

기후환경 변화예측 위한 데이터 표준화에 관한 연구¹⁾

Study on Data Standardization for Predicting Climate and Environment Change

김무준* · 김계현 · 남기범 · 김나영

Mu Jun Kim* · Kye Hyun Kim · Gi Beom Nam · Na Young Kim

인하대학교 지리정보공학과

{nhageo8 · gib2om}@naver.com · kye Hyun@inha.ac.kr · 0213olive@hanmail.net

요 약

전 세계적인 지구 온난화 현상으로 해수면 상승과 생태계 변화가 발생하여 기후변화에 대한 사회적관심이 증가하고 있다. 이와 더불어 기후변화와 지구환경시스템의 대기, 수권, 생물권, 지표면 등 각 권역간의 상호작용과 피드백을 고려한 연구가 증가하고 있는 실정이다. 기후와 환경을 통합적으로 분석하여 기후변화에 따른 지구환경시스템의 변화특성을 이해하고 이러한 피드백 과정을 파악하기 위해서는 분석 자료의 원활한 공유와 연계를 위한 통합 데이터베이스 구축이 필요하다. 이를 위해서는 먼저 다양한 기후/환경 연구 분야의 자료를 관리하기 위한 데이터의 의미, 명칭, 정의 등에 대한 원칙의 수립이 요구된다. 따라서 본 연구에서는 기후/환경 변화예측 연구 자료의 원활한 공유와 관리를 위한 데이터 표준화 연구를 수행하였다. 기후/환경 변화예측 연구 분야의 자료 현황을 조사 및 분석하였고 그에 따른 자료 관리 방안을 마련하였다. 그 결과 관리할 오브젝트를 기준으로 기후/환경 연구 분야의 데이터 표준화를 수행하였고 표준단어, 표준도메인, 표준용어를 정의하였다. 데이터 표준화 결과는 기후/환경 변화예측 자료를 관리하고 공유하는데 있어 데이터의 의미를 효율적으로 파악하고, 데이터베이스 설계과정에서 데이터의 품질과 생산성을 향상 시킬 수 있다. 향후 연구에서는 데이터베이스 개념적 엔티티의 속성설계 단계부터 데이터 표준을 적용한 통합 데이터베이스 구축이 필요하다.

1. 서론

오늘날 기후변화로 인하여 자연생태계가 파괴되고, 해빙과 해수면 상승 등의 환경변화가 발생하고 있다. 이에 2009년 코펜하겐에서는 전 세계적으로 진행되고 있는 이상기온의 문제를 해결하기 위한 총회가 개최되고, 우리나라에서도 각각의 기관마다 기후변화와 관련한 연구를 진행하기 위한 센터의 설립 등 다양한 활동이 진행 중에 있다. 특히, 기후변화는 지구환경 시스템을 구성하는 모든 권역에 영향

을 끼치며, 또한 지구환경시스템을 구성하는 대기권, 수권, 생물권, 육지표면 등은 다시 기후변화에 영향을 끼친다. 따라서 기후변화가 환경에 미치는 영향을 파악하고 환경 변화를 종합적으로 예측을 하는 연구가 요구되고 있다. 하지만 기후변화에 따른 환경 변화의 영향을 예측하기 위한 자료는 매우 방대하고 다양하게 존재하기 때문에 통합적 연구를 수행하는데 어려움이 존재한다. 따라서 기후변화에 따른 지구환경시스템의 변화특성을 이해하고, 이러한 피드백 과정을 파악하기

1) 이 논문은 교육과학기술부의 재원으로 한국연구재단(No. 2010-0001906)과 국토해양부의 공간정보 전문인력 양성사업의 지원을 받아 수행된 연구임

위해서는 분석 자료의 원활한 공유와 연계를 위한 방안이 요구된다. 이를 위해 각 권역에서 사용되는 자료의 의미를 명확히 파악하고 원활히 공유할 수 있는 데이터 관리 체계의 마련이 필수적이다.

본 연구에서는 기후와 환경의 변화예측 및 상호작용을 지원하기 위해 기후/환경 변화예측 연구 분야의 데이터 표준화에 대한 연구를 수행하였다.

2. 연구내용

2.1 기후/환경 분야의 자료 현황 파악

기후/환경 변화예측 분석을 위해 기후, 대기, 생물계절, 탄소유출, 수 환경, LSPM, RS/GIS분야를 선정하였다. 각각의 분야는 기후변화가 지구시스템의 각 권역에 미치는 영향과 피드백 과정을 예측하기 위하여 선정되었다. 기후와 대기 분야는 모델을 사용하여 시나리오의 종류에 따라 자료를 생산한다. 특히 기후 자료는 타 분야에 기초자료로 사용된다. 생물계절과 탄소 유출 분야는 기후 변화에 따른 생물종의 변화 양상과 탄소 순환 시스템에 대한 자료를 생성한다. 또한 수 환경 분야는 기후변화가 유역과 호소에 변화에 미

치는 영향에 대한 자료를 생성한다. LSPM(Land Surface Process Model)의 경우에는 육지표면 과정 모델을 사용하여 다른 분야에 기초 자료를 제공한다. RS분야는 MODIS영상을 사용하여 지표면의 온도 값을 관측하고 다른 분야에 입력 자료를 제공한다. 마지막으로 GIS분야는 분야별 공간자료를 구축하고 공간분석을 수행하며 데이터를 시각적으로 표현한다. 종합적으로 살펴보면 기후의 모델 시나리오 결과자료는 타 분야의 적용되고, 타 분야에서 생성된 데이터는 다시 기후 모델의 입력 자료로 사용된다. 이러한 과정으로 기후와 지구시스템 각 권역의 상호작용을 파악할 수 있게 하고, 미래 기후 및 지구환경시스템을 예측할 수 있다. 하지만 각 분야에서 서로 다른 컴퓨터 운영체제와 소프트웨어를 사용함에 따라 다양한 형태의 자료가 생산될 뿐만 아니라, 관리방법이 상이하게 존재하여 원활한 연구의 진행에 어려움이 발생한다. 따라서 자료의 공유 및 관리방안을 마련하고 각 분야에 생산된 자료의 의미를 명확히 파악하고 이를 구별할 수 있는 데이터 표준화를 수립하는 것이 필요하다.

표 1. 기후/환경 변화예측을 위한 데이터 현황

분야	용도	축척/해상도	범위	파일포맷	좌표계
기후	미래기후 시나리오	140km	전 세계	NC	WGS84
대기	대기화학 변화 분석	-	한반도 및 동아시아	NC	WGS84
생물계절	생물종의 변화 양상 파악	-	전국 74개 지점	XLS	-
탄소유출	탄소 순환의 시스템	-	강원도 양구군	XLS	-
수환경	유량 및 물질 순환 분석	-	용담댐	TXT	Bessel
LSPM	LSPM 모델 수행	1km	한반도	OUT	WGS84
RS	LSPM입력 자료 생성	5.6km	전 세계	HDF	WGS84
GIS	공간자료 생성	다양함	전 세계	SHP	WGS84

2.2 자료 관리 방안

기후/환경 각각의 분야마다 다양한 포맷의 자료가 파일형태로 존재하고 있다. 각 분야에서 분석을 위해 취득한 자료 및 모델 결과로 나온 자료는 바이너리 코드 의 파일 형태로 되어 있다. 일부 자료는 4차원의 복잡한 구조로 구성되어 있어 자료의 형태를 분석하여 데이터베이스 내 테이블 형태로 관리하기가 쉽지 않다. 또한 기후/환경 분야는 사용되는 데이터가 많고 대부분 대용량의 데이터이므로, 파일을 기반으로 연구를 진행하고 있는 경우가 많다. 따라서 파일 단위의 자료 관리가 요구되며 각 분야의 파일을 관리하기 위한 데이터 정보 요소에 대한 원칙을 수립하여, 원활한 자료의 공유와 관리가 이루어 질 수 있도록 표준화를 진행하였다.

2.3 데이터 표준화

데이터 표준화는 통합 데이터베이스 구축에 필요한 데이터 요소의 명칭, 정의, 형식, 규칙에 대하여 원칙을 수립하여, 이를 전사적으로 적용하는 것을 의미한다. 데이터 표준화는 데이터 모델 및 데이터 베이스에서 정의할 수 있는 오브젝트를 대상으로 수행하는 것이 이상적이나 주로 관리해야 할 필요성이 있는 오브젝트만 대상으로 데이터 표준화를 진행하는 것이 효율적이다(한국데이터베이스진흥센터, 2006). 따라서 본 연구에서는 앞서 도출한 자료 관리방안에 따라 파일정보를 효과적으로 저장하고 관리하기 위한 데이터 표준화를 진행하였다. 데이터 표준화는 데이터 활용상의 문제점을 최소화 할 수 있으며, 데이터의 정확한 의미를 파악하고, 데이터에 대한 상반된 시각을 조정하는 역할을 수행할 수 있다.

기후/환경 변화 연구에 사용되는 용어에 대해 표준 단어를 구성함으로써 용어에 대한 한글명과 영문명을 일관되게 정의할 수 있도록 하였다. 표준단어는 표준

용어를 구성하는 최소한의 의미를 가지는 단어로, 이는 표준용어의 영문명 작성 기준이 된다. 또한 동일한 성질의 컬럼에 대해 그룹핑한 표준 도메인을 설정하여 데이터 타입 및 데이터 길이를 일관되게 관리할 수 있도록 하였으며, 공통적인 데이터 검증 규칙의 적용이 가능하도록 하였다.

3. 기후/환경 데이터 표준화 결과

기후/환경 변화예측 각 분야의 연구 방법 및 내용과 수집한 자료의 분석을 통해서 데이터 표준화를 수행하였다. 기후/환경 분야의 파일을 관리하고 공유에 사용될 모든 단어를 추출하여 표준 단어를 정의하였다. 표준단어는 정확한 용도, 출처, 사용여부를 판단 할 수 있는 의미 있는 정보로서 정의되었다. 표 2와 같이 기후 변화 환경에서 데이터 관리를 위한 일정한 의미와 뜻을 가질 수 있는 최소단위로 구성하였다. 표준 단어를 구성함으로써 동일한 개념을 의미하는 표준용어의 생성을 방지하도록 하였다.

표 2. 기후/환경 분야 표준 단어(예)

	표준단어	영문명	영문 약어명
1	기후	Climate	CLI
2	인덱스	Index	IDX
3	순번	Number	NUM
4	날짜	Date	DATE
5	시나리오	Scenario	SCEN
6	해상도	Resolution	RESOL
7	수직차원	Vertical	VERT
8	시간간격	Time Interval	INTER
9	파일	File	F
10	명	Name	NM
11	목록	List	LIS
12	코드	Code	CD
..

예를 들어 “기후” 라는 표준단어를 선별하고 그에 맞는 영문명을 작성하였고, 영문 약어명을 작성하여, 약어명을 통해서도 데이터의 의미를 쉽게 파악할 수 있도록 하였다. 이와 더불어 데이터를 표현하기 위한 컬럼의 용도, 사용빈도와 데이터의 물리적인 특성을 고려하여 도메인을 분류하고, 도메인별 타입을 부여하였다. 표준 도메인은 문자형, 숫자형, 코드성 도메인으로 구성되어 정의하였다(표 3). 정의한 표준 단어와 도메인을 기준으로 기후/환경 변화의 통합 DB를 구축하기 위한 표준용어 사전을 구성하였다. 표 4는 기후 분야의 표준용어 사전의 작성결과를 보여준다. 각각의 표준용어는 유일하게 작성되었으며, 사용자들이 데이터의 의미를 명확히 확인할 수 있도록 정의하였다. 표준용어의 영문명은 표준단어의 영문 약어를 이용하여 조합하였다. 예를 들어 기후분야의 파일명은 표준단어의 기후(CLI)+파일(F)+명(NM)의 조합으로 파일명

(CLI_F_NM)을 정의하였다. 이러한 방법으로 기후/환경 변화예측 연구를 수행하는 각 분야에 대해서 데이터 표준화를 수행하였다.

표 3. 표준 도메인

	표준 도메인명	데이터 타입	길이
1	순번	Number	5
2	코드	Text	6
3	해상도	Text	10
4	파일	Text	50
5	년월일	Text	10
6	년	Text	4
7	간격	Text	5
8	유무	Yes/No	1
9	대상지역	Text	20

표 4. 데이터 표준용어(기후분야 예시)

	표준용어명	영문약어명	유형	용어정의	표준 도메인	데이터타입
1	기후 테이블	TB_IDX_CLI	테이블	기후 데이터 파일에 대한 정보를 담고 있는 테이블	-	-
2	순번	CLI_NUM	컬럼	새로운 파일의 자동 증분으로 저장되는 번호	순번	Number(5)
3	날짜	CLI_DATE	컬럼	해당 데이터의 모의 날짜	년월일	Text(10)
4	시나리오	CLI_SCEN	컬럼	기후 분야에서 지정한 경제 성장 시나리오	-	Text(5)
5	해상도	CLI_RESOL	컬럼	해당 파일에서 적용하는 타일(격자)의 크기	해상도	Text(10)
6	수직차원	CLI_VERT	컬럼	대상 공간에 대한 수직 모의 단위	-	Text(5)
7	시간간격	CLI_INTER	컬럼	모의를 통해 생산되는 데이터의 저장 간격	간격	Text(5)
8	파일명	CLI_F_NM	컬럼	해당 파일의 이름	파일	Text(50)
9	데이터목록 코드	CLI_LIS_CD	컬럼	해당파일 내에 데이터의 종류를 나타내는 코드	코드	Text(6)

4. 결론

본 연구에서는 통합 데이터베이스 구축 과정에서 데이터의 의미파악과 설계시의 생산성을 향상시킬 수 있도록 하기 위해 기후환경 변화 분야의 데이터 표준화 연구를 수행하였다. 이를 위하여 자료 현황 파악 및 특성을 분석하여 도출된 자료 관리 방안에 따라 관리될 필요성이 있는 오브젝트를 대상으로 표준화가 수행되었다. 아울러 파일단위의 자료 관리를 위한 데이터 정보 요소에 대한 명칭, 정의, 형식, 규칙을 수립하여 사용자들에게 데이터의 의미를 파악하고 원활한 의사소통이 가능하게 하였다. 이러한 표준화 결과로 기후/환경 변환예측 분야의 표준단어, 표준도메인, 표준용어를 정의하였다.

본 연구에서 수립된 데이터 표준화의 결과는 연구 자료의 원활한 공유와 다양한 정보 제공이 가능하여 기후/환경 분야의 다양한 연계 연구를 지원 가능하다. 또한 연구에서 생산되는 다양한 결과를 효율적으로 관리할 수 있으며, 데이터의 명칭과 의미를 통일화하여 기후와 환경변화의 연구를 최적으로 지원할 것으로 사료된다. 향후 연구에서는 데이터베이스 설계과정에서 개념적 엔티티의 속성 설계 단계부터 데이터 표준을 적용한 통합 데이터베이스 구축을 진행하는 과정이 필요하다.

참고문헌

- [1] 한국데이터베이스진흥센터,
“데이터베이스백서”, 2006.06
- [2] 한국데이터베이스진흥센터,
“데이터아키텍처 전문가 가이드”,
2006.01
- [3] 배창호, “국방데이터베이스 통합을
위한 데이터표준화 방안”,
충실대학교, 2000.12
- [4] 이석호, “데이터베이스 시스템”,
정익사, 2007
- [5] 행정안전부, 행정데이터관리시스템,