

연구데이터 서비스의 유용성 평가 모형 연구*

A Study on Evaluation Model for Usability of Research Data Service

박진호 (Jin Ho Park)**

고영만 (Young Man Ko)***

김현수 (Hyun Soo Kim)****

초 록

본 연구의 목적은 연구데이터 서비스 자체의 유용성과 연구데이터에 대한 사용경험 기반의 유용성 측면에서 평가 모형을 개발하는 것이다. 다양한 사례에서 도출한 데이터 서비스의 유용성 평가 요소로부터 연구데이터에 내재된 평가척도인 검색성, 접근성, 상호운용성, 재활용성 4개와 각각의 측정지표 총 20개를 도출하였다. 그리고 Google Analytics, YouTube 광고료 책정 기준, 서울특별시, Altmetrics의 사례를 분석하여 연구데이터에 대한 이용자 경험 기반의 유용성 측정지표 12개를 도출하였다. 평가척도와 측정지표에 대한 타당성과 신뢰성 검정을 위해 연구데이터의 잠재적 이용자 164명을 대상으로 설문조사를 실시하였다. 평가척도의 타당성 검정을 위해 KMO Bartlett 분석을 하였으며, 측정지표의 성분분석을 위해 주성분 분석과 베리맥스 회전분석법을 사용하였다. 내재적 평가척도의 경우 4개 척도 모두 KMO Bartlett의 타당성 값을 충족시켰으며, 평가척도에 대한 측정지표의 성분분석 결과 모두 단일 성분으로 나타나 현재의 척도로 해당 지표에 대한 설명이 가능하였다. 그러나 이용자 경험 기반의 12개 측정지표의 성분분석 결과는 2개 성분으로 나누어지는 것으로 나타나 각각을 활용도와 참여도라는 개념의 2개 평가척도로 구분하였다. Cronbach's alpha 계수에 의한 신뢰도 측정 결과 6개의 평가척도 모두 0.6 이상의 측정치를 충족시키는 것으로 나타났다.

ABSTRACT

The Purpose of this study is to develop an evaluation model for usability of research data service from the angles of evaluating usefulness of research data service itself and research data use experience-based usability. First, the various cases of evaluating usability of data services are examined and 4 rating scales and 20 measuring indicators for research data service are derived as a result of comparative analysis. In order to verify validity and reliability of the rating scale and the measuring indicators, the study conducted a survey of 164 potential research data users. KMO Bartlett Analysis was performed for validity test, and Principle Component Analysis and Verimax Rotating Method were used for component analysis on measuring indicators. The result shows that the 4 intrinsic rating scales satisfy the validity criteria of KMO Barlett: A single component was determined from component analysis, which verifies the validity of measuring indicators of the current rating scale. However, the result of 12 user experience-based measuring indicators analysis identified 2 components that are each classified as rating scale of utilization level and that of participation level. Cronbach's alpha of all 6 rating scales was 0.6 or more for the overall scale.

키워드: 연구데이터, 데이터 품질, 사용성 평가, 평가척도, 측정지표

research data, data quality, usability evaluation, rating scale, measuring indicator

* 본 연구는 2019년 한국과학기술정보연구원(KISTI)의 위탁연구 과제로 수행한 “국가연구데이터플랫폼의 메타데이터 품질 및 유용성 평가모형 연구”의 일부분임.

** 주식회사 리스트 사업개발본부장(jinhopark.lis@gmail.com) (제1저자)

*** 성균관대학교 문과대학 문헌정보학과 교수(ymko@skku.edu) (교신저자)

**** 성균관대학교 정보관리연구소 연구원(lurgee81@skku.edu) (공동저자)

■ 논문접수일자: 2019년 11월 17일 ■ 최초심사일자: 2019년 12월 10일 ■ 게재확정일자: 2019년 12월 13일

■ 정보관리학회지, 36(4), 129-159, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.4.129>

1. 서론

1.1 연구 목적

최근 들어 학술연구 활동에 있어서 나타나고 있는 큰 변화 요구의 하나는 연구 과정과 결과물에 대한 개방화이다. 이 흐름을 대표하는 오픈 사이언스(open science)라는 용어는 오픈 액세스(open access), 오픈 데이터(open data), 개방형 학술 커뮤니케이션(open scholarly communication)을 포괄하는 개념이다. 이 중 오픈 액세스와 개방형 커뮤니케이션에 대해서 비교적 오랜 기간 논의가 이루어져 왔으나 연구데이터 개방은 그러지 못했다. 특히 국내의 경우 실제 구현된 연구데이터 서비스는 아직 없으며, 한국과학기술정보연구원(이하 KISTI)에서 2018년 12월 국가연구데이터플랫폼(Research Data Platform, 이하 RDP)을 구축하여 시범 서비스를 운영하고 있다. RDP는 “국가연구개발사업의 관리 등에 관한 규정(2019년 3월 19일 개정, 2019년 9월 1일 시행)”의 법적 근거를 구축한 것으로 공적 자금을 투입한 연구의 데이터 보존과 서비스, 활용을 위한 제반 서비스 제공을 목표로 한다.

본 연구는 KISTI의 RDP에서 제공하는 연구데이터와 그 서비스를 실험 대상으로 삼아 연구데이터 서비스의 유용성 평가 모형을 제시하는 것을 목적으로 한다. 연구데이터와 관련하여 국내에서는 연구데이터의 수집과 공유 및 활용 등에 관한 연구(김은정, 남태우, 2012; 박미영, 안인자, 남승주, 2018), 연구데이터의 관리 및 서비스를 위한 기능적 요소 연구(김주섭, 김선태, 최상기, 2019), 연구데이터에 대한 연

구자의 인식 연구(유사라, 2019), 연구데이터의 특성과 인용 경향 연구(조재인, 2016), 연구데이터의 재활용성과 상호운용성 가이드라인 연구(최리진, 정영미, 2019) 등이 수행되었다. 그러나 실제 시스템과 이용자가 존재하는 상태에서 가능한 연구데이터에 관한 유용성 평가 연구는 아직 이루어지지 않은 것으로 나타났다. 국제적으로는 2016년 G20 Hangzhou summit에서의 “FAIR principle for research”, Elsevier (2015)의 연구데이터 관리 프로세스 로드맵, Research Information Network(RIN)의 데이터셋 품질 보증 요소, 계량지표 기반의 연구성과 평가를 위한 Leiden Manifesto, OECD의 연구데이터 원칙 등이 있으나, 연구데이터에 대한 합의된 유용성 평가 모형은 부재한 것으로 나타났다.

따라서 본 연구에서는 RDP의 연구데이터 시범서비스와 잠재고객을 대상으로 두 가지 측면에서 유용성 평가 모형을 개발하고자 하였다. 하나는 연구데이터에 내재된 유용성 평가를 개발하는 것이며, 다른 하나는 이용자 경험을 기반으로 한 유용성 평가 모형을 개발하는 것이다. 연구데이터에 내재된 유용성 평가 모형을 개발하기 위해 데이터 서비스의 유용성 평가에 관한 다양한 사례로부터 평가척도와 그 측정지표를 추출하였으며, 이용자 경험 기반의 평가와 관련해서는 해당 평가척도가 부재하다는 점을 감안하여 Google Analytics, YouTube, Altmetrics, 서울특별시 등의 데이터 서비스로부터 관련 측정지표를 추출하여 연구데이터 적용의 적합성을 검토하였다. 잠재적 이용자를 대상으로 하는 설문조사를 통해 적합성 검토 결과 선정된 평가척도와 측정지표에 대하여 타당성

과 신뢰성 검정을 하였으며, 이를 토대로 연구데이터 서비스의 유용성 평가 모형을 최종적으로 제시하였다.

1.2 연구 방법

본 연구에서는 서비스 측면과 이용자 경험 측면을 고려하여 연구데이터의 유용성에 관한 평가 모형을 개발하였다. 연구데이터 서비스 측면에서의 유용성은 상호운용이나 재활용과 같이 서비스되는 연구데이터에 내재된 유용성을 말하며, 이용자 경험 기반의 유용성은 관심도나 열람수와 같은 연구데이터 이용자의 이용 패턴에 따른 유용성을 의미한다. 평가 모형은 평가척도와 측정지표로 구성되며, 선정된 평가척도와 측정지표에 대해 타당성과 신뢰성을 검정하여 본 연구의 연구데이터 유용성 평가 모형으로 제시하였다.

먼저 연구데이터의 유용성에 관한 합의된 국제적 평가 모형이 없는 점을 고려하여 국내 법률, EC(European Commission), EPSRC(Engineering and Physical Sciences Research Council), OECD, ISO의 연구데이터에 관한 정의를 검토하였으며, 이를 토대로 연구데이터 및 오픈데이터에 관한 기준들을 분석하여 연구데이터에 내재된 특성을 찾아내 평가척도의 기준으로 삼고자 하였다. 연구데이터의 유용성 평가척도와 측정지표 도출을 위해서는 FAIR Principle for Research, Elsevier(2015)의 연구데이터 관리 프로세스 로드맵, Research Information Network(RIN)의 데이터셋 품질 가이드라인, 계량지표 기반의 연구성과 평가를 위한 Leiden Manifesto, OECD의 연구데이터 품질 가이드라인, Obama

행정부의 오픈 데이터 선언, 서울특별시 오픈데이터 평가 모형을 비교하였다. 비교 분석에 의한 평가척도의 병합과 중복 제거 등의 정제를 거쳐 도출된 평가척도와 측정지표를 대상으로 타당성 검정과 신뢰성 분석을 실시하였다.

연구데이터에 관한 이용자 경험 기반의 유용성 평가 기준은 아직 마련되어 있지 않으며, 방문수, 열람수, 다운로드 수 등의 단순한 정량적 계수들이 일반적인 정보서비스 이용과 관련된 유용성 측정지표로 사용되고 있다. 본 연구에서는 보다 심층적인 지표를 도출하기 위해 Google Analytics, YouTube 광고료 책정 기준, 서울특별시, Altmetrics의 사례를 분석하였으며, 도출된 측정지표에 대하여 타당성을 검정하였다. 검정 결과 동일 성분으로 집합되는 측정지표들을 아우를 수 있는 용어로 개념화하였으며, 개념화한 용어를 평가척도 명칭으로 삼아 이에 대한 신뢰도를 검정하였다.

평가척도와 측정지표에 대한 타당성과 신뢰성 검정을 위해 연구데이터의 잠재적 이용자 164명을 대상으로 설문조사를 실시하였다. 평가척도에 대한 신뢰성은 chronbach's alpha 계수를 사용하여 검정하였으며, 사회과학 분야에서의 일반적 기준인 0.6 이상을 검정 기준치로 삼았다. 평가척도와 측정지표의 타당성 검정을 위해서는 KMO Bartlett 분석을 하였으며 유의도 수준은 $p \leq 0.05$ 이다.

그리고 평가척도에 대한 해당 측정지표의 보편성과 차별성을 탐색하기 위해 주성분 분석(Principle Component Analysis)과 베리맥스 회전분석법(Verimax Rotating Method)을 사용하여 요인분석을 실시했다. 요인분석에서의 고유값은 요인의 설명력을 말하는 것으로 일반

적으로 1 이상을 기준으로 하며, 수치가 높을수록 요인이 측정요소들의 분산을 잘 설명한다(강병서, 김계수, 2009; 송지준, 2008). 따라서 본 연구에서도 주성분분석을 통해 초기 고유값(eigen value)이 1 이상이고, 공통성이 0.4 이상인 지표를 추출하여 타당성을 검정하였으며, 베리맥스 회전법을 활용하여 요인적재량(factor loading)을 검토한 후, 요인적재량의 일반적 판단 기준인 0.3 이상을 기준으로 평가척도와 측정지표의 상관관계 정도를 파악하였다.

2. 연구데이터의 유용성 평가척도와 측정지표

2.1 연구데이터에 요구되는 특성

유용성의 한국어대사전 정의는 “쓸모가 있고 이용할 만한 특성”이며, 영역인 usability에 대한 Webster 사전의 정의는 “the quality or state of being usable”이다. 그러나 연구데이터에 관한 공유된 정의는 나와 있지 않으며(University of Leicester, 2012), 연구데이터의 유용성에 대한 정의 역시 모호하다(Abran, Khelifi, Suryan, & Seffah, 2003). 따라서 본 연구에서는 공신력 있는 기관의 연구데이터에 관한 정의, 연구데이터 또는 오픈데이터에 요구되는 기준들을 탐색하여 연구데이터의 “쓸모 있고 이용할만한 특성”과 관련된 항목을 확인하고, 이를 토대로 평가척도를 도출하고자 하였다. 이를 위해 먼저 EC (n.d), EPSRC(n.d), OECD(Pilat, Fukasaku, 2007), ISO/IEC 9126-1 2001 및 국내 법률의 연구데이터에 관한 정의를 분석하였

으며, 이들 정의에 포함된 연구데이터의 특성은 형식성(디지털 형식), 연결성(원 연구와의 연결), 공유성, 개방성, 사용용이성, 학습가능성인 것으로 나타났다.

연구데이터에 관한 기준으로는 FAIR Principle(Wilkinson et al., 2016), OECD의 연구데이터 품질 가이드라인(Pilat & Funkasaku, 2007), ISO-9126, Elsevier(Elsevier, 2015), RIN 가이드라인(Swan & Brown, 2008)을 조사하였으며, 오픈 데이터에 관한 기준으로 OKI(Dietrich, Gray, McNamara, Poikola, Pollock, Tait, & Zijlstra, 2009), Obama 행정부의 오픈 데이터 선언(Executive Office of the President of the United States, 2013), 서울특별시 오픈 데이터 평가모델(서울특별시 2015)을 검토하였다. 이들 기준으로부터 연구데이터와 관련된 특성을 탐색하여 확인하였으며, 영어 표현은 다르지만 한국어 표현이 동일한 경우 FAIR 원칙에서 사용하는 용어를 기준으로 정리하였다(〈표 1〉 참조). 분석 결과 연구데이터와 관련된 정의와 기준 모두에서 요구되는 특성으로는 검색성(findability), 연결성(원연구와의 연결/접근성(accessibility)), 공유성(shared)/상호운용성(interoperability), 개방성(openness)/활용성(availability), 사용용이성(ease of use)/재활용성(reusability), 학습가능성(learnability)/검토성(reviewability)으로 나타났다.

2.2 연구데이터의 내재적 유용성 평가척도와 측정지표

연구데이터에 요구되는 특성을 비교하여 중복되는 특성을 제거하고 특성 항목의 명칭을

<표 1> 연구데이터에 요구되는 특성

정의	출처	기준	출처	정의/사례 공통
형식성(디지털 형식인가) / format	EC, OECD, EPSRC, 국내법률 ISO/IEC 9126-1	검색성(findability)	FAIR, RIN, Elsevier	검색성(findability)
연결성(원연구와의 연결)		접근성(accessibility)	FAIR, Elsevier, RIN, OKI, Obama, 서울시	연결성(원연구와의 연결) / 접근성(accessibility)
공유성(shared)		상호운용성(interoperability)	FAIR, OECD	공유성(shared) / 상호운용성(interoperability)
개방성(openness)		개방성(openness)	OECD	개방성(openness) /
사용용이성(ease of use)		활용성(availability)	OKI, 서울시	활용성(availability)
학습가능성(learnability)		재활용성(reusability)	FAIR, RIN, OKI, Obama, 서울시	사용용이성(ease of use) / 재활용성(reusability)
		검토성(reviewability)	Elsevier	학습가능성(learnability) / 검토성(reviewability)
운영성(operability) 매력성(attractiveness) 준수성(compliance)	ISO/IEC 9126-1			
		재현가능성(reproducible) 저장성(storability) 보존성(preservability) 인용가능성(citability)	Elsevier	
		이해성(comprehensible)	Elsevier, ISO-9126	
		수집분석기법 공개 분석과정 공개	Leiden Manifesto	
		재배포 범용적 참여	OKI	
		참여와 소통, 관리체계	서울시	
		완전성	서울시, Obama	
		상세설명, 공공성, 적시성, 사후관리	Obama	
		유연성(flexibility) 투명성(transparenty) 법률준수(legal conformity) 지적재산권 보호(protection of intellectual property) 공식적 책임(formal responsibility) 전문성(professionalism) 품질(quality) 보안(security) 효율성(efficiency) 책임성(accountability) 지속가능성(sustainability)	OECD	

정제하는 절차를 거쳐 연구데이터의 내재적 유용성 평가척도 항목을 도출하였다. 특성의 명칭은 다르지만 측정방법 즉 측정지표가 동일할 경우 하나의 명칭으로 통합하고, 동일한 측정지표를 포함하고 있으나 특성 명칭이 다를 경우 FAIR Principle에서 제시하는 명칭을 사용하는 것을 기준으로 삼아 중복되는 특성을 통합하고 명칭을 정제하였다. 정제된 특성과 해당 지표에 대해서 연구데이터의 평가척도, 측정지표, 측정요소로서의 적절성 및 계량화 가능 여부,

유용성과의 관련 여부 등을 검토하였다. 본 연구에서 개발된 평가척도와 해당 측정지표는 RDP의 잠재적 이용자를 대상으로 하는 타당성과 신뢰성 검정을 충족시킬 경우 확정되므로, 검토 결과가 RDP 시스템에 적용시킬 경우 적합한지를 다시 검토하였으며([부록 1] 중복제거 상세 내역 참조), 검토 결과에 따라 검색성, 접근성, 상호운용성, 재활용성의 네 가지 특성과 해당 지표가 연구데이터의 내재적 평가척도와 측정지표로 선정되었다(〈표 2〉 참조).

〈표 2〉 연구데이터의 내재적 유용성 평가척도와 측정지표

평가척도	측정지표
검색성	<ul style="list-style-type: none"> • 정의: 연구데이터를 이용자가 잘 검색할 수 있는지를 측정 ✓ 측정지표1. 연구데이터 검색이 가능한 적합한 메타데이터를 사용하고 있는가? ✓ 측정지표2. 연구데이터에는 고유하고 영구적인 식별자를 할당하고 있는가? ✓ 측정지표3. 연구데이터 관리, 서비스에 적용한 메타데이터는 식별자 요소를 포함하고 있는가? ✓ 측정지표4. 연구데이터, 메타데이터에 대한 색인이 이루어지고 있는가?
접근성	<ul style="list-style-type: none"> • 정의: 연구데이터가 모든 이용자와 사용목적에 부합하는지를 평가하기 위한 척도 ✓ 측정지표1. 연구데이터에 자유로운 접근이 가능한가(별도의 로그인 등 제한이 없는지 확인)? ✓ 측정지표2. 연구데이터는 기계가독형인가? ✓ 측정지표3. 연구데이터는 개방형 포맷(open format)인가?
상호운용성	<ul style="list-style-type: none"> • 정의: 연구데이터는 글로벌 상호운용성을 확보하고 있는지 평가하기 위한 척도 ✓ 측정지표1. 해당 연구데이터, 메타데이터를 가지고 지식표현(knowledge representation)이 용이한 형태를 유지하고 있는가? ✓ 측정지표2. 관리, 서비스 대상 연구데이터와 메타데이터가 검색성(findability), 접근성(accessibility), 상호운용성(interoperability), 재활용성(reusability)을 충족하는가? ✓ 측정지표3. RDP 운영조직에 기술적 상호운용성 보장을 위한 조직이 존재하는가? 혹은 관련 연구를 지속적으로 수행하고 공표하고 있는가?
재활용성	<ul style="list-style-type: none"> • 정의: 연구데이터는 이용자들이 재활용하기 위한 기술적, 제도적 제한에서 자유로운지를 평가하기 위한 척도 ✓ 측정지표1. 연구데이터의 벌크 다운로드가 가능한가? ✓ 측정지표2. 연구데이터는 적시, 최신성을 유지하고 있는가? ✓ 측정지표3. 연구데이터가 무엇에 관한 것인지 상세설명을 하고 있는가? ✓ 측정지표4. 연구데이터의 출처(원 저작물, 원 연구자료)를 포함하고 있는가? ✓ 측정지표5. 원천데이터를 변환한 데이터라면 변환과정에 대한 상세설명을 포함하고 있는가? ✓ 측정지표6. 연구데이터를 기술하는 메타데이터는 풍부한 속성을 가지고 있는가? ✓ 측정지표7. 연구데이터를 기술하는 메타데이터는 해당 지식분야(커뮤니티)에서 활용하고 있는 표준 메타데이터인가? ✓ 측정지표8. 연구데이터는 해당 지식분야(커뮤니티)에서 일반적으로 활용하고 있는 형식(format)을 준수하고 있는가? ✓ 측정지표9. 연구데이터를 기술하는 메타데이터는 출처정보를 표현하기 위한 요소를 포함하고 있는가? ✓ 측정지표10. 연구데이터를 활용한 이용자의 의견(평가)을 수용할 수 있는 기술적장치(댓글 등)가 존재하는가? ✓ 측정지표11. 연구데이터에 오픈 라이선스를 부여했는가?(재활용에 비용을 소모하지 않는 오픈 라이선스)

2.3 연구데이터의 이용자 경험 기반 유용성 측정지표

본 연구에서는 이용자 측면의 유용성 측정을 위해 기본적으로 사용되는 방문수, 다운로드 수, 인용수와 같은 단순한 지표가 아니라 보다 심층적으로 이용자의 이용 경험을 토대로 하는 유용성을 측정할 수 있는 계량 지표를 개발하고자 하였다. 이를 위해 Google Analytics의 로그 관리시스템, Altmetrics의 대안적 연구성과 측정지표, YouTube의 동영상 광고 이용 측정지표, 서울시의 공공데이터 이용 측정 방식을 분석하였다. 서울시 측정 방식의 경우 계량적 지표를 이미 포함하고 있고, Altmetrics는

연구 성과 측정과 직접 관련이 있어 분석 대상으로 삼았다. Google Analytics는 RDP에서 이미 사용하고 있는 분석도구로 통계 결과를 보다 적극적으로 활용하기 위해, YouTube는 가장 대중적인 영상 검색 매체로 실제 사용자들의 활용 정도에 따라 광고료를 집행하므로 연구데이터의 실질 이용을 측정하는데 차용 가능할 것으로 판단하여 채택하였다.

그러나 이들의 측정지표는 연구데이터와는 상이한 이용자 및 매체적 특성에 기반을 둔 것이므로, 본 연구에서는 분석을 통해 확인된 지표가 연구데이터에 적용될 수 있는지 검토하였다(〈표 3〉 참조).

〈표 3〉 계량적 측정지표 도출 상세 내역

사례	측정요소	검토내용
서울시	<ul style="list-style-type: none"> 접근성의 방문율: 다양한 국가, 언어별로 접근이 가능한 지를 측정하기 위한 요소로 웹 브라우저, 언어, 국가/지역 별 접근 로그 통계를 확인하여 측정 활용성의 실질 이용률: 이용자들이 실제로 활용한 데이터의 빈도를 측정하기 위한 지표로 단순열람, 다운로드, API 호출의 연평균 증감율로 산정하며 산정식은 $\{(열람 연평균증감율 * 0.1) + (다운로드 연평균증감율 * 0.4) + (API호출연평균증감율 * 0.5)\}$임 재활용성의 재생산율: 첫째, 제공 데이터가 학술연구정보에 활용되었는가?, 둘째, 제공 데이터가 참고용으로(기사 작성, 블로그 등) 인용되거나 활용되었는가? 셋째, 활용데이터가 웹 서비스, 모바일 애플리케이션 등 시각적 확인이 가능한 응용서비스로 개발되었는가를 측정 재활용성의 재사용율: 첫째, 데이터셋 갱신 일자별 재사용률(다운로드율), 둘째, 한 번 활용한 데이터의 갱신이 이루어진 후 재 다운로드가 이루어지는지 측정 	<ul style="list-style-type: none"> 유용성 관점에서 단순 방문율이 아니라 유효한 방문인지 확인할 수 있는 방안과 해당 데이터가 계속 사용되고 있는지는 측정할 수 있는 제방문 측정이 가능한 요소 개발 필요 재생산 관점은 연구데이터 측면에서 보면 인용 정도를 측정하는 것이 가장 근사하며, 측정 시 단순 이용수 누적보다는 얼마나 인용되었는지를 측정할 수 있는 요소 개발 필요 간접적인 측정도 중요하지만 이용자가 직접 참여하여 유용성을 평가할 수 있는 지표 필요
Altmetrics	<ul style="list-style-type: none"> 소셜미디어를 활용하여 연구성과의 영향력을 '논문 단위(Article Level)'로 계량화하기 위한 평가지표임 인용정보를 기반으로 한 활용성 측정으로 불가능한 다양한 관점에서의 학술 논문 영향도를 측정하기 위한 것으로 학계가 아닌 일반 대중이 해당 논문을 추천, 저장, 활용하는 영향도를 측정하기 위한 것임 과학적 성과에 대한 대안적인 지표로 다양한 대중과 소통 가능한 플랫폼에 가중치를 두어 합산하는 방식을 취하고 있음 	<ul style="list-style-type: none"> Altmetrics가 기존의 영향력 평가 방식에 대한 문제점을 명확히 인식하고 있으며 다양한 정보 유통 플랫폼을 모두 고려하여 다양한 이용자에 대한 유용성 평가에 활용할 수 있는 도구임 단, 국내 이용자들 대상으로 할 경우에는 Altmetrics가 대상으로 하는 플랫폼이 적합한지에 대한 검토가 필요하며, 가중치에 대한 적합성 여부를 측정하기 위한 별도의 논의 필요 때문에 본 연구에서는 유용성 평가에 있어서 가중치를 부여하여 합산하여 측정방식과 소셜 영향력을 평가할 수 있는 지표 개발 필요

사례	측정요소		검토내용	
Google Analytics	•대분류 기준 Google Analytics 측정 요소		<ul style="list-style-type: none"> • Google Analytics에서 확인할 수 있는 점은 단순히 누적 수만을 중요하게 측정하지 않는다는 점, 즉 누적량도 중요하지만 전체 대비 해당 측정항목이 차지하는 비율과 평균을 중요하게 간주함 • 본 연구에서도 각 측정지표 개발 시 전체 대비 해당 지표가 차지하는 비율을 중요한 측정방식으로 고려 	
	구분	측정요소		
	방문	방문현황(UV)		사용자수
				신규사용자
				재사용자
		사이트 콘텐츠		평균페이지에 머문 시간
				이탈률
		방문페이지		세션 평균세션시간
	페이지	페이지뷰(PV)		페이지뷰수
				신규사용자
				재사용자
				세션
세션당 페이지수				
소셜	네트워크 추천	소셜 추천을 통한 세션수 전체 세션수		
	데이터 허브활동			
	트랙백			
검색	검색어	전체 순 검색량		
		검색당 결과 페이지 뷰 수		
		검색 후 이탈률		
		재검색율		
		검색 후 시간		
		평균 검색 심도		
YouTube 동영상 광고 측정항목	<ul style="list-style-type: none"> • 조회수: 사용자가 동영상 광고를 시청하거나 광고에 참여한 횟수 • 조회율: 광고에서 발생한 조회수 또는 참여수를 광고가 노출된 횟수(동영상 및 미리보기 이미지 노출수)로 나눈 값. • 조회당비용(CPV): 사용자가 동영상의 30초 지점까지(동영상 광고가 30초 미만인 경우 광고 전체를) 계속 봤을 때 또는 동영상과 상호작용할 때(둘 중 이른 시점 적용) 지불하는 평균 비용 • 시청시간: 사용자가 동영상 광고를 시청한 총 시간(초) • 평균시청시간/노출수: 사용자가 광고 노출 1회당 동영상 광고를 시청한 평균 시간(초) • 클릭 실적: 사용자가 동영상을 클릭한 횟수 • 클릭률(CTR): 광고에서 발생한 클릭수를 광고가 게재된 횟수로 나눈 값 • 참여수: 동영상의 카드를 확장하기 위해 티저나 아이콘과 같은 양방향 요소를 클릭한 횟수(외부로 연결되지 않음) • 참여율: 광고에서 발생한 참여수를 광고가 게재된 횟수로 나눈 값 • 순이용자: 일정 기간 동안 광고를 본 총 사용자 수 • 참여율: 광고에서 발생한 참여수를 광고가 게재된 횟수로 나눈 값 			
	<ul style="list-style-type: none"> • YouTube 동영상 광고료 책정 기준에서 도출할 수 있는 핵심적인 사항은 모든 계량적 측정지표가 단순 누적분이 아니라 비율을 함께 측정하여 실제로 광고 효과가 있었는지를 측정 • 요소 측면에서 참여수와 참여율은 직접 본 연구에 적용이 가능한 요소로 연구데이터 이용자가 직접 참여한 수와 비율을 측정하는 것이 가능 • 조회수, 클릭수는 본 연구에 있어서 다운로드, 참여, 열람 등 보다 다양한 방법으로 측정이 가능하여 제외 • 순이용자는 별도 개인정보 취득이 가능해야 하는 구조로 측정요소로 부적합하여 제외 			

이러한 과정을 통해 연구데이터의 이용자 경험 기반 유용성을 측정하기 위한 12개의 지표 (열람수, 다운로드수, 다운로드 비율, 인용수, 인용비율, 참여수, 참여율, 공유수, 공유율, 유효 방문률, 재방문수, 별점)가 도출되었다(〈표 4〉 참조).

〈표 4〉 연구데이터의 이용자 경험 기반 유용성 측정지표

측정지표	측정방식	출처
열람수	<ul style="list-style-type: none"> • 개별 연구데이터의 다운로드 및 이용이 가능한 랜딩페이지까지 달한 방문자 수를 측정 ✓ 개별 연구데이터의 다운로드 및 이용 가능한 페이지까지 도달한 방문자 수 	기본 요소 기본 항목
다운로드수	<ul style="list-style-type: none"> • 실제 활용하고 있는 연구데이터 숫자를 측정 ✓ 연구데이터를 다운로드한 수 	
다운로드 비율	<ul style="list-style-type: none"> • 실제 적극적으로 활용하고 있는 연구데이터의 비율을 측정 ✓ 연구데이터 다운로드 수 / 전체 연구데이터 수 	
인용수	<ul style="list-style-type: none"> • 데이터의 재활용 결과를 간접적으로 측정 ✓ 해당 연구데이터를 실제 논문, 프로젝트 등 학술 연구 활동 등에 활용한 수 	
인용비율	<ul style="list-style-type: none"> • 데이터의 재활용 결과를 간접적으로 측정 ✓ 다른 연구, 저작물에 활용된 수 / 전체 연구데이터 수 	YouTube
참여수	<ul style="list-style-type: none"> • active user를 간접적으로 측정 ✓ 측정방식: 댓글, 별점이 부여된 연구데이터 랜딩페이지 수 	
참여율	<ul style="list-style-type: none"> • 이용자가 연구데이터 랜딩페이지에 댓글 혹은 별점을 부여한 수를 측정 • 적극적인 이용자의 참여가 이루어지고 있는 데이터 즉 활용성이 높은 데이터라는 점을 간접적으로 측정 ✓ 댓글, 별점이 부여된 연구데이터 랜딩페이지/전체 연구데이터 랜딩 페이지(전체 연구데이터수) 	
공유수	<ul style="list-style-type: none"> • 소셜 미디어로 공유된 연구데이터의 수를 측정 ✓ RDP 시스템에서 제공하는 소셜 플러그인으로 해당 데이터셋 페이지를 공유한 숫자 	
공유율 (Social Impact)	<ul style="list-style-type: none"> • 소셜 미디어로 공유된 연구데이터의 비율을 측정 ✓ 소셜 플러그인으로 공유한 연구데이터 수 / 전체 연구데이터 수 	서울특별시 Altmetrics
유효방문률	<ul style="list-style-type: none"> • 실제 연구데이터를 활용한 유효 이용자를 측정 ✓ 연구데이터 열람, 다운로드가 가능한 랜딩페이지까지 도달한 방문자 / 전체 방문자 	서울특별시 Google Analytics
재방문수	<ul style="list-style-type: none"> • 지속성을 갖는 이용자를 측정하여 연구데이터의 활용도를 간접적으로 측정 ✓ 특정 기간에 방문한 방문자 중 이전 기간에 방문한 적이 있는 방문자의 수 	서울특별시
별점 (직접입력)	<ul style="list-style-type: none"> • 이용자가 직접 연구데이터의 가치와 활용성에 대해 응답하고 측정 ✓ 5점 기준 별점을 부여할 수 있는 기능 제공 	

3. 평가척도와 측정지표의 타당성, 신뢰성 검정

3.1 평가 개요

연구데이터의 유용성 평가를 위해 본 연구에서 도출된 내재적 평가척도 4개와 해당 측정지표 및 이용자 경험 기반의 측정지표 12개에 대하여 RDP의 잠재적 이용자를 대상으로 설문 조사를 통해 타당성과 신뢰성을 검정하였다. 설문지는 각 평가척도와 측정지표에 대한 적합성 정도를 측정하는 Likert 5점 척도로 구성되었으며([부록 2] 참조), 온라인 웹 설문 링크를 발송하여 10월 7일부터 10월 18일까지 총 12일에 걸쳐 조사를 실시하였다. 전체 응답자는 225명, 분석 가능한 유효 응답자는 164명이었으며, 설문 대상자는 한국과학기술정보연구원 소속 연구원, 한국과학기술정보연구원이 제공한 초청 이용자 그룹, 국가연구데이터플랫폼 협의체, UX/UI평가그룹, S 대학 기초학문자료센터 실무자, 교수, 문헌정보학 박사급 강사, 문헌정보학 및 인터랙션사이언스 대학원생, 대학도서관 사서로 구성되었다.

3.2 평가척도와 측정지표의 타당성

3.2.1 내재적 유용성 평가척도와 측정지표의 타당성

KMO Bartlett 방식을 사용하여 연구데이터의 내재적 평가척도인 검색성, 접근성, 상호운용성, 재활용성의 척도별 측정지표가 적합하게 구성되었는지를 검정하였다. 검정결과 검색성은 KMO 0.813, Bartlett 0.000, 접근성은 KMO 0.686, Bartlett 0.000, 상호운용성은 KMO 0.734, Bartlett 0.000, 재활용성은 KMO 0.942, Bartlett 0.000로 4개 척도 모두 KMO 및 Bartlett의 검정 값이 기준치인 $KMO > 0.5$, Bartlett의 유의도 $p \leq .05$ 를 충족시켜 평가척도의 구성이 적절한 것으로 검정되었다(<표 5> 참조).

3.2.2 내재적 유용성 측정지표의 타당성

타당성이 검정된 4개의 내재적 평가척도의 측정지표에 대해 주성분 분석과 베리맥스 회전분석법을 사용하여 척도와 지표 구성 간의 타당성을 검정하였다. 성분 추출 결과 4개 척도 모두 하나의 성분만 추출만 되었으며 회전 성분 행렬 값은 추출되지 않아 측정지표들이 해당 척도를 잘 설명하고 있는 것으로 검정되었다. 척도별 성분 추출 결과는 <표 6>과 같다.

<표 5> 내재적 유용성 평가척도의 KMO Bartlett 검정 결과

	검색성	접근성	상호운용성	재활용성	
표본 적절성의 Kaiser-Meyer-Olkin 척도	0.813	0.686	0.734	0.942	
Bartlett의 구형성 검정	근사 카이제곱	435.909	223.535	281.239	1432.081
	자유도	6	3	3	45
	유의확률	0.000	0.000	0.000	0.000

〈표 6〉 내재적 평가척도의 측정지표 성분 분석 검정 결과

구분	총분산 결과						
	성분	초기 고유값			추출 제곱합 적재량		
		전체	% 분산	누적 %	전체	% 분산	누적 %
검색성	1	3.115	77.884	77.884	3.115	77.884	77.884
	2	0.415	10.363	88.247			
	3	0.296	7.406	95.653			
	4	0.174	4.347	100.000			
접근성	1	2.306	76.857	76.857	2.306	76.857	76.857
	2	0.457	15.246	92.103			
	3	0.237	7.897	100.000			
상호운용성	1	2.453	81.758	81.758	2.453	81.758	81.758
	2	0.334	11.141	92.899			
	3	0.213	7.101	100.000			
재활용성	1	7.027	70.266	70.266	7.027	70.266	70.266
	2	0.618	6.181	76.447			
	3	0.533	5.325	81.772			
	4	0.382	3.817	85.590			
	5	0.376	3.759	89.349			
	6	0.338	3.377	92.726			
	7	0.195	1.950	94.676			
	8	0.191	1.907	96.582			
	9	0.179	1.787	98.369			
	10	0.163	1.631	100.000			

3.2.3 이용자 경험 기반 유용성 측정지표의 타당성

연구데이터의 이용자 경험 기반 유용성 평가를 위해 도출된 12개 측정지표에 대해 KMO Bartlett 방식을 사용하여 지표 구성의 타당성을 검정하였으며, 주성분 분석과 베리맥스 회전분석법을 사용하여 지표의 동일성을 검정하

였다. 검정 결과 KMO 0.734, Bartlett 0.000으로 기준치인 $KMO > 0.5$, Bartlett의 유의도 $p \leq .05$ 를 만족하여 지표 구성이 적절한 것으로 검정되었다. 이에 비해 지표의 동일성에 관한 성분 분석 결과는 12개 측정지표가 2개의 성분으로 나누어졌다(〈표 7〉 참조).

2개 성분으로 나누어졌다는 것은 측정지표들

〈표 7〉 이용자 경험 기반 측정지표 성분 분석 검정 결과

성분	초기 고유값			추출 제곱합 적재량			
	전체	% 분산	누적 %	전체	% 분산	누적 %	
1	7,137	59.472	59.472	7,137	59.472	59.472	
2	1,023	8.523	67.995	1,023	8.523	67.995	
3	0,824	6.867	74.862				
4	0,795	6.626	81.488				
5	0,534	4.450	85.938				
6	0,455	3.791	89.728				
7	0,330	2.753	92.481				
8	0,241	2.005	94.487				
9	0,233	1.939	96.425				
10	0,181	1.511	97.937				
11	0,153	1.278	99.215				
12	0,094	0.785	100.000				
표본 적절성의 Kaiser-Meyer-Olkin 측도						0.734	
Bartlett의 구형성 검정						근사 카이제곱	281.239
						자유도	3
						유의확률	0.000

을 2개의 단위로 나누어야 각 단위를 구성하는 지표간의 동일성이 유지된다는 것을 의미한다. 회전된 성분행렬 분석 결과 5번, 6번, 3번, 4번 10번, 9번 11번 지표와 11번, 7번, 8번, 1번, 2번, 12번 지표로 구분되는 것으로 나타났다(〈표 8〉 참조).

3.2.4 이용자 경험 기반 유용성 평가척도의 개념화

〈표 9〉는 이용자 경험 기반의 측정지표를 성분 1과 성분 2로 구분하여 동일 성분별로 정리한 것이다. 각 성분별 지표의 분석 결과 성분1의 지표는 연구데이터의 인용, 다운로드, 방문, 열람에 관한 것으로 활용의 측정을 지향하며,

성분2의 지표는 공유, 참여, 별점 등 이용자의 참여 측정을 지향하는 것으로 나타났다. 따라서 성분1을 활용도로 성분2는 참여도로 개념화하여, 측정지표로만 구성되었던 이용자 경험 기반 유용성의 평가척도로 삼았다.

3.3 평가척도의 신뢰성

연구데이터 유용성의 내재적 평가척도와 이용자 경험 기반 평가척도에 대해 Cronbach's alpha 계수를 측정하여 척도의 신뢰성을 측정하였으며, 측정 결과 모든 척도가 기준치 0.6을 상회하여 신뢰성을 가지는 것으로 검정되었다(〈표 10〉 참조).

〈표 8〉 이용자 경험 기반 측정지표의 회전 성분 행렬

성분	초기 고유값			추출 제곱합 적재량			
	전체	% 분산	누적 %	전체	% 분산	누적 %	
1	7,137	59.472	59.472	7,137	59.472	59.472	
2	1,023	8.523	67.995	1,023	8.523	67.995	
3	0,824	6,867	74.862				
4	0,795	6,626	81.488				
5	0,534	4,450	85.938				
6	0,455	3,791	89.728				
7	0,330	2,753	92.481				
8	0,241	2,005	94.487				
9	0,233	1,939	96.425				
10	0,181	1,511	97.937				
11	0,153	1,278	99.215				
12	0,094	0,785	100.000				
표본 적절성의 Kaiser-Meyer-Olkin 측도						0.734	
Bartlett의 구형성 검정						근사 카이제곱	281.239
						자유도	3
						유의확률	0.000

	성분	
	1	2
UQ_05	0.823	0.211
UQ_06	0.800	0.329
UQ_03	0.787	0.294
UQ_04	0.781	0.362
UQ_10	0.718	0.343
UQ_09	0.700	0.426
UQ_11	0.522	0.463
UQ_07	0.224	0.883
UQ_08	0.256	0.869
UQ_01	0.470	0.658
UQ_02	0.503	0.612
UQ_12	0.405	0.598

〈표 9〉 이용자 경험 기반 측정지표의 성분별 구분

구분	코드	측정지표
성분1	UQ_05	인용수: 사이트 내 개별 연구데이터의 실제 인용 건수를 측정
	UQ_06	인용비율: 전체 연구데이터 구축 건수 대비 사이트 내 실제 인용 건수의 비율을 측정
	UQ_03	다운로드 수: 사이트 내 개별 연구데이터 실제 다운로드 건수를 측정
	UQ_04	다운로드 비율: 전체 연구데이터 구축 건수 대비 사이트 내 실제 다운로드 건수의 비율을 측정
	UQ_10	유효방문률: 실제 연구데이터를 활용한 유효 이용자 수의 비율을 측정
	UQ_09	열람수: 개별 연구데이터의 다운로드 및 이용 가능한 페이지까지 도달한 방문자수를 측정
	UQ_11	재방문수: 연구데이터를 지속적으로 활용하는 이용자의 규모를 측정
성분2	UQ_07	공유수: 소셜 미디어를 활용하여 공유되어진 개별 연구데이터의 건수를 측정
	UQ_08	공유율: 전체 연구데이터 구축 건수 대비 사이트 내 소셜미디어로 공유되어진 연구데이터의 비율을 측정
	UQ_01	참여수: 실제 사용자 (active user)의 확인을 위해 댓글, 별점 등이 부여된 연구데이터 랜딩페이지 수를 측정
	UQ_02	참여율: 전체 연구데이터 구축 건수 대비 실제 사용자 (active user)에 의한 댓글, 별점 등의 결과로 (참여 /이용) 활성화가 확인되어지는 연구데이터의 비율을 측정
	UQ_12	별점(직접입력): 개별 연구데이터에 대해 이용자가 평가한 가치 및 활용성 정도를 별점 부여 방식을 통해 측정

〈표 10〉 연구데이터 유용성 평가척도의 Cronbach's alpha 계수 측정 결과

구분		Cronbach의 alpha 계수
내재적 평가척도	검색성	0.905
	접근성	0.849
	상호운용성	0.888
	재활용성	0.951
이용자 경험 기반 평가척도	활용도	0.917
	참여도	0.883

3.4 최종 평가모델

타당도와 신뢰도 검증을 마친 연구데이터 유용성 평가척도, 측정지표, 측정방식은 〈표 11〉과 같이 정리된다.

4. 결론

본 연구에서는 연구데이터의 유용성 평가척도와 측정지표를 개발하고 그 타당성과 신뢰성을 검증하여 유용성 평가모형으로 제시하였다. 모형은 연구데이터의 내재적 유용성과 관련된 검색성, 접근성, 상호운용성, 재활용성 등 4개

〈표 11〉 연구데이터 유용성 평가척도, 측정지표, 측정방식

구분	평가척도	측정지표	측정 방식
내재적 측면	검색성	• 연구데이터 검색이 가능한 적합한 메타데이터를 사용하고 있는가?	• 연구데이터 검색을 위해 표준 용어(메타데이터)를 사용하고 있는지 확인 • DCMI, ISO, W3C 등 공인기관 혹은 학술 연구기관, 집단에서 온라인에 온톨로지, 스키마를 공개한 표준 메타데이터 중 기술용(descriptive metadata)을 사용하고 있는지 확인
		• 연구데이터에 고유하고 영구적인 식별자를 할당하고 있는가?	• 연구데이터에 고유하고 영구적인 식별자를 사용하고 있는지 확인
		• 연구데이터 관리, 서비스에 적용한 메타데이터는 식별자 요소를 포함하고 있는가?	• 다양한 목적별 메타데이터 요소가 모두 식별자 요소를 포함하고 있는지 확인
		• 연구데이터, 메타데이터에 대한 색인이 이루어지고 있는가?	• 연구데이터, 메타데이터 색인이 이루어지고 있는지 확인
	접근성	• 연구데이터에 자유로운 접근이 가능한가?	• 연구데이터 서비스와 데이터 활용을 위한 접근에 제한사항이 존재하는지 확인
		• 연구데이터는 기계가독형인가?	• 연구데이터는 기계처리가 가능한 형식인지 확인
		• 연구데이터는 개방형 포맷인가?	• 연구데이터가 누구나 제한없이 접근해서 활용할 수 있는 형식인지 확인
	상호운용성	• 해당 연구데이터, 메타데이터를 가지고 지식표현이 용이한 형태를 유지하고 있는가?	• 연구데이터와 메타데이터는 데이터 간, 메타데이터 간 활용과 배포가 용이한 형태인지 확인
		• 관리, 서비스 대상 연구데이터와 메타데이터는 검색성, 접근성, 상호운용성, 재활용성을 충족하는가?	• 연구데이터와 메타데이터가 누구나 접근, 검색, 활용이 가능한 목적 달성을 충족시킬 수 있는 구조를 갖추고 있는지 확인
		• RDP 운영조직에 기술적 상호운용성 보장을 위한 조직이 존재하는가? 혹은 관련 연구를 지속적으로 수행하고 공표하고 있는가?	• RDP 운영에 책임있는 기관은 기술적 상호운용성을 보장하기 위한 노력을 지속하고 있는지 확인
	재활용성	• 연구데이터의 벌크 다운로드가 가능한가?	• 연구데이터의 일괄 다운로드가 가능한지 확인
		• 연구데이터는 적시, 최신성을 유지하고 있는가?	• 연구데이터에 대한 업데이트 주기 등을 확인할 수 있는 요소가 존재하는지 확인
		• 연구데이터에 대한 상세 설명을 하고 있는가?	• 이용자가 직접 다운로드 받기 전 연구데이터에 대한 충분한 이해가 가능한 정보를 제공하고 있는지 확인
		• 연구데이터의 출처(원저작물, 원연구자료)를 포함하고 있는가?	• 연구데이터를 생산한 원본 저작물에 대한 정보를 제공하고 있는지 확인
		• 별도 형식의 원천데이터가 존재하는 변환 데이터인 경우 변환과정에 대한 상세설명을 포함하고 있는가?	• 연구데이터가 비표준형태 혹은 비디지털형태의 원본이 존재하는 경우 이를 변환한 과정에 대한 상세설명을 포함하고 있는지 확인
		• 데이터와 데이터셋을 기술하는 메타데이터는 해당 지식분야(커뮤니티)에서 활용하고 있는 표준 메타데이터인가?	• 연구데이터를 이해하고 활용하기 위해 별도의 메타데이터에 대한 이해를 위한 학습과정과 가공과정이 필요한지 확인
		• 연구데이터가 해당 지식분야(커뮤니티)에서 일반적으로 활용하고 있는 형식(format)을 준수하고 있는가?	• 연구데이터를 이해하고 활용하기 위해 별도 형식에 대한 이해를 위한 학습과정과 가공과정이 필요한지 확인

구분	평가척도	측정지표	측정 방식
		• 연구데이터를 기술하는 메타데이터는 출처정보를 표현하기 위한 요소를 포함하고 있는가?	• 연구데이터에서 사용하는 메타데이터 요소에 출처정보를 확인할 수 있는 요소가 있는지 확인
		• 연구데이터를 활용한 이용자의 의견(평가)을 수 용할 수 있는 기술적 장치(댓글 등)가 존재하는가?	• 연구데이터의 품질과 향후 개선을 위한 이용자 의견 수렴이 가능한 기술적 장치의 존재 여부 확인
		• 연구데이터에 오픈 라이선스를 부여했는가? (재 활용에 비용을 소모하지 않는 오픈 라이선스)	• 연구데이터의 재활용이 불가능한 라이선스를 부여했는지 확인
이용자 경험 기반	활용도	• 참여수	• active user를 간접적으로 측정 - 댓글, 별점이 부여된 연구데이터 랜딩 페이지수
		• 참여율	• 이용자가 연구데이터 랜딩페이지에 댓글 혹은 별점을 부여한 수, 적극적인 이용자의 참여가 이루어고 있는 데이터 즉 활용성이 높은 데이터 라는 점을 간접적으로 측정 - 댓글, 별점이 부여 된 연구데이터 랜딩 페이지 / 전체 연구데이터 랜딩 페이지(전체 연구데이터 수)
		• 다운로드수	• 실제 활용하고 있는 연구데이터 숫자를 측정 - 연구데이터를 다운로드한 수
		• 다운로드 비율	• 실제 적극적으로 활용하고 있는 연구데이터의 비율을 측정 - 연구데이터 다운로드 수 / 전체 연구데이터 수
		• 인용수	• 데이터의 재활용 결과를 간접적으로 측정 - 해당 연구데이터를 실제 논문, 프로젝트 등 학술 연구 활동 등에 활용한 수
		• 인용비율	• 데이터의 재활용 결과를 간접적으로 측정 - 다른 연구, 저작물에 활용된 수 / 전체 연구데이터 수
	참여도	• 공유수	• 소셜 미디어로 공유된 연구데이터의 수 - RDP 시스템에서 제공하는 소셜 플러그인으로 해당 데이터셋 페이지를 공유한 숫자
		• 공유율	• 소셜 미디어로 공유된 연구데이터의 비율 - 소셜 플러그인으로 공유한 연구데이터 수 / 전체 연구데이터 수
		• 열람수	• 개별 연구데이터의 다운로드 및 이용이 가능한 랜딩페이지까지 도달한 방문자 수 - 개별 연구 데이터의 다운로드 및 이용 가능한 페이지까지 도달한 방문자 수를 측정
		• 유효방문률	• 실제 연구데이터를 활용한 유효 이용자를 측정 • 연구데이터 열람, 다운로드가 가능한 랜딩 페이지까지 도달한 방문자 / 전체 방문자
		• 재방문수	• 지속성을 갖는 이용자를 측정하여 연구데이터의 활용도를 간접적으로 측정 • 특정 기간에 방문한 방문자 중 이전 기간에 방문한 적이 있는 방문자의 수
		• 별점(직접입력)	• 이용자가 직접 연구데이터의 가치, 활용성에 대해 응답하고 측정

의 평가척도와 이용자 경험 기반의 유용성에 관한 활용도, 참여도의 2개 평가척도로 구성되어 있으며, 각 평가척도는 해당 측정지표를 통해 측정된다.

내재적 유용성 평가 모형을 개발하기 위하여 국내 법률, EC, EPSRC, OECD, ISO의 연구데이터에 관한 정의를 검토하였으며, FAIR, Elsevier, RIN, Leiden Manifesto, OECD, 오바마 행정부, 서울특별시의 관련 가이드라인과 기준들을 분석하였다. 분석을 통해 도출된 평가척도와 측정기준에 대하여 RDP 시스템에 적용할 수 있는지 검토하였다. 이용자 경험 기반 유용성 평가 모형은 Google Analytics, Altmetrics, YouTube, 서울시의 관련 지표를 분석하여 12개의 측정지표를 도출하였으며, 지표에 대한 성분 분석 결과 2개로 나누어진 지표 성분을 활용도와 참여도로 개념화한 후 이를 평가척도의 명칭으로 삼았다.

평가척도와 측정지표에 대한 타당성과 신뢰성 검정을 위해 연구데이터의 잠재적 이용자 164명을 대상으로 온라인 설문조사를 실시하여 평가척도와 측정지표에 대한 타당성 및 평가척도에 대한 신뢰성을 검증하였다. 평가척도와 측정지표의 타당성 검정을 위해서는 KMO Bartlett 분석을 하였으며, 측정지표의 보편성과 차별성을 탐색하기 위해 주성분분석과 베리맥스 회전분석법에 의한 요인분석을 하였다. 통계분석 결과 제시된 내재적 유용성의 평가척도인 검색성, 접근성, 상호운용성, 재활용성과 해

당 측정지표 모두 통계적으로 타당성이 있는 것으로 검증되었다. 이용자 경험 기반의 측정지표는 2개 성분으로 나누어 설명이 가능한 것으로 나타나 각 성분을 활용도와 참여도로 개념화하여 평가척도로 삼았으며, 두 평가척도 모두 타당성이 있는 것으로 검증되었다. 평가척도에 대한 신뢰성은 cronbach's alpha 계수를 사용하여 검증하였으며, 제시된 6개의 평가척도 모두 통계적으로 신뢰성이 있는 것으로 검증되었다.

본 연구는 크게 두 가지 측면에서 의미를 지닌다. 하나는 통계적인 측면에서 유의미한 연구데이터 유용성 평가 모형을 구축함으로써 계량적 측정의 실제 응용이 용이하다는 것이며, 다른 하나는 개발된 평가척도와 측정지표의 많은 부분이 FAIR Principle을 따르고 있어서 RDP 외의 다른 연구데이터 서비스에서 FAIR 준수여부(FAIRness)를 측정하는 도구로 사용될 수 있다는 점이다. 그리고 개발된 유용성 평가모형을 충족시키기 위해 시범서비스 되고 있는 RDP의 연구데이터 및 서비스 시스템에 대한 개선 지침으로 사용될 수 있을 것이다. 그렇지만 본 연구에서 개발된 유용성 평가 모형은 RDP의 시범서비스에 대한 잠재적 이용자를 대상으로 검증한 것이므로, 안정화된 서비스에 적용할 수 있도록 유용성 평가 모형에 대한 지속적인 개선 작업과 이용자 환류 체계를 구축하는 것이 필요하다.

참 고 문 헌

- 강병서, 김계수 (2009). (SPSS 17.0) 사회과학 통계분석. 서울: 한나래.
- 김은정, 남태우 (2012). 연구데이터 수집에 영향을 미치는 요인 분석. 정보관리학회지, 29(2), 27-44.
<http://dx.doi.org/10.3743/KOSIM.2012.29.2.027>
- 김주섭, 김선태, 최상기 (2019). 연구데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건. 한국문헌정보학회지, 53(3), 317-344. <http://dx.doi.org/10.4275/KSLIS.2019.53.3.317>
- 김지현 (2015). 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 한국문헌정보학회지, 49(3), 413-436. <http://dx.doi.org/10.4275/KSLIS.2015.49.3.413>
- 박미영, 안인자, 남승주 (2018). 과학기술분야 출연연구기관 연구데이터 관리 및 공유 사례 분석 연구. 한국비블리아학회지, 29(4), 319-344. <http://dx.doi.org/10.14699/kbiblia.2018.29.4.319>
- 서울특별시 (2015). 서울시 공공데이터 이용 지표 발굴. 서울: 서울특별시.
- 송지준 (2008). 논문작성에 필요한 SPSS/AMOS 통계분석방법. 서울: 21세기사.
- 유사라 (2019). 차세대 학술연구데이터 공유 활성화를 위한 연구기록의 구조적 요건에 대한 연구. 한국문헌정보학회지, 53(3), 101-120. <http://dx.doi.org/10.4275/KSLIS.2019.53.3.101>
- 조재인 (2016). Data Citation Index를 기반으로 한 연구데이터 인용에 관한 연구. 한국문헌정보학회지, 50(1), 189-207. <http://dx.doi.org/10.4275/KSLIS.2016.50.1.189>
- 최리진, 정영미 (2019). 연구데이터의 법적 상호운용성 가이드라인에 관한 연구. 한국도서관·정보학회 하계 학술발표회 논문집, 241-250.
- Abran, A., Khelifi, A., Suryan, W., & Seffah, A. (2003, April). Consolidating the ISO usability models. In Proceedings of 11th international software quality management conference. 23-25.
- Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, P., Tait, J., & Zijlstra, T. (2009). Open data handbook. Retrieved from <http://opendatahandbook.org/guide/en/what-is-open-data/>
- Elsevier (2015). 10 aspects of highly effective research data. Retrieved from <https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data>
- Engineering and Physical Sciences Research Council (n.d.). EPSRC policy framework on research data. Retrieved from <https://epsrc.ukri.org/about/standards/researchdata/>
- European Commission (n.d.). Facts and Figures for open research data. Retrieved from https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en

- Executive Office of the President of the United States (2013). Open Data Policy-Managing Information as an Asset. Washington, D.C.
- Hicksa, D., Woutersb, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The leiden manifesto for research metrics. *Nature*, 520, 429-31.
- ISO/IEC 9126 (2001). Quality characteristics and guidelines for the user. Geneva: International Organization for Standardization.
- Pilat, D., & Fukasaku, Y. (2007). OECD principles and guidelines for access to research data from public funding. *Data Science Journal*, 6, OD4-OD11.
- Swan, A., & Brown, S. (2008). To share or not to share: Publication and quality assurance of research data outputs. A report commissioned by the research information network. University of Leicester (2012.09.04.). Research Data - Definitions. Retrieved from <https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3. <https://doi.org/10.1038/sdata.2016.18>

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

- Cho, Jane (2016). Study about research data citation based on DCI (Data Citation Index). *Journal of the Korean Society for Library and Information Science*, 50(1), 189-207. <http://dx.doi.org/10.4275/KSLIS.2016.50.1.189>
- Choi, Li-Jin, & Jung, Young-Mi (2019). A study on the legal interoperability guidelines for research data. *In Proceedings of Summit, Meeting of Korean Library And Information Science Society*, 2019. 5. 24, Gyeongsangbuk-do: Kyungpook National University Global Plaza, 241-250.
- Kim, Eun-Jung, & Nam, Tae-Woo (2012). Factor analysis of effects on research data collection. *Journal of the Korean Society for Information Management*, 29(2), 27-44. <http://dx.doi.org/10.3743/KOSIM.2012.29.2.027>
- Kim, Ji-Hyun (2015). A study on the perceptions of university researchers on data management and sharing. *Journal of the Korean Society for Library and Information Science*, 49(3), 413-436. <http://dx.doi.org/10.4275/KSLIS.2015.49.3.413>
- Kim, Jun-Seop, Kim, Sun-Tae, & Choi, Sang-Ki (2019). The functional requirements of core

elements for research data management and service. *Journal of the Korean Society for Library and Information Science*, 53(3), 317-344.

<http://dx.doi.org/10.4275/KSLIS.2019.53.3.317>

Park, Mi-Young, Ahn, In-Ja, & Nam, Seung-Joo (2018). A study on the analysis of research data management and sharing of science & technology government-funded research institutes. *Journal of the Korean Society for Library and Information Science*, 29(4), 319-344. <http://dx.doi.org/10.14699/kbiblia.2018.29.4.319>

Seoul Metropolitan Government (2015). *Developing indicators for public data use in seoul city*. Seoul: Seoul Metropolitan Government

You, Sa-Rah (2019). Reconsideration of research framework for RRM in the perspective of linked open data. *Journal of the Korean Society for Library and Information Science*, 53(3), 101-120. <http://dx.doi.org/10.4275/KSLIS.2019.53.3.101>

[부록 1] 중복제거 상세 내역

척도	측정요소	출처	검토내용
검색성	<ul style="list-style-type: none"> • 데이터를 활용, 재활용하게 하는 첫 번째 단계는 데이터를 찾을 수 있도록 하는 것임 • 메타데이터와 데이터 모두는 반드시 사람과 컴퓨터가 모두 쉽게 찾을 수 있도록 해야 함 • 기계가독형 메타데이터는 데이터셋과 해당 서비스를 자동화로 찾기 위한 필수요소로 중요성이 높음 F1. 데이터와 메타데이터에는 전역적으로 고유하고 영구적인 식별자를 할당 F2. 데이터는 풍부한 메타데이터로 기술해야 함(R1 참조) F3. 메타데이터는 명확하고 명시적으로 데이터를 기술할 수 있는 식별자를 포함해야 함 F4. 데이터와 메타데이터는 검색 가능한 자원에 등록하거나 색인되어야 함 	FAIR	채택
	<ul style="list-style-type: none"> • 사람이 검색할 수 있어야 함 	RIN	미채택: 측정요소로 부적절
	<ul style="list-style-type: none"> • 현재는 학술 논문 탐색은 어렵지 않지만 연구데이터의 경우는 아님 • 데이터를 찾기 위해서는 해당 연구를 찾는 것이 쉬운 방법임 • Elsevier는 데이터에 DOI를 붙이거나 또 다른 식별번호를 부여함으로써 이를 해결하고자 함 	Elsevier	미채택: FAIR에서 언급
접근성	<ul style="list-style-type: none"> • 이용자가 필요한 데이터를 검색하면 인증, 권한을 포함한 접근에 필요한 방법을 알 수 있도록 해야 함 A1. 데이터와 메타데이터는 표준통신프로토콜을 사용해서 데이터와 메타데이터에 부여한 식별자로 검색할 수 있어야 함 A1.1 이 프로토콜은 개방형이어야 하고, 무료여야 하며, 보편적으로 구현 가능한 것이어야 함 A1.2 이 프로토콜은 필요한 경우 인증이나 허가 절차를 허용할 수 있어야 함 A2. 데이터를 더 이상 사용할 수 없더라도 메타데이터에 대한 접근은 가능해야 함 	FAIR	채택: HTTP 중심으로 활용
	<ul style="list-style-type: none"> • 데이터를 저장, 보존했다는 것이 자동으로 접근을 보장하는 것은 아님 • 연구자와 기계는 각자의 목적에 맞게 데이터를 재사용하기 위한 접근이 필요함 • 연구자 개인이 이런 접근을 가능하게 하는 것은 어려우며 Mendeley Data (https://data.mendeley.com)와 같은 공유 저장소를 활용하는 것이 필요함 • Elsevier도 Open Data Pilot (https://www.elsevier.com/authors/author-resources/research-data/open-data)으로 데이터 저장, 보존, 접근, 탐색을 모두 지원하고자 노력하고 있음 	Elsevier	미채택: RDP는 이미 저장소
	<ul style="list-style-type: none"> • 사람이 접근할 수 있어야 함 	RIN	미채택: 측정요소로 부적절
	<ul style="list-style-type: none"> • 데이터는 인터넷을 통해서 다운로드 받을 수 있어야 하고, 원천데이터를 변환하거나 새로운 저작물 생산에 합리적인 비용 소모가 가능해야 함 	OKI	미채택: FAIR에서 언급
	<ul style="list-style-type: none"> • 오픈 데이터는 검색, 다운로드, 지표화가 가능하도록 편리하고 수정 가능하며 오픈 포맷(open format)과 기계가독형 형태(machine readable format)를 가져야 함. • 개방형 데이터는 특정한 개인이나 단체를 차별하지 않고 가능한 한 모든 이용자와 사용 목적을 고려하여 만들어지고 이에 따라 다양한 형태로 배포되며 그 형태(format)는 공공적으로 접근가능하고 규제를 받지 않아야 함. 	Obama	채택
<p>2.1. 데이터를 온라인상에서 활용 가능한가?</p> <p>2.1.1. 개방과 관련된 기관 홈페이지의 서비스 메뉴 혹은 독립된 사이트가 존재하는가?</p> <p>2.1.2. 데이터 활용을 위한 자유로운 접근이 가능한가?</p> <p>2.2. 제공되는 데이터는 기계가독형인가?</p> <p>2.2.1. 기계가독형(Machine Readable)으로 제공하는가?</p> <p>2.2.2. 개방형 포맷(Open Format)으로 제공하고 있는가?</p>	서울시	일부채택: 2.3 방문율을 제외한 나머지 측정 요소는 FAIR, OKI, 오바마행정부에서 언급 RDP의 경우 링크드데이터서비스를 고려하고 있지 않으므로 2.2.3, 2.2.4는 부적절함	

척도	측정요소	출처	검토내용
	2.2.3. 개체식별을 위해 URI를 활용하고 있는가? 2.2.4. Linked Data 원칙을 준수하고 있는가? 2.3. 방문율 2.3.1. 웹 브라우저 / 2.3.2. 언어 / 2.3.3. 국가, 지역		
상호 운용성	<ul style="list-style-type: none"> 일반적으로 데이터는 다른 데이터와 통합이 필요하고, 분석, 저장, 처리를 위해서 혹은 응용프로그램과 상호운용이 가능해야 함 I. 데이터와 메타데이터는 지식표현이 가능하도록 형식적이고, 접근가능하고, 공유 가능하며 광범위하게 적용 가능한 언어를 사용해야 함 II. 데이터와 메타데이터는 FAIR 원칙을 따르는 용어집을 사용해야 함 III. 데이터와 메타데이터는 다른 데이터, 메타데이터에 대한 한정참조를 포함해야 함 	FAIR	채택: 단, format은 접근성에서 언급
	<ul style="list-style-type: none"> 기술과 의미적 상호운용성 보장은 연구데이터의 국제, 학제간 접근과 사용을 가능하게 하는 핵심적인 요소임 데이터 접근을 보장하기 위해 국제표준을 활용해야 하며, 회원국과 연구기관은 새로운 표준을 개발할 국제기구와 협력해야 함 학술연구와 과학은 고도화, 세계화를 위해 노력하고 있지만, 기술표준과 절차표준의 비호환성은 데이터셋 활용을 심각하게 저해할 수 있음 표준은 상호운용성의 첫 번째 요구사항이므로 명시적으로 언급해야 함 학술연구는 가장 진보적인 분야의 관행을 채택하여 데이터의 수집과 보존을 위한 방법을 강구해야 함 	OECD	채택: 데이터셋 활용을 위한 표준화 절차
	데이터는 편리하게 수정 가능한 형태로 제공해야 함.	OKI	미채택: 접근성에서 언급
활용성	3.1. 제공되는 데이터는 무료로 활용 가능한가? 3.2. 제공되는 데이터를 가공하기 위한 추가적인 비용이 발생하는가? 3.3. 제공되는 데이터는 별크 형태로 활용이 가능한가? 3.4. 데이터는 적시, 최신성을 유지하면서 제공하고 있는가? 3.5. 서비스하고 있는 데이터가 무엇에 관한 것인가에 대한 상세한 설명을 포함하고 있는가? 3.6. 서비스하고 있는 데이터의 출처(source)에 대한 설명(원데이터에 대한 링크(위치정보)을 포함하고 있는가? 3.7. 제공하고 있는 데이터로 이용자가 할 수 있는 것에 대한 설명(데이터를 제공하는 이유에 대한 설명 포함)을 포함하고 있는가? 3.8. 원천데이터의 변환 과정에 대한 설명을 제공하고 있는가? 3.9. SPARQL Endpoint를 제공하고 있는가? 3.10. 데이터셋을 제공받을 수 있는 구문에 대한 정보를 제공하고 있는가?	서울시	미채택: FAIR에서 언급
	<ul style="list-style-type: none"> FAIR의 궁극적인 목표는 데이터를 재활용하는데 최적화하는 것임이며, 이를 위해서 메타데이터와 데이터는 잘 기술되어서 다른 설정으로 복제나 결합이 가능해야 함 R1. 데이터와 메타데이터는 복수의 정확하고 관련 있는 속성으로 풍부하게 기술해야 함 R1.1. 데이터와 메타데이터는 명확하고 접근 가능한 라이선스로 배포해야 함 R1.2. 데이터와 메타데이터는 상세한 출처정보를 포함하고 있어야 함 R1.3. 데이터와 메타데이터는 도메인과 관련하고 있는 커뮤니티의 표준을 충족해야 함 	FAIR	채택
재활용성	<ul style="list-style-type: none"> 향후 검증을 위한 재사용이 가능해야 함 	RIN	미채택: 측정요소로 부적절
	<ul style="list-style-type: none"> 데이터는 다른 데이터셋과의 조합적 사용을 포함하는 재사용, 재배포가 가능한 형태(가능하면 관련 라이선스나 법/제도적 조항 포함)여야 함 	OKI	채택
	<ul style="list-style-type: none"> 오픈 데이터는 오픈 라이선스(Open Licence)로 배포되어야 하며, 이용에 규제가 따라서는 안 됨 	Obama	채택
	4.1. 서비스하고 있는 데이터의 라이선스, 저작권에 대한 설명을 제공하고 있는가 4.2. 상업적 목적으로 자유로운 활용이 가능한가? 4.3. 적절한 표준을 활용한 값 표현 구조를 가지고 있는가? 4.3. 기존 표준용어집을 재활용하고 사용하고 있는가? 4.4. 신규 개발용어집, 용어를 잘 정의하고 사용하고 있는가?	서울시	채택

척도	측정요소	출처	검토내용
	<ul style="list-style-type: none"> • 재생산율: 열린 데이터 광장에서 제공하는 데이터셋이 실제로 학술연구, 개발 등에 활용되었는지를 측정하기 위한 것으로 이용자의 참여가 존재하지 않으면 평가하기 어려운 항목임. 때문에 배점으로 평가하지 않고 이용자와의 별도 커뮤니티 채널 혹은 제공 데이터셋에 댓글 및 이용자 참여서비스를 통해 지속적으로 관련 데이터가 누적될 수 있도록 조치가 필요함. • 제공 데이터가 학술연구정보에 활용되었는가? • 제공 데이터가 참고용으로(기사작성, 블로그 등) 인용되거나 활용되었는가? • 활용데이터가 웹서비스, 모바일 애플리케이션 등 시각적 확인이 가능한 응용서비스로 개발되었는가? • 재사용율: 재사용을 역시 배점을 통해 평가하기 어려운 항목으로 기존의 분석틀이 제공하는 기능을 활용하여 측정하고 별도정량평가요소로 활용이 필요함. • 데이터셋 갱신일자별 재사용률(다운로드율) • 한 번 활용한 데이터의 갱신이 이루어진 후 재다운로드가 이루어지는지 측정 		
검토성	<ul style="list-style-type: none"> • 논문과 마찬가지로 데이터에 대한 동료평가가 필요함 • 데이터형식, 해당 분야의 표준인지, 메타데이터를 첨부하고 있는지 여부를 검토 • Elsevier의 경우 Open Data Pilot에서 검토자는 제출 파일의 구문분석과 관련된 도메인 내에서 일반적으로 사용하는 원시 데이터 유형인지 확인함 	Elsevier	미채택: 재활용성에서 언급
재현 가능성	<ul style="list-style-type: none"> • 연구 결과의 재현성은 신뢰도는 높이기 위해 꼭 필요함 • Reproducibility Initiative는 연구결과와 신뢰도를 높이기 위해 데이터의 복제와 실험결과를 검증하고 있음(유료). • 그러나 도메인별로 특히 생물학 관련 연구의 경우 재현성 혹은 재사용하기에 충분한 조건을 충족하지 못하는 경우가 있음 • Force11 Resource Identification Initiative는 생물 의학 분야에서 고유 연구 리소스 식별자(RRID)사용 • 연구데이터 공유의 가장 큰 장점은 데이터의 재사용에 있음 • 연구데이터의 신뢰성과 재현성이 높을 때 재사용이 가능함 • 가장 적극적인 방법은 라이선스를 데이터셋에 부여하여 모든 이용자가 활용할 수 있도록 하는 것으로 크리에이티브커먼즈(CreativeCommons)가 가장 일반적인 방식임 	Elsevier	미채택: 재활용성에서 언급
저장성	<ul style="list-style-type: none"> • 수집 데이터를 저장하는 것 • 많은 연구 조직들이 데이터저장방법을 명확하게 정의하지 않고 있어 초기 실험의 목적으로 데이터 재사용(reusable)이 어려움 • 전자연구노트 등을 사용하는 방법이 대안이 될 수 있음 • 일반 데이터 저장소와 특수데이터저장소를 구분하여 데이터에 대한 접근과 공개 여부를 선택할 수 있는 방법 등을 제공하는 것도 필요함 	Elsevier	미채택: 연구데이터와 관련성 없음
보존성	<ul style="list-style-type: none"> • 보존은 단순한 지장이 아니라 장기 보존을 의미함 • 형식에 독립적(format-independent)인 방식으로 보존이 필요함 • 데이터복구, 재처리, 디지털화 등 향후 사용에 필요한 정보를 포함한 보존이 필요하며, 네덜란드의 DANS(http://www.dans.knaw.nl)가대표적임 • 데이터 저장소가 사라지더라도 데이터가 손실 없는 보존을 보장하는 다크 아카이브(dark archive)도 염두 해야 함 	Elsevier	미채택: 기관 정책 관련 요소
인용 가능성	<ul style="list-style-type: none"> • 데이터 인용은 데이터 제출, 재사용에 있어서 기록과 추적을 가능하게 함 • 연구자는 데이터에 고유한 영구식별자를 부여하는 방법을 고려해야 함 • FORCE11이 제공하는 데이터 인용법도 참고할 필요가 있음 	Elsevier	미채택: 재활용성에서 언급
이해성	<ul style="list-style-type: none"> • 이해성은 데이터 재사용이 목적이며, 측정단위, 데이터 수집 방법, 약어, 매개변수를 명확하게 기술했는지에 대한 것임 • 특히 메타데이터가 중요함 	Elsevier	미채택: 재활용성에서 언급
	<ul style="list-style-type: none"> • 쉽게 이해할 수 있는가를 측정 	ISO-9126	미채택: 측정요소로 부적절
수집 분석 기법 공개	<ul style="list-style-type: none"> • 평가를 위한 데이터베이스 구축은 확실하게 명시된 규칙을 따라야 하며 그 규칙은 평가를 완료하기 전에 제시해야 함 • 이해하기 쉬운 명확하고 간단한 연구성과 평가지표는 평가의 투명성을 높여줌 • 그러나 단순화된 평가지표가 연구활동 전반을 반영할 수 없으므로, 왜곡의 소지가 있음 • 따라서 평가자는 연구과정의 복잡성과 평가지표의 단순성 간의 균형을 반드시 유지해야 함 	Leiden	미채택: 재활용성에서 언급

척도	측정요소	출처	검토내용
분석과정 공개	<ul style="list-style-type: none"> 연구성과 평가에 사용하는 데이터의 품질을 확실히 하기 위해 평가에 관여한 모든 연구자는 연구성과를 정확하게 반영하였는지 확인할 수 있어야 함 평가과정을 감독하고 관리하는 평가담당자는 자체검증 또는 외부감사로 데이터의 정확성을 담보해야 함 대학의 연구정보시스템(RIS)에는 데이터 검증모듈을 포함해야 하며 이를 연구정보시스템공급업체 선정의 기준으로 활용해야 함 정확한 고품질데이터를 수집하고 처리하기 위해서는 시간과 재원이 필요하므로 이를 위한 예산 할당이 필요함 	Leiden	미채택: 재활용성에서 언급
재배포	<ul style="list-style-type: none"> 데이터는 다른 데이터셋과의 조합적 사용을 포함하는 재사용, 재배포가 가능한 형태(가능하면 관련 라이선스나 법/제도적 조항 포함)여야 함 	OKI	미채택: 재활용성에서 언급
범용적인 참여	<ul style="list-style-type: none"> 모든 사람이 사용, 재사용, 재배포할 수 있어야 함. 예를 들어, '상업적', '비상업적' 사용에 대한 제한 혹은 특정 목적(교육용 등)을 위한 사용의 제한 등이 없어야 함 	OKI	미채택: 재활용성에서 언급
참여와 소통	<ol style="list-style-type: none"> 소셜 미디어, 데이터개방서비스 사이트에 이용자 참여가 가능한 채널이 존재하는가? 데이터 개방 관련 정책 수립 등에 이용자가 참여하고 검증할 수 있는 체계, 사례가 존재하는가? 기관의 열린 데이터를 실제 활용하고 있는 이용자그룹(시민단체, 학계등)이 존재하는가? 기관 데이터를 활용한 정기적인 공동창작행사(해커톤 등)가 존재하는가? 기관 데이터 이용자 간의 커뮤니케이션이 가능한 장이 존재하는가? 외부 국내외 다양한 데이터 개방사이트에서 현재 제공하고 있는 데이터의 동일 수준 활용이 가능한가? 브랜드 로고와 링크 정보를 제공하고 있는지와 서비스를 지원, 운영하고 있는 주체에 대한 정보 제공하고 있는가? 	서울시	미채택: 연구데이터에 부적절
상세설명	<ul style="list-style-type: none"> 오픈 데이터는 이용자가 데이터를 처리하는 데 필요한 정보뿐만 아니라 데이터의 강점, 약점, 한계, 보안 요구사항 등에 관한 상세하고도 충분한 설명을 갖추고 있어야 함. 따라서 오픈 데이터는 메타데이터에 대한 상세 문서와 데이터 사전뿐만 아니라 샘플 데이터 등의 다양한 정보를 제공해야 함. 	Obama	미채택: 재활용성에서 언급
완전성	<ul style="list-style-type: none"> 오픈 데이터는 처음 수집한 원천 형태 그대로 실행 가능하도록 제공되어야 함. 2차 저작물인 경우는 항상 원 데이터도 개방되어야 함 	Obama	미채택: 재활용성에서 언급
	<ol style="list-style-type: none"> 제공하고 있는 데이터에 대한 부족한 부분, 향후 업데이트 수정사항 등에 대한 상세정보를 제공하고 있는가? 개방된 데이터는 즉시 실행 및 활용 가능한 형태로 제공되고 있는가? 2차 저작물(이용자 혹은 데이터 개방자에 의해 원천데이터가 변경된 경우)의 경우 항상 원천데이터도 함께 제공하고 있는가? 여러 데이터셋이 모여 하나의 유의미한 데이터셋을 완전성은 보장되는가? 	서울시	미채택: 재활용성에서 언급
공공성	<ul style="list-style-type: none"> 정부 기관과 관련 단체 등이 구축한 정보와 데이터 중 개인정보, 비밀성, 보안과 같은 최소한의 규제를 제외한 모든 정보와 데이터는 공공성을 가짐 	Obama	미채택: 이미 공공성 확보
적시성	<ul style="list-style-type: none"> 오픈 데이터는 데이터의 가치 유지를 위해서 가능한 한 빠르게 배포되어야 하고 배포의 빈도는 주요 이용자와 최종 이용자의 필요에 따라 결정되어야 함 	Obama	미채택: 재활용성에서 언급
사후관리	<ul style="list-style-type: none"> 오픈 데이터의 이용이나 이용자의 사후 요구에 응대할 수 있도록 담당자의 연락처가 명시되어야 함. 	Obama	미채택: 측정요소로 부적절
관리체계	<ul style="list-style-type: none"> 개방형 데이터를 뒷받침해 줄 수 있는 정책, 예산, 홍보 등이 이루어지고 있는가? 	서울시	미채택: 측정요소로 부적절
개방성	<ul style="list-style-type: none"> 누구나 접근할 수 있는 제한이 없는 데이터 	EU	미채택: 재활용성에서 언급
	<ol style="list-style-type: none"> 서울시 예산 데이터가 개방되었는가? 서울시 소재 기업에 대한 목록정보가 개방되었는가? 실행 위반 코드 데이터가 개방되었는가? 건축 허가 데이터가 개방되었는가? 범죄 데이터가 개방되었는가? 구획 데이터가 개방되었는가? 조달 계약 데이터가 개방되었는가? 	서울시	미채택: 측정요소로 부적절

척도	측정요소	출처	검토내용
	1.8. 재산 평가데이터가 개방되었는가? 1.9. 공공 빌딩 데이터가 개방되었는가? 1.10. 식당 검사 데이터가 개방되었는가? 1.11. 서비스 요청 데이터가 개방되었는가? 1.12. 지출 데이터가 개방되었는가? 1.13. 운송 데이터가 개방되었는가? 1.14. 지대 정보가 개방되었는가? 1.15. 웹 분석 데이터가 개방되었는가?		
	<ul style="list-style-type: none"> 가능한 최저 비용으로 국제적인 연구 공동체가 동등한 조건으로 연구데이터에 접근할 수 있어야 함 공적 기금이 들어간 연구데이터라면 인터넷을 기반으로 쉽고, 시기적절하고 이용자 친화적으로 제공해야 함 	OECD	미채택: 접근성에서 언급
유연성	<ul style="list-style-type: none"> 정보통신 기술의 예측 불가능한 변화와 각 연구 분야의 특성 및 국가별 연구기관별 연구 시스템, 법률 시스템, 문화의 다양성을 고려해야 함 연구데이터 서비스와 같이 접근 관련 계약이나 약정을 개발할 경우 정부의 원칙, 지침, 사회, 경제적인 규제 등을 검토하여 구체적으로 제시해야 함 	OECD	미채택: 측정요소로 부적절
투명성	<ul style="list-style-type: none"> 연구데이터와 이를 만든 개인, 조직에 대한 정보와 데이터에 대한 문서, 사용조건 등을 명세화하여 인터넷으로 투명하게 공개하고 국제적으로 누구나 사용할 수 있어야 함 투명성 확보를 위한 핵심요소는 다음과 같음 <ul style="list-style-type: none"> ✓ 데이터 생산조직, 보유현황, 사용 가능한 데이터셋 및 사용조건정보를 인터넷에서 쉽게 찾을 수 있어야 함 ✓ 공개적으로 자금을 지원받는 연구 수행 절차에 있어서는 개별연구자, 학회, 대학 및 기타 이해 관계자에게 연구데이터 정책에 대한 정보를 적극적으로 보급해야 함 ✓ 관련 연구 커뮤니티의 모든 구성원은 언제든지 데이터를 목록화하기 위한 표준화에 참여하고 합의하여 결과를 도출해야 함. 단, 연구자와 연구기관의 업무 부담을 줄이기 위해 기존기준을 적절하게 고려해야 함 ✓ 데이터 관리 및 접근조건에 대한 정보를 데이터저장소와 데이터생산 기관에 전달하여 모범 사례를 공유할 수 있어야 함 	OECD	미채택: 재활용성에서 언급
법률준수	<ul style="list-style-type: none"> 데이터 접근에 관한 약정(규약)은 공공 연구기관의 모든 이해 관계자의 법적 권리와 정당한 이해를 존중해야 함 특정 연구데이터에 대한 접근과 사용은 다양한 법적 요구사항으로 제한할 수 있으며 다음과 같음 국가안보: 정보, 군사활동 또는 정치적 의사결정과 관련된 데이터는 분류하여 제한접근의 대상이 될 수 있음 개인정보 및 기밀유지: 인적사항 및 기타 개인정보 데이터는 기밀성 및 개인정보를 보호하기 위해 국내 법률 및 정책에 따라 제한적으로 접근할 수 있음. 그러나 기밀유지수준을 보장하는 익명성 또는 기밀유지 절차는 연구자에게 최대한 많은 데이터 효용성을 유지하기 위해 관리자가 고려해야 함 영업비밀 및 지적재산권: 기밀정보를 포함한 비즈니스 또는 타사의 데이터는 연구를 위해 접근할 수 없음 희귀종, 위협 또는 멸종위기에 처한 생물 종의 보호, 보존을 위해 생물자원에 대한 접근을 제한할 수 있음 법적절차: 법적 조치(하위판단)에서 고려중인 데이터에 접근할 수 없음 	OECD	미채택: 측정요소로 부적절
지적 재산권 보호	<ul style="list-style-type: none"> 공공 기금으로 운영하는 연구데이터베이스는 관련 저작권, 지적재산권법의 적용을 고려해야 함 공공 기금으로 수행한 연구라도 민간과 협력할 경우 데이터 사용에 제한이 생길 수 있으나 이 경우라도 연구데이터에 대한 광범위한 접근을 용이하게 해야 함 상업적 이익을 보호하면서 데이터에 대한 접근을 허용할 수 있는 정책과 라이선스 개발 등 대책이 있어야 함 특정 관할지역에서 발생할 수 있는 지적재산권 문제를 피하여 공공연구, 기타 공익 목적을 위해 데이터에 접근할 수 있도록 해야 함 	OECD	미채택: 측정요소로 부적절

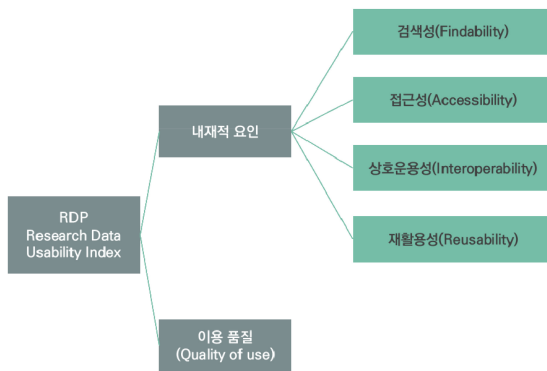
척도	측정요소	출처	검토내용
공식적 책임	<ul style="list-style-type: none"> • 데이터 접근을 보장할 수 있도록 당사자의 책임에 관한 규칙, 규정을 개발하여 공식적으로 제도화해야 함 • 제도적 규칙은 저작, 제작자, 소유권, 보급, 사용제한, 재정적 고려사항, 윤리규범, 라이선스조건, 보존 및 서비스 책임과 장기 보존까지 관련 가질 수 있어야 함 • 정부 또는 기관 차원에서 영향을 받을 수 있는 모든 이해관계자와의 합의 과정을 거쳐야 함 • 공동연구나 프로젝트 특히 국제적인 공동개발연구로 지역간 기관간 차이가 발생할 수 있는 경우에는 조기에 연구데이터 공유약정을 협상해야 함 • 기타 다음과 같은 사항을 고려해야 함 <ul style="list-style-type: none"> ✓ 데이터 접근, 보급, 공유 관련 문제는 접근 및 사용조건에 대한 명시적인 제도적 합의가 없을 때 발생함. 그간의 관행으로 이루어지던 비공식 합의는 공식적인 합의 절차로 보완해야 함 ✓ 데이터 접근과 관리의 다양한 측면에 대한 책임은 기관의 공식업무, 보조금신청, 연구계약, 출판계약, 라이선스와 같은 문서에서 공식적으로 확립해야 함 ✓ 데이터 접근에 필요한 인프라의 장기적 지속가능성은 특히 중요하므로 연구기관 및 정부 기관은 장기간에 걸쳐 효율적이고 적절히 연구데이터를 사용할 수 있도록 효과적으로 보존, 관리 및 접근에 책임이 있음을 공식화해야 함 	OECD	미채택: 측정요소로 부적절
전문성	<ul style="list-style-type: none"> • 연구데이터 관리를 위한 제도적 장치는 학술 연구 공동체에서 제도화된 구체적인 표준과 가치에 기초해야 함 • 고려해야 할 요소는 다음과 같음 <ul style="list-style-type: none"> ✓ 전문연구자와 지역 사회 기여를 제도화에 고려하면 연구데이터 접근을 규제해야 하는 부담을 줄일 수 있음 ✓ 연구자와 연구기관, 기타 조직 간의 신뢰는 이런 성문화한 제도적 장치의 수립과 유지에 중요한 역할을 함 ✓ 현재의 연구 관행에서는 초기데이터를 생산한 연구자 혹은 기관은 데이터를 임시로 독점적으로 사용하는 것으로 보장받을 수 있음. 이러한 인센티브 제도에 대한 규칙은 영향을 받는 연구공동체와 협력하여 자금원에 의해 개발되고 명시적으로 명시해야 함 ✓ 프로젝트 및 프로그램 기획 시 데이터 큐레이션에 대한 자금과 기술지원을 고려하여 명시적으로 기술해야 하며, 연구데이터 관리의 모든 분야에서 인센티브와 전문성 개발에 주의해야 함 	OECD	미채택: 측정요소로 부적절
품질	<ul style="list-style-type: none"> • 연구데이터의 가치와 유용성은 데이터 자체의 품질이 크게 좌우함 • 데이터 관리자와 데이터 수집 조직은 명시적인 품질 표준을 준수하는지 주의해서 확인해야 함 • 기준이 아직 존재하지 않는 경우, 기관 및 연구 단체는 연구 커뮤니티와 협력하여 개발해야 함 • 모든 학술연구분야가 동일한 수준의 품질을 요구하는 것은 아니므로 국제적인 표준 혹은 표준연구원과의 협의 하에 데이터품질요건을 마련하는 것이 중요함 • 구체적인 확인사항은 다음과 같음 <ul style="list-style-type: none"> ✓ 데이터저장소는 데이터 접근에 필요한 수집, 보급, 보존에 활용한 방법, 기술과 도구에 대한 우수사례를 기술하여 동료검토가 가능하도록 하여 품질관리에 도움을 줄 수 있어야 함 ✓ 출처는 입증 가능한 방법으로 문서화하고 명시해야 하며 이 문서는 데이터를 사용하는 모든 사람들이 쉽게 이용할 수 있어야 하고 데이터셋, 메타데이터와 통합할 수 있어야 함. 특히 메타데이터의 개발은 연구자, 과학자가 데이터셋의 의미를 이해할 수 있는데 중요함 ✓ 데이터셋에 대한 접근은 원본연구자료에 대한 접근과 항상 연결되어 있어야 하며 복사한 데이터셋은 원본 데이터셋과 연결되어야 함. 이는 데이터의 유효성 확인과 데이터셋 내의 오류식별을 용이하게 함 ✓ 연구기관 등은 데이터 인용에 필요한 적절한 방법을 개발해야 함 	OECD	미채택: 접근성, 재활용성에서 언급
보안	<ul style="list-style-type: none"> • 연구데이터는 무결성과 보안을 보장하기 위한 기술과 도구 지원에 주의를 기울여야 함 • 데이터셋의 무결성을 보장하기 위해 데이터의 완전성과 오류 부재 보장을 위한 노력을 기울여야 함 • 보안과 관련하여 관련 메타데이터 및 설명과 함께 데이터는 명시적 보안프로토 	OECD	미채택: 측정요소로 부적절

척도	측정요소	출처	검토내용
	<ul style="list-style-type: none"> • 쿨에 따라 고의 또는 의도하지 않은 손실, 파괴, 수정 및 무단 액세스로부터 보호해야 함 • 데이터셋과 관련 장비는 열, 먼지, 전기, 자기 및 정전기 방전과 같은 환경위험으로부터 보호해야 함 		
효율성	<ul style="list-style-type: none"> • 데이터 접근과 공유 촉진의 목표는 비싸고 불필요한 데이터 수집 작업의 중복을 피하기 위해 공공기금으로 조성한 과학 연구의 전반적인 효율성을 향상시키는 것 • 구체적인 고려사항은 아래와 같음 <ul style="list-style-type: none"> ✓ 데이터접근장치는 데이터 관리 및 전문 지원 서비스에 대한 우수사례를 기술함으로써 글로벌과학시스템 내에서의 비용 효율성을 증진시켜야 함 ✓ 공공기금으로 조성한 연구자료는 개방성 원칙에 의거하지만 모든 자료를 영구보존해야 하는 것은 아님 ✓ 데이터 보관 커뮤니티는 주기적으로 비용편익평가를 수행해야하며 잠재적인 효율이 가장 큰 데이터셋을 보존하고 접근가능 하도록 보장하기 위해 지속적으로 보존 프로토콜을 개발하고 세분화해야 함 ✓ 허용한 보존 프로토콜과 데이터의 철저한 문서화로 보존에 필요한 선택성을 확립하고 불필요한 중복을 줄여야 함 ✓ 특정 연구프로젝트 또는 데이터관리전문가조직의 비학술전문가와의 협력을 통한 전문화된 지원서비스는 연구데이터의 비용효율적인 생산, 사용, 관리 및 보관을 보장하기 위한 수단으로 고려할 필요가 있음 ✓ 연구자나 데이터베이스 제작자에게는 임금 및 승진 심사에서의 데이터 관리 활동의 인정을 포함하여 새로운 보상체계의 개발 등이 필요함 	OECD	미채택: 측정요소로 부적절
책임성	<ul style="list-style-type: none"> • 데이터 접근 성능은 사용자 그룹, 담당 기관 및 연구 자금 지원 기관의 정기 평가를 받아야 함 • 각 당사자는 다소 다른 평가기준을 사용할 가능성이 있지만, 결과의 총계는 데이터의 가치와 데이터접근구조에 대한 포괄적인 그림을 제공해야 함 • 이러한 평가는 개방형 접근에 대한 지원을 높이는데 도움이 됨 • 다음 사항을 고려해야 함 <ul style="list-style-type: none"> ✓ 연구데이터의 생산 및 관리에 대한 전반적인 공공투자 ✓ 데이터 수집 및 보관기관의 관리성능 ✓ 기존 데이터셋의 재사용범위 ✓ 기존 데이터를 재사용하여 생성된 지식 	OECD	미채택: 측정요소로 부적절
지속 가능성	<ul style="list-style-type: none"> • 공공자금이 지원된 연구데이터에 대한 접근은 지속성을 가져야 함 • 이는 장기보존이 필요한 것으로 판단된 데이터에 영구적으로 접근할 수 있도록 조처에 대한 관리책임은 지는 것을 의미함 	OECD	미채택: 측정요소로 부적절
학습성	<ul style="list-style-type: none"> • 쉽게 배울 수 있는지를 측정 	ISO9126, 정의	미채택: 측정요소로 부적절
운영성	<ul style="list-style-type: none"> • 최소한의 노력으로 사용할 수 있는지를 측정 	ISO9126	미채택: 측정요소로 부적절
매력성	<ul style="list-style-type: none"> • 인터페이스가 이용자 친화적인지를 측정 	ISO9126	미채택: 측정요소로 부적절
준수성	<ul style="list-style-type: none"> • 특정표준, 법률 등을 준수하고 있는지 측정 	ISO9126	미채택: 측정요소로 부적절
형식성	<ul style="list-style-type: none"> • 디지털 형식(format)인지 측정 	EPSRC	미채택: 재활용성에서 언급
연결성	<ul style="list-style-type: none"> • 데이터를 활용한 원 연구결과물과 연결을 유지해야 함 	EU	미채택: 접근성에서 언급
공유성	<ul style="list-style-type: none"> • 비용 없이 온라인상에서 활용 	OECD, EU, KISTI	미채택: 접근성, 재활용성에서 언급
사용 용이성	<ul style="list-style-type: none"> • 쉽게 이용할 수 있는지 측정 	ISO	미채택: 측정요소로 부적절

[부록 2] 설문지

다음은 연구데이터의 유용성 평가모델을 구성하는 각 지표들의 적합성을 평가하기 위한 것입니다. 제안된 모델은 FAIR* 원칙과 같이 연구데이터에 대한 정의와 기준을 제시하고 있는 사례, 오픈 데이터 관련 품질평가 모델 등 신뢰할 수 있는 기구, 기관에서 제시하고 있는 여러 기준들을 참조하여 구성하였습니다. 아래 그림은 연구데이터 유용성 평가 모델 전체를 도식화한 것입니다.

* FAIR: 연구데이터의 검색성(Findability), 접근성(Accessibility), 상호운용성(Interoperability), 재사용성(Reusability)을 향상시키기 위해 제시된 연구데이터의 공개 원칙



본격적인 설문 응답전에 평가모형의 구조 파악에 참조하시기 바랍니다.

응답시 각 측정지표가 해당 차원을 측정하기에 적합한 요소인지를 5점 기준으로 평가하여 제출하시면 됩니다.

1. 검색성(Findability)

검색성은 이용자들이 찾고자 하는 연구데이터에 대한 원활한 검색의 수행이 가능한 구조인가를 측정하기 위한 차원입니다. 검색성 측정을 위한 다음의 총 4가지 측정지표의 적합한 정도를 평가해 주십시오.

평가문항	평가척도				
	매우 부적합	부적합	보통	적합	매우 적합
측정지표1: 연구데이터 검색의 가능성을 높일 수 있는 적합한 메타데이터를 사용하고 있는가? 연구데이터 혹은 데이터셋의 원활한 검색을 위해 공인 기관에서 제시한 표준 용어 및 메타데이터를 사용하고 있는가?					
측정지표2: 연구데이터에 고유하고 영구적인 식별자를 할당하고 있는가? 각각의 연구데이터에 고유하고 영구적인 식별자를 사용하고 있는가?					
측정지표3: 연구데이터의 관리, 서비스에 적용된 메타데이터는 식별자 요소를 포함하고 있는가? 연구데이터에서 사용된 다양한 목적별 메타데이터 요소는 각각의 식별자 요소(예: 생성자 식별체계 - ORCID, ISNI, 과학기술인번호 등)를 포함하고 있는가?					
측정지표4: 연구데이터 및 메타데이터에 대한 색인이 이루어지고 있는가? 데이터 및 메타데이터는 시스템의 검색엔진에 의해 색인이 이루어지고 있는가?					

2. 접근성(Accessibility)

접근성은 이용자들이 연구데이터를 활용함에 있어 용이한 접근이 가능한지를 측정하기 위한 차원입니다. 접근성 측정을 위한 다음의 총 3가지 측정지표의 적합한 정도를 평가해 주십시오.

평가문항	평가척도				
	매우 부적합	부적합	보통	적합	매우 적합
측정지표1: 연구데이터에 자유로운 접근이 가능한가? 연구데이터의 서비스 및 활용을 위한 접근에 제한사항(IP제한, 인가자 접근 등)이 존재하는가?					
측정지표2: 연구데이터는 기계 가독형인가? 연구데이터는 별도 추가적인 처리 과정의 필요없이 즉시 기계처리가 가능한 형식인가?					
측정지표3: 연구데이터는 개방형 포맷(open format)인가? 연구데이터는 누구나 제한없이 접근해서 활용할 수 있는 형식인가?					

3. 상호운용성(Interoperability)

상호운용성은 연구데이터가 다양한 서비스와의 연계 및 유통이 가능하도록 데이터 및 메타데이터의 상호운용성을 확보하고 있는지 측정하기 위한 차원입니다. 상호운용성 측정을 위한 다음의 총 3가지 측정지표의 적합한 정도를 평가해 주십시오.

평가문항	평가척도				
	매우 부적합	부적합	보통	적합	매우 적합
측정지표1: 해당 연구데이터, 메타데이터는 지식표현(knowledge representation)이 용이한 형태를 유지하고 있는가? 연구데이터와 메타데이터는 데이터간, 메타데이터간 활용과 배포가 용이한 형태인가?					
측정지표2: 연구데이터와 메타데이터는 검색성(findability), 접근성(accessibility) 상호운용성(interoperability), 재활용성(reusability)의 관리 및 서비스 기준을 충족하는가? 연구데이터와 메타데이터가 누구나 접근, 검색, 활용이 가능하며, 목적 달성을 충족시킬 수 있는 구조를 갖추고 있는가?					
측정지표3: 연구데이터 플랫폼 운영조직에 기술적 상호운용성 보장을 위한 조직이 존재하는가? 혹은 관련 상호운용성 보장 연구를 지속적으로 수행하고 공표하고 있는가? 연구데이터 플랫폼 운영에 참여하는 기관은 기술적 상호운용성을 보장하기 위한 노력을 지속하고 있는가?					

4. 재활용성(Reusability)

재활용성은 이용자들이 자신의 연구에 연구데이터를 활용함에 있어 기술적, 제도적 제한에서 얼마나 자유로운지를 측정하기 위한 차원입니다. 재활용성 측정을 위한 다음의 총 10가지 측정지표의 적합한 정도를 평가해 주십시오.

평가문항	평가척도				
	매우 부적합	부적합	보통	적합	매우 적합
측정지표1: 연구데이터의 벌크 다운로드가 가능한가? 연구데이터의 일괄 다운로드가 가능한가?					
측정지표2: 연구데이터는 적시성, 최신성을 유지하고 있는가? 연구데이터에 대한 업데이트 주기 등을 확인할 수 있는 요소가 존재하는가?					
측정지표3: 연구데이터에 대해 상세한 설명을 하고 있는가? 연구데이터를 직접 다운로드 받기 전 이용자에게 연구데이터에 대한 충분한 정보를 제공하고 있는가?					
측정지표4: 연구데이터의 원 저작물의 출처를 포함하고 있는가? 연구데이터의 생산과 관련한 원본 저작물의 출처 정보를 제공하고 있는가?					
측정지표5: 별도 형식의 원천데이터가 존재하는 변환 데이터인 경우 변환과정에 대한 상세설명을 포함하고 있는가? 연구데이터에 비표준형태 혹은 비디지털형태의 원본이 존재하는 경우 변환 과정에 대한 상세설명을 포함하고 있는가?					
측정지표6: 데이터와 데이터셋을 기술하는 메타데이터는 해당 지식분야(연구커뮤니티)에서 활용하고 있는 표준 메타데이터인가? 해당 연구데이터는 지식 및 학문분야에서 표준으로 활용되고 있는 표준메타데이터를 사용하는가?					
측정지표7: 연구데이터가 해당 지식분야(커뮤니티)에서 일반적으로 활용하고 있는 형식(format)을 준수하고 있는가? 해당 연구데이터는 지식 및 학문분야에서 표준으로 활용되고 있는 표준 형식을 사용하는가?					
측정지표8: 연구데이터를 기술하는 메타데이터는 출처정보를 표현하기 위한 요소를 포함하고 있는가? 연구데이터의 출처정보를 확인할 수 있는 메타데이터 요소가 있는가?					
측정지표9: 연구데이터를 활용한 이용자의 의견(평가)을 수용할 수 있는 기술적 장치(댓글, 별점 등)가 존재하는가? 연구데이터 시스템 상 특정 연구데이터의 품질과 향후 개선 목적으로 이용자 의견 수렴이 가능한 댓글, 별점 등의 기술적 장치가 있는가?					
측정지표10: 연구데이터에 재활용의 제약없는 오픈 라이선스를 부여했는가? 연구데이터의 재활용이 가능한 라이선스를 부여하고 있는가?					

5. 이용 품질(Quality of use)

이용 품질 관련 지표들은 이용자가 서비스 시스템을 활용하면서 남긴 흔적을 정량적으로 측정하여 유용성 평가에 활용하기 위한 지표들입니다. 측정에 활용되는 데이터는 Google Analytics와 사용자 로그(log)로부터 수집됩니다.

평가문항	평가척도				
	매우 부적합	부적합	보통	적합	매우 적합
측정지표1: 참여수 - 실제 사용자(active user)의 확인을 위해 댓글, 별점 등이 부여된 연구데이터 랜딩페이지 수 측정 댓글, 별점이 부여된 연구데이터 랜딩페이지 수					
측정지표2: 참여율 - 전체 연구데이터 구축 건수 대비 실제 사용자(active user)에 의한 댓글, 별점 등의 결과로 (참여/이용)활성화가 확인되어지는 연구데이터의 비율 측정 댓글, 별점이 부여된 연구데이터 랜딩페이지의 수(참여수) / 전체 연구데이터 랜딩페이지(전체 연구데이터의 수)					
측정지표3: 다운로드 수 - 사이트 내 개별 연구데이터 실제 다운로드 건수 측정 개별 연구데이터의 다운로드 수					
측정지표4: 다운로드 비율 - 전체 연구데이터 구축 건수 대비 사이트 내 실제 다운로드 건수의 비율 측정 전체 연구데이터의 다운로드 수 / 전체 연구데이터 수					
측정지표5: 인용수 - 사이트 내 개별 연구데이터의 실제 인용 건수 측정 사이트에서 제공받은 연구데이터를 실제 논문, 프로젝트와 같은 학술연구활동 등에 인용하여 활용한 수					
측정지표6: 인용비율 - 전체 연구데이터 구축 건수 대비 사이트 내 실제 인용 건수의 비율 측정 전체 연구데이터의 인용 수 / 전체 연구데이터 수					
측정지표7: 공유수 - 소셜 미디어를 활용하여 공유되어진 개별 연구데이터의 건수 측정 연구데이터 시스템에서 제공하는 소셜 플러그인(페이스북, 트위터 등)으로 해당 데이터셋 페이지를 공유한 수					
측정지표8: 공유율 - 전체 연구데이터 구축 건수 대비 사이트 내 소셜 미디어로 공유되어진 연구데이터의 비율 측정 소셜 플러그인으로 공유된 연구데이터의 수 / 전체 연구데이터 수					
측정지표9: 열람수 - 개별 연구데이터의 다운로드 및 이용 가능한 페이지까지 도달한 방문자수 측정 개별 연구데이터의 다운로드 및 이용이 가능한 랜딩페이지까지 도달한 방문자 수					
측정지표10: 유효방문률 - 실제 연구데이터를 활용한 유효 이용자 수의 비율 측정 연구데이터 열람, 다운로드가 가능한 랜딩페이지까지 도달한 전체 방문자 수(전체 열람수) / 사이트 전체방문자 수					
측정지표11: 재방문수 - 연구데이터를 지속적으로 활용하는 이용자의 규모 측정 특정 기간(년, 분기, 월) 방문자 중 이전 기간(년, 분기, 월)에 방문을 한 방문자의 수					
측정지표12: 별점(직접입력) - 개별 연구데이터에 대해 이용자가 평가한 가치 및 활용성 정도를 별점 부여 방식을 통해 측정 이용자가 직접 연구데이터의 가치 및 활용성 정도에 대해 부여한 별점의 정도를 측정					

