

[Oceanography] Note

해양과학 데이터저장소-JOISS는 FAIR한가?

송태윤^{1,*} · 이지윤¹ · 김우람¹ · 박소예나¹ · 노태근²

해봄데이터㈜, 서울 08503, 대한민국¹

한국해양과학기술원 해양기기개발·운영센터, 부산 49111, 대한민국²

Is the ocean science data repository, JOISS able to be FAIR?

Tae Yoon Song^{1,*} · Ji Yoon Lee¹ · Woo Ram Kim¹ · Soyeona Park¹ and Tae Keun Rho²

Haebom Data Inc., Seoul 08503, Republic of Korea¹

Instrumental Development and Management Center, Korea Institute of Ocean Science and Technology, Busan, 49111, Republic of Korea²

Received: 14 June 2022, revised: 8 July 2022, accepted: 13 July 2022

요약문 전 세계적인 오픈사이언스 운동은 최근 코로나 팬데믹 대응을 통해 그 효용성이 입증되면서 국내에서도 연구 데이터의 개방과 공유를 위한 연구분야별 또는 기관별 리포지토리 연구와 서비스 플랫폼 구축이 활발하다. 연구데이터 리포지토리의 목적은 데이터 관리 그 자체에 있기 보다는 지식 발견과 혁신, 후속 데이터 및 지식의 통합과 재사용에 있다. 이에 따라 최근의 리포지토리 관련 연구들은 관측에서부터 데이터 문서화(documentation), 데이터 결합, 품질관리, 데이터 출판에 이르는 협력적인 과정에서 FAIR 원칙의 구현을 강조하고 있다. 특히 다학제적인 협업연구가 필요한 해양관측에서는 리포지토리의 FAIR 구현을 통한 높은 수준의 데이터 상호운용성이 필수적이다. 국내에서는 해양과학데이터저장소 JOISS를 포함하여 여러 해양관측 기관들의 리포지토리가 있으나 아직까지 FAIR 원칙의 측면에서 데이터 리포지토리의 구축과 운영체계를 평가한 연구는 전무하다. 따라서 본 연구에서는 데이터 수명주기적 측면에서 JOISS 리포지토리의 구축과정, 데이터 관리현황 그리고 그 웹플랫폼의 주요 기능과 서비스를 살펴보고, 이러한 운영체계에서 오픈사이언스의 FAIR 원칙이 어떻게 작동하는지 파악하고 그 한계점을 평가하고자 하였다. 본 연구는 해양관측의 진화와 함께 데이터의 다양성과 볼륨이 급증하는 환경속에서 JOISS를 비롯한 국내 해양 리포지토리들의 데이터 관리와 서비스 개선방향에 대한 시사점을 제공한다.

주요어: 오픈사이언스, 데이터 리포지토리, JOISS, FAIR, 해양관측

Abstract As the global open science movement has recently proven its effectiveness in responding to the corona pandemic, research on disciplinary or institutional data repositories and establishing service platforms for the open and sharing research data are also active in Korea. The purpose of the research data repository is not to manage data per se but to discover and innovate knowledge and to integrate and reuse subsequent data and knowledge. Therefore, recent repository-related studies emphasize implementing the FAIR principle in this collaborative process, from observation to data documentation, data combination, quality control, and data publication. In particular, high-level data interoperability through the FAIR implementation of the repository is

*Corresponding author: tysong@haebomdata.com

essential for ocean observation that requires multidisciplinary collaborative research. In Korea, ocean observatory organizations have repositories, including the ocean science data repository, JOISS; however, no studies evaluate the establishment and operation of data repositories in the FAIR principle. Therefore, this study aims to examine the construction process and data management status of the JOISS repository and the main functions and services of the web platform in terms of the data lifecycle and evaluate The FAIR principle of Open Science works in such an operating system and its limitations. The study provides implications for the improvement direction of data management and services of domestic marine repositories, including the JOISS, in an environment where the diversity and volume of data are rapidly increasing along with the evolution of ocean observation.

Keywords: Open Science, data repository, JOISS, FAIR, ocean observation

1. 서론

오픈사이언스 정책은 공공기금 지원으로 창출된 연구성과물을 공개·공유함으로써 그 성과를 확산하고 극대화하기 위해 2010년대 중반부터 유럽, 미국을 비롯하여 OECD, UNESCO 등을 중심으로 구체화되기 시작했다. 요컨대, 오픈사이언스 정책의 핵심은 공공기금으로 지원된 연구의 성과물(출판물 및 데이터)을 디지털 방식으로 공개하여 이용자가 보다 쉽게 접근하고 활용할 수 있도록 하는데 있다(신은정 외, 2016). 이를 실현하기 위해서는 연구데이터의 재사용성과 재현성 확보를 위한 데이터 리포지토리(data repository)가 필수적이다. 국내에서도 최근 수년간 연구데이터 리포지토리에 대한 연구와 구축 노력이 활발하게 일어났고, 특히 covid-19로 인한 팬데믹 상황에서 오픈사이언스 정책의 효용성이 확인되면서 공공정책으로 더욱 가속화되고 있다(Besançon et al., 2021). 해양과학데이터저장소 JOISS (이하 JOISS)는 2017년부터 오픈사이언스 정책에 기반하여 해양분야 리포지토리(ocean disciplinary repository)로서 서비스를 시작하였다(<https://joiss.kr>). JOISS의 주요 서비스는 연구자들이 국가연구개발과제 수행으로 생산한 해양조사 및 관측데이터와 국내·외 공공기관이 생산한 해양관측데이터 등을 수집하여 1) 표준화 및 품질관리, 2) 식별화 및 검색·조회, 3) 데이터 큐레이션(curation), 4) 중개(brokerage) 서비스를 수행하고 있다. 특히 해양데이터 공급이 거의 생산기관 위주로 기관간 연계성이 없는 우리나라 실정에서 JOISS의 데이터 큐레이션과 중개의 역할은 다학제적인 (multidisciplinary) 해양데이터들을 결합·융합으로써 데이터 생태계 구성에 기여할 수 있다는 점에서 기존 기관 리포지토리와 차별화된다. 우수한 연구데이터 관리는 그 자체가 목표가 아니라 지식 발견과 혁신, 후속 데이터 및 지식의 통합과 재사용으로 이어지는 핵심 과정이다. 따라서 이러한 데이터 관리의 목적을 위해서는 관측에서부터 데이터 문서화(documentation), 데이터 결합, QA/QC, 데이터 출판에 이르기까지, 이러한 협력적인 과정에서 FAIR 원칙을 구현하는 것이 매우 중요하다(Silva et al., 2017; Tanhua et al., 2019). Wilkinson et al. (2016)의 '과학데이터 관리를 위한 FAIR 데이터 원칙 지침'에 따르면, FAIR한 데이터는 검색 가능하고(Findable), 접근 가능하며(Accessible), 상호운용 가능하면서(Interoperable), 재사용 가능(Reusable)한 4가지 특성을 만족하도록 하는 포괄적인 개념을 담고 있다. 그러므로 FAIR 원칙의 구현은 연구프로젝트 수명주기의 모든 단계에서 연구자의 다양한 데이터 큐레이션 활동을 촉진할 수 있다. 일반적으로 해양관측은 다학제적이고 대규모로 협업연구가 빈번한 분야의 특징을 갖는다. 이러한 특징때문에 해양관측분야가 그 어느 분야 보다 FAIR의 구현을 통한 데이터의 상호운용성에 매우 큰 관심을 가져왔다(Silva et al., 2017; Tanhua et al., 2019). 하지만 아직까지 국내에서는 FAIR의 측면에서 연구데이터 리포지토리 구축과 운영과정을 평가한 연구는 알려진 바 없다. 본 연구에서는 데이터 수명주기적 측면에서 JOISS 리포지토리의 구축과정과 데이터 관리현황 그리고 그 웹 플랫폼의 주요 기능을 살펴보았다. 또한 이러한 운영체계에서 오픈사이언스의 FAIR 원칙이 어떻게 작동하는지 평가하고 그 한계점을 파악하고자 하였다.

2. 본론

2.1 데이터 표준화

JOISS가 국내 해양연구자와 관측데이터 생산기관으로부터 수집한 데이터를 살펴보면 그 데이터 구조와 포맷이 서로 다르고 용어를 혼용하거나 오용하고 데이터세트명을 무의미하게 짓는 경우가 많으며 메타데이터는 아예 제공하고 있지 않다. 따라서 이러한 다종다양한 해양관측데이터를 기반으로 국내·외 해양데이터 센터들을 통합하기 위해서 데이터 표준(용어 및 메타데이터 표준)과 유통포맷(데이터 구조, 파일포맷, 데이터세트 명명규칙)을 마련하는 것이 선결되어야 한다(National Ecological Observatory Network, 2020). 용어의 경우 빈번하게 혼용과 오용이 생기는 관측분야(discipline), 항목(variables), 장비(equipment), 해역(sea region), 플랫폼(platform), 단위(unit)에 대하여 시스템과 사용자 모두 일관성 있게 사용할 수 있도록 표준화 및 체계화를 진행하였다(Table 1).

Table 1. Summary of relevant ocean variables and retained data by discipline in the JOISS

Discipline	Ocean Variables	Number of Data (1,000)
Physical Oceanography	Water column temperature and salinity, Acoustics, Currents, Optical properties, Sea level, Waves etc. (38 in total)	290,056
Chemical Oceanography	Dissolved Gases, Carbon, nitrogen and phosphorus, Nutrients, PCBs and organic micropollutants, Isotopes etc. (79 in total)	63,966
Biological Oceanography	Bacteria and viruses, Biota abundance, Biomass and diversity, Pigments etc. (86 in total)	7,945
Marine geology	Field geophysics, Geothermal measurements, Gravity, magnetics and bathymetry, Rock and sediment chemistry, Rock and sediment lithology and mineralogy, Rock and sediment physical properties, Rock and sediment pore water chemistry, Suspended particulate material etc. (195 in total)	6,649
Atmosphere	Atmospheric chemistry, Meteorology etc. (21 in total)	125,605

해양지리정보 메타데이터 표준 설계는 지리정보 메타데이터 표준인 ISO19115를 기반으로 국내·외 해양데이터 센터(예: SeaDataNet, CSIRO)의 메타데이터 XML (eXtensible Markup Language)을 참고하였다. 이를 바탕으로 국내 관측데이터에 적합한 메타데이터 패키지와 개체정보를 도출하고 이를 JOISS의 메타데이터 XML 작성에 적용하였다(Boldrini & Naivi, 2019). 특히 기존 ISO19115 메타데이터 스키마의 식별정보(IdentificationInfo)에 연락처(pointOfContact), 기술주제어(descriptiveKeywords), 자원제약(resourceConstraints) 개체정보를 추가하여 보완하였다. 연락처 개체정보는 공동조사와 관측이 많은 해양데이터의 특성 상 다양한 기관과 연구자에 대하여 세분화된 역할코드(CI_RoleCode)를 부여하여 기술할 수 있었다. 기술주제어 개체정보에서는 관측항목, 장비, 해역, 플랫폼 등에 대한 상세 주제어를 사용하여 데이터세트를 설명하도록 하였고, 자원제약 개체정보를 통해 데이터세트 이용자에게 직관적으로 저작권 정보를 제공하였다(Fig. 1). 데이터세트 구조는 데이터 프레임 상단에 필수적인 데이터 설명과 컬럼 헤더를 배치하고, 데이터 프

레이미에는 메타성 데이터, 기본변수, 데이터변수의 값과 플래그를 쌍으로 배치하였다(Fig. 2). 이러한 구조와 함께 관측 데이터 파일포맷은 텍스트형 오픈 포맷인 CSV를 사용함으로써 오픈 소프트웨어인 ODV (Ocean Data View)에서 쉽게 불러올 수 있도록 하였다. 이용자에게 제공되는 데이터세트는 그 이름 자체에서 데이터의 맥락을 직관적으로 식별할 수 있도록 하기 위해서 명명규칙을 사업명, 년도, 데이터 형식, 도메인 분야, 장비로 조합하여 생성되도록 하였다(예를 들어, 사업명_년도_데이터 형식_분야_장비.csv).

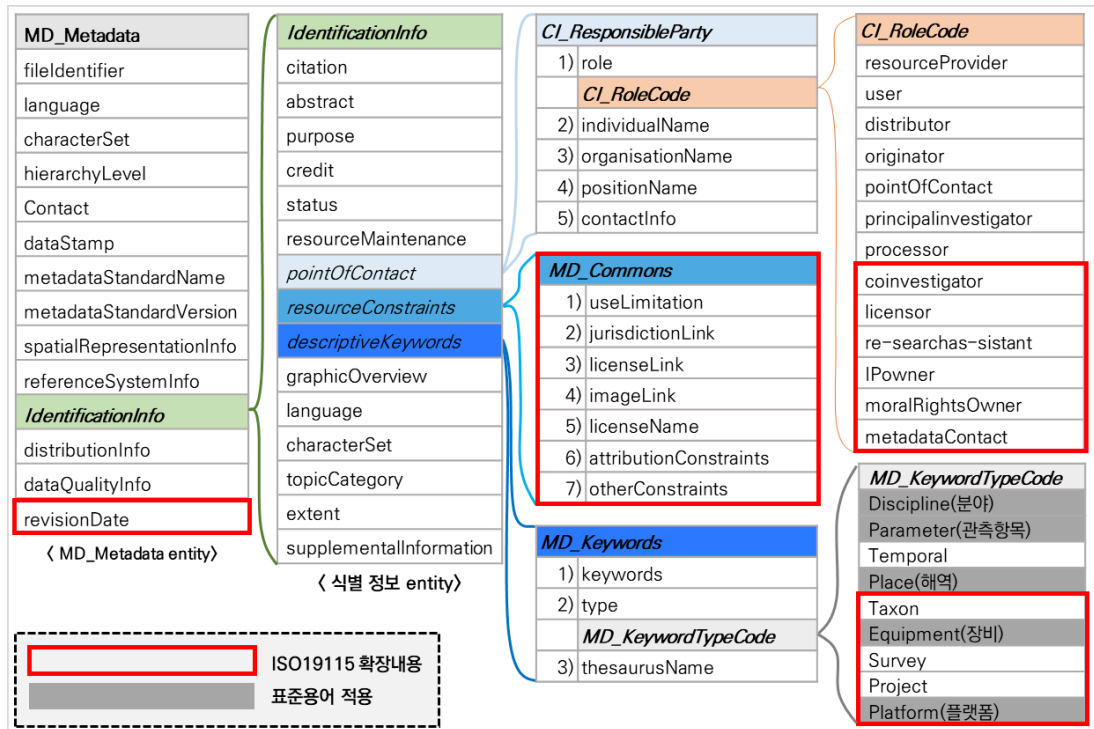


Figure 1. Metadata standard of the JOISS system that extends the metadata schema and vocabularies based on ISO19115

project_name	station	yyyy-mm-dd hh:mm:ss	Longitude [degrees_east]	Latitude [degrees_south]	botdepth[m]	수심[m]	염분	sal_qc	수온[°C]	wtrmp_qc	potential density-1000[kg/m³]	ptden_qc
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:00	129.4941667	37.2481667	0	0						
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:37	129.4941667	37.2481667	0	8.84	33.07		16.52	1	24.15	2
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:38	129.4941667	37.2481667	0	19.73	33.9		9.07	1	26.24	2
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:38	129.4941667	37.2481667	0	29.79	33.95		7.65	1	26.5	2
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:39	129.4941667	37.2481667	0	49.39	34.02		4.6	1	26.95	2
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:40	129.4941667	37.2481667	0	74.26	34.01		2.76	1	27.11	2
2006년 7월 관측역역 집중조사	B1	2006-07-13 0:40	129.4941667	37.2481667	0	100.35	34		2.28	1	27.15	2

Figure 2. Structure of data frame in the dataset files served via JOISS web platform

2.2 데이터 품질

JOISS는 데이터 생산자가 아닌 중개자이기 때문에 JOISS의 데이터 정책은 데이터 품질에 대한 책임은 지지 않는다고 밝히고 있다(JOISS, 2020). 중개자로서 과도한 데이터 품질처리는 오히려 품질을 훼손시킬 우려가 있다고 보기 때문에 자동화가 어렵고 도메인 지식에 의존하는 scientific QC의 경우 연구자의 몫으로 두었다. 하지만 실제 등록된 연구 데이터에는 많은 품질오류가 포함되어 있기 때문에 이러한 오류 데이터에 대하여 플래그를 부여하는 방식으로 JOISS 자체의 품질 프로세스를 마련하고 이용자로 하여금 품질 신뢰성을 제고하도록 하였다. JOISS의 품질관리 정책 및 절차는 WOD, SeaDataNet 등에서 국제적으로 통용되고 있는 해양데이터 품질관리체계를 참고하여, JOISS의 데이터 관리·공유 프로세스 특성에 적합한 체계를 구축하였다(Garcia et al., 2018; SeaDataNet, 2010). Fig. 3과 같이 데이터 품질 테스트 및 플래그 처리절차를 살펴보면, 데이터 수집 이후 1차 QC로 1) 날짜와 시간 형식, 2) 위경도 좌표형식, 3) 위치 점검(on land test)을 필수로 수행하고 있고, 데이터베이스에 저장 후 2차 QC로 4) 범위(ranges) 테스트, 5) 튀값(spike) 테스트, 6) 변화도 테스트를 선택적으로 적용하고 있다. 특히 범위 테스트는 전지구적 범위와 지역적(동해, 서해, 남해) 범위의 두 단계에 걸쳐 실시하는데, 그 설정범위는 WOD, SeaDataNet, 장비의 측정한계치, JOISS가 보유한 전체 관측 데이터로부터 도출하여 사용하고 있다.

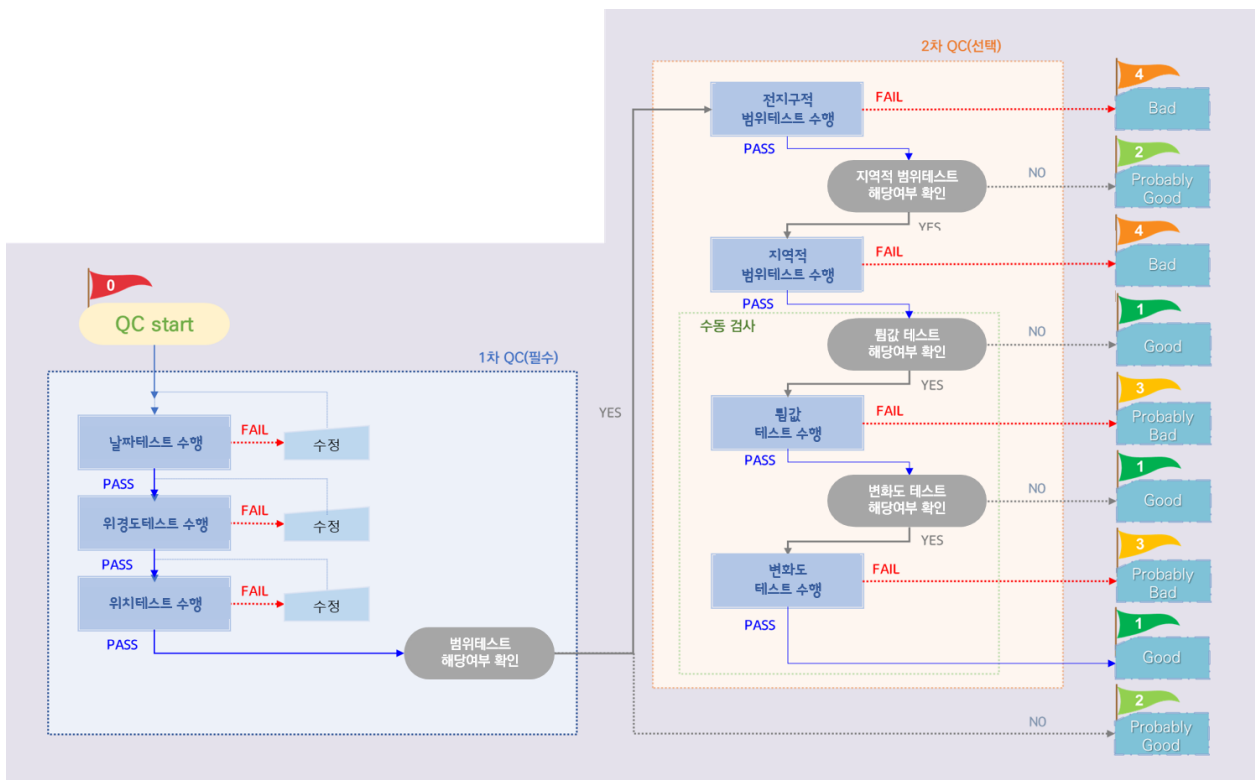


Figure 3. Workflow of 1st & 2nd quality controls and flagging procedures in the JOISS

최근 국내에서는 해양화학데이터의 정도관리가 강화되면서 QC에 대한 관심이 높아지고 있다. 하지만 해양물리데이터에 비하면 아직까지 자동화된 QC 처리기법은 충분히 알려져 있지 않다. JOISS는 이용자들의 품질 요구를 충족하기 위해 현재 영양염류 데이터에 대해서는 화학성분들 간의 상호관계(염분-화학성분, 수심-화학성분, 화학성분-화학성분) 특성을 이용하여 화학자료의 특이값을 찾아내는 품질처리를 추가적으로 진행하고 있으며, 이 데이터세트에 대해서는

추후 별도로 서비스할 예정이다. 이러한 내부 데이터 품질관리와는 별도로 한국데이터산업진흥원의 '데이터 품질인증 제도'를 활용하여 객관적인 대외 품질검증을 실시하였으며(한국데이터산업진흥원, 2010), 그 결과 JOISS의 DB 품질지수는 국내 평균(94.79%, 3.13 σ) 보다 높은 99.998% (5.71 σ)로 Platinum Class에 해당하였다.

2.3 데이터 큐레이션 및 데이터 현황

데이터 큐레이션이란 데이터 발견, 추출, 품질, 가치 부여를 통한 재사용성을 증대시키는 활동이라고 정의할 수 있다 (University of Melbourne Library, 2021). 여기에는 과거 문헌으로 존재하는 해양자료의 복원도 포함될 수 있다. JOISS는 자체 수립한 데이터 관리정책, 품질관리 정의서, 데이터 표준가이드에 따라 큐레이션 활동을 수행하였다. 그 결과 8개 국가연구개발과제의 연구데이터를 비롯하여 국립수산과학원, 국립해양조사원, 해양환경공단 등에서 생산하는 공개자료 9종에 대해 약 5억 건 이상의 데이터를 수집하였고 주기적으로 데이터를 최신화하고 있다(Table. 1). 해양화학데이터의 경우 우리나라 주변해역의 해양환경변화를 지속적으로 감시하기 위해서 국립수산과학원, 해양환경공단, 대학 및 연구기관이 정기적으로 또는 연구과제별로 다양한 항목을 관측하여 생산하고 있는데, 이러한 해양화학 데이터들은 기관별 고유 목적에 따라 수집자료의 종류, 명칭, 단위 등이 서로 다른 경우가 매우 많다. 따라서 이용자의 편의를 위해서 먼저 관측항목의 명칭을 통일하고 다양하게 사용된 농도단위를 $\mu\text{mol/L}$ 로 변환하여 통합하였다. 그 밖에도 가로림해역처럼 과거 장기간에 걸쳐 종이매체로 남아 있는 해양관측자료는 기후변화와 해양환경변화 연구에 있어서 중요한 가치를 갖는다. 이러한 자료는 해양자료유산의 복원이라는 측면에서 데이터와 메타데이터를 디지털로 변환, 추출하여 영구보존 작업을 수행하였다.

3. 결론 및 토의

3.1 리포지토리의 FAIR

연구데이터의 재현성과 투명성을 확보하기 위해서는 데이터 관리에 있어서 FAIR 원칙의 준수가 매우 중요하다. FAIR 원칙이란 데이터가 검색, 접근, 상호운용, 재사용 가능(Findable, Accessible, Interoperable, Reusable)하도록 하기 위한 일련의 지침을 의미한다(Wilkinson et al., 2016; FORCE11, 2016; Helliwell et al., 2019; GO FAIR, 2021). 구체적으로 보면, 검색가능한 데이터를 위해서는 데이터에 고유하고 영구적인 식별자(예를 들어 DOI, Handle, ARK)가 할당되고, 데이터를 충실히 기술한 메타데이터로 검색가능한 자원(포털)에 색인화 되어야 함을 말한다. 접근가능한 데이터를 위해서는 표준화된 통신 프로토콜(예: HTTP, HTTPS, FTP 등)을 통해 식별자로 검색되고, 권한 인증이 필요한 경우(예: 개인 정보데이터, 민감데이터, 옴바고 지정 데이터)에는 사용자계정을 통해 인증한 후 데이터세트에 접근하며, 이때 데이터 접근이 불가능한 경우라도 메타데이터 접근은 보장되어야 함을 말한다. 상호운용가능한 데이터는 메타데이터와 데이터 세트 모두 기계가 처리할 수 있고(machine-actionable), 파싱이 가능하고(parsable), 공식적이고 공통적으로 적용가능한 통제어휘, 온톨로지를 사용하는 것을 의미한다. 마지막으로 재사용가능한 데이터는 메타데이터에 정확하고 관련성 높은 속성을 포함하고, 재사용을 위한 라이선스를 제공하며, 도메인 관련 커뮤니티의 표준을 사용하는 것을 말한다(Fig. 4).



Figure 4. Elements of 'FAIR principles' needed to make data FAIR

Fig. 5과 같이 JOISS에서 수집된 데이터는 정제와 품질처리를 거쳐 저장된 후, 웹 플랫폼을 통해 공유된다. 이러한 일련의 데이터 관리 프로세스와 더불어 웹 플랫폼의 다양한 연계와 공유 기능은 FAIR의 핵심요소가 될 수 있다(Boeckhout et al., 2018).

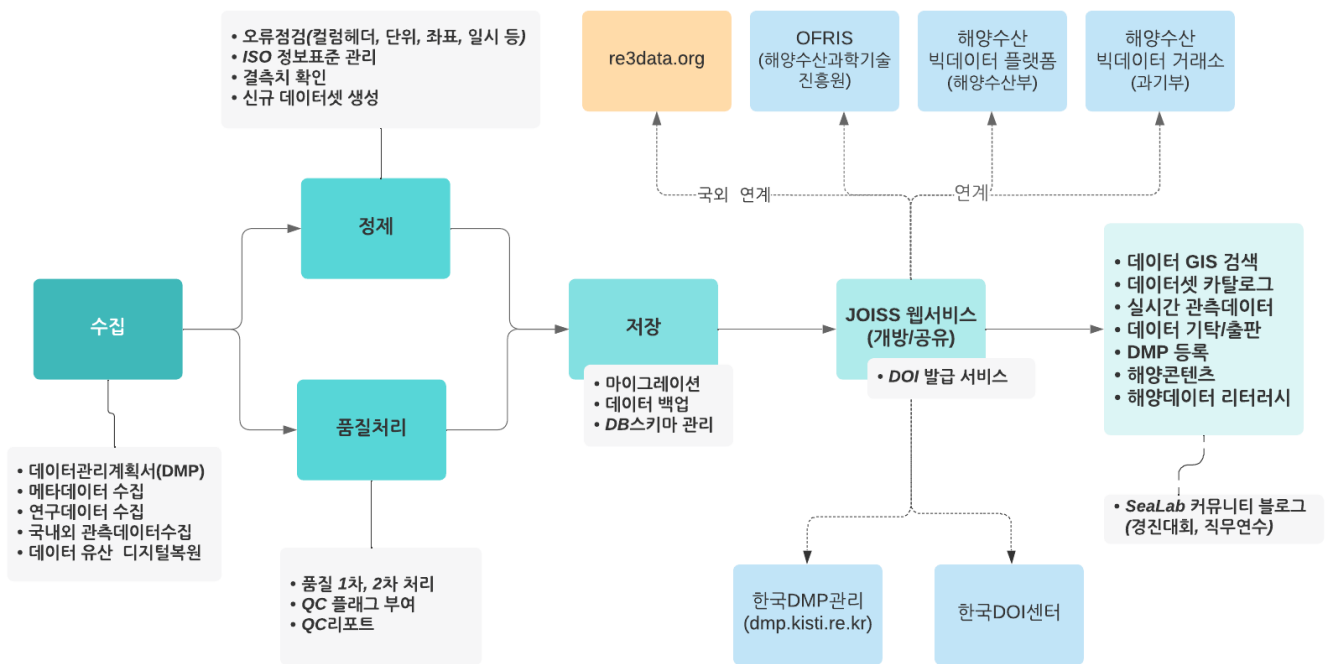


Figure 4. Data management processes and web platform functionality in the JOISS

예를 들어 한국DOI센터와 연계하여 데이터셋 및 데이터 관리 계획서에 대한 디지털 객체식별자(Digital Object Identifier, DOI)를 부여함으로써 데이터 자원을 식별하고, 웹 검색기능으로 지리공간정보(GIS) 검색방식인 '관측자료 검색'과 메타데이터 브라우징 방식인 '메타검색'을 통해 검색가능성을 향상시켰다. 또한 시스템 접근 인증과 이용 승인의 측면에서 개방과 라이선스 보호라는 양립하기 어려운 문제를 해소하기 위해 사용자계정을 ResearcherID와 ORCID를

활용하여 일반이용자, 연구자, 데이터 기탁자로 구분하고, 기능적 접근과 권한을 부여함으로써 접근가능성을 높이고자 하였다. 이밖에 상호운용성과 재사용성 향상을 위해 데이터세트 파일은 기본 메타정보와 품질처리 정보를 담아 사용자 친화적인 데이터프레임으로 구조화하였고, 확장된 메타데이터 스키마와 표준용어를 적용하여 해양관측데이터의 특성이 잘 반영되도록 메타정보를 XML로 구축·제공하였다(Fig. 1, Fig. 2). 또 Tim (2012)가 제안한 오픈데이터 발전모델에 따라 현재 JOISS의 데이터 유통포맷은 특정 소프트웨어에 의해 종속되지 않도록 CSV, TXT같은 개방형 파일포맷으로 제공하고 있다(Fig. 6). 이 밖에 데이터기탁 과정에서 기탁자에게 저작권 정보 및 인용에 관한 크리에이티브 커먼즈 라이선스(Creative Commons License, CCL)을 권고하고 있으며, 이용자에게도 CCL을 JOISS 정책으로 제시하고 있다.

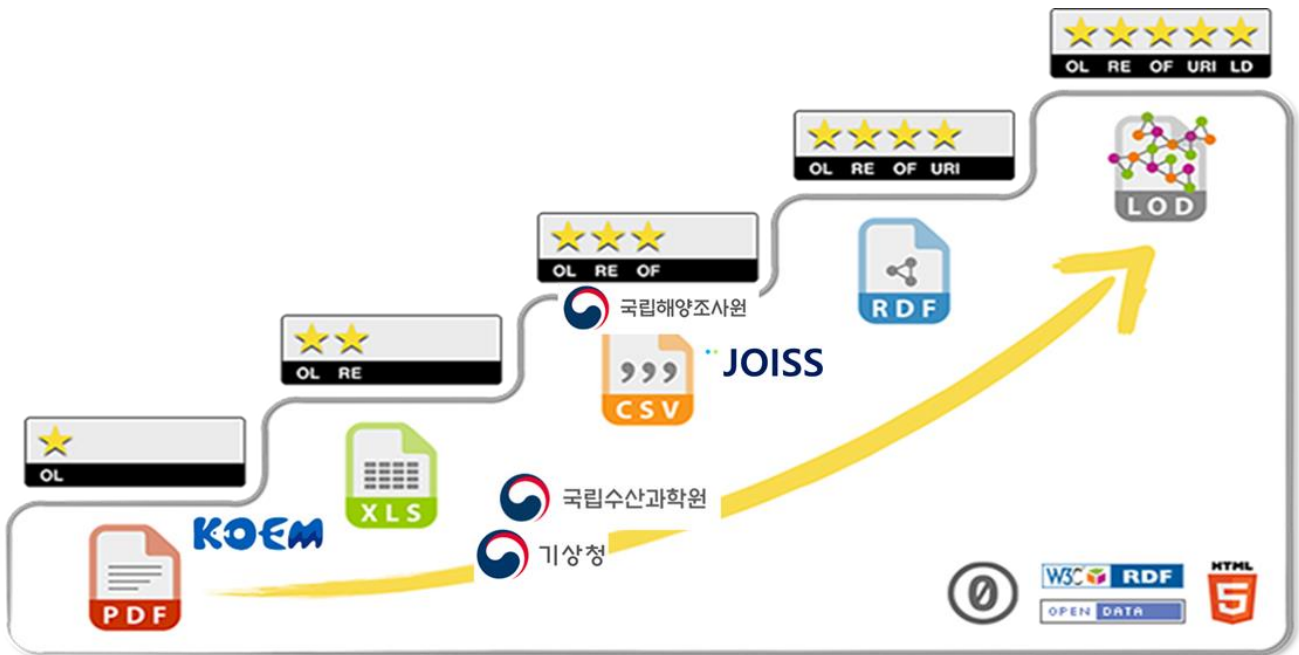


Figure 5. JOISS's efforts to reach a 5-star deployment scheme for Open Data by providing an open file format (JOISS is equivalent to 3-star.)

JOISS는 이러한 리포지토리 서비스 구축으로 리포지토리 국제인증과 색인서비스를 제공하는 re3data.org에 2021년 등재되었다(<http://doi.org/10.17616/R31NJMXU>). Pampel et al. (2013)의 re3data.org 등재를 위한 권고 또는 요건사항에 대하여 이에 해당하는 JOISS의 적용사항은 Table. 2와 같다. 현재 re3data.org에는 전세계 2,874개 리포지토리가 색인되어 있고, 이 가운데 국내 해양분야 리포지토리는 한국극지데이터센터와 JOISS가 유일하다.

Table 2. Requirements for registration on re3data.org and the corresponding application of the JOISS

등록요건	설명	JOISS 적용내용
General Information	(리포지토리)기본정보, 인증 및 표준, 데이터 정책, 접근권한, 라이선스 등이 확인가능한 페이지의 유무	해당 정보에 대한 한글 및 영문 페이지 별도제공
Access	데이터베이스 및 데이터 접근권한에 대한 세분화 기준 유무	일반사용자, 연구자, 데이터 기탁자로 구분

Licenses	데이터 및 데이터베이스 라이선스 적용 여부 (CCL, copyright 등)	CCL 적용
Pid systems	영구 식별체계 적용 여부 (DOI, ARK, hdl 등)	DOI 적용
Aid systems	연구자 식별체계 적용 여부 (ORCID, ResearcherID 등)	ORCID 적용
Policy	연구데이터 관리 활동과 관련된 정책수립 유무	데이터 큐레이션, 품질관리, 데이터 기탁 등 정책 수립 및 제시
Certificates and Standards	API적용 (OAI-PMH, SPARQL 등)	없음
	데이터 모델 및 메타데이터 표준 적용	ISO19115/19139 적용
	데이터 버전 관리 수행 여부	없음

최근 해양관측은 다양한 자동화 플랫폼과 센서들의 등장과 함께 빠르게 진화하고 있다. 이러한 변화속에서 핵심해양변수들(EOV)이 확대되고 데이터 다양성과 볼륨이 급격하게 증가하고 있다. 따라서 데이터 작업흐름의 자동화와 효과적인 표준화의 요구는 데이터 수명주기에 걸쳐 점점 더 커지고 있으며, 이는 물리관측에서부터 생·지화학 관측으로 까지 모든 데이터 유형들로 확대되는 추세다. 현재 JOISS가 보유하는 데이터는 해양과학의 전반적인 주제분야를 다루고 있으나 지연모드(Delayed-mode) 해양관측데이터가 주를 이루고 있다. ISO19115 메타표준은 이러한 데이터를 관리하는데 비교적 유용하지만 생·지화학 및 생물(또는 생물지리학)·생태학 자료의 특성까지 충분히 설명하기에는 어려움이 있다. 자동화 플랫폼과 센서 기반의 데이터 유형 역시 장비특성을 설명하고 스트리밍 데이터를 관리하는데 적합하지 않다. 이러한 다양한 유형의 데이터 표준을 해결하기 위해 해외 해양데이터 리포지토리들은 Darwin Core, Ecological Model Language, Sensor Model Language을 채택하려는 시도가 늘어나고 있다(Partescano et al., 2017). 또한 해양데이터 관리의 측면에서 데이터의 상호운용성은 FAIR의 원칙을 관통하는 매우 중요한 원칙이라 할 수 있다. 높은 수준의 상호운용성을 달성하기 위해서는 CSV나 TXT 유통포맷에서 더 나아가 통제어휘(Controlled Vocabulary)에 기반하여 RDF (Resource Description Framework)나 LOD (Linked Open Data)를 구축함으로써 리포지토리 간의 시맨틱 검색(예: SPARQL)과 유통이 가능한 API 구현이 필요하다. 따라서 JOISS가 국내·외 해양관측데이터 리포지토리들의 유통 생태계에서 중개자로 그 역할을 충실히 수행하려면 해양관측기술의 진보에 맞추어 지속적인 표준화 개발과 플랫폼의 FAIR 요소를 강화하는 것이 중요하고 시급한 과제다.

4. 사사

이 논문은 2022년 해양수산부 재원으로 해양수산과학기술진흥원의 지원사업 '관할해역 해양정보 공동활용 체계 구축(2 단계) [과제번호 1525010982]'의 지원을 받아 수행된 연구입니다.

5. 참고문헌

Boeckhout M, Zielhuis GA, Bredenoord AL (2018) The FAIR guiding principles for data stewardship: fair enough?.

- European journal of human genetics: EJHG 26(7): 931-936 <https://doi.org/10.1038/s41431-018-0160-0>
- Boldrini E, Nativi S (2019) SeaDataNet metadata profile of ISO 19115–XML encoding (Version 11.0.0) Accessed 14 June 2022 <http://dx.doi.org/10.25607/OBP-572>
- Besaçon L et al. (2021) Open science saves lives: lessons from the COVID-19 pandemic. BMC Med Res Methodol 21(117): 1-18 <https://doi.org/10.1186/s12874-021-01304-y>
- FORCE11 (2016) The FAIR Data Principles. <https://www.force11.org/group/fairgroup/fairprinciples> Accessed 14 June 2022
- Garcia HE et al. (2018) World Ocean Database 2018: User's Manual (prerelease). <https://www.ncei.noaa.gov/products/world-ocean-database> Accessed 14 June 2022
- GO FAIR (2021) FAIRification process. <https://www.go-fair.org/fair-principles/fairification-process> Accessed 14 June 2022
- Helliwell JR et al. (2019) Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography. Journal of applied crystallography 52(Pt3): 495-497 <https://doi.org/10.1107/S2052252519005918>
- JOISS (2020) Data Policy. <https://www.joiss.kr/joiss/stnd.JoissDataPolicyPage.do> Accessed 14 June 2022
- National Ecological Observatory Network (2020) Data Formats and Conventions. <https://www.neonscience.org/data-samples/data-management/data-formats-conventions> Accessed 14 June 2022
- Pampel H et al. (2013) Making Research Data Repositories Visible: The re3data.org Registry. PLoS ONE 8(11): e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Partescano E et al. (2017) From heterogeneous marine sensors to sensor web: (near) real-time open data access adopting OGC sensor web enablement standards. Open geospatial data, softw. stand. 2(22) <https://doi.org/10.1186/s40965-017-0035-2>
- SeaDataNet (2010) Data Quality Control. <https://www.seadatanet.org/Standards/Data-Quality-Control> Accessed 14 June 2022
- Silva FCCd, Abadal E, Wulff E (2017) Oceanographic data repositories: An analysis of the international situation. Publications 5(2): 1-9 <https://doi.org/10.3390/publications5020008>
- Tanhua T et al. (2019) Ocean FAIR Data Services. Frontiers in Marine Science, 6. <https://doi.org/10.3389/fmars.2019.00440>
- Tim BL (2012) 5 stars principles of Linked Open Data. <http://5stardata.info> Accessed 14 June 2022
- University of Melbourne Library (2021) Research Data Curation. https://library.unimelb.edu.au/digital-stewardship/research_data_curation Accessed 14 June 2022
- Wilkinson MD et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3(160018) <https://doi.org/10.1038/sdata.2016.18>
- 신은정, 안형준, 정원교 (2016) 오픈사이언스를 위한 연구성과물 공개정책과 과제, 과학기술정책연구원
- 한국데이터산업진흥원 (2010) 데이터 품질점검 사항. https://www.kdata.or.kr/_img/web/DQCV_check.pdf Accessed 14 June 2022