

지오빅데이터 오픈플랫폼 모니터링단 최종 보고서

데이터의 신뢰도, 정확도, 지속가능한 확장성을 위한, 지오빅데이터
오픈플랫폼 개선 사항 및 발전방향

차 례

| | |
|---|-----------|
| ㄱ. 들어가는 말..... | 2 |
| ㄴ. 마당 글..... | 2 |
| I. 오픈플랫폼 활용 사례 및 활성화 제안..... | 2 |
| 1. <u>위치정보 기반 통합적 암석 및 지질 데이터 구축 제안</u> | 2 |
| (1) 서론 | |
| (2) 본론 | |
| 1) GPS 정보가 있는 사진 자료와 연계하여 GIS 지질 정보 처리 | |
| 2) 데이터 사용자 및 생산자의 수요에 맞추어 데이터 생산 및 보완 | |
| (3) 소결론 | |
| 2. <u>용어 순화적 관점에서 바라본 ‘지오빅데이터 오픈플랫폼’에 대한 제안</u> | 5 |
| (1) 서론 | |
| (2) 본론 | |
| (3) 소결론 | |
| 3. <u>‘지오빅데이터 오픈플랫폼’에 접근성 향상 방안에 대한 제안</u> | 7 |
| (1) 서론 | |
| (2) 본론 | |
| 1) 구글(Google), 네이버(Naver), 다음(Daum)에서 국내외 지도 공유 서비스를 통한 접근성 향상 방안 | |
| 2) ‘지오빅데이터 오픈플랫폼’ 영문 양식 추가 보완 운영 제안 | |
| (3) 소결론 | |
| 4. <u>‘지오빅데이터 오픈플랫폼’에 신규 등록 Raw Data의 4단계 검수 절차에 대한 제안</u> | 9 |
| (1) 서론 | |
| 1) 데이터 신뢰도 향상을 위한 검증 절차 보완의 필요성 | |
| (2) 본론 | |
| 1) 빅데이터 산업 검증 절차의 도입을 통한 자료 신뢰성 향상 | |
| 2) 1차 검수자, 2차 검수자, 전문가 검수자, 총괄 검수자 및 총괄 검수자의 제안 | |
| (3) 소결론 | |
| II. 오픈플랫폼 오류 개선..... | 10 |
| 1. <u>기존에 등록된 데이터 오류 보고</u> | 10 |
| (1) 서론 | |
| (2) 본론 | |
| 1) 기존 데이터 오류 내용의 보고 | |
| (3) 소결론 | |
| ㄷ. 나오는 말..... | 13 |

| | |
|-------------|----|
| <참고문헌>..... | 13 |
|-------------|----|

그림 목차

| | |
|--|----|
| 그림 1 ‘지오빅데이터 오픈플랫폼’ 홈페이지 메뉴 구성 화면..... | 6 |
| 그림 2 사례 ㄱ. 암편사진을 현미경 사진으로 잘못 기재..... | 11 |
| 그림 3 사례 ㄴ. 암편사진을 현미경 사진으로 잘못 기재..... | 11 |
| 그림 4 사례 ㄷ. 등록 자료의 제목 표시에 있어서 단순 오탈자..... | 12 |

ㄱ. 들어가는 말

2016년 3월 9일부터 10일, 12일, 13일, 15일 열린 ‘알파고(AlphaGo)’라는 구글 딥마인드(Google DeepMind)가 개발한 인공지능 바둑프로그램이 당시 최강 바둑 기사 중에 하나인 이세돌 9단을 4대 1일로 완승했다.¹ 수많은 데이터가 적절히 프로그래밍 되어서 올린 성과였다. 정확한 데이터가 많은 능력을 발휘함을 전 세계가 목격했다. 데이터를 가진 그룹이 많은 능력을 가진 세상이 도래하는 것은 분명해 보인다.

인터넷을 통해 많은 정보에 손쉽게 접근가능한 세상이 되었지만 그 정보를 어떻게 가공하느냐에 따라 그 효용은 다를 것이다. 많은 정보가 공유되는 세상 속에서, 앞으로, 정보 공유화가 세계를 유토피아로 만들지, 디스토피아로 만들지는 그 쓰임에 따라 달려 있다고 본다. 다만, 정보의 영향력이 막대한 세상 속에서 살고 있음을 부정할 수 없기에, **많은 나라의 과학기술분야에서 체계적인 정보시스템을 확장성에 기반을 두고 확장 개방형 과학 정보 시스템을 구축해 나가고 있다.**

‘지오빅데이터 오픈플랫폼’도 그러한 흐름 안에서 기획되는 지질 정보 시스템일 것이다. 이러한 시스템이 잘 정착하기 위해서 많은 요소들이 필요하겠지만, 지속적인 모니터링을 하면서, **데이터의 신뢰도 향상과 정확도 향상, 지속가능한 확장성의 측면이 중요하다고 보고 이를 염두에 두고, 이 플랫폼의 개선 방안을 고찰해 보았다.**

크게 5개의 마당글 중 4개는 ‘오픈플랫폼 활용 사례 및 활성화 제안’으로서, **위치정보 기반 통합적 암석 및 지질 데이터 구축, 용어 순화적 관점에서 바라본 ‘지오빅데이터 오픈플랫폼’, ‘지오빅데이터 오픈플랫폼’에 접근성 향상 방안, ‘지오빅데이터 오픈플랫폼’에 신규 등록 원(原)자료(Raw Data)의 4단계 검수 절차에 대한 제언 및 검토를 담고 있다.** 마지막 1개의 마당글은 ‘지오데이터 오픈플랫폼 오류 개선’에 대한 내용으로 **기존 데이터의 오류를 보고하면서, 신뢰성 향상을 위한 총체적인 데이터의 검수 보완 절차를 제안하고 있다.**

ㄴ. 마당 글

I. 오픈플랫폼 활용 사례 및 활성화 제안

1. 위치정보 기반 통합적 암석 및 지질 데이터 구축 제언

(1) 서론

‘지오빅데이터 오픈플랫폼’ 사업은 여러 가지 유용한 지질 학술정보를 전문가 영역부터 대중적 영역까지 포괄하여 이용할 수 있는 플랫폼으로 활용될 수 있다고 본다. 그러기 위해서는 무엇보다도 대중적인 측면을 고려하여, 데이터의 접근성 향상이 중요하다고 본다. 데이터를 쉽게 접근할 수 있는 방안으로는 지도 서비스와 연계하여 내가 필요한 지역의 논문 및 암석 정보를 직관적으로 쉽게 접근할 수 있는 지도 기반의 정보 구축을 해 놓는다면 더 쉽게 자료를 찾을 수 있고, 활용 가능하리라 본다.

현재도 다양한 주제로 데이터들에 접속가능하다. 하지만, 지도서비스와 보다 직관적인 접근이 가능하도록 홈페이지 메인 화면에 지도 정보 서비스로 바로 링크해서 접근할 수 있도록 개선하고, 지도 정보 안에서 다양한 지질정보를 규격화하여, 접근할 수 있게 개선한다면, 보다 쉽게 정보를 활용할 수 있을 것으로 본다.

‘지오빅데이터 오픈플랫폼’이 지향하는 바는 학술적인 정보의 검색과 더불어, 대중적인 영역까지 함께 아우를 수 있는 플랫폼으로 그 방향성을 설정하고 가는 것이 확장성 면에서 유용하다고 본다. 다시 말해 전문가부터 학생들 그리고 일반 대중까지도 쉽게 접근가능하고 이해할 수 있는 정보들을 꾸준히 지속적으로 생산하는 체계가 필요하다고 본다.

한국의 지질 전반부터 지역적인 지질정보의 특성까지 한눈에 볼 수 있게 지속적으로 개선하면서, 세부적으로 접근할 수 있는 ‘지도 정보 시스템’ 기반의 화면을 신설한다면, 그 지역에 대한 개략적인 정보부터, 자세한 정보까지 상세한 정보까지 정보의 수위를 계층적으로 관리하는 것은 유용하리라 본다. 그러기 위해서는 데이터 플랫폼에 올라온 정보들을 위치정보에 기반하여 다시 분류하고, 링크해서 찾아갈 수 있도록 데이터베이스화하여 관리한다면 쉽게 전반적인 사항부터 세부적인 사항까지 찾아 들어갈 수 있도록 다양한 기준선을 마련하여, 분류하는 체계가 필요할 것이다.

대중적인 수준에 정보부터 시작하여, 전문적인 영역의 설명까지 다양한 눈높이로 체계적으로 자료들에 대한 사전 설명을 첨부해 놓는 것이 플랫폼을 활성화하는 데 도움이 된다고 본다.

오픈 플랫폼의 특징은, 많은 사람들이 정보를 올리는 데, 그 원(原)자료들을 중장기적으로 지속적으로 검토하고 개선하면서, 데이터의 신뢰도와 정확도를 높이는 것이 무엇보다 중요하다고 본다.

국내외에 이미, 많은 오픈 소스 지질 자료가 있는 만큼 그 연계성을 고려하여 지질 정보 구축을 제언하고 싶다. 암석 정보에 대한 지질학적이고 체계적인 오픈 소스 구축할 때, 이미 시행하고 있는 국내외 다양한 지질 관련 오픈플랫폼의 양식을 참고하는 것을 제안하고 싶다.

(2) 본론

1) GPS 정보가 있는 사진 자료와 연계하여 GIS 지질 정보 처리

이미 구글어스 및 여러 지리정보 기반의 위도와 경도값이 사진정보 안에서 핸드폰 및 카메라에 사진을 찍을 때, 위치 정보가 예전에 비해 손쉽게 정확하게 입력된다. 그 정보값에 기초하여, 체계적으로 암석 및 노두 하천 및 토양의 특성을 일정한 등록 양식을 정해주어 등록하게 한다 면, 지형학적 암석학적 학술 조사 및 대중적 지질정보 파악에 기초가 될 것으로 보인다.

이용자가 올린 자료는 크게 2가지로 목적에 맞게 양식화를 해 놓으면 이용에 편의성이 있을 것으로 보인다. 첫째는 지질정보를 담은 사진의 세부 GPS 정보와 사진에 대한 설명을 함께 첨부할 수 있도록 일정한 양식화를 해서 자료를 정리한다면 좋을 것이다.

예를 들어, 사진의 GPS 정보, 최소 및 최대 해상도, 스케일바의 유무 표시, 사진을 찍은 일시와 환경 정보 표시, 노두에서 있는 암석인지, 암석의 종류, 전석인지 노두에 붙은 암석인지 같은 지질정보, 박편사진과 XRD 및 화학 분석 자료의 유무와 같은 것들을 확인할 수 있는 양식지를 제작하여, 자료를 정리하면 좋을 것이다.

자료를 정리하는 파일 양식은 널리 이용되는 엑셀 파일 또는 QGIS로 할 것인지, 아니면 또 다른 기본 프로그램을 제작·지정해 주고, 그에 맞는 양식에 지질정보를 구축할 수 있도록 하면, 자료가 방대하게 생성되었을 때, 정량적으로 검색 가능한 데이터베이스로 구축할 수 있을 것이다.

예를 들어, 일정한 양식에 등록된 사진 중에, 엑셀 또는 QGIS를 활용하여, 화강암이 몇 개소에 있는지 한눈에 검색할 수 있는 시스템을 구축하면, 암석 정보를 효율적으로 데이터베이스화하는 기초자료가 될 것으로 보인다.

다른 하나의 양식으로는 일정한 규격 없이 올릴 수 있는 지질정보 시스템을 만들어 놓아서, 정보의 다양성을 열어 놓는 것이 필요하다고 본다. 하지만, 일정한 규격으로 통일되지 않은 양식의 지질 정보를 등록한다면, 그 정보에 대해서는, 어떤 기준으로 자료를 정리했는지 명확한 설명과 함께, 자료 등록 전에 적절성에서 대한 시스템 운영자의 세부 검토 및 수정 절차를 거치는 쪽으로 시스템을 만들었으면 좋겠다.

2) 데이터 사용자 및 생산자의 수요에 맞추어 데이터 생산 및 보완

이러한 지질 정보 시스템은 발전하는 기술에 맞게 자료의 생산자와 소통하면서, 계속 보완하면서, 지질정보를 일정하게 구축해 나가면서 지속적으로 관리하고 개선할 필요가 있어 보인다.

한국적인 환경에서는 지리정보를 찾는데, '네이버'(Naver)나 '다음'(Daum)의 '맵'(Map) 서비스를 많이 활용하고, 국제적인 환경에서는 구글 지도를 많이 활용하고 있다. 이러한 환경에서, '지오빅데이터 오픈플랫폼'의 정보를 관련 업체와의 업무 협약과 자료 공유를 통해서, 그 소스를 대형 검색 엔진에서 직접 검색 가능하게 시스템을 구축한다면, 지질정보에 대한 접근성 향상의 측면에서 매우 유용할 것으로 보인다.

더 나아가 다른 지질정보와 연계성이 높은 다른 환경학적 또는 지리학적 데이터를 접목해서 볼 수 있도록 다른 기관의 오픈 데이터와도 접목하려는 노력을 기관들 차원에서 이루어진다면, 더 유용할 것으로 보인다.

처음에 이러한 다양한 정보를 융합하여 표시할 때는, 링크를 통해서, 지질데이터와 접목해서 유용한 데이터를 접근할 수 있는 입구만 통합하는 것으로부터 시작하여, 생산된 데이터의 호환성을 향상시키는 방향으로, 점점 하나의 시스템으로 묶어 나가는 작업을 장기적으로 추구한다면 좋을 것이다.

(3) 소결론

'지오빅데이터 오픈플랫폼'을 활용해 본 결과, 지질도를 검색하고, 그 자료와 연계된 지질

도 및 여러 자료를 찾는 데에는 다른 학술 검색 엔진보다 우수한 측면이 있다고 본다. 하지만 다른 학술 검색 엔진은 이미 검증된 논문 위주로 검색된 만큼 신뢰성이 보장된 측면이 있지만, 지오빅데이터는 정식으로 발행된 논문뿐만 아니라 오픈 플랫폼의 특성상 검증이 필요한 다양한 정보를 수집하는 측면이 있는 만큼, 그 신뢰성을 향상시키고 담보할 수 있는 도구적 장치가 절대적으로 필요해 보인다. 그 방법으로는 정보 생산 주체에 대한 관리와 더불어, 일정한 등록 양식의 규격화와 생산 내용의 검토, 전문가 및 조사자 및 감수자를 통해 다방면으로 자료의 신뢰성 확인이 필요해 보인다.

2. 용어 순화적 관점에서 바라본 ‘지오빅데이터 오픈플랫폼’에 대한 제언

(1) 서론

‘지오빅데이터 오픈플랫폼’을 지속적으로 모니터링하면서 그 이름에서부터 느껴졌던 것은 외래어 사용에 있어서 순화의 노력이 필요하다는 인식이었다. 이미 우리 사회에 영어에 근원을 둔 외래어가 많이 사용되고, 이미 어색하지 않은 정도까지 들어와 있다는 것을 부정할 수 없다. 하지만 공공기관은 용어를 생성하고 정립하는 데에서 하나의 주체적인 역할을 하는 만큼, 조금 더 순화한 표현을 쓰는 노력이 필요하다.

표현을 ‘생성하는 대표적인 주체라 할 수 있는 정부, 언론, 학계, 민간’² 중에 한 축을 이루고 있다는 각성과 더불어, 용어 선정 및 사용에 신중을 기할 필요가 있다. 홈페이지 내에서 사용되는, 용어 사용에 있어서 국문법적 검토와 학술적 검토를 더 세심하게 고려하는 것이 알기 쉽고 이해하기 쉬운 환경을 만들어, 이용의 편의성이 향상되는 만큼, 장려해야 할 것이다.

특히, 지질자원연구원은 지질학 분야의 영향력 있는 연구기관으로서 위상을 고려해 볼 때, 학계의 일부라 봐도 무방할 것이다. 그 영향력을 생각할 때, 용어의 엄밀성을 생각하면서 용어를 쓰는 것이, 그 위상에 맞는 신뢰성을 얻어가는 하나의 가시적인 방법 가운데 하나이다.

정제된 용어의 사용, 특히 학술적 용어 사용과 용어 순화는 의미 전달의 정확성의 측면에서 중요하고 자료의 신뢰성을 높여 줄 것이다. ‘지오빅데이터 오픈플랫폼’의 너무 많은 외래어 사용에 초점을 맞추어 순화할 방법을 생각해 보고자 한다.

(2) 본론

1) ‘지오빅데이터 오픈플랫폼’의 용어 순화

먼저 ‘지오빅데이터 오픈플랫폼’이라는 이 서비스의 이름은 5개의 일상적으로 쓰는 외래어 또는 외국어로만 구성되어 있다. 올바른 국문으로 순화하자면, ‘지질 빅데이터 개방형 플랫폼’이 적절하다고 본다. 외래어를 대체하면서 의미가 변화하지 않는 것이 있다면, 의미가 변화하지 않는 순화어나 통용되는 국문 학술 용어를 쓰는 것이 더 적합할 것이다. 그 채택에 있어서는 국문법에 있어 공신력 있는 ‘국립국어원’에서 권고하는 순화어나 학계 용어집에서 채택하고 있는 용어를 사용하는 것이 더 적합할 것이다.

용어 사용에 있어서 가장 기본적으로 검토되어야 할 것은 『표준국어사전』과 『지질학사전』에서 쓰려는 단어가 등재된 적합한 용어인지 검토하는 것을 시작으로, 만약에 적합한 용어가 없다면, 국문법에 맞는 조어법을 바탕으로 적절한 말로 신중을 다하여, 용어를 만들어 사용해야 할

것이다.

‘외래어와 외국어를 구별하는 기준으로서 1) 동화의 정도, 2) 국어 문장 속에서의 자연스러움의 정도, 3) 외국어 의식이 있는가 없는가 4) 우리 사회에서 널리 쓰이는가 그렇지 않은가 등을 든다.’³ 그러한 기준에 맞추어 ‘지오’(Geo)는 ‘지질’로 ‘오픈’(Open)은 ‘개방형’으로 국어 조어법에 맞추어 순화할 수 있을 것이다.

이미 ‘지오빅데이터 오픈플랫폼’이라는 용어를 사용하는 만큼, 순화된 표현으로 ‘지질 빅데이터 개방형 플랫폼’을 병기(併記)하고, 반대로, 우리말로 표현된 것들은 영어를 병기하는 것(Geo Bic Data Open Platform)은 영어 문화권에 있는 많은 인터넷 사용자가 국내 지질자료에 접근성을 향상시키는 방법으로써, 적절하게 활용할 수 있는 창구적인 역할을 할 것이다

2) ‘지오빅데이터 오픈플랫폼’의 ‘메인 메뉴’의 용어 순화.

홈페이지 모니터링에 있어서 눈에 들어오는 상단의 메뉴에 ‘데이터셋’, ‘정보서비스’, ‘My 데이터’, ‘소개’로 메뉴를 만들어 놓았다.

여기서 ‘데이터셋’은 그 의미가 무엇인지 그 단어만 들었을 때, 정확히 구분이 힘들다. ‘셋’이라는 용어가 무엇을 뜻하는지, 영어를 병기하거나, ‘데이터세트’와 같은 외래어로 표기하거나, 좀 더 순화된 표현으로 ‘자료 정리’ 또는 ‘자료 모음’이 하위 메뉴의 구성(하위 메뉴로, ‘검색’과 ‘활용 갤러리’가 있다.)을 보았을 때 더 적합할 것으로 보인다. 영문 홈페이지를 참고하면 ‘Dataset’이라는 용어를 활용했는데, 그렇다면 ‘데이터세트’ 또는 ‘자료 모음’ 또는 ‘자료 정리’가 적절한 외래어 표기 방법일 것이다.



그림 1 ‘지오빅데이터 오픈플랫폼’ 홈페이지 메뉴 구성 화면

그러한 국문 사용의 기준 아래, ‘My 데이터’ 또한, ‘나의 자료(My Data)’ 정도로 순화하는 것이 더 정제된 표현이라고 본다.

이렇게 홈페이지에서 빈번히 드러나는 영어식 외래어 및 외국어 표현의 사용은 몇 가지 원인이 있을 것이다.

먼저, 우리 학계 및 일상생활에 영어가 깊숙이 들어오면서, 그에 대한 순화 노력이 필요 없다는 인식도 한몫을 한다고 본다. 그럼에도 불구하고, 영어가 국제적인 표준언어로 많이 사용되고 있지만, 그 언어를 우리말로 순화하려는 노력은 여러모로 단점보다 장점이 많기 때문에, 필요하다. 여러 나라에서 국가적인 차원에서 학술 용어를 순화하려는 노력을 지속하고 있다. 또한 ‘지오’

빅데이터 오픈플랫폼'의 기본적인 목적을 많은 사람들이 이용 가능한 플랫폼을 목표로 한다면, 영어가 익숙지 않은 세대에 대한 접근성 및 이해도의 향상적인 측면에서 보다 일상적인 언어로 국문법적인 오류가 없는 표현으로 순화할 필요성이 있다고 본다.

또한 제도적으로, 홈페이지의 용어 사용에 있어서, 다양한 전문가들의 검토가 있어야 할 것이다. 특히, 지질 및 지구물리 전문가뿐만 아니라, 용어 선정에 있어서, 국문법적인 오류가 없는지도 주기적으로 국문법 전문가 그룹에 검토를 받을 필요성이 있어 보였다.

(3) 소결론

'지오빅데이터 오픈플랫폼'이 많은 계층과 사람들이 이용할 수 있는 개방형 플랫폼으로 사용되면서, 신뢰성을 높이는 하나의 방법으로, 외래어 및 외국어 사용에 있어서 순화적인 표현을 쓰려는 노력을 한다면 신뢰성 향상의 측면에서 더 개선될 것으로 보인다.

현재의 홈페이지에서 용어 사용에 있어서, 적절한 국문법적 검토가 다소 미약했던 것으로 보인다. 홈페이지 용어 사용에 있어서 앞서, 통용되는 국문 검토 기관, 예를 들어, '국립국어원'의 '외래어 사용 표준 지침'에 나와 있는 순화 표현을 사용하는 것이, 정부출연연구기관의 하나인 지질자원연구원이 좀 더 신뢰할 수 있는 정보를 생산하는 기관으로 인식될 수 있는 수단이 될 것이라고 본다.

'지오빅데이터 오픈플랫폼'은 많은 사람들이 쉽게 사용하면서도 **신뢰성을 향상**시키는 것을 목표로 삼아야 된다고 본다. 그러기 위해서는 **하나의 용어 사용에 있어서도 대중의 눈높이에 맞춰서 쉽게 표현하려는 노력과 전문성을 잃지 않으려는 노력을 병행하면서, 홈페이지를 개선하고 자료를 개선하려는 노력을 지속해야 한다.** 동시에 예민하게 대중의 흐름을 보아가면서 필요성을 충족시키면서 서비스를 제공할 때 유용한 지질 정보 서비스로 살아남을 수 있을 것이다.

무심코 사용되고 있는 영어에 근원을 둔 외래어 표현들은 아름다운 우리말 표현 및 국어 표현으로 대체하는 것이 신뢰성과 정확성 및 전달력 향상에 기여할 것이다. 올바른 국문 사용의 장점은 크다. 작은 부분이라고 생각될지도 모르겠지만, 그 사회가 쓰는 말과 글이 그 사회를 지배하는 인식이 된다. 홈페이지에서 사용되는 용어 선정에 있어서, **전체적으로 국문법적 검토와 연구를 제안하는 바이다.**

3. '지오빅데이터 오픈플랫폼'에 접근성 향상 방안에 대한 제언

(1) 서론

유용한 정보를 어떻게 홍보하고 활용할지 단기 계획부터 중장기적 관점의 계획이 필요해 보인다. 현재 모니터링단을 통한, 사용자의 개선방안을 도출하는 것도 실사용자들의 좋은 홍보 방안 중에 하나로 생각되고 이러한 프로그램을 보다 다양한 계층이 참여할 수 있도록 여러 계층의 취향과 흥미에 맞게 만드려는 노력이 필요해 보인다.

그러기 위해서는 '지오빅데이터 오픈플랫폼'의 홍보를 전담하는 부서나 인원을 만들 필요가 있을 것이다. 더 나아가 '전문가 그룹의 빅데이터 관리 및 검증 시스템의 구축' 및 그 근거가 될 만한 '법적 근거'를 마련하는 것은 지속적으로 데이터를 관리할 수 있는 근거가 되면서, 동시에, 대

중적인 필요성 인식과 대중의 수요에 발맞추어서, 서비스를 제공하는 입장에서, 기여하는 바가 있을 것으로 본다.

(2) 본론

1) 구글(Google), 네이버(Naver), 다음(Daum)에서 국내외 지도 공유 서비스를 통한 접근성 향상 방안
현재, ‘지오빅데이터 오픈플랫폼’은 구글어스 기반의 정보를 바탕으로 구축한 것으로 보인다. 전세계적인 표준을 찾는 데는 구글어스 기반의 플랫폼이 유용하겠지만, 실질적으로 실사용자들이 국내 사용자라는 점을 감안한다면, 네이버 지도 서비스 및 카카오맵 서비스와 연계와 제휴를 검토하는 것도 사용자의 접근성 측면과 홍보 측면에서 유용할 것으로 보인다.

현재 주제별 지질도를 검색하면 지도 위에 GIS 기반으로 여러 지질정보가 표시된다. 하지만, 이용하면서, 그 심벌(Symbol) 및 색인(Index) 정보가 함께 표기되지 않아, 해당 심벌이 의미하는 바를 찾는 데 수고가 들어가는 어려움이 있었다. 표시되는 지질도 화면상 많이 사용되는 심벌 및 색인 정보를 표시해준다면 더 유용할 것으로 보인다.

‘지오빅데이터 오픈플랫폼’을 좀 더 체계화하기 위해서는 **법제화 및 독립적인 예산 마련 방안을 구축하기 위한 노력도 필요해 보인다**. 수많은 단발성 프로젝트들을 많은 국내외에서 실시하고 있는 상황에서 우수한 학술 데이터 기반 자료를 제공하면서 경쟁력을 확보하기 위해서는 법제화를 바탕으로 체계적인 **중장기 시스템 마련**이 필요해 보인다.

지질정보는 국내 지진정보와도 많은 연계성을 가지고 있다. 최근 몇 년 사이, 포항 및 경주 일대에서 발생한 지진으로, 지질학과 지진학에 대한 사회적·정치적 연구 필요성이 시의적(時宜的)으로 어느 때보다 강조되고 있는 시점으로 보인다. 그전까지 상대적으로 지진의 안전지대로 인식되던 한반도도 세부적으로 탐구하다보면 결코 그렇게 안전하다고 단정할 수 없음이 관련 연구를 통해, 드러나고 있고, 이에 관한 **보충적인 연구를 하는데, 지진과 더불어 지구물리 데이터와의 연계성 가지고, 일반지질에 대한 파악이 중요해 보인다**.

그와 같은 연구를 하기 위해서는 **원(原)자료(Raw data)에 대한 검증과 더불어, 그 접근성을 향상하는 것이 중요하다고 생각된다**. 지질자원연구원이라는 공적인 연구기관에서 그에 대한 명성과 신뢰성을 바탕으로, 양질의 정보를 제공하기 위해 **제도적 지원책**도 보완해 가면서, 사업을 계속 발전해 나갈 필요가 있어 보인다.

2) ‘지오빅데이터 오픈플랫폼’ 영문 양식 추가 보완 운영 제언

이미 국내외에서 많은 각 나라의 국책 연구기관을 통해, 지질정보의 축적이 이루어지고 있다. 대부분 영문 바탕으로, 자료를 검색 및 다운로드 가능하다. 지질정보의 축적 및 검색 공유에 있어서 가능한 일정 부분이라도, **사이트의 유용한 정보들은 영문화 사이트를 보완 운영한다면, 국외 연구자들도 자유롭게 자료에 대한 접근성이 향상될 것으로 본다**.

자료 등록 양식에 있어서도 영문 양식을 보완하여, 추가해서 제공한다면, 많은 국외 연구자들이 수집한 자료들을 공유할 수 있는 기본적인 조건은 충족된다고 본다. 또한 비슷한 목적으로 지질 자료를 수집하는 국외 지질정보 사이트에 대한 링크를 통해서, 유용한 정보의 상호 교류 및 협력을 추진하는 것이 필요하다고 본다. 또한 관련 시스템을 활성화하기 위한 연구도 필요하다고

본다.

(3) 소결론

이 플랫폼의 접근성 향상 방안으로, 국내적인 요인으로는 다양한 메인 정보 제공 검색 엔진들과 업무 제휴를 통한 접근성 향상을 제안했고, 국내적으로는 자료접근의 용이성을 위해서, 홈페이지의 영문화 개선과 더불어, 국제적인 오픈 빅데이터 수집 양식과의 통일성 및 호환성을 검토하는 연구를 제안했다. 이를 중장기적으로 지원할 수 있는 부서 및 인원, 법적인 지원 근거의 마련을 제도적으로 마련한다면 질 좋은 지질 정보들이 많이 생산되고, 공유되면서, 더욱 효과적인 지질 정보의 창구로서 ‘지오빅데이터 오픈플랫폼’이 이용될 수 있다고 보고, 정보의 접근성 측면에서 발전 방향을 제안한다.

4. ‘지오빅데이터 오픈플랫폼’에 신규 등록 Raw Data의 4단계 검수 절차에 대한 제언

(1) 서론

1) 데이터 신뢰도 향상을 위한 검증 절차 보완의 필요성

많은 지질학 관련된 데이터들을 ‘지오빅데이터 오픈 플랫폼’은 제공하고 있다. 현재는 일정 검수 절차를 걸쳐서 데이터가 등록할 수 있도록, 구성되어 있지만, 일괄적으로 올린 기존 자료에 대해서는 적잖은 오류가 관찰된다. ‘현미경 사진’에 ‘암편 사진’이 올라가 있는 경우가 모니터링 결과 월간 보고서에서 올린 2건(이미 보고된, ‘10월 보고서’와 본 보고인 ‘2월 보고서’) 이외에도 많았다. 기존 등록된 자료들에 대한 지속적인 수정 검증이 필요해 보인다.

(2) 본론

1) 빅데이터 산업 검증 절차의 도입을 통한 자료 신뢰성 향상

빅데이터 산업은 여러 국가에서 실시하고 있고, 그 자료의 활용을 위해서 기본적으로 6단계에서 8단계의 검증 절차를 거쳐서 유용하고 신뢰성 있는 정보로 공유될 수 있는 장치를 마련해 놓았다. 그러한 과정을 우리도 벤치마킹해서, 유용한 학술 및 대중 빅데이터화 정보로 만들어야 필요성이 있다.

그 과정은 크게, 사전 조사, 자료 수집, 자료 검증, 자료 분석, 자료에 대한 심층적 용역 연구 등으로 나누어서 체계적이고, 면밀하게 검토하는 시스템을 만들 필요가 있다.

2) 1차 검수자, 2차 검수자, 전문가 검수자, 총괄 검수자 및 총괄 검수자의 제안

등록된 자료의 신뢰성을 향상하는 방안은 다양하고 지속적인 사업을 통한 효율성을 고려하면서도, 자료의 복수 검토이다. 단순 오류 사항에 대한 수정은 1차 검수자, 2차 검수자를 통해 이루어지도록 고안하고, 심층적인 검토는, 전문가 검수자, 총괄 검수자 및 총괄 검수자의 제안으로 역할을 나누어, 4단계의 시스템을 거쳐서, 자료에 대한 생산 및 검증 절차를 지속해 나가는 것이 자료의 신뢰성과 효율성을 높이는 방법이 될 수 있을 것이라 본다.

그러기 위해서 매뉴얼화 된 자료 업로드 양식을 규격화해서 통일하여 목적성을 가지고, 빅데이터화 할 구상을 가지면서 자료를 검증하면서 생산하게 해야 한다. 그렇게 신뢰성이 보장된

자료들을 모을 때 유용한 역할을 할 수 있으리라 본다.

이 목적을 달성하기 위해서는 전체적인 구상을 짜는 전문가 및 총괄 검수자가 개선 방안 및 지질 자료의 목적성을 정확히 지정하고, 그에 맞는 데이터를 증장기적으로 데이터를 모으려는 노력이 필요하다고 본다.

(3) 소결론

신뢰성 향상의 측면에서는 일반 검수자부터 지질 및 지구물리 전문가 검수자와 데이터 전문가 까지 투입되어, 신뢰할만한 자료를 양산하는 것이 무엇보다 중요하다. ‘앞으로는 ‘D-N-A(Data, Network, AI)가 주도하는 사회가 될 것이고, 데이터의 중요성을 강조할 때 가장 흔히 비유되는 말로 ‘Garbage In, Garbage Out(GIGO)’이라는 말이 널리 쓰이고 있다. ‘쓰레기가 들어가면 쓰레기만 나온다’는 뜻이다.’⁴ 거꾸로 생각하면, ‘좋은 것을 넣으면 좋은 것이 나온다’는 말도 될 것이다. 대중이 생산하면서도 데이터의 질을 보장할 수 있도록 그 방법을 찾아야만 할 것이다.

지질자원연구원에서 지속적으로 좋은 원(原)자료를 생산하는 노력을 할 필요성이 있어 보인다. 그러기 위해서는 이용자 및 자료 생산자에게 적절한 효용성을 주면서, 유용한 플랫폼으로 인식되고 활용될 때 가능할 것이다. 양질의 정보를 생산하는 사람들에게 적절한 이점을 주는 것이 필요할 것이다. 가장 큰 이점은 신뢰할 수 있는 정보를 찾을 수 있고, 장기적으로 그 정보에 접근가능하다는 것이 되어야 한다고 본다. 그러기 위해서는 시스템을 운영하는 전문가들의 시스템을 어떤 체계로 운영할지에 대한 빅데이터 전문가들의 조언을 참고할 때, ‘지오빅데이터 오픈플랫폼’이 지속적으로 유용한 지질 자료를 제공하는 수단으로 자리매김할 수 있을 것이다.

이미 다양한 영역에서 빅데이터화 사업이 많이 이루어졌고 지금도 계속되고 있다. 그 사업들 중에는 성공한 것도 있고, 실패한 것들도 있다. 대표적인 빅데이터 성공 모델로 ‘위키피디아’를 언급한다. ‘위키피디아’는 다수의 사용자가 검증하면서, 자료들을 보완하면서 공유하는 시스템을 구축했다.⁵ 그것을 통해 신뢰할 수 있는 정보가 무엇인지 시스템적으로 선별할 수 있는 빅데이터 모델링을 했고, 그것이 지난 10여년 간에는 성공 모델로 인식되었지만, 앞으로 변화하는 환경 속에서 어떤 것이 지질자료를 모으고 활용하는 것이 적당한 것인지는 계속 연구하고 투자를 지속해야 한다. 과연 유용하고 신뢰할 만한 지질정보를 수집하고 생산하는데 적절한 시스템이 무엇일지 늘 고민하면서, 대중의 참여를 이끌어야만 하고, 여러 사업 모델을 시도하면서 지속적으로 운영해야 한다고 본다.

II. 오픈플랫폼 오류 개선

1. 기존에 등록된 데이터 오류 보고

(1) 서론

많은 지질 관련된 데이터들을 ‘지오빅데이터 오픈 플랫폼’은 제공하고 있다. 현재는 소정(所定)의 검수 절차를 거쳐서 데이터가 등록할 수 있도록 보완되어 있지만, 일괄적으로 올린 기존 자료에 대해서는 적잖은 오류를 찾을 수 있다. 모니터링을 하면서 발견된 오류들을 보고하면서, 이미 등록된 자료에 대한 보완 및 수정 절차를 할 필요성과 방향성에 관해서 논하겠다. 오류에 대한 공통점을 파악하면서 비슷한 오류를 막을 방안도 검토해 보고자 한다.

(2) 본론

1) 기존 데이터 오류 내용의 보고



그림 2 사례 ㄱ. 암편사진을 현미경 사진으로 잘못 기재

내용정보: '현미경사진_I-030.jpg (2.00 Mb)' -> '암석사진 I-030(2.00Mb)' 수정 필요.

데이터명: 화성암(안산암(Andesite))_I-030)

Url: <https://data.kigam.re.kr/data/6d7b376b-6354-47ea-9284-2f6ab389afb5>

오류 내용: 암편사진을 현미경 사진으로 잘못 기재



그림 3 사례 ㄴ. 암편사진을 현미경 사진으로 잘못 기재

내용정보: 데이터 오류

데이터명: '현미경사진_M-033.jpg (2.00 Mb)' -> '암석사진 M-033(2.00Mb)' 수정 필요.

Url: <https://data.kigam.re.kr/data/7cbd8084-6dbd-45ea-8ebd-361158607d3d>

오류 내용: 암편사진을 현미경 사진으로 잘못 기재



그림 4 사례 ㄷ. 등록 자료의 제목 표시에 있어서 단순 오타자

내용정보: 데이터 오류

데이터명: [퇴적암] 석탄기 홍점통 분광측정자료

Url: <https://data.kigam.re.kr/search?type=bdfb096f-5bae-4d67-9c3e-a135d9166ee8&page=1>

오류 내용: 빨간색 박스 부분, '석탕기'-'>'석탄기'로 수정 요망.

(3) 소결론

처음부터 완벽한 자료는 없을 것이다. 지속적인 모니터링은 소소한 오류들을 수집할 수 있었다. 여러 오류를 최소화할 수 있는 장치들을 만들 필요성과 기존 자료들을 지속적으로 검토할 수 있는 지질자원연구원의 노력이 필요하다고 본다. 외부적으로 여러 프로그램을 통해서, 많은 사람들의 참여를 독려하면서, 수정과 보안을 거치면서, 점차 데이터의 완성도와 신뢰도를 높여 갈 수 있으리라 본다. 그러면서, 동시에 대중 및 전문가의 필요성 및 수요도 증가할 것으로 보인다.

이러한 선순환이 이루어지기 위해서는 지속적인 데이터의 수정 및 보완 작업이 중요하다. 그럼으로써 양질의 데이터의 효용성의 보장될 때, 많은 사람들이 계속 데이터를 수정하고 보완하려는 노력을 지속할 것이다.

데이터의 신뢰성, 데이터의 정확성을 향상시키기 위해서, 단순한 모니터링부터 시작해서, 데이터베이스 이론에 접목해서, 방대한 학술데이터를 사용자가 사용하기 적합한 데이터로 가공하려는 노력이 지속적으로 필요할 것이다. 그러기 위해서 단순한 모니터링을 통한 데이터의 오류 보고도 무시할 수 없는 중요한 부분이 될 것이고, 그러한 프로그램들을 지속적으로 실시하면서, 자발적인 참여자들의 이점을 주었을 때, 지속가능하게 시스템이 운영될 것으로 본다. 시스템 기여 정도에 대한 정량적이고 정성적인 평가와 인정이 될 수 있는 방안을 계속 논의하고 검토해야 할 것으로 보인다.

㉔. 나오는 말

크게 5개의 마당글로, 1) 위치정보 기반 통합적 암석 및 지질 데이터 구축, 2) 용어 순화적 관점에서 바라본 ‘지오빅데이터 오픈플랫폼’, 3) ‘지오빅데이터 오픈플랫폼’에 접근성 향상 방안, 4) ‘지오빅데이터 오픈플랫폼’에 신규 등록 원(原)자료(Raw Data)의 4단계 검수 절차에 대한 제언 및 검토와 함께, 5) ‘기존 데이터의 오류를 보고’하고 개선방안에 대해서 검토해 보았다.

이와 같은 검토는 궁극적으로 데이터의 신뢰도, 정확도, 지속가능한 확장성을 지향하고 있다. 제안한 5가지 개선방안 중에 바로 도입 가능한 것부터 추가적인 검토와 고려를 통해서, 도입 가능한 부분까지, 지질자원연구원에서 적극적으로 반영하려는 노력이 있을 때, 시스템을 보다 안정화하고 유용한 플랫폼으로 자리매김하는 데 실질적인 도움이 될 것이다. 그렇게 완성된 시스템이 완전해지면 완전해질수록 유용한 정보가 되고, 많은 사람들이 이용하는 선순환이 이루어지면서 보다 유용한 플랫폼으로 거듭날 수 있을 것이다.

그러기 위해서, 무엇보다 중요한 것은, 이 플랫폼의 목적성이 무엇인지 현실적인 면을 고려하면서 신중히 검토하고 중장기적인 계획(5~10년 단위) 하에서 시스템을 구축하려는 노력이 필요할 것이고, 많은 인력과 예산이 효율을 발휘하기 위해서, 어떻게 빅데이터 개선을 위한 다양한 전문가들의 노력을 모을지에 대한 다양한 계층의 견해와 반응을 모으려는 노력도 함께 이루어져야 할 것이다.

‘지오빅데이터 오픈플랫폼’이 신뢰성 있고, 정확한 정보로서 자리매김한다면, 그 확장성은 지질 분야에 국한되지 않고 다양한 과학 연구의 기초자료로도 충분히 그 가능성이 있을 것이다. 제안한 다섯가지 보고를 바탕으로, ‘지오빅데이터 오픈플랫폼’에 대한 보다 과감한 투자와 시스템 구축을 위한 심도있는, 다양한 연구와 사업을 지속하기를 제안한다.

<참고문헌>

1. 네이버 지식백과, 시사상식사전, ‘알과고-이세돌 대국’, 참조.
(2022년 3월 20일 검색)
(<https://terms.naver.com/entry.naver?docId=3408537&cid=43667&categoryId=43667>)
2. 송철의, 1998, ‘외래어의 순화 방안과 수용 대책’ 『새국어생활 제8권 제2호(‘98년 여름)』, 37쪽 인용.
3. 송철의, 1998, ‘외래어의 순화 방안과 수용 대책’ 『새국어생활 제8권 제2호(‘98년 여름)』, 26쪽 인용.
4. 김복철, 2022, ‘[기고] 지오 빅데이터로 사회문제 해결한다’ 기사, 디지털타임즈, 인용.
(입력: 2021-01-18 19:18)(2022년 3월 19일 검색)
(http://www.dt.co.kr/contents.html?article_no=2021012702102369660001)
5. 바라트 아난드, 콘텐츠의 미래, 2017, 리더스북, 참조.

<끝>