

# 식품연구 데이터 플랫폼 구축과 데이터 검색

KFRI 한국식품연구원  
안기택



# CONTENTS

- 01 한국식품연구원 소개
- 02 식품 연구데이터 플랫폼
- 03 식품연구데이터 관리계획
- 04 식품 연구데이터 검색기술 개발
- 05 Q&A





# 01

## 한국식품연구원 소개

# 01

## 한국식품연구원 소개

### 설립근거

『과학기술분야 정부출연 연구기관 등의 설립·운영 및 육성에 관한 법률』 (제8조) (1987.12.31.)

식품 분야의 연구개발, 공익가치 창출, 성과확산 및 기술지원 등을 통해  
국가산업발전과 국민 삶의 질 향상에 기여

### 설립목적



### 임무

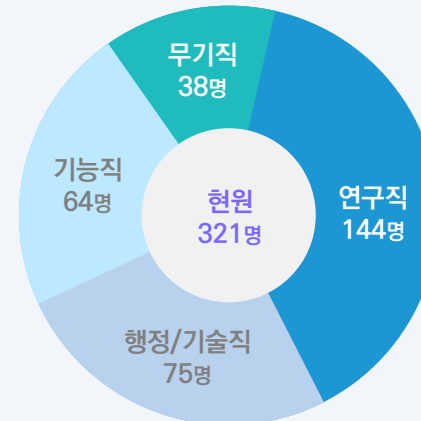
1. 식품 기능성 규명, 신소재·신공정 연구개발
2. 식품 저장·유통·안전성 기술 연구개발
3. 전통식품의 세계화 연구개발
4. 식품 분석, 정보, 표준화 및 기반조성 연구개발
5. 정부, 민간, 법인, 단체 등과 연구개발 협력 및 기술용역 수탁·위탁
6. 중소·중견기업 등 관련 산업계 협력·지원과 기술사업화
7. 주요 임무 분야의 전문인력 양성 및 관련 기술정책 수립 지원

### 인력현황



'24년 1월 기준 321명

(정규직 283명, 무기계약직 38명)



### 위치



본원 완주군 이서면 전북혁신도시 내





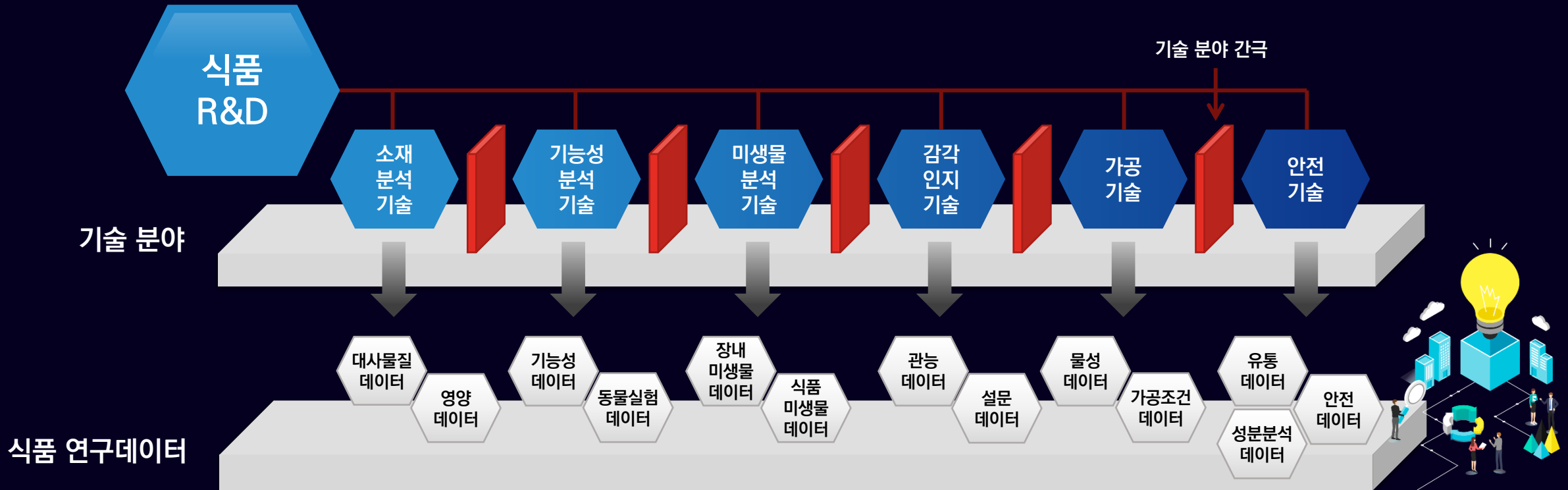
02

**식품 연구데이터 플랫폼**

# 02

## 식품 연구데이터 플랫폼 | 배경

- 식품 연구개발은 다양한 기술을 활용하고 다양한 종류가 생산됨



# 02

## 식품 연구데이터 플랫폼 | 연구목표 및 핵심기술

개별 연구자가 보유하고 있었던 산재된 식품 연구 정보를  
상호연계 · 융복합 · 표준화하여

1단계

내부연구자에게 데이터기반 연구 플랫폼 제공

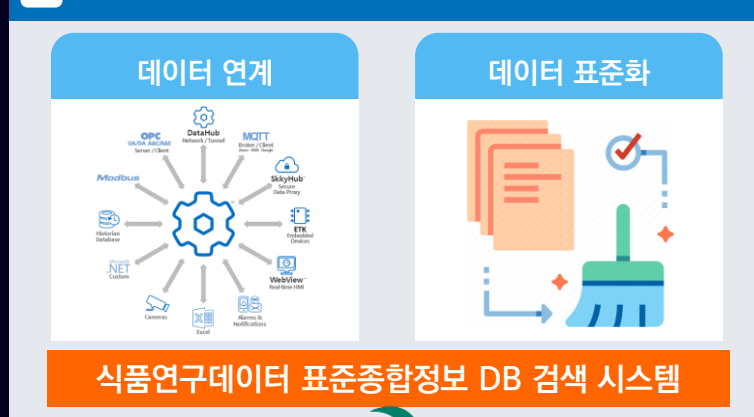
2단계

외부에 개방 · 공동활용 오픈 플랫폼 구축

### 1 식품 연구데이터 통합 인프라 구축



### 2 데이터 상호연계 · 융복합 · 표준화/정제



분석

시각화

### 3 클라우드 기반 식품데이터 분석 플랫폼 개발



### 데이터 수요자

1단계

내부사용자 (KFRI)

2단계

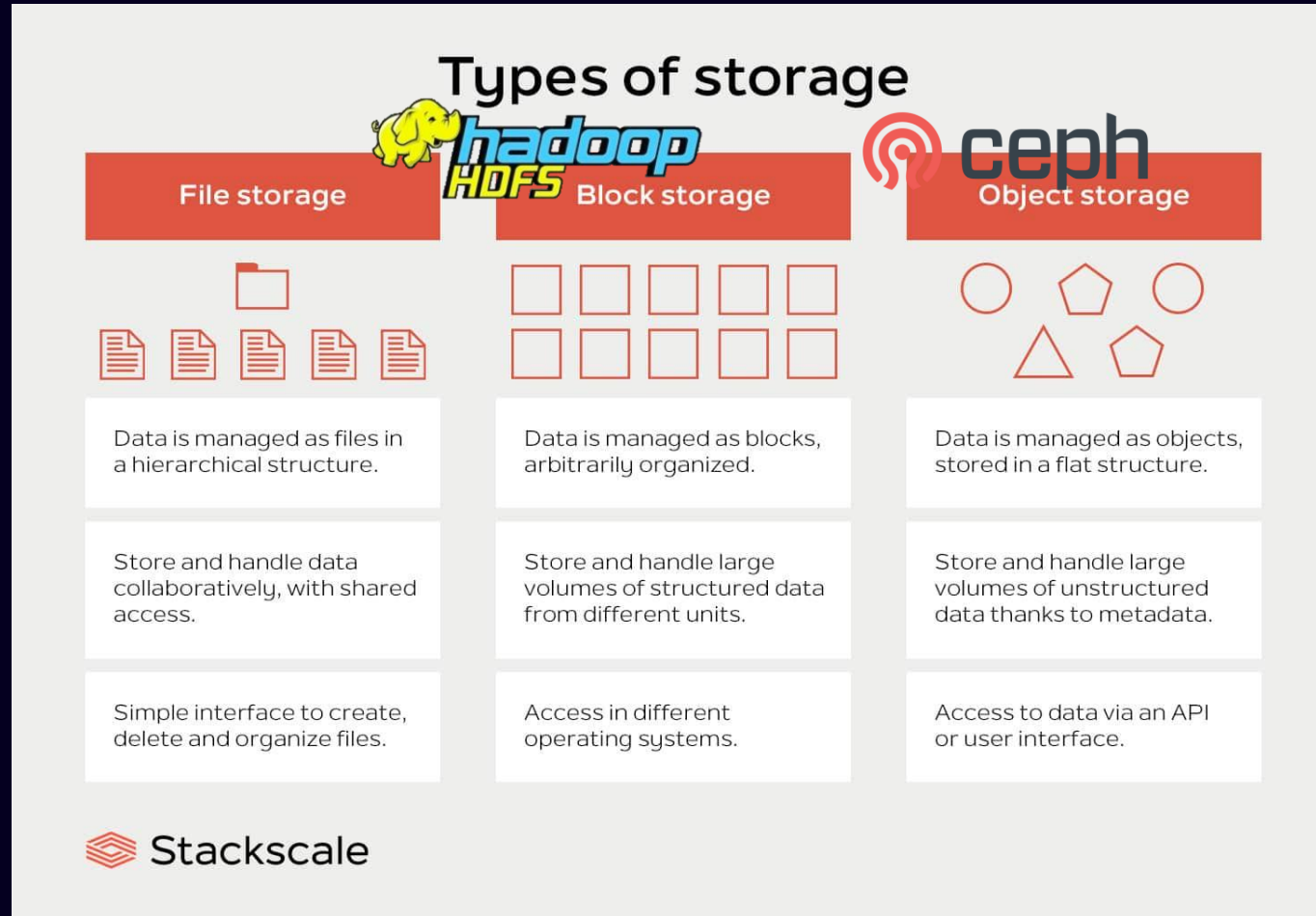
대국민 (외부), 산업체 (외부), 학교 (외부), 연구기관 (외부)

수요자의견 반영

연구자 지향

### 4 데이터 공동활용 공유 · 정책

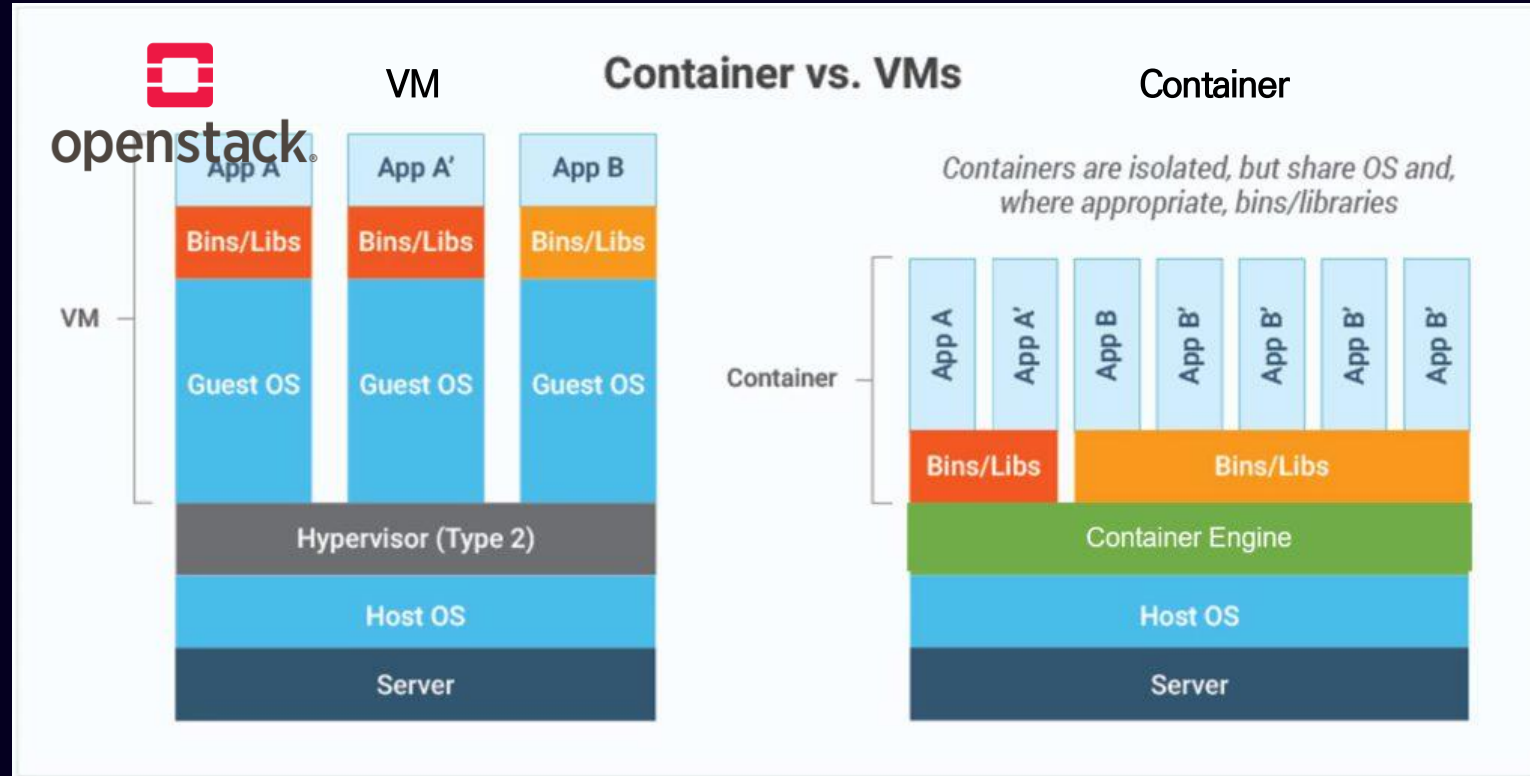






## 02

# 식품 연구데이터 플랫폼 | 서버가상화 기술 적용





03

## 식품연구 데이터 관리계획



# 식품연 연구데이터 관리계획

[별지 제1호 서식] 데이터관리계획(DMP) 양식

(n/N차년도)

## 데이터관리계획

과제: **국문 과제명**

연구책임자: **홍길동**

\* 세부 목차 및 항목은 변경될 수 있으므로 웹플랫폼 참고  
식품데이터플랫폼(FoodInsight),

XXXX. XX. XX.

**KFRI** 한국식품연구원

### 1 과제 소개

#### 1.1. 과제 소개

- ☐ 과제번호: 과제번호
- ☐ 과제명: 국문 과제명
- ☐ 연구책임자: 홍길동
- ☐ 연구기간: XXXX.XX.XX ~ XXXX.XX.XX

### 2 연구데이터 정의

#### 2.1. 연구데이터 구성

- ☐ 데이터 관계도

시료-방법-결과 데이터 관계 표현

- ☐ 연구데이터 요약

구분	DATATYPE_NAME	DATATYPE_DESC
시료/방법/결과		

#### 2.2. 시료 데이터

##### 2.2.1. Datatype name (Category): Datatype description

FIELD	TYPE	UNIT	DESCRIPTION
INDEX			- 필수 필드: INDEX - TYPE: STRING/NUMERIC/FI - UNIT: NUMERIC의 경우 필수

##### 2.2.2. Datatype name (Category): Datatype description

FIELD	TYPE	UNIT	DESCRIPTION

#### 2.3. 결과 데이터

##### 2.3.1. Datatype name (Category-추가예정): Datatype

FIELD	TYPE	UNIT	DESCRIPTION
INDEX			- 필수 필드: INDEX, METHOD - TYPE: STRING/NUMERIC/FI - UNIT: NUMERIC의 경우 - METHOD_INDEX: 방법 - 기제
METHOD_INDEX			

##### 2.3.2. Datatype name (Category-추가예정): Datatype

FIELD	TYPE	UNIT	DESCRIPTION

#### 2.4. 방법 데이터

##### 2.4.1. Datatype name (Category-추가예정): Datatype

연구방법 자체의 기술

##### 2.4.2. Datatype name (Category-추가예정): Datatype

연구방법 자체의 기술

### 3 연구데이터 생산 계획

DATATYPE_NAME	1차년도	2차년도	...	N차년도
시료, 결과데이터만 해당	>=0			

### 4 연구데이터 공개 계획

DATATYPE_NAME	원래 공개 시점	원래 공개 시점
시료, 결과데이터만 해당	일자 지정	일자 지정

### 5 연구데이터 개별관리 계획 (개별관리가 필요한 연구데이터에 대해서만 작성)

DATATYPE_NAME	개별관리 사유	데이터 보관 방법	데이터 공개 방법	데이터 공개 시점
개별관리 대상 연구 데이터만 작성	- 위치: - 기간: - 담당자:			- 원래: - 일회:

\* 세부 목차 및 항목은 변경될 수 있으므로 웹플랫폼 참고  
식품데이터플랫폼(FoodInsight),

# 식품연 연구데이터 관리계획

☑ (적용범위) 기본사업 연구활동을 통해 획득한 데이터 중 데이터관리계획에 포함된 연구데이터

④ (관리체계) 위원회 설치, 관리·운영부서 지정

☑ (데이터관리계획 및 연구데이터 제출)

- 작성대상은 신규과제 공고 시 명시
- 연구개발계획서와 함께 데이터관리계획 제출
- 연구데이터 등록 마감은 연차(최종) 종료일로부터 60일 이내

☒ (공개) 메타데이터 공개\* 후 이용 요청에 따라 심의

\* 관리부서는 데이터관리계획의 공개시점 3개월 전에 변경여부 확인

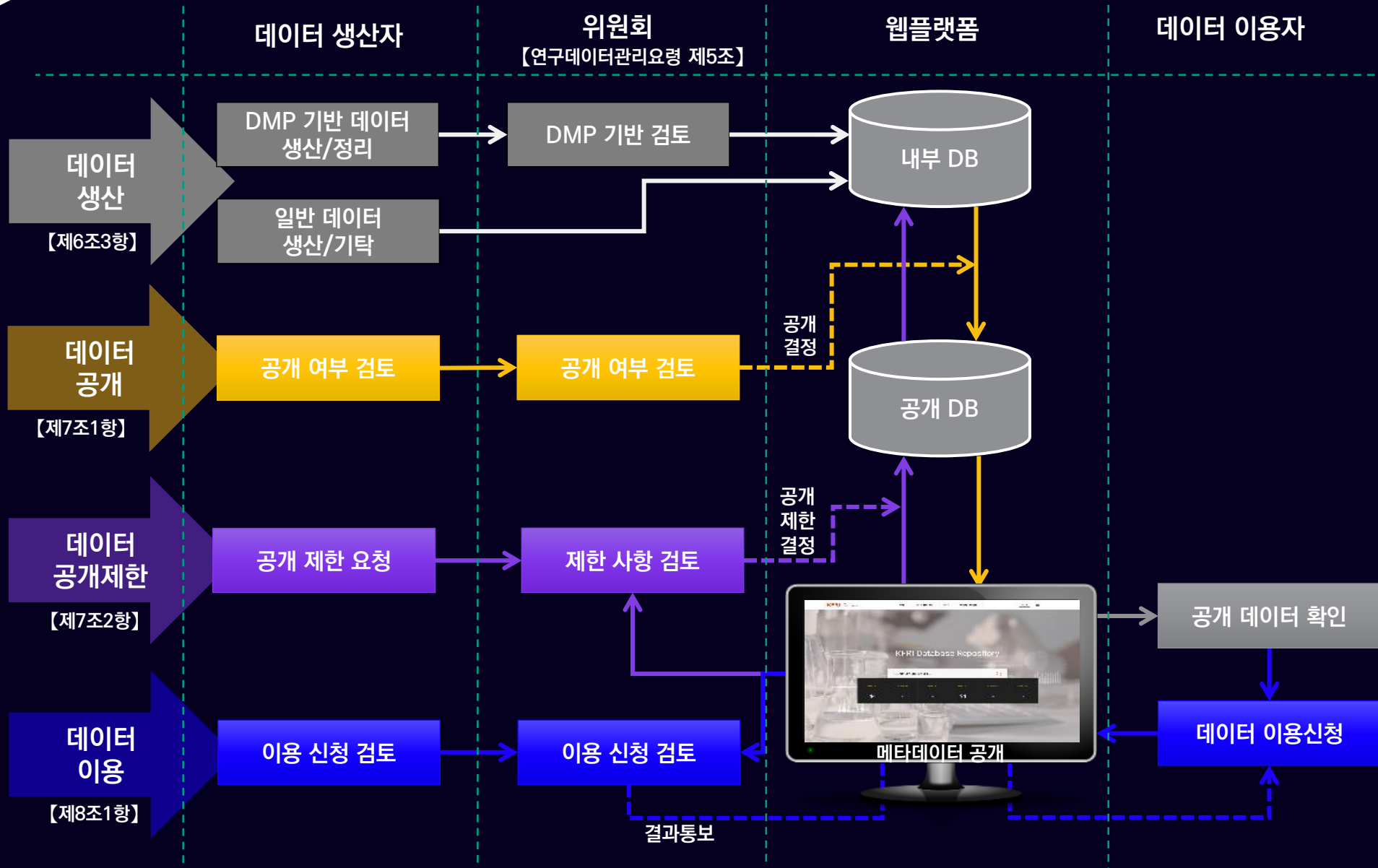
☒ (이용조건) 과제참여자가 아닌 경우 이용 승인 필요

## 제정 2023. 05. 01.

③ 이용하고자 하는 연구데이터에 라이선스가 적용된 경우 반드시 해당 라이선스 이용조건을 확인하고, 저작자, 출처 등을 명시하는 등 이용조건을 준수해야 한다.

# 03

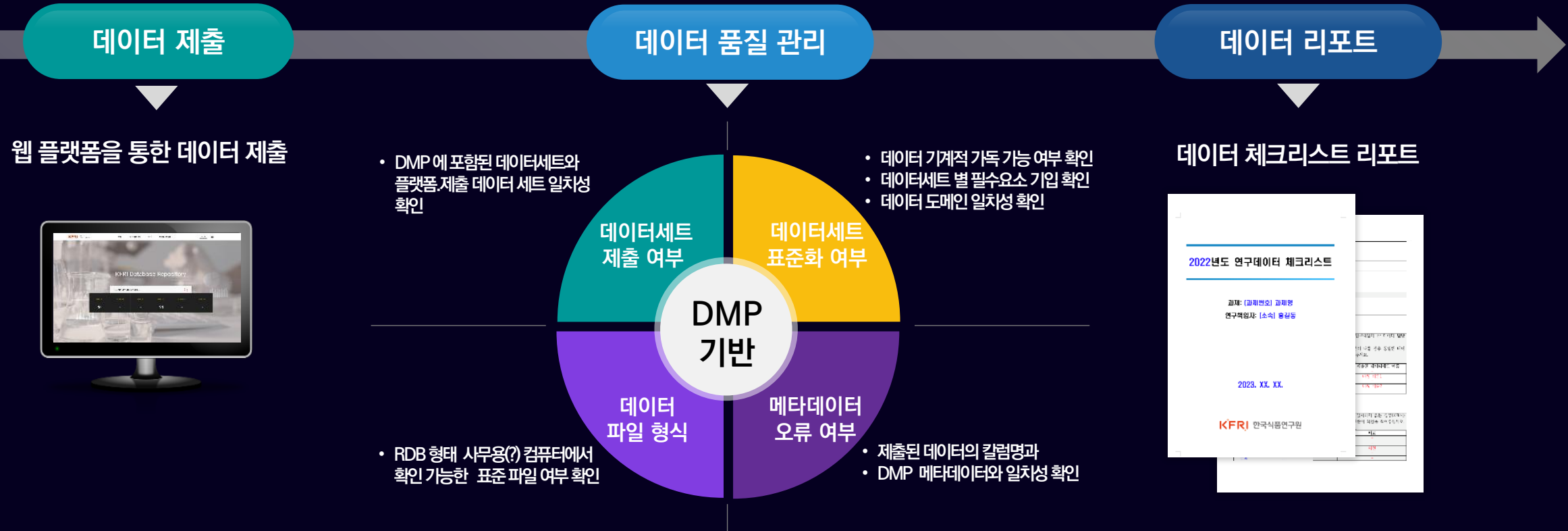
## 식품연구데이터 관리계획



# 03

## 식품연구데이터 관리계획 | 연구데이터 품질관리

- 데이터 품질관리 프로세스는 양식검증과 내용검증의 2단계로 추진
- 양식검증은 데이터 표준화 내용과 과제의 DMP에 기반하여 진행





# 03 식품연 연구데이터 관리계획 | 연구데이터 품질관리

## 데이터관리계획 체크리스트

연구책임자	0
과제정보	0
점검결과	<input checked="" type="checkbox"/> 확정 <input type="checkbox"/> 보완 요청

### □ 점검의견

-  
.

#### 항목

1. 연구계획 대비 생산 예정 연구데이터의 적절성	
1.1. 연구데이터 누락	<input type="checkbox"/>
1.2. 「연구방법 요약」 보완 필요	<input type="checkbox"/>
1.3. 「연구데이터 구성」 보완 필요	<input type="checkbox"/>
1.4. 위탁/용역 연구데이터 미포함	<input type="checkbox"/>
1.5. 데이터세트 변경 필요	<input type="checkbox"/>
1.6. 승인되지 않은 데이터세트 변경 존재	<input type="checkbox"/>
2. 데이터세트별 메타데이터의 적절성	
2.1. 데이터세트 명칭 중복	<input type="checkbox"/>
2.2. 시료/방법/결과 데이터 구분 부정확	<input type="checkbox"/>
2.3. 시료데이터의 표준 'category' 검토 필요	<input type="checkbox"/>
2.4. 방법데이터 보완 필요	<input type="checkbox"/>
2.5. 시료데이터 보완 필요	<input type="checkbox"/>
2.6. 결과데이터 보완 필요	<input type="checkbox"/>
3. 연구데이터 생산·관리·공개 계획의 적절성	
3.1. 연구데이터 생산계획 적절도	<input type="checkbox"/>
3.2. 연구데이터 개발관리 부적절	<input type="checkbox"/>
3.3. 연구데이터 외부공개 시기 적절도	<input type="checkbox"/>
3.4. 연구데이터 외부공개 시기 적절도	<input type="checkbox"/>
3.5. 연구데이터 비공개/개발공개 부적절	<input type="checkbox"/>

### □ 항목별 검토/보완 가이드

\*연실 확정된 최신 연구계획 대비 DMP 작성 및 검토

항목	검토/보완 가이드
1. 연구계획 대비 생산 예정 연구데이터의 적절성	
1.1. 연구데이터 누락	- 차년도 연구계획 대비 생산 예정 연구데이터를 모두 포함 - 세계보존 시료/방법/결과 데이터의 누락이 없는지 확인 - 「연구방법 요약」, 「연구데이터 구성」, 데이터세트 리스트가 일치해야 함
1.2. 「연구방법 요약」 보완 필요	- 생산되는 데이터세트들을 중심으로 연구방법 기술 - 데이터세트 간의 연관성을 그림(연구데이터 구성도)로 표현 - 「연구방법 요약」, 「연구데이터 구성」, 데이터세트 리스트가 일치해야 함
1.3. 「연구데이터 구성」 보완 필요	- 위탁/용역 기관에서 생산한 데이터도 DMP 대상임 - 위탁/용역 생산 데이터는 「연구방법 요약」에서 별도 표기
1.4. 위탁/용역 연구데이터 미포함	- 데이터세트 분리, 병합 등 구성에 관한 변경 의견 (ex. Index만 포함된 시료데이터는 다른 시료데이터와 통합하거나, 결과데이터에 편입)
1.5. 데이터세트 변경 필요	- 과거 DMP 대비 승인되지 않은 데이터세트의 변경, 삭제 등 - 연실에서 확정된 연구계획의 변경과 관련 없는 경우
1.6. 승인되지 않은 데이터세트 변경 존재	
2. 데이터세트별 메타데이터의 적절성	
2.1. 데이터세트 명칭 중복	- 여러 개의 데이터세트에 동일한 이름 사용 금지 - 동일한 형태의 데이터가 아닐 경우, 세계보존에 다른 이름의 데이터세트 정의 필요
2.2. 시료/방법/결과 데이터 구분 부정확	- 시료/결과 데이터 혼동 주의
2.3. 시료데이터의 표준 'category' 검토 필요	- 표준 'category' 선택이 적절인지 확인 - 'etc' 사용 시 표준 'category' 사용이 불가능한 사유 검토 필요
2.4. 방법데이터 보완 필요	- 방법데이터: 실험, 측정, 가공 방법 및 조건 등에 관한 설명, 결과에 영향을 줄 수 있는 사용 장비, 실험 방법, 실험 조건, 측정 단위 등 - 다른 연구에서 생산한 동일 종류의 데이터를 통합하여 사용하고자 할 때 고려되어야 함 사항들이 포함되어야 함 - 유사 연구를 수행하는 연구자가 이해할 수 있는 수준으로 작성
2.5. 시료데이터 보완 필요	- Description 누락, 설명 부족, 무분별한 약어 사용 등 - 유사 연구를 수행하는 연구자가 이해할 수 있는 수준으로 작성 - 데이터세트 내 중복 필드 불가 (다른 데이터세트 내 동일 필드는 가능) - Index 필드 작성 시 시료/방법/결과 데이터의 연결성 고려 - 필수 필드 중 누락된 필드가 있는지 확인 - 동일한 샘플/실험대상어에 대한 시료/결과데이터들의 연결성을 Index를 통해 확인 가능해야 함 - 원고 필드 활용 권장 - Index 명명 형식/규칙/외의 설명 권장
3. 연구데이터 생산·관리·공개 계획의 적절성	
3.1. 연구데이터 생산계획 적절도	- 과거 생산계획 변경 불가 - 연구계획 대비 작성된 생산계획의 적절성 확인 - 당해년도 비생산 연구데이터는 삭제하지 말고 '필드 표기' (ex. 계수: 0) - 생산되는 데이터의 개수는 샘플 수(ex. 역설의 행 수) 기준으로 작성
3.2. 연구데이터 개발관리 부적절	- 개발관리 사용, 관리방법 등 자세히 기술 필요
3.3. 연구데이터 외부공개 시기 적절도	- 최대 기준: 연차종료 시점 2년 후 공개 - 최대 기준 초과 시 사유 검토 필요 - 원고 기준: 연차종료 시점에 공개
3.4. 연구데이터 외부공개 시기 적절도	- 최대 기준: 단계/최종 종료 시점 2년 후 공개 - 최대 기준 초과 시 사유 검토 필요 - 원고 기준: 단계/최종 종료 시점에 공개
3.5. 연구데이터 비공개/개발공개 부적절	- 비공개 사유 자세히 기술 필요

## 연구데이터 체크리스트

연구책임자	(소속) 성명
과제정보	(과제번호) 과제명
점검결과	<input type="checkbox"/> 확정 <input type="checkbox"/> 보완 요청

### □ 점검의견

-  
.

종류	명칭	점검 결과	비고
시료/방법/결과 (DMP 기준)	DATATYPE_NAME (DMP 기준)	<input type="checkbox"/> 미제출/확인불가	
		<input type="checkbox"/> 메타데이터 불일치	
시료/방법/결과 (DMP 기준)	DATATYPE_NAME (DMP 기준)	<input type="checkbox"/> 메타데이터 보완	
		<input type="checkbox"/> 기타	
시료/방법/결과 (사건도 DMP 대상 데이터)	DATATYPE_NAME (사건도 DMP 대상 데이터)	<input type="checkbox"/> 미제출/확인불가	
		<input type="checkbox"/> 메타데이터 불일치	
시료/방법/결과 (DMP 이외 데이터)	DATATYPE_NAME (DMP 이외 데이터)	<input type="checkbox"/> 확인불가	
		<input type="checkbox"/> 메타데이터 보완	
시료/방법/결과 (DMP 이외 데이터)	DATATYPE_NAME (DMP 이외 데이터)	<input type="checkbox"/> 메타데이터 보완	
		<input type="checkbox"/> 기타	

### □ 보완 가이드

검토의견	보완 가이드
데이터 미제출/확인불가	
파일 미제출	- DMP에 작성된 데이터세트나 연구데이터 파일이 미제출된 경우 제출 필요
잘못된 파일 제출	- DMP 과제번호가 잘못 매칭되었을 경우 수정 필요
파일 확인 불가	- 제출된 파일이 파일번호가 걸려있는 등 확인이 불가능한 경우 비밀번호 해독하여 제출
데이터 기재내용 불가	
파일 내 이미지가 삽입됨	- 장형 데이터(엑셀 등) 내 이미지, 지도 등이 포함된 경우 수정 필요 • 이미지는 별도 파일로 제출하고 기존 엑셀 파일에 새로운 필 추가하거나 별도의 엑셀 파일로 작성하여 이미지 파일에 대한 메타데이터 함께 제출 • 이미지가 장형데이터에 대한 참조일 경우 해당 설명(메타데이터)에 포함
파일 내 시트가 하나 이상 포함됨	- 파일 내 시트가 여러 개 포함되어 있는 경우 각 시트를 개별 파일로 분리 제출 - 하나의 시트 내에 여러 개의 데이터가 포함된 경우 각 데이터를 개별 파일로 분리 제출
시트 내 데이터가 하나 이상 포함됨	- 데이터를 구역 외에 (ex. 언어설명, 사용한 방법 등) 정보가 기입되어 있는 경우 데이터를 구역 내에 기입하거나 개별 파일로 제출
병합된 셀 존재	- 병합된 셀을 분리하고, 분리한 각 셀에 동일한 값을 기재
필드 하나 행 이상 포함	- 필드명은 하나의 행으로 구성하여 기재
데이터데이터 보완 (당해년도 DMP 누락)	
데이터세트 명칭 중복	- 동일한 데이터세트가 아닐 경우, 다른 명칭으로 정의 (최소 '세토/법/세부 구분')
시료/방법/결과 데이터 구분 부정확	- 시료데이터: 연구, 실험, 분석의 대상에 대한 정보 - 결과데이터: 시료데이터에 대하여 '방법데이터'에 따라 획득한 측정, 결과 정보 - 방법데이터: 실험, 측정, 가공 방법 및 조건 등에 관한 설명, 결과에 영향을 줄 수 있는 사용 장비, 실험 방법, 실험 조건, 측정 단위 등
방법데이터 보완 필요	- 유사 연구를 수행하는 연구자가 이해할 수 있도록 작성 - 다른 연구에서 생산한 동일 종류의 데이터를 통합하여 사용하고자 할 때 고려되어야 할 사항이 모두 포함하도록 작성
시료데이터 보완 필요	- 유사 연구를 수행하는 연구자가 이해할 수 있도록 작성 • Description 설명 부족 및 누락, 사소한 약어 사용 등 확인 및 수정 • 시료/방법/결과 데이터 간의 연결성을 확인할 수 있도록 Index 필드 작성 Index에 원고 필드명 활용을 권장하며, 아닐 시 방법 형식/규칙/외의 설명 작성 - 하위 데이터세트 내에서 필드명이 중복되지 않도록 작성 - 필수 필드(시료: Index, 결과: Index, Method, Index) 포함 - 각 필드명 데이터의 누락이 없도록 작성, 불일치 여부 확인 및 수정
결과데이터 보완 필요	
기타	
데이터세트 리스트 변경 필요	- 시료와 결과 정보가 존재한 데이터세트는 시료데이터와 결과데이터로 분리 - 삭제정보만 포함된 시료데이터는 다른 데이터세트로 편입
raw 데이터 미제출	- raw 데이터가 아닌 파일이 제출된 경우 raw 데이터 제출 • FASTQ 등의 용류되는 표준양식의 데이터, 이미지, 영상 등의 비정형 데이터는 엑셀 등의 형태로 변환된 필요없이 raw 파일로 제출 가능



04

**식품 연구데이터 검색기술 개발**

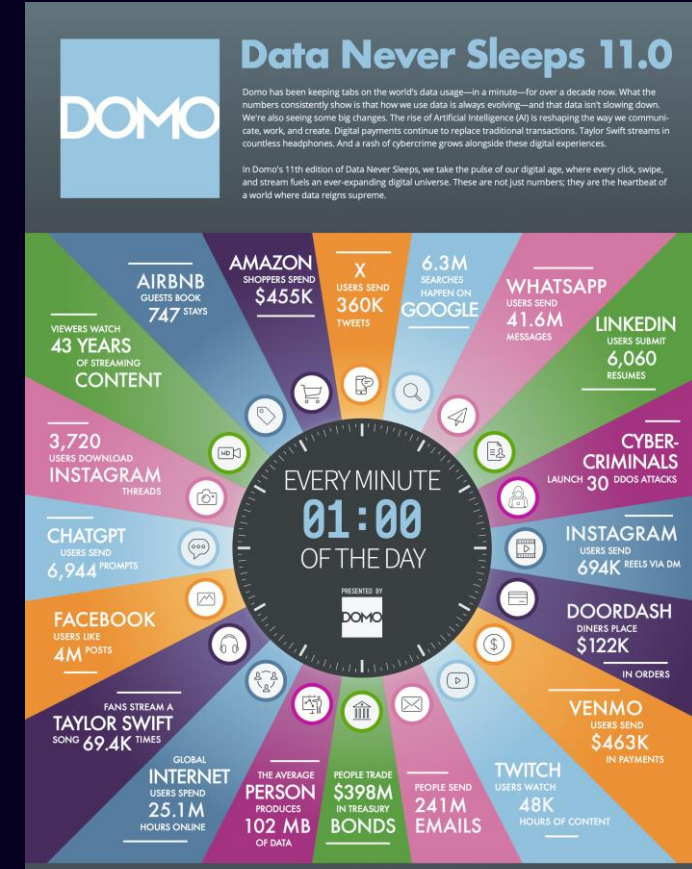


# 04

## 식품 연구데이터 검색기술 개발

### :: 연구데이터 관리 요령 제정 (2023.5.1. 시행)

- ✓ 검색을 하는 이유는
  - 다양한 정보를 찾고,
  - 질문에 대한 답을 얻으며,
  - 지식을 확장하고,
  - 의사결정을 내리기 위해서
- 이를 통해 궁금증을 해결하고,  
새로운 아이디어를 발견하며,  
학습과 연구를 지원하는 등 다양한 목적을 달성



뉴스와이어

### 삼성경제연구소 '정보홍수 속에서 金脈찾기:빅 데이터(Big Data) 분석과 활용'

디지털 공간에서의 활동 일상화와 모바일 기기의 사용 확대로 개인과 조직의 활동기록이 축적되면서 경영에 유용한 정보도 폭발적으로 증가하고 있다. 웹사이트의 방문...

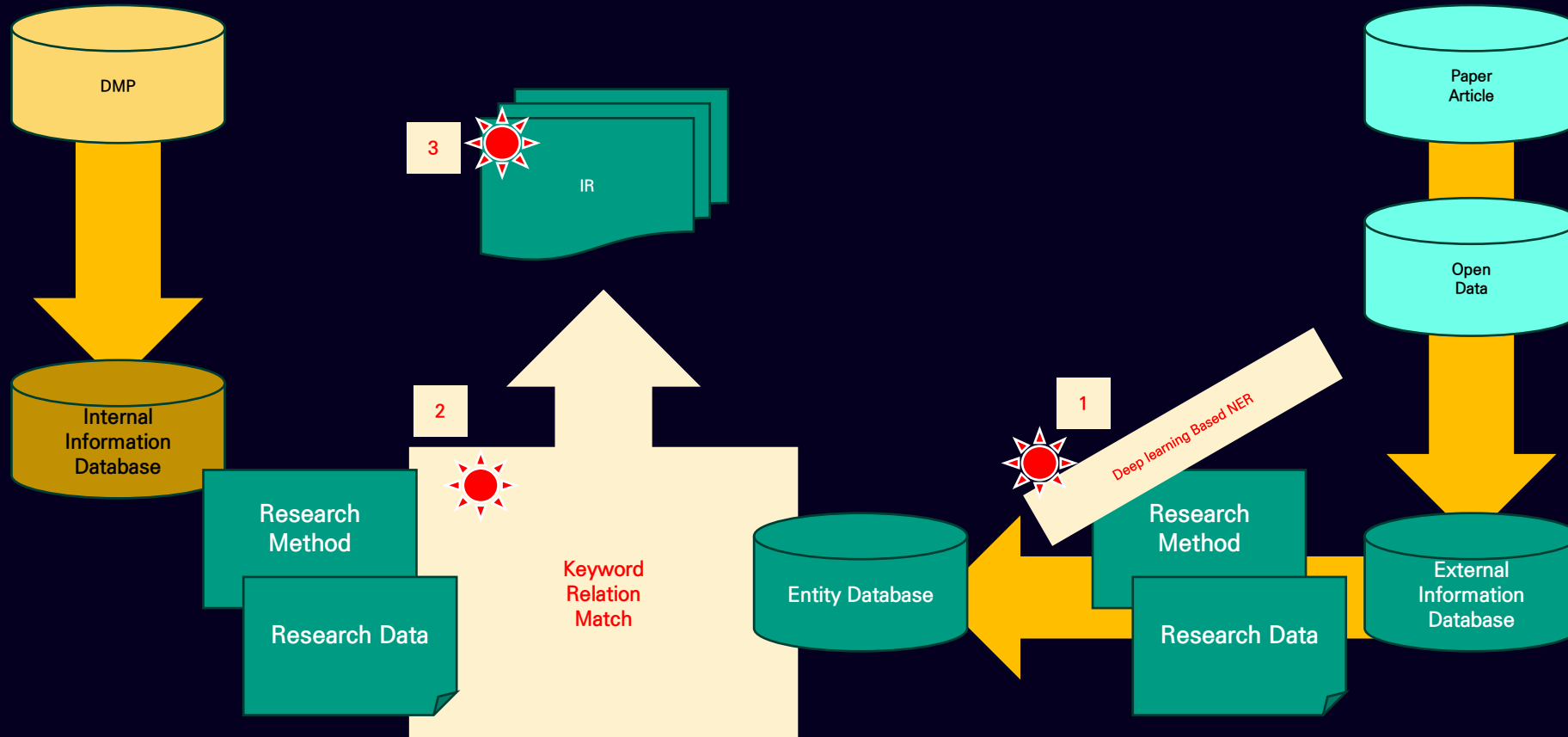
2011. 2. 10.

구분	주요 내용
연구 목적	빅데이터의 활용 가능성
연구 방법	데이터 마이닝, 통계 분석
연구 결과	빅데이터의 활용 가능성
연구 결론	빅데이터의 활용 가능성

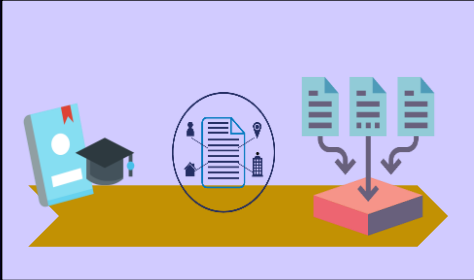
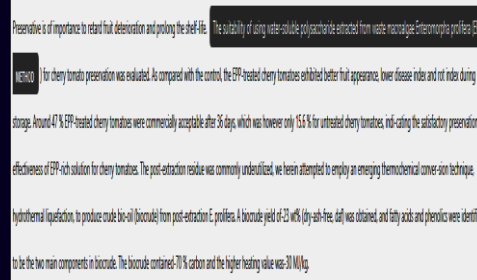
# 04

## 식품 연구데이터 검색기술 개발

식품연구에서는 ...



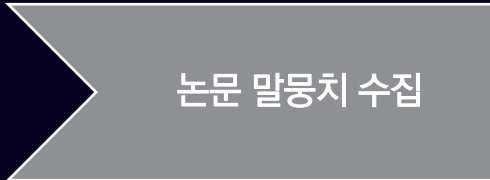
검색 Query

NER 기반 Query  
Expansion

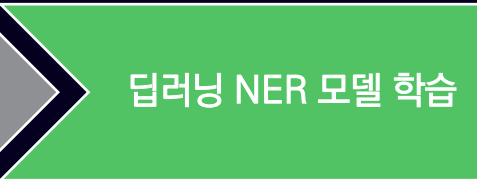
DB Search



논문 말뭉치 수집



딤러닝 NER 모델 학습



Entity DB 구축





## 식품 연구데이터 검색기술 개발

33개 식품 연구 분야 학술지, 360,087개의 논문 정보 수집

5,998개의 논문 초록 데이터를 ‘소재’, ‘연구방법’, ‘연구결과’ Tagging

ID	소재
KFRI-001274	whey and milk fat globule membrane (MFGM)
KFRI-001275	leguminous seeds and oilseeds samples
KFRI-001276	non-dairy cream milk fat cream
KFRI-001277	zein-pectin nanoparticle-stabilized corn
KFRI-001278	Amaranthus L. 'Gartenfuchsschwanz' (A. hybridus L.) 'Red Callaloo' (A. tricolor L.) 'Gemuse-Amaranth' (A. tricolor L.)
KFRI-001279	electrochemically active bacteria (EAB)
KFRI-001280	beta-lactoglobulin (BLG) aggregates
KFRI-001281	norfloxacin (NFX)
KFRI-001282	soy products, pork and milk samples for BAs
KFRI-001283	corn, potato, and pea starches subjected to
KFRI-001284	Soy protein hydrolysates
KFRI-001285	ripened Pu-erh tea (PE-21) and tea heads
KFRI-001286	pea starch (PS)
KFRI-001287	high-purity BCFA-enriched structured TAG

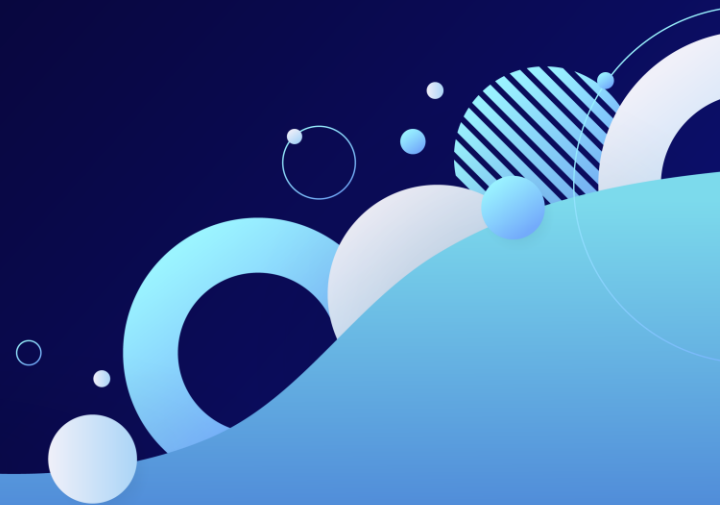
ID	연구 방법
KFRI-001309	microencapsulated spray-drying process
KFRI-001310	toasting, dehulling, sterilization (sterilized jarred lupins), oven baking (cookies), frying
KFRI-001311	storage temperature, duration, and
KFRI-001312	dynamic stabilization mechanism Ultra-high-resolution microscopy
KFRI-001313	analyzed for their chemical characteristics XGBoost machine learning algorithm
KFRI-001314	An indirect and ultrasensitive ratiometric molecularly imprinted (MIP) sensor, based
KFRI-001315	ultrasound treatment
KFRI-001316	transfer behaviour of 61 pesticides and elaborated a predictive model
KFRI-001317	ratiometric fluorescent probe

ID	연구 결과
KFRI-001294	developed sensor exhibited remarkable sensitivity towards GA, featuring a low LOD of 1.30 nM and a linear range spanning from
KFRI-001295	Brewing water with high pH values (>8.10) and high TDS content (>140 ppm) resulted in a lower overall aroma acceptability for tea
KFRI-001296	successfully for visually distinguishing three
KFRI-001297	desirable pH, reducing sugar, protein, total phenolic content and isoflavone for achieved milk than conventional soy milk
KFRI-001298	EU-CF and CAR-CF NPs were bactericidal to E. coli at dosage of 0.09% and 0.07% (v/v) they were just bacteriostatic to
KFRI-001299	Staphylococcus aureus at 0.10% and 0.08% YNITGW (IC50 = 9.78 & PLUSMN: 0.13 mM) showed the strongest activity
KFRI-001300	the sequence of affinity for CPI was puerarin > apigenin > naringenin > epigallocatechin gallate

# Q & A

gt@kfri.re.kr

데이터 공유 문의: data@kfri.re.kr



# 감사합니다.

식품연구 데이터 플랫폼 구축과 데이터 검색

