

보도 일시	2022. 10. 31.(월) 12:00	배포 일시	2022. 10. 31.(월) 09:00
담당 부서	기상서비스진흥국 국가기후데이터센터	책임자	과 장 김진석 (042-481-7470)
		담당자	사무관 황동익 (042-481-7475)

'기상기후 거대자료(빅데이터)', 사회적 융합·활용 쉬워진다 - 기상청, 내년까지 「기상기후데이터허브」 구축 -

- 100년 전 관측 자료부터 100년 후 미래 전망 자료까지 빠짐없이 모은다.
 - 한국에서 원하는 자료를 막힘없이 흐르게 한다.
 - 단순 업무에서 인공지능 분석까지 어려움 없이 사용한다.
-
- 기상청(청장 유희동)은 올여름 집중호우, 태풍 힌남노 등 기후 위기가 눈앞의 현실로 다가온 시점에서 기상기후 거대자료(빅데이터)를 재난 방지 및 4차산업 분야에서 편리하고 자유롭게 분석·활용할 수 있는 「기상기후데이터허브」 구축을 시작한다고 밝혔다.

【참고】 기상기후데이터 활용 사례

- ▲ (재난) 5년간의 기상관측자료와 피해 이력 정보를 분석으로 침수 피해 예측 체계 개발
- ▲ (에너지) 기상데이터 기반 신재생에너지 발전량 예측 고도화
- ▲ (농업) 기상데이터 기반 농업재해 예측 및 온실 환경 최적화 서비스 개발
- ▲ (4차산업) 드론 및 자율주행차 운영 체계에 기상정보를 활용하여 운행 안정성 확보

- 「기상기후데이터허브」는 ▶ 기상청과 관계기관이 축적해 온 방대한 (수백 페타바이트급) 전 세계의 기상기후데이터가 ▶ 재난 현장과 4차산업 융합(인공지능, 거대자료(빅데이터)) 등에 널리 활용되도록 데이터를 ▶ 수요자 맞춤형 에이피아이*(API)를 기반으로 통합 연계·제공함으로써 국가 및 사회의 ▶ 기상기후 의사결정을 지원하여 부가가치를 창출할 수 있는 공개 운영 체계로 구축된다.

* API(Application Program Interface)란 공개된 데이터 운영 체계를 활용해 외부에서 애플리케이션(앱)이나 서비스를 만들 수 있는 공개 응용 프로그램이다. 공개된 API를 활용해 누구나 앱과 서비스를 개발할 수 있다. API는 날씨나 교통 정보 등 갱신이 빈번한 대용량의 데이터를 서비스에 연계할 때 유용하다.

- 기상청은 기상기술 발전에 따라 최근 기상기후데이터 용량이 기하급수적으로 증가하여 수십 페타바이트를 상회하고 있으며 이를 국가와 사회의 기상기후데이터 활용 수요를 충족시키는데 기여하기 위해 「기상기후데이터허브」 구축을 추진한다고 설명했다.
- ※ 기상기후데이터 용량은 수치모형, 위성, 레이더의 공간 해상도가 향상되면서 거대자료화되어 가고 있으며, 최근 5년간 수치모형에 의해 생산된 데이터는 5년 전보다 10배, 기상위성은 70배, 기상레이더는 5배 증가되었다.
- 기상청은 「기상기후데이터허브」 구축을 데이터 통합관리체계부터 화재 등 시스템 재해에 대비한 보관체계 구축까지 4단계로 나누고 2025년까지 완료할 계획이지만, 국민이 실질적 혜택을 받는 시점은 1단계 구축이 완료되는 2023년 하반기라고 밝혔다.
- 2023년부터 국민은 맞춤형 에이피아이(API) 방식으로 기상기후데이터를 선별하여 실시간으로 서비스 받을 수 있고 사회 각 부문의 정부 기관, 공공기관, 산업계, 학계 등은 데이터 활용에 필요한 대규모 시스템 구축이나 데이터 처리에 필요한 기초 지식만으로도 업무를 처리할 수 있을 것으로 기대된다.

< 100년 전 관측 자료부터 100년 후 미래 전망 자료까지 빠짐없이 모은다 >

- 호우, 폭염, 한파 등 국민의 93%가 체감하는 기후변화에 대하여 국가 및 사회의 대응이 본격화됨에 따라, 기상청은 방대한 기후변화 시나리오 데이터와 기상청의 관측 및 예보자료, 그리고 관계기관과 전 세계의 다양한 기상자료를 한곳에서 제공하는 서비스(One-Point 서비스)를 추진한다.
- 기상청은 기후변화 감시 자료와 기후변화 시나리오를 기상자료개방포털 (<http://data.kma.go.kr>)에서 제공하기 위해 실시간 데이터베이스 동기화 등 대내외 데이터 수집·처리 체계를 정비하고, 대용량의 기후변화 시나리오에 대해서 수요자가 원하는 영역(행정구역 등), 요소(기온, 강수량 등) 및 기간을 선택할 수 있도록 기능을 구현한다.

< 한곳에서 원하는 자료를 막힘없이 흐르게 한다 >

- 기상청은 기상기후데이터 활용에 대한 높은 수요에 부응하여 수요자가 필요로 하는 데이터를 한층 편리하게 수집하고 원활히 가공·처리할 수 있도록 데이터 공유·활용체계를 맞춤형 에이피아이(API) 방식으로 확대 전환해 나갈 것이라고 설명했다.

[참고] 기상기후데이터 이용 건수 현황

- ▲ (내려받기) 2016년 1월부터 2022년 9월까지 내려받기 건수 **4천2백만 건 이상**
- ▲ (API) 개방형 API 호출 건수는 2016년 1월부터 2022년 9월까지 **209억 건**, 특히 기상청 단기예보 데이터는 공공데이터포털의 개방형 API 중 활용신청 순위 2위 기록
- ▲ (대용량 데이터) 수치모형, 위성, 레이더 등 대용량 격자데이터에 대한 내려받기 건수는 2020년 2백5십만 건에서 2022년 9월 말 기준 1천1백만 건으로 대폭 상승
- ▲ (활용) 2022년(1~9월) 기준 분야별 기상기후데이터 활용률은 '학술·연구'가 57.1%로 1위, 이어 '토목·건축', '교육·행정', '농업' 순

- 기존에는 관측, 예보(동네예보, 특보 등) 및 생활기상지수 등만 에이피아이(API) 방식을 제공하였으나, 앞으로 데이터 형식 변환 기능까지 적용된 에이피아이(API)를 개발하여 수치모형, 위성, 레이더 등 대용량 격자데이터*까지 점차 확대해 나간다는 방침이다.

* 격자데이터란 특정 영역(한반도 등)을 위도, 경도 기반으로 바둑판 눈금 모양의 일정한 간격으로 나누어 각 교차점에 데이터를 연계시킨 자료를 말한다.

- 또한 에이피아이(API) 방식은 데이터 중 수요자가 원하는 변수 또는 원하는 지역에 대한 데이터만을 실시간으로 받을 수 있는 맞춤형 형식으로 구성될 예정이다.
- 기상청은 이를 통해 수요자들이 별도의 데이터 저장·관리 체계를 갖추지 않고도 앱 등을 편리하게 개발하고, 이에 따라 데이터 처리 시간도 획기적으로 단축되어 결과적으로 데이터 활용이 더욱 활성화될 것으로 기대하고 있다.

< 단순 업무에서 인공지능 분석까지 어려움 없이 사용한다 >

- 이와 함께, 기상청은 민간의 기상기후데이터 분석 및 활용을 유도하기 위해 인터넷 기반 자원 공유 데이터 분석·활용 체계를 구축하고 시범 운영할 계획이다.
- 데이터 분석·활용 체계는 기상청의 외부 사용자용 데이터 분석 운영 체계로 인터넷 기반 자원 공유 기술을 적용하여 사용자에게 전산 자원, 기상기후데이터, 데이터 분석에 필요한 인공지능 소프트웨어 및 시각화 분석 기능을 일괄 제공한다.
- 이에 따라, 민간 사용자는 기상청 내부에 있는 다양한 기상기후데이터를 보다 쉽게 검색·저장하고, 다양하고 완성도 높은 데이터 분석을 수행하고 시각화할 수 있을 것으로 기대된다.
- 유희동 기상청장은 “기상기후데이터가 국민의 일상생활과 직결될 뿐만 아니라 방재, 농업, 교통 등 사회 각 분야에서 자주 찾는 공통 기초자료로서 그 중요성이 한층 강조되고 있습니다.” 라며,
 - “민간을 중심으로 기상기후데이터 활용이 활발해지고 있는 것은 고무적인 현상으로 앞으로 민간에서 데이터를 더욱 편리하고 효과적으로 활용할 수 있도록 체계를 다져 나가겠습니다.” 라고 밝혔다.
- 붙임 1. 기상기후데이터 개방 서비스 현황
 - 2. 공공데이터포털 활용신청 순위
 - 3. 기상기후데이터허브 구축 추진 방향
 - 4. 기상기후데이터 특성

□ 기상자료개방포털(data.kma.go.kr) 서비스(2015.8. 오픈)

○ 기상청 데이터 서비스 종류(누계)

구 분	'16년	'17년	'18년	'19년	'20년	'21년	'22년(1~9월)
데이터종류(누계)	75종	82종	99종	124종	130종	136종	140종

○ 회원수

구 분	'16년	'17년	'18년	'19년	'20년	'21년	'22년(1~9월)
회원수 (누적가입수)	6,211명	16,475명	27,246명	37,842명	57,820명	80,031명	96,159명

○ 연도별 오픈 API 호출 건수(백만건)

구 분	'16년	'17년	'18년	'19년	'20년	'21년	'22년(1~9월)
호출건수	937	1,557	2,691	2,818	3,918	6,011	2,992

○ 2022년도 기상청 데이터 활용분야 비율(%)

연도	사용분야							
	학술연구	토목건축	농업	환경정화	교육행정	전기통신	보건·의료	기타
'22년(1~9월)	57.1	15.7	7.0	3.5	9.1	2.6	1.2	3.8

○ 연도별 데이터 내려받기 건수

구 분	'16년	'17년	'18년	'19년	'20년	'21년	'22년(1~9월)
일반	367,769	561,480	1,461,461	2,456,563	5,040,368	2,809,805	1,812,348
대용량	355,317	961,874	1,579,096	2,443,739	2,590,162	6,103,223	11,234,421
기후 통계분석	-	-	663,757	372,360	434,901	504,343	409,372
합계	732,086	1,523,354	3,704,314	5,272,662	8,065,431	9,417,371	13,456,141

2022. 01. 01 ~ 2022. 09. 30 기준
 '22년 누적 공공 데이터
 활용신청 TOP 20



파일 데이터 TOP 20

순위	제공기관	파일데이터 목록명	확장자
1	한국고전번역원	한국고전번역원_한국문집총간_월고집(月叢集)	xml
2	소상공인시장진흥공단	소상공인시장진흥공단_상기(상권)정보	csv
3	한국고전번역원	한국고전번역원_고전원문	xml
4	한국언론진흥재단	한국언론진흥재단_뉴스빅데이터_메타데이터_코로나	csv
5	인천광역시	인천광역시_도로 내 IMU 센서	csv
6	농림수산식품교육문화정보원	농림수산식품교육문화정보원_농림축산식품부 농업기상	csv
7	국세청	국세청_근로소득 간이세액표	hwp
8	농림수산식품교육문화정보원	농림수산식품교육문화정보원_팜랩정보_충청남도	shp
9	대검찰청	대검찰청 범죄의 발생 검거상황	csv
10	문화재청	문화재청 문화유산 3D 프린팅 데이터 현황	ply

오픈 API TOP 20

순위	제공기관	오픈 API 명	서비스 유형
1	국토교통부	국토교통부_연속지적도	LINK
2	기상청	기상청_단기예보 ((구)_동네예보) 조회서비스	REST
3	한국수출입은행	한국수출입은행 환율 정보	LINK
4	국토교통부	국토교통부_CCTV 화상자료	LINK
5	한국환경공단	한국환경공단_에어코리아_대기오염정보	REST
6	국토교통부	국토교통부_교통CCTV	LINK
7	국토교통부	국토교통부_시군구	LINK
8	국토교통부	국토교통부_자동차종합정보 API서비스	LINK
9	식품의약품안전처	식품의약품안전처_건강기능식품 기능성 원료인정 현황	LINK
10	농림축산식품부	농림축산식품부_인심식당 정보	LINK

※ 출처: 공공데이터포털(www.data.go.kr) > 데이터찾기 메뉴 > 이슈 및 추천데이터

□ 기상기후데이터허브란?

- ① 기상청과 관계기관이 첨단 기술을 통해 수 십 년 전부터 축적해 온 방대한(수 백 페타바이트급) 기상기후데이터가
- ② 재난 현장과 4차산업 융합(인공지능, 빅데이터) 등에 널리 활용되도록
- ③ 수요자 맞춤형 API를 기반으로 빠르게 통합 연계·제공함으로써
- ④ 국가·사회의 기상기후 의사결정을 지원하는 부가가치 창출형 오픈 플랫폼

▶ 국가, 공공기관, 산업계, 학계 등은 기상기후 데이터 활용에 필요한 대규모 시스템 구축과 전문 기상데이터 처리 부담을 해소하여 국가적 시너지 창출 및 수 천 억원의 예산 절감

〈기상기후데이터허브 구성 개념도〉

2023년,
기상기후데이터허브가 구축되면...



기상기후데이터 허브

방대한 기상기후자료를 국가, 산업, 학계, 국민들 모두 자유롭게 활용·분석

□ 날씨예보 생산

○ 관측부터 예보, 통보까지 전 과정에서 과학적, 체계적 데이터 생산



< 기상청 데이터 7가지 특성 >

특성	내용
① Value	· 일상생활과 직결되고 사회 전 분야 의사결정에 많은 영향력 · 각 분야 융합 시너지 효과가 큰 고부가가치 데이터
② Volume	· 과거 100년부터 누적되며, 미래 100년까지의 자연현상을 설명·예측 · 수 십만 테라바이트급(TB) 전지구 차원의 초대용량 데이터
③ Variety	· 첨단 관측장비·예측기술을 통해 다양한 형태의 정보로 가공·유통 · 국내는 물론 전세계 기상기관의 데이터를 수집·보관
④ Velocity	· 빠르게 변하는 대기현상에 따라 실시간으로 끊임없이 생산 · 국민안전을 위해 생산 즉시 전달되어야 하는 데이터
⑤ High Dimension	· 과학기술을 근간으로 자연현상을 촘촘한 고해상도 데이터로 표현 · 자연(날씨)과 사회(인간)를 이어주는 매개체 역할
⑥ High Uncertainty	· 자연과학 한계로 인한 예측 불확실성을 내포 · 다양한 실험·검증을 통해 끊임없이 증명과정을 거치는 데이터
⑦ High Complexity	· 매우 복잡한 과학모델 속에서 복합 처리와 분석을 통해 생산 · 고성능 슈퍼컴퓨터, 수치모델링 소프트웨어 필수