

maDMP 구현 사례와 적용방안에 관한 연구*

A Study on the maDMP (machine-actionable DMP) Implementation Cases and its Application Method

김 주 섭 (Juseop Kim)** , 김 선 태 (Suntae Kim)***
한 연 중 (Yeonjung Han)**** , 유 원 재 (Won-Jae Youe)*****

초 록

최근 국내의 출연연을 중심으로 DMP 작성 및 제출이 점차적으로 의무화되고 있다. 그러나 DMP 작성이 서면 또는 자유 텍스트로 기술되다 보니 표준 및 형식 그리고 관리 측면에서 비표준화 및 불충분한 작성으로 인하여 연구데이터 관리를 제대로 설명하지 못하는 문제점이 발생하고 있다. 따라서 본 연구에서는 기계가 자동으로 생성하고 유지할 수 있는 기계가독형 DMP에 대하여 사례조사를 진행하였으며 maDMP를 적용할 수 있는 방안에 대해서 제안하였다. 조사된 maDMP 사례에는 RDCS, Argos, Haplo Repository 그리고 DMap을 포함하였다. 또한 maDMP를 적용할 수 있는 방안으로 영구 식별자의 사용, 통제어휘 적용 그리고 온톨로지와 같은 시멘틱 기술의 적용을 들 수 있다.

ABSTRACT

Recently, the preparation and submission of DMP is gradually becoming compulsory, centering on domestic government-funded research institutes. However, as DMP preparation is described in written or free text, there is a problem that research data management cannot be properly explained due to non-standardization and insufficient preparation in terms of standards, formats, and management. Therefore, in this study, a case study was conducted on a machine-readable DMP that can be automatically generated and maintained by a machine, and a method for applying maDMP was proposed. Examples of maDMP investigated included RDCS, Argos, Haplo Repository, and DMap. In addition, the use of permanent identifiers, application of controlled vocabulary, and application of semantic technologies such as ontology can be mentioned as possible ways to apply maDMP.

키워드: 오픈 사이언스, 데이터 관리 계획, 기계 가독형 데이터 관리 계획, RDA, RDCS
Open Science, DMP (Data Management Plan), maDMP (Machine-Actionable Data Management Plan), RDA (Research Data Alliance), RDCS (RDA Commons Standard)

* 이 논문은 2021년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.
본 연구는 2021년도 국립산림과학원의 연구지원에 의해 이루어진 것임.
** 전북대학교 문헌정보학과 강사(kimjuseop@jbnu.ac.kr) (제1저자)
*** 전북대학교 문헌정보학과 부교수(kim.suntae@jbnu.ac.kr) (교신저자)
**** 국립산림과학원 임산자원이용연구부 임업연구사(yeonjungh@korea.kr) (공동저자)
***** 국립산림과학원 임산자원이용연구부 임업연구사(sngkgk@korea.kr) (공동저자)
논문접수일자 : 2021년 11월 22일 논문심사일자 : 2021년 11월 22일 게재확정일자 : 2021년 12월 14일
한국비블리아학회지, 32(4): 111-134, 2021. <http://dx.doi.org/10.14699/kbiblia.2021.32.4.111>

© Copyright 2021 Korean Biblia Society for Library and Information Science
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

1.1 연구의 필요성 및 목적

최근 해외 연구비 제공기관(Funder, 이하 펀더)을 중심으로 데이터 관리 계획(Data Management Plan, 이하 DMP) 제출을 의무화하고 있다. 이러한 배경에는 연구데이터에 제약 없이 자유롭게 개방, 공유, 접근할 수 있는 오픈 사이언스가 자리 잡고 있다. 오픈 사이언스는 국가의 세금으로 생산한 연구 결과물에 대한 접근 확대가 핵심으로서 연구 결과와 이점이 가능한 한 널리 퍼지도록 만들고 있다.

DMP는 연구 수행 중 수집하거나 생성한 데이터에 대하여 체계적으로 관리하기 위한 문서로서 일반적으로 데이터 접근, 공유 그리고 보존 등과 같은 데이터 관리 방법에 대한 안내라고 볼 수 있다. 상술한 오픈 사이언스의 저변 확대에 따라 미국 NSF(미국 과학 재단, 이하 NSF) 그리고 영국 DCC(데이터 큐레이션 센터, 이하 DCC) 등과 같은 해외 주요 연구 기관을 중심으로 DMP 제출이 의무화되기 시작했다. 국내에서도 한국연구재단, 정보통신기획평가원, 한국과학기술정보연구원, 한국지질자원연구원 등 정부출연기관을 중심으로 DMP 제출을 점차 확대해가고 있는 추세이다.

이러한 DMP 작성에 대한 중요성이 증대되고 있지만 실제로 DMP가 데이터 관리 활동에 효과적으로 사용되지 못하고 있다는 문제점이 나타나고 있다. 실제로 DMP가 데이터 관련한 설명이 부족하다는 점, DMP가 가지는 실제 이점을 증명하지 못하는 점, 보존 측면에서의 문제점 그리고 DMP 표준화와 업데이트 부족 등이

문제점으로 제시되고 있다(Loon et al., 2017; Smale et al., 2018; Bakos, Miksa, & Rauber, 2018; Cardoso, Proença, & Borbinha, 2020).

이와 같은 문제점을 개선하기 위하여 기계가독형 데이터 관리 계획(machine-actionable Data Management Plan, 이하 maDMP)이 대안으로 등장하였다. maDMP는 DMP를 상호운용 가능하고 자동화하여 점차 표준화하는 것을 목표로 하는 개념이다(Cardoso, Proença, & Borbinha, 2020). 이러한 maDMP는 DMP 내용의 일부에 대하여 생성 및 유지를 자동화할 수 있으며 작성 누락 문제의 해결 및 주기적인 DMP 업데이트 등이 가능하다는 특징을 가지고 있다(Cardoso, Proença, & Borbinha, 2020; Bakos, Miksa, & Rauber, 2018; Simms et al., 2017).

이러한 배경에 따라 본 연구에서는 DMP 작성의 문제점과 이를 해결하기 위한 방안으로 제시된 maDMP에 대하여 알아보고 RDA 공통 표준과 같은 구체적인 구현 사례를 통해 DMP를 기계가독형 DMP로 적용할 수 있는 방안에 대해서 논의하고자 한다.

1.2 연구방법

아직 maDMP에 대한 연구는 해외를 중심으로 이루어지고 있지만 많은 연구가 수행되지는 않았고 특히 국내의 경우 Kim(2020)을 제외하고는 해당 연구가 전무한 실정이다. 따라서 본 연구의 목적을 달성하기 위하여 maDMP의 이론적 배경과 관련 연구를 검토하고 RDA 공통 표준 등과 같은 구체적인 maDMP 구현 사례를 통해 국내에서 maDMP를 적용할 수 있는 방법을 제안하고자 한다.

본 연구에서 포함될 maDMP의 구현사례는 다음과 같다.

- RDA Commons Standard(RDCS)
- Argos
- Haplo repository
- The Machine-actionable Data Management planning application (DMap)

이번 연구에서는 먼저 DMP와 maDMP에 대한 개념적 정의를 확인하고 특히 maDMP에 대한 선행연구 분석을 통해 특징과 이점 및 발전방향에 대해서 소개하고자 한다. 다음으로 maDMP의 실제 구현 사례로서 상술한 RDA 공통 표준 등에 대한 내용을 제시하고 마지막으로 maDMP 적용방안에 대해 제안하고자 한다.

본 연구의 제한점으로 아직 maDMP에 대한 연구가 미흡하여 현황에 대한 조사 및 분석이 주를 이루고 있으며 마지막으로 적용방안도 개념적인 부분에 그치고 있다는 것이다. 그러나, 이번 연구를 기회로 국내에서도 maDMP에 대한 연구가 활성화될 것이라 기대되며 DMP 자동화에 대한 구체적인 논의가 이루어질 것이라 판단된다.

2. 이론적 배경 및 관련 연구 검토

2.1 DMP

DMP란 연구데이터를 체계적으로 관리할 목

적으로 연구 중에 획득하거나 생성할 데이터에 대하여 데이터 관리, 기술 및 저장 방법 그리고 데이터를 보존하는 방법에 대하여 설명하는 공식 문서를 말한다(Stanford Libraries, n.d.). 해외의 경우 주요 펀더를 중심으로 연구비 신청서와 함께 DMP를 제출하도록 요구하고 있으며 국내에서도 DMP 적용을 위한 이행방안 마련을 시작으로 연구데이터 정의와 DMP 도입 및 정보등록의 근거가 법령에 신설되었으며 국가 출연연을 중심으로 DMP 작성 및 제출이 의무화되고 있는 추세이다.

이러한 DMP의 기원과 역사를 살펴보면 다음과 같다(Smale et al., 2018).

- 1966년: 항공 및 공학 프로젝트에서 처음 사용
- 1970년대 후반~1980년대: 다양한 공학 및 과학 분야로 확대
- 2000년~2010년: e-리서치와 경제정책
- 2010년~2017년: 미국 NSF를 시작으로 공적 펀딩에 대한 DMP 제출 요구

DMP는 1966년에 항공 및 공학 프로젝트에 사용되었으며 초기에는 예상되는 연구개발 활동을 요약하는 문서로 활용되었다. 2000년부터 2010년까지는 e-리서치와 경제정책이 DMP 작성을 동인하는 대표적인 분야가 되었다. 2010년부터 2017년까지는 미국 국립과학재단의 펀딩의 결과로 제시된 연구데이터의 보존 및 공유를 위해 DMP 제출이 요구되는 시기였고 2013년에 미국 OSTP(Office of Science and Technology Policy) 메모에서는 펀딩을 받는 연구자들에게 DMP 작성을 의무화하는데 필요한 정책을 제

시하였다.

이러한 DMP를 작성하여 얻을 수 있는 이점을 정리하면 다음과 같다(Stanford Libraries, n.d.; UNIVERSITY OF EDINBURGH, n.d.; Jones, 2011).

- 계획의 약점을 파악하여 개선할 수 있음
- 프로젝트 전반에 걸쳐 데이터 작업을 위한 프레임워크 구축 및 연구데이터 지원을 위한 리소스 식별 가능
- 프로젝트 도중 연구 인력이 변경되더라도 연속성 유지 가능
- 연구의 중복 방지 및 데이터 공유에 따른 협업 연구 가능
- 공유된 데이터의 인용으로 데이터에 대한 신뢰도 확보

DMP의 세부 내용에는 데이터 유형, 데이터 기술을 위한 문서화와 메타데이터, 스토리지, 백업 및 보안, 접근 및 공유 그리고 데이터 아카이빙 및 접근 방법 등이 포함된다(Harvard Medical School, 2021).

2.2 maDMP

DMP 작성에 대한 의무화가 점차 확대되고 있는 시점에도 불구하고 기존 DMP 작성이 데이터 관리에 있어 효과적으로 사용되지 못하는 것으로 나타났다.

다음의 <표 1>은 기존 DMP가 가지는 한계점을 정리한 것이다(Loon et al., 2017; Smale et al., 2018; Bakos, Miksa, & Rauber, 2018; Cardoso, Proença, & Borbinha, 2020).

본 연구의 핵심인 maDMP는 기존 DMP가 가지는 단점을 극복하기 위한 대안으로 제시되고 있다. 이러한 maDMP는 기계 가독형 또는 실행 가능한 데이터 관리 계획이라고 부르며 DMP를 상호 운용 가능하고 자동화하며 점차 표준화하는 것을 목표로 하는 개념이다. 다음은 기존 연구에서 살펴본 maDMP의 이점을 정리한 것이다(Cardoso, Proença, & Borbinha, 2020; Bakos, Miksa, Rauber, 2018; Miksa et al., 2019).

- 연구데이터 관리 작업의 편이성 및 관리 부담 최소화

<표 1> 기존 DMP가 가지는 한계점

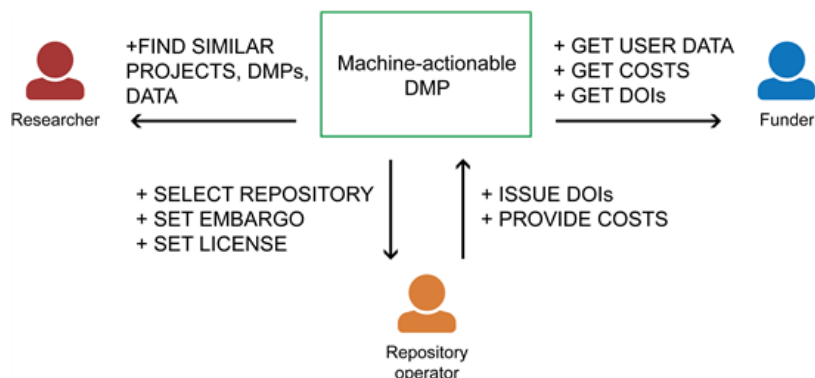
구분	한계점
표준 및 형식	<ul style="list-style-type: none"> • 생성 데이터, 데이터 관리법, 보존 및 공유법 설명 불충분 • 자유 텍스트로 기록하거나 미기입하는 경우 발생 • 비표준화 및 일반적 지침의 부재 • 연구데이터 관리를 실제로 설명하지 못함
데이터 관리	<ul style="list-style-type: none"> • 기관 준수 요건을 형식적으로만 준수하는 경우 발생 • 데이터 관리자 개인을 명시하지 않음 • 무지로 인한 데이터 관리 미숙 문제 • 관리 부담 및 비생산성 • 미흡한 보존 관리작업 및 워크플로우 문제
공개성	<ul style="list-style-type: none"> • 기대에 비해 정부 및 펀더 측에서 얻는 이득이 불확실 • DMP 작성에 참고할 모범사례의 부족

- DMP 내 정보 품질 개선
- 고품질의 정보로 인한 풍부한 데이터 및 부가가치 창출
- DMP 내 일부 정보의 작성 누락 문제 해결 및 주기적 업데이트 가능
- 시스템 및 도구 간 교환 용이성으로 접근, 검색 및 재사용 가능
- 시스템 간 연계를 바탕으로 자동화 서비스 실행 가능
- 연구 생태계 전반에 따른 이용자 경험 개선

maDMP의 주요 목표 중 하나는 일부 정보의 생성 및 유지를 자동화하는 것으로 DMP 정보 생성을 자동화할 시, 다양한 장점을 얻을 수 있다. 연구자가 수행하던 복잡한 연구데이터 관리 작업 및 과정을 간소화시켜 관리 작업의 편의성을 도모할 수 있으며, 특히 여러 위치에 동일한 정보를 입력해야 하는 관리 부담을 완화할 수 있다. 또한 시간과 노력이 많이 소모되었던 연구데이터 관리 작업을 축소함으로써, 연구자가 온전히 연구 활동에 전념할 수 있어 정보 자체의 품질을 향상시킬 수 있다. 고품질의 정보는

풍부한 데이터와 부가가치를 창출할 수 있도록 하며, 이러한 품질 향상은 기계가 자동으로 작성하는 maDMP의 특성에서 비롯되는데 과거에는 수동으로 행해졌던 DMP 작성을 자동화하는 것으로 작성의 누락과 주기적인 업데이트의 부재를 해결할 수 있으며, 더 효과적으로 데이터 관리 활동을 지원할 수 있다. 기존 DMP는 연구비 신청 단계에서만 작성하고 제출하면 되지만 maDMP의 경우 펀더 신청 외의 단계에서도 유용하게 이용할 수 있도록 업데이트할 수 있다는 것이다. 또한 maDMP는 시스템 또는 도구 간 교환이 용이하므로 액세스, 검색 및 재사용이 원활하게 이루어질 수 있으며 또한, 시스템 간 교환을 바탕으로 자동 평가 및 모니터링과 같은 추가적인 자동화 서비스를 연계할 수 있다. 마지막으로 연구 프로젝트 과정 중 이해관계자뿐만 아니라 연구 생태계 전반에 이르러 이용자의 경험을 개선하는 기회가 될 수 있다.

다음 <그림 1>은 연구 프로젝트의 이해관계자인 연구자, 펀더, 그리고 Repository 관리자가 maDMP를 통해 상호작용하는 DMPTool의 예시를 나타낸 것이다(DMPTool Blog, 2021).



<그림 1> maDMP 생태계의 이해관계자 상호작용의 예시

maDMP 생태계 내에서, 이해관계자는 DMP를 통해 정보를 교환하고 서로 소통한다. Repository 운영자는 적절한 Repository의 선택, 엠바고 기간을 설정하고 연구자가 제출한 데이터에 올바른 라이선스를 할당할 수 있다. 시스템은 데이터에 할당된 DOI 목록과 저장 및 보존 비용에 대한 정보를 제공한다. 펀더는 DMP가 어떻게 구현되었는지 확인할 수 있다. 또한 maDMP에는 이해관계자가 연구 프로젝트 주기 동안 업데이트된 정보를 질의할 수 있는 변경 내역과 함께 해당 프로젝트 및 결과물에 대한 주요 정보가 포함된다.

2.3 선행연구

이번 절에서는 maDMP와 관련된 국내외 선행연구를 살펴보고 해당 연구에서 나타난 시사점을 분석하여 본 연구가 가지는 의의를 제시하고자 한다. 다음의 <표 2>는 DMP 및 maDMP와 관련된 연구 목록을 나타낸 것이다.

Kim(2020)은 연구데이터의 다양한 이해관계자의 편의성을 높이는 것을 목적으로, maDMP

구현을 위한 참고자료로써 RDA 공통 표준을 분석하였다. RDA 공통 표준에서 제안한 모델의 클래스를 분석한 뒤 학술문헌 클래스 확충 등의 보완 방안을 추가하였고, DMPOnline에 등록된 DMP를 검토하여 매핑이 누락된 속성(Selection 속성)을 발견하였다. 추가적으로, DMP에 프로젝트에 참여하는 연구원의 소속 기관 정보를 설명하는 속성 추가와, 속성에서의 네임스페이스와 DOI 사용을 확대할 것을 제안하였다. 해당 연구에서 저자는 연구원의 복수 조직 정보를 포함하기 위하여 기여자(contributor) 엔티티와 연결된 조직 엔티티에 대한 추가적인 연구 및 Shema.org와 FAOF(Friend of a Friend) 용어의 조사 분석을 후속 조치로 제시되었다.

Cardoso, Proença, Borbinha(2020)는 maDMP를 표현하고 그 기능을 활용하기 위하여 시멘틱 기술을 사용하는 방법을 설명하고, 온톨로지로 표현된 DMP가 펀더의 요구사항 평가에 어떻게 영향을 미칠 수 있는지에 대하여 기술하였다. 저자는 DMP 공통 표준 모델(DMP Common Standards model)의 구현을 위해 생성된 DCSO

<표 2> 선행연구 목록 및 주요 내용

논문명	시사점
maDMP 모델 분석 및 개선 방향(2020)	RDA 공통 표준의 개선방향
maDMP: 펀더의 요구 사항 평가를 자동화하기 위한 지식 접근 방식(2020)	기존 DMP 표준화 부족 및 업데이트 부재
maDMP를 위한 10가지 원칙(2019)	maDMP 이점 실현을 위한 10가지 원칙
DMP 역사, 옹호 및 효과(2018)	DMP의 기원, 역사 및 문제점
프로세스 엔진 및 maDMP를 이용한 연구데이터 보존(2018)	maDMP를 활용한 보존 측면에서 DMP 보완 방안 제시
기계-가독형 DMP(2017)	maDMP 구현을 위한 공통 표준 필요성 제안
연구중심대학에서 DMP 품질평가(2017)	기존 DMP의 문제점과 해결 방안 제시
DMP의 미래: 도구, 정책 및 참여자(2016)	연구자, 펀더, 출판물 등 연계 가능한 DMP를 위한 글로벌 프레임워크 구축 제안
우수한 DMP 작성하기 위한 10가지의 간단한 규칙(2015)	DMP 작성을 위한 10가지 규칙 제시

(DMP Common Standard Ontology, 이하 DCSO)가 특정한 DMP 템플릿에 적용하기에는 너무 일반적이라고 지적하며, 이러한 특정 문맥을 충족시키는 DCSO 확장(extensions)을 만들 것과 기존 DMP 프레임워크를 이용하여 인스턴스화된 DCSO를 자동 생성할 것을 주장하였다.

Miksa et al.(2019)는 maDMP을 실행하고 그 이점을 실현하기 위한 10가지 원칙을 제시하였다. 다음은 10가지 원칙을 정리한 것이다.

- 이해관계자를 대신하여 작동하는 자동화 시스템 도입
- 인간과 기계에 모두 적용되는 정책 고안
- 데이터 관리 생태계(인간 및 기계를 모두 포함하는) 구성요소 기술
- 영구식별자(Persistent Identifier, 이하 PID)와 통제어휘 사용
- maDMP의 공통 데이터 모델 준수
- 인간과 기계가 모두 이용할 수 있는 DMP 수립
- 데이터 관리 평가와 모니터링 지원
- 수정 가능하고 현재 사용할 수 있는, 버전 관리가 된 DMP로 작성
- DMP의 공공 이용 지원

Smale et al.(2018)는 DMP의 기원, 역사 및 DMP 사용 시 이점에 대하여 연구를 수행하였다. 저자는 DMP를 통해 서비스와 연구자를 연결하면 연구자가 해당 연구 결과를 더 효율적으로 관리할 수 있으며 지적 재산권, 윤리의식 및 데이터 출판 관련 문제를 다룰 수 있는 일반적인 지침이 제공될 수 있는 등 여러 장점이 있

다고 분석하였다. 또한 저자는 DMP의 역사와 기관 및 펀더의 정책, DMP를 사용할 때의 이점 및 현황에 대한 연구 결과를 제시하면서 정부 및 펀더 차원에서 공유와 재사용으로 인한 편익은 명확히 나타나지 않았다고 지적하였다. 기관 차원에서는 기관 준수 요건을 갖춘 DMP에 대한 이점이 있으나, 때때로 학자들이 DMP 요구사항을 해결한 것처럼 보이도록 다양한 전략을 사용하기도 한다. 따라서 저자는 펀더가 DMP 평가 및 연구 후 관리를 위한 효율적인 프로세스에 투자하지 않는 한 DMP가 성공적으로 결과를 달성하였는지 판단할 수 없다고 주장하면서 DMP의 목적과 사례를 재차 검토하고, DMP를 결과 중심적이며 사례에 일관된 방식으로 구현할 것을 제안하였다.

Bakos, Miksa, Rauber(2018)는 연구데이터의 관리 작업 중 보존 측면에서 DMP를 보완하는 방안에 대한 연구를 수행하였다. 데이터 기반 연구 프로젝트에서 이전 실험 및 결과물에 대한 접근, 인용, 연결 및 참조하기 위하여 Repository 사용이 요구되지만, 실제로는 네트워크 드라이브 또는 공유 폴더에 저장하는 경우가 발생하고 있다고 지적하였다. 보존 측면에 대한 또 다른 문제로는, 보존 워크플로우의 복잡성을 언급하였고 또한 연구자 측면의 문제로, 연구자들이 관리 작업의 모든 단계를 거치지 않는 경향이 있다는 것을 확인하였다. 이러한 문제로 인해 보존 관리 작업의 복잡성이 심화되므로, 저자는 사전 정의된 보존 워크플로우 실행 프로세스 엔진을 이용하여 데이터 관리의 일상적인 워크플로우에 보존 기능을 통합하는 매키니즘을 제시하였다.

Simms et al.(2017)는 DMP가 가지고 있는 문제점을 해결하기 위하여 maDMP 구현을 위

한 공통 표준의 필요성을 제안하였다. maDMP는 동적(active 또는 dynamic) 또는 기계 가독형(machine-readable) DMP라고 불리기도 하며, 인간이 읽을 수 있는 형식을 유지하는 동시에 컴퓨터 또한 해당 DMP를 읽을 수 있도록 작성된다. 이러한 형태의 DMP는 기계 실행이 가능하며 연구자, 펀더, Repository 관리자, 프로젝트 관리자 그리고 데이터 사서(data librarians) 등과 같은 이해관계자들을 위한 부가가치를 창출할 수 있는 풍부한 잠재력을 가질 수 있음을 시사하였다.

Loon et al.(2017)는 웨인 주립대 DMP 내용 분석을 통해 NSF 지침의 전반적인 품질과 준수 여부를 평가하고, NSF 지침의 준수 여부가 학술 단위에 따라 다른지 살펴보았다. 평가 결과, 많은 DMP가 생성 대상 데이터, 데이터 관리법 또는 보존 및 공유법을 적절하게 설명하지 못하고 있음을 밝혀냈다. 본 연구결과는 도서관이 연구데이터 관리 및 보급에 있어 DMP 자체의 품질 개선 및 모범 사례 개발에 대하여 보다 많은 의견수렴과 교육 및 컨설팅을 제공해야 하며 이러한 노력이 각 연구자 집단에 맞게 조정되어야 함을 시사하였다.

Simms et al.(2016)는 DMP의 미래를 도구, 정책, 참여자를 중심으로 제시하였다. 특히 DCC의 DMPOnline과 CDL(California Digital Library)의 DMPTool을 언급하면서 해당 도구가 데이터 공유 및 재사용이 가능할 수 있도록 가이드라인을 제공한다는 점을 언급하였다. 여기에 저자는 연구자, 펀더, 출판물, 데이터 및 연구 라이프 사이클의 기타 구성 요소와 연결하는 DMP를 위한 새로운 글로벌 프레임워크 구축할 것을 제안하였다.

Michener(2015)는 그의 연구에서 DMP 작성을 위한 10가지 규칙을 제안하였다. 10가지 규칙에는 펀더의 요구사항 분석, 수집 대상 데이터의 식별, 데이터 조직, 데이터 도큐멘테이션, 데이터 품질 보증 및 품질 관리, 데이터 저장 및 보존, 프로젝트 데이터 정책, 데이터 배포, 역할 및 책임 할당 마지막으로 예산 준비가 포함된다.

이상과 같이 DMP와 maDMP 관련 연구 9편을 살펴본 결과 기존 DMP의 문제점을 지적하고 해결방안을 제시한 논문이 3편, DMP 작성 규칙과 DMP의 미래를 제안한 논문이 2편, 기존 DMP의 문제점을 해결하기 위해 고안된 maDMP 구현과 이점 실현을 위한 10가지 원칙 그리고 개선방향 등 4편으로 나타났다. 상술한 선행연구와 다르게 본 연구에서는 maDMP의 구체적인 사례를 분석하고 maDMP 적용방안에 대하여 개념적 측면에서 논의하였다는 점에서 의의가 있다고 할 수 있다.

3. maDMP 사례

본 장에서는 maDMP의 사례로서 RDA Commons Standard(이하, RDCS) 그리고 RDCS를 기반으로 구현한 Argos, Haplo Repository 그리고 DMap에 대하여 살펴보려고 한다.

3.1 RDA Commons Standard(RDCS)

maDMP의 궁극적 목표는 연구자, 펀더, Repository 관리자, 정보통신기술(ICT) 제공업체 그리고 사서 등 연구데이터 라이프 사이클과 관련한 이해관계자를 대신하여 시스템 간의 상

호운용성을 달성하는 것이다. 따라서 시스템 간의 상호운용성을 달성하기 위하여 공통 표준 및 프로토콜의 필요성이 제기되었다(Simms et al., 2017).

이러한 일환으로 2019년 RDA¹⁾에서는 DMP 공통 표준 워킹 그룹(DMP Commons Standards WG)을 통해 기존 DMP의 정보를 기계에서 실행할 수 있는 방식으로 표현할 수 있는 애플리케이션 프로파일을 개발했다. 이것을 RDA 공통 표준(RDA Commons Standard, 이하 RDCS)이라고 하며 DMP에 제공된 정보의 자동 교환 통합 및 유효성 검사를 허용한다. RDCS는 SBA Research²⁾ 소속의 Tomasz Miksa를 포함한 200여 명의 인원이 개발하였다. 또한 2020년 기준으로 RDCS는 전 세계의 11개 프로젝트에서 공통 표준으로 채택하였다. 채택 기관은 다음과 같다.

- Norwegian Open Research Data Infrastructure(노르웨이)
- EasyDMP(노르웨이)
- RDMO(독일)
- DMP Tool(미국)
- DMP Online(영국)
- F1000Research open research publisher(영국)
- Haplo repository(영국)
- TU Wien, TU Graz, Uni Wien via FAIR Data Austria project(오스트리아)
- Data Stewardship Wizard(EU)

- Argos - OpenDMP(EU)
- DMP OPIDoR(프랑스)

국가를 기준으로 RDCS를 표준으로 채택한 기관을 살펴보면 영국 3곳, 노르웨이 2곳, EU 2곳, 독일 1곳, 미국 1곳, 오스트리아 1곳 그리고 프랑스 1곳으로 나타났다. 다음은 RDCS의 특징과 기능에 대하여 정리한 것이다(Simms et al., 2017).

- 이해관계자의 역할을 자동화: 시스템 간 정보의 자동 교환
- 정보 통합 및 유효성 검사 가능
- 모든 이해관계자의 이용 고려: 포괄성 / 확장 가능

RDCS는 먼저, 이해관계자의 역할이었던 시스템 간 정보를 자동으로 교환할 수 있으며 또한 정보통합 및 유효성 검사를 할 수 있다. 덧붙여 데이터와 관련된 모든 이해관계자 이용을 고려하여 개발되었다. 펀더와 같은 특정 이해관계자 그룹의 요구사항을 반영하지 않았지만 요구사항을 반영할 경우, 확장이 가능하다는 특징을 갖는다. 그러나 확장 요소에 대한 상호운용성을 보장하지 않는다. 다음은 RDCS의 최상위 요소에 대하여 나타낸 것이다(Miksa et al., 2021).

- DMP: DMP 제목, 수정날짜 등을 기술하는 최상위 요소

1) RDA는 2013년에 유럽 위원회, 미국 정부의 국립 과학 재단, 국립 표준 기술 연구소, 호주 정부의 혁신부에 의해 데이터의 공유 및 재사용을 가능하게 하는 사회적 및 기술적 인프라를 위해 커뮤니티 주도 이니셔티브로 출범했다.
2) SBA Research는 COMET Competency Centers for Excellent Technologies의 국가 이니셔티브에서 부분적으로 펀딩을 제공받는 정보 보안 연구 센터임.

- Contact(담당자): DMP와 관련된 정보를 제공할 개인 또는 조직
- Contributor(기여자): 데이터 관리 프로세스 및 DMP 생성과 관리에 관련된 모든 개인
- Cost(비용): DMP와 관련된 비용을 기술
- Dataset(데이터셋): 데이터에 관한 정보를 기술
- Metadata(메타데이터): 데이터를 기술하는데 사용된 메타데이터 표준에 관한 포인터를 기술
- Security and Privacy(보안 및 개인정보): 특정 데이터셋의 보안 및 개인정보보호와 관련된 사항을 기술
- Technical resource(기술적 자원): 현미경과 같이 데이터 생성 및 처리에 필요한 장비 등을 기술
- Distribution(배포): 데이터셋과 관련된 구체적인 정보를 기술
- License(라이선스): 데이터셋을 사용할 수 있는 라이선스를 기술, 엠바고 기술 가능
- Host(호스팅): 현재 데이터가 저장되어있는 시스템, 연구 중에 저장된 시스템에 관한 정보를 기술
- Funding(펀딩): NSF와 같은 프로젝트 중 펀딩 프로젝트에 관한 세부 정보
- Project(프로젝트): 펀딩 프로젝트, 연구 프로젝트를 포함한 DMP와 관련된 모든 유형의 프로젝트

이중 '데이터셋'의 경우 W3C의 DCAT 명세서에서 제공하는 '접근 또는 다운로드가 가능한 단일 에이전트가 제시하거나 선별한 데이터 컬렉션'을 의미하며 또한 '배포'의 경우도 같은 출처를

인용하여 '데이터셋에 관한 구체적인 설명, 자연어, 미디어 유형 그리고 시간 및 공간 해상도 등 다양한 속성을 기술'한다고 제시하고 있다(W3C, 2020). 다음의 <표 3>은 상술한 RDCS의 최상위 요소에 대한 하위 속성을 나타낸 것이다.

먼저, DMP는 '담당자', '기여자', '비용', 'DMP 생성일자', '데이터셋', '설명', 'DMP ID', '윤리 이슈 설명', '윤리이슈 존재여부', '윤리 이슈 보고서', '언어', '수정날짜', '프로젝트' 그리고 'DMP 제목' 등 14개의 하위 속성으로 구성되어 있다. 다음으로 '담당자' 최상위요소는 '담당자 식별자', '담당자 이메일' 그리고 '담당자명' 등 3개를 '기여자'는 '기여자 ID', '기여자 이메일', '기여자명' 그리고 '역할' 등 4개의 하위 속성을 포함하고 있다. 4번째 최상위요소인 '비용'은 '통화 코드', '비용 설명', '비용명' 그리고 '비용 값' 등 4개로 '데이터셋'은 '데이터 품질 보장', '데이터셋 ID', '설명', '배포', '데이터셋 생성일자', '키워드', '데이터셋 기술 언어', '메타데이터', '개인정보 포함 여부', '보존 상태', '보안과 개인정보', '민감한 데이터 존재여부', '기술 자원', '데이터셋 제목' 그리고 '데이터셋 유형' 등 15개의 하위 속성으로 구성되어 있다. 다음으로 '메타데이터'는 '설명', '메타데이터 언어' 그리고 '메타데이터 표준 ID' 등 3개를 '보안 및 개인정보'는 '설명' 그리고 '보안 및 개인정보 계획명' 등 2개를 '기술 자원'은 '설명' 그리고 '리소스명' 등 2개의 하위요소를 포함하고 있다. 다음 '배포'의 하위속성에는 '접근 URL', '이용 기한', '데이터셋 크기', '데이터 접근성', '설명', '다운로드 URL', '포맷', '호스팅 기관', '라이선스' 그리고 '배포명' 등 10개가 '라이선스'는 '라이선스 문서 URL' 그리고 '라이선스 적용 날짜' 등 2개가 포함되어 있

〈표 3〉 RDCS 상위요소 및 하위요소

DMP	담당자 (Contact)	기여자 (Contributor)	비용 (Cost)	데이터셋 (Dataset)	메타데이터 (Metadata)	보안 및 개인정보 (Security and privacy)	기술 자원 (Technical resource)	배포 (Distribution)	라이선스 (License)	호스팅 (Host)	프로젝트 (Project)	펀딩 (Funding)
<ul style="list-style-type: none"> • 담당자 • 기여자 • 비용 • DMP • 생산일자 • 데이터셋 • 설명 • DMP ID • 용리 이슈 • 설명 • 용리 이슈 • 존재여부 • 용리 이슈 보고서 • 언어 • 수정날짜 • 프로젝트 • DMP 제목 	<ul style="list-style-type: none"> • 담당자 ID • 담당자 이메일 • 담당자명 	<ul style="list-style-type: none"> • 기여자 ID • 기여자 이메일 • 기여자명 • 역할 	<ul style="list-style-type: none"> • 통화 코드 • 비용 설명 • 비용명 • 비용 값 	<ul style="list-style-type: none"> • 데이터 품질 • 데이터셋 보장 • 데이터셋 ID • 설명 • 배포 • 데이터셋 생성일자 • 키워드 • 데이터셋 기술 언어 • 메타데이터 • 개인정보 포함 여부 • 보존 상태 • 보안과 개인정보 • 민감함 • 데이터 존재여부 • 기술 자원 • 데이터셋 제목 • 데이터셋 유형 	<ul style="list-style-type: none"> • 설명 • 메타데이터 언어 • 메타데이터 표준 ID 	<ul style="list-style-type: none"> • 설명 • 보안 및 개인정보 계획명 	<ul style="list-style-type: none"> • 설명 • 리소스명 	<ul style="list-style-type: none"> • 접근 URL • 이용 기한 • 데이터셋 크기 • 데이터 접근성 • 설명 • 다운로드 URL • 포맷 • 호스팅 기관 • 라이선스 • 배포명 	<ul style="list-style-type: none"> • 라이선스 • 문서 URL • 라이선스 적용 날짜 	<ul style="list-style-type: none"> • 이용가능성 • 백업 빈도 • 백업 유형 • Repository 인증 표준 • 설명 • 호스팅 기관 위치 • PID 시스템 • 스트리지 유형 • 버전 관리 지원 여부 • 호스팅 기관명 • 시스템 URL 	<ul style="list-style-type: none"> • 설명 • 프로젝트 시작 일자 • 프로젝트 종료일자 • 펀딩 정보 • 프로젝트명 	<ul style="list-style-type: none"> • 펀딩 ID • 펀딩 상태 • 지원금 ID

다. 다음으로 '호스팅' 요소에는 '이용가능성', '백업 빈도', '백업 유형', 'Repository 인증 표준', '설명', '호스팅 기관 위치', 'PID 시스템', '스토리지 유형', '버전 관리 지원 여부', '호스팅 기관명' 그리고 '시스템 URL' 등 11개로 '프로젝트' 요소는 '설명', '프로젝트 시작 일자', '프로젝트 종료일자', '펀딩 정보' 그리고 '프로젝트명' 등 5개로 마지막으로 '펀딩' 요소에는 '펀더 ID', '펀딩 상태' 그리고 '지원금 ID' 등 3개로 구성되어 있다.

이와 같이 RDCS를 분석해본 결과 13개의 상위요소와 78개의 하위요소를 포함하고 있음을 알 수 있다. 이 중에서 PID를 사용하는 요소는 7개로 contact ID(담당자), contributor ID(기여자), Dataset ID(데이터셋), DMP ID, funder ID(펀더), grant ID(연구비 지원) 그리고 metadata standard ID(메타데이터 표준)이다. 다음은 각 요소에 대하여 PID 서비스를 적용하여 통제어휘를 나타낸 것이다.

- contact ID(담당자), contributor ID(기여자): Open Researcher and Contributor ID(ORCID), International Standard Name Identifier(ISNI), OpenID 등
- Dataset ID(데이터셋), DMP ID: handle, Digital object Identifier(DOI), Archival Resource key(ARK), URL 등
- funder ID(펀더): funderRef ID, URL 등
- grant ID(연구비 지원), metadata standard ID(메타데이터 표준): URL 등

또한, 78개의 하위요소의 데이터 유형을 살펴본 결과 통제어휘로 사용할 수 있는 요소는 12개이다. 다음의 <표 4>는 요소별 통제어휘값과 예시를 나타낸 것이다.

통제어휘값을 가지는 하위요소는 12개로서 'ethical issues exist', 'currency code', 'language', 'personal data', 'sensitive data', 'language', 'data access', 'certified with', 'geo location', 'PID system', 'support versioning' 그리고 'funding status'이다. 이 중 'yes/no/unknown'과 같이 여부를 확인하기 위한 통제어휘값을 갖는 요소는 'ethical issues exist', 'personal data', 'sensitive data' 그리고 'support versioning'이다. 데이터 접근성을 식별하기 위한 'data access'는 'open/shared/closed'와 같은 통제어휘값을 Repository 인증 표준을 선택하기 위한 'certified with'는 'din31644' 등 8개의 통제어휘값을 가지고 있다. PID 시스템의 경우 'ark' 등 19개 값을 펀딩 상태를 나타내는 'funding status'는 'planned/applied/granted/rejected' 등 4개 값을 가지고 있다. 마지막으로 'currency code', 'language' 그리고 'geo location'은 각각 'ISO 4217', 'ISO 639-3' 그리고 ISO 3166-1에서 정의한 코드값을 가진다.

3.2 Argos OpenDMP

ARGOS는 FAIR(Findable, Accessible, Interoperable and Reusable)(GOFAIR, 2016)³⁾ 및 Open 모범 사례를 다루고 사용 및 채택에

3) 2016년에 '과학데이터 관리 및 관리를 위한 FAIR 지침'이 『Scientific Data』에 게재되었다. 저자들은 디지털 자산의 검색, 접근성, 상호운용성 및 재사용을 개선하기 위한 지침을 제공하고자 했으며, 이 원칙은 인간의 데이터의 볼륨, 복잡성 및 생성 속도의 증가로 인해 데이터를 처리하기 위해 컴퓨터 지원에 점점 더 의존하기 때문에 기계 작동성(즉, 인간의 개입 없이 데이터를 검색, 액세스, 상호 운용 및 재사용할 수 있는 컴퓨터 시스템의 용량)을 강조하였다.

〈표 4〉 RDCS의 요소 중 통제어휘로 표현 가능한 요소

maDMP 요소		설명	통제어휘값	예시
DMP	ethical issues exist	DMP의 윤리적 이슈 존재 여부	yes / no / unknown	
Cost	currency code	ISO 4217에서 정의한 화폐 코드값		KRW
Dataset	language	ISO 639-3에서 정의한 언어 코드값		kor
	personal data	데이터셋의 개인 정보 포함 여부	yes / no / unknown	
	sensitive data	데이터셋의 윤리적 이슈 존재 여부	yes / no / unknown	
Metadata	language	ISO 639-3에 따른 메타데이터 표준에서 기술한 언어		
Distribution	data access	데이터 접근성을 식별	open / shared / closed	Open
Host	certified with	Repository를 인증한 인증 표준	din31644 / dini-zertifikat / dsa / iso16363 / iso16919 / trac / wds / coretrustseal	coretrustseal
	geo location	ISO 3166-1(알파벳 두 글자)에 따른 데이터의 물리적 위치		KO
	PID system	호스팅 기관이 사용하는 PID 시스템	ark / arxiv / bibcode / doi / ean13 / eissn / handle / igsn / isbn / issn / istc / lissn / lsid / pmid / purl / upc / url / urn / other	lsid, doi
	support versioning	버전 관리 지원 여부	yes / no / unknown	yes
Funding	funding status	프로젝트 생명주기의 단계 표현	planned / applied / granted / rejected	granted

장벽이 없다고 가정하는 DMP 작성에 대한 개방형 플랫폼을 제공하기 위하여 OpenAIRE(Open Access Infrastructure for Research in Europe) (KOREA-EU RESEARCH CENTRE, 2018)⁴⁾와 EUDAT(European Commission, 2021)⁵⁾의 공동 노력으로 만들어진 온라인 도구다. Argos OpenDMP는 독립형 서비스(OpenDMP) 및 OpenAIRE 서비스(ARGOS)로 사용 가능한 확장 가능 개방형 서비스로 DMP의 관리, 검증,

모니터링 및 유지 관리를 간소화할 수 있으며 이를 통해 행위자(연구원, 관리자, 감독자 등)는 데이터 소유자의 의도와 약속에 따라 데이터 관리 프로세스의 특정 측면을 수행하기 위해 인프라 간에 자유롭게 교환할 수 있는 실행 가능한 DMP를 만들 수 있다(Argos, n.d.).

Argos의 경우 RDA DMP 공통 표준을 준수하고 있어 생산된 DMP가 다른 호환 DMP 플랫폼에서 활용 가능하고 반대의 경우도 마찬가지로

4) OpenAIRE는 오픈 사이언스 클라우드를 위한 범유럽 네트워크 구축과 혁신적 디지털 기반시설 건설을 목표로 유럽집행위원회의 지원을 받아 진행되는 프로젝트이다.
 5) EUDAT는 연구를 위한 데이터 서비스의 생성을 목적으로 하는 EU 프로젝트로, 특히 범유럽의 연구 협력을 위한 것이다. 또한 미래 연구자의 요구를 지속 가능한 방식으로 충족할 수 있는 역량과 능력을 갖춘 CDI(Collaborative Data Infrastructure)를 제공하는 3년 프로젝트이다.

리는 장점이 있다. 그러나 Argos의 모델은 다른 DMP 템플릿과는 달리 DMP와 Dataset이라는 두 가지의 Entity로 구성되어 있다. DMP 엔티티는 연구 수명 동안 수집, 생산 또는 재사용되는 데이터 세트의 관리, 처리 및 큐레이션에 대한 심층적인 정보와 함께 생성된 연구 프로젝트에 대한 중요한 정보를 포함한다. 데이터셋 엔티티에는 연구데이터 수명 주기 동안 데이터가 처리, 관리 및 큐레이션된 방식에 대한 설명 작성을 지원하는 템플릿의 질문이 들어 있다. Argos의 DMP는 하나 이상의 데이터셋으로 구성될 수 있다. 이러한 방식으로 데이터셋은 데이터셋 또는 연구 커뮤니티의 유형별로 각기 다른 템플릿을 따라 별도로 설명할 수 있는 유연성을

제공하며, 여러 DMP에서 복사하여 사용할 수도 있다. 그런 다음 데이터셋은 DMP에 번들로 포함되어 더 광범위하게 공유될 수 있다. 재사용되고 있는 데이터의 취급에 각별한 주의가 요구된다. 다음의 <표 5>는 Argos의 DMP 구성 항목을 나타낸 것이다.

Argos의 DMP는 크게 DMP 타이틀, DMP의 간략한 설명, 언어, 공개방법, 연구자, 생성 조직 마지막으로 DMP 소유자 등으로 구성된 메인 정보와 펀딩 기관의 정보, 보조금 번호 등이 포함된 펀딩 정보, 라이선스 목록을 나타내는 라이선스 정보 그리고 데이터셋 템플릿을 선택하여 기술하는 데이터셋 정보로 구성되어 있다. 이 중에서 통제어휘로 기술이 가능한 항목은 언어

<표 5> Argos DMP 구성

Argos DMP 엔티티		설명	통제어휘값	예시
Main info	Title of DMP	• DMP의 제목		
	Description	• DMP에 대한 간략한 설명, 범위 및 목표		
	Language	• DMP를 작성한 언어	• ISO language name	Korean
	Visibility	• Argos에서 DMP가 표시되는 방식. 'Public'을 선택하면 "Public DMP" 컬렉션의 모든 사용자가 DMP를 자동으로 이용할 수 있음	• Public • Restricted	Public
	Researchers	• DMP에 기술된 데이터를 생산, 처리, 분석한 사람들	• ORCID	0000-0002-8726-6367
	Organizations	• DMP 생성 및 개정에 기여한 조직 명칭 • OpenAIRE에 등록된 기관 선택 가능		
Contact	• DMP 소유자의 연락처 정보			
Funding info	Funding organizations	• Argos에서 펀딩 기관의 이름을 찾을 수 없는 경우, 이용자는 펀딩 기관의 이름과 세부 정보로 새 레코드를 생성할 수 있음 • OpenAIRE에 등록된 펀딩 기관 선택 가능	• Academy of Finland 등 24개 기관	Academy of Finland
	Grants	• 해당 펀딩 기관의 연구 프로젝트와 관련된 보조금 선택 • Argos에서 보조금을 찾을 수 없는 경우 사용자는 그랜트의 번호와 이름으로 새 레코드를 생성할 수 있음		Computational Nanotechnology (118377)
	Project	• 여러 보조금(grants)이 적용되는 프로젝트일 경우에 작성		
License Info	License	• DMP에 할당할 라이선스 목록 중 선택		Creative Commons Attribution 4.0
Dataset Info		• Argos에 등록된 Dataset 템플릿을 선택하여 해당 Dataset에 대한 정보를 기술	• Academy of Finland 등 17개 템플릿	Horizon 2020

(ISO language name), 공개방법, ORCID⁶⁾와 같은 식별자를 사용하는 연구자, OpenAIRE에 등록된 펀딩 기관을 선택할 수 있는 펀딩 기관 마지막으로 Argos에 등록한 데이터셋 템플릿을 선택할 수 있는 데이터셋 정보 등 5개이다. 다음의 <표 6>은 펀딩 기관과 데이터셋 템플릿

에 대한 목록을 정리한 것이다.

적합한 펀딩기관이 없는 경우, 식별자와 펀더명을 수동으로 입력할 수 있으며 데이터셋도 커뮤니티 또는 요구 사항에 맞는 맞춤형 템플릿을 Argos와 협의 후 등록할 수 있도록 되어 있다.

<표 6> Argos의 펀딩 기관 및 데이터셋 템플릿

펀딩 기관	데이터셋 템플릿
<ul style="list-style-type: none"> • Academy of Finland • Australian Research Council • Austrian Science Fund • Canadian Institutes of Health Research • Croatian Science Foundation • European Commission • European Environment Agency • French National Research Agency • Fundação para a Ciência e a Tecnologia, I.P. • Ministry of Education, Science and Technological Development of Republic of Serbia • Ministry of Science, Education and Sports of the Republic of Croatia(MSES) • National Health and Medical Research Council (NHMRC) • National Institutes of Health • National Science Foundation • Natural Sciences and Engineering Research Council of Canada • Netherlands Organisation for Scientific Research (NWO) • Research Council UK • Science Foundation Ireland • Social Sciences and Humanities Research Council • Swiss National Science Foundation • Tara Expeditions Foundation • Türkiye Bilimsel ve Teknolojik Araştırma Kurumu • UK Research and Innovation • Wellcome Trust 	<ul style="list-style-type: none"> • Academy Of Finland • ARIADNEPlus Horizon 2020 • CHIST-ERA • Clarin-D • FWF-Austian Science Fund • Horizon 2020 • IOSSG • National Science Center Poland • Netherlands Organisation For Scientific Research • Részletes kutatási adatkezelési terv • Rövid kutatási adatkezelési terv • Science Europe • Serbian Science Fund • Swedish Research Council • The Research Council Of Norway • TÜBİTAK DMP template • UK Research Innovation

6) ORCID(Open Researcher and Contributor ID)는 학술 커뮤니케이션에서 저자 이름 모호성 문제를 해결하기 위하여 개별 연구자에게 고유 식별자를 부여하는 체계

Argos는 일반적으로 전체 DMP 수명 주기⁷⁾ 따른다. 이용자는 DMP를 만들고 데이터셋에 대한 설명을 추가할 수 있다. DMP 및 데이터셋은 기본적으로 비공개 모드로 생성되며 동료와 공유하여 작성 프로세스를 용이하게 할 수 있다. 또한 데이터셋은 관련 유형 및/또는 학문 분야별로 DMP에 추가된다. 생성된 DMP 및 해당 데이터셋은 초안 상태이며 Argos 환경에서는 살아있는 문서로 취급된다. 즉, 언제든지 버전을 관리하고 업데이트할 수 있으며, 준비가 되면 이용자는 해당 데이터셋을 포함하여 DMP를 검증 및 완료하고 가시성 상태를 비공개에서 공개로 변경할 수 있다. Argos의 개방형 DMP는 공개 대시보드에서 모든 사람이 사용할 수 있고 액세스할 수 있다. DMP 수명 주기를 종료하기 위해 Argos는 Zenodo를 통합하여 이용자가 직접 DMP를 게시할 수 있도록 한다.

3.3 Haplo repository maDMP

Haplo Repository는 maDMP에 대한 RDCS를 구현한 세계 최초의 Repository 제공업체이다. RDA Helsinki 및 IDCC20의 연구 커뮤니티에 maDMP 프로토타입을 발표하였으며 이후 RDA 및 이용자의 피드백을 수용하여 2020년 4월에 London South Bank 대학에 Haplo maDMP 모듈을 오픈 소스로 출시하였다. 현재 Haplo Repository는 온라인 데모 출시를 준비

중이다. Haplo maDMP 모듈의 기능은 다음과 같다(Renner, 2020).

- 독립적으로 사용될 경우, 공통 데이터 교환 포맷 지원을 통해 타 기관 시스템과 공유 가능
- 단일 통합 시스템에서 전체 연구 수명 주기 동안 작업 및 프로세스 정보 제공
- RDCS 기반 확장 가능한 DMP 양식 사용, 개방형 커뮤니티 표준의 이점 유지 및 기관별 필드 활성화
- 기관에서 Haplo의 윤리, 펀딩 및 Repository 모듈을 사용 중인 경우, 이를 바탕으로 DMP 정보의 일부를 자동 기입
- DMP 템플릿의 버전 제어 및 변경 추적
- 프로비저닝(provisioning) 계획을 위해 대량의 데이터가 생성될 경우, Repository 직원에게 자동 알림
- Repository 보관기간 동안 데이터셋 접근 수준을 DMP로 확인 가능
- 기관 DMP 범위 및 내용 포괄적 제공

특히, Haplo Repository는 RDCS를 이용하여 Haplo가 개발한 모든 모듈을 하나의 어플리케이션 내에서 공존이 가능하도록 하였다. 또한 앞으로 Haplo Repository는 연구 관리 작업을 자동화하기 위한 시스템 내 알림 추가, 사용 가능한 보고 확장 그리고 전체 연구 생태계에서 DMP의 가치를 극대화하기 위한 외부 시스

7) Argos의 DMP 수명 주기는 생성 및 추가(Create & Add), 검증(Validate), 마무리(Finalize)로 정리된다. 생성 및 추가 과정은 DMP와 데이터셋 엔티티를 작성하는 것이고, 검증에서는 필수 필드 작성 여부를 확인한다. 마지막으로 마무리는 DMP 초안에 대한 잠금 과정 그리고 Argos 플랫폼이나 DOI가 있는 Zenodo에 게시하는 프로세스를 구성된다. 특히, 민감한 콘텐츠가 있는 경우, DMP를 게시하기 전에 중요한 데이터셋을 핵심에서 분리할 수 있게 하여 보호할 수 있도록 하는 기능을 제공한다.

템과의 통합을 준비하고 있다.

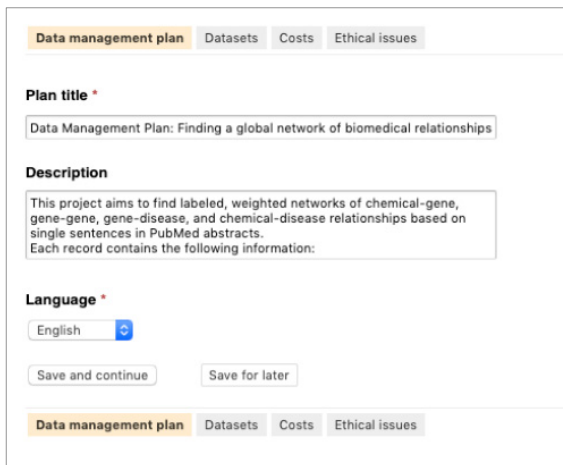
아직은 프로토타입인 Haplo repository maDMP는 크게 DMP, 데이터셋, 비용 그리고 윤리 이슈 등 4개의 카테고리로 구성되어 있다.

DMP 카테고리에서는 DMP 타이틀, 간략 설명(description) 그리고 언어에 대한 세부 내용을 입력하고 데이터셋에서는 데이터셋 타이틀, 데이터셋 간략 설명, 데이터셋 유형, 데이터의 예상 크기 마지막으로 데이터셋을 기술하기 위한 키워드를 등록할 수 있다. 다음으로 비용에서는 비용 타이틀, 간략 설명 그리고 비용과 통화단위를 입력할 수 있으며 마지막으로 윤리 이슈에서는 데이터와 관련된 윤리 이슈 여부를 확인하고 윤리 이슈보고서의 위치(URL 등록)와 간략한 기술을 등록할 수 있도록 하였다. Haplo Repository maDMP는 프로젝트별로 DMP와 연동이 가능하게 하여 생성된 날짜, 수정된 날짜, 대규모 파일, 민감 데이터 예상, 민감 데이터 보안, 예상되는 데이터셋 개수 마지막으로

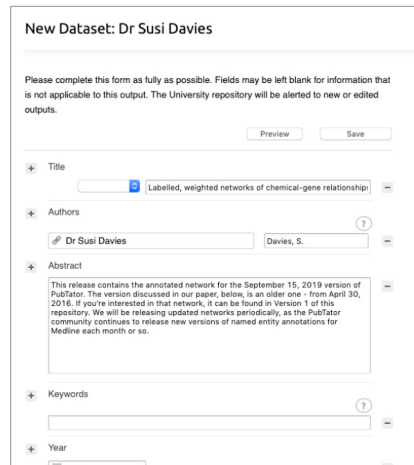
데이터셋이 DMP와 연결되어 있는지를 확인할 수 있다.

3.4 The Machine-actionable Data Management Planning application

The Machine-actionable Data Management Planning application(이하 DMap)은 Simon Oblasser(2020)가 워크플로우 및 시스템 간 정보 자동 교환을 제공하기 위하여 개발한 maDMP 어플리케이션이다. RDCS를 최초로 지원하는 도구이며, 그래픽 모형을 통한 이해 관계자 피드백을 반영하고 있지만 현재는 프로토타입 상태이다(Github, n.d.). DMap의 DMP 기술 과정은 프로젝트 선택, 데이터 관리 관련 연구자 선택, 실제 연구데이터 업로드 및 기술, 데이터 품질 및 문서화 기술, 법과 윤리적 측면 기술, 라이선스 기술, Repository 기술, 내보내기 등 총 8개의 카테고리로 구성되어 있다.



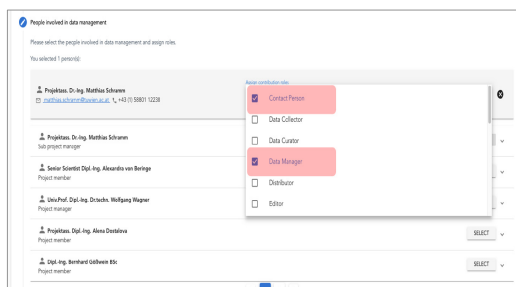
<그림 2> Haplo maDMP 모듈 - DMP 작성 화면



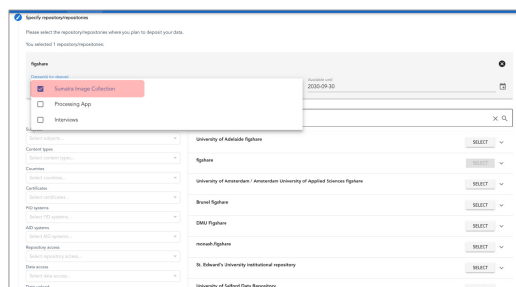
<그림 3> Haplo maDMP 모듈 - 데이터셋 세부 내용 작성 화면

먼저, 영구 식별자(Persistent Identifier, 이하 PID) 서비스와 연동하여 프로젝트를 검색한 후 해당 DMP와 관련된 프로젝트를 선택할 수 있다. 다음으로 연구자명을 검색할 수 있는 기능이 제공되어 검색 결과에서 해당하는 연구자를 입력할 수 있다. 연구자는 연구자 역할⁸⁾을 선택하여 기술할 수 있다. 일반적으로 연구자 정보에는 연구자명, 이메일 주소 그리고 연락처 정보가 제공된다. 다음으로 연구데이터 기술 단계에서는 데이터셋 타이틀을 입력한 후, 데이터셋 포맷과 데이터셋 크기를 등록한다. 실제 연구데이터 생성 완료 단계에선 샘플 데이터 업로드, 데이터 포맷 및 크기 그리고 개수에 대한 자동 입력이 가능하며, 연구자 재량으로 포맷 및 크기 그리고 개수는 수정이 가능하도록 구현되었다. 데이터 품질 및 문서화 기술 단계에서는 사용 메타데이터, 데이터 생성 방법 및 메소드, 구조화 및 버전 관리 계획 그리고 이용 대상을 자유 텍스트로 기술한다. 법과 윤리적 측면 기술 단계에서는 개인 정보 포함 여부, 민감 데이터 여부, 데이터 처리 및 공유 시 기타

법적 제한 존재 여부, 윤리적 이슈 여부를 통제 어휘(Yes/No)로 기술한다. Yes에 해당하는 경우, 상단에 기술한 데이터셋 중 해당하는 데이터셋을 선택하여 기술할 수 있다. 라이선스 기술 단계에서는 데이터셋에 해당하는 라이선스를 선택한 후 라이선스 종료일을 날짜 형식으로 기술할 수 있도록 되어 있다. 다음으로 Repository 기술 단계에서는 Repository 검색 후 해당하는 Repository를 선택하여 이용 기간을 Date 형식으로 기술할 수 있으며, 각 Repository마다 포함되어있는 프로젝트 내 데이터셋을 선택할 수 있다. Repository 정보를 통해 리포지토리의 URL, 유형, 인증 여부, 사용 PID 시스템, 버전 관리 여부, Repository 접근성, 데이터 접근성, 데이터 업로드, 데이터 라이선스, 사용 메타데이터 표준, 사용 가능한 데이터 포맷 그리고 주제 정보를 제공하여 리포지토리 검색과 브라우징이 가능하다. 마지막으로 내보내기 단계에서는 내보낼 DMP를 선택하여 다운로드가 가능하도록 구현되었다.



〈그림 4〉 DMap 데이터 관리 관련 연구자 선택 화면



〈그림 5〉 DMap Repository 기술 화면

8) 담당자, 데이터 수집자, 데이터 큐레이터, 데이터 매니저, 배포자, 편집자

4. maDMP 적용방안

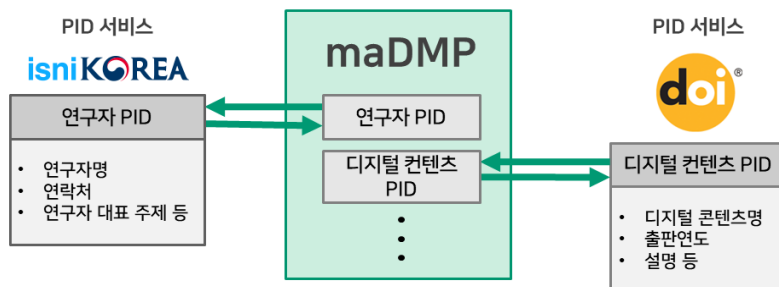
이번 장에서는 2장과 3장에서 조사 및 분석한 maDMP와 해당 구현사례를 종합하여 적용 가능한 방안을 논의하고자 한다.

기존 DMP를 maDMP로 구현하기 위한 방안으로서 PID 시스템의 연동과 통제어휘 및 온톨로지를 들 수 있다. 일반적으로 maDMP는 PID와 PID 서비스를 이용하여 DMP와 관련된 모든 리소스를 모아 연결한 것이다(DataCite Metadata Working Group, 2021). 따라서 본 연구에서는 다음 <그림 6>과 같이 maDMP를 구현하기 위한 개념적인 모형을 PID 시스템을 연계하여 표현하였다.

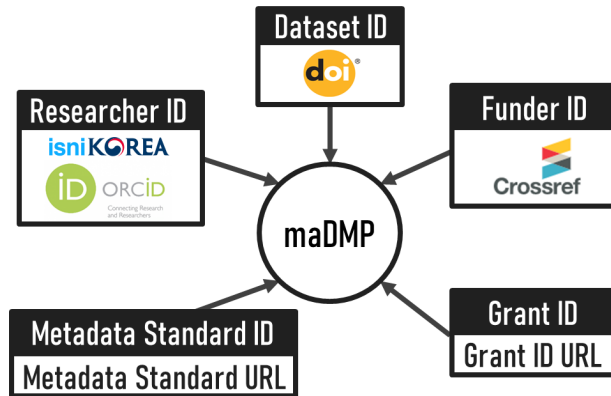
<그림 6>은 PID 서비스를 이용하여 maDMP를 구현한 개념적 모형이다. 먼저, maDMP에서는 DMP마다 연구자명, 연구자 연락처를 기술하지 않아도, 연구자 PID만을 기술하면 연구자명, 연락처 등은 PID 서비스를 통해 자동으로 가져올 수 있다. 즉, maDMP 내에 '연구자 PID'라는 요소가 있다면 이를 PID 서비스인 isniKOREA와 연동할 수 있으며, isniKOREA가 갖고 있는 연구자 PID 내 정보(연구자명, 연락처, 연구자 대표 주제 등)와의 자동 연계가

가능하다. 또한, 연구자 정보가 수정된 경우에도 maDMP에서 변경할 필요없이 자동으로 수정된다. 마찬가지로 연구자의 디지털 콘텐츠에 대해서도 대표적인 PID 서비스인 DOI를 maDMP에 적용하면 해당 콘텐츠명, 출판연도 등에 대한 정보를 자동으로 연계할 수 있다. 이러한 PID 서비스는 다양한 PID 서비스와 연동 가능한데 예를 들어 다음의 <그림 7>과 같이 연구자 ID, 데이터셋 ID, 펀더 ID 그리고 보조금 ID 등을 maDMP 구현하는데 활용할 수 있을 것이다.

<그림 7>에서와 같이 PID에는 ISNI와 ORCID를 활용한 연구자 ID, DOI가 부여된 데이터셋 ID, Crossref에 등록되어 있는 펀더 ID 그리고 각 기관별 보조금 ID 등이 포함된다. 또한 메타데이터 표준 ID로는 메타데이터 표준 URL을 적용할 수 있다. 특히 연구자 ID의 경우 VIAF (Virtual International Authority File), Publons 그리고 Scopus author-id 등도 등록이 가능하도록 PID 연계 서비스를 확대해야 할 것이다. 마찬가지로 펀더 ID에도 Crossref 등록 기관뿐만 아니라 GRID(Global Research Identifier Database), ROR(Research Organization Registry) 그리고 Ringgold 등 다양한 기관 식별자도 추가할 수 있도록 서비스 연계가 확대되어야 한다.



<그림 6> maDMP 구현을 위한 개념적 모형



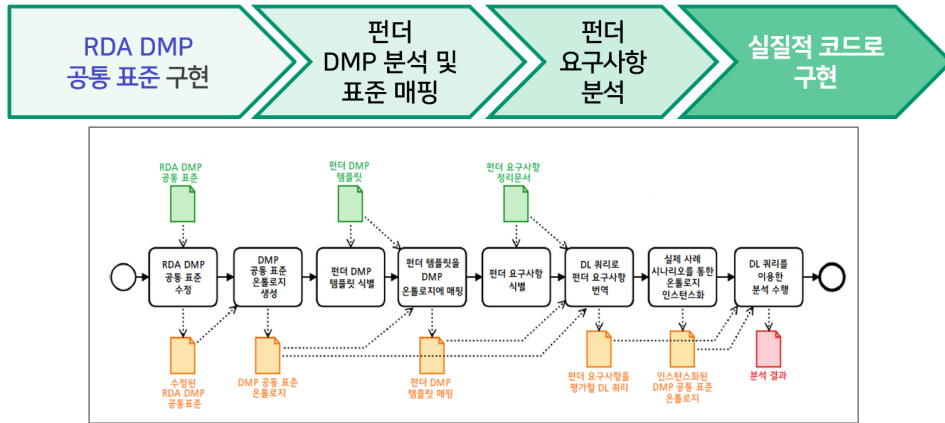
〈그림 7〉 PID 서비스를 활용한 maDMP 구현 방안

다음으로 maDMP를 구현하기 위한 구체적인 방안으로서 ‘통제어휘’와 ‘온톨로지’를 제시할 수 있다. 먼저, 통제어휘의 경우, 선별된 단어만 한정하여 사용함으로써 상호운용성을 보장할 수 있다. RDCS의 경우, PID 서비스 연계를 제외하고 전체 78개의 하위요소 중 12개 요소에서 통제어휘로 허용값을 적용할 수 있다. 대표적인 통제어휘로는 ‘ISO 4217’, ‘ISO 639-3’ 그리고 ‘ISO 3166-1’처럼 국제표준을 이용하여 기술할 수 있지만 포함 여부나 존재 여부를 확인하기 위한 허용값으로 ‘yes / no / unknown’과 같은 자연어도 통제어휘로 허용하고 있다. 마지막으로 ‘온톨로지’와 같은 시멘틱 기술의 적용이다. 다음의 〈그림 8〉은 온톨로지(Sheth, Ramakrishnan, 2003)⁹⁾로 공식화된 DMP가 특정 조직에 대한 펀더의 요구 사항 평가를 자동화하기 위해 어떻게 사용될 수 있는지를 나타낸 것이다(Cardoso, Proença, & Borbinha, 2020). maDMP 생성을 위해 위의 〈그림 8〉과 같이

분석할 수 있는 8가지 작업으로 구성되며, 세 가지 주요 부분으로 해석될 수 있다. 첫 번째 부분은 DMP 공통 표준 모델을 준수하는 RDCS 공통 표준으로 온톨로지를 생성한 후, 펀더 템플릿을 생성한 온톨로지에 반영한다. 두 번째 부분은 온톨로지와 식별된 DMP 템플릿 간의 필요한 매핑 수집과 펀더의 요구 사항에 기반한 DL(Data Language, 이하 DL) 쿼리 생성의 결과로 이어진다. 마지막 부분은 마지막 두 작업으로 구성되며, 이후에 이전에 생성된 DL 쿼리를 사용하여 분석되는 온톨로지가 생성된다. 이 프로세스의 핵심은 DMP가 기계 작동 가능하고 DMP 공통 표준 모델을 준수하며 의미 기술의 사용을 통해 표현되도록 하는 것에 있다.

종합하자면 maDMP를 적용하기 위해서는 영구 식별자의 사용, 통제어휘 적용 그리고 온톨로지와 같은 시멘틱 기술의 적용으로 요약할 수 있을 것이다.

9) 사람의 마음 속에 존재하는 내재적 생각이나 외재적 세계의 현상에 대하여 공유하는 개념을 컴퓨터가 이해할 수 있는 형식으로 명확하고 명시적으로 정의하고 규정하는 것으로 최근에는 시멘틱 기술로서 시스템 상호운용성을 증진하고 웹 인프라를 향상시키기 위한 정보의 의미론 활용에 초점을 맞추고 있다.



〈그림 8〉 maDMP 생성 프로세스

5. 결론

최근 오픈 사이언스에 대한 유네스코 권고안 이 회원국들에 의해서 유네스코 과학위원회에서 만장일치로 채택되었다. 이러한 맥락에서, 국제과학위원회는 과학의 세계적 목소리로서 오픈 사이언스를 옹호하고 발전시키는 것이 과학의 세계적 공공재라는 위원회의 비전을 달성하기 위한 기본 작업을 재확인하였다. 이러한 오픈 사이언스를 구체적으로 실행하기 위한 방안으로 DMP가 해외를 중심으로 작성되기 시작했다. 이러한 DMP는 해외의 주요 연구기관을 중심으로 의무화가 되기 시작했으며 국내에서도 법령에 DMP 도입 및 정보등록의 근거를 신설하여 DMP 작성 및 제출이 의무화는 추세로 접어들었다. 그러나 이러한 DMP가 자유 텍스트로 기록되거나 미기입하는 경우가 발생하고 기관 준수 요건을 형식적으로 준수하는 경우로 귀결되다 보니 생산하고자 하는 연구데이터를 제대로 설명하지 못하거나 정부 또는 편더가 갖는 실효성이나 효과성이 떨어지는 것

으로 나타났다. 따라서, 자유 텍스트로 기술되는 기존 DMP를 시스템이 이해할 수 있도록 하는 기계가독형 DMP에 대한 연구가 진행되고 있으며 일부 구현이 되어 이용 촉진과 발달이 제안되고 있다.

본 연구에서는 국내에서는 생소한 maDMP에 대한 현황을 알아보고 해당 구현 사례와 적용방안에 대해서 연구를 진행하였다. 특히, RDA DMP 공통 표준인 RDCS를 분석하여 그 적용가능성을 확인하였다. 이러한 RDCS는 maDMP를 생성하거나 사용하는 시스템 간의 기본 상호 운용성을 제공하기 위한 메타데이터 응용 프로파일이다. 여러 DMP를 비교하여 공통적인 요소를 추출한 후 이 요소들을 메타데이터 응용 프로파일로 개발한 것이다. 이외에도 Argos, Haplo repository 그리고 DMap에 대해서도 살펴보았다. Haplo repository와 DMap은 아직 프로토타입 버전이지만 향후 국내에서 maDMP를 구현할 때 시사점을 얻을 수 있도록 분석대상에 포함하였다.

maDMP를 적용하기 위한 방안으로서 본 연

구에서는 2가지를 제안하였다. 먼저 PID 서비스와의 연계이다. DMP를 기계가독형으로 구현하기 위해서는 PID 서비스와 같은 영구 식별자를 반드시 활용해야 한다. PID 서비스로 제안할 수 있는 것은 ORCID 등을 활용한 연구자 ID, DOI를 활용한 데이터셋 ID 그리고 Crossref에 등록되어 있는 펀더 ID 등이 포함될 수 있다. 다음으로 기계가 인식하기 위한 수단으로서 통제어휘와 온톨로지의 활용이다. 통제어휘는 선별된 단어만 한정하여 사용할 수 있도록 하여 상호운용성을 확보할 수 있으며 온톨로지를 통해 maDMP 속성들의 관계를 자연어가 아닌 프로그래밍 언어 형태로 기술하여 시멘틱 기술

을 적용할 수 있을 것이다.

이번 연구는 maDMP 적용방안에 대한 개념적 연구라 구체적인 방안에 대한 논의로서는 부족하다고 판단된다. 그러나 국내에서는 생소한 개념으로서 maDMP에 대한 현황과 개념적인 적용방안에 대해서 논의했다는 것만으로 그 의의를 가질 수 있을 것이다. 향후, 기계가독형 DMP 템플릿을 개발하려는 연구가 진행될 것이다. 따라서 본 연구가 maDMP 개발을 위한 선행연구로서 활용될 수 있기를 기대하며 또한 maDMP를 구현하기 위한 기술적인 논의가 구체적으로 이루어지기를 바란다.

참 고 문 헌

- 심희원 (2018.10.12.). OpenAIRE(Open access infrastructure for research in Europe). KOREA-EU RESEARCH CENTRE. Available:
<https://k-erc.eu/openaireopen-access-infrastructure-for-research-in-europe/>
- Argos [n.d.]. How it works. Available:
<https://argos.openaire.eu/splash/about/how-it-works.html>
- Bakos, A., Miksa, T., & Rauber, A. (2018). Research Data Preservation Using Process Engines and Machine-Actionable Data Management Plans. Proceedings of 22nd International Conference on Theory and Practice of Digital Libraries.
- Cardoso, J., Proença, D., & Borbinha, J. (2020). Machine-actionable data management plans: a knowledge retrieval approach to automate the assessment of funders' requirements. *Advances in Information Retrieval*, 12036, 118-125.
- DataCite Metadata Working Group (2021, March 30). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs Version 4.4. DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>
- DMP Tool Blog (2021, July 21). What's new with our machine actionable DMP work?. Available:

- <https://blog.dmpool.org/category/madmps/>
- European Commission (2020). EDUDAT. Available:
https://ec.europa.eu/eurostat/cros/content/eudat_en
- Github [n.d.]. DMap. Available: <https://github.com/oblacters/dmap>
- GO FAIR (2016). FAIR Principles. Available: <https://www.go-fair.org/fair-principles/>
- Harvard Medical School (2021). What is a Data Management Plan? Available:
<https://datamanagement.hms.harvard.edu/plan/data-management-plans>
- Jones, S. (2011). How to Develop a Data Management and Sharing Plan: DCC How-to Guides. Available: <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>
- Kim, S. (2020). Machine-actionable data management plans model analysis and improvement direction. *Journal of Information Science Theory and Practice*, 8(4), 20-28.
<https://doi.org/10.1633/JISTaP.2020.8.4.2>
- Loon, V. J., Akers, K. G., Hudson, C., & Sarkozy, A. (2017). Quality evaluation of data management plans at a research university. *International Federation of Library Associations and Institutions*, 43(1), 98-104. <https://doi.org/10.1177/0340035216682041>
- Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLoS Computational Biology*, 11(10), e1004525. Available:
<https://doi.org/10.1371/journal.pcbi.1004525>
- Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLoS Computational Biology*, 15(3), e1006750.
<https://doi.org/10.1371/journal.pcbi.1006750>
- Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M., Cardoso, J., Kvamme, T., Praetzelis, M., Suchánek, M., Hooft, Rob., Faure, B., Moa, H., Hasan, A., & Jones, S. (2021). Application profile for machine-actionable data management plans. *Data Science Journal*, 20(1), 32. <http://doi.org/10.5334/dsj-2021-032>
- Renner, T. (2020, May 22). Haplo Repository breaks new ground with Machine Actionable Data Management Plans. Haplo. Available:
<https://www.haplo.com/news/haplo-repository-dmp>
- Simms, S., Jones, S., Mietchen, D., & Miksa, T. (2017). Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3, e13086.
<https://doi.org/10.3897/rio.3.e13086>
- Simms, S., Strong, M., Jones, S., & Ribeiro, M. (2016). The future of data management planning: tools, policies, and players. *International Journal of Digital Curation*, 11(1).

<https://doi.org/10.2218/ijdc.v11i1.413>

Smale, N., Unsworth, K., Denyer, G., & Barr, D. (2018, October 17). The History, Advocacy and Efficacy of Data Management Plans. bioRxiv. Available:

<https://doi.org/10.1101/443499>

Stanford Libraries [n.d.]. Data management plans. Available:

<https://library.stanford.edu/research/data-management-services/data-management-plans>

UNIVERSITY OF EDINBURGH [n.d.]. Writing a DMP. Available:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/before/writing-a-data-management-plan>

W3C (2020, February 4). Data Catalog Vocabulary (DCAT) - Version 2. Available:

<https://www.w3.org/TR/vocab-dcat-2/>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

Shim, H. (2018, October 12). OpenAIRE(Open access infrastructure for research in Europe). KOREA-EU RESEARCH CENTRE. Available:

<https://k-erc.eu/openaireopen-access-infrastructure-for-research-in-europe/>