

공공기관 해양공간정보 서비스 기반 데이터 처리 시스템 구축 가능성 연구

이정현^{1,*}, 하일운², 임효혁³, 김가원⁴
¹(주)한국해양기상기술

A Feasibility Study of Developing Data Processing System based on Marine Geospatial Open Government Data

Cheonghyun Lee^{1,*}, Ilwoon Ha², Hyohyuc Im³, Gawon Kim⁴
¹Korea Oceanic and Atmospheric System Technology, Seoul, Korea

요 약

웹 기술이 발전함에 따라 정부에서는 열린 정부 정책인 정부 3.0 의 일환으로 각 공공기관에서 취득한 데이터를 공개하여 민간인들이 이를 활용할 수 있도록 하였다. 이러한 데이터는 정형화되지 않은 원시 데이터이기에 사용 목적에 맞게 가공이 필요한데, 본 연구에서는 공공 데이터 중 해양 공간 정보 관련 데이터를 가공하는 처리 시스템 구축에 대한 타당성을 분석하고, 향후에 보완 및 추가적으로 연구해야 할 점들을 기술하였다. 분석을 위해 해양공간정보를 기반으로 하는 서비스 중 한 사례로 해양 민감 자원 표출 시스템을 적용하였다.

ABSTRACT

As web technology improves, Korean government decided to open government data into public so that people can use the government data for various purposes. To use those open data, those data must be refined to meet the system's own purpose. This paper presents a feasibility study of developing data preprocessing system of marine geospatial open government data and presents the future studies in conclusion. To analyze our approach, we applied the marine environmental sensitive data viewer system as a case study.

Keywords : Open Data (공공데이터), Data Processing System (데이터 처리 시스템), Geographic Information System (GIS), Marine Environmental Sensitive Resource (해양 민감 자원)

1. 서 론

웹 기술이 발전함에 따라 기존에는 활용하지 않았던 데이터들을 기반으로 하여 편리함을 제공하는 다양한 서비스들이 생겨나고 있다. 이러한 IT 발전 동향에 맞게 국내 정부에서는 정부 3.0 정책 하에 각 정부 기관에서 취득한 공공 데이터를 대중에 공개하였다 (Nam[2015]). 이렇게 공개된 원시 데이터는 매쉬업 (Engard[2009]) 을 통해 새로운 서비스를 창출하는 것이 가능한데, 본 연구에서는 해양정보 서비스에 활용 가능한 공공 데이터 서비스 현황에 대하여 분석하고, 이 데이터를 기반으로 하여 민감 자원 표출 시스템 구축에 대한 타당성을 검토한다. 그 결과, 원시 데이터는 구축하고자 하는 시스템에 바로 사용하기에 일관성이 없어 데이터 정제 및 사용 목적에 맞는 전처리 과정이 필요하다는 것을 알 수 있었고, 기존에 흔히 사용해온 관계형 데이터베이스인 오라클 데이터베이스만으로는 불필요하게 복잡한 설계 과정을 거쳐야 하는 불편함이 있다는 것을 알 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 기초 데이터로 사용할 공공데이터와 구축하고자 하는 해양 민감 자원 표출 시스템에 대하여 서술한다. 3장에서는 공공기관 해양정보 서비스 데이터 현황과 본 시스템과의 연계 방안을 분석하고, 데이터 처리 방안과 이를 위한 처리 시스템 설계에 고려해야 할 사항을 기술한다. 마지막으로, 4장에서는 결론 및 향후 연구를 기술한다.

2. 배경

2. 1 공공데이터

공공데이터는 정부 운영과정의 산출물로 공공기관이 생성 또는 취득하여 관리하고 있는 데이터를 말하며, 정부의 투명성 보장과 산업적 활용성을 극대화 하는 등 열린 정부 (Janssen, Charalabidis, and Zuiderwijk[2012]) 정책의 일환으로 국민들에게 공개하는 것을 의미한다 (이현정[2014]). 국내에서는 정부 3.0 (Nam[2015]) 정책 하에 행정자치부에서 공공 데이터를 통합적으로 관리 및 배포하는 공공데이터포털 (www.data.go.kr)을 운영하고 있고, 이 외에도 서울 열린 데이터광장 (data.seoul.go.kr), 경기데이터드림 (data.gg.go.kr) 등 각 행정구역에서 주관하여 데이터를 공개하고 있다. 또한, 해양수산부 등 각 정부 기관에서도 각 기관에서도 취득한 데이터를 공개하고 있다. 공공 데이터를 활용한 서비스의 예로는 버스 위치 데이터를 활용하여 실시간 버스 정보를 제공한 ‘서울버스’ 앱, 보건복지부 건강 보험심사평가원 병원정보 DB 를 활용하여 사용자의 입력 정보를 토대로 원하는 병원을 찾아주는 ‘메디라떼’ 등이 있다 (권성아[2015]). 본 연구에서는 해양 공간정보위주의 공공 데이터를 기초로 하여 이를 표출해주는 시스템을 사례로 활용한다.

2. 2 해양 민감 자원 표출 시스템

해양 민감 자원 표출 시스템은 해양 유류유출 오염 사고 발생 시 사회, 경제적인 관점에서 상대적으로 민감하게 영향을 받는 지역 (노영희[2016])을 시각적으로 표출해 주는 시스템이다. 이 시스템은 유류유출 확산 예측 시스템의 일부로서, 유류유출 사고 발생 시에 사회적으로 영향을 받는 해수욕장이나 환경적으로 영향을 받는 갯벌 등을 일괄적으로 정리 및 통합하여 사용자의 요청에 맞게 하나의 시스템에서 표출한다. 이 시스템은 정부의 열린 정부 정책 하에 공개된 다양한 해양, 기상 및 사회 공공 데이터 정보를 한 곳에 모아 표출하여 주고, 필요 시 이러한 자료를 기초로 하여 민감한 정도를 정량화 하는 역할 (노영희 and 김충기[2016])을 수행한다.

3. 본 론

3. 1 공공기관 해양 정보 서비스 데이터 현황

시스템 구축에 앞서 민감도 평가를 위해 수집 가능하고 평가에 의미가 있는 자료를 정리하였다. Table 1은 해양 민감 자원 표출 시스템 구축을 위해 수집한 데이터를 연계 가능성과 함께 표로 정리한 자료이다. 총 135건의 데이터를 수집하였고 원시데이터 항목 중 내부 자료에 해당하는 “기존 DB”를 제외한 항목들은 정부의 데이터 공개 정책에 의거하여 공공데이터로 공개된 자료이다. 예를 들면, Web API 의 해파리정보, 해양조석정보는 공공데이터포털에서 수집하였고, 양식장, 산업단지, 환경보전해역 정보는 해양수산부에서, 기상 데이터는 기상청의 기상자료개방포털에서 수집한 자료이다. 자료 수집은 구축하고자 하는 시스템의 요구사항 분석을 거쳐, 수집자가 직접 각 공공기관마다 어떠한 데이터를 보유하고 있는지 확인 후 해당 데이터의 형태 및 상세 정보를 일일이 확인하는 방식으로 진행하였다.

Table 1. 원시 데이터 수집 및 연계 가능성

원시데이터	데이터 종류	항목	연계 방안	항목 건수
Web API	HTTP (REST)	해파리정보, 해양조석정보 등	API 연계 (주기적 갱신)	27건
Web API	HTTP (Raw)	해양지명정보, 해양관측정보 등	HTTP 파싱 (주기적 크롤링)	5건
기존 DB	Oracle	RIMS (일반특성, 생물자원 등)	DB 연동 (주기적 갱신)	23건
기존 DB	xml	RIMS (관할구역 좌표 등)	수동 입력	4건
기존 DB	shp	재난정보공동활용시스템 (사고선박정보 등)	수동 입력	11건
기타	shp 파일	양식장, 산업단지, 환경보전해역 등	수동 입력	30건
기타	기상 데이터 (csv, nc 등)	단기예보, 관측자료 등	FTP 연계 (업로드 후 자동처리)	20건
기타	csv, zip, png, hwp 등	파고부이, 해양기상부이 등	수동 다운로드 후 입력	15건

Table 1의 데이터 종류 항목은 수집한 데이터의 형태 정보를 기입한 것으로, HTTP (REST)는 웹 서비스 API 형식으로 제공하는 서비스를, HTTP (Raw)는 웹 페이지 형식, Oracle은 데이터베이스 형태로, 그 외에 xml, shp 등 각 파일로 제공되는 데이터를 의미한다. 연계 방안 항목은 최초 조사 단계에서 수동으로 수집한 이후, 시스템 구축 시에 어떻게 연계를 할지에 대한 방안을 기입하였다. 예를 들어, API 연계 및 DB 연동은 시스템에 자동으로 Web API 형식으로 연계하여 자동으로 주기적 갱신하는 것을 의미하며, FTP 연계는 시스템에 데이터 서버를 열어 두어 FTP로 직접 혹은 타 시스템과 연동하여 주기적으로 자동 업로드 하는 방식을, 그리고 수동 입력 및 수동 다운로드 후 입력은 사용자가 직접 수집하여 입력하는 것을 의미한다. 이렇게 조사를 거친 데이터는 적용하고자 하는 시스템에 맞게 처리 과정을 거쳐 사용한다.

3.2 데이터 처리 방안

앞서 수집한 원시 데이터를 시스템에서 사용하기 위해서는 데이터 처리 시스템에 데이터를 입력 및 가공하는 선행 작업이 필요하다. 이를 위해, 조사한 원시 데이터를 기초로 하여 해양 민감 자원 표출 시스템의 목적에 맞게 가공하기 위한 처리 방법에 대한 정보를 Table 2에 정리하였다. 처리 방안 구상을 거치면서 원시 데이터의 24건이 중복 데이터이거나 시스템의 목적에 맞지 않는 것으로 확인되어 사용하지 않기로 결정하여 총 135건에서 110건으로 축소되었다. Table 2의 분류 항목은 해양 민감 자원 표출 시스템의 목적에 맞게 데이터 분류를 정의한 것이다. 데이터 구조 항목은 각 데이터의 파일 형태 등 구조를 기입한 항목으로, 각 데이터 구조에 따라 처리 방법이 달라진다. 원시 데이터 간에 보유하고 있는 정보는 다양한 출처에서 다양한 목적으로 제공된 데이터이기, 데이터 간의 일관성이 없고 최종 시스템에 사용하기에 부합하지 않는 항목들이 많다. 따라서, 최종적으로 시스템에서 사용할 수 있도록 일관성 있게 정리하는 과정이 필수적이다.

Table 2. 원시 데이터 구조별 처리 방안

분류	데이터 구조	건수	처리방법	합계
사회경제	Xml 파일	2건	Xml 파싱 후 입력	14건
	ESRI Shapefile	10건	Shp 파일 내부 구조 확인 후 입력	
	Web API	2건	웹 연동하여 주기적으로 읽어옴	
기상정보	Zip 파일 (구조 제각각)	29건	파일 구조에 따라 맞는 방법 사용	29건
	Web API	2건	웹 연동하여 주기적으로 읽어옴	14건
생태	RIMS DB	1건	DB에 직접 접속 연동하여 입력	
	ESRI Shapefile	11건	Shp 파일 내부 구조 확인 후 입력	
환경	RIMS DB	3건	DB에 직접 접속 연동하여 입력	19건
	ESRI Shapefile	13건	Shp 파일 내부 구조 확인 후 입력	
	Web API	3건	웹 연동하여 주기적으로 읽어옴	
재난안전	CSV 파일	1건	CSV 파싱 후 입력	5건
	Hwp 파일	4건	수동으로 입력	
해양정보	Web API	20건	웹 연동하여 주기적으로 읽어옴	37건
해양정보	Http	5건	웹 파싱으로 연동하여 입력	
해양정보	DB	12건	DB에 직접 접속 연동하여 입력	
기타	Xml 파일	2건	Xml 파싱 후 입력	2건
합계	-	-	-	110건

3. 3 데이터 처리 시스템 구성 및 고려 사항

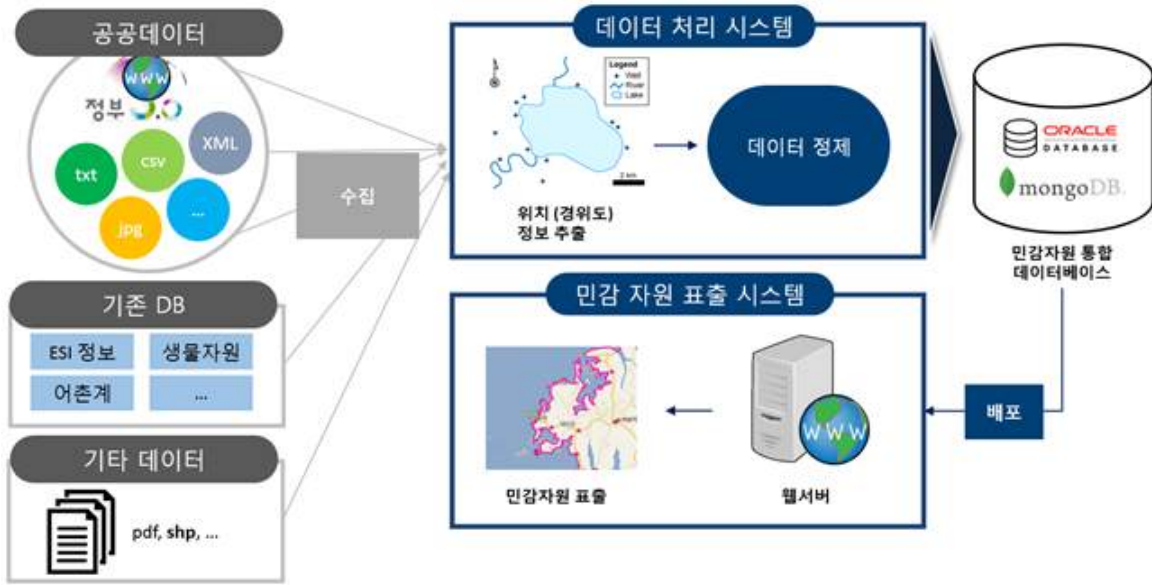


Fig. 1. 데이터 처리 시스템을 포함한 전체 시스템 구성.

수집한 원시 데이터를 체계적으로 정리하고, 정리된 데이터를 민감 자원 표출 시스템에 표출하는 일련의 과정을 Fig. 1 과 같이 구성하였다. 수집한 원시 데이터는 데이터 처리 시스템에 입력되어 해양오염사고 데이터 정제 과정을 거쳐 민감 자원 통합 데이터베이스에 저장된다. 저장된 데이터는 민감 자원 표출 시스템에서 정의된 방식에 따라 지도 위에 표출한다.

민감 자원 표출 시스템에 맞게 데이터를 정제하는 과정은 아래와 같이 진행된다.

1. 수집한 데이터를 데이터 처리 시스템에서 불러온다.
2. 원시 데이터에서 공간 정보를 추출한다.
3. 원시 데이터에 공간 정보를 경위도 정보로 일관화 작업을 거친 후 저장한다. 예를 들어, 원시 데이터에 주소 정보만 있을 경우 Geocoding 기법 (Svennerberg[2010])을 사용하여 경위도 값을 산출하여 저장한다.
4. 공간 정보 외의 부가적인 정보는 데이터 내용 및 형태에 상관없이 바로 저장한다.

정제된 데이터는 경위도 정보만을 일관화 하여 MongoDB 에 저장되는데, 각 원시데이터에 일관성이 없는 특성상 기존 관계형 데이터베이스인 오라클만을 사용하면 데이터베이스 설계가 불필요하게 복잡해지고 성능 저하가 우려되어 일관성이 없는 원시 데이터를 공간 정보를 추출 후 형태에 상관없이 JSON 에 key 값을 유연하게 지정하여 데이터를 입력할 수 있는 MongoDB 를 함께 사용하였다 (Han et al.[2011]). 따라서, 관계형 데이터베이스인 오라클 DB 에는 원시 데이터 수집 정보와 같은 메타 데이터 정보만을 저장하고, 정제된 데이터는 MongoDB 에 저장하도록 민감 자원 통합 데이터베이스 설계하였다. 또한, MongoDB 는 지형공간정보를 색인하고 검색하는 기능을 지원하기에 (MongoDB[n.d.]), 구축하고자 하는 시스템의 데이터베이스 설계에 적합하다고 판단하였다.

3. 결론 및 향후 연구

본 논문에서는 공공기관에서 제공하는 해양공간정보 데이터를 기반으로 하여 해양 민감 자원 표출 시스템에 활용할 데이터를 처리하는 시스템 구축에 필요한 것이 무엇인지 분석하고, 구축 가능성에 대하여 검토하였다. 그 결과 공공데이터는 다양한 기관에서 다양한 목적으로 생성되었기에, 본 시스템 구축에 사용하기 위해서는 전처리 과정이 필요하다는 것을 알게 되었고, 각 데이터의 형태 및 구조를 파악하여 그에 맞는 처리 방법을 정의하였다. 또한, 처리 시스템에 입력되는 데이터는 공간 정보를 추출하고 그 외의 데이터를 일관화 하여 저장하는 작업이 추가로 필요하다는 것을 알게 되었으며, 데이터가 저장될 데이터베이스는 널리 사용되는 관계형 데이터베이스만으로는 부족하다는 것을 알 수 있었다. 향후 연구로는 원시 데이터 처리를 더 효율적으로 할 수 있는 방법을 조사하고, 현재 구성한 민감 자원 통합 데이터베이스가 성능, 효율적인 면에서 타당한지 여부를 검토할 것이며, 나아가 정제된 해양공간

정보 데이터 추가적인 활용 방안에 대하여 연구하고자 한다.

후 기

본 연구는 해양경찰청의 재원으로 재난안전기술개발사업단의 지원을 받아 수행된 연구임[KCG-01-2017-05]

References

- [1] 권성아 and 이지연, “국내 공공데이터 개방현황 및 활용사례 분석,” in 한국정보관리학회 학술대회 논문집, 2015, pp. 37 - 45.
- [2] 노영희 and 김충기, “해양 유류유출 오염으로 인한 환경 및 사회·경제적 민감자원 선정 방안,” in 한국해양환경·에너지학회 학술대회논문집, 2016, pp. 153 - 154.
- [3] 노영희 and 김충기, “해양 유류유출 오염으로 인한 사회·경제적 민감자원 선정 및 지수화 방안,” 환경영향평가, vol. 25, no. 6, pp. 402 - 413, 2016.
- [4] 이현정 and 남영준, “우리나라 공공데이터의 이용활성화 방안에 관한 연구,” 정보관리학회지, vol. 31, pp. 249 - 266, 2014.
- [5] G. Svennerberg, “Chapter 10: Location, Location, Location,” in Beginning Google Maps API 3, Apress, 2010, pp. 211 - 242.
- [6] N. C. Engard, “What is a Mashup?,” in Library Mashups: Exploring New Ways to Deliver Library Data, Information Today, Inc., 2009, pp. 3 - 17.
- [7] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, “Benefits, adoption barriers and myths of open data and open government,” Inf. Syst. Manag., vol. 29, no. 4, pp. 258 - 268, 2012.
- [8] T. Nam, “Challenges and Concerns of Open Government: A Case of Government 3.0 in Korea,” Soc. Sci. Comput. Rev., vol. 33, no. 5, pp. 556 - 570, 2015.