepted), 출판일(Published), 데이터 가용성 진술(Data Availability Statements), 연구

비 지원 기관(Funding), 인용 수(Citation) 등의 정보를 수집했으며, 다음의 3단계에 걸친 표본 추출 작업이 진행되었다.

먼저, ‘제출일’을 기준으로 2014년 3월 3일 이전에 투고된 논문 141건을 제외하였다. 이는 PLOS의 “현 데이터 정책은 2014년 3월 3일부터 시행되었고, 해당 날짜 이전에 제출된 모든 문서에는 DAS가 없으며, 그 이전에 제출되거나 출판된 모든 원고의 경우 합당한 요청이 있을 시 데이터를 제공해야 한다”는 명시(PLOS, 2019)를 반영한 것이다. 앞선 Web of Science에서의 표본 수집 단계에서 ‘제출일’에 따른 검색이 불가했으므로 이와 같은 추출 작업을 진행하였다. 선행연구(Federer et al., 2018; Jiao, Li, & Fang, 2022)의 경우 ‘출판일’을 기준으로 2014년 3월부터 발행된 논문을 다루었으나 본 연구에서는 제출일을 고려함으로써 보다 실질적인 분석이 이루어질 수 있도록 하고자 하였다. 다음으로 저자의 ‘국내 기관 소속 연구자’ 해당 여부를 조사하고자 DOI를 통해 2,135건의 PLOS ONE 개별 논문 페이지에 접속하여 저자의 소속기관을 확인하였다. 본 연구에서의 ‘국내 연구자’를 한국인으로 특정하지 않고 ‘국내 기관 소속 연구자’로 선정한 것은 PLOS 시스템상 저자의 국적에 대한 식별이 불가함에 따라 한국인 특정에 한계가 있고, ‘국내 기관 소속 연구자’ 출판 논문의 ‘국내 연구성과’로서의 의의를 고려한 것이다. 이에 따라 단독·공저 논문의 모든 저자가 국적과 무관하게 국내 기관 소속 연구자에 해당될 경우로 범위를 국한하고 국내·국외의 다중 기관 소속 연구자는 포함되도록 한 결과, 분석대상에서 434건의 논문이 제외되었다. 마지막으로, ‘연구비 지원 기관’

에 National Research Foundation Of Korea가 해당되지 않는 논문 1건을 Web of Science의 검색결과 오류로 파악하여 이를 제외하였다. 해당 1건 이외의 논문 모두가 한국연구재단 한 기관만을 통해 연구비를 지원받은 것은 아니며, 한국연구재단 이외에도 국내외 연구비 지원 기관의 복수 지원을 받은 경우가 다수 존재한다. 본 연구에서는 한국연구재단의 지원이 적용된 경우에 중점을 두어 표본의 동질성을 최대한 고려하고자 하였다.

이상의 표본 추출 작업에 따라 최종적으로 총 1,700편의 논문이 분석대상으로 선정되었으

며, 해당 논문들의 DAS에 대한 내용분석을 실시하였다. 분석을 위한 코드(Code) 및 정의는 Federer et al.(2018)의 연구 결과로 도출된 <표 1>의 DAS 데이터 공유 메커니즘 유형을 적용하였다.

본 연구에서는 <표 1>을 분석의 틀로 삼되, '㉑ 제한적 접근(access restricted)' 코드에 대해서는 별도의 범주로 구분하여 분석을 수행하였다. 이는 나머지 9가지 코드의 경우 모두 '데이터의 소재정보 혹은 존재여부'에 중점을 둔 반면, 해당 코드는 '데이터의 공개여부'에 보다 초점을 맞추고 있기 때문이다. 또한 실제 사전

<표 1> Federer et al.(2018)의 DAS 데이터 공유 메커니즘 유형

코드	정의	해당 사례
① 논문 내 (in paper)	데이터가 표, 그림 등을 포함해서 논문 내에 있음을 명시함	연구의 결과를 뒷받침하는 최소한의 데이터 세트는 논문 내에 있음.
② 보충자료 내 (in SI)	데이터가 보충자료(Supporting Information)내에 있음을 명시함	모든 데이터 및 분석 코드는 보충자료로 제공됨.
③ 논문 및 보충자료 내 (in paper and SI)	데이터가 논문 및 보충자료의 두 가지 모두에 존재함을 명시함	모든 관련 데이터는 논문과 보충자료를 통해 이용할 수 있음.
④ 리포지터리 (repository)	데이터의 이용이 가능한 공개적 접근 위치(리포지터리 및 웹사이트)를 명시함	모든 샘플에 대한 데이터 세트는 <리포지터리>의 <접근번호(accession number)>를 통해 이용할 수 있음.
⑤ 요청 시 (upon request)	데이터에 접근하기 위하여 저자, 개인, 그룹 등에 연락을 취해야 함을 명시함	데이터는 <저자명>의 <이메일 주소>로 요청하여 이용할 수 있음.
⑥ 위치 불명 (location not stated)	데이터의 이용 가능함을 언급하고 있으나 데이터의 위치 또는 접근 방법은 명시되지 않음	저자는 연구 결과의 기초가 되는 모든 데이터를 제한 없이 완전히 이용할 수 있음을 확인함. 데이터 기탁.
⑦ 해당 없음 (N/A)	DAS가 작성되어 있으나, 해당 없음(Not Applicable 또는 N/A)이 함께 언급됨	저자는 연구 결과의 기초가 되는 모든 데이터를 제한 없이 완전히 이용할 수 있음을 확인함. 해당 없음.
⑧ 기타 (other)	9가지 범주의 어디에도 해당되지 않음	저자는 연구 결과의 기초가 되는 모든 데이터를 제한 없이 완전히 이용할 수 있음을 확인함. 이 논문은 이론적 논의에 해당되므로 관련 데이터가 없음.
⑨ 제한적 접근 (access restricted)	윤리적·법적 문제 및 개인정보 보호와 관련한 접근 제한을 언급하거나, 접근을 제한하는 제3자가 데이터를 소유함을 언급함	심각한 범죄로부터 식별 가능한 정보를 보호하기 위하여, 데이터 접근을 위한 윤리적 승인이 필요함. 기밀 데이터에 대한 접근 기준을 충족하는 연구원을 위해 <출처>에서 데이터를 사용할 수 있음. 자세한 내용은 <웹페이지>를 참조.
⑩ 조합 (combination)	데이터 공유 메커니즘을 한가지 이상 언급함	논문 또는 보충자료에 포함되지 않은 모든 데이터는 <repository>(<URL>)을 통해 이용할 수 있음.

코딩 단계에서 '⑤ 요청 시(upon request)' 코드와 '⑨ 제한적 접근(access restricted)' 코드가 중복되는 다수의 사례가 확인되었으며, 코드 간 경계가 모호하여 이를 '⑩ 조합(combination)' 범주로 코딩하기에는 적절치 않은 것으로 판단하였다. 따라서 '⑨ 제한적 접근(access restricted)'을 제외한 9가지 코드를 중심으로 DAS 데이터 공유 메커니즘 유형을 파악하였으며, '데이터 공개여부'에 대해서는 '공개, 제한적 공개, 비공개, 복합, 확인불가'의 기준을 적용하여 전체 표본에 대한 별도의 분석을 실시하였다.

한편, DAS 데이터 공유 메커니즘에서 제시된 데이터 소재/존재 정보에 따른 실제 해당 데이터의 이용 가능 여부를 파악하고자, 공개 데이터 범주에 해당되는 '② 보충자료 내(in SI)', '③ 논문 및 보충자료 내(in paper and SI)' 및 '④ 리포지터리(repository)' 유형에 대한 실사를 진행하였다. '① 논문 내(in paper)' 유형의 경우, PLOS ONE의 게재 논문에 해당되는 본 연구의 표본에 대하여 이미 앞선 분석 과정에서 데이터(논문)가 공개되어 있음을 확인하였고, 사실상 대다수의 논문에서 표, 그림 등의 데이터를 다루고 있어 논문이 데이터의 기본적인 정보원이 될 뿐 아니라 다양한 연구 분야에 대한 맥락적 파악에는 한계가 있기 때문에 실사 대상에서 제외하였다. 이에 '② 보충자료 내(in SI)' 및 '③ 논문 및 보충자료 내(in paper and SI)' 유형에서의 SI 존재 여부, '④ 리포지터리(repository)' 유형에서의 데이터 접근점 제시 방식(DOI, URL, Accession Number: 고유접근번호, None: 제시 접근점 없음, Other: 기타), 제시 건수, 접근 가능 건수에 대한 확인을 진행하였다.

분석은 2명의 연구자에 의해 수행되었으며, 보다 일관성 있는 분석을 위하여 코드북 초안을 대상으로 두 연구자가 다르게 코딩을 한 경우 추가적인 논의를 진행하여 해당 유형을 재분류 및 정리하였다. 이를 통해 표본 전체에 대한 개별 분석 시 발생될 수 있는 차이를 최소화할 수 있도록 하였으며, 사후 논의 및 검토, 정제 작업을 통하여 분석결과의 신뢰성을 높였다.

4. DAS 분석 결과

4장은 DAS에 기술된 데이터 공유 메커니즘에 대한 분석 결과와 DAS에 언급된 리포지터리에 대한 분석 결과로 구성된다. 데이터 공유 메커니즘의 유형 구분을 보다 명확하게 제시하기 위하여 <표 1> 코드의 영어 표기에 따라 후술한다.

DAS에 기술된 데이터 공유 메커니즘을 설명하기에 앞서, 전체 분석 대상 논문 중 DAS가 존재하지 않는 논문을 조사하고 데이터의 공개 여부 분포를 확인하여 기술하였으며 DAS가 작성된 논문의 인용 수 평균에 대한 산출 결과를 설명하였다. 이후 본 연구와 선행연구의 데이터 공유 메커니즘 유형별 비율을 비교하고 시간의 흐름에 따른 언급 횟수 변화를 파악하여 제시하였다. 또한 전체 분석 대상 논문의 인용 수 평균을 기준으로 데이터 공유 메커니즘 유형별 인용 수 평균을 비교하여 기술하였다.

DAS의 공유 메커니즘에 대한 분석 이후, DAS에 언급된 리포지터리를 대상으로 진행한 세부적인 연구 내용을 상술하였다. 'repository'로 분류된 세부 유형을 조사하고 가장 많이 언급된

상위 10개 리포지터리를 파악하였으며 상위 리포지터리의 시계열적 변화 추이를 조사하였다. 리포지터리 자체에 대한 연구 뿐만 아니라 리포지터리에 명시된 접근점의 유형과 실제 접근 가능 비율에 대해 기술하였으며, 추가적으로 리포지터리와 접근점의 인용 수 및 인용 평균 분석 결과를 제시하였다.

4.1 DAS가 포함된 논문 분석 일반

4.1.1 DAS 존재 논문 비율 분석

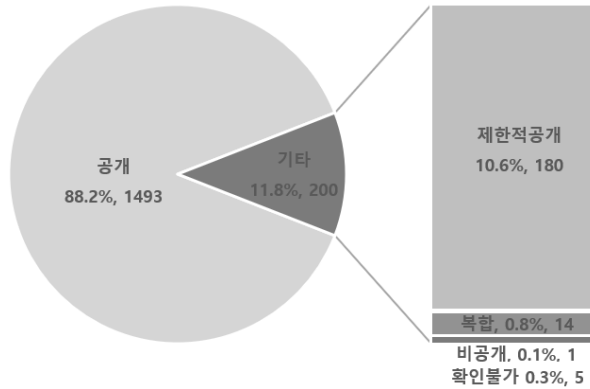
DAS 분석의 시작점으로, 분석의 대상이 된 전체 논문 수 중 DAS가 없는 논문의 수를 파악하였다(〈표 2〉 참조). PLOS ONE에서 DAS 제도를 도입한 2014년은 과도기였으므로 DAS가 작성되지 않은 논문도 소수(5건) 있으나, 이후 2015년 1건, 2021년 1건을 제외한 모든 논문에 DAS가 존재하는 것을 확인하였다. 분석의 대상이 된 1,700편의 논문 중 99.6%(1,693편)의 논문에 DAS가 명시되어 있으므로 현재 PLOS ONE에 DAS 제도가 정착된 것으로 간주할 수 있다.

4.1.2 DAS 존재 논문 공개 현황 분석

〈그림 1〉은 분석 대상 논문의 DAS에 명시된 정보를 Federer et al.(2018)의 데이터 공유 메커니즘에 따라 분류하고 이를 기반으로 연구 데이터 공개 현황을 유형화한 것이다(공개, 제한적 공개, 비공개, 복합, 확인불가). 'in paper'와 'in SI', 'in paper and SI' 메커니즘에 해당하는 경우 데이터에 대한 접근 제한이 없다는 측면에서 '공개'로 구분하였으며, 'upon request'는 요청이라는 접근 제한이 존재하므로 '제한적 공개'로 구분하였다. 상이한 공개 구분의 데이터 공유 메커니즘이 DAS에 함께 명시된 경우도 소수 존재하였으며, 이는 '복합'으로 구분하였다. 'repository' 메커니즘에 해당하는 DAS는 개별적인 확인 작업을 거쳤으며, 최종적으로 데이터에 제한 없이 접근이 가능한 경우 '공개'로, 구매나 회원가입 등의 제한이 있는 경우 '제한적 공개'로 분류하였다. DAS에는 데이터에 접근이 가능하다고 명시했음에도 불구하고 구체적인 접근 방법이 제시되지 않았거나, 접근점을 통해 실질적인 데이터 접근 가능 여부가 확인되지 않는 소수 사례에 대해서는 '확인불가'

〈표 2〉 DAS 부존재 논문 수 및 존재비율

연도	DAS 부존재	전체 논문	DAS 부존재비율	DAS 존재비율
2014	5	81	6.2	93.8
2015	1	261	0.4	99.6
2016	0	233	0.0	100.0
2017	0	262	0.0	100.0
2018	0	224	0.0	100.0
2019	0	183	0.0	100.0
2020	0	182	0.0	100.0
2021	1	180	0.6	99.4
2022	0	94	0.0	100.0
전체	7	1,700	0.4	99.6

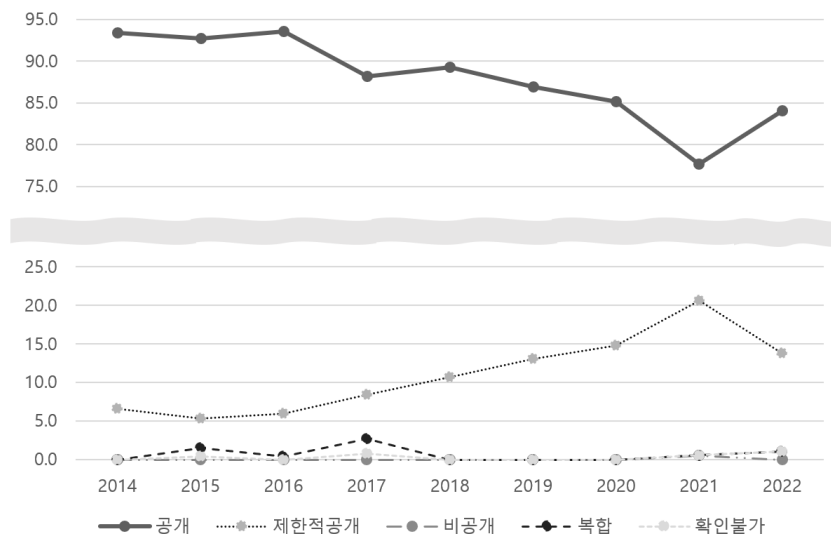


〈그림 1〉 DAS 존재 논문의 데이터 공개 현황

로 분류하였다. 그 결과 표본의 88.2%(1,493건)에 해당하는 DAS가 공개에 해당되었으며, 제한적 공개가 10.6%(180건), 접근 및 확인이 아예 불가능한 경우가 0.3%(5건) 존재했다. 이는 데이터의 공유와 활용을 보장하려는 DAS의 취지에 부합하는 결과이다. 다만 'in paper'와 'in SI' 유형을 본문에 대한 확인 없이 '공개' 유

형에 포함했다는 측면에서, 연구자가 실질적으로 논문에 활용할 만큼의 데이터 공개가 이루어지고 있는가에 대한 질적인 고려는 부족하다는 한계가 있다.

〈그림 2〉는 DAS상에 명시된 연구데이터 공개 유형의 시계열적 비율 변화 추이를 분석한 것이다. '공개'에 해당하는 논문의 비율은 점차



〈그림 2〉 연도별 논문 공개 유형 비율 변화 추이

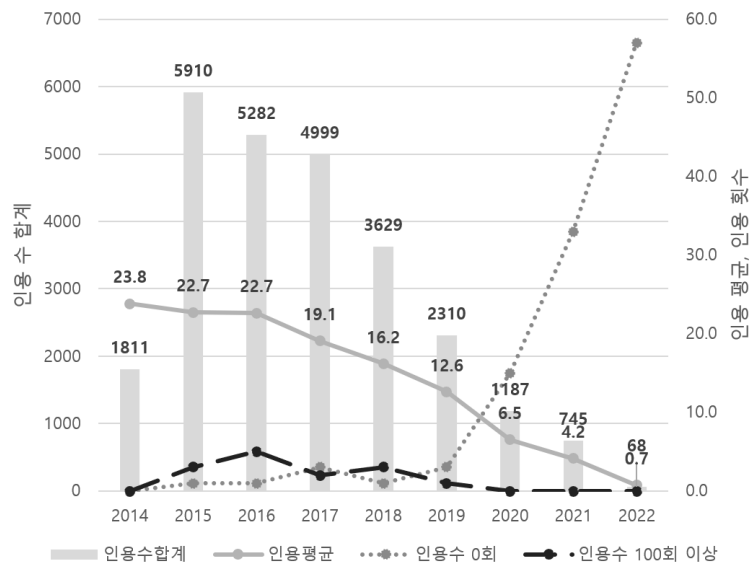
감소하는 반면 '제한적 공개'에 해당하는 논문의 비율은 증가하는 추세를 보이고 있다. 사실상 데이터의 공개로 간주할 수 없는 '제한적 공개'(Jiao, Li, & Fang, 2022)가 증가하고 있는 것은 연구데이터의 공유 및 활용의 저해를 야기할 우려가 있는 현상이다. 따라서 연구자가 연구데이터를 '제한적 공개'로 설정할 수밖에 없는 상황에 대해 파악하고 이를 개선하기 위한 방향을 모색할 필요가 있다.

4.1.3 DAS 작성 논문의 인용 수 분석

DAS가 제시된 논문의 인용 추이 및 특성을 살펴보고자, 분석 대상 데이터에서 개별 논문을 단위로 하여 논문의 인용 수를 분석하였다(〈그림 3〉 참조). DAS가 없는 논문 7편을 제외한 1,693편의 전체 인용 총합은 25,941회, 인용 평균은 15.3회였다. 인용이 한 번도 되지 않은 논문이 전체의 6.7%(114편)으로 가장 많았고, 인

용 수 1회(6.3%, 107편), 인용 수 2회(5.6%, 94편)가 뒤를 이었다. 인용 수가 0회 이상, 10회 미만에 해당하는 논문이 전체의 49.5%(842편), 10회 이상 20회 미만에 해당하는 논문은 25.9%(424편)로 75%가량의 논문이 인용 수 20회 미만이라는 점을 확인할 수 있었다. 인용수가 100회 이상인 논문은 전체의 0.8%(14편)이며, 가장 높은 인용 수를 기록한 것은 2016년에 출간된 논문으로 373회 인용되었다.

인용횟수가 0회인 논문은 간행일이 비교적 최근인 2020년 이후로 크게 증가하였고, 연도별 인용 평균과 인용 수 합계 또한 시간의 흐름에 따라 지속적으로 감소하였다. 이는 최근에 발행된 논문의 경우 충분한 인용수가 축적될 만큼의 시간이 경과하지 않았기 때문에 발생하는, 일반적인 논문의 인용 추이라고 할 수 있다. 다만 100회 이상으로 인용이 많은 논문은 시간의 영향을 상대적으로 적게 받는 것으로 드러



〈그림 3〉 연도별 인용 변화 추이

나며, 이는 출간 이후 시간이 많이 흐를수록 논문에 대한 인용이 무조건적으로 높아지지는 않는다는 것을 의미한다.

4.2 DAS의 데이터 공유 메커니즘 유형 분석

4.2.1 선행연구와 주요 데이터 공유 메커니즘 유형 비율 비교

본 연구에서는 DAS에 명시된 데이터 공유 메커니즘의 유형을 분류하고자 Federer et al. (2018)과 Jiao, Li, Fang(2022)의 선행연구를 참고하였으며, 본 연구에서도 선행연구의 결과와 동일하게 <표 3>에 명시된 6가지가 주요 데이터 공유 메커니즘으로 확인되었다. 본 연구의 결과를 선행연구(Federer et al., 2018; Jiao, Li, & Fang, 2022)와 비교하기 위하여 선행연구에서의 분류와 마찬가지로 두 가지 이상 유형이 적용된 'combination'을 1회로 분류하고 다른 유형에는 포함시키지 않았다.

Federer et al.(2018), Jiao, Li, Fang(2022)과 비교했을 때 본 연구의 표본이 크지 않음에도 불구하고 공유 메커니즘 유형별 비율은 두 가지 선행연구와 큰 차이를 보이지 않았다. 본

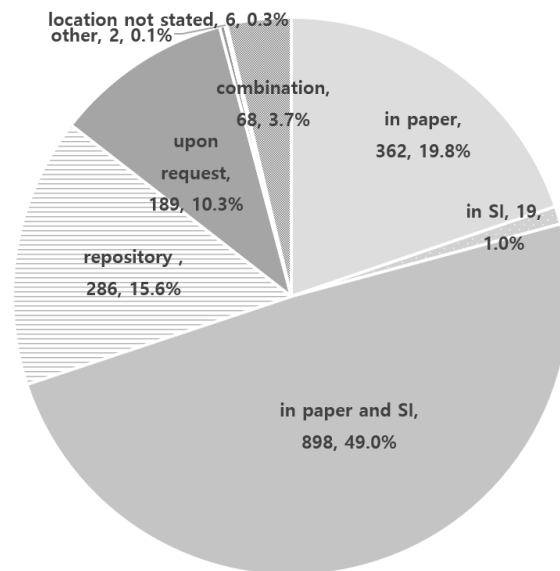
연구와 선행연구의 공유 메커니즘 유형별 비율 차이는 대체로 5% 이내이며, 'combination' 유형의 경우 세 개의 연구가 모두 1% 이내의 차이를 보였다. 가장 큰 차이를 드러낸 것은 'upon request' 유형으로, Federer et al.(2018)에서는 1.4%, Jiao, Li, Fang(2022)은 5.1%인 것에 비해 본 연구에서는 9.6%로 비교적 높게 나타났다. 본 연구의 메커니즘 유형별 분류 결과 약 99.8%의 논문이 여섯 가지 주요 공유 메커니즘에 해당한다는 것 또한 선행연구와는 구분되는 특성이다. 주요 공유 메커니즘에 해당하지 않는 DAS는 데이터의 존재는 인정하되 위치를 명시하지 않는 경우('location not stated', 2건, 0.12%)와 연구와 관련하여 생산된 데이터셋이 없다고 명시한 경우('other', 1건, 0.06%)에 해당된다.

4.2.2 데이터 공유 메커니즘별 비율

<그림 4>는 본 연구에서 분석 대상이 된 데이터의 전체적인 데이터 공유 메커니즘 분포를 확인한 것이다. <그림 4>에서는 <표 3>에서와 달리 두 가지 이상의 공유 메커니즘에 해당하는 'combination'에 대해서도 각 메커니즘 유형에 개수를 포함함으로써 전체적인 언급 양상을 확인하고자 하였다. 가장 큰 비중을 차지한 공

<표 3> 주요 데이터 공유 메커니즘 유형 비율 비교

데이터 공유 메커니즘	Federer et al.(2018)	Jiao, Li, Fang(2022)	본 연구(비율)	본 연구(편수)
in paper and SI	45.3	48.2	51.3	868
in paper repository	24.3	17.2	20.4	346
combination	15.4	13.6	13.8	233
upon request	4.5	4.7	4.0	68
in SI	1.4	5.1	9.6	163
Total	1.4	2.1	0.7	12
	92.3	90.8	99.8	1,690



〈그림 4〉 데이터 공유 메커니즘별 비율

유 메커니즘은 'in paper and SI'이며, 전체의 49.0%(898건)에 해당하였다. 'in paper(19.8%, 362건)'와 'repository(15.6%, 286건)', 'upon request(10.3%, 189건)'가 뒤를 이었으며, 그 외의 공유 메커니즘은 전체의 5% 이하로 매우 소수였다. 추가적으로, DAS에 'in SI'로 명시하고 실제로 SI가 존재하지 않는 경우는 없었으나, 'in paper and SI'로 명시된 898건 중 20건은 논문상 SI가 존재하지 않는 오류도 발견되었다.

〈표 4〉는 'combination'으로 분류된 DAS에 어떤 데이터 공유 메커니즘이 복합되어 있는지를 분석한 것이다. 'in paper and SI'와 'repository'가 복합되어 있는 경우가 24건으로 가장 많았으며, 'repository'와 'upon request'의 복합이 13건, 'in paper'와 'repository'의 복합이 10건이었다. 'in paper and SI'와 'repository', 'upon request' 3개 유형이 복합된 경우도 1건 존재하

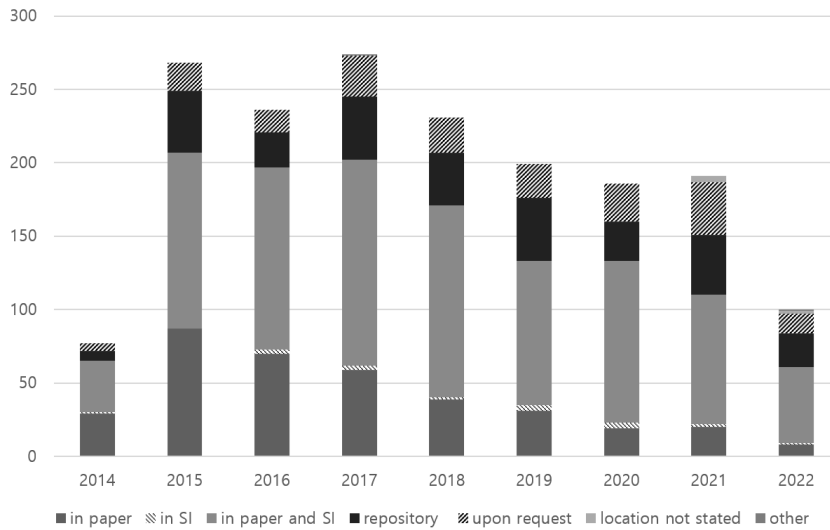
였다. 'combination'에 포함된 데이터 메커니즘 유형 중 가장 빈번한 것은 'repository'로 53번이며, 'in paper and SI'는 30번, 'upon request'는 26번 복합되어 있었다.

4.2.3 전체 공유 메커니즘 비율

〈그림 5〉는 연도별 공유 메커니즘 변화 양상을 분석한 것이다. 데이터 공유 메커니즘 언급의 전체 횟수는 처음 DAS 제도가 도입된 2014년에는 그 수가 많지 않으나, 2015년에 크게 증가하였다. 2017년 가장 많은 언급 이후 점차 감소하는 추세를 보이는 것은 분석 대상 논문 수 자체가 감소한 것의 영향이라고 할 수 있다. 전체 비율 분석에서의 경향과 마찬가지로 매년 'in paper and SI' 유형이 가장 많은 비중을 차지했으며 꾸준히 일정 정도가 유지되고 있다. 'in paper'는 전체 비율 분석에서는 두 번째로 많은 비중을 차지하고 있었으나 시간이 흐를수록 감

〈표 4〉 ‘combination’으로 분류된 DAS의 데이터 공유 메커니즘 구성

데이터 공유 메커니즘 복합	편수	비율
in paper and SI + repository	24	35.3
in paper and SI + upon request	3	4.4
in paper and SI + other	1	1.5
in paper and SI + location not stated	1	1.5
in paper + repository	10	14.7
in paper + upon request	5	7.4
in paper + location not stated	1	1.5
in SI + repository	5	7.4
in SI + upon request	2	2.9
repository + upon request	13	19.1
upon request + location not stated	2	2.9
in paper and SI + repository + upon request	1	1.5



〈그림 5〉 연도별 데이터 공유 메커니즘 언급 변화 추이

소하는 추세이며, ‘repository’의 비중은 상대적으로 고르게 분포하고 있었다. 실질적으로 연구데이터가 공개되는 방식이라고 간주하기 어려운 ‘upon request’의 경우 높은 비중을 보이지는 않으나 매년 일정 부분을 차지하고 있었다. DAS가 ‘upon request’ 유형으로 작성된 논문은 연구자가 제3자의 데이터를 활용하였거

나, 연구 과정에서 생산 혹은 활용된 데이터가 개인정보 및 민감정보에 해당하는 사례가 대부분이었다. 따라서 제3자의 데이터가 활용된 연구의 DAS 작성을 위하여 별도의 접근점 제시 표준을 마련하고, 개인정보 및 민감정보의 비식별화 기술 및 관련 제도를 보완함으로써 연구데이터의 실질적인 공유를 확대시킬 수 있는

방안을 모색할 필요가 있다.

4.2.4 데이터 공유 메커니즘별 인용수 분석

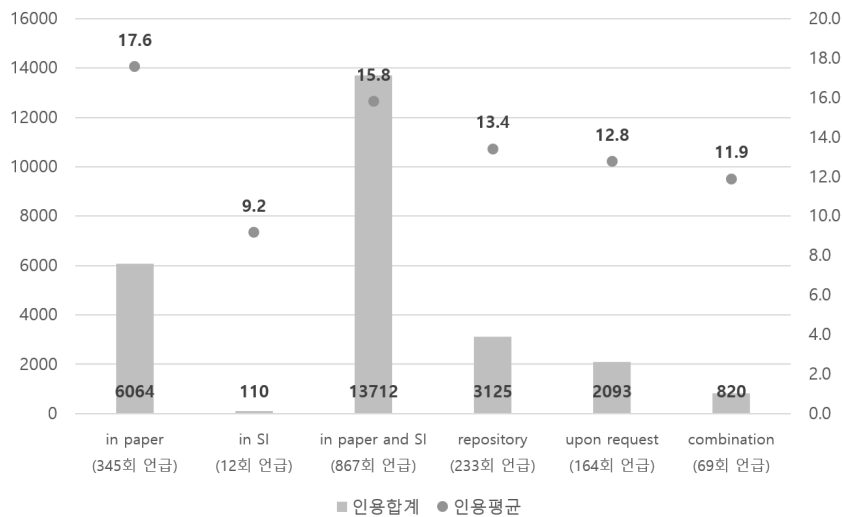
〈그림 6〉은 데이터 공유 메커니즘 유형별 인용 수 합계와 평균을 분석한 것이다. 가장 높은 비율을 차지하는 'in paper and SI' 유형이 가장 높은 인용 합계를 드러냈으며, 각 유형에 해당하는 논문의 수가 많을수록 인용 합계 또한 높게 나타났다. 인용 평균이 가장 높은 메커니즘 유형은 'in paper'(17.6회)로, 전체 평균(15.3회)을 크게 상회했으며, 'in paper and SI' 유형은 15.8의 인용평균으로 전체 평균보다 조금 높게 나타났다. 그 외 'repository', 'upon request', 'combination' 유형은 각각 13.4회, 12.8회, 11.9회로 전체 평균보다 낮았다. 'in SI' 유형의 경우 인용 합계 110회, 인용 수 평균 9.2회로 매우 낮게 나타났는데, 'in SI' 유형에 해당되는 표본이 소수이므로 보다 큰 표본에서 재확인할 필요가 있다.

4.3 DAS에 언급된 리포지터리 분석

4.3.1 'repository' 세부 구성 비율

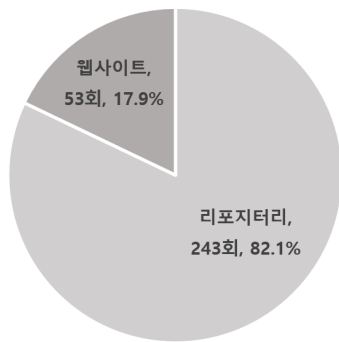
본 연구에서 분석의 기준으로 활용한 Federer et al.(2018)에 따르면, 데이터에 공적으로 접근할 수 있는 경우 모두 'repository' 유형으로 분류하고 있으며, 이에는 데이터 리포지터리뿐만 아니라 상용 웹사이트도 포함된다. 본 연구에서는 연구자들이 주로 사용하는 리포지터리에 대한 세부적인 분석을 진행하고자 하므로 연구 데이터의 검색과 공유를 지원하는 리포지터리(이하 리포지터리)와 상용 웹사이트(이하 웹사이트)를 구분하였다. DAS에 리포지터리와 웹사이트가 모두 명시된 논문(9건)도 존재하였으나, 대부분 둘 중 한 종류에 대해서만 언급하고 있었으며 하나의 DAS에서 복수의 리포지터리나 웹사이트가 명시된 경우가 다수 존재하였다.

하나의 DAS에 웹사이트와 리포지터리가 모



〈그림 6〉 데이터 공유 메커니즘 유형별 인용 분석

두 언급된 경우, 복수의 리포지터리 혹은 웹사이트 언급된 경우를 모두 고려하여 'repository' 유형 구분 내 웹사이트와 리포지터리의 언급 비율을 파악한 결과는 <그림 7>과 같다. 'repository' 유형으로 구분된 DAS 중 82.1%가 리포지터리에 해당했으며, 웹사이트는 전체의 20% 미만으로 적은 비율만 차지했다. 웹사이트의 경우 연구 과정에서 분석의 대상이 되거나 데이터 수집의 대상이 된 사이트의 URL이 명시된 사례가 대부분이었다.

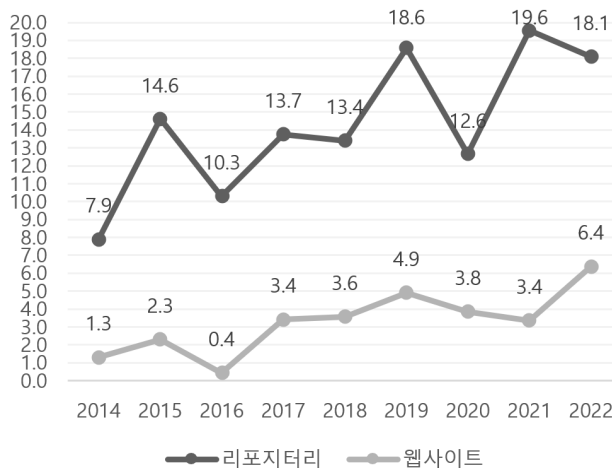


<그림 7> 'repository' 유형 구성

<그림 8>은 전체 DAS 중 리포지터리와 웹사이트 언급된 비율의 변화 추이를 시계열적으로 분석한 것이다. DAS가 도입된 초기에는 리포지터리와 웹사이트 두 종류 모두 전체 논문에서 차지하는 비율이 높지 않았으나, 시간의 흐름에 따라 등락하며 점차 증가하는 추세를 확인할 수 있다. 이는 연구자들이 점차 적극적으로 데이터를 공유하고자 하고 있음을 의미한다고 할 수 있다.

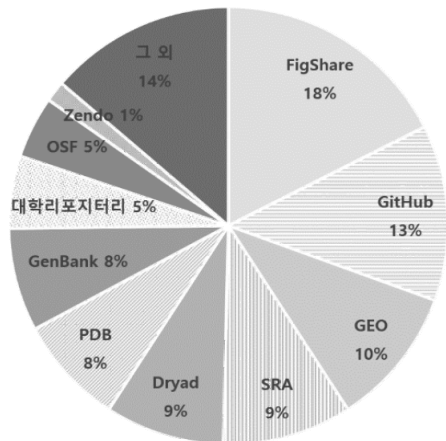
4.3.2 상위 10개 리포지터리

<그림 9>는 DAS에 언급된 리포지터리의 빈도를 분석하여 상위 10개 리포지터리를 추려낸 것이다. 학문분야의 구분 없이 도표, 이미지, 시뮬레이션 영상 등 다양한 유형의 연구데이터를 포괄할 수 있는 FigShare가 가장 많이 언급되었으며(17.56%, 46회), 이를 통해 PLOS 정책의 FigShare 데이터 업로드 명시에 따른 영향의 가능성을 추정해 볼 수 있다. 두 번째로 높은 비율을 차지하는 GitHub는 연구데이터로서 코드가



<그림 8> 연도별 리포지터리 및 웹사이트 언급 비율 변화 추이

존재하는 대부분의 경우에 이용되었다. FigShare와 GitHub 외 리포지터리 중 GEO(10.31%, 27회), SRA(9.54%, 25회), PDB(8.02%, 21회), GenBank(7.63%, 20회)는 모두 NCBI에서 제공하는 생물공학분야 데이터베이스에 해당하며, Dryad(8.78%, 23회)와 OSF(4.58%, 12회), Zendo(1.53%, 4회)는 학문분야 구분 없이 이용 가능한 범용 리포지터리이다. 전체 리포지터리 중 범용 리포지터리의 비율이 더 높기는 하나, 생물공학분야 리포지터리가 다수를 차지하는 것은 PLOS ONE 학술지 자체의 경향으로부터 영향을 받은 것으로 추정된다.



〈그림 9〉 DAS 언급 빈도 상위 10개 리포지터리

리포지터리에 대한 분석을 통해 연구 분야에 기반을 둔 리포지터리 사용과 연구데이터의 자료 유형에 기반을 둔 리포지터리 사용 양상이 복합적으로 드러나고 있음을 확인하였다. 특히 GitHub의 경우 연구데이터 게시를 위해 타 리포지터리를 사용하고 있는 상황일지라도 코드가 존재한다면 GitHub를 추가적으로 이

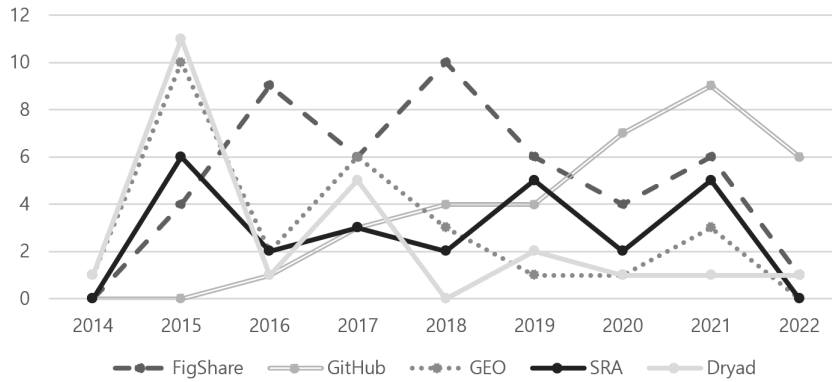
용하는 경향이 드러났으며, 이를 통해 GitHub가 코드를 관리하는 리포지터리로서 갖는 영향력이 상당하다는 점을 파악할 수 있었다. 마지막으로, Harvard Diverse나 Snap.Stanford 등 대학에서 제공하고 관리하는 리포지터리를 사용하는 경우도 존재하였으며, 이는 소속 기관에서 제공하는 연구데이터 관리 서비스의 영향을 받은 리포지터리 사용 양상이라고 할 수 있다. 국내 기관 소속 연구자가 국가 연구비를 지원받아 생산한 데이터를 기탁하는 것임에도 불구하고 언급빈도 상위 10개 리포지터리 중 9개가 해외의 리포지터리였으며, '대학 리포지터리' 중에서도 국내 대학 리포지터리가 언급된 사례는 한 건 뿐이었다.

4.3.3 상위 5개 리포지터리 언급 횟수 시계열적 변화 분석

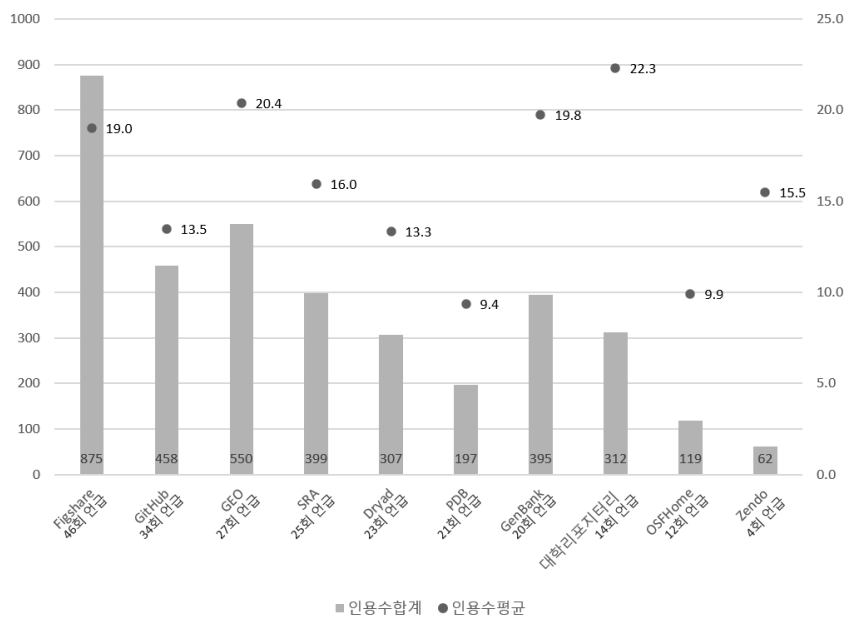
〈그림 10〉은 상위 5개 리포지터리의 시계열 등락 추이를 분석한 것이다. PDB, GenBank, 대학리포지터리, OSF, Zendo는 상위 10개 리포지터리에 해당하지만 표본의 크기가 작아 연간 언급 빈도가 지나치게 소수이므로 연도별 변화 추이 분석에는 적합하지 않다고 판단하여 제외하였다. DAS에서 가장 많이 언급된 FigShare는 2016년과 2018년에 빈번하게 사용된 후 다소 하락세에 있으나 꾸준히 사용되는 추세인 반면, GEO와 Dryad, SRA는 2015년도에 많이 사용된 후 등락을 거치며 점차 줄어들고 있다. GitHub의 경우 2015년에 처음 사용된 후 꾸준히 증가하는 추세를 드러내고 있다.

4.3.4 상위 10개 리포지터리별 인용수 분석

〈그림 11〉은 DAS에서 많이 언급된 상위 10



〈그림 10〉 상위 5개 리포지터리 연도별 언급 변화 추이



〈그림 11〉 상위 10개 리포지터리별 인용 분석

개 리포지터리를 대상으로 각 리포지터리의 인용수 합계와 인용수 평균을 분석한 것이다. 한 논문의 DAS에 복수의 리포지터리가 제시된 경우 인용수와 관계가 있는 리포지터리가 무엇인지 확인하기 어려우므로, 리포지터리를 언급한 243개 DAS중 복수의 리포지터리가 제시된

15개의 DAS는 분석에서 제외하였다.

데이터 공유 메커니즘은 언급 횟수가 많을수록 인용 수 합계도 높아지는 추세를 보였으나, 리포지터리의 경우 언급 횟수와 인용 수 합계가 비례하지 않았다. 언급이 가장 많은 FigShare는 인용 수 합계도 805회로 가장 높았으나, GEO

(528회 언급)의 인용 수 합계가 GitHub(404회 언급)의 인용 수 합계보다 높았고, GenBank(287회 언급)와 대학리포지터리(312회 언급)의 인용 수 합계는 전체 언급 수에 비해 매우 높게 나타났다.

절반 이상의 리포지터리가 전체 데이터 인용 평균(15.3회)보다 높은 인용 평균을 드러냈으며, 특히 대학리포지터리(22.3회)나 GEO(22.0회)의 경우 평균 인용 수가 20회 이상이었다. 그러나 GitHub과 Dryad, PDB, OSF는 전체 평균 인용 수보다 낮은 인용 수가 나타났다. GitHub은 14.4회로 전체 평균에 약간 미치지 못하는 수준이었으나, PDB와 OSF는 10회 미만으로 전체 평균보다 굉장히 동떨어지는 인용 수 평균을 나타냈다. 다만 대학리포지터리나 OSF, Zendo의 경우 표본이 소수라는 점이 고려될 필요가 있다.

4.4 DAS에 명시된 접근점 분석

데이터 공유 메커니즘으로서 리포지터리 혹은 웹사이트를 언급한 경우, 단순히 리포지터리

의 명칭에 대한 명시에서만 그치지 않고 리포지터리(웹사이트) 내에서 어떻게 해당 연구의 데이터로 접근할 수 있는지 접근점이 제시된다. 접근점은 데이터의 실질적인 접근과 직접적인 관련이 있는 요소이므로, 본 연구에서는 분석 대상 논문의 DAS에서 접근점이 제시되는 양상과 사용되는 접근점의 종류, 그리고 실제 접근 가능 건수에 대해 분석을 진행하였다.

4.4.1 연도별 리포지터리, 웹사이트 접근가능 비율 시계열 분석

〈표 5〉는 연구데이터 리포지터리와 웹사이트의 접근점 제시 건수와 실제로 확인한 접근가능 건수, 그리고 접근가능 비율에 대한 분석 결과이다. 리포지터리의 경우 DAS에 제시된 접근점의 실제 접근 가능 비율이 점차 증가하는 추세를 보이는 반면, 웹사이트 접근점(URL)의 경우 등락폭이 크고 불규칙하였다. 이를 통해 리포지터리가 아닌 일반 웹사이트를 데이터의 접근 방법으로서 제시했을 때 향후 데이터 접근의 안정성이 떨어질 수 있다는 점을 유추할 수 있었다.

〈표 5〉 리포지터리, 웹사이트 접근가능 비율

연도	리포지터리			웹사이트		
	접근점 제시 건수	접근가능 건수	접근가능 비율	접근점 제시 건수	접근가능 건수	접근가능 비율
2014년	15	14	93.3	2	2	100.0
2015년	84	80	95.2	12	8	66.7
2016년	40	39	97.5	2	0	0.0
2017년	102	99	97.1	13	8	61.5
2018년	45	44	97.8	16	15	93.8
2019년	261	259	99.2	15	9	60.0
2020년	358	358	100.0	10	10	100.0
2021년	74	73	98.6	14	10	71.4
2022년	72	72	100.0	6	5	83.3

4.4.2 연도별 리포지터리 접근점 시계열 분석

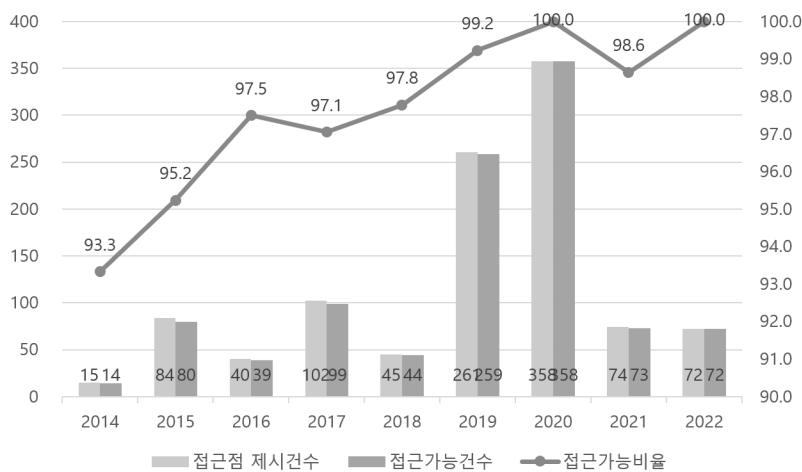
〈그림 12〉는 리포지터리와 함께 제시된 접근점 언급 건수와 실제 접근 가능 건수를 연도별로 제시하고 있으며, 이를 기반으로 접근 가능 비율을 확인하였다. 접근점 제시 건수는 2014년 이후 등락을 반복하며 2020년까지 지속적으로 증가하다가, 2021년에 크게 감소하였다. 2022년의 경우 표본 자체가 소수이므로 접근점 제시 건수도 많지 않으나, 2021년의 경우 전체 논문 수 대비 접근점 언급 수가 크게 감소했다. 표본에서 DAS를 통해 제시된 접근점 건수는 1,051건이며 이 중 1,038건이 접근 가능한 것으로 파악되어 전체 접근점 제시 건수 대비 실제 접근가능 비율은 98.8%이다. 접근 가능 비율은 DAS가 처음 도입된 2014년에도 93.3%로 대부분의 접근점이 접근 가능한 것으로 확인되었으며, 헛수가 거듭됨에 따라 대체적으로 증가하는 추세를 보였다. 2020년에는 접근점 제시 건수가 가장 많음에도 불구하고 모든 접근점이 유효한 것으로 파악되며 100%의 접근가능 비

율을 기록했다.

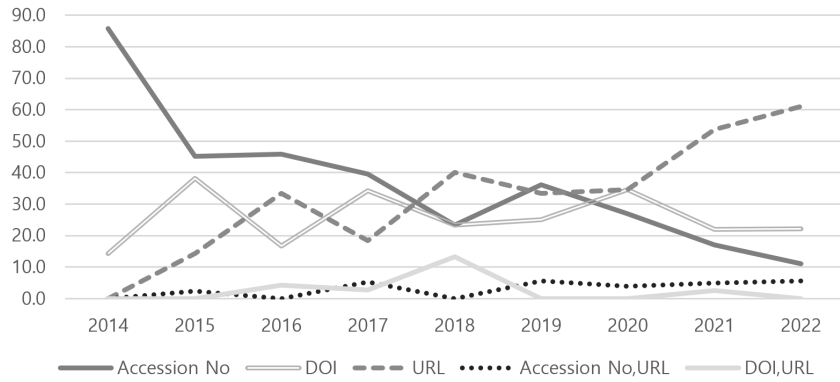
접근가능비율의 증가 추세는 DAS 제도가 정착되며 연구자가 데이터에 접근할 수 있는 구체적이고 정확한 정보를 제시하고 있다는 긍정적인 측면으로 볼 수 있는 한편, 접근점이 제시된 후 시일이 지남에 따라 접근점이 유효하지 않게 되는 문제가 발생하고 있다고도 해석할 수 있다. 따라서 DAS에 영구적으로 이용 가능한 접근점이 명시될 수 있도록 제도를 마련해야 하며, 영구적인 접근점의 개발 및 활성화가 필요하다.

4.4.3 시계열 변화에 따른 접근점 유형별 언급 비율 변화 추이

DAS에 제시된 접근점 유형은 크게 Accession No, DOI, URL로 구분할 수 있으며, 하나의 리포지터리에서 두 가지 유형의 접근점이 복합적으로 사용되는 경우(Accession No+URL, DOI+URL)도 있었다. 〈그림 13〉은 각 연도별로 언급된 전체 접근점 수 대비 각 유형의 접근점 제



〈그림 12〉 연도별 접근가능 비율 변화 추이



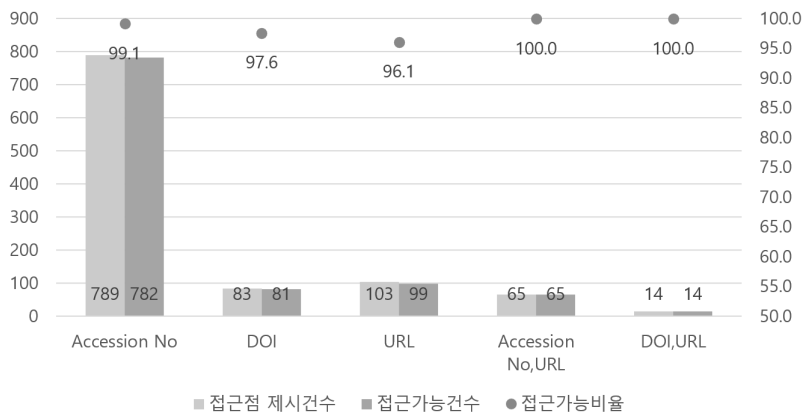
〈그림 13〉 접근점 유형별 언급 비율 변화 추이

시 건수를 기반으로 산출한, 시계열 변화에 따른 접근점 유형별 언급 비율 변화 추이이다. 두 가지 이상의 접근점이 복합적으로 사용되는 경우는 그 수가 많지 않으며 따라서 등락도 크지 않았다. Accession No의 경우 2014년 85.7%부터 꾸준히 언급 비율이 줄어드는 반면 URL은 언급이 많아지고 있는 추세인데, 이는 접근점으로 URL을 사용하는 리포지터리인 GitHub의 사용이 지속적으로 증가하고 있는 것의 영향으로 볼 수 있다. DOI의 경우 등락은 있으나 꾸준히 일

정 수준 이상의 언급 비율을 유지하고 있었다.

4.4.4 접근점 유형별 접근가능 비율 분석

〈그림 14〉는 접근점 유형별 제시 건수와 실제 접근가능 건수, 이에 따른 접근가능 비율을 조사한 결과이다. 전체 분석 대상 데이터에서 가장 많이 제시된 접근점 유형은 Accession No (789건)이며, URL(103건)과 DOI(83건)의 제시 건수를 크게 상회했다. 하나의 리포지터리에 여러 개의 연구데이터가 게시되면서 개별 데이



〈그림 14〉 접근점 유형별 접근가능 비율

터 단위마다 Accession No가 부여되는 경우가 다수 발견되었고, 이러한 접근점 부여 방식의 영향으로 Accession No의 제시 건수가 가장 높게 나타났다. Accession No와 DOI, URL 세 가지 유형의 접근 가능 비율은 각각 99.1%, 97.6%, 96.1%로 높게 나타났으며, 하나의 리포지터리에서 두 가지 유형의 접근점을 복합적으로 사용하는 경우, 복합적으로 사용된 접근점의 유형과 무관하게 제시된 모든 접근점이 유효했다.

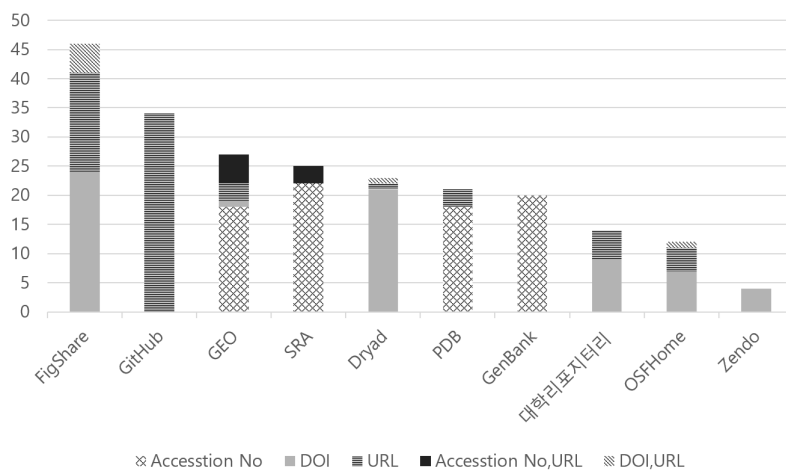
4.4.5 리포지터리별 접근점 이용 분석

〈그림 15〉는 리포지터리별 접근점 이용 양상을 분석한 것이다. GEO와 같이 세 가지 유형의 접근점이 모두 사용되는 경우도 있는 반면, 대부분 리포지터리별로 이용하는 접근점의 특징이 분명하게 드러났다. FigShare와 Dryad, OSF 경우 DOI와 URL이 개별적 혹은 복합적으로 사용되고 Accession No는 전혀 언급되지 않은 반면, PDB와 SRA의 경우 Accession No와 URL만을 사용하였다. GitHub는 URL만,

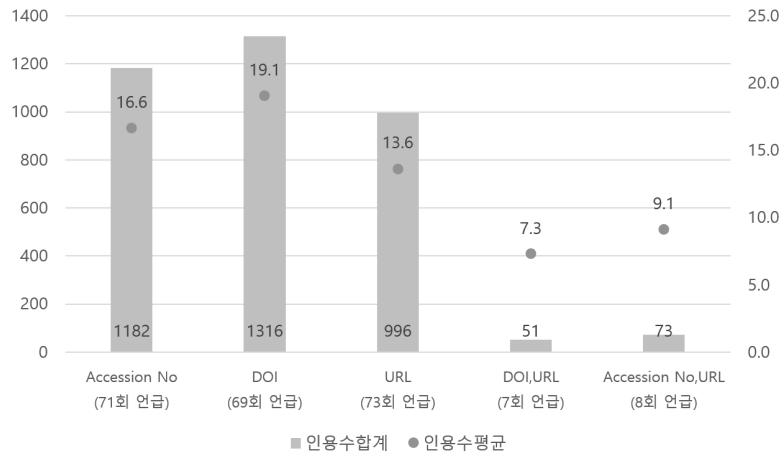
GenBank는 Accession No만, Zendo는 DOI만을 사용하는 등 리포지터리에서 제공하는 서비스와 리포지터리의 데이터베이스 구성 특성에 따라서 사용되는 접근점이 상이하였다.

4.4.6 접근점 유형별 인용 수 분석

〈그림 16〉은 DAS에 제시된 접근점의 종류에 따른 인용수의 차이를 분석한 것이다. 접근점별 인용수는 리포지터리별 인용수 분석의 경우와 마찬가지로 해당하는 논문 수와 인용수 합계가 비례하지 않았다. 인용 수 합계가 가장 높은 접근점 유형은 DOI(1,316회)이며, Accession No(1,182회)와 URL(996회)가 뒤를 이었다. DOI와 Accession No의 경우 전체 데이터 인용 수 평균(15.3회)을 상회하는 인용수 평균이 나타났으며, 특히 DOI는 19.1회로 인용수 평균이 매우 높았다. 반면 URL의 경우 다른 접근점 유형과 달리 13.6회로 전체 평균보다 낮았다.



〈그림 15〉 리포지터리별 접근점 이용



〈그림 16〉 접근점 유형별 인용 분석

5. 결론 및 제언

본 연구는 국내 연구자의 논문에 수록된 DAS에서 확인되는 데이터 공유 메커니즘과 데이터 리포지터리를 분석하고, 이를 통해 국내 연구자의 데이터 공유 현황과 특징을 살펴보는 것을 목적으로 수행되었다. 2014년 3월부터 DAS 제도를 도입한 PLOS에 주목하여 제도 시행 이후부터 2022년 12월 31일까지 PLOS ONE 학술지에 게재된 논문 중, 국내 기관 소속 연구자가 저술한 논문을 분석의 대상으로 선정하였다.

우선 분석 대상 데이터에 대한 일반적인 조사를 진행하였으며, 이를 통해 PLOS ONE에서 DAS 제도를 도입한 초기 시점의 일부 출판 논문을 제외하고 대부분의 논문에서 DAS가 작성되었음을 파악하고 각 논문의 데이터 공개 여부와 인용 수 특성을 확인하였다. 이후 DAS에 명시된 데이터 공유 메커니즘의 유형 분류를 진행하였다. 이를 통해 데이터 공유 메커니즘 중 선호되는 유형은 'in paper and SI'와 'in

paper', 'repository'순이라는 점을 확인했으며, 선행연구와의 비교를 통해 우리나라의 데이터 공유 메커니즘 작성 양상이 국제적인 양상과 유사하다는 사실이 드러났다. 데이터 공유 메커니즘의 선호 양상은 시간의 흐름에 따라 변화하고 있었으며, 'in paper' 유형에 대한 언급은 점차 감소하는 추세인 반면 'repository'와 'upon request' 유형은 항상 일정 수준이 유지되는 현상이 관찰되었다. 또한 데이터 공유 메커니즘별 인용 수 분석 결과, 가장 권장되는 공유 방법인 'repository'의 인용 수 평균보다 'in paper and SI'와 'in paper' 유형의 인용 수 평균이 높게 나타났다.

DAS의 데이터 공유 메커니즘 중에서도 가장 이상적인 공유 방식이라고 할 수 있는 'repository' 유형에 대한 구체적인 분석을 진행하였다. 'repository' 유형에는 데이터 리포지터리와 일반 웹사이트가 모두 포함되지만 80% 이상이 데이터 리포지터리에 해당되므로 데이터 리포지터리에 분석의 초점을 맞추었다. DAS상

에 많이 언급된 상위 10개 리포지터리를 조사하였으며, 상위 10개 리포지터리가 전체 리포지터리 언급의 88%를, 상위 5개 리포지터리가 전체의 60%를 차지하는 양상을 확인하였다. 상위 리포지터리의 시계열적 변화 추이 분석 결과, PLOS 정책상 명시된 FigShare는 매년 일정 비율의 언급이 유지되고 있었으며, 코드라는 특정 유형의 연구데이터를 다루는 GitHub의 경우 점차 증가하는 추세를 보였다. 리포지터리는 DAS 언급 수와 인용 수, 인용 수 평균이 큰 관계를 보이지 않았으며, 언급 수가 적은 리포지터리라고 할지라도 인용 수 평균이 높게 나타나기도 했다.

리포지터리에 대한 분석의 일환으로 데이터의 구체적인 접근점도 조사하였다. 접근점 제시 건수는 점차 증가하는 추세를 보이다가 2021년부터 크게 감소하였으나, 접근점을 이용한 실제 접근 가능 비율은 해가 거듭됨에 따라 대체적으로 증가하였다. 접근점 유형별 세부 분석을 위해 전체 접근점 언급 횟수 대비 각 유형의 언급 횟수 비율을 확인하였으며, 이를 통해 Accession No는 언급 비율이 낮아지는 반면 URL은 증가하는 경향이 있음을 파악하였다. 학술지 리포지터리별 특성에 따라 사용되는 접근점의 유형이 상이하다는 사실이 드러났으며, 제시된 접근점의 실제 접근 가능 비율은 모든 유형에서 95% 이상으로 매우 높았다. 마지막으로, 접근점 유형별 인용 수 분석 결과 Accession No와 DOI의 인용 수 평균은 데이터 전체 평균보다 높게 나타난 반면 URL은 전체 평균에 미치지 못했다.

이상과 같이 본 연구는 PLOS ONE 학술지를 대상으로 국내 기관 소속 연구자 논문의

DAS 게시 현황과 DAS에 기술된 리포지터리를 조사함으로써 학술지 데이터 정책의 영향력과 중요성을 확인하였다. 국내의 데이터 공개·공유 확대 및 발전을 위한 방향 모색의 일환으로서 다음과 같은 제언을 하고자 한다.

국내 기관 소속 연구자의 데이터 공유 메커니즘 선택 경향은 선행연구에서 제시된 해외 연구자의 데이터 공유 메커니즘 분석 결과와 유사한 양상을 드러냈으며, 이러한 현상의 원인으로 학술지의 DAS 정책이 저자의 데이터 공유에 미치는 영향을 추정할 수 있다. 데이터 정책의 영향은 리포지터리의 시계열적 변화 추이를 통해서도 확인되었다. 많은 논문의 DAS에서 언급된 리포지터리일지라도 시간의 흐름에 따라 언급 비율의 등락 폭이 큰 반면 FigShare의 경우 매년 일정 비율의 언급이 유지되고 있는데, 이는 PLOS의 데이터 정책 상 제출된 모든 '보충자료'를 논문 출판 시 FigShare에 업로드하고 있다는 명시와 관련이 있을 수 있다. PLOS 출판 논문의 데이터를 FigShare를 통해 이용한 연구자가 본인의 데이터 기탁 시 동일한 리포지터리를 선택할 가능성도 있을 것이다. 이처럼 연구자가 선택하는 데이터 공유의 메커니즘과 리포지터리는 학술지가 제시하는 정책의 영향을 받으므로, 데이터 공유를 활성화하기 위해서는 체계적이고 실현가능한 정책의 수립이 중요할 것이다.

데이터 공유와 이용 확대를 위한 기술 및 제도에 대해서도 고려해볼 수 있다. 국내 기관 소속 연구자가 선호하는 데이터 공유 메커니즘 유형은 시간의 흐름에 따라 변화하는 양상이었으나, 사실상 데이터 비공개에 해당하는 'upon request' 유형의 언급은 매년 일정 수준 유지되

었으며 'upon request' 유형을 택하는 경우는 대부분 기존에 존재하는 데이터를 활용하거나 개인정보 및 민감정보가 포함된 데이터를 다루는 논문이었다. 따라서 'upon request' 유형의 감소와 실질적인 데이터 공유 확대를 위해서는 제3자 데이터 활용 제도의 마련과 데이터의 비식별화 기술 개발을 위한 노력이 요구된다는 점을 알 수 있다. 연구자의 리포지터리 이용에 대한 분석 결과, 코드리는 특정 유형의 연구데이터를 다루는 리포지터리인 GitHub만이 시간이 흐를수록 언급도 증가하는 추세를 보였으므로 이를 통해 특정 데이터 유형을 다루는 리포지터리에 대한 수요가 증가할 것이라는 점을 예측할 수 있다. 이와 더불어 DAS를 통해 언급되는 리포지터리의 대부분이 국외의 다양한 리포지터리에 해당되는 반면, 국내 리포지터리가 사용되는 경우는 거의 없다는 것은 국내의 데이터 기탁 환경이 미비함을 의미한다. 연구자가 국가에서 연구비를 지원받아 생산된 연구데이터는 국가의 자산이므로, 이상의 주요 사항을 반영하여 국내 리포지터리에서 저장·관리·공유될 수 있도록 지원하는 데이터 거버넌스 체

제의 마련이 필요할 것이다.

본 연구는 연구 대상 범위가 실질적이고 세부적으로 설정됨에 따라 표본의 크기가 작다는 것이 단점이며, 데이터의 성격과 DAS 작성에 큰 영향을 미칠 수 있는 학문분야별 특성에 따른 차이를 살펴보기 못했다는 한계가 있다. 또한 저자의 DAS 작성과 리포지터리 선택, 접근점 이용은 저널과 리포지터리에서 제시하는 정책의 영향을 받기 때문에, PLOS ONE을 대상으로 한 본 연구의 결과는 특정 학술지와 리포지터리의 정책 특성이 반영될 수밖에 없다는 제한점도 존재한다. 그러나 국내에 아직 잘 알려지지 않은 데이터 공유 수단으로서 DAS의 개념을 소개하고, 국내 연구자의 DAS 작성 내용과 리포지터리의 이용에 대해 조사함으로써 우리나라의 DAS 정책 도입과 연구데이터 공유의 발전 방향을 모색하기 위한 단초를 제공했다는 점에서 의의가 있다. 본 연구를 기반으로 향후 학문분야적 특성을 고려하여 DAS를 분석하거나 DAS를 통해 제시된 데이터의 실제 이용 경향에 대해 파악하는 후속 연구가 진행될 수 있을 것으로 기대된다.

참 고 문 헌

- 권나현, 표순희 (2022). 오픈엑세스를 위한 리포지터리 시스템에 대한 수요 분석과 경제적 가치 측정. 한국비블리아학회지, 33(3), 179-203. <https://doi.org/10.14699/kbiblia.2022.33.3.179>
- 김소영, 김지현, 최희석, 황혜경 (2016). 해외 공공연구기금에 의한학술논문의 오픈 액세스 정책 분석. 한국문헌정보학회지, 50(4), 209-229. <https://doi.org/10.4275/KSLIS.2016.50.4.209>
- 김우중, 김형준, 박희진, 최상욱 (2021). 자연과학 분야 연구 데이터 리포지터리 속성분석: 운영수준, 주제, 유형, 활용도를 중심으로. 기술혁신학회지, 24(4), 577-595.

- <https://doi.org/10.35978/jktis.2021.8.24.4.577>
- 김지현 (2016). 연구데이터 레포지터리의 데이터 접근 및 이용 통제 정책 요소에 관한 연구. *한국도서관·정보학회지*, 47(3), 213-239. <https://doi.org/10.16981/kliss.47.3.201609.213>
- 박규리, 안병군 (2017). 연구데이터 국가 리포지터리 구축에 관한 연구. *한국정보관리학회 제24회 학술대회 논문집*, 33-38.
- 신영란, 정연경 (2012). 국내 인문사회 연구데이터 아카이브의 개선방안에 관한 연구. *한국기록관리학회지*, 12(3), 93-115. <https://doi.org/10.14404/JKSARM.2012.12.3.093>
- 신은정, 정원교 (2017). 오픈사이언스정책의 확산과 시사점. *STEPI Insight*, 216, 1-39.
- 심원식 (2021). 오픈액세스 메가 학술지 논문 게재 경험이 있는 국내 기관 소속 저자의 오픈액세스 출판 인식 및 경험에 관한 연구. *한국문헌정보학회지*, 55(4), 45-65.
<https://doi.org/10.4275/KSLIS.2021.55.4.045>
- 심원식, 안병군, 박성은, 김현수 (2020). PLoS ONE 학술지 게재 국내 기관 소속 연구자 논문의 계량적 분석. *정보관리학회지*, 37(2), 47-69. <https://doi.org/10.3743/KOSIM.2020.37.2.047>
- 심원식, 안혜연, 변제연 (2015). 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략 연구. *한국문헌정보학회지*, 49(3), 155-183. <https://doi.org/10.4275/KSLIS.2015.49.3.155>
- 정경희, 이재윤, 정은경, 최상희 (2021). 한국연구재단 오픈액세스 정책 실행방안 연구. *정보관리학회지*, 37(4), 255-286. <https://doi.org/10.3743/KOSIM.2020.37.4.255>
- 정영미, 이상기 (2010). 성공적인 리포지터리의 운영정책에 관한 연구. *정보관리학회지*, 27(4), 131-152. <https://doi.org/10.3743/KOSIM.2010.27.4.131>
- 조재인, 박종도 (2019). re3data를 기반으로 한 인문사회 RDR 연구. *한국비블리아학회지*, 30(2), 69-87. <https://doi.org/10.14699/kbiblia.2019.30.2.069>
- 황혜경, 이지연 (2016). 연구자의 인구사회학적 특성이 기관리포지터리 수용의도에 미치는 영향. *한국정보관리학회 제23회 학술대회 논문집*, 49-55.
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *Plos One*, 15(4). e0230416. <https://doi.org/10.1371/journal.pone.0230416>
- Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., & Clark, T. (2018). A data citation roadmap for scientific publishers. *Scientific Data*, 5, 180259. <https://doi.org/10.1038/sdata.2018.259>
- Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *Plos One*, 13(5), e0194768. <https://doi.org/10.1371/journal.pone.0194768>
- Gómez, N. D., Méndez, E., & Hernández-Pérez, T. (2016). Social sciences and humanities

- research data and metadata: a perspective from thematic data repositories. *El Profesional de la Información*, 25(4), 545-555. <https://doi.org/10.3145/epi.2016.jul.04>
- Gorraiz, J., Kraker, P., Lex, E., Gumpenberger, C., & Peters, I. (2017). Zenodo in the spotlight of traditional and new metrics. *Frontiers in Research Metrics and Analytics*, 2(13), 1-14. <https://doi.org/10.3389/frma.2017.00013>
- Graf, C., Flanagan, D., Wylie, L., & Silver, D. (2019). The open data challenge: an analysis of 124,000 data availability statements and an ironic lesson about data management plans. *Data Intelligence*, 2(4), 554-568. https://doi.org/10.1162/dint_a_00061
- Grant, R. & Hrynaszkiewicz, I. (2018). The impact on authors and editors of introducing data availability statements at Nature journals. *International Journal of Digital Curation*, 13(1), 195-203. <https://doi.org/10.1101/264929>
- Helliwell, J. R., Minor, W., Weiss, M. S., Garman, E. F., Read, R. J., Newman, J., Raaij, M. J., Hajdu, J., & Baker, E. N. (2019). Findable accessible interoperable re-usable (FAIR) diffraction data are coming to protein crystallography. *Journal of Applied Crystallography*, 52(3), 495-497. <https://doi.org/10.1107/S2052252519005918>
- Jiao, C., Li, K., & Fang, Z. (2022). Data sharing practices across knowledge domains: a dynamic examination of data availability statements in PLOS ONE publications. *Journal of Information Science*, OnlineFirst, 1-17. <https://doi.org/10.1177/01655515221101830>
- Khan, N., Thelwall, M., & Kousha, K. (2022). Are data repositories fettered? A survey of current practices, challenges and future technologies. *Online Information Review*, 46(3), 483-502. <https://doi.org/10.1108/OIR-04-2021-0204>
- Kindling, M. & Strecker, D. (2022). Data quality assurance at research data repositories. *Data Science Journal*, 21(18), 1-17. <http://doi.org/10.5334/dsj-2022-018>
- Löffler, F., Wesp, V., König-Ries, B., & Klan, F. (2021). Dataset search in biodiversity research: do metadata in data repositories reflect scholarly information needs?. *Plos One*, 16(3), e0246099. <https://doi.org/10.1371/journal.pone.0246099>
- Magazine, D. L. (2017). The landscape of research data repositories in 2015: a re3data analysis. *D-Lib Magazine*, 23(3-4). <https://doi.org/10.1045/march2017-kindling>
- Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist*, 63(5), 643-664. <https://doi.org/10.1177/0002764218784991>
- Marcial, L. H. & Hemminger, B. M. (2010). Scientific data repositories on the Web: an initial survey. *Journal of the American Society for Information Science and Technology*, 61(10),

- 2029-2048. <https://doi.org/10.1002/asi.21339>
- McDonald, L., Schultze, A., Simpson, A., Graham, S., Wasiak, R., & Ramagopalan, S. V. (2017). A review of data sharing statements in observational studies published in the BMJ: a cross-sectional study. *F1000Res*, 19(6). <https://doi.org/10.12688/f1000research.12673.2>
- McGuinness, L. A. & Sheppard, A. L. (2021). A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts. *PLoS One*, 16(5), e0250887. <https://doi.org/10.1371/journal.pone.0250887>
- OECD (2016). *OECD Science, Technology and Innovation Outlook 2016*. Paris: OECD Publishing. https://doi.org/10.1787/sti_in_outlook-2016-en
- Page, M. J., Nguyen, P. Y., Hamilton, D. G., Haddaway, N. R., Kanukula, R., Moher, D., & McKenzie, J. E. (2022). Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. *J Clin Epidemiol*, 147, 1-10. <https://doi.org/10.1016/j.jclinepi.2022.03.003>
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H. J., Gundlach, J., Schirmbacher, P., & 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>
- PLOS (2019, December 5). Data Availability. <https://journals.plos.org/plosone/s/data-availability>
- Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, 570(7759), 27-29. <https://doi.org/10.1038/d41586-019-01720-7>
- Thelwall, M. & Kousha, K. (2016). Figshare: a universal repository for academic resource sharing?. *Online Information Review*, 40(3), 333-346. <https://doi.org/10.1108/OIR-06-2015-0190>
- Thelwall, M. & Kousha, K. (2017). Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib Journal of Information Management*. <https://doi.org/10.1108/AJIM-09-2016-0159>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E.,

Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, Jun., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

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

- Cho, Jane & Park, Jong-Do (2019). A study on analysis of research data repository in humanities and social sciences. *Journal of the Korean BIBLIA Society for library and Information Science*, 30(2), 69-87. <https://doi.org/10.14699/kbiblia.2019.30.2.069>
- Hwang, Hyekyong & Lee, Jeeyeon (2016). The influence of socio-demographic characteristics on adoptive intention of institutional repositories. *Proceedings of the 23th Korean Society for Information Management Conference*, 49-55.
- Joung, KyoungHee, Lee, Jae Yun, Chung, EunKyung, & Choi, Sanghee (2021). A study on methods of implementation for the NRF open access policy. *Journal of the Korean Society for Information Management*, 37(4), 255-286. <https://doi.org/10.3743/KOSIM.2020.37.4.255>
- Jung, Youngmi & Lee, Sang-Gi (2010). A study on the operational policies and best practice of repository. *Journal of the Korean Society for Information Management*, 27(4), 131-152. <https://doi.org/10.3743/KOSIM.2010.27.4.131>
- Kim, Jihyun (2016). A study on policy components of data access and use controls in research data repositories. *Journal of Korean Library and Information Science Society*, 47(3), 213-239. <https://doi.org/10.16981/kliss.47.3.201609.213>
- Kim, So-Young, Kim, Jihyun, Choi, Heeseok, & Hwang, Hyekyong (2016). An analysis on open access policies on publications funded by overseas public institutions. *Journal of the Korean Society for Library and Information Science*, 50(4), 209-229. <https://doi.org/10.4275/KSLIS.2016.50.4.209>
- Kim, Woojoong, Kim, Hyungjoon, Park, Heejin, & Choi, Sangok (2021). An analysis on operational level, subjects, content types and utilization of research data repositories in natural science. *Journal of Korea Technology Innovation Society*, 24(4), 577-595. <https://doi.org/10.35978/jktis.2021.8.24.4.577>
- Kwon, Nahyun & Pyo, Soon Hee (2022). Measuring the economic value of a self archiving and journal repository system for open access. *Journal of the Korean BIBLIA Society for library*

- and Information Science, 33(3), 179-203. <https://doi.org/10.14699/kbiblia.2022.33.3.179>
- Park, Kyu-Ri & An, Byoung-Goon (2017). A study on developing national research data repository. Proceedings of the 24th Korean Society for Information Management Conference, 33-38.
- Shim, Wonsik (2021). A study on the awareness and experience of open access publishing among korean authors of an open access mega journal. Journal of the Korean Society for Library and Information Science, 55(4), 45-65. <https://doi.org/10.4275/KSLIS.2021.55.4.045>
- Shim, Wonsik, An, Byoung-Goon, Park, Seong-Eun, & Kim, Hyun Soo (2020). Quantitative analysis on PLoS ONE articles published by authors affiliated with korean institutions. Journal of the Korean Society for Information Management, 37(2), 47-69. <https://doi.org/10.3743/KOSIM.2020.37.2.047>
- Shim, Wonsik, An, Hye-yeon, & Byun, Jeayeon (2015). Strategies for improving the collection and use of research data in the humanities. Journal of the Korean Society for Library and Information Science, 49(3), 155-183. <https://doi.org/10.4275/KSLIS.2015.49.3.155>
- Shin, Eunjung & Jung, Wonkyo (2017). Expansion and implications of open science policy. STEPI Insight, 216, 1-39.
- Shin, Young-Ran & Chung, Yeon-Kyoung (2012). A study on the improvement plans of the humanities and social sciences research data archives in Korea. Journal of Korean Society of Archives and Records Management, 12(3), 93-115. <https://doi.org/10.14404/JKSARM.2012.12.3.093>