

극한해양공간 실시간 데이터 수집 및 분석을 위한 통합 데이터 플랫폼 구축 방안

한현경¹ · 이철용² · 박용길^{3*}

¹한국해양과학기술원 해양빅데이터센터 기술원

²한국해양과학기술원 해양빅데이터센터 선임기술원

^{3*}한국해양과학기술원 해양빅데이터센터 연수연구원

A Plan to Build an Integrated Data Platform for Real-Time Data Collection and Analysis of Extreme Ocean Spaces

Hyeon-Gyeong Han¹ · Cholyoung Lee² · Yong-Gil Park^{3*}

¹Research Specialist, Marine Bigdata center, Korea Institute of Ocean Science & Technology, Busan, Korea

²Senior Research Specialist, Marine Bigdata center, Korea Institute of Ocean Science & Technology, Busan, Korea

^{3*}Post Doctoral Scientist, Marine Bigdata center, Korea Institute of Ocean Science & Technology, Busan, Korea

[요 약]

최근 데이터 공유, 활용에 관한 정책 및 법률이 재정되는 등 데이터 수집 및 분석의 중요성이 대두되고 있다. 특히 극한해양공간의 경우 높은 데이터 수집 비용과 기술이 필요하므로 수집된 데이터의 유실 없는 관리 및 분석 체계의 구축이 요구된다. 이에 본 연구는 국내외 연구 데이터 정책 동향, 빅데이터 수집 관리 기술 동향, 관련 기관 및 조직의 수요를 조사하고 이를 종합하여 극한해양공간 실시간 데이터 플랫폼 구축 방안을 제시하는 것을 목표로 하였다. 먼저, 극한해양공간 데이터의 실시간 수집, 저장, 처리, 분석, 가시화의 기능 구조를 정의하였고, 이를 연계하는 방식으로 통합 데이터 플랫폼 구성도를 도시함으로써 빅데이터 플랫폼 구축 방안을 제안하였다. 향후 본 연구의 결과를 활용하여 실질적인 극한해양공간 통합 데이터 플랫폼 구축 연구가 이루어져야 할 것이다.

[Abstract]

Recently, policies and laws related to data sharing and utilization have been revised, and the importance of data collection and analysis is emerging. In particular, in the case of extreme marine spaces, high data collection costs and technology are required, so it is required to establish a management and analysis system without loss of collected data. Therefore, this study aims to present a plan to establish an extreme ocean space data platform by investigating domestic and foreign data policy trends, big data collection management technology trends, and demands of related organizations. First, the functional structure of real-time collection, storage, processing, analysis, and visualization of oceanic extreme space data was defined. In addition, a plan for building a big data platform was proposed by schematically drawing the configuration of the integrated data platform in a way that connects it. In future research, a study on the establishment of a practical extreme marine space integrated data platform should be conducted using the results of this study.

색인어 : 극한해양공간, 빅데이터, 데이터 플랫폼, 통합 시스템 설계, 데이터 정책 동향

Key word : Extreme Ocean Space, Big Data, Data Platform, Integrated System Design, Data Policy Trend

<http://dx.doi.org/10.9728/dcs.2021.22.6.989>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 23 March 2021; **Revised** 22 March 2021

Accepted 05 March 2021 April May June

***Corresponding Author; Yong-Gil Park**

Tel: 82-51-644-3779

E-mail: ygpark32@kiost.ac.kr

1. 서론

최근 연구개발의 패러다임은 데이터 중심의 활용 기술 및 융합 연구 환경에 집중되는 양상으로 변화하고 있다[1]. 또한 국가예산의 투입으로 생산된 데이터를 개별 소유 및 관리하던 과거와 달리 오픈사이언스 정책 확산에 따라 연구데이터를 공공재로 인식되어가는 추세이다[2].

국외의 경우 1990년도에 인간게놈지도 생성을 위해 미국, 영국, 일본, 독일, 프랑스, 중국 등이 참여하는 국제 연구컨소시엄을 구성하여 연구를 추진하였다. 이때 논문작성을 보류하고 목표달성을 위해 연구 결과를 공개 및 공유하여 대규모 연구협력을 진행한 결과 최종 목표연도인 2005년보다 2년 빠른 2003년 인간게놈지도가 완성되었다[3]. 또한 2009년부터 수행된 유럽 입자물리학연구소의 Large Hadron Collider(LHC) Project는 LHC를 이용한 실험데이터 공동분석 연구를 위해 유럽입자물리학연구소 산하 36개국 네트워크를 통해 연구데이터 공유 및 공동연구를 진행하였다. 이를 통해 2012년 희스입자를 발견하고 380여 편의 논문을 생산하였다. 2013년 미국과학재단의 보고서에서는 데이터 관리, 보존하고 이를 다른 연구자들이 이용할 수 있게 데이터 리포지토리를 구축하는 것을 사이버 인프라의 핵심요소로 제안하고 있다[2],[4].

이처럼 국내에서도 다양한 연구기관에서 과거부터 현재까지 보유하고 있는 방대한 양의 데이터를 수집하고 가치를 추출해내는 것이 향후 우리나라의 과학발전에 중요한 역할을 할 것으로 기대된다. 그중에서도 해양의 경우 사람의 접근이 어려워 위성, 부이, 연구선 등을 이용한 데이터 수집이 이루어져 높은 데이터 수집 비용과 고난이도 관측기술이 요구된다[5]. 따라서 수집된 데이터의 유실 없는 보관과 효율적인 활용이 필요하며 특히 심해, 극지 빙저 수중환경, 해상상태 8(Wave height 9 to 14 metres)의 황천, 고염 해역 등 극한해양공간에서 수집되는 데이터는 데이터의 수집이 어려울 뿐 아니라 새로운 해양현상 발견과 높은 기술의 해양기술 개발을 위해 필수적인 수집과 공유체계가 구축되어야한다.

이에 본 연구에서는 극한해양공간 데이터 통합관리 체계 구축을 위해 빅데이터(Big data)의 정책동향과 기술동향 조사를 수행하였고 관련 연구자들의 수요조사를 통해 극한해양공간에 최적화된 빅데이터 플랫폼 구축방안을 제시하였다.

II. 국내외 연구 데이터 정책 동향

2-1 국외 연구 데이터 정책

미국은 2013년 오바마 정부의 Office of Science and Technology Policy(OSTP) memorandum은 공공부문 오픈액세스 정책 확대의 기폭제 역할을 하였다. OSTP memorandum의 최종목표는 연구의 성과를 연구자뿐만 아니라 기업과 사회가

활용해 혁신을 도모하고 경제성장과 고용창출에 기여하는 것이다[6]. 이와 관련된 세부내용은 표 1과 같으며 학술논문에 관한 지침과 연구데이터에 관한 지침을 간단하게 설명하면 다음과 같다. 먼저 학술논문에 관한 지침은 연방정부의 지원을 받은 공공연구에서 생산된 논문은 출판 이후 일반인이 접근가능하며 다운로드 받아 분석할 수 있도록 공개할 것을 권고하고 있다. 또한 각 논문에 관한 메타데이터가 제공되어 완전한 공공액세스가 구현될 수 있도록 지원하고 오픈액세스체계를 구축하여 민관과 협력하여 저자, 저널, 출판 등의 권리를 보장할 수 있게 하고 있다. 다음으로 연구데이터에 관한 지침은 연방정보 지원으로 생산된 연구데이터에 대한 접근성과 함께 국가 안보, 프라이버시, 지식 재산권 등의 문제도 고려하도록 권고하고 있다. 또한 연방정보 지원으로 수행된 학술연구와 정부기관 자체연구에서 생산된 연구데이터에 대해서는 데이터 관리계획(DMP; Data Management Plan)을 작성하고 체계적인 관리, 보존을 권고하고 있다. 데이터의 관리와 보존을 위해 데이터 액세스 및 호환체계 마련, 교육 및 전문 기술인 양성을 위한 민관협력, 타 기관의 협력을 장려하고 있다.

영국의 경우 2014년 European Commission에서 “Open science”라는 용어의 사용을 시작하였다. 오픈사이언스는 과학의 효율성과 신뢰성을 증대시키며 사회발전에 기여할 수 있어 유럽에 최적화된 오픈사이언스 활용을 위해 이해관계자, 공동설계, 개발자와 함께 정책을 개발하고 있다. Horizon Europe에서 연구데이터는 가능한 개방적이며 필요한 경우 폐쇄를 원칙으로 하고 있으며 연구데이터를 공개하지 않더라도 DMP 작성을 의무화 하고 있다[7].

일본의 경우 2013년 G8 정상회의 이후 오픈사이언스 정책수립 검토를 시작하였고 2015년 오픈사이언스 정책에 관한 제안을 작성을 수행하였다. 이후 2016년부터는 일본 과학기술정책 차원에서 오픈사이언스 추진방안의 구체화를 시작하였고 일본 내각부에서 주관하여 관계부처가 협력하여 시행중이다[8].

2-2 국내 연구 데이터 정책

국외의 주요 국가는 연구 성과와 연구과정을 공유하는 오픈 사이언스 정책을 추진하고 있으나 국내에는 연구 데이터 관리 및 공유의 제도와 인프라가 부족한 실정이다. 이에 2018년부터 과학기술정보통신부에서 혁신성장 촉진을 위한 연구데이터 공유 및 활용전략을 추진하고 있다[9]. 이를 통해 국가 연구개발 과제 수행 중 생산, 축적되는 연구데이터를 체계적으로 관리 및 공유하여 새로운 지식과 가치창출에 연구데이터를 적극 활용하도록 지원하는 것을 목표로 하고 있다. 또한 2019년부터 2023년까지 데이터와 Artificial Intelligence(AI)를 가장 안전하게 잘 쓰는 나라라는 비전을 가지고 관계부처 합동으로 데이터·AI 경제 활성화 계획을 추진하고 있다[10]. 이를 위해서는 데이터의 수집과 제공이 필수이기 때문에 약 743억을 투자하여 빅데이터 플랫폼과 빅데이터 센터를 설립할 예정이다.

표 1. 미국 OSTP(Office of Science and Technology Policy) memorandum 주요 내용[6]

Table 1. Main contents of the US OSTP(Office of Science and Technology Policy) memorandum[6]

Objectives for Public Access to Scientific Publications	Objectives for Public Access to Scientific Data in Digital Formats
<p>a) Ensure that the public can read, download, and analyze in digital form final peer-reviewed manuscripts or final published documents within a timeframe that is appropriate for each type of research conducted or sponsored by the agency. Specifically, each agency:</p> <p>b) Facilitate easy public search, analysis of, and access to peer-reviewed scholarly publications directly arising from research funded by the Federal Government;</p> <p>c) Ensure full public access to publications' metadata without charge upon first publication in a data format that ensures interoperability with current and future search technology. Where possible, the metadata should provide a link to the location where the full text and associated supplemental materials will be made available after the embargo period;</p> <p>d) Encourage public-private collaboration to:</p> <p>i) maximize the potential for interoperability between public and private platforms and creative reuse to enhance value to all stakeholders,</p> <p>ii) avoid unnecessary duplication of existing mechanisms,</p> <p>iii) maximize the impact of the Federal research investment, and</p> <p>iv) otherwise assist with implementation of the agency plan;</p> <p>e) Ensure that attribution to authors, journals, and original publishers is maintained; and</p> <p>f) Ensure that publications and metadata are stored in an archival solution that:</p> <p>i) provides for long-term preservation and access to the content without charge,</p> <p>ii) uses standards, widely available and, to the extent possible, nonproprietary archival formats for text and associated content (e.g., images, video, supporting data),</p> <p>iv) enables integration and interoperability with other Federal public access archival solutions and other appropriate archives.</p>	<p>a) Maximize access, by the general public and without charge, to digitally formatted scientific data created with Federal funds, while:</p> <p>b) Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why longterm preservation and access cannot be justified;</p> <p>c) Allow the inclusion of appropriate costs for data management and access in proposals for Federal funding for scientific research;</p> <p>d) Ensure appropriate evaluation of the merits of submitted data management plans;</p> <p>e) Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies;</p> <p>f) Promote the deposit of data in publicly accessible databases, where appropriate and available;</p> <p>g) Encourage cooperation with the private sector to improve data access and compatibility, including through the formation of public-private partnerships with foundations and other research funding organizations;</p> <p>h) Develop approaches for identifying and providing appropriate attribution to scientific data sets that are made available under the plan;</p> <p>i) In coordination with other agencies and the private sector, support training, education, and workforce development related to scientific data management, analysis, storage, preservation, and stewardship; and</p> <p>j) Provide for the assessment of long-term needs for the preservation of scientific data in fields that the agency supports and outline options for developing and sustaining repositories for scientific data in digital formats, taking into account the efforts of public and private sector entities.</p>

빅데이터 플랫폼은 주요 분야별 각종 빅데이터의 수집, 분석, 유통을 지원하고 빅데이터 센터는 기업, 대학 등 주요 기관별 빅데이터를 체계적으로 생산하고 관리하는 역할을 가지고 있다. 이처럼 다양한 국가기관에서 데이터 공유 및 활용에 대한 계획을 발표하며 향후 연구데이터의 공유, 활용 체계 구축 및 운영에 관한 지속적인 압력이 예상된다.

특히 해양 빅데이터의 경우 문재인 대통령 대선 공약집 제10번 안전한 대한민국에서 해양지진, 지진해일 등 대형 해양재난 대응체계 강화를 위한 빅데이터 기반 해양예측 및 자동경부 시스템 구축 추진에서 빅데이터 구축의 필요성이 강조되고 있다 [11]. 이에 2021년 해양정보의 체계적인 관리 및 활용과 해양산업의 발전 및 해양재해의 예방 등을 위해 해양조사와 해양정보 활용에 관한 법률이 제정되었다. 해당 법에서는 해양정보의 보관과 일반인의 열람이 가능하여야 하며 해양조사 계획이나 해양정보 제출을 요구할 수 있음을 명명하고 있다.

또한 생산된 해양정보 등을 수집·가공·분석·예측하고 이를 총괄하여 관리·제공하는 국가해양정보시스템의 구축·운영이 가능함이 명시되어있다(표 2).

III. 빅데이터 수집 및 관리 기술 동향

빅데이터는 기존 데이터베이스로는 수집·저장·관리·분석을 수행하기 어려울 만큼 방대한 양의 데이터로 대량의 정형, 비정형 데이터를 포함하고 있다. 극한해양공간 자료 역시 부이, Cable, ROV(Remotely Operated Vehicle), AUV(Autonomous Underwater Vehicle) 등 다양한 관측 장비를 이용해 수집되므로 극한해양공간 실시간 데이터 플랫폼에서는 많은 양의 데이터를 분석하여 유의미한 정보를 추출하고 결과분석을 가능하게 하는 것은 중요하다. 이를 위한 관련 요소기술로는 수집, 공유,

저장, 처리, 분석, 시각화 기술이 있다. 따라서 극한해양공간 데이터 플랫폼 구축을 위해서는 이러한 요소기술들을 조합하여 적절한 배치와 연계가 필요하다. 이에 빅데이터 수집, 저장, 처리, 분석 및 시각화 기술을 분석하였다.

표 2. 해양조사와 해양정보 활용에 관한 법률

Table 2. Law on Ocean Survey and Ocean Information (<https://www.law.go.kr/법령/해양조사와해양정보활용에관한법률>)

해양조사와 해양정보 활용에 관한 법률
제4장 해양정보의 활용
제1절 해양정보
제42조(해양정보의 보관 및 열람 등) ① 해양수산부장관은 해양정보를 보관하고 일반인이 열람할 수 있도록 하여야 한다.
제43조(해양정보의 품질관리) ① 해양수산부장관은 해양정보의 정확도를 확보하기 위하여 해양정보의 품질관리에 필요한 시책을 추진하여야 한다.
② 제1항에 따른 품질관리의 대상, 범위, 기준 및 절차 등에 관한 사항은 해양수산부령으로 정한다.
제44조(관계 기관의 해양정보 활용 등) ① 해양수산부장관은 관계 기관이 다음 각 호의 어느 하나에 해당하는 해양조사를 할 때에는 그 해양조사 계획이나 해양정보를 제출할 것을 요구할 수 있다.
1. 조석·조류·해류의 관측 및 해수의 물리적 특성 조사
2. 해저지형, 해상 지구자기, 해상 중력 및 해저지질의 조사
3. 인공어초 등 해저위험물의 조사
4. 그 밖에 해양수산부장관이 정하여 고시하는 사항에 관한 조사
② 해양수산부장관은 제1항에 따라 해양조사 계획을 제출한 관계 기관과 조사자료의 공동활용, 공동조사 및 기술협력을 위하여 노력하여야 한다.
③ 제1항에 따른 해양조사 계획이나 해양정보의 제출 등에 필요한 사항은 해양수산부령으로 정한다.
제45조(국가해양정보시스템) ① 해양수산부장관은 생산된 해양정보 등을 수집·가공·분석·예측하고 이를 총괄하여 관리·제공하는 국가해양정보시스템을 구축·운영할 수 있다.
② 해양수산부장관은 국가해양정보시스템의 운영을 위하여 필요한 경우 관계 행정기관 및 해양조사 관련 기관 등에 관련 자료의 제공을 요청할 수 있다.
제46조(해양정보활용센터의 설치 등) ① 해양수산부장관은 해양정보의 수집·가공·분석·예측 업무를 효율적으로 수행하고, 정보이용자에게 해양정보를 원활하게 제공하기 위하여 대통령령으로 정하는 해양수산부 소속 기관에 해양정보활용센터를 설치·운영할 수 있다.
② 제1항에 따른 해양정보활용센터의 설치·운영에 필요한 사항은 대통령령으로 정한다.

3-1 빅데이터 수집기술

빅데이터 수집은 다양한 데이터 소스로 부터 필요한 데이터를 검색하여 단순한 데이터 수집이 아닌 검색, 수집, 변환을 통해 정제된 데이터를 수집하는 기술이다(그림 1).

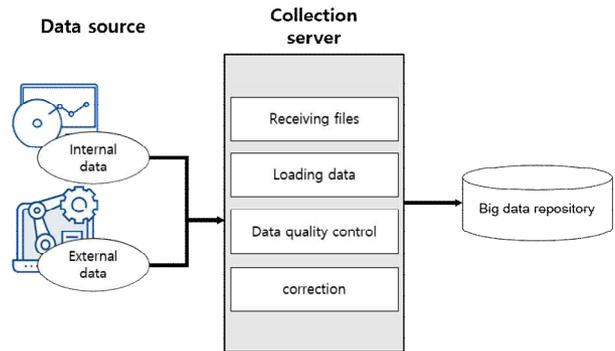


그림 1. 데이터 수집흐름
Fig. 1. Data collection flow chart

이를 위해 대표적인 로그 수집기술로 아파치(Apache)의 플룸(Flume)과 척와(Chuckwa), 페이스북에서 스트리밍 데이터를 처리하기 위해 개발한 스크라이브(Scribe) 등이 존재한다. Flume은 최근 국내의 빅데이터 수집기술로 많이 사용되고 있으며 다양한 소스로부터 데이터를 수집해 데이터를 전송할 수 있다. 이는 아키텍처가 단순하고 유연하여 실시간 분석에 유용하다[13]. Chukwa는 수집된 로그파일을 Hadoop Distributed File System(HDFS)에 저장하여 HDFS의 장점을 활용할 수 있으며 실시간 분석에도 활용이 가능하다[14]. 하지만 하둠에 의존적이라는 단점이 존재한다. Scribe는 서버 타입에 상관없이 다양한 방식으로 로그를 읽어 들일 수 있으며[15] 현재 페이스북의 수백억건의 메시지를 처리하고 있다[16].

3-2 빅데이터 저장 및 처리 기술

빅데이터 저장은 수집된 데이터를 모두 저장하여 분석을 효율적으로 수행하기 위해 만들어진 기술이며 빅데이터 처리기술은 기존방식으로 처리할 수 없는 방대한 양의 데이터를 처리하는 기술이다. 이와 같은 대용량 데이터 처리는 컴퓨팅 자원을 이용하는 병렬처리를 기본으로 하고 있다. 테라바이트 이상으로 늘어나는 데이터에 대한 병렬처리는 기존 Database management system(DBMS)에서는 거의 불가능했던 대용량의 데이터를 저비용으로 처리를 가능하게 해준다[17]. 최근 저렴한 데이터 저장과 처리를 위해 오픈소스 기술을 많이 사용하고 있으며 관련 기술은 하둠(Hadoop)의 HDFS/Hbase, Mango Database(DB), NoSQL DB등이 있다.

그중 최근 많은 각광을 받고 있는 하둠의 분산저장 파일시스템(HDFS)은 IBM, Amazon, Yahoo 등의 세계적인 IT기업의 클라우드 컴퓨팅 플랫폼 기반 분산 파일 시스템으로 사용되고 있다. 또한 Markets and markets의 하둠 빅데이터 분석 시장 보고서에 따르면 2025년까지 연평균 13%의 성장이 예상되며 그 가치는 2020년 132억 달러에서 2025년 235억 달러의 가치를 가질 것으로 예상된다[18]. HDFS는 대표적인 하둠의 파일시스템으로 네임노드와 데이터 노드로 구성되어 테라바이트, 페타바이

트 이상의 대용량 파일을 분산 저장하여 저장된 데이터를 빠르게 처리할 수 있는 장점이 있다(그림 2)[19].

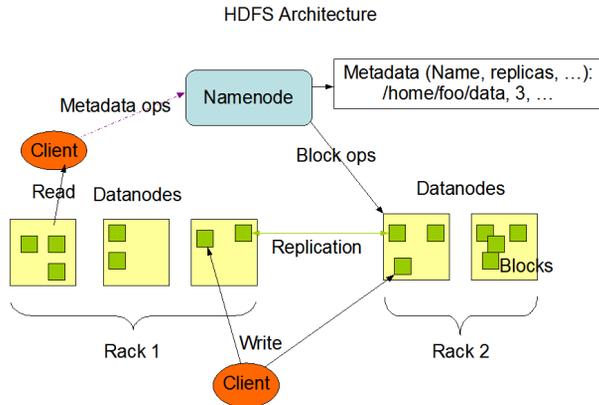


그림 2. HDFS 구조[19]
Fig. 2. HDFS architecture[19]

3-3 빅데이터 분석 및 시각화 기술

빅데이터 분석이 최근 주목받는 이유는 과거 분석할 수 없던 방대한 양의 데이터를 정확하게 분석하고 새로운 가치를 창출하여 적시에 활용 가능하기 때문이다[20]. 과거 기가바이트(GB; Gigabyte) 크기의 데이터 처리가 가능했다면 현재는 분산 병렬 처리와 인메모리 기술을 이용해 수십, 수백 테라바이트(TB; Terabyte)이상의 데이터를 처리할 수 있고 정형 데이터와 반정형, 비정형 데이터까지도 처리가 가능하다. 이를 통해 샘플링 과정에서 발생했던 오차를 줄일 수 있고 기하급수적으로 늘어나는 데이터의 계산을 수행 할 수 있다. 이는 그동안 불가능했던 분석이 가능하게 되었을 뿐만 아니라 시간의 단축과 결과의 정확도를 향상 시키는 효과를 가져 올 수 있다[21].

빅데이터분석 기술은 크게 4가지 텍스트 마이닝(Text mining), 평판분석(Opinion mining), 소셜 네트워크 분석(Social network analysis), 클러스터 분석(Cluster analysis)으로 나눌 수 있다[22]. 텍스트 마이닝은 비정형, 반정형 텍스트에서 의미 있는 정보를 추출하여 문서 분류, 문서군집, 정보추출, 문서 요약 등의 분야에 사용가능하다. 평판분석은 텍스트 마이닝의 관련 분야로 소셜미디어 등의 정형, 비정형 텍스트의 선호도를 판별하는 기술로 서비스 및 상품의 시장 규모 예측, 소비자 반응 예측 등에 활용되고 있다. 소셜네트워크 분석은 소셜네트워크의 연결 구조 및 연결 강도 등을 바탕으로 영향력의 중심인 사용자를 찾는 데 주로 사용된다. 마지막으로 클러스터 분석은 데이터의 특성을 고려해 최종적으로 집단을 대표할 수 있는 특성을 발굴하는데 주로 사용된다.

IV. 기술개발 목표 및 내용

본 연구에서는 극한해양공간에 최적화된 빅데이터 플랫폼 구축방안 제시를 위하여 한국해양과학기술원(KIOST; Korea Institute of Ocean Science and Technology), 극지연구소(KOPRI; Korea Polar Research Institute), 선박해양플랜트연구소(Korea Research Institute of ships and ocean engineering)를 대상으로 수요조사를 실시하였다. 각 기관은 해양 탐사, 조사, 관측, 극한환경 관측 기술과 첨단 장비 개발, 연구선 및 해저관측 케이블 연구관련 많은 성과를 가지고 있는 전문 기관이다.

수요조사에는 극한해양공간에서 관측예상 되는 데이터의 종류, 수집주기, 전송방식, 용량 등에 대해 데이터 통합관리 및 활용 차원에서 요구되는 기능과 서비스, 데이터 공개 및 공동 활용 가능 여부 등을 조사하였다. 수요조사 결과는 다음과 같다(표 3). 먼저 데이터 전송방식은 실시간 통신 자료의 수집 및 서비스 수요가 많은 것을 확인하였고 데이터 생산주기는 일별로 일별자료 수집을 통한 빅데이터 구축, 품질 및 데이터의 체계적인 관리, 분석 및 가시화 지원체계 구축 등이 요구되고 있었다. 또한 비공개 데이터에 대한 데이터 보안강화 요구가 있었으며 관련하여 비공개 데이터 활용 시 대책 방안 모색이 필요하다.

이에 실시간 극한해양공간의 기장비에서 생산되는 빅데이터를 체계적이고 효율적인 수집-저장-관리-분석-서비스를 수행하기 위한 플랫폼 구축 방안을 제시하였다. 이를 위해 첫째, 실시간 통합관측 데이터 수집/관리 체계, 둘째, 해양과학 빅데이터 통합 관리 시스템 및 서비스 체계, 셋째, 해양과학 데이터 사이언스 플랫폼, 넷째, 극한해양공간 통합관측 기장비 거버넌스 플랫폼 구축방안을 제시하였다.

4-1 실시간 극한해양공간 데이터 수집/관리 체계 구축

극한지 해양 공간 데이터의 실시간 수집, 관리를 위해 요구되는 기능은 다음과 같다. 먼저, 실시간 자료 수집 수요조사 및 연계 대상 기장비의 정의가 수행되어야 한다. 이때 사람의 접근이 어려운 극한지의 데이터 획득을 위해 실시간 데이터 송수신 환경 분석 및 물리적 연결 구조를 설계하여야한다. 이를 위해 Internet of Things(IoT) 센서 및 Information & Communications Technology(ICT)기반 실시간 데이터 수집 기술의 검토 및 도입이 필요하며 시스템의 유지보수와 관리를 위해 Microsoft .net Framework, Java 플랫폼 등을 활용한 Client/Server(C/S) 시스템 개발이 필요하다.

또한 실시간 데이터 수집, 관리, 제어 등의 기능을 상세 설계하고 앞장의 빅데이터 수집기술에 언급되었듯 단순 데이터 수집이 아니라 실시간으로 수집되는 데이터의 정제와 표준화 방안을 수립하여 Real-Time streaming DB를 구축해야한다. 표준화를 통해 수집된 데이터의 인덱싱 처리 및 검색, 조회 기능과 데이터 수집 현황 모니터링 및 이상 징후 탐지 기능이 필요하다. 이때 극한지 환경 특성상 많은 내외부 장애 요소에도 지속적인 서비스 제공을 위해 DB 이중화가 요구된다. 이후 실시간 데이터 수집, 처리 결과에 따른 통계적 가시화 기능이 요구되어 대시보드 형식의 정보 가시화 기술 검토가 필요하다.

표 3. 극한환경 데이터 수요조사 결과

Table 3. Extreme Environment Data Demand Survey Results

Detailed Research and Development Content	Survey Observed ocean	Data Production Device/Equipment Name	New development of observation equipment	Data production method	Data production projected items	Data format	Data production cycle	
Submarine Observation Cable Sensor Field	East Sea, the sea area around the Korean Peninsula, etc.	Submarine Observation Cable	Complement/improvement of existing equipment	Fixed point time series observation	Submarine, water temperature, Pressure change, Underwater acoustic signal	CSV	Continuous observation	
Developed Top-Tensioned Buoy	Rough sea	Buoy	New development	Fixed point time series observation	Water temperature, salinity, water pressure, pH, DO, underwater acoustics	CSV	Continuous observation	
Ice bottom exploration hybrid AUV development	pol, Under ice	Hybrid AUV and sensor	Etc	Tracking linear observation	Water temperature, salinity, water pressure, pH, DO, underwater acoustics	Undefined	Etc	
Own data storage method	Data transmission method	Data production/transmission per day	Raw Data Offered or not	Post-processing after data provision	Need to utilize big data analysis	Data visualization function required	Real-time display service required	Public/private
File format	Real-time wired communication	100~ 1000 MB/Day	Yes	Yes	Yes	Yes	Yes	Private
Undefined	Real-time wireless communication	100~ 1000 MB/Day	Yes	Yes	Yes	Undefined	Yes	Public
Undefined	Undefined	Undefined	Yes	Yes	Yes	Yes	Undefined	Undefined

마지막으로 전문기관의 수요조사에서 요구된 시스템 보안 및 정보 보안 대책의 수립과 적용 고려를 위해 DB보안 및 백업용 DB 도입 방안 검토와 시스템 보안 및 정보 보안 기술 검토 및 도입이 필요하다. 해당 체계가 구축되면 시스템 관리자와 해양과학연구자가 주된 사용자가 될 것이며 이를 정리한 시스템 구성도는 그림 3과 같다.

4-2 빅데이터 통합 관리 시스템 및 서비스 체계 구축

극한해양공간 빅데이터의 통합 관리와 서비스 시스템 체계 구축을 위해 데이터 및 메타 데이터의 표준화 방안을 모색해야 하며 수집된 데이터 기반의 통합 Operation DB 검색 및 조회기능을 설계하여야 한다. 데이터 검색 및 조회 편의를 위해 단순한 데이터 리스트 조회가 아닌 Geographic Information System(GIS) 기반 가시화 기능과 사용자 요구에 맞는 엑셀, Comma-separated values(CSV), 보고서 데이터 등 제공기능이 필요하다. 이때 효율적인 가시화 기능을 위해 QGIS, gdal과 같은 오픈소스 기반 GIS 소프트웨어 및 라이브러리 사용이 고려되어야 한다. 이와 함께 대용량 로그 및 통계 분석을 위한 Apache Hadoop 기반의 빅데이터 분석 플랫폼과 연계하여 시스템 이용자, 데이터 통계, 수집 관리 기능의 검토가 필요하다. 또한 관계 기관과의 데이터 공유를 위해 Comprehensive Knowledge Archive Network(CKAN), Socrata

등 기술 검토가 필요하고 필요시 Microsoft, Google의 데이터 분석 플랫폼 연계방안 고려도 필요하다. 해당 서비스 체계 구축이 완료되면 해양과학자, 외부 연구자, 정부지자체와 일반국민들 등 다양한 사용자가 이용 가능하다(그림 4).

4-3 극한해양과학 데이터 플랫폼 구축 및 운영

플랫폼 구축을 위해 해양과학 빅데이터 통합관리 시스템 DB등 Data science DB 관련 시스템과의 DB연계가 필요하며 이를 위해 Application Programming Interface(API), 크롤링 기술 등 검토가 필요하다. 또한 앞서 언급한 빅데이터 저장 및 처리 기술에 기술되어 있는 HDFS 기반 인프라 및 오픈소스 S/W 기술 도입을 통한 분산저장 및 병렬처리 지원이 필요하다. 그리고 클라우드 기반 데이터의 접근과 활용을 통한 빅데이터 분석을 위해 Platform as a service(PaaS), Infrastructure as a Service(IaaS), Software as a Service(SaaS) 등 아키텍처와 R, Python, Scala등 통계분석 프로그램 도입의 검토가 필요하다. 분석된 빅데이터를 사용하는 해양 연구자들의 플랫폼 활용 효율을 위해 데이터의 가시화를 위해 웹 브라우저 기반의 Graphical user interface(GUI) 기술 검토와 더불어 보안을 위한 이중화 기술 등의 검토가 필요하다. 이를 정리한 시스템 구성도는 그림 5와 같다.

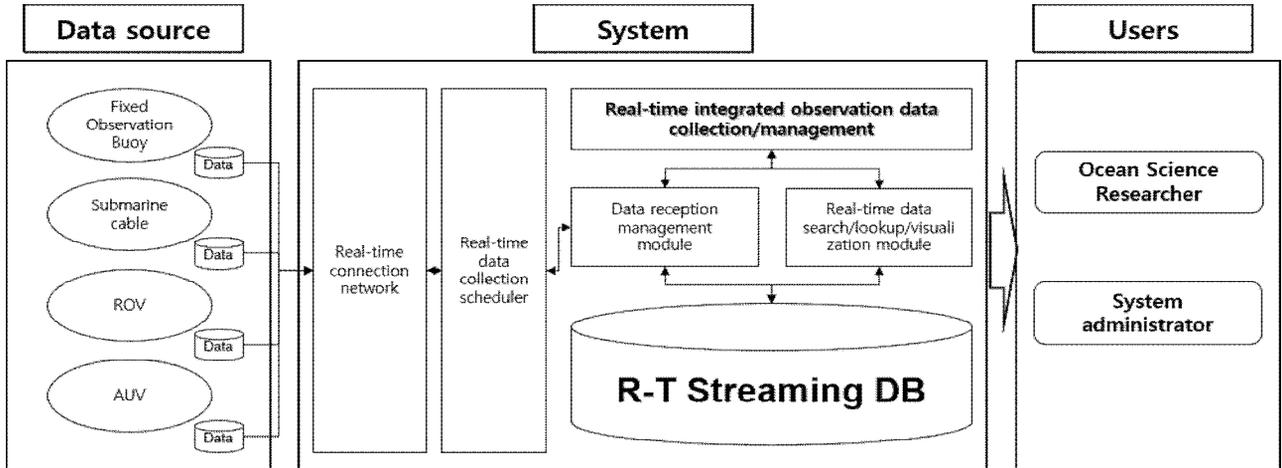


그림 3. 실시간 통합관측 데이터 수집/관리 체계 구축 및 운영 구성도

Fig. 3. Real-time integrated observation data collection/management system construction and operation diagram

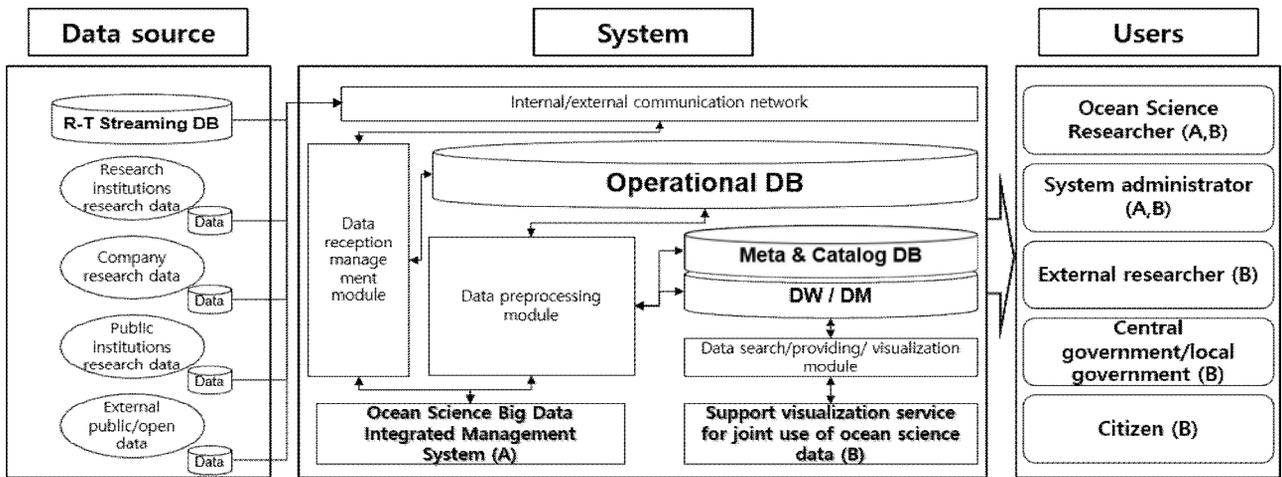


그림 4. 해양과학 빅데이터 통합관리 시스템 및 서비스 체계 구축 및 운영 구성도

Fig. 4. Ocean science big data integrated management system and service system construction and operation diagram

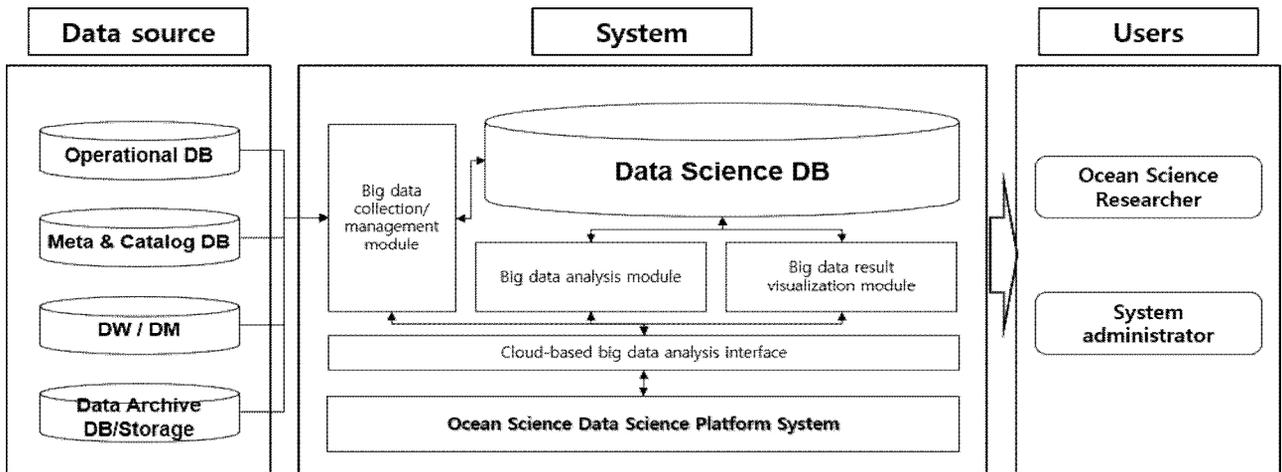


그림 5. 해양과학 데이터 플랫폼 구축 및 운영 구성도

Fig. 5. Ocean science data platform construction and operation diagram

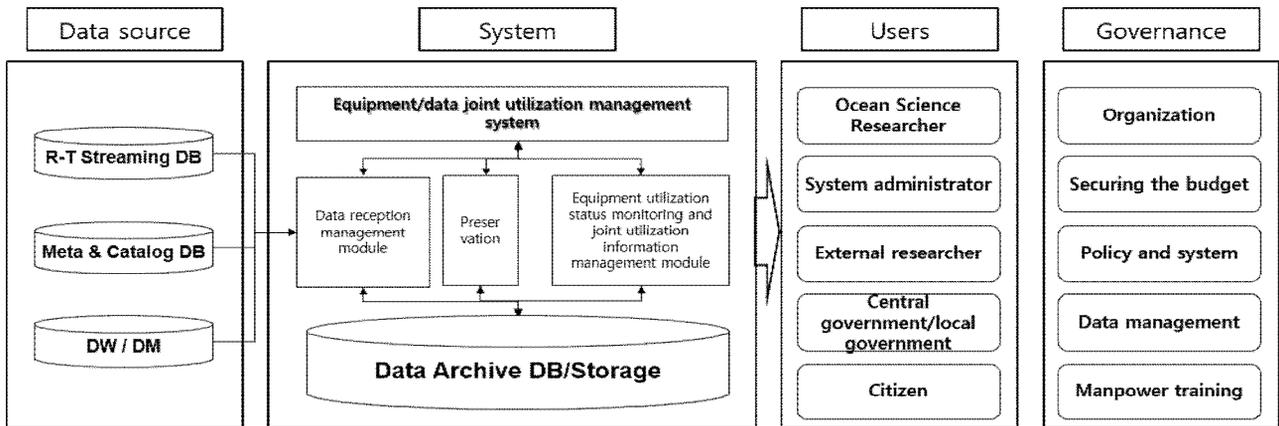


그림 6. 극한해양공간 통합관측 기장비 거버넌스 구축 및 운영 구성도

Fig. 6. Construction and operation diagram of integrated monitoring equipment and governance for extreme ocean spaces

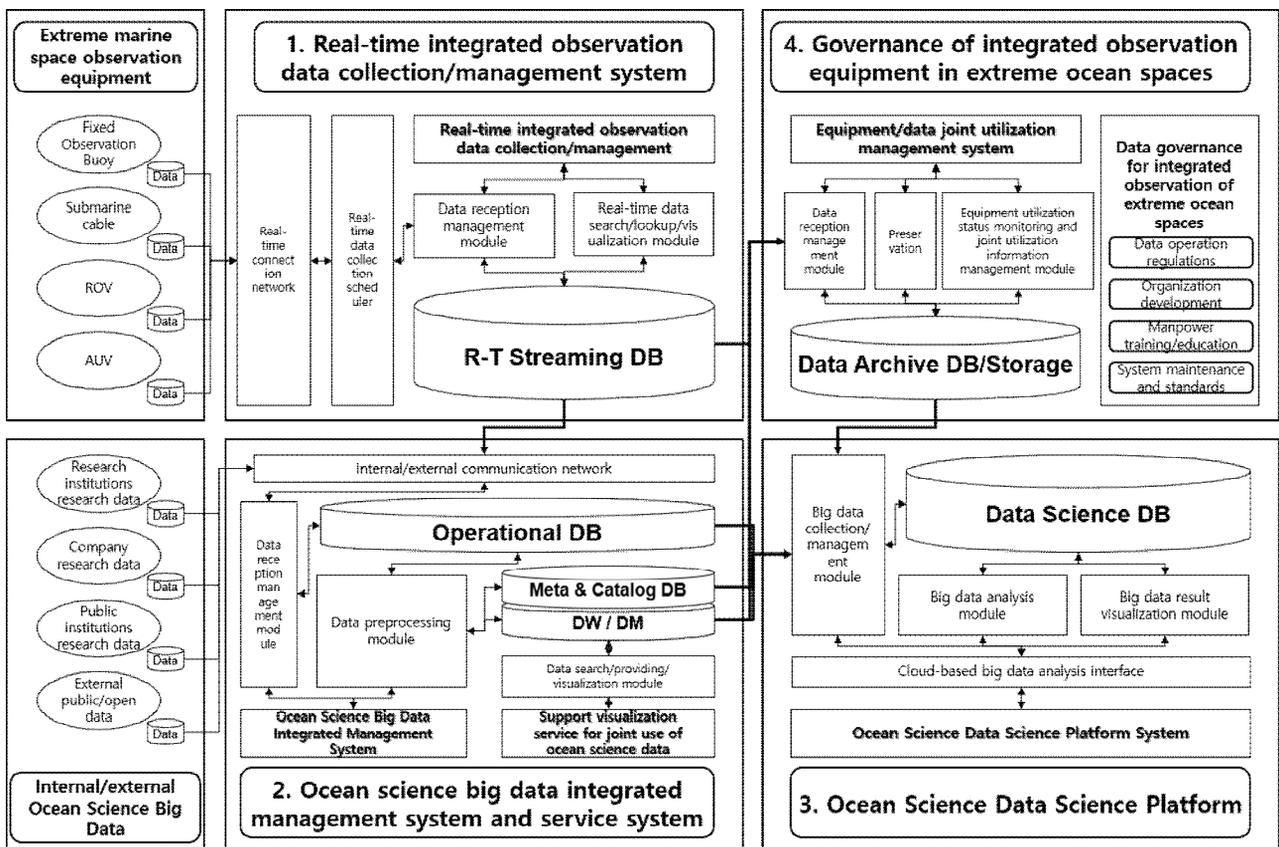


그림 7. 극한해양공간 빅데이터 플랫폼 전체 목표 체계 구성도

Fig. 7. Extreme marine space big data platform overall goal configuration diagram

4-4 극한해양공간 통합관측 기장비 거버넌스 구축 및 운영

실시간 데이터 수집을 위해 실시간 극한해양공간 데이터 수집/관리 시스템의 Data Archive DB/Storage 연계가 필요하며 외부 연계를 위한 데이터 표준화 및 API 기술의 검토가 필요하다. 또한 통합관측 기장비 거버넌스 구축을 위해 크게 3가지, 시스템 관리 기능 개발, 통합 관측 데이터 수집 및 처리 기능개발, 통합관측 기장비 정보 관리 기능 개발이 이루어져야한다.

먼저 시스템 관리 기능 개발을 위해 사용자별 접근 제어, 권한 관리, 이력관리, 자원관리, 시스템 구성관리, 외부 시스템 연동 관리가 필요하다. 통합관측 데이터 수집/처리 기능 개발을 위해서는 데이터 수신 프로토콜, 데이터 표준화(분류/그룹화), 데이터 품질관리, DB 입출력 제어, 데이터 아카이브가 고려되어야한다. 통합관측 기장비 정보 관리 기능 개발을 위해서는 장비그룹, 목록, 이력, 상태, 운용환경 관리가 필요하다. 이를 위해서는 전담 조직 구성과 예산확보, 미국, 영국과 같이 이해관계

자, 개발자와의 합의를 통한 DMP작성 등과 같이 구체적이고 현실적인 정책 및 제도 수립이 필요하다(그림 6).

최종적으로 실시간 극한해양공간 데이터 수집/관리, 빅데이터 통합 관리 시스템 및 서비스 체계 구축, 데이터 플랫폼 구축 및 운영, 기장비 거버넌스 구축 및 운영 구성도를 통합한 극한 해양 공간 실시간 데이터 수집 및 분석을 위한 통합 플랫폼 구축방안은 그림 7과 같다.

V. 결 론

최근 연구 환경과 정책이 데이터 중심으로 변화되어가는 추세이다. 또한 국가예산이 투입된 데이터의 경우 개별소유가 아닌 공공재로 인식되어가고 있다. 이에 국내외적으로 데이터 공유와 활용확대에 대한 노력이 진행 중이다.

그 중 해양의 경우 높은 데이터 수집 비용과 고난이도 관측 기술이 요구되어 수집된 데이터의 효율적인 보관과 활용이 필요하다. 특히 심해, 극지, 황천 등의 해양공간은 수집이 어려울 뿐만 아니라 새로운 해양현상의 발견과 높은 해양 기술의 발전을 위해 데이터 수집 및 공유체계가 구축되어야 한다. 이를 위해 본 연구에서는 빅데이터 관련 정책동향과 기술동향을 조사하고 관련연구자들의 수요조사를 통해 극한해양공간에 최적화된 빅데이터 플랫폼을 구축방안을 제시하였다.

수요조사 결과 데이터 전송방식은 실시간 통신자료의 수집 및 서비스 수요가 많았고 일별 자료 수집을 통한 빅데이터 구축, 품질관리 및 데이터 관리, 분석, 시각화 지원체계 구축 등이 요구되었다. 또한 비공개 데이터에 대한 보안 강화가 요구되었다. 이에 실시간 극한해양환경의 기장비에서 생산되는 데이터의 수집, 저장, 분석, 서비스를 위한 플랫폼 구축 방안을 제시하였다.

이를 위해 첫 번째, 실시간 자료 수집 수요조사 및 연계 대상 기장비 정의를 수행하고 실시간 데이터 송수신 환경 분석 및 물리적 연결 구조 설계를 위한 실시간 통합관측 데이터 수집/관리 체계 구축 방안을 제시하였다. 두 번째, 극한해양에서 수집된 데이터 및 메타데이터 표준화 방안 모색과 수집된 데이터 기반의 통합 DB 조회 및 검색기능 설계를 위한 해양 과학 빅데이터 통합 관리 시스템 및 서비스 체계 구축방안을 제시하였다. 세 번째, 해양과학 빅데이터 통합관리 시스템 DB 등 유관 시스템 DB 연계, 분산저장 및 병렬 처리, 분석 및 처리된 데이터의 가시화 서비스를 위한 해양과학 데이터 플랫폼 구축 및 운영 방안을 제시하였다. 네 번째, 실시간 데이터의 수집 관리 시스템 DB 연계와 외부 연계를 위한 데이터 표준화 및 API 기술 검토와 시스템 관리 기능 개발, 통합 관측 데이터 수집 및 처리 기능 개발, 통합 관측 기장비 정보 관리 기능 개발을 통한 극한해양공간 통합관측 기장비 거버넌스 구축 및 운영 방안을 제시하였다.

본 연구는 극한해양공간 데이터의 실시간 수집, 표준화, 작업절차 방안 제시를 통한 플랫폼 구축의 기반을 마련하였다. 이는 향후 스마트 해양 관측 체계 및 운용 인프라 구축을 통한 극한지 해양 환경 데이터 수집과 통합관리 체계 구축에 이용 될

수 있으며 실시간 극한 해양 데이터의 개방으로 연구 범위 확대와 질 향상에 도움이 될 것으로 기대된다. 더 나아가 극한지 뿐만 아니라 다양한 정보 시스템에서 데이터를 수집, 정제 및 가공을 통한 분석으로 새로운 가치를 제공하며 계속해서 변화하는 문화와 환경에 선제적 대응이 가능할 것으로 사료된다.

감사의 글

본 연구는 해양수산부/해양수산과학기술진흥원의 지원으로 수행되었음(Project No. 1 525011610, KIMST-20210629)

참고문헌

- [1] Ministry of Science and ICT. The research data in the drawer is reborn as big data that is used together[Internet]. Available: <https://eiec.kdi.re.kr/policy/materialView.do?num=173274>
- [2] E. J. Shin and W. G. Jung, Spreading and Implications of Open Science Policy, STEPI Insight, pp.1-39, 2017
- [3] Y. S. Kim, "Human Genome Project" *Korean association of occupational health nurses*, Vol. 7, No.2, pp.35 - 41, 2000
- [4] LHCb publications. LHC The LHC Public results [Internet]. Available: https://lhcbproject.web.cern.ch/Publications/LHCbProjectPublic/Summary_all.html
- [5] H. G. Han, C. Y. Lee, T. H. Kim, J. R. Han and H. W. Choi, "Establishment of A WebGIS-based Information System for Continuous Observation during Ocean Research Vessel Operation", *Journal of the Korean Association of Geographic Information Studies*, Vol. 24, No. 1, pp. 40-53, March, 2016.
- [6] Trusted Data Services for Global Science. ostp public access memo 2013 [Internet]. Available: <https://www.worlddatasystem.org/files/ostp-public-access-memo-2013.pdf/view>
- [7] Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open science, open data, and open scholarship: European policies to make science fit for the Twenty-First Century. *Frontiers in Big Data*, 2, 43. <https://doi.org/10.3389/fdata.2019.00043>
- [8] Korea open access platform for researchers, Japan's Open Science Report[Internet]. Available: <https://www.koar.kr/oaintro/trend/news/newsDetail.do?recordNo=4955>
- [9] Ministry of Science and ICT. Promotion of research data sharing and utilization strategies[Internet]. Available: <https://now.k2base.re.kr/portal/trend/dmstcTrend/view.do?poliTr>

ndId=TRND000000000032710&menuNo=200044&pageIndex

[10] Joint Ministry. Data·AI Economy Revitalization Plan ('19~'23): Strategic Investment for Innovative Growth[Internet]. Available: https://policy.nl.go.kr/search/searchDetail.do?rec_key=SH2_PLC20190231106

[11] Together with the Democratic Party. Make the country a country-the 19th presidential election together with the Democratic Party's policy commitments[Internet]. Available: <https://theminjoo.kr/board/view/policyreference/90440>

[12] Korea Communications Agency, Promotion trends and implications for each element technology according to the stage of using big data, *Broadcasting and Communication Technology Issues & Prospects*, pp. 1-19, 2013

[13] Flume, Apache Flume [Internet]. Available: <http://flume.apache.org/>.

[14] R. P. Padhy, "Big data processing with Hadoop-MapReduce in cloud systems." *International Journal of Cloud Computing and Services Science* Vol. 2, No.1, pp. 16-27, February 2013.

[15] C. Rubén and M. Younas. "Emerging trends and technologies in big data processing." *Concurrency and Computation: Practice and Experience*, Vol. 27, pp. 2078-2091, October 2015.

[16] Github, scribe [Internet]. Available: <https://github.com/facebookarchive/scribe>.

[17] Y. Zhang, T. Cao, S. Li, X. Tian, L. Yuan, H. Jia and A. V. Vasilakos. "Parallel processing systems for big data: a survey." *Proceedings of the IEEE*, Vol. 104, No. 11, pp. 2114-2136, August 2016.

[18] Market and market, Hadoop Big Data Analytics Market worth \$23.5 billion by 2025 [Internet]. Available: <https://www.marketsandmarkets.com/PressReleases/hadoop.asp>

[19] Apach hadoop, HDFS Architecture [Internet]. Available: <https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>.

[20] B. J. Choi, H. J. Kim, J. H. Kim and S. H. Jin. "Data Analytics for CRM in the Age of Big Data" *Entrue Journal of Information Technology*, Vol. 11, No. 1, pp. 19-27, 2012.

[21] B. Y. Lee, J. T. Lim and J. S. Yoo "Utilization of Social Media Analysis using Big Data", *Journal of the korea contents association*, Vol. 13, No. 2, pp. 211-219, January 2013.

[22] M. M. Kang, S. R. Kim and S. M. Park. "Analysis and Utilization of Big Data", *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 30, No.64, pp. 25-32, 2012.



한현경(Hyeon-Gyeong Han)

2019년 : 부경대학교 대학원 (공학석사)

2019년~2020년: 한국환경정책평가연구원
2020년~현 재: 한국해양과학기술원
※관심분야 : 공간정보, 빅데이터 등



이철용(Cholyoung Lee)

2007년 : 인하대학교 대학원 (공학석사)
2018년 : 인하대학교 대학원 (공학박사
-공간정보공학)

2014년~2016년: (주)환경정보연구원 대표
2016년~현 재: 한국해양과학기술원 해양빅데이터센터 선임
기술원
※관심분야 : 해양빅데이터 분석/기획, 공간분석, GIS DB
설계/구축 등



박용길(Yong-Gil Park)

2011년 : 인하대학교 대학원 (공학석사)
2020년 : 인하대학교 대학원 (공학박사
-공간정보공학)

2020년~현 재: 한국해양과학기술원 박사후연구원
※관심분야 : 빅데이터(Bigdata), 공간정보시스템
(Geographical Information System)