

의안번호	제 146 호
보 고 년 월 일	2011. 11. 2 (제 8 회)

보 고 사 항

빅데이터를 활용한 스마트 정부 구현(안)

제 출 자	국가정보화전략위원회 위원장 이각범
제출년월일	2011. 10. 26

1. 보고주문

- 빅데이터를 활용한 스마트 정부 구현(안)을 별지와 같이 보고함

2. 보고이유

- 폭증하는 데이터가 경제적 자산이 되는 ‘빅데이터 시대’의 정부 역량 강화를 위한 빅데이터를 활용한 스마트 정부 구현 방안을 보고하고, 위원회의 의견을 반영하여 추진하고자 함

3. 주요내용

가. 추진배경

- 모바일 인터넷과 소셜 미디어 등장으로 폭증하는 데이터가 경제적 자산이 되는 ‘빅데이터 시대’ 도래
- 단순히 정보를 전달하던 웹1.0에서 참여와 소통으로 집단 지성을 창출하는 웹 2.0을 지나 데이터 분석으로 융합지식을 창출하는 웹 3.0시대 도래
- 융합 지식 창출을 통한 정부 운영 효율화 및 투명성 제고, 맞춤형 국민서비스 제공 요구 증대

※ 美 미통령 과학기술자문위원회는 ‘Designing a Digital Future’ 보고서에서 ‘모든 미국 연방 정부 기관은 빅 데이터 전략이 필요하다’고 강조(‘10.12)

< 빅데이터의 정의 >

◆ 사회영역에 활용되는 빅데이터란?

대용량 데이터를 활용, 분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술

나. 빅데이터의 활용 효과

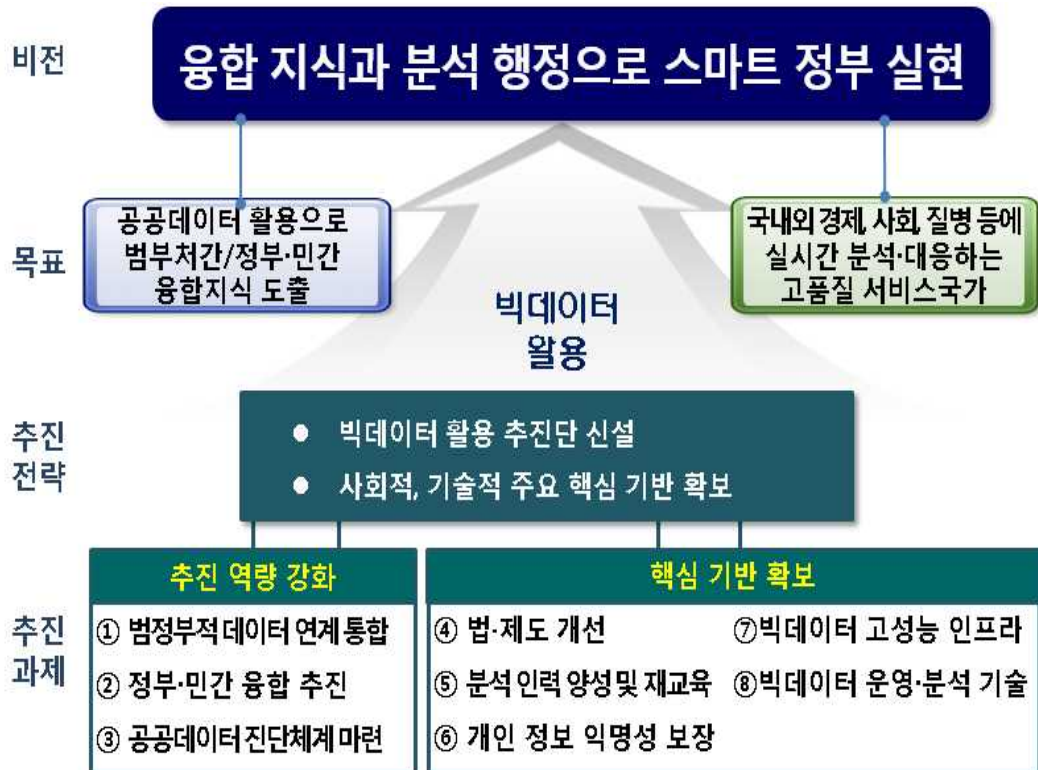
- 스마트폰과 소셜 네트워크 효과로 디지털 공간의 데이터 빅뱅이 발생하여 정부 차원의 데이터 지식 확보 및 활용 방안 요구
- 정부의 빅 데이터 활용시 예산 절감, 대내외 변화에 대한 신속한 대처, 삶의 질과 정부 신뢰도 향상이 가능
 - ※ 오바마 정부의 'Pillbox 프로젝트'로 연간 5,000만 달러 비용 절감 효과
 - ※ 독일 연방 노동기관에서 빅데이터 활용 맞춤형 고용으로 3년간 백억 유로 비용 절감
- 공공 데이터와 소셜 데이터 등을 분석하여 대내외의 이슈와 변화를 감지하고 적시 대책을 수립
- 모바일 혁명시대, 스마트 사회에서는 빅데이터가 핵심 자원이며 이를 수집·분석하는 새로운 빅데이터 기업과 新산업 출현
- 공공 분야 빅데이터 활용시 경제 효과는 국내의 경우 10.7조원 이상
 - ※ EU는 비용절감, 부정 및 오류에 따른 손실 감소, 세수 증대에 따른 비용 효과가 1,500~3,000억 유로 규모, 한국은 최소 10.7조원 추산

시사점

- ▶ 분석·예측 시대의 도래로 빅데이터를 활용하는 효율적인 정부 운영을 도모하는 선진국들의 현황으로 볼 때, 빅데이터를 활용하지 못하는 정부는 국제 사회에서 국가 경쟁력 하락의 위험과 대면하게 될 것임
- ▶ 데이터 활용에 따른 우려 불식을 위하여 '정보 익명성(Anonymizing)'과 '수집 데이터 비저장' 등을 포함하는 공공데이터 활용 가이드라인 수립 필요

다. 비전 및 추진과제

- 정부가 빅데이터 활용 및 인프라 기반조성을 선도적으로 추진하기 위한 '빅데이터 활용추진단' 신설
- 빅데이터 경쟁력 함양을 위한 국가적, 사회적 기반을 정비하고, 산·학·연 협력을 통한 원천기술 개발로 빅데이터 핵심 기반 확보



라. 과제별 주요내용

① 범정부적 데이터 연계·분석체계 구축

- 기존 행정정보공동이용센터 등을 확대하여 부처 및 공공기관의 정보를 지속적으로 연계·수집할 수 있는 체계 수립
- 예측기반의 국정 운영 혁신을 위한 국가 전반의 데이터 수집·분석 체계 마련

② 정부·민간 데이터 융합 추진

- 소셜 미디어, 포털 데이터 등과 같은 민간 데이터와 공공 데이터의 연계·활용을 위한 체계 및 기술 확립
- 산·학 등이 보유한 각종 정보를 범국가적으로 연계·수집할 수 있도록 저장소 구축·운영
- 공공 데이터의 단계적 개방으로 민간의 가치창출 및 기업 활동 강화 지원
- 민간 기업의 공공 데이터 접근에 따른 ‘데이터 라이선스’ 규약을 신설하여 데이터 활용의 공익성을 확보

③ 공공데이터 진단체계 구축

- 공공데이터의 효율적 관리를 위한 범정부적 MDM 체계 구축

※ MDM(Master Data Management): 다수의 시스템에서 활용되고 중앙에서 통제·관리되어야 하는 표준 참조데이터

- 범부처 데이터 공동 활용을 위한 품질관리 기준 및 체계 마련

④ 법·제도 개선

- 범국가적 빅데이터 활용 추진 기본계획 수립 및 관련 법령 개정
- 공공부문 빅데이터 분석 서비스 활용 촉진 방안 및 성과관리 체계 확립

수 분석 인력 양성 및 재교육

- 빅데이터 시대 新수요에 대응한 빅데이터 분석 전문인력 양성
- 빅데이터 활용 역량 강화를 위한 다양한 재교육 프로그램 제공

⑥ 개인정보 익명성 보장체계 확립

- 안전하고 신뢰할 수 있는 공공데이터 공개 및 활용을 위해 개인정보와 프라이버시 보호 관련 기본 원칙을 체계화
- 데이터의 안전한 공유와 유통을 위한 강화된 보안대책 수립
- 데이터의 개방·공유·활용에 따른 정부 공공데이터 활용 가이드라인 마련

⑦ 기술개발

- 빅데이터 관련 인프라 기술 개발
- 빅데이터 운영·분석 기술 개발

【 별 지 】



빅데이터를 활용한 스마트 정부 구현

2011. 11

국가정보화전략위원회

목 차

I. 빅데이터의 시대의 도래	1
1. 추진배경	1
2. 빅데이터란?	2
3. 빅데이터의 활용 가치	3
가. 국가적 가치	3
나. 경제적 가치	4
II. 해외 빅데이터 활용 현황	5
1. 국가 안보	5
2. 국가 인프라 혁신	6
3. 정부 효율성 향상	7
III. 빅데이터를 활용한 정책 프로세스 혁신	9
1. 정부 의사결정 구조의 변화 방향	9
2. 빅데이터 활용 시나리오	10
가. 재난 전조 감지	10
나. 구제역 예방	12
다. 맞춤형 복지 실현	14
라. 물가 관리	16
마. 과학기술·의료 선진화	19
IV. 빅데이터 활용 추진 방안	21
1. 비전 및 목표	21
2. 세부 추진 과제	22
가. 추진 역량 강화	22
나. 핵심기반 확보	24
3. 향후 계획	27

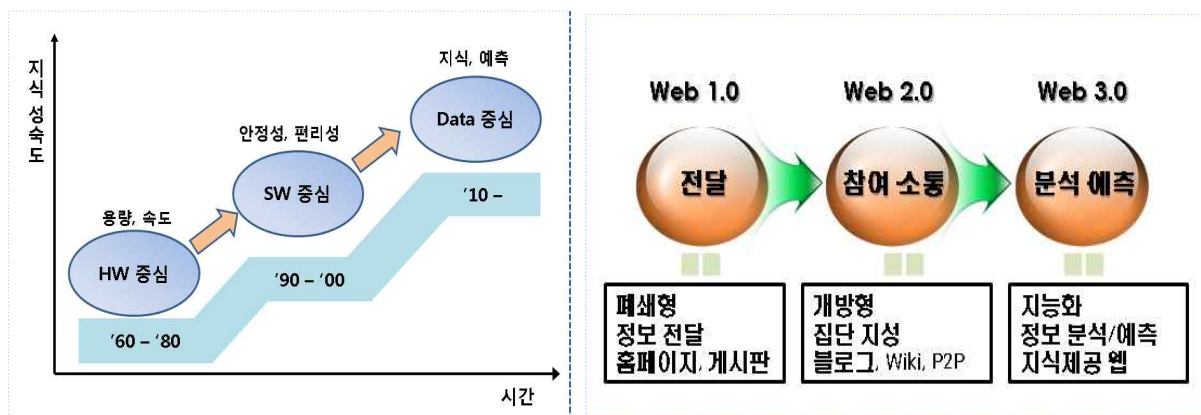
1. 추진 배경

□ 모바일 인터넷과 소셜 미디어 등장으로 폭증하는 데이터가 경제적 자산이 되는 '빅데이터 시대' 도래

- 과학·IT 패러다임은 기존 HW(과거)·SW(현재) 중심에서 데이터 빅뱅을 활용, 정치·사회·경제 등 제반 이슈와 연계된 분석·예측으로 진화

※ MS는 제4패러다임으로 '데이터 인텐시브 컴퓨팅' 지목하고, 자연현상표현(1세대), 모델/일반화(2세대), 현상시뮬레이션(3세대), 데이터분석/예측(4세대)로 표현 ('The 4th Paradigm,'10)

- 단순히 정보를 전달하던 웹1.0에서 참여와 소통으로 집단 지성을 창출하는 웹 2.0을 지나 데이터 분석으로 융합지식을 창출하는 웹 3.0시대 도래



□ 융합 지식 창출을 통한 정부 운영 효율화 및 투명성 제고, 맞춤형 국민서비스 제공 요구 증대

- 국내·외 경제·사회 정세를 실시간으로 파악하는 분석력을 확보하여, 질병·금융위기 등 위협에 대한 빠른 대응이 필요
- 데이터 공개·활용으로 인건비 절감, 세수 투명성 확보, 국민 복지 향상 등 효율적인 정부 구현 가능

※ 영국, 미국, 캐나다, 호주 등 다수 국가들이 오픈 데이터 정책에 참여

2. 빅데이터란?

◆ 사회영역에 활용되는 빅데이터란?

대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술

- ※ (기술적 정의) 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 (IDC, '11)
- ※ (규모적 정의) 일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터 (매킨지, '11)
- ※ (방법적 정의) 빅 데이터는 당초 수십~수천 테라바이트에 달하는 거대한 데이터 집합 자체만을 지칭하였으나, 점차 관련 도구, 플랫폼, 분석기법까지 포괄하는 용어로 변화 (삼성경제연구소, '10)

< 빅데이터 활용에 따른 분석 방법의 변화 >

분류		현재(AS-IS)	융합지식기반(TO-BE)
활용 변화	데이터개방	웹기반 인터페이스	원본/분석/가시화 3계층
	이슈접근	후집계/원인 파악하는 사후대책반	실시간 이슈 탐지를 통한 선대응적 기획반
	활용형태	부처별 수직적 활용	범부처/민간 수평적 분석
	주체	업무 운영담당자	분석, 기획 담당자
분석 변화	분석대상	정형화된 DB 데이터	정형데이터 + SNS, 질문 등의 비정형 데이터
	규모	기가~테라바이트급	페타~제타 바이트급
	분석범위	단일 저장소(Silo)	다중 저장소
	적용시간	일괄(batch) 처리	인타임 처리
	데이터	저장 후 분석	흐름(on-the-fly) 분석

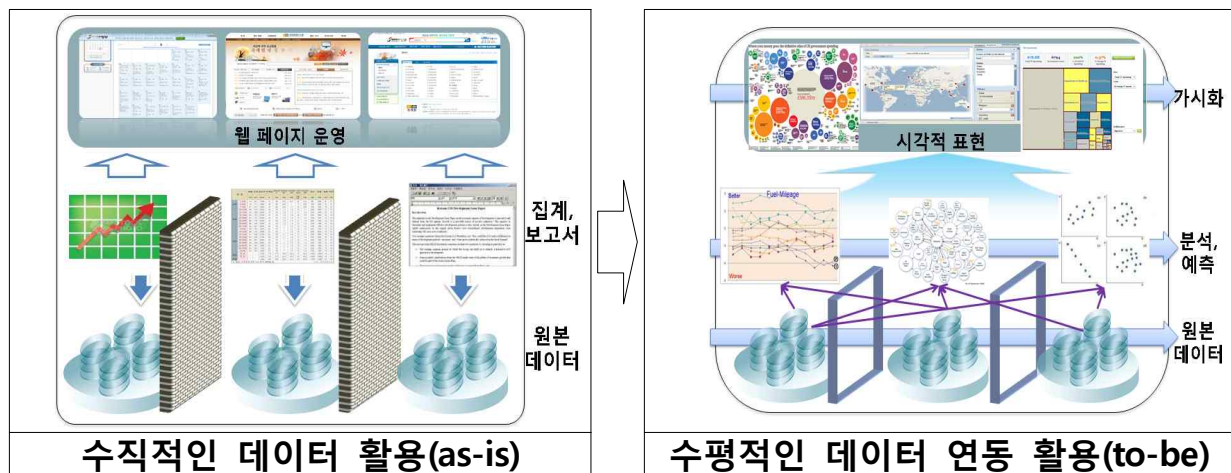
3. 빅데이터의 활용 가치

가. 국가적 가치

□ 미래 국가 경쟁력은 '빅 데이터'의 활용에 달렸음

- 美 대통령 과학기술자문위는 '모든 미국 연방 정부 기관은 빅 데이터 전략이 필요하다'고 강조('10.12, Designing a Digital Future 보고서)
- 스마트폰과 소셜 네트워크 효과로 디지털 공간의 데이터 빅뱅이 발생하여 정부 차원의 데이터 지식 확보 및 활용 방안을 요구
 - ※ 2010년 1.2 제타바이트 정보 생산 (제타바이트 크기: 미의회도서관 4백만배)
 - ※ 디지털 데이터 규모 성장: '09년 800 엑사바이트(10^{18}) → '20년 35 제타바이트(10^{21})
- 빅 데이터는 정부, 기업, 의료, 학술연구 분야에서 그 가치가 입증됨에 따라 다양한 분야에서 도입·활용이 확대되는 추세

□ 기존 정부·공공데이터의 수직적·단절적 활용에서 기관간 연동을 통한 수평적 활용으로 혁신하여 실시간 이슈를 탐지하고 선대응하는 정부를 구현



- 기관간 데이터의 상호접근으로 포괄적 정보를 생성
- 기관별 웹페이지를 통해 제공되는 정보를 '보는' 형태에서 실시간으로 변화하는 데이터를 '활용·분석'하는 형태로 전환

□ 정부의 빅 데이터 활용시 예산 절감, 대내외 변화에 대한 신속한 대처, 삶의 질과 정부 신뢰도 향상이 가능

- 빅데이터 기술을 활용한 공공 데이터 분석·운영으로 정부 비용 절감
 - ※ 오바마 정부의 'Phillbox 프로젝트'로 연간 5000만 달러 비용 절감 효과(미국 연간 100만건의 알약 문의 X 건당 확인 비용 평균 50달러)
 - ※ 독일 연방 노동기관에서 빅데이터 활용 맞춤형 고용으로 3년간 백억 유로 비용 절감
- 공공 데이터와 소셜 데이터 등을 분석하여 대내외의 이슈와 변화를 감지하고 인타임 대책을 수립
 - ※ 구제역, 조류독감, 국제금융위기 등의 이슈를 사전에 인자·분석하여 피해를 최소화
- 공공 데이터 공개로 국가 운영을 투명화·효율화하여 정부 신뢰도 향상

나. 경제적 가치

□ 공공 분야 빅데이터 활용시 경제 효과는 국내의 경우 10.7조원 이상

- ※ EU는 비용절감, 부정 및 오류에 따른 손실 감소, 세수 증대에 따른 비용 효과가 1,500~3,000억 유로 규모, 한국은 최소 비율 적용 시 10.7조원 추산

□ 모바일 혁명시대, 스마트 사회에서는 빅데이터가 핵심 자원이며 이를 수집·분석하는 새로운 빅데이터 기업과 新산업 출현

- 구글 등 글로벌 선진기업들은 웹사이트 방문기록, 검색통계, 소셜 미디어 기록 등 빅데이터 분석을 활용한 新비즈니스 모델에 주목
 - ※ 일본에서는 트위터에서의 반응을 모니터링한 후 빠른 주기로 마케팅 방법을 변경

<구글 독감 예상치와 실측데이터>



시사점

- ▶ 분석·예측 시대의 도래로 빅데이터를 활용하는 효율적인 정부 운영을 도모하는 선진국들의 현황으로 볼 때, 빅데이터를 활용하지 못하는 정부는 국제 사회에서 국가 경쟁력 하락의 위험과 대면하게 될 것임
- ▶ 데이터 활용에 따른 우려 불식을 위하여 '정보 익명성(Anonymizing)'과 '수집 데이터 비저장' 등을 포함하는 공공데이터 활용 가이드라인 수립 필요

Ⅱ 해외 빅데이터 활용 현황

1. 국가 안보

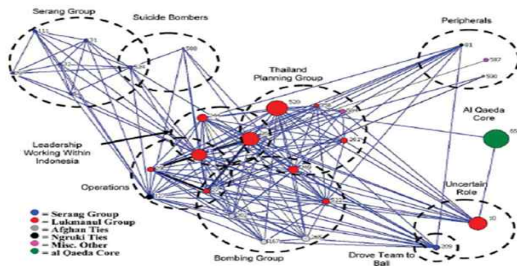
□ (재난방재) 국가 안전을 위협하는 글로벌 요인이나 질병, 위기 등을 수평적 리스닝을 통해 분석하여 선제 대응

- 기후변화, 해양오염방지, 방사능 유출탐지 분야 등 광범위한 지역에 걸쳐 생성되는 대용량 정보를 실시간 처리
 - ※ M2M확산으로 현재 3000만개 이상의 사물인터넷 센서가 설치되어 있으며, 향후 5년간 연평균 35% 증가예상(KT경제경영연구소, '11)
 - ※ 미국, 영국, 일본 등 방재선진국들은 첨단장비나 센서를 활용 다양한 감시체제를 연계, 구축하여 재난 예방 상호 협력(SERI, '08)
- 금융사기, 자금세탁 방지 및 고객 행동 분석 솔루션 등에도 활용
 - ※ 테라데이터와 SAS는 데이터웨어하우징 기술과 분석기술을 결합하여 '자금세탁방지', '신용위험관리' 등 위험관리와 금융사기 예방을 위한 솔루션 공급(ZDNet Korea, '11)
- 태풍, 강수량, 지진 등의 정보를 분석하여 쓰나미, 홍수 등의 가능성을 예측
 - ※ IBM은 물보유량과 기상 데이터, 현재 강수량등을 종합하여 홍수, 가뭄 대책 수립

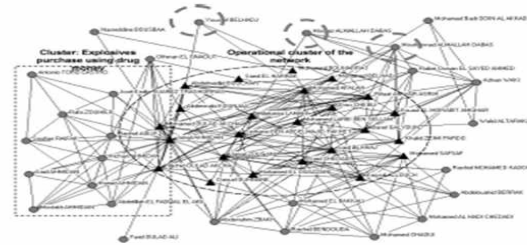
□ (테러감지) 국가 안보 위협 요인을 사전에 감지·대응

- 상황인지 및 공유 정보 분석을 통해 정부·민간의 정보활용 효과를 극대화
 - 세계적으로 국가안전(Homeland Security)에 대한 관심 및 투자 증대
 - ※ 미국은 홈랜드 안보를 위해 정보(intelligence) 부분에 연간 200조원 이상 투입
- 소셜네트워크, 신문, 잡지, 기사 등으로부터 수집된 정보를 분석하여 테러 동향을 파악, 테러 징후 사전 예측
 - ※ 미국 국가테러방지센터는 24시간 운영되며 매일 10,000개 이상의 테러관련정보를 수신
- 5세대 전쟁으로 정의되는 안보위협이 복잡 다양성을 네트워크 분석을 통해 해결하고자 노력
 - ※ 5세대전쟁 : 전통적인 전쟁에 정치, 경제, 사회문화적 요소와 사이버공간 등이 추가되어 전쟁과 범죄가 융합되는 현상

- 테러 네트워크 파악을 위한 정보 수집에 대한 근거법안 마련 추진



출처 Maguire, Atran & Sageman, 2008
<2002 발리 폭탄테러 네트워크>



출처 Jordan, Manas & Horsburgh, 2008
<글로벌 자하드 네트워크>

2. 국가 인프라 혁신

□ (국가 경제) 물가 안정화 및 글로벌 금융 위기 대응력 확보

- 정유, 통신 등 독과점 산업의 가격 결정 구조개선 등에 빅데이터 분석을 활용하여 물가안정에 기여
 - ※ 영국의 국립 연구소에서는 대규모 임상 데이터를 사용하여 신약의 임상과 비용 효과를 조사하여 제약 및 의료 회사들과 가격 및 시장 접근 조건을 협상하는 가이드라인으로 활용
- 글로벌 금융 위기에 대응력을 높이기 위해 시장 감시 및 국제 금융 정보 수집 및 분석
 - 국내외 금융시장 동향 분석과 전망, 위험요인에 대한 정책 대응 지원
 - 국가별 금융·경제 동향 및 위험요인에 대한 분석과 대응방안 제시

□ (의료·과학기술) 데이터분석을 통한 새로운 연구 성과 도출로 국가 경쟁력 제고



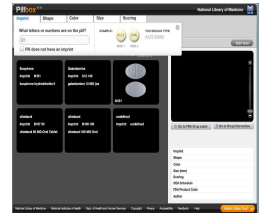
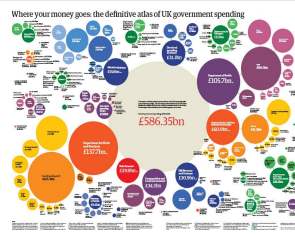

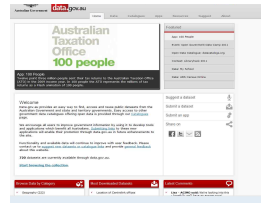
- 유전자·질병·진료 데이터 분석을 통해 의료기술 향상
 - 유전자 데이터 분석을 통해 질병 발생 확률을 예측
 - ※ 카디오DX(미국)는 관상 동맥 질환을 예측하기 위해 900명 이상의 환자에서 1억개 이상의 유전자 샘플 데이터를 추출, 23번의 테스트를 거쳐 예측되는 유전자를 확인
 - 검색 데이터 및 소셜 데이터 분석을 통해 질병 확산 상황 분석
 - ※ 구글은 지역별 검색어 분석을 통해 전세계 독감 확산 지도를 서비스 하고 있음. 이 서비스는 의료 당국에서 조사를 통해 알게 되는 정보보다 빠르고 정확한 것으로 알려짐

- 기존 시스템으로는 분석 불가능한 대규모 데이터에 대한 분석 처리
 - ※ 세계 최대의 가속기인 LHC(Large Hardron Collider)는 연간 15 페타바이트(10^{15})의 관측 데이터 생성
 - ※ Pan-STARRS 프로젝트는 전체 밤하늘의 사진을 촬영하여 다양한 관측활동을 수행 중. 연간 산출되는 전체 관측 데이터의 양은 2.5 페타 바이트
- (교통·항만) 차량, 선박 교통량 및 환경 데이터 분석 및 예측을 통한 교통망·항만 시스템 최적화
 - 도로에 설치된 카메라, 소음측정기, 오염도 측정기 등에서 수집된 데이터를 분석하여 교통량 측정 및 교통 정책 수립
 - ※ NoiseTube 프로젝트는 개인의 휴대폰을 사용하여 위치정보 및 노이즈 양을 측정해서 지도상에 노이즈 맵을 제공
 - 다양한 데이터를 융합 처리함으로써 선박 추적 및 선박 운항 정밀도를 개선하여 선박 운항 효율성을 개선
 - ※ VTS(Vessel Traffic Service) 레이더, CCTV, AIS(Automatic Identification)등 다양 데이터 융합 처리를 통한 선박 추적 및 운항 정밀도 향상
 - ※ 흩어져 존재하는 선박 운항 정보 및 항구 정보를 통합 분석함으로써 최적화된 VTM(Vessel Traffic Management)을 실현하여 선박의 안정성을 높이고, 선박과 항구 운영의 비효율성 제거

3. 정부 효율성 향상

- 빅데이터 활용을 통한 정부의 효율적 운영 등으로 국가 경쟁력 제고
 - 공공 업무에 필요한 정보의 양이 급격히 증가하고 있어 효율적인 공공 업무 처리 시스템을 위해 빅데이터 기법을 활용
 - ※ 독일 BA(연방 노동기구)에서는 실업자 이력, 고용 중재, 사람들이 직장을 구하는데 걸리는 시간과 고용주에 대한 방대한 량의 데이터를 분석, 분류하여 고용 중재에 활용
 - ※ 매킨지 보고서에서도 정부와 부동산 분야가 GDP 영향 크기가 가장 큰 것으로 집계되었으며, 유럽의 공공 분야에 적용시 경제 가치는 150억~300억 유로로 추산
 - 지방자치 단체의 정보시스템과 중앙 정부를 연계하는 접근 채널 다중화 인프라 구축
 - ※ 영국은 지방정부의 전자정부 구축 촉진을 위하여 'National CRM Programme'를 추진하여 Brent 의회의 전자상점 Pathfinder 프로젝트, Bromley Pathfinder 프로젝트, Camden Pathfinder 프로젝트 등 10여개 이상의 CRM 구축 사업을 진행

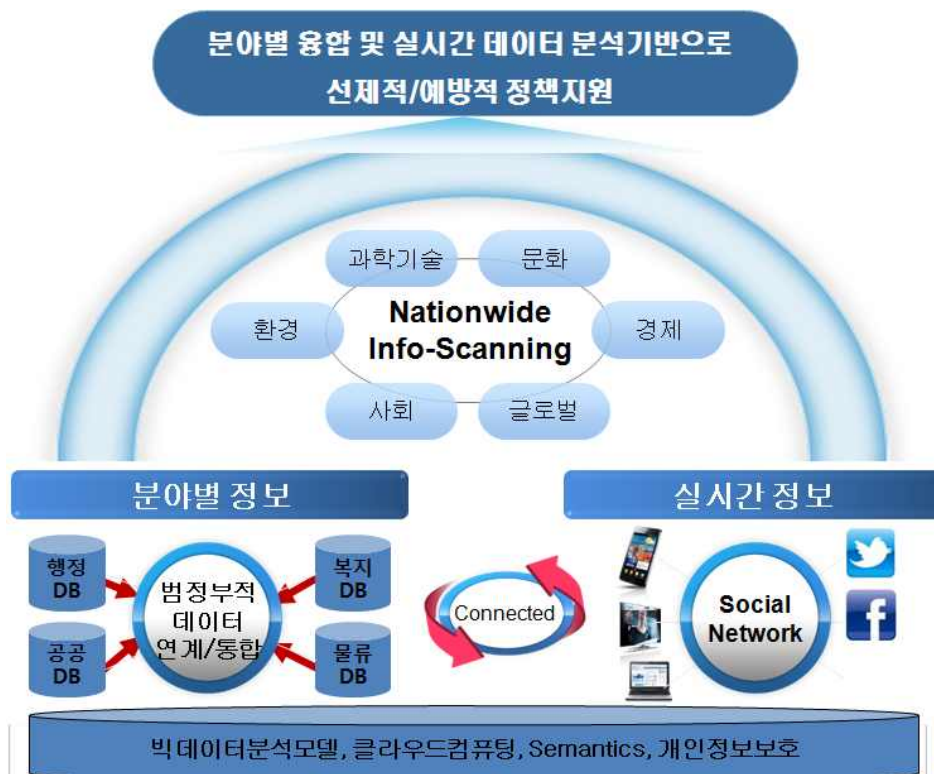
< 주요국 빅데이터 활용 현황 >

국가	활용분야	내용	비고
미국	국토보안	<ul style="list-style-type: none"> 9.11 이후 미국은 국토안보부를 중심으로 테러·범죄 방지를 위한 범정부적 빅데이터 수집, 분석 및 예측체계를 도입 부시행정부의 국토안보부 장관인 Michael Chertoff는 국토보안을 위한 빅데이터 추진현황 언급 국내외 금융 시스템의 개인, 기관의 금융거래 감시로 자금 세탁 및 테러 자금 조달 색출 강화 	<p>Big data analysis can thwart sec</p> <p>By Ryan Kim Jun. 22, 2011, 1:13pm PT Comments Off</p>  <p>Michael Chertoff, Department of Homeland Security, is seen in a video frame. The text mentions his role in intelligence and the explosion of Internet and the government security is to for the kinds of signals and traffic that yield important insight</p> <p>< 국토 보안을 위한 빅데이터 추진현황 보도 ></p>
	치안	<ul style="list-style-type: none"> FBI의 종합 DNA 색인 시스템(CODIS) DNA포렌직, 클라우드DNA분석 등 “빅DNA데이터”의 활용을 통해 2007년 45,400건의 범인 DNA Hit rate 달성 1시간 안에 범인 DNA 분석을 위한 주정부 데이터 연계 및 빅데이터 실시간 분석 솔루션 확보 	 <p>< Arkansas주 CODIS 분석 시스템 ></p>
	의료	<ul style="list-style-type: none"> 오바마 Health.20 - 필박스 프로젝트(Pillbox) 국립보건원(National Library of Medicine)의 사이트로 약 검색을 서비스 Pillbox를 통해 수집된 빅데이터를 통해 후천성면역결핍증(HIV) 등 관리대상 주요 질병의 분포, 연도별 증가 등에 대한 통계치 확보 가능 	 <p>< PillBox 사용 화면 ></p>
영국	정보공개	<ul style="list-style-type: none"> 영국은 정부 사이트(data.gov.uk)를 통해 공공부문의 정보 공유 및 활용을 위한 데이터 원스톱 서비스 제공 정부의 투명성 제고, 국민의 권리 향상, 데이터의 공개를 통한 경제적 사회적 가치 증대, 차세대 웹(web of data)에서 주도권 획득 목표 일반인들의 참여를 장려하고 아이디어 수렴, 앱 개발, 데이터 공개 등의 주제에 대한 커뮤니티 제공 	 <p>< Where did my tax go? ></p>
싱가포르	국가 위험관리	<ul style="list-style-type: none"> 싱가포르 정부는 빈번히 발생하는 테러 및 전염병으로 인한 불확실한 미래 대비를 위하여 2004년부터 빅데이터 기반 위험 관리 계획을 추진 RAHS(Risk Assessment & Horizon Scanning) 시스템을 통해 질병, 금융위기 등 모든 국가적 위험을 수집 및 분석하여 위험을 선제적으로 관리 수집된 위험 정보는 시뮬레이션, 시나리오 기법 등을 통해 분석되어 사전에 위험을 예측하고 대응 방안을 모색함 	 <p>< RAHS 2.0 시스템: 세계 신종 인플루엔자 발병 현황 ></p>
호주	정보공개	<ul style="list-style-type: none"> 호주 정보관리청은 정부 2.0을 통한 정보 개방 방대한 양의 정보를 검색하고 분석 및 재사용할 수 있도록 자동화된 툴을 활용하여 시간과 자원을 절감 AGIMO 산하 정부 2.0 전략/서비스 팀에서는 정부 데이터에 대한 리포지터리 및 검색 툴을 서비스하는 data.gov.au 웹사이트 운영 	 <p>< data.gov.au ></p>

III 빅데이터를 활용한 정책 프로세스 혁신

1. 정부 의사결정 구조의 변화 방향

- (기존) 기관별로 단절된 정보의 제한적 활용으로 합리적 의사결정을 지원하는데 한계
 - 정보의 부족과 단절로 인한 공백을 경험치이나 결단으로 메우게 되어 현실과 괴리된 의사결정 가능성 상존
- (향후) 분야별 정보의 수평적 융합과 실시간 데이터 분석에 기반한 의사결정으로 스마트 정부 구현
 - 범정부적인 데이터 융합과 소셜 미디어 등 민간 데이터 활용으로 빅데이터 분석의 토대 구축
 - 실시간 수집되는 방대한 실제 정보에 대한 빅데이터 분석을 통해 적시성 있고 현실적이고 선제적인 의사결정 및 정책집행 체계 마련



2. 빅데이터 활용 시나리오

가. 재난 전조 감지

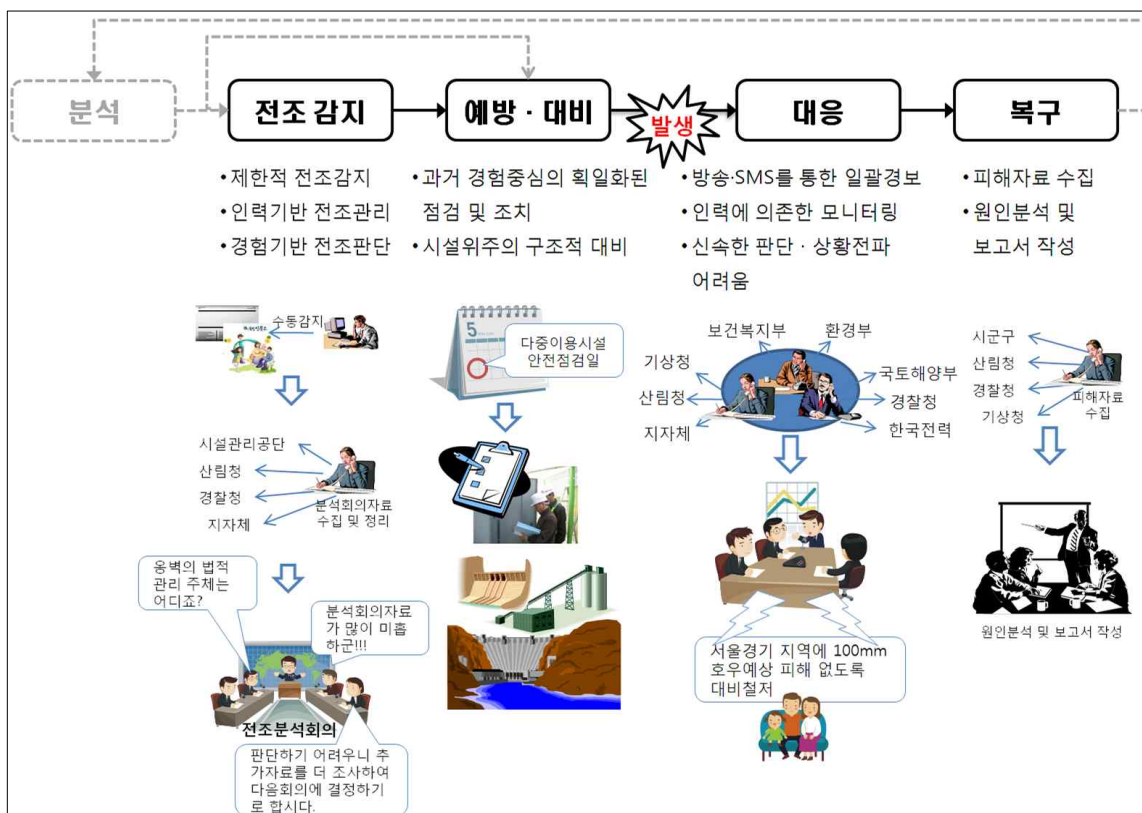
□ 개요

- 피해·위험 사전예측을 통한 분석 기반의 예방·대비 체계를 구축으로 실시간 상황분석 및 신속·정확한 대응·복구 가능

□ 현황

- 경험기반의 획일화된 점검과 조치, 시설위주의 구조적 대비, 인력에 의존한 상황모니터링 등으로 신속한 판단이 어려움
- ※ '11년 7월 서울시 집중호우...붕괴우려지역의 주민에게 사전대피를 지시하지 않았고, 상습침수지역에 대한 교통통제도 늦어져 피해가 커졌음 (내일신문,'11.7.28.)
- ※ 현재 전조감지는 인적재난(화재, 교통사고 제외)에 한해 인력으로 신문기사 및 방송 보도, 민원 등에서만 수행- 연간 1,800여 수집 건의 약 50%만 전조정보(효율성 저하)

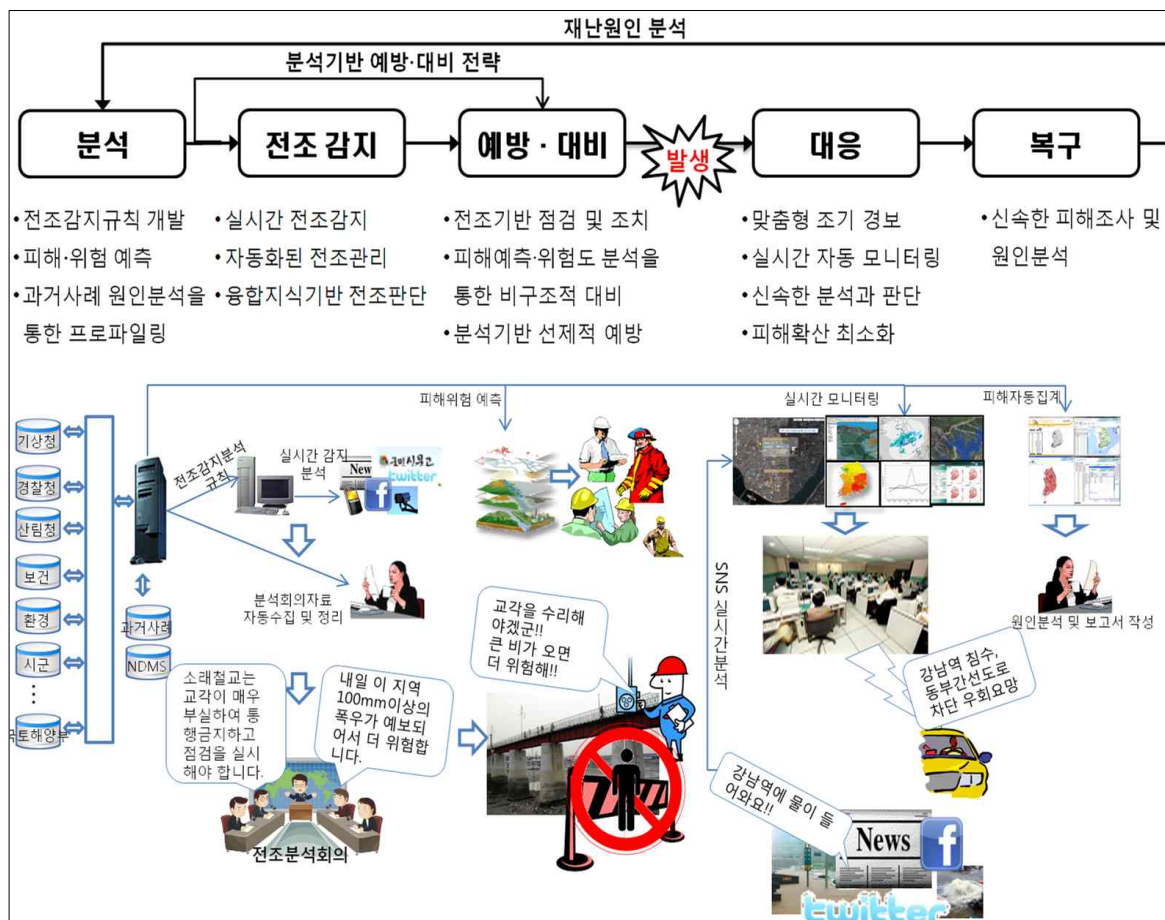
< 경험기반 재난관리 >



□ 빅데이터 활용 시나리오

- 실시간 전조감지를 통해 대형재난으로의 확산을 사전에 차단
 - 과거 피해기록과 예상 강우에 따른 도시 침수지역·구간을 사전 예측
 - 유관기관의 실시간 정보의 자동분석으로 신속한 상황판단
 - 소셜데이터·뉴스 등의 실시간 분석으로 피해지역 맞춤형 경보 및 대응
 - 융·복합 피해원인 분석 및 환류체계 구축으로 지능형 재난관리
 - ※ 실시간 전조감지와 융합분석을 통한 위험등급 예측으로 신속·정확한 예방
 - ※ 전조감지 및 대응을 위한 빅데이터 : 기상정보(기상청), 수문(국토해양부), 절개지(산림청), 센서(기상청, 산림청), CCTV(경찰청), 계측기, 교통정보(경찰청), 인공위성자료(국정원, 기상청), 재해위험 예측, 과거 피해기록, 언론(신문기사, 방송보도), 민원 및 여론, 소셜네트워크, 지리·사회·환경적 특성 등

< 분석기반 재난관리 >



나. 구제역 예방

□ 개요

- 각 기관별로 산재되어 있는 가축방역 및 농장정보 관련 데이터를 융합·분석하여 구제역 등가축질병예방 및 통제를 위한 의 선제적 의사결정지원

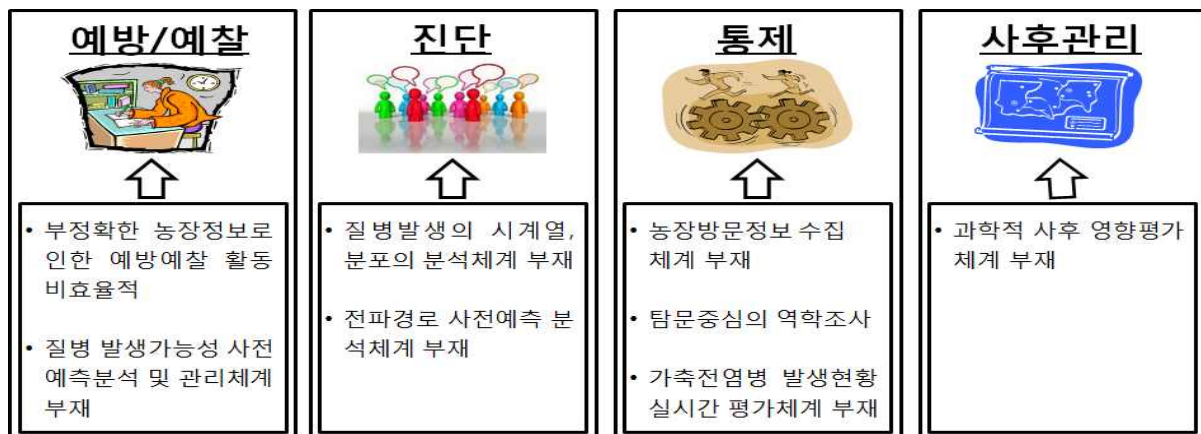
□ 현황

- 가축질병 관련 방역정보가 산재되어 있고, 방역의 핵심정보인 농장 방문(자)정보 등이 수집되지 않아 신속한 의사결정 및 대책 수립에 한계
 - 축산농장을 관리하는 시스템마다 농장DB를 기관별로 관리하고 있어 연계 및 공동활용의 문제 발생
 - 부정확한 농장DB 및 탐문중심 역학조사 등으로 예방·예찰부터 진단·통제까지 프로세스가 비효율적

<축산관련 기관 및 시스템 현황>

기관명	시스템명	주요 정보
농림수산물식품부	AGRIX(농업경영체, 소부르셀라검사시스템)	농장정보
행정안전부	축산행정시스템, 주민등록시스템	농장·주민등록정보
축산물품질평가원	소이력관리시스템	농장·가축이동정보
가축위생방역지원본부	축산농가방역정보시스템	농장정보, 예찰실적
농림수산검역검사본부	국가동물방역통합시스템(KAHIS) 축산물안전관리시스템	농장정보, 도축장, 도계장
종축개량협회	종돈혈통관리시스템	농장·가축이동정보

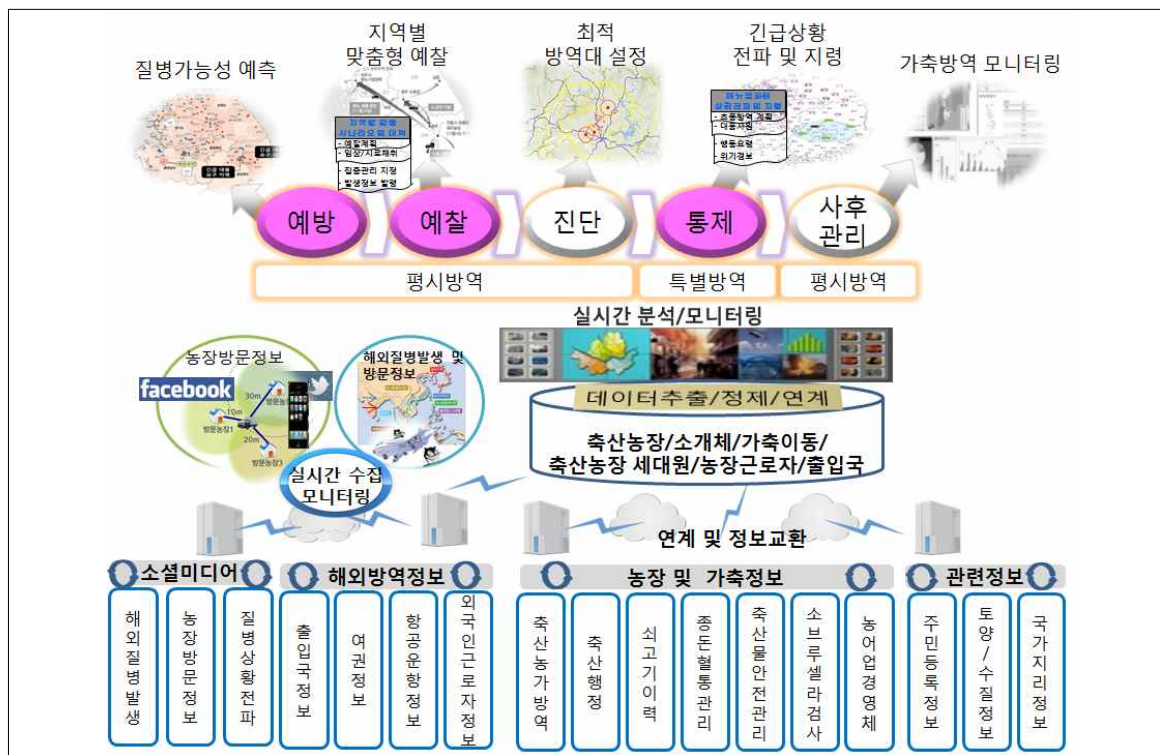
<현행 가축질병 관리체계의 현황>



□ 빅데이터 활용 시나리오

- 기관별 농장정보, 축산관계자, 가축이동 정보는 물론 해외 질병발생, 출입국자, 농장방문 정보 등을 융합·분석하여 사전예방 및 선제적 의사결정 지원
- **[예방]** 해외 질병발생 동향, 출국자, 축산관련 차량이동, 소셜 미디어 정보 등을 실시간 자동수집·분석하여 질병 가능성을 예측하고 소독·예방접종 등을 통해 사전예방
- **[예찰]** 가축전염병 이상 징후 발견 시, 확산 및 전파경로 사전분석을 통해 예찰활동, 방역조치 등 지역별 맞춤형 예찰 매뉴얼 체계 마련
- **[진단]** 국가지리정보, 토양, 수질관리 등 시스템과 데이터의 연계를 통해 신속하고, 효과적으로 진단
- **[통제]** 발생지역에 대한 살처분, 매몰, 백신접종 등을 위한 최적의 인력, 장비, 물자 등 대응자원을 확보하고 농장방문 및 이동 통제, 위기경보 등을 매뉴얼화된 지령체계 및 SNS를 통해 긴급 전파
- ※ 구제역 상황시에는 농장방문정보 뿐 아니라 차량의 이동경로, 농장세대원, 농장근로자 등에 대한 개인정보 수집 및 활용수준을 높임

<데이터 분석기반의 구제역 방역사례>



다. 맞춤형 복지 실현

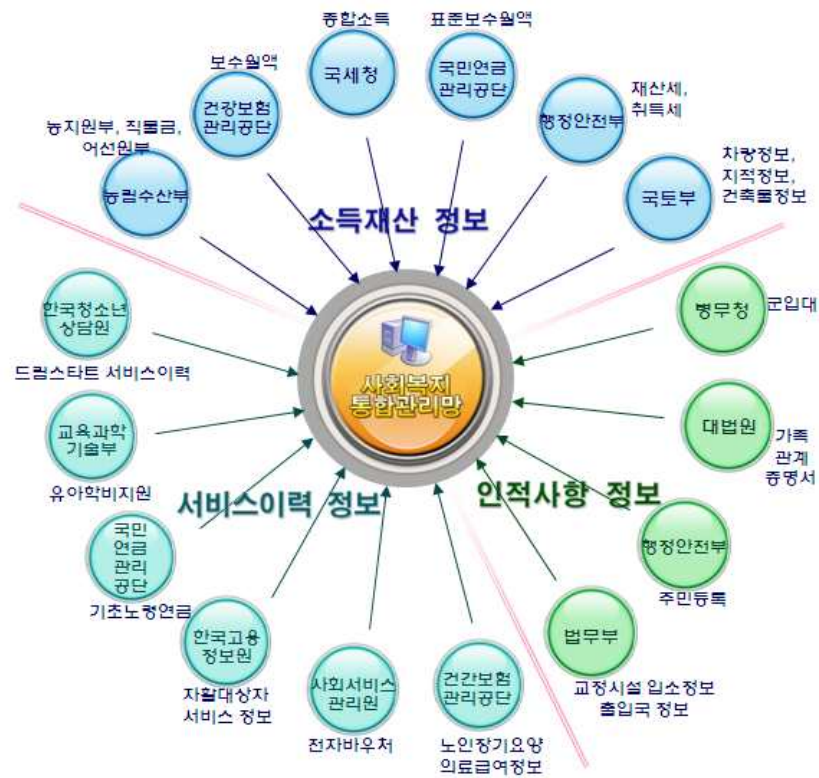
□ 개요

- 다양한 복지 정보 수집을 통한 맞춤형 복지 서비스를 제공

□ 현황

- 사회복지통합관리망 구축으로 복지사업별 기준에 맞는 적정 수급자 관리에 필요한 소득, 재산, 인적사항 등 219종의 정보를 행안부, 국세청 등 27개 기관으로부터 수집하여 자격 및 중복수급 여부 판단에 활용

<사회복지통합관리망 정보 수집 네트워크>

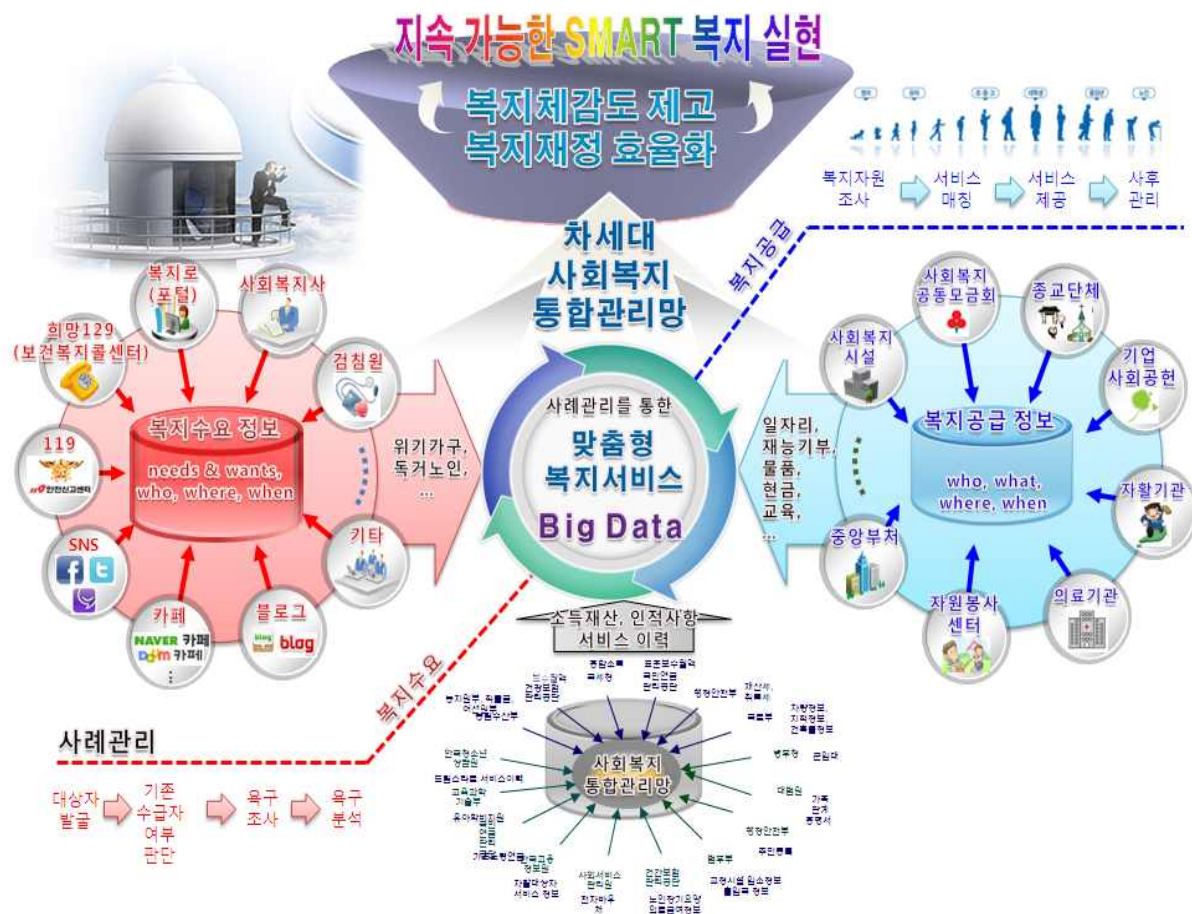


- 사회복지통합관리망을 통해 적정 수급자를 관리하여 금년 8월말 기준, 3,847억 원의 복지재정 누수를 방지
- 그러나 현재 시스템은 적정 수급자 관리에 필요한 정보만 갖고 있어 복지사각지대 발굴 및 맞춤형 복지서비스 제공에 한계

□ 빅데이터 활용 시나리오

- 사회복지사, 각종 포털, 상담센터 등 복지수요 정보와 사회복지시설, 종교 단체, 사회복지공동모금회 등 복지공급 정보를 수집하여 생애주기별, 복지 서비스 유형별 필요로 하는 맞춤형 복지서비스 제공
- 수집된 복지수요 정보를 활용하여 잠재적 복지 대상자를 발굴하고, 욕구 조사 및 분석을 통하여 필요로 하는 복지 서비스가 무엇인지를 식별
- 복지자원을 조사하여 복지 수요와 매칭시켜 맞춤형 서비스를 제공해주고, 서비스 결과에 대한 분석을 통하여 복지수요와 복지공급을 예측

<차세대 사회복지통합관리망 정보 수집 네트워크>



라. 물가 관리

□ 개요

- 실시간 생산·소비 물류 정보 모니터링을 통해 물가 안정 대책을 수립

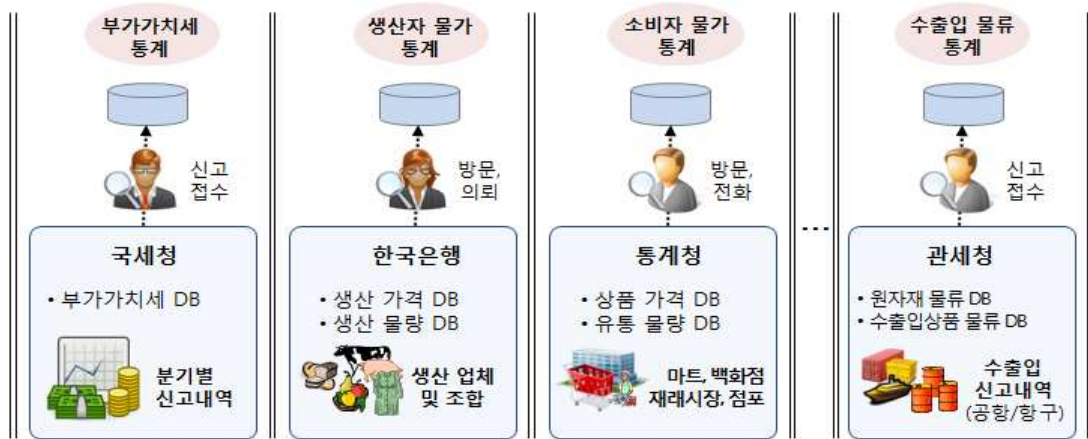
□ 현황

- 생산·소비·물류 정보를 부처별로 인력에 의존하는 표본 조사 기반으로 산출하고 있어, 부처간 정보 공유가 미흡하고 조사주기가 길어 신속한 대응에 한계
 - 기관별로 특정 물가관리 대상품목을 선정하고 정기적으로 오프라인 방식(방문, 전화, 서면 등)을 통해 물가지수를 조사·발표
 - ※ 물가지수 현황 파악 및 경제 상황 관리를 위한 거시적 지표로만 활용되는 실정
 - ※ 온라인 조사 및 입력 방식 활용이 낮아 비용 효과적이지 못하고 물가지수 값 산출 주기가 길어 실시간적 활용이 어려움
 - ※ 통계 방식의 표본선정 오류나 대체효과에 의한 왜곡으로 공시 물가와 체감 물가와의 괴리가 큼 (장바구니 물가 불일치)
 - 관세청의 수출입 물동량, 국세청의 생산유통 물동량, 공정거래위원회의 거래 물류정보 등은 개개 기관별로 신고자의 신고 내역으로만 물류 정보를 파악하므로 잠재적 오류 또는 불성실 신고 가능성 상존

<기관별 물류관리 방법>

기 관	항 목	주 기	대 상	방 법
통계청	소비자 물가	매월	•상품 및 서비스 (489개 품목) •전국 37개 주요도시 점포	•월1~3회 방문조사 (일부 품목은 전화 조사)
한국은행	생산자 물가	매월	•상품 및 서비스 (884개 품목)	•전화/이메일/팩스/방문조사 •관련 업체 및 조합 의뢰
관세청	물류 정보	수시	•관세기반 수출입 물류 및 가격	•단위 신고내역기반 세금 징수 (분기별 또는 단위 신고)
국세청		분기	•부가가치세기반 물동량 및 가격	
공정거래 위원회	물류 정보	수시	•불공정거래, 하도급, 부당표시, 전자상거래, 국내외 카르텔 등	•공정위 수시조사 및 민원

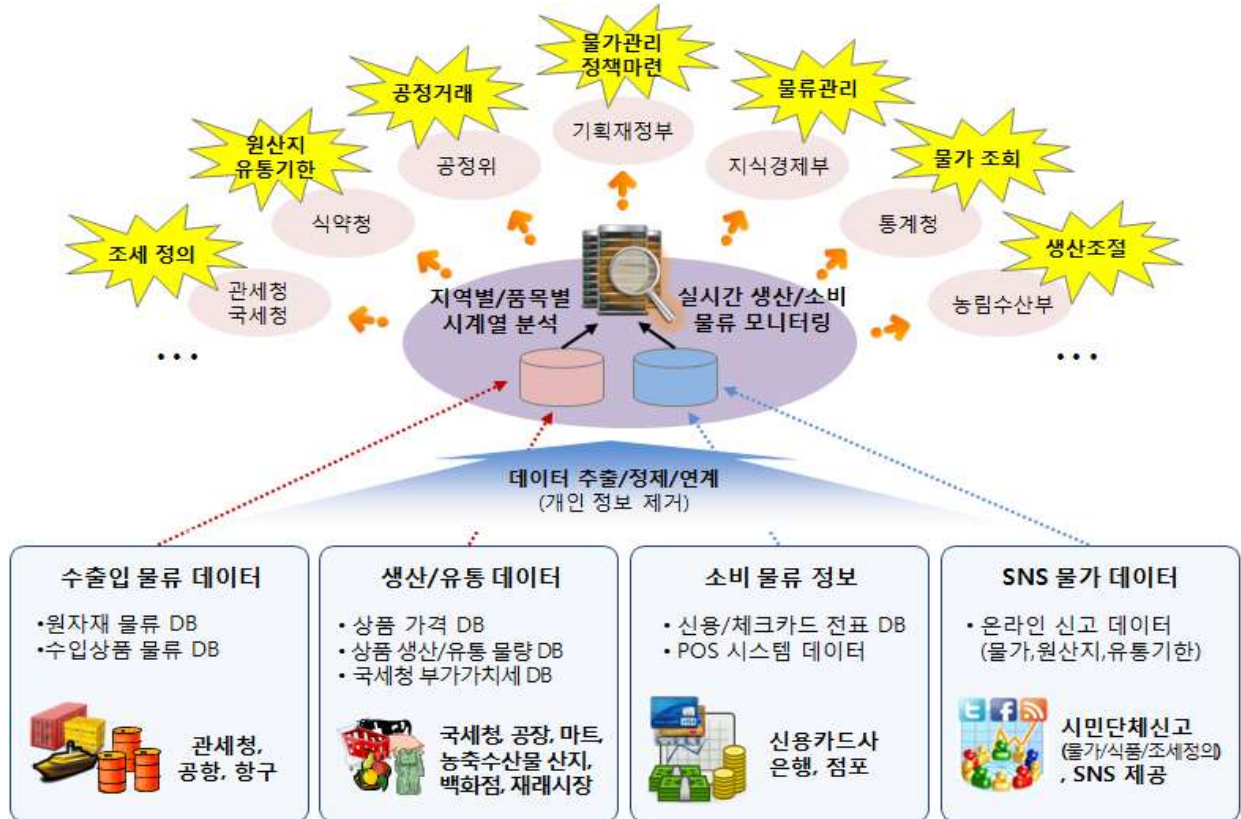
<현행 물류 관리>



□ 빅데이터 활용 시나리오

- 생산자부터 소비자까지의 가격 및 물동량 정보를 대용량 데이터 베이스 기술을 활용하여 물류 유통 전체 단계에 대해 전수 조사 기반 연계 분석을 수행
 - ※ 관련 부처 : 기획재정부, 지식경제부, 관세청, 식약청, 농림수산부, 공정거래위원회, 통계청, 국세청, 지자체 등
 - ※ 생산/소비/물류 모니터링 빅데이터 : 생산자 가격 및 물량, 관세 데이터 (수출입 물품 데이터), 시장 물가 데이터 (전국 마트, 재래시장, 백화점, 경매 POS 시스템 데이터 등), 소비물류정보(신용/체크카드 전표 DB), SNS제공정보(시민단체, 개인)
- 지역·품목별 생산자·소비자 가격 변동을 실시간으로 제공하여 생산 및 수입 규모를 조절하고, 물가 안정화를 위한 거시적 대책 수립 등에 활용
- 지역·품목별 물류관리를 통해 유통 구조를 최적화 및 물류 재분배로 물가 안정에 기여
 - ※ 농축수산물의 원산지 관리, 재고 관리, 수입 또는 생산 물량 사전경고 및 예측에 활용
- 실시간으로 불공정거래 및 불법물량을 탐지하여 세수 확대 및 조세 투명성 제고
- 데이터 개방을 통해 민간 분석을 활성화하고 지식을 축적하여 상품 기획, 생산 및 수입 계획 등에 활용

<융합지식기반 생산/소비 물류 정보 모니터링>



< 기관별 융합지식기반 생산/소비 물류 정보 활용 예시 >

기 관	활용 방법 및 예시
지식경제부	<ul style="list-style-type: none"> • 물가 안정화를 위한 거시적 대책 마련 (대체상품, 연관 상품 분석) • 물가 차이에 따른 생산지 재분배 또는 유통 구조 재구성 • 물가정책 투명성 확보 (장바구니 물가 반영)
기획재정부	<ul style="list-style-type: none"> • 복지 지원 금액의 현실 물가 반영 • 공공 상품별 지역 물가 지표로 세분화 활용
식품의약품안전청	<ul style="list-style-type: none"> • 농수산물 원산지 관리 • 식품 유통기한 관리 (불량식품 적발)
농림수산부	<ul style="list-style-type: none"> • 물가 예측을 통해 특정 품목의 생산 및 수입 규모 조정 • 농산물 파동 예방 및 사전 경고
공정거래위원회	<ul style="list-style-type: none"> • 공정거래 관리 (시간 담합 인상 탐지) • 불법 물동량 파악 (가짜 휘발유 물량 파악 등)
관세청/국세청	<ul style="list-style-type: none"> • 조세 정의 및 세원 확보 (부가가치세)
통계청	<ul style="list-style-type: none"> • 지역별, 품목별 소비자 가격 변동 실시간 제공 (휘발유 값 검색 사이트 등)
민간기업	<ul style="list-style-type: none"> • 품목별 유통비용 분석 및 재고관리

마. 과학기술 · 의료 선진화

□ 개요

- DNA·의료 데이터 공유 및 활용 촉진으로 개인 맞춤형 의료시대 실현

□ 현황

- 국가생명연구자원정보센터는 국가 차원에서 DNA 정보를 관리하고 있으나 데이터의 폭증으로 효과적인 활용에 한계
 - ※ 바이오 데이터 중 DNA 정보는 2003년 10GB에서 2010년 10^5 GB로, 2020년에는 10^9 GB로 증가할 것으로 예상(Nature Biotechnology, Vol. 28, 2010)
 - ※ 특히 차세대염기서열 해독장치 개발 및 기기 가격하락으로 저비용으로 대량의 인간 유전체 해독이 가능해짐
- 선진국은 DNA데이터의 중요성을 인식하여, 국가 차원에서 정책을 수립하고 통합관리체계를 구축
 - ※ NCBI(National Center for Biotechnology Information)는 미국 국립보건원(National Institute of Health)산하의 DNA 데이터 통합관리센터로서 유전자의 검색, 분석, 다운로드 기능을 제공(근거법 Public law 100-007)
 - ※ 1000 Genomes Data Project : 미국 정부의 지원으로 DNA 데이터 분석을 통하여 인간의 유전적 다양성과 질병과의 상관관계 연구(Univ. of Harvard, MIT, Harvard Medical School, Dana-farber Cancer Institute, Massachusetts General Hospital 등 '09년)
- 국내 DNA 데이터의 보존·활용은 선진국의 약 1/100 수준(KOBIC, 2010)
- '08년부터 보건복지부에서 한국인체자원은행사업을 통해 약 36만명의 인체자원을 확보하였으나 선진국에 비해 정보수집과 활용실적이 부족
 - ※ 전국 16개 병원에서 인체자원정보(혈액, 조직, 임상정보 등)를 수집하여 인체자원중앙은행에서 관리하고 있으나 DNA데이터 확보가 미흡하고 병원간 협력과 활용체계가 부족

□ 빅데이터 활용 시나리오

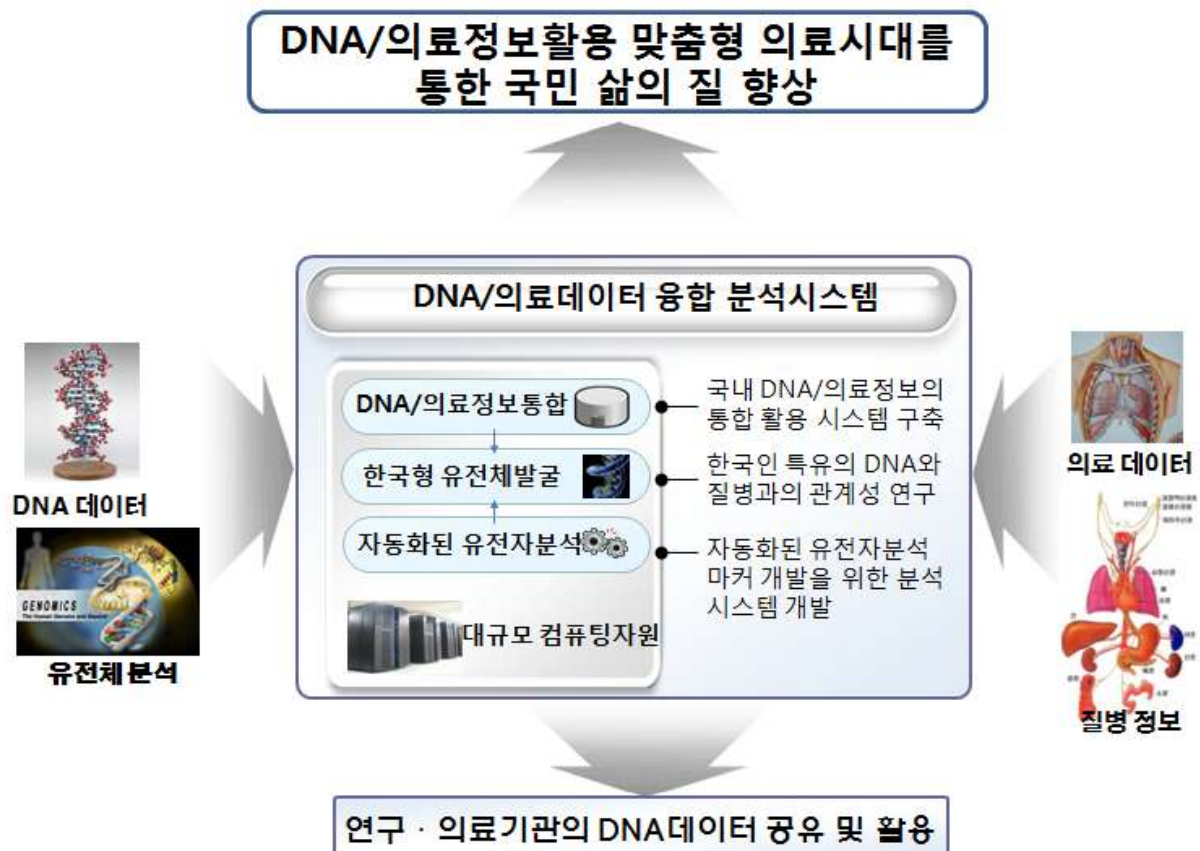
- 국가 차원의 체계적인 DNA 데이터 관리체계를 구축하여 DNA와 의료 데이터의 연계 시스템을 구축
- DNA데이터와 다양한 의료정보를 융합을 통해 개인별 질병 진단·치료 등 개인 맞춤형 의료를 실현

- ※ 개인 맞춤형 의료를 통하여 암, 당뇨, 비만, 고혈압 등 100여개 질환의 유전적인 위험도를 예측하고 사전 예방할 수 있음
- ※ 향후 개인별로 유전체정보를 획득 했을 시 이를 분석할 수 있도록 한국인의 유전체 분석시스템이 요구됨

○ 데이터 수집 및 활용방안(예)

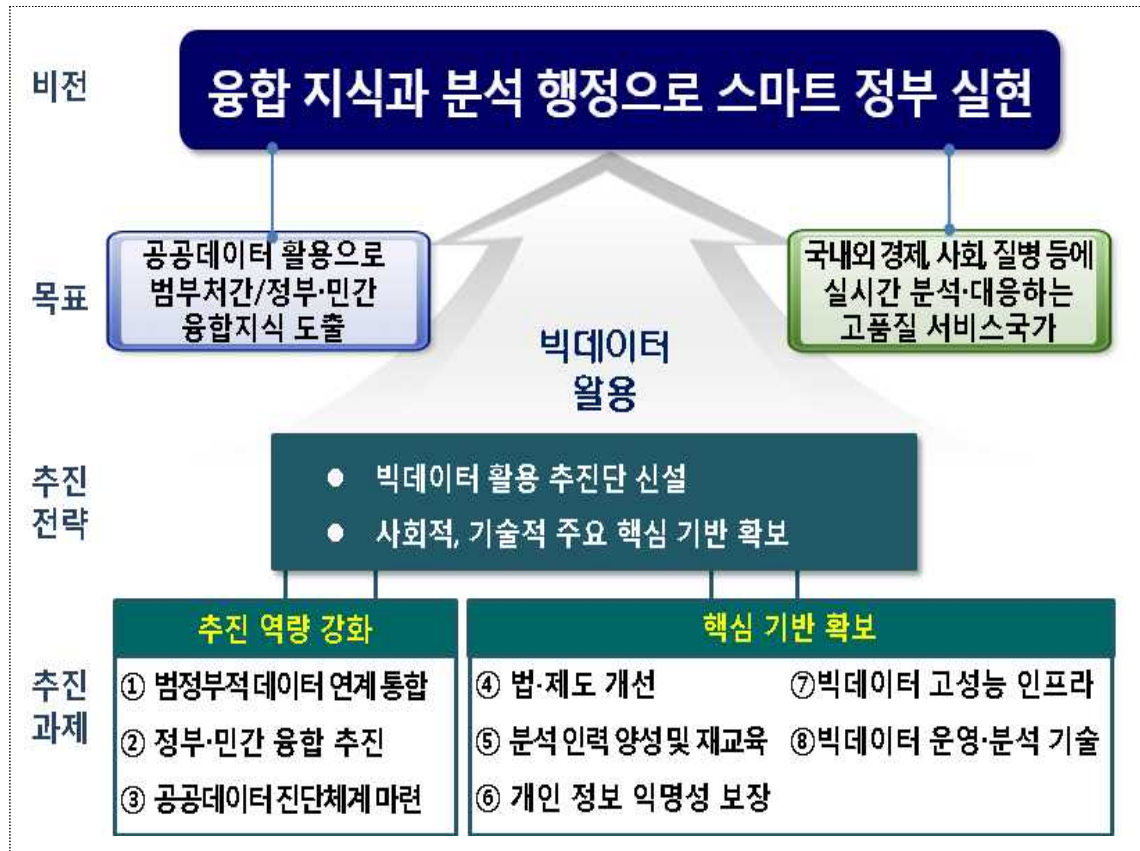
- 1) 정부출연금으로 진행된 연구결과(DNA 정보, 유전자/단백질 정보 등)의 공유를 활성화 할 수 있는 제도 마련
- 2) 개인연구자, 병원의 자발적 데이터 기부
- 3) 연구·의료기관의 데이터 공유체제를 구축하여 한국인의 DNA·의료정보의 활용시스템 구축

- ※ 선진국은 이미 공공기관을 중심으로 TCGA, ICGC 등이 설립되어 암 유전체의 대용량 DNA데이터를 확보하여 공유·활용함
 - TCGA(The Cancer Genome Atlas), ICGC(Int'l Cancer Genome Consortium)
 - 1명의 유전체 분석에 소요되는 비용과 시간변화 :
2003년(\$30억, 13년) → 2010년(\$5000, 1주) → 2020년(\$100, 1시간)



IV 빅데이터 활용 추진 방안

1. 비전 및 목표



□ 추진전략

- 정부가 빅데이터 활용 및 인프라 기반조성을 선도적으로 추진하기 위한 빅데이터 활용추진단 신설
 - 부처 데이터의 연계/협력채널 구축, 개인정보보호, 데이터 품질관리, 지능형 분석 등의 업무를 총괄하고, 다부처 분석조직을 총괄
- 빅데이터 국가 경쟁력 함양을 위한 국가적, 사회적 기반을 정비하고, 산·학·연 협력의 산업원천 기술 개발로 빅데이터 주요 핵심 기반 확보

2. 세부 추진 과제

가. 추진 역량 강화

① 범정부적 데이터 연계·분석체계 구축

- 기존 행정정보공동이용센터 등을 확대하여 부처 및 공공기관의 정보를 지속적으로 연계·수집할 수 있는 체계 수립
 - ※ (주요기능) 기관별 보유하고 있는 정형화된 DB 및 비정형화된 DB에 대한 시계열적 축적 및 관리, 추출 및 변환(ETL), 정제, 융합 관리 등
- 행정업무 수행관련 시스템들을 연계하여 범정부적 차원에서 업무와 지식을 융합·분석할 수 있는 지식행정 체계 구축
- 예측기반의 국정 운영 혁신을 위한 국가 전반의 데이터 수집·분석 체계 마련
 - 다양한 사회 현상을 데이터 기반으로 분석하여 정확한 정책을 수립·이행
 - 국내외 경제·사회·과학기술 관련 주요 예측치, 전망, 사회변화 트렌드 등 데이터 축적 및 분석 시스템 구축
 - ※ 시점별(예: 2015년, 2020년, 2025년, 2030년 등), 영역별(경제, 사회, 기술, 산업, 환경 등), 지역별(글로벌/국내 등) 등으로 정보 탐색

② 정부·민간 데이터 융합 추진

- 소셜 미디어, 포털 데이터 등과 같은 민간 데이터와 공공 데이터의 연계·활용을 위한 체계 및 기술 확립
- 산·학 등이 보유한 각종 정보를 범국가적으로 연계·수집할 수 있도록 저장소 구축·운영
 - 각 기관별 소셜미디어 정보, 각종 보고서, 정책, 전자화문서 등 주요 비정형 데이터 수집 및 가공
- 공공 데이터의 단계적 개방으로 민간의 가치창출 및 기업 활동 강화 지원
 - 원시 데이터, 분석 데이터, 가시화 단계로 구분하여 데이터 개방 수준을 상세화하고, 각 수준별로 데이터 공개 및 활용

- 공공 데이터를 활용한 중소기업 및 개인 창업자 등에 비즈니스 기회 제공을 위한 다양한 사례 제시
- 민간 기업의 공공 데이터 접근에 따른 ‘데이터 라이선스’ 규약을 신설하여 데이터 활용의 공익성을 확보

③ 공공데이터 진단체계 구축

- 공공데이터의 효율적 관리를 위한 범정부적 MDM 체계 구축
 - ※ MDM(Master Data Management): 다수의 시스템에서 활용되고 중앙에서 통제·관리되어야 하는 표준 참조데이터
- 각 요소별, 영역별 등으로 다양한 분류체계를 통한 정보 접근 및 키워드 검색 및 분석 등이 가능한 시스템 구축
- 범부처 데이터 공동 활용을 위한 품질관리 기준 및 체계 마련
 - 국가 차원의 공공정보 품질 관리활동의 선순환 환류체계를 평가하기 위한 품질성과 평가모델 구축
 - 공유서비스 고도화를 위한 공공데이터 품질관리 거버넌스 체계 확보 및 인식제고, 활동 참여 등 교육 지원방안 마련
- 기 구축 데이터에 대한 범정부 차원의 집중 품질정비 사업 추진
 - 불일치, 낮은 품질, 현행화 지연 등으로 인해 국가적 데이터 활용에 저해가 되는 데이터에 대한 품질 보정활동 추진
 - ※ 범정부 대상으로 데이터 품질 현황조사 및 대책수립을 위한 품질진단을 수행하여 품질정비 활동의 실효성 증대를 도모

나. 핵심 기반 확보

④ 법·제도 개선

- 범국가적 빅데이터 활용 추진 기본계획 수립 및 관련 법령 개정
 - 빅데이터 활용을 위한 핵심전략 및 비전, 추진과제, 범부처 추진체계, 중장기 로드맵 등을 포함한 '융합지식기반구축 기본계획' 수립
 - 빅데이터 활성화 전문조직(빅데이터 활용 추진단)의 설립 등 빅데이터 추진 및 활성화를 위한 법적 근거 마련
 - 각 부처 및 분야의 정보의 연계·분석 활성화를 위한 관련 법제도 정비
- 공공부문 빅데이터 분석 서비스 활용 촉진 방안 및 성과관리 체계 확립
 - 기관별 정책수립, 현업 업무수행, 업무평가 등에 빅데이터 분석 서비스 활용 방안을 마련하고, 빅데이터 기반으로 업무프로세스를 재설계
 - 빅데이터 분석 등을 활용하여 정부업무 개선에 대한 기관별 지속적인 모니터링, 보고체계 등 구성

수 분석 인력 양성 및 재교육

- 빅데이터 시대 新수요에 대응한 빅데이터 분석 전문인력 양성
 - 데이터의 관리자와 전문 분석가, 기술인력 등 전문인력 양성과 빅데이터 관련 신규 일자리 창출 추진
 - 정보분석 전문인력 교육 및 정보분석사 자격증 운영 등 시행
 - ※ 미국의 경우 2018년에 분석 전문가가 44만~49만명 정도 필요하나, 실제 졸업생은 18만명 정도로 추정(매킨지, 2011)
- 빅데이터 활용 역량 강화를 위한 다양한 재교육 프로그램 제공
 - 빅데이터를 활용한 다양한 우수사례 발굴, 기업 업무별 적용 지원, 분석기법 교육 등으로 공공·민간부문의 빅데이터 활용 역량 강화

⑥ 개인정보 익명성 보장체계 확립

- 안전하고 신뢰할 수 있는 공공데이터 공개 및 활용을 위해 개인정보와 프라이버시 보호 관련 기본 원칙을 체계화
 - 공공데이터 활용에 따른 국민 불안 해소를 위해 공공데이터 활용시 개인정보침해 최소화 장치 마련
 - ※ 목적명확화의 원칙, 수집제한의 원칙 등의 개인정보보호법상의 원칙은 물론 빅데이터의 수집·이용·제공 등에 있어서 개인 식별 정보의 변환 및 삭제의 원칙, 데이터 원본보다는 집계(agggregation) 값의 활용 원칙 등 마련
 - 민감한 개인정보에 대해 정보보안 수준별 권한 및 관리체계를 명확히 제시
- 빅데이터의 안전한 공유와 유통을 위한 강화된 보안대책 마련
 - 불필요한 데이터 사용을 원천적으로 차단하고, 빅데이터 불법 수집 및 오남용 방지를 위한 처벌규정 도입 및 보안성 모니터링체계 상시화
 - ※ 빅데이터 센터 내 전문기술과 경험을 갖춘 보안전담조직의 설치·운영
- 데이터의 개방·공유·활용에 따른 정부 공공데이터 활용 가이드라인 마련
 - 개인정보가 노출되지 않도록 개인식별 속성 제거, 암호화 등에 관한 원칙 수립
 - 소셜 데이터를 연계하여 개방·공유·활용 등의 경우는 원본 데이터, 위치정보 등에 대한 비저장 및 개인 비식별 등의 원칙 마련

⑦ 빅데이터 관련 인프라 기술 개발

- 외산 솔루션의 국내 빅데이터 시장 잠식을 방지하기 위해 대용량 분산 시스템 관련 핵심 기술 확보
 - ※ 대용량 정형 데이터를 위한 RDBMS (SQL) 기술과 대용량 실시간 스트림 데이터를 위한 비정형 분산DB (NoSQL) 기술
 - ※ 비정형 고속 데이터 처리를 위한 메모리 기반 데이터베이스 기술
 - ※ 비정형 대용량 스트림 데이터의 분산 병렬 처리를 위한 클라우드 컴퓨팅 기반 기술
- 개인의 익명성 보장 등 빅데이터 보안 기술 확보
 - ※ 개인 프라이버시 보호 지침에 따른 개인 정보의 대체, 은닉, 삭제 등 정보의 익명화
 - ※ 데이터 연계 시 보안성을 확보하기 위한 정책 기반의 프레임워크

○ 부처간, 공공·민간 데이터간 산재된 데이터 간의 연계·공유를 위한 융합·분석 프레임워크 기술 개발

※ 예) 노동·교육·복지 등과 같이 관련된 분야의 데이터 모델간의 연계를 통한 지식 융합

- 공공 서비스별 특화 서비스 제공을 위한 분석 기술 제공

※ 예) 기상 데이터와 소셜 웹 상의 독감데이터의 상관관계 분석, 고객들의 소비 패턴과 의료/바이오 데이터 상관관계 등

⑧ 빅데이터 운영·분석 기술 개발

○ 빅데이터 국내 기술 경쟁력 강화를 위한 핵심 기술 국산화

- 대용량 데이터의 수집, 해석, 정보 추출, 정보 분석, 정보 저장과 관리, 정보 검색과 질의, 가시화 등 빅데이터 관련 핵심 기술의 국산화

※ 핵심기술들은 마이크로소프트, IBM, 오라클, SAS 등 글로벌 대표 기업들이 독점

- 비정형 정보의 효과적인 분석을 위한 고도화된 최신 분석 요소 기술

※ 한국어 및 다국어 언어 처리, 텍스트 마이닝, 이미지 및 영상 분석, 음향 분석, 소셜 네트워크 분석, 의미적 데이터 마이닝 등과 같은 다양한 비정형 정보에 대한 분석 기술 개발과 적용을 통해 사실 간의 연관성 추적과 미래 예측이 가능

○ 급속도로 성장하는 소셜 미디어 환경에 대응하기 위한 소셜 스트림 데이터의 관리 및 분석 플랫폼 구축

- 웹 상의 다양한 소셜 스트림 데이터에 대한 즉시적인 취합, 분석, 해석, 관찰 및 정보 추출을 위한 실시간, 대용량 분산 프레임워크 개발

※ 다양한 온라인 서비스와 소셜 미디어 서비스의 콘텐츠 스트림, 관련 메타데이터, 사용자의 행위 기록 등이 대상

- 효과적인 빅데이터 관리 체계 구축을 위한 빅데이터의 전주기 관리, 데이터 중복 제거 및 연결 기술 개발

○ 도메인별 빅데이터 운영 기술 및 사용자 중심의 정보 시각화 기술 개발

- 공공 데이터의 분석 결과와 소셜 미디어의 분석 결과를 융합하여 도메인별로 특화된 지식을 창출하는 지식 융합·운영 기술 개발

- 분석 결과의 효과적 활용을 위한 사용자 친화적 정보 시각화 기술 개발

3. 향후 계획

<추진 일정>

추진내용	'12	'13	'14	'15	'16	비고
1. 추진 역량 강화						
① 범정부적 데이터 연계 통합						
② 정부민간 융합 추진						
③ 공공데이터 진단체계 마련						
2. 핵심 기반 확보						
④ 법제도 개선						
⑤ 분석인력 양성 및 재교육						
⑥ 개인 정보 익명성 보장						
⑦ 빅데이터 고성능 인프라						
⑧ 빅데이터 운영·분석 기술						

※ 소요 예산은 관계부처 및 기재부와의 협의과정을 거쳐 추후 확정