

FAIR Principle-Based Metadata Assessment Framework

Park Jin Hyo[†] · Kim Sung-Hee^{††} · Youn Joosang^{†††}

ABSTRACT

Development of the big data industry, the cases of providing data utilization services on digital platforms are increasing. In this regard, research in data-related fields is being conducted to apply the FAIR principle that can be applied to the assessment of (meta)data quality, service, and function to data quality evaluation. Especially, the European Open Data Portal applies an assessment model based on FAIR principles. Based on this, a data maturity assessment is conducted and the results are disclosed in reports every year. However, public data portals do not conduct data maturity evaluations based on metadata. In this paper, we propose and evaluate a new model for data maturity evaluation on a big data platform built for multiple domestic public data portals and data transactions, FAIR principles used for data maturity evaluation in Europe's open data portals. The proposed maturity evaluation model is a model that evaluates the quality of public data portal datasets.

Keywords : Metadata, FAIR Principle, Data Managements

FAIR 원칙 기반 메타데이터 평가 프레임워크

박진효[†] · 김성희^{††} · 유주상^{†††}

요약

최근 빅데이터 산업의 발전으로 디지털 플랫폼에서 데이터 활용 서비스를 제공하는 사례가 증가하고 있다. 이와 관련해 데이터 관련 분야에서 (메타)데이터 품질, 서비스, 기능 등의 평가에 적용할 수 있는 FAIR 원칙을 데이터 품질 평가에 적용하여 활용하는 연구가 진행되고 있다. 특히, 유럽 오픈 데이터 포털에서는 FAIR 원칙 기반의 평가 모델을 적용하여 이를 기준으로 데이터 성숙도 평가를 시행하고 그 결과를 매년 보고서로 공개하고 있다. 이에 반해 공공데이터 포털에서는 메타데이터를 기반으로 한 데이터 성숙도 평가를 시행하고 있지 않다. 따라서 본 논문에서는 유럽 오픈 데이터 포털에서 데이터 성숙도 평가를 위해 사용되고 있는 FAIR 원칙을 국내 여러 공공데이터 포털 및 데이터 거래를 위해 구축된 빅데이터 플랫폼에 데이터 성숙도 평가를 위한 새로운 모델 제안하고 평가를 시행한다. 제안한 성숙도 평가 모델은 공공데이터 포털 데이터셋 품질을 평가하는 모델이다.

키워드 : 메타데이터, FAIR원칙, 데이터 관리

1. 서론

최근 다양한 산업 분야에서 빅데이터 기술들이 개발 중이다. 그중에서도 공공데이터와 산업 데이터를 수집하여 데이터 플랫폼을 구축하고 이러한 데이터 플랫폼을 통합 사이트로 활용하는 데이터 인프라를 구축 중이다. 이러한 데이터 통합 사이트에서 서로 다른 산업의 데이터를 결합하여 활용해 서비스를 제공하면서 새로운 서비스를 창출하려는 목적을 가진다. 정부에서도 이러한 데이터의 중요성을 인식하여 데이

터 산업의 육성을 위해 정책을 수립하고 사업을 추진 중이다 [1,2]. 이를 통해 빅데이터 거래·유통 플랫폼을 구축하면서 빅데이터 산업을 활성화하기 위해 노력했지만, 플랫폼의 활용도와 플랫폼 내의 데이터 활용도는 낮은 수준이다.

데이터 산업 분야에서 이러한 문제를 해결하고자 (메타)데이터 품질, 서비스, 기능 등의 평가에 적용할 수 있는 FAIR (Findable, Accessible, Interoperable, Reusable) 원칙[3]을 데이터 플랫폼 기능과 데이터 품질 평가에 적용하는 연구가 활발히 진행 중이다. 특히, 유럽 오픈 데이터 포털(data.europa.eu)에서는 데이터의 재사용성을 향상하기 위해 제안된 FAIR 원칙을 기반으로 메타데이터의 품질을 평가하여 데이터 포털의 메타데이터 평가지표를 상시로 공개하고 있다[4].

하지만 우리나라의 빅데이터 플랫폼에서는 (메타)데이터 성숙도를 평가하여 활용하고 있지 않다. 이는 데이터 공급자나 사용자의 데이터 활용도를 떨어뜨려 빅데이터 산업의 활

※ 이 논문은 과학기술정보통신부에서 시행한 해양수산 빅데이터 플랫폼 및 센터 구축사업과 2022 BB21+ Project 지원을 받아 수행된 결과임.

† 준회원 : 동의대학교 IT융합학과 석사과정

†† 정회원 : 동의대학교 산업융합시스템공학부 조교수

††† 종신회원 : 동의대학교 산업융합시스템공학부 교수

Manuscript Received : September 2, 2022

Accepted : October 24, 2022

* Corresponding Author : Youn Joosang(jsyou@deu.ac.kr)

성화를 저해하는 요인이 된다. 이를 해결하기 위해 FAIR 원칙을 기반으로 메타데이터 평가지표를 활용한다면 정부와 민간에서 활용하고 있는 빅데이터 플랫폼의 데이터의 품질을 올리기 위한 지표로써 활용할 수 있을 것이다.

본 논문에서는 우리나라 빅데이터 플랫폼에서 데이터의 품질을 향상을 위해 FAIR 원칙을 기반으로 한 메타데이터 평가 지표를 제시하여 우리나라에서 제공하는 빅데이터 포털에서 충분한 메타데이터를 제공하는지 평가한다[5].

2. 관련 연구

2.1 FAIR 원칙

FAIR 원칙은 2016년 연구데이터를 관리하기 위해 탐색성, 접근성, 상호운용성, 재사용성을 개선하고자 제안되었다. 또한 데이터 리소스, 도구, 어휘 및 인프라가 제삼자의 검색 및 재사용을 지원하기 위해 나타내야 하는 특성을 정의한다 [3]. 데이터의 양, 복잡성 및 생성 속도의 증가로 인해 사람이 데이터를 처리하기 위해 컴퓨터 지원에 점점 더 의존하고 있기 때문에, FAIR 원칙은 기계 작동 가능성(machine-actionability: 사람의 개입 없이 기계 스스로가 데이터를 찾고, 액세스하고, 상호 운용하고, 재사용할 수 있는 컴퓨터 시스템의 능력)을 강조한다[3]. Table 1은 FAIR 원칙과 그 세부 사항을 나타낸다.

이러한 4가지 특징과 핵심 요구사항을 기반으로 현재는 오픈 사이언스 측면으로 확산 중이다[8-11]. 특히, 데이터 포털 사이트에서는 데이터(셋)의 영구식별자, 메타데이터 규격, 데

이터 관리 정책, 데이터의 탐색을 위한 저장소 핵심 기능 등을 정의 할 수 있으며, FAIR 세부 원칙은 디지털 자원의 생산, 축적, 공유에 포괄적 평가가 가능하며 세부 평가지표로는 데이터 탐색성, 접근성, 상호운용성, 재사용성 대한 품질을 평가할 때도 적용 가능하다[12].

2.2 FAIR 평가 지표

FAIR 원칙은 데이터/메타데이터를 기계가 판별하도록 하여 데이터를 다양한 방면에서도 활용할 수 있도록 만들어졌다. 최근 데이터 관련 산업 분야에 데이터 품질, 서비스, 기능 등의 평가에 적용할 수 있는 원칙을 참조하여 데이터 플랫폼 기능과 데이터 품질 평가에 적용하는 연구가 활발히 진행 중이다.

유럽 오픈 데이터 포털에서는 FAIR 원칙을 적용한 데이터 성숙도 평가를 진행하고 있다[4]. FAIR 원칙을 기반으로 23 가지(탐색성 4개, 접근성 3개, 상호운용성 6개, 재사용성 6개, 맥락성 4개) 지표로 구성되어 있다. 이 지표를 항목별로 가중치를 부여해 4단계(Excellent, Good, Sufficient, Bad)로 (메타)데이터의 품질을 평가하고 있다. 평가는 지속해서 진행하여 결과를 상시로 게시하고 있으며 매년 데이터 포털 품질 보고서[13]를 작성하고 배포하며 메타데이터 품질 보증에 노력을 기하고 있다. 유럽 오픈 데이터 포털의 메타데이터 성숙도 평가지표와 그 지표에 대한 설명은 Table 2와 같다.

Table 2와 같이 유럽연합에서는 공공데이터의 관리를 위해 FAIR 원칙 기반의 평가지표를 만들어 데이터 접근성과 사용성을 확보하였다. 이에 반해 우리나라의 데이터 포털에서

Table 1. FAIR Principle[3, 6]

<p>Findable</p> <p>F1. (Meta)data are assigned a globally unique and persistent identifier</p> <p>F2. Data are described with rich metadata (defined by R1 below)</p> <p>F3. Metadata clearly and explicitly include the identifier of the data they describe</p> <p>F4. (Meta)data are registered or indexed in a searchable resource</p> <p>Accessible</p> <p>A1. (Meta)data are retrievable by their identifier using a standardised communications protocol</p> <p>A1.1 The protocol is open, free, and universally implementable</p> <p>A1.2 The protocol allows for an authentication and authorisation procedure, where necessary</p> <p>A2. Metadata are accessible, even when the data are no longer available</p> <p>Interoperable</p> <p>I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</p> <p>I2. (Meta)data use vocabularies that follow FAIR principles</p> <p>I3. (Meta)data include qualified references to other (meta)data</p> <p>Reusable</p> <p>R1. (Meta)data are richly described with a plurality of accurate and relevant attributes</p> <p>R1.1. (Meta)data are released with a clear and accessible data usage license</p> <p>R1.2. (Meta)data are associated with detailed provenance</p> <p>R1.3. (Meta)data meet domain-relevant community standards</p>

Table 2. Metadata Maturity Assessment Table for the European Data Portal[14]

Indicator	Description	Weight
Findability(Weight: 100)		
Keyword usage	Increase the findability of the data dataset	30
Categories	Categories help users to explore datasets thematically	30
Geo search	Usage of spatial information would enable users in order to find the dataset with a geo faceted search	20
Time based search	Usage of temporal information would enable users for a timely based faceted search	20
Accessibility(Weight: 100)		
AccessURL accessibility	Refers to a URL that gives access to the dataset or where more information about the dataset is available	50
DownloadURL	The downloadURL is a direct link to the referenced data	20
DownloadURL accessibility	If a downloadURL exists, the accessibility is checked	30
Interoperability(Weight: 110)		
Format	This field specifies the file format of the distribution	20
Media type	This field specifies the media type of the distribution	10
Format / Media type from vocabulary	Checks whether format and media type belong to a controlled vocabulary	10
Non-proprietary	Checks if the format of the distribution is non-proprietary	20
Machine readable	Checks if the format of the distribution is machine-readable	20
DCAT-AP compliance	This check is only performed if the metadata is originally harvested as DCAT-AP or as a valid derivate	30
Reusability(Weight: 75)		
License information	A license is valuable information for the reuse of data	20
License vocabulary	To limit the indication of incorrect license information	10
Access restrictions	This field indicates whether the access to the data is public or restricted	10
Access restrictions vocabulary	The use of a controlled vocabulary increases reusability	5
Contact point	The contact point contains inform whom to address in case of questions regarding the data	20
Publisher	It is checked whether the property is set or not	10
Contextuality(Weight: 20)		
Rights	Indicate the rights of a dataset	5
File size	Specifies the size of the file in bytes	5
Date of issue	The date on which the dataset or distribution was released	5
Modification date	The date on which the dataset or distribution was last changed	5

는 데이터 접근성과 사용성을 확보하기 위한 메타데이터 작성 기준과 평가지표가 미비하다. 그러므로 유럽 오픈 데이터 평가 모델을 국내에서 사이트별 특성을 반영하여 적용하기에는 어려움이 많다.

RDA에서는 FAIR 원칙이 적용된 데이터 성숙도 모델의 지표로 총 41개 (탐색성 7개, 접근성 12개, 상호운용성 12개, 재사용성 10개)의 항목으로 정의하고 있다[3,9]. 항목별로 지표 간의 중요성을 강조하고 공정성을 평가하기 위해 지표별 우선순위를 3단계(Essential, Important, Useful)로 구분했다[7]. 자세한 성숙도 평가표에 대한 설명은 Table 3과 같다.

3. 메타데이터 평가 모델

3.1 메타데이터 품질 평가 프레임워크

본 논문에서는 우리나라의 빅데이터 플랫폼에서 데이터를

통합(Integration)하고 데이터의 접근성(Accessibility) 강화 및 사용성(Usability) 증가를 유도하기 위해 메타데이터 품질 프레임워크를 제안한다. Table 4는 제안하는 FAIR 원칙 기반 메타데이터 평가 모델을 기술하고 있다. FAIR 원칙의 기본 원칙에서 탐색, 접근, 상호운영, 재사용성을 평가할 수 있도록 평가 모델의 지표를 구성하였다. 기본 원칙의 세부 고려 사항과 FAIR 원칙을 기반으로 한 품질 평가 연구를 참조하여 해당 지표의 중요성에 따라 총점을 부여했다. 지표의 세부 항목에 대해 충족한다면 해당 항목에 할당된 점수를 부여한다. 지표별 총점의 70%를 넘으면 데이터셋의 해당 지표가 만족한다고 평가한다.

논문에서 제시하는 평가 프레임워크는 다른 메타데이터 성숙도 평가 모델과는 다르게 빅데이터 플랫폼의 특성에 따라 항목별 점수를 변경할 수 있다. 이는 데이터 공급자와 사용자가 사이트 성격에 맞는 데이터셋을 쉽게 찾을 수 있도록 유도

Table 3. RDA Metadata Maturity Assessment Table[7]

ID	Indicator	Priority
Findability		
RDA-F1-01M	Metadata is identified by a persistent identifier	□□□
RDA-F1-01D	Data is identified by a persistent identifier	□□□
RDA-F1-02M	Metadata is identified by a globally unique identifier	□□□
RDA-F1-02D	Data is identified by a globally unique identifier	□□□
RDA-F2-01M	Rich metadata is provided to allow discovery	□□□
RDA-F3-01M	Metadata includes the identifier for the data	□□□
RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	□□□
Accessibility		
RDA-A1-01M	Metadata contains information to enable the user to get access to the data	□□
RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	□□□
RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	□□□
RDA-A1-03M	Metadata identifier resolves to a metadata record	□□□
RDA-A1-03D	Data identifier resolves to a digital object	□□□
RDA-A1-04M	Metadata is accessed through standardised protocol	□□□
RDA-A1-04D	Data is accessible through standardised protocol	□□□
RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	□□
RDA-A1.1-01M	Metadata is accessible through a free access protocol	□□□
RDA-A1.1-01D	Data is accessible through a free access protocol	□□
RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	□
RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	□□□
Interoperability		
RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	□□
RDA-I1-01D	Data uses knowledge representation expressed in standardised format	□□
RDA-I1-02M	Metadata uses machine-understandable knowledge representation	□□
RDA-I1-02D	Data uses machine-understandable knowledge representation	□□
RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	□□
RDA-I2-01D	Data uses FAIR-compliant vocabularies	□
RDA-I3-01M	Metadata includes references to other metadata	□□
RDA-I3-01D	Data includes references to other data	□
RDA-I3-02M	Metadata includes references to other data	□
RDA-I3-02D	Data includes qualified references to other data	□
RDA-I3-03M	Metadata includes qualified references to other metadata	□□
RDA-I3-04M	Metadata include qualified references to other data	□
Reusability		
RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	□□□
RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	□□□
RDA-R1.1-02M	Metadata refers to a standard reuse licence	□□
RDA-R1.1-03M	Metadata refers to a machine- understandable reuse licence	□□
RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	□□
RDA-R1.2-02M	Metadata includes provenance information according to a cross-community language	□
RDA-R1.3-01M	Metadata complies with a community standard	□□□
RDA-R1.3-01D	Data complies with a community standard	□□□
RDA-R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	□□□
RDA-R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	□□

□□□:Essential , □□: Important, □: Useful

Table 4. Proposed FAIR-based Metadata Maturity Assessment Model

Indicator	Description	Type	Score
Findability Assessment(Total: 120)			
Unique Identifier	(Meta)Data identifier	U	30
Keywords	(Meta)Data keywords for searching (e.g., region, time)	I	30
Data Basics	Provide creation/update date information for dataset in Metadata	U	30
Data Type	Data type of metadata	U	30
Accessibility Assessment(Total: 110)			
Access	Information to confirm providing access authority for dataset	A	20
Access Manual	Manual access function to (meta)data	A	30
Data Persistence	If data is unavailable, continue to provide information and data	I, A	30
DownloadURL	Direct download links to dataset	A	30
Interoperability Assessment(Total: 80)			
Open File Format	Standardized open file format	I, A	20
Standardised Format	Use standard vocabulary where metadata are compatible with other data	A	20
Machine-Readable	Use machine-readable expressions	I, A	20
Web Data Model	Information about the web data model	I, A	20
Reusability Assessment(Total: 100)			
License	(meta) Data provides licenses for datasets	I, A, U	30
License URL	(meta) data provides license links for datasets	I, A, U	30
Sources	(meta) data provides the source of the dataset	A	20
Contact Point	Information to inquiries about data	A, U	20

A: Accessibility, I: Integration, U: Usability

하여 데이터 활용도를 늘려 빅데이터 플랫폼 활성화에 도움을 준다. Table 4에서는 공공데이터셋을 제공하는 사이트들 기준으로 메타데이터 품질을 평가하여 각 항목의 점수를 부여하였다.

3.2 메타데이터 품질 평가

앞서 제안된 메타데이터 품질 평가 프레임워크를 기준으로 3곳(공공데이터포털, 통합데이터지도, 해양수산빅데이터 거점소)의 데이터 통합 플랫폼의 메타데이터의 품질을 평가했다. 비교한 사이트는 모두 공공데이터를 기반으로 한 사이트이므로 Table 4의 평가 점수를 기준으로 평가하였다. 평가는 사이트의 데이터셋의 정보를 확인해 평가 기준에 적합한지 여부를 확인. 공공데이터포털(data.go.kr)은 공공기관에서 수집된 공공데이터를 한곳에 모아 운영하고 있다. Table 5는 메타데이터 평가 모델을 활용해 공공데이터포털의 메타데이터를 평가한 결과를 보여준다.

공공데이터포털이 공공데이터에만 국한되어 있다면, 통합데이터지도(www.bigdata-map.kr)는 공공/민간 기관에서 수집된 데이터셋을 메타데이터 기반으로 통합하여 운영하고 있다. Table 6은 통합데이터지도 사이트의 메타데이터 품질을 평가한 결과를 보여준다.

Table 5. Assessment of Metadata Quality of Public Data Portal

Indicator	Score	Evaluation
Findability Assessment		
Unique Identifier	30	✓
Keywords	30	✓
Data Basics	30	✓
Data Type	30	✓
Accessibility Assessment		
Access	20	✓
Access Manual	30	✓
Data Persistence	30	
DownloadURL	30	✓
Interoperability Assessment		
Open File Format	20	✓
Standardised Format	20	✓
Machine-Readable	20	✓
Web Data Model	20	✓
Reusability Assessment		
License	30	✓
LicenseURL	30	✓
Sources	20	✓
Contact Point	20	✓

Table 6. Assessment of Metadata Quality of Bigdata-map

Indicator	Score	Evaluation
Findability Assessment		
Unique Identifier	30	
Keywords	30	✓
Data Basics	30	✓
Data Type	30	✓
Accessibility Assessment		
Access	20	
Access Manual	30	
Data Persistence	30	
DownloadURL	30	✓
Interoperability Assessment		
Open File Format	20	✓
Standardised Format	20	
Machine-Readable	20	
Web Data Model	20	
Reusability Assessment		
License	30	
LicenseURL	30	
Sources	20	✓
Contact Point	20	

통합데이터지도 사이트가 다양한 기관에서 실제 데이터셋 보다는 각 사이트의 메타데이터 기반으로 사이트를 운영하고 있다면, 해양수산빅데이터 거래소(www.bigdata-sea.kr)는 해양 수산과 관련된 빅데이터를 수집하여 운영 중이다. Table 7은 해양수산빅데이터 거래소의 메타데이터 품질을 평가한 결과를 보여준다.

이렇게 공공데이터 사이트의 메타데이터 품질 평가를 진행했다. Fig. 1은 품질 평가를 진행한 세 사이트의 점수를 보여준다. 공공데이터포털의 경우 실제 데이터셋에 대한 메타데이터 정보를 가장 많이 제공하고 있는 사이트로 메타데이터 품질 평가에서도 대부분의 세부 지표에 만족하며 높은 점수를 보여줬다.

하지만 다른 두 사이트는 공공데이터포털에 비해서 낮은 점수를 보여줬다. 통합데이터지도는 탐색성 평가지표를 제외한 평가지표에서 낮은 점수를 보였다. 접근성 평가 지표에서 액세스에 대한 정보와 데이터 지속성에 대한 정보가 미비하였다. 상호운용성 평가에서는 개방형 파일 형식을 지켰지만, 웹데이터 모델 관련 정보나 기계가 이해할 수 있는 표현을 사용하지 않는 등 다른 항목에 대해서는 부족한 모습이었다. 또한 재사용성 평가 지표에서는 라이선스에 대한 정보가 부족하고 데이터에 대한 문의를 할 수 있는 정보를 제공하지 않았다. 이는 통합데이터지도 사이트 내에서 제공하는 데이터셋의 정보가 원본 데이터셋이 있는 사이트보다 적어 메타데이터를 제대로 활용하고 있지 못한 문제로 보인다. 이는 사이트

Table 7. Assessment of Metadata Quality of BigdataSea

Indicator	Score	Evaluation
Findability Assessment		
Unique Identifier	30	
Keywords	30	✓
Data Basics	30	✓
Data Type	30	✓
Accessibility Assessment		
Access	20	
Access Manual	30	
Data Persistence	30	
DownloadURL	30	✓
Interoperability Assessment		
Open File Format	20	✓
Standardised Format	20	
Machine-Readable	20	
Web Data Model	20	
Reusability Assessment		
License	30	✓
LicenseURL	30	
Sources	20	✓
Contact Point	20	✓

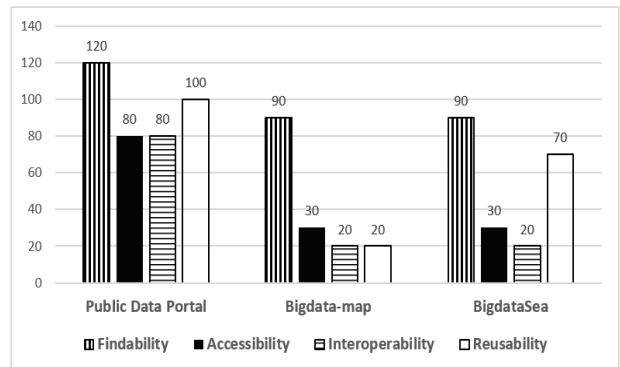


Fig. 1. Metadata Quality Assessment Comparison

가 추구하는 데이터 통합에 한계가 있어 보였다. 이러한 문제를 해결하기 위해 원본 사이트에서 제공하는 메타데이터를 통합데이터지도 사이트에서도 활용한다면 지금보다는 나은 데이터 품질을 보일 것이다.

해양수산빅데이터 거래소의 경우 사이트 내에서는 데이터셋의 사용성을 증가시키기 위해 다양한 메타데이터를 활용해 데이터 정보를 제공하기 위해 노력했다. 항목별로 보면 접근성과 상호운용성 지표에서는 빅데이터 지도와 같이 낮은 점수를 보여주었다. 액세스 정보를 제공하고 기계가 이해할 수 있는 표현을 사용하는 등 접근성과 상호운용성을 높이기 위한 개선이 필요할 것으로 보인다.

4. 결 론

본 논문에서는 FAIR 원칙을 기반 메타데이터 평가 모델을 제시하여 데이터셋의 데이터 통합, 접근성 강화, 사용성 증가를 유도할 방안을 제시했다. 이를 공공데이터포털, 통합데이터지도 사이트, 해양수산빅데이터 거래소에 적용해서 메타데이터 품질 평가를 진행하였다. 평가에서 좋은 점수를 받은 빅데이터 통합 사이트도 있지만, 우리나라에서 데이터(셋)를 제공하는 사이트에서의 메타데이터 품질이 낮아 사용자나 데이터 공급자가 원하는 수준의 데이터 통합의 한계가 있어 빅데이터 플랫폼의 데이터 활용도가 낮은 것으로 파악된다. 이를 위해 지속적인 메타데이터 품질 개선을 위한 노력이 필요해 보인다.

향후 제한한 메타데이터 품질 평가 프레임워크를 기반으로 자동화된 평가 도구를 개발하고자 한다. 평가 도구를 개발함으로써 국내외 빅데이터 플랫폼에서 간편하게 메타데이터 성숙도를 평가할 수 있을 것으로 기대한다. 이는 메타데이터 품질 향상으로 이어져 다른 산업 데이터와의 유기적인 결합이 원활해지면서 빅데이터 플랫폼의 데이터 활용도가 증가할 것으로 기대한다.

References

- [1] Open Data Strategy Council, “- Expansion of 「Public Data Provision Standards」 for Convenient Data Utilization”, Oct. 2021.
- [2] The 4th Industrial Revolution Committee under the direct control of the President, ‘Data platform activation plan: from system platform to user service platform,’ Jun. 2021.
- [3] “fair-principles” go-fair. n.d. [Internet], <https://www.go-fair.org/fair-principles>. accessed 10 Aug. 2022.
- [4] “Metadata Quality Dashboard,” data.europa.eu, n.d., [Internet], <https://data.europa.eu/mqa>. accessed 10 Aug. 2022.
- [5] J. H. Park, E. S. Seong, K. H. Kim, and J. S. Youn, “FAIR principle-based metadata quality assessment framework,” *Korea Multimedia Society Conference Papers*, Mar. 2022. <https://doi.org/10.3745/PKIPS.y2021m11a.491>.
- [6] M. Wilkinson et al., “The FAIR guiding principles for scientific data management and stewardship,” *Scientific Data*, Vol.3, Article 160018, 2016. <https://doi.org/10.1038/sdata.2016.18>
- [7] FAIR Data Maturity Model Working Group. FAIR Data Maturity Model. Specification and Guidelines. Jun. 2020. <https://doi.org/10.15497/rda00050>
- [8] N. P. Chue Hong et al., “FAIR Principles for Research Software (FAIR4RS Principles),” Research Data Alliance. <https://doi.org/10.15497/RDA00065>.
- [9] Katz, Daniel S., Pollard, Tom, Psomopoulos, Fotis, Huerta, Eliu, Erdmann, Chrisand Blaiszik, Ben, “FAIR principles for Machine Learning models,” Zenodo, 2020. <https://doi.org/10.5281/zenodo.4271996>.
- [10] Bahime, Christophe, Dekkers, Makx, Herczog, Edit, Russell, Keith, Stall, Shelleyand RDA FAIR Data Maturity Model Working Group, “Member survey on bridging the gap between funders and communities – perspectives on benefits and challenges of FAIR assessments V2.0,” Jan. 14, 2021. <https://doi.org/10.15497/RDA00061>.
- [11] D. S. Katz, F. E. Psomopoulos, and L. J. Castro, “Working towards understanding the role of FAIR for machine learning,” *Presented at the 2nd Workshop on Data and Research Objects Management for Linked Open Science (DaMaLOS 2021)*, Oct. 24, 2021. <https://doi.org/10.5281/zenodo.5594990>.
- [12] H. L. Kim, “FAIR principles: Considerations for implementing digital archives from a data perspective,” *Journal of Korean Society of Archives and Records Management*, Vol.21, No.2, pp.155-172, 2021. <https://doi.org/10.14404/JKSARM.2021.21.2.155>
- [13] FAIR Data Maturity Model Working Group, “FAIR Data Maturity Model. Specification and Guidelines,” Jun. 25, 2020. doi: 10.15497/rda00050.
- [14] “Metadata Quality Assessment Methodology,” data.europa.eu, n.d. [Internet], <https://data.europa.eu/mqa/methodology>. accessed 10 Aug 2022.
- [15] D. S. Katz, T. Pollard, F. Psomopoulos, E. Huerta, C. Erdmann, and B. Blaiszik, “FAIR principles for Machine Learning models,” Research Data Alliance Virtual Plenary 16 (RDA VP16). Zenodo. <https://doi.org/10.5281/zenodo.4271996>.
- [16] C. Bahim et al., “The FAIR data maturity model: An approach to harmonise FAIR assessments,” *Data Science Journal*, Vol.19, No.1, pp.41, 2020. <http://doi.org/10.5334/dsj-2020-041>.



박진효

<https://orcid.org/0000-0003-1646-6035>

e-mail : jhpark3679@gmail.com

2021년 동의대학교 응용소프트웨어공학
(학사)

2021년 ~ 현 재 동의대학교 IT융합학과
석사과정

관심분야 : 빅데이터, 인공지능, 강화학습, IT융합



김 성 희

<https://orcid.org/0000-0002-9716-8349>

e-mail : sh.kim@deu.ac.kr

2006년 이화여자대학교

컴퓨터정보통신공학과(학사)

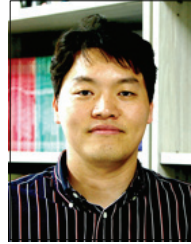
2008년 이화여자대학교

컴퓨터정보통신공학과(석사)

2014년 Purdue University 산업공학(박사)

2017년 ~ 현재 동의대학교 산업융합시스템공학부 조교수

관심분야: 데이터시각화, 기계학습, 인간-컴퓨터 상호작용



윤 주 상

<https://orcid.org/0000-0001-9952-9649>

e-mail : jsyoun@deu.ac.kr

2001년 고려대학교 전기전자전파공학부

(학사)

2003년 고려대학교 전자공학과(석사)

2008년 고려대학교 전자컴퓨터공학과(박사)

2008년~현재 동의대학교 산업융합시스템공학부 교수

관심분야: 사물인터넷, 엣지컴퓨팅, 클라우드, 강화학습