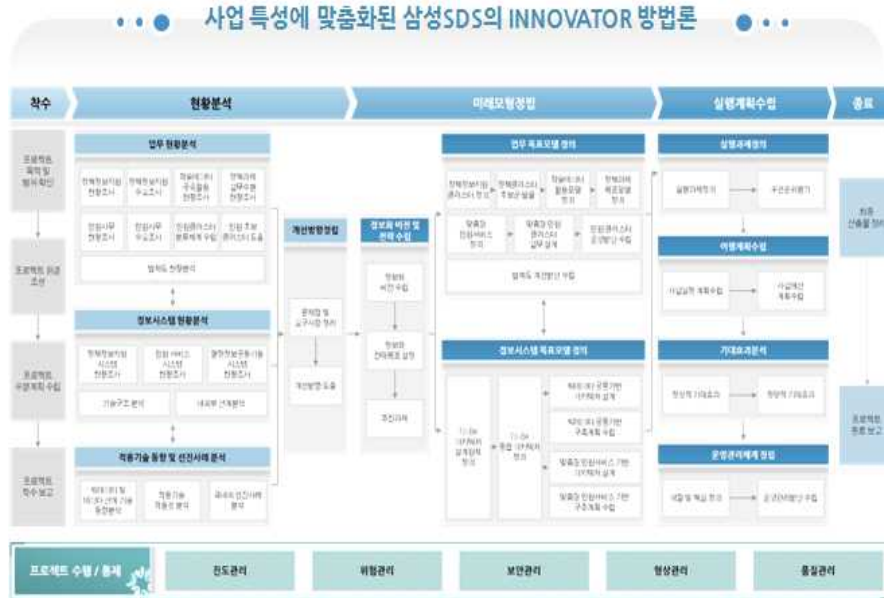


제1절. 적용방법론

1. 방법론 적용내용

『빅데이터 공통기반 마련 및 활용을 위한 BPR/ISP』 사업의 추진을 위해 삼성SDS가 보유하고 있는 방법론인 INNOVATOR 방법론을 근간으로 본 사업수행에 적합하도록 맞춤형된 컨설팅 방법론을 적용함



2. 사업 추진 절차

| 단계명 (phase) | 세그먼트명 (segment) | 단위업무명 (task) | 수행업무 | 산출물 |
|-------------|-----------------|--------------|--|----------|
| 프로젝트 착수 | 계획 수립 | 사업수행 계획 수립 | <ul style="list-style-type: none"> 프로젝트 범위, 추진방법/절차/일정 정의 프로젝트 관리 및 통제 방법, 품질보증활동, 인수기준 정의 추진조직 및 역할 정의 | 사업수행 계획서 |
| 환경분석 | 환경분석 | 일반환경 분석 | 본 사업에 대한 정책, 경제, 사회 환경 분석 | 환경분석서 |

| 단계명 (phase) | 세그먼트명 (segment) | 단위업무명 (task) | 수행업무 | 산출물 |
|-------------------|--------------------------|-------------------|--|-------|
| | | 기술환경 분석 | <ul style="list-style-type: none"> IT기술환경 변화 조사 분석 대상 기술 정의 대상 정보기술 동향 조사 요소기술 적용성 검토 | |
| | | 선진사례 분석 | 빅데이터 활용 선진사례 분석 | |
| | | 환경분석 종합 | 환경분석 시사점 종합 | |
| 현황분석 | 빅데이터 공통기반 및 활용과제 업무현황 분석 | 개념 정립 및 구성요소 도출 | 빅데이터 공통기반 정의 | 현황분석서 |
| | | 클러스터 분류 방안 수립 | <ul style="list-style-type: none"> 정책정보 현황분석 정책 클러스터 분류기준 정의 | |
| | | 정책과제 업무현황 분석 | <ul style="list-style-type: none"> 범죄 예측 정책과제 업무현황 분석 재난 대응 정책과제 업무현황 분석 민원 환류 정책과제 업무현황 분석 교통사고 개선 정책과제 업무현황 분석 국가 미래전략 및 위기 대응전략 수립 | |
| | | 학술데이터 공유/활용 방안 수립 | <ul style="list-style-type: none"> 학술데이터 공유/활용 현황 분석 해외사례 분석 | |
| | | 기반 모델 개념 정립 | <ul style="list-style-type: none"> 기반 모델 구성요소 도출 구성요소 정의 업무 모델 개념 정립 | |
| 맞춤형 민원서비스 업무현황 분석 | | 민원사무 현황 분석 | <ul style="list-style-type: none"> 민원사무 현황 조사 행정서비스 분류체계 현황 조사 수요자 관점 서비스 분류 | |
| | | 주제별 클러스터 분류방안 수립 | <ul style="list-style-type: none"> 클러스터 정의 민원서비스지원 클러스터 분류 서비스 수요조사 | |
| | | 서비스 수요조사 | 맞춤형 민원서비스 수요조사 | |
| | | 정책수립지 | 응용시스템 | |

| 단계명 (phase) | 세그먼트명 (segment) | 단위업무명 (task) | 수행업무 | 산출물 |
|-------------|------------------------|---------------------------|---|----------|
| | 원 및 맞춤형 민원서비스 정보화 현황분석 | 분석 | •민원 유관 시스템 현황 분석 | |
| | | 데이터 아키텍처 분석 | •정책정보 유관 데이터 현황 분석 •민원 유관 데이터 현황 분석 | |
| | | 인프라 아키텍처 분석 | •정책정보 유관 인프라 현황 분석 •민원 유관 인프라 현황 분석 | |
| | | 정보관리체계 현황 분석 | •행정정보공동이용센터 정보관리체계 현황 •개인정보보호 관리체계 현황 | |
| | | 법제도 현황 분석 | •관련 법제도 현황 검토 | |
| | | 현황분석 종합 | •시사점 및 개선기회 종합 •개선 방향 도출 | |
| 미래모형 정립 | 빅데이터 공통기반 및 활용과제 업무 설계 | 정보화 비전 및 전략 수립 | •개선방향 검토 •정보화 비전 및 전략 수립 •정보화 과제 정의 •목표시스템 개념도 정의 | 미래모형 정의서 |
| | | 주제별 클러스터 분류방안 수립 및 후보군 도출 | •정책 클러스터 후보군 도출방안 수립 •정책 클러스터 관계 정의 •정책 클러스터 후보군 발굴 | |
| | | 공통기반 업무 설계 | •정책 단계별 정보수집 체계 정의 •정책 단계별 정보 활용 방안 정의 •공통업무 지원 모델 수립 •정책정보 지원 서비스 제공 방안 수립 | |
| | | 정책과제별 업무 설계 | •범죄 예측 정책과제 업무 설계 •재난 대응 정책과제 업무 설계 •민원 환류 정책과제 업무 설계 •교통사고 개선 정책과제 업무 설계 •국가 미래전략 및 위기 대응 전략 수립 정책과제 업무 설계 | |
| | | 학술데이터 공유 활용 방안 | •학술데이터 공유 및 활용 모델 정의 •모델 분석 및 평가 | |
| | | 맞춤형 민원 | •설계 목적 및 범위 정의 | |
| | | | | |

| 단계명 (phase) | 세그먼트명 (segment) | 단위업무명 (task) | 수행업무 | 산출물 |
|-------------|---------------------------------|---------------------------|--|----------|
| | 민원서비스 업무 설계 | 설계 방안 마련 | •업무 설계 Framework 정의 •업무 모델 개념 정립 •선정 클러스터별 지원 서비스 Pool 현황 분석 | |
| | | 목표 업무모델 설계 | •민원사무 기본사항 설계 •지원 서비스 기본사항 설계 •선·후행 관계성 설계 •제약/전제 조건 설계 •중복필터 설계 | |
| | | 클러스터 운영방안 마련 | •민원서비스 운영관리 현황 분석 •클러스터 운영관리 모델 설계 •클러스터 문제점 및 부작용 분석 | |
| 미래모형 정립 | 정책수립지원 및 맞춤형 민원서비스 정보화 구축 방안 수립 | 목표시스템 구축 모형 설계 | •빅데이터 구축 모형 설계 •맞춤형 민원 서비스 구축 모형 설계 | 미래모형 정의서 |
| | | 빅데이터 공통기반 및 정책과제 구축 방안 수립 | •응용 아키텍처 설계 •데이터 아키텍처 설계 | |
| | | 인프라 아키텍처 설계 | •범정부 클라우드 활용방안 수립 •인프라 구축 방안 수립 •보안 체계 방안 수립 | |
| | | 범정부 통합관리체계 수립 | •클러스터 통합 관리 방안 수립 •범국가적 협력체계 구성 및 운영 방안 수립 | |
| | | 법제도 정비방안 수립 | •빅데이터 공통기반을 위한 신규 법제도 제정 방안 | |
| 이행계획 수립 | 일정계획 수립 | 이행과제 정의 | •이행과제 도출 및 정의 •이행과제 우선순위 도출 | 실행계획서 |
| | | 로드맵 정의 | •로드맵 정의 •통합추진일정계획 수립 | |

| 단계명 (phase) | 세그먼트명 (segment) | 단위업무명 (task) | 수행업무 | 산출물 | |
|-------------|--------------------------------|--------------|--|-------|---|
| | | | •과제별 세부일정계획 수립 | | |
| | | | 소요예산 산정 기준 정의 | | •소요예산 산정 기준 정의 •과제별 소요예산 산정 |
| | | | 추진조직 정의 | | •추진조직 정의 •추진조직 역할 정의 |
| | | | 기대효과 산정 | | •기대효과 도출 프레임워크 정의 •과제별 기대효과 도출 •기대효과 종합 |
| | | | 향후 발전방안 수립 | | •향후 미래모형별 발전 방향 정의 |
| 프로젝트 종료 | 프로젝트 산출물 종합정리 종료보고회 준비 및 실시 | | •프로젝트 추진절차상에 작성된 산출물을 최종적으로 정리 •종료보고회 준비 및 실시 •검토된 산출물에 대한 사업 최고 책임자의 승인을 득함 | 종료보고서 | |

3. 산출물

| 제출산출물 | 업무구분 | 산출물내역 | 제출시기 |
|------------|-----------|---|--------|
| 착수산출물 | - | • 착수계(사업수행계획서) | 계약체결 후 |
| 최종산출물 | 환경 및 현황분석 | • 환경분석서 • 현황분석서 | 검사요청 시 |
| | 미래모형정립 | • 미래모형정립서 | |
| | 실행계획수립 | • 이행계획서 | |
| 사업추진 결과보고서 | - | • 사업추진결과보고서 - 사업개요 및 개발내용 - 운영계획 및 발전방향 | 사업완료 후 |

제2절. 개발내용

1. 사업환경 분석

가. 환경분석 개요

- 빅데이터 공통기반 마련 및 활용 체계 구축사업의 추진 필요성과 적용가능 기술을 제시하고, 사업을 추진하기 위한 방향성을 도출하기 위한 일반환경 기술환경, 선진사례를 분석하고 시사점을 도출함

[환경 분석 목적 및 범위]

| | |
|-----------|---|
| 분석목적 | • 빅데이터 공통기반 마련 및 활용 체계 구축 사업 추진 필요성, 적용가능기술, 사업추진 방향성 도출 |
| 분석범위 | 일반환경 분석 • 빅데이터 공통기반 마련 및 활용 체계 구축 사업의 필요성 및 타당성 제시를 위한 정책, 시장, 사회환경 분석 및 시사점 도출 |
| | 기술환경 분석 • 빅데이터, 데이터 연계통합, 클라우드 관련 기술동향 파악 및 적용 가능 기술 제시 |
| | 선진사례 분석 • 사업 추진 방향, 서비스, 빅데이터 관련 정보기술 적용 측면의 선진사례분석 및 시사점 도출 |
| 분석대상 및 방법 | • 분석대상 : 국가정보화 추진현황, 국가정보화 기본계획, 빅데이터 마스터플랜 등 • 분석방법 - 문헌조사 : 인터넷, 신문, 논문 등 관련 자료 - 인터뷰 : 중앙정부기관, 공공기관, 시도/시군구, 행정정보공동이용센터 |

나. 일반환경 분석

- 빅데이터 공통기반 및 활용 그리고 수혜자 맞춤형 행정 서비스 관련 정책, 경제, 사회 문화 환경을 분석하였으며, 분석결과는 아래와 같음

[일반환경 분석 시사점]

| | | |
|-------|------|---|
| 일반 환경 | 정책환경 | <ul style="list-style-type: none"> • 빅데이터 등 범국가적으로 고도화된 ICT 인프라 기반 체계 마련 필요 • 범정부 차원의 데이터 연계-통합, 빅데이터 활용과제 발굴 및 확산 등을 담당할 운영 및 관리체계 마련 필요 • 여러 부처에 편재되어 있는 정책정보의 탈 경계적 연계 및 활용체계 마련 필요 • 주민번호 등 민감한 개인정보가 노출되지 않기 위한 정보보호체계 기반 구축 필요 |
| | 경제환경 | <ul style="list-style-type: none"> • 지속적으로 성장하고 있는 빅데이터 시장('15년 169억달러, 연평균 40% 성장, ICT 시장성장률 7배)에 대비한 정책 및 데이터활용기반 마련 필요 • 해외 공공기관 빅데이터 도입 및 전략적 활용에 발맞추어 빅데이터 활용 기반 마련, 데이터 분석가 확보 등 빅데이터 기반의 우리나라 정책 수립 필요 |
| | 사회환경 | <ul style="list-style-type: none"> • 효과적인 정책 수립 지원을 위한 민간 SNS 데이터 활용체계 마련 필요 • 공개된 공공데이터를 한곳에서 쉽게 활용할 수 있는 기반체계 마련 필요 • 국가적 지원 서비스 정보에 상대적으로 취약한 다문화 가정과 외국인을 고려하여 미리 알려주는 채널형 서비스 제공 필요 • 제공자 위주의 일방적인 지원 서비스 제공이 아니라 지원 대상에 맞춘 입체적 지원 서비스 제공 필요 |

다. 기술환경 분석

1) 대용량 데이터 기술(빅데이터)

기존 데이터베이스 관리도구의 데이터 수집·저장·관리·분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 셋 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술



○ 데이터 수집

- 비정형 데이터 수집 : 데이터 발생원으로부터 안정적인 저장소로 저장하는 기능을 수행하며, 오픈 소스로는 Flume, Scribe, Chukwa 등이 있음
- 정형 데이터 수집 : 다양한 저장소에 대용량 데이터를 신속하게 전송할 수 있는 정형 데이터를 수집하는 기술로 Sqoop, Hiho 등과 같은 하둡 기반 기술과 기존 방식의 ETL, Web Robot, EAI, OPEN API 등이 있음

○ 데이터 저장

- 원본 데이터 저장 : 수집된 데이터를 안정적으로 저장하는 저장소이며, 비구조적 데이터 저장소로 Hadoop FileSystem, MogileFS 등과 같은 대용량 파일 저장소를 사용함
- 트랜잭션 데이터 저장 (NoSQL) : 원본 데이터를 실시간으로 저장, 조회 처리를 하기 위한 저장소이며, NoSQL(Clouddata, HBase, Cassandra) 등과 같은 구조적 저장소 또는 Katta, ElasticSearch 등과 같은 검색 엔진 기술을 활용함
- 트랜잭션 데이터 저장 (검색 엔진) : 통합색인 검색엔진과 메타 검색엔진, 분산 검색엔진 에이전트 등으로 나뉘어질 수 있으며, 하둡기반의 오픈소스로는 Katta, ElasticSearch 등이 있음
- 메모리 저장 : 데이터베이스를 메모리에 상주시켜 운영하는 DBMS로 디스크 접근 없이 메모리에 직접 접근하여 데이터 읽기 연산을 수행하며, 오픈 소스로는 Redis, Membase 등이 있고 하드웨어로는 메모리와 SSD 저장 기술이 있음

- 데이터 보안 (Dynamic data masking) : 실시간으로 민감한 정보를 동적으로 마스킹 하고, 민감한 정보에 액세스하는 사용자에게 차단, 감사 및 경고를 보내며, 동시에 데이터 기밀 유지 규정을 신속하게 준수하는 데이터 보안 기술임

○ 데이터 처리

- 배치 처리/분석 (MapReduce) : 전체 또는 부분 데이터에 대해 복잡하고 다양한 분석 수행 및 대용량 처리를 위해 분산, 병렬처리를 수행하고, 단순 텍스트 분석부터 그래 프 분석까지 다양한 분석 모델을 지원하며, 오픈소스로는 Hadoop MapReduce(Hive, Pig), Giraph, GoldenOrb 등이 있음
- 실시간 처리/분석 (Complex Event Processing) : 데이터 수집과 동시에 분석을 수행하고 복잡한 분석보다 count, sum 등 단순한 aggregation 연산 정도를 수행 하며, 별도의 분석 클러스터를 구성하는 S4, Storm 등과 데이터 수집기에 탑재하 여 분석하는 Esper 등이 있음

○ 데이터 분석

- 텍스트 분석 : 비정형 텍스트 데이터에서 가치 있는 정보를 추출하는 기술이며 문서분류, 문서군집, 정보추출, 문서요약 기술분야로 구성됨
- 소셜네트워크 분석 : 클라우드 컴퓨팅 기술의 발전으로 대용량 데이터의 실시간 분석이 가능해지면서 비정형화된 소셜 데이터를 분석하여 다양한 분야의 비즈니스 에서 활용하고 있음
- 통계분석 (R) : 오픈소스 통계 프로그래밍 언어로 통계 분석에 필요한 강력한 기능을 제공하며, 이미 2백만명의 애널리스트들이 사용중임
- Machine Learning (Mahout) : 인공 지능의 한 분야로, 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야이며, Mahout은 아파치 오픈소스이고, Hadoop을 활용할 수 있는 Machine learning library임
- 예측분석(알고리즘) : 과거의 데이터나 사건으로부터 미래에 발생 가능한 상황이나 사건을 예측하여 선제적인 의사결정을 지원하는 분석 기술임
- Data Mining : 방대한 데이터에 숨겨져 있는 패턴(Pattern), 상호 관련성, 경향 등을 발견하고 규칙을 추론함으로써 유용한 의사결정정보를 찾는 프로세스로서, 비즈니스 문제에 대한 통찰을 발견하고 새로운 비즈니스 기회 창출 가능

○ 데이터 표현

- 시각화 : 분석 결과를 사용자에게 시각적으로 다양하게 제공하고 데이터의 진정한 가치를 알려주는 기술이며, 오픈소스로는 Prefuse, d3.js 등이 있음

2) 데이터 연계·통합 기술

- MDM(Master Data Management) : 기준데이터의 일관성과 정확성을 보장하고 활용 및

공유하기 위한 데이터 관리체제로 프로세스, 조직 등 비즈니스와 IT를 결합한 기술임

- 메타데이터 관리 : 메타데이터는 데이터 요소들에 대한 설명으로 데이터를 분류, 수집하고 속성에 대한 표현을 돕는 데이터의 데이터로 정보자원에 대한 통합과 분류를 지원할 수 있음
- DQM(Data Quality Management) : 운용 상의 비즈니스 프로세스와 결점을 확인하고 수정하기 위한 프로세스와 기술을 의미하며, 데이터 품질의 신뢰성을 확보하기 위해 활용됨
- 웹서비스(Web Service) : 온라인 협업 및 사용자간의 데이터 공유를 강조하여 사용자 참여, 경험 공유, 데이터의 중요성, 느슨하게 연결되어 웹서비스의 개념이 포함된 차세대 인터넷 기반 서비스임
- EAI(Enterprise Application Integration) : 비즈니스 프로세스를 중심으로 기관 내 또는 기관간 이질적인 어플리케이션과 비즈니스 프로세스를 통합함
- ESB(Enterprise Service Bus) : 서비스 연동환경에서의 Web Services 전용 미들웨어로서 연동 프로세스의 정의, 연계시스템 어댑터, 연동 시 오류보정, 부하분산 등의 기능을 통해 데이터의 연계를 가능하게 함
- ETL(Extract, Transform, Load) : 데이터 원천으로부터 필요한 데이터를 추출하여 데이터웨어하우스 또는 데이터마트에 입력하는 기능을 담당하는 기술로 품질 보장과 신뢰성 보장을 위해 필수적인 기술적 고려사항으로 데이터 웨어하우스 운영 과정에 중요한 역할을 수행함

3) 서비스 편의 향상 기술

- 모바일 : 무선환경을 이용하여 각종 정보를 시간, 장소에 관계없이 모바일 기기를 활용하여 실시간으로 사용자가 활용할 수 있도록 하는 서비스
- OPEN API(Application Programming Interface) : 수많은 개발자가 API를 바탕으로 개발한 어플리케이션의 서비스를 가능하게 하는 플랫폼이며, 오픈플랫폼 하에서 개발자들은 다양한 어플리케이션을 만들어 API 공급자인 Container에게 제공
- Web 2.0 : 다양한 분야의 개방형, 참여 기반 모델을 지향하는 웹기반의 서비스와 응용환경을 위해 필요한 기술요소

4) 정보자원 관리 및 이용 기술

- 클라우드 컴퓨팅 : 서로 다른 컴퓨팅 자원들을 가상화 기술로 연결하여 시간과 장소에 구애받지 않고 사용할 수 있는 환경을 제공함
- Semantic Web : 컴퓨터가 정보를 읽고 스스로 처리하여 이를 상황에 맞게 사람과 컴퓨터, 컴퓨터와 컴퓨터가 상호 협력할 수 있도록 돕는 기술
 - RDF(Resource Description Framework) : 메타데이터를 처리하기 위한 기초이며, 웹에서

기계가 이해할 수 있는 정보를 교환하는 어플리케이션간의 상호운용성을 제공함

- 온톨로지 :실세계에 존재하는 모든 개념들과 개념들의 속성, 그리고 개념들이 상호간에 의미적으로 어떻게 연결되어 있는지를 컴퓨터가 이해할 수 있도록 기술해 놓은 어휘집
- LOD (Linking Open Data) : 인터넷 상에서 거대한 시스템 간의 공유가 가능한 지식베이스를 구현하기 위한 기반적 기술로 등장했으며, 이러한 LOD는 공공기관에서의 정보 공개, 시맨틱 메시업 등과 함께 관심을 끌고 있음

라. 선진사례 분석

- 빅데이터 공통기반의 선진사례는 거버넌스, 공동인프라, 정책서비스를 추진 중에 있는 싱가포르, 미국, EU, 일본 국가 사례를 분석함

[선진사례 분석결과]

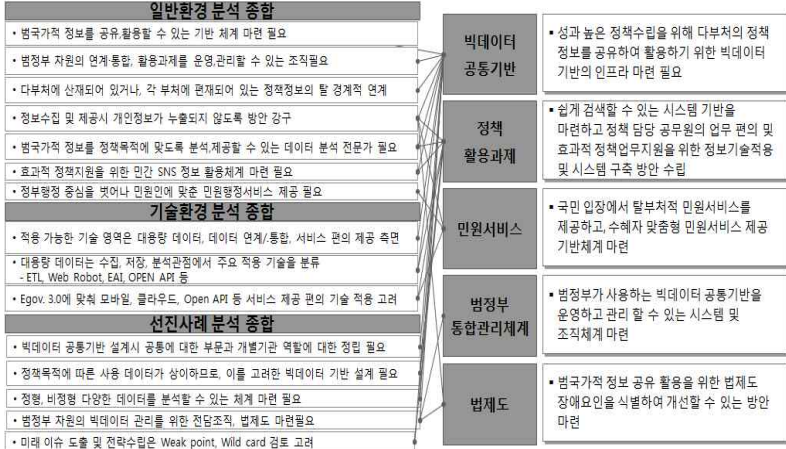
| | 싱가포르 | 미국 | EU | 일본 |
|----------|--|---|---|--|
| 추진사업 | 국가 위기관리를 위한 지능형 정책지원체계 확립 | 빅데이터 전략수립 및 법무 협력거버넌스 구축 | 미래예측을 위한 iknow 프로젝트 추진 | 빅데이터 활용으로 Active Japan 구현 |
| 거버넌스 | • 종리실 산하 국가안보조정국에서 | • 대통령실 내 과학기술 정책실 설치, 운영 | • EU 추진 | • 중무성 산하 빅데이터 활용 특별 부회 추진 운영 |
| 빅데이터 인프라 | • 정치, 경제 등 환경스캐닝 및 Emerging 이슈 발굴 - RAHC 실험 센터 운영 - 데이터 수집, 중앙저장 | • 빅데이터기반 R&D추진 - 국방 등 5개 기관 참여 - 빅데이터 활용 통합 인프라 구성되지 않음 | • 유럽 초 국가환경스캐닝 - 전세계웹사이트,저널탐구 - 전세계 약신호 분석 - 유럽전역과학기술스캐닝 | • 빅데이터 기본전략수립 - 민간 협업 촉진 - 혼다, 후지쯔 등 10개 민간기관 참여 |
| 정책 서비스 | • 국가 위기관리 시나리오 제공 | • 연방정부 데이터관리, 빅데이터 관련 인력 향상 | • 44개 정책 브리프 제시 • 분야별 전문가 보고서 제공 | • 빅데이터 추진과제 수립 |
| 변화방향 | <ul style="list-style-type: none"> > [거버넌스] 빅데이터는 범국가적 차원의 추진체계를 조직하여 정책을 기획하고 법제도를 정비함 > [공동인프라] 범부처가 참여하여 공동 활용할 수 있는 빅데이터 활용 기반의 공동인프라 구축이 필요함 > [정책서비스] 국가 미래전략 수립은 우리나라에 한정하지 않고, 전세계 환경으로 시야를 넓혀 데이터를 수집하며, 수집시 약신호(weak signal)에서 emerging 이슈를 도출하는 방향으로의 고려가 필요함 | | | |

* Weak signal(미래 발생가능성이 있는 사건(예, 15년 전기자 백만대 보급(미국), 빈번한 홍수(유럽))

마. 환경분석 종합

- 일반환경, 기술환경, 선진사례분석을 종합하고 추진방향을 제시함

[환경분석 종합]



2. 빅데이터 공통기반 및 활용과제 설계 방안

가. 현황분석

1) 현황분석 개요

- 빅데이터 공통기반 및 활용과제 방안 수립을 위한 현황분석은 빅데이터 공통기반 및 활용과제 업무, 빅데이터 공통기반 및 활용과제 관련 정보시스템 현황 분석을 통해 목표 업무 설계 및 정보화 방향성을 도출함

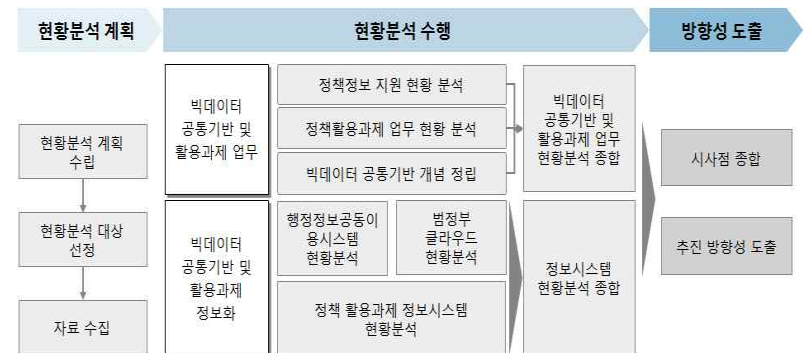
[현황분석 목적 및 범위]

| | | | | | |
|----------------------|---|---------------------|---|----------------------|--|
| 분석 목적 | <ul style="list-style-type: none"> • 빅데이터 공통기반 마련 및 활용과제 업무 설계 방향성 도출 • 정책수립지원 및 맞춤형 민원서비스 정보화 설계 방향성 도출 | | | | |
| 분석 범위 | <table border="1"> <tr> <td>빅데이터 공통기반 및 활용과제 업무</td> <td> <ul style="list-style-type: none"> • 빅데이터 공통기반 개념 정립 및 구성요소 도출 • 정책 정보 현황 분석을 통해 데이터 연계통합 및 활용도 제고 방안 도출 • 활용과제별 업무절차, 환경요인, 참조정보, 요구사항 도출 </td> </tr> <tr> <td>빅데이터 공통기반 및 활용과제 정보화</td> <td> <ul style="list-style-type: none"> • 행정정보공공이용시스템 운용기능, 데이터, 내외부 연계, 운영현황 • 정책 활용과제 관련 정보시스템 현황(범죄발생최소화, 주민참여형 교통사고 감소 관련시스템 등) • 법정부클라우드 추진 현황 및 계획 </td> </tr> </table> | 빅데이터 공통기반 및 활용과제 업무 | <ul style="list-style-type: none"> • 빅데이터 공통기반 개념 정립 및 구성요소 도출 • 정책 정보 현황 분석을 통해 데이터 연계통합 및 활용도 제고 방안 도출 • 활용과제별 업무절차, 환경요인, 참조정보, 요구사항 도출 | 빅데이터 공통기반 및 활용과제 정보화 | <ul style="list-style-type: none"> • 행정정보공공이용시스템 운용기능, 데이터, 내외부 연계, 운영현황 • 정책 활용과제 관련 정보시스템 현황(범죄발생최소화, 주민참여형 교통사고 감소 관련시스템 등) • 법정부클라우드 추진 현황 및 계획 |
| 빅데이터 공통기반 및 활용과제 업무 | <ul style="list-style-type: none"> • 빅데이터 공통기반 개념 정립 및 구성요소 도출 • 정책 정보 현황 분석을 통해 데이터 연계통합 및 활용도 제고 방안 도출 • 활용과제별 업무절차, 환경요인, 참조정보, 요구사항 도출 | | | | |
| 빅데이터 공통기반 및 활용과제 정보화 | <ul style="list-style-type: none"> • 행정정보공공이용시스템 운용기능, 데이터, 내외부 연계, 운영현황 • 정책 활용과제 관련 정보시스템 현황(범죄발생최소화, 주민참여형 교통사고 감소 관련시스템 등) • 법정부클라우드 추진 현황 및 계획 | | | | |
| 분석 방법 | <ul style="list-style-type: none"> • 문헌조사: 인터넷, 신문, 논문, 분석 대상 정보시스템 관련 문서 등 관련 자료 • 인터뷰 : 중앙정부기관, 공공기관, 시도/시군구, 행정정보공공이용센터, 정부통합전산센터 | | | | |

- 빅데이터 공통기반 및 활용과제 방안 수립을 위한 현황분석 Framework를 제시함

- 분석 Framework는 현황분석 계획, 분석수행, 방향성 도출 순으로 추진함

[현황 분석 Framework]



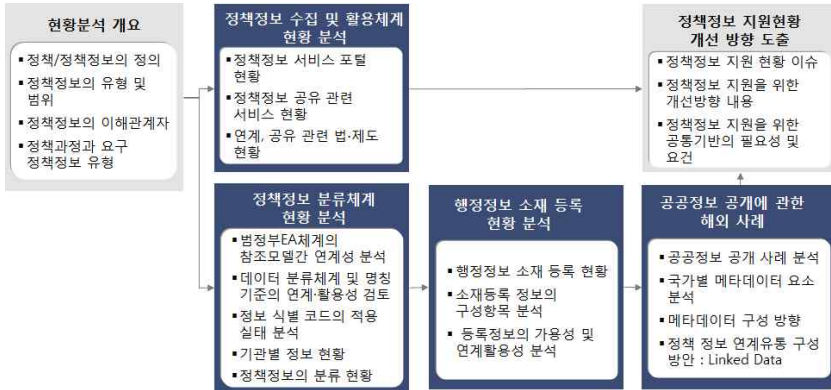
2) 빅데이터 공통기반 및 활용과제 업무 현황분석

가) 정책정보 지원 현황

(1) 정책정보 지원현황 분석 Approach

정책정보에 대한 분석, 정책정보 서비스 포털, 공유 서비스 현황, 범정부 EA 체계의 활용성 분석, 행정정보 소재등록 현황 및 해외 사례를 통해 현행 정책정보 지원 현황을 검토하고 이슈 및 개선방향을 도출함

[현황분석 Approach]



(2) 정책정보의 정의

정책결정자가 정책결정 과정에 필요로 하는 정보로 구성된 자료이며, 정책의 집행, 정책의 성과평가 등 정책 전 과정에서 항상 요구되는 자료임



* 일반적으로 정책결정은 정책과정에서 생성된 산출물 중심의 정책자료로 한정되는 경우가 많으므로 주로 문헌정보 형태로 존재하며, 존재하지 않더라도 정책정보를 생산하기 위한 기본 자료인 원시행정데이터도 정책정보에 포함시켜 고가함

(3) 정책정보의 유형 및 범위

정책정보는 크게 정부 간행물, 정책연구자료, 정책정보 웹문서, 정책정보관련 DB 등의 문헌정보와 통계 및 메타데이터와 같은 요약정보, 그리고 정책정보를 가공할 수 있는 원시행정데이터와 민간포함 비정형데이터로 구분할 수 있음

| 정책정보 유형 | 정의 | 예시 | |
|---------------------------------|---|---|---|
| 정형 | 원시행정데이터 | <ul style="list-style-type: none"> • 요약정보 형태의 정책정보를 만드는 데 원재료가 되는 행정데이터로 각 기관의 행정정보시스템 운영 과정에서 생성 • 그 자체로서는 정책정보로서의 가치가 없으나 가공, 분석을 통해 의미 있는 정책정보로 만들어질 수가 있음 | <ul style="list-style-type: none"> • 각종 인허가, 신고, 신분/자격확인, 사업계획 확인, 소유권 증명, 고종차리 등 민원행정 처리의 과정과 결과에서 생성되는 대장, 전표, 승인서, 증명서, 확인서 등의 원시정보로 구성 |
| | 가공데이터 | <ul style="list-style-type: none"> • 원시행정데이터 또는 원문 정보를 가공, 요약하여 도출한 정책정보 | <ul style="list-style-type: none"> • 정책통계정보, DW정보, 정책관련 메타데이터(원문요약, 원문위차) 등 |
| 비정형 | 정부부처 간행물 | <ul style="list-style-type: none"> • 정책을 입안하거나 정책을 집행하고 그 성과를 평가하는 모든 부처에서 발행되는 자료 | <ul style="list-style-type: none"> • 국내외 정부 간행물, 국내외 정책홍보자료 |
| | 정책연구자료 | <ul style="list-style-type: none"> • 정책관련 유관기관 및 연구자가 발간한 자료 | <ul style="list-style-type: none"> • 정책관련 연구논문 및 제언, 정책관련 보고서 |
| | 정책정보 웹문서 | <ul style="list-style-type: none"> • 정부 또는 관련 기관, 관련자 웹 사이트에 산재되어 있는 웹문서 형식의 정책자료 | <ul style="list-style-type: none"> • 정부부처 웹사이트, 정책정보를 생산하는 기관 웹사이트, 정책정보를 담고 있는 웹사이트 |
| | 정책정보 관련 데이터베이스 | <ul style="list-style-type: none"> • 간행물, 연구자료, 웹문서 등을 체계적으로 정리하여 검색 및 조회, 다운로드 서비스를 제공하는 DB | <ul style="list-style-type: none"> • 국내외 정책관련 자료의 데이터베이스, 국내외 정책정보 관련 부처의 서지 DB |
| 민간포함 비정형 데이터 (SNS, 뉴스, 영상/음성 등) | <ul style="list-style-type: none"> • 정책의제를 발굴하거나 새로운 정책수단을 개발하기 위한 착안점을 위기 위해 분석하는 대용량 데이터로 빅데이터 분석의 대상이 됨 | <ul style="list-style-type: none"> • CCTV나 SNS 데이터와 같은 비정형 정보, RFID와 같은 사물센싱으로 생성되는 반정형 데이터 등 | |

(4) 정책정보 서비스 포털 현황

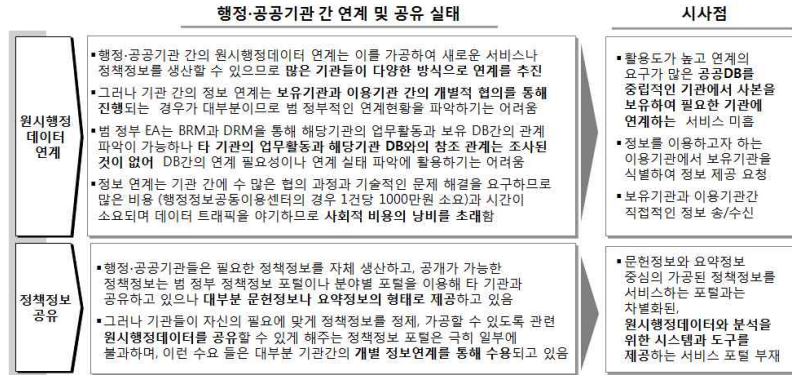
현재 공공부문의 정책정보는 주로 문헌정보 형태로 유통되고 있으며, 종합 또는 분야별 정책정보 포털에서 제공하는 정책정보 뉴스를 열람하거나 간행물, 연구자료, 웹문서 등을 검색, 조회하거나 정책통계 정보를 조회하는 것이 일반적인 형태임

[주요 종합정책정보포털 현황]

| 서비스명 | 서비스 소개 | 정책정보 서비스 내용 |
|---|--|--|
| 문화체육관광부 공공코리아 정책정보 아카이브 (www.korea.kr/arc) | 정부의 정책정보, 보도자료, 브리핑, 전문자료 등 정책 자료를 검색 및 열람하는 서비스 | <ul style="list-style-type: none"> • 174개 정책주제별로 정책보고, 연구보고, 정책홍보, 연감/백서 등 문헌정보와 통계/지표 등 요약정보 중심의 정책정보 서비스 제공 있음 • 원시행정데이터와 같은 정책정보는 제공하지 않으며 통계지표 등 요약정보도 역설파일 수준이어서 가공이 용하지 않음 |
| 통계청 국가통계포털 (kosis.kr/index/index.jsp) | 국가승인통계를 국민들이 편리하게 이용할 수 있도록 하는 One-Stop 통계정보서비스 | <ul style="list-style-type: none"> • 529종의 국가승인통계를 통계청 자체의 16개 주제별로 분류하여 액셀 형태의 요약정보로 제공하고 있어 일부 가공이 가능함 • 원시데이터는 제공하지 않으므로 이용자들이 다양하고 풍부한 정책정보(요약정보)를 만들어 내기 어려움 |
| 국립중앙도서관 디브리리 정책정보 (policy.dlibrary.net) | 정책관련 원천에 대한 주소정보 제공과 정부기관 등 주요 정책관련 기관에서 발행되는 최신자료와 간행물을 선별하여 원문 링크 서비스를 제공 | <ul style="list-style-type: none"> • 정부부처, 청, 위원회 및 산하기관들의 정책정보 원천 링크 서비스 제공 • 최신 정책자료의 메타정보와 정부간행물 추천자료의 서지정보를 제공하고 있으며 원문데이터를 직접 보유하지는 않음 • 요약통계정보나 원시행정데이터와 같은 유형의 정책정보는 제공 않음 |
| 행정안전부 대한민국정부포털 (www.korea.go.kr) | 행정 공공기관의 사이트, 서비스, 웹문서 등의 정보를 단일창구를 통해 쉽고 빠르게 제공 | <ul style="list-style-type: none"> • 국정 뉴스, 자치단체 뉴스와 정책정보를 보유하고 있는 정부기관 포털의 메타데이터, 각종 테마별, 사용자별 서비스안내 메타정보를 제공 • 정책과 관련된 문헌정보 원본을 보유하고 있지는 않음 |
| 행정안전부 정부통합전원포털 국가공유자원포털 (www.data.go.kr) | 국가 보유의 다양한 국가공유자원(서비스, 소프트웨어, 컴포넌트형태)에 대한 검색/활용 지원 | <ul style="list-style-type: none"> • 공공정책, 통계, 법률, 정치, 국토관리, 환경, 기상, 재해예방, 치안 등의 주제영역에서 텍스트, 이미지, 동영상, 데이터셋 등의 매체 유형으로 구성된 데이터의 열람 및 다운로드, 배송 서비스와 Open API 연계 통한 행정데이터 2차 가공 지원 서비스를 제공하고 있음 • 데이터셋과 Open API 서비스는 원시행정데이터에 가장 가깝지만 데이터셋은 학술연구 분야에 국한되고, Open API는 용도가 제한적임 |
| 행정안전부 프리즘 (www.prism.go.kr) | 중앙부처에서 수행하는 정책연구과제를 포괄적으로 관리하고 정책연구보고서를 공유하는 시스템 | <ul style="list-style-type: none"> • 정책연구보고서에 국한된 문헌정보 중심의 정책정보를 제공하고 있음 • 요약통계정보나 원시행정데이터와 같은 유형의 정책정보는 제공 않음 |
| 행정안전부 정부통합지시행정 시스템 (www.gkmc.go.kr) | 중앙행정기관 업무관리시스템, 중앙기관과 지방자치단체별로 운영중인 지식관리시스템의 연계와 법령, 통계, 교육훈련, 정책연구, 해외정보 등 정책수립을 지원하는 다양한 지식정보를 제공하고 있음 | <ul style="list-style-type: none"> • 중앙행정기관 업무관리시스템, 중앙기관과 지방자치단체별로 운영중인 지식관리시스템의 연계와 법령, 통계, 교육훈련, 정책연구, 해외정보 등 정책수립을 지원하는 다양한 지식정보를 제공하고 있음 • 실무중심의 정책관련 정보가 풍부하게 유통되나 대부분 문헌 정보이며 원시행정데이터와 같은 유형의 정보는 유통되지 않음 |

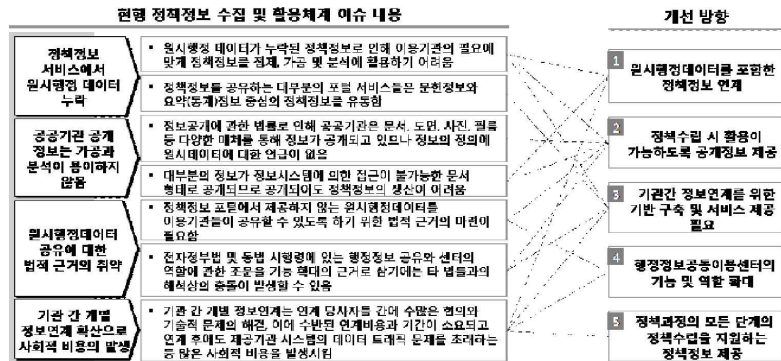
(5) 행정-공공기관 간 정책정보 연계/공유 실태 및 시사점

전체 정책영역에 대해 원시행정데이터나 분석 도구를 유관기관에 공식적으로 제공하는 서비스나 기관은 극소수임



(6) 현행 정책정보 수집 및 활용체계의 문제점

정책정보를 생산하는데 있어 타 기관의 원시행정데이터를 이용할 수 있도록 하는 법제도적 여건의 개선이나 서비스 기능 확대에 대한 의지가 강하지 않음



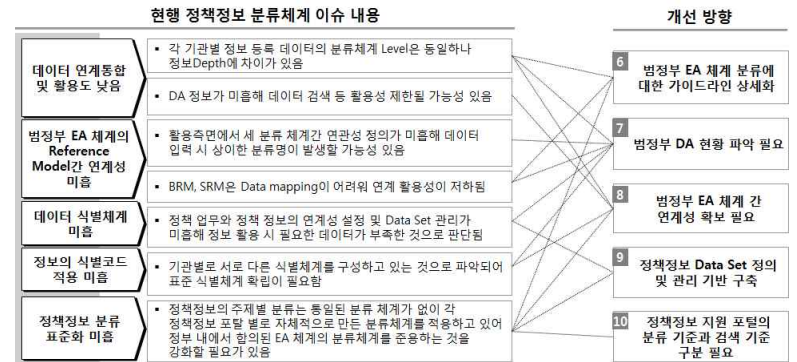
(7) 정책정보 분류 체계 현황 분석

- 정책정보는 주제별, 유형별, 생산기관별 등 다양한 기준으로 분류가 가능하지만 국내외를 막론하고 확립된 분류기준이 있는 것은 아니고 서비스 기관 나름대로의 기준을 세워 분류하는 것이 일반적임
- 실제로 국내의 정책정보 서비스 포털의 현황을 살펴보면, 자료의 생산기관과 주제별 분류를 제공하고 있으나 자체적인 임의 분류 기준도 사용하고 있음

- 업무기준 분류체계인 BRM은 정책연구관리 시스템, 국가재정운용계획분야 등과 매칭되므로 정책정보 분류체계로서 준용하는 것이 타당할 것으로 판단됨

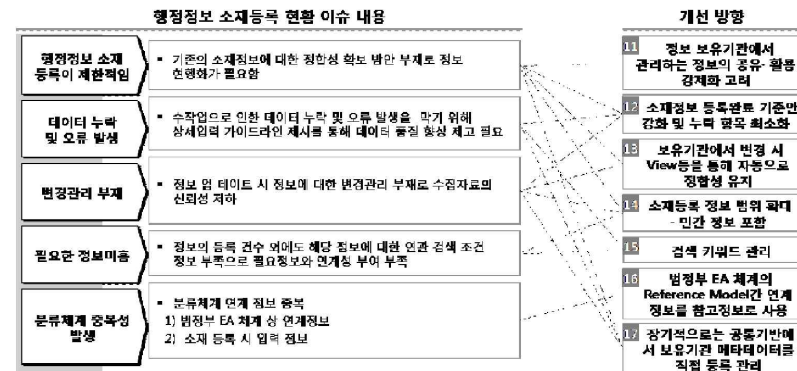
○ 현행 정책정보 분류 체계의 이슈 및 개선방향

분류 체계 자체는 기준으로서 적용이 용이하나 활용성 측면에서 주제별 검색 등의 용도에 부적합하므로 별도의 검색/활용 방안이 필요함



(8) 행정정보 소재 등록 현황 분석

구조적으로 등록 정보의 미비와 현행화 부족의 문제를 겪고 있어 공통기반에서 활용할 수준에 미흡하므로 장기적으로는 협약을 맺은 보유기관의 메타데이터를 직접 등록하고 갱신하는 체계로 운영해야 할 것임



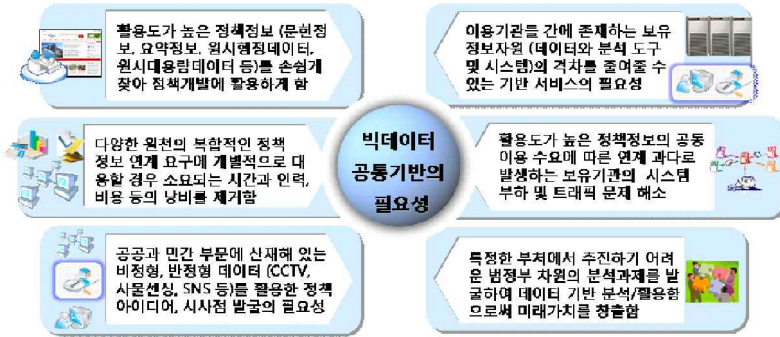
(9) 공공정보 공개에 관한 해외사례 분석

미국, 영국 및 호주의 공공정보 공개방식의 주요 특징은 데이터의 연계-통합 및 데이터의 유통 생태계 조성을 위해 Linked Data의 도입을 들 수 있으며, 국내 공공정보 공개방식에도 도입이 고려되어야 하는 것으로 판단됨

나) 정책정보 지원을 위한 빅데이터 공통기반의 개념

(1) 정책정보 지원을 위한 빅데이터 공통기반의 필요성

현행 정책정보 수집 및 활용체계를 개선하고 빅데이터 활용기반을 조성하기 위해서는 공통적으로 필요한 기능을 통합하여 행정·공공기관 빅데이터 활용 플랫폼을 구축하는 빅데이터 공통기반이 필요하다



(2) 정책정보 지원을 위한 빅데이터 공통기반의 요건

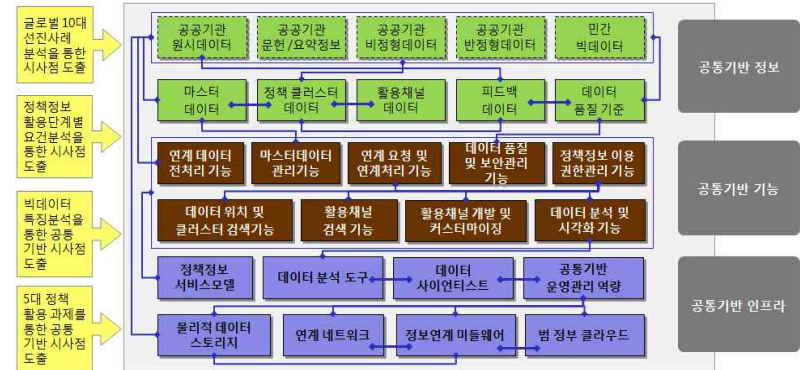
정책정보 수집 및 활용체계를 개선하기 위해 필요한 빅데이터 공통기반은 정책단계의 전 과정을 지원하는 정책정보를 공급할 수 있도록 물리적 플랫폼과 서비스 포털, 분석 도구와 시스템이 유기적으로 상호작용하는 것이어야 함



(3) 빅데이터 공통기반 구성요소 도출

선진사례 분석, 정책정보 활용단계별 요건분석, 빅데이터 특징분석, 5대 정책활용과제 분석을 통해 도출된 공통기반 요소들을 정보, 기능, 인프라의 관점에서 분류 및 구조화하여 아래와 같은 빅데이터 공통기반 구성 개념도를 작성하였음

[빅데이터 공통기반 구성 개념도]



(4) 정책정보 활용 표준모델 정의

공통기반의 구성요소를 활용하여 다양한 조합의 정책정보활용 표준모델을 도출하고, 이 중에서 이용기관의 유형과 요구정보 특성을 고려하여 4가지 표준모델을 도출하였으며, 각각의 서비스 개념도는 다음과 같음

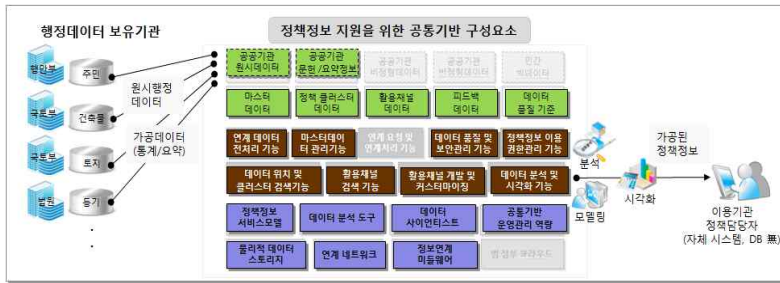
[문헌정보 기반의 정책정보 검색 서비스]



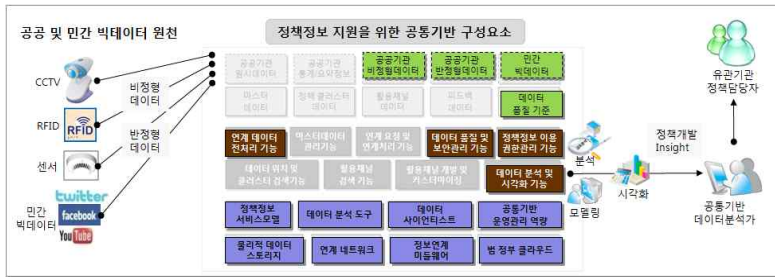
[공공데이터를 활용한 정책개발 서비스]



[타 기관 보유 원시행정데이터의 연계 서비스]



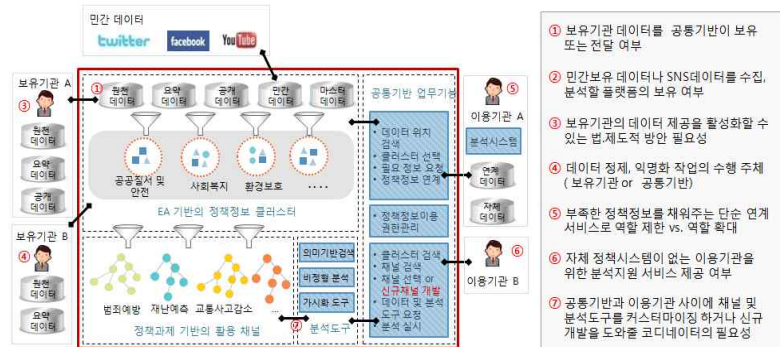
[공공데이터 및 민간 빅데이터를 활용한 정책개발 서비스]



(5) 빅데이터 공통기반 서비스의 개념 및 고려사항

빅데이터 공통기반은 공공부문의 정책정보 지원을 목적으로 추진되는 범 정부적 정책 정보 서비스 체계로, 서비스 모델은 공통기반을 가운데 두고 보유기관과 이용기관이 각자의 여건에 맞는 다양한 서비스를 제공하거나 활용할 수 있도록 설계함

빅데이터 공통기반의 정의
공공 정책의 품질과 생산성 향상을 위해 공공과 민간 부문에 산재한 다양한 형식과 내용의 정보를 체계적으로 수집하여 사용자 지향적으로 구조화하고 다양한 분석의 관점을 정책담당자에게 제시 해 주는 범 정부적 정책정보 서비스 체계임



다) 정책활용과제 업무 현황

(1) 범죄예측 정책 활용과제 현황 분석

(가) 목적 및 범위

- 경찰청의 범죄발생 예측분석 업무를 중심으로 주요 업무 절차 및 이해관계, 참조정보, 요구사항 등의 상세 현황분석을 통해 범죄발생 최소화 구현을 위한 시사점 및 개선방향을 도출함

(나) 업무 현황 분석

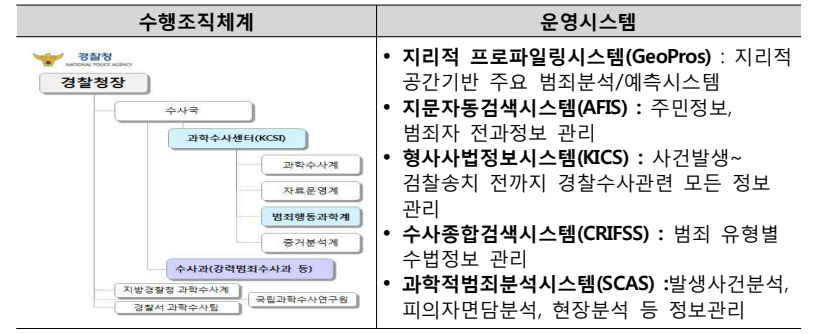
○ 범법 유형별 분석

- 최근 10년간 5대 범죄 및 성폭력 등 범죄통계를 기반으로 범죄발생 유형 중 사회적 이슈가 가장 큰 범법유형을 대상으로 분석함

| 범법 유형별 현황분석 | 시사점 |
|--|---|
| <p>범법 유형별 분포 현황</p> <p>성범의 증가 현황</p> <p>범죄 유형별 분포 현황</p> <ul style="list-style-type: none"> - 01년~11년 중범죄(5대 범죄 중상위)의 범법유형별 발생순위는 1) 절도, 2) 폭행, 3) 강간으로 도출됨 (여 성폭력범죄가 2위를 탈락하였고, 10년 이후 1위로 부상하며, 11년 현재 전체 범법유형 중 281,561건으로 전체 범죄 대비 34.7%를 차지함) - 절도범죄는 매년 증가하는 추세이며, 11년 현재 전체 범법유형 중 281,561건으로 전체 범죄 대비 34.7%를 차지함 - 평균발생빈도 : 11년 평균 절도범죄 하루 771.4건, 시간당 32.1건 발생, 1년 이내 재범률 63.2% 발생 - 성범의 증가 현황(최근 3년간) - 전체 범법유형 중 성범의는 발생비중은 낮으나, 다른 범법이 줄어드는 것에 비해 성범죄(강간, 강제추행 등)의 경우 3년 평균 36.4%의 증가 추세를 헤미다. 가장 높은 증가추세를 보이고 있음 - 평균발생빈도 : 11년 평균 성폭력 범죄 하루 60.4건, 시간당 2.5건 발생, 1년 이내 재범률 43% 매우 높음 <p><small>주:1) Source : 범법분석 통계, 대검청, 2009~2012</small></p> | <ul style="list-style-type: none"> • 범법유형 중 절도와 성범죄의 범죄가 매년 증가 추세로 사회적 문제로 부각 • 범죄발생건수 및 범죄재범률이 가장 높은 절도(침입절도) 경우 사전 범죄예방 활동 수행 시 범죄발생을 감소 실효성 기대 |

○ 현황 업무 수행체계 분석

- 범죄대응을 위한 수행조직체계 및 운영시스템의 구성은 다음과 같음



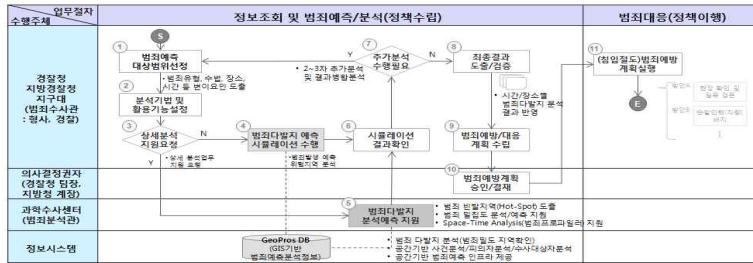
※ 과학수사센터에서는 범죄발생 최소화 정책정보 지원을 위한 범죄예측분석 업무지원 및 운영시스템을 보유하고 있으며, 실제 범죄대응 업무는 경찰청/지방경찰청/지구대의 수사담당관이 범인검거 및 범죄대응을 수행함

(다) 범죄예측 정책업무 프로세스 분석

| 유형 | 업무 | 방식 | 수행내용 | 발생주기 | 주무부서 | 운영시스템 | 기타 |
|--------|---------------------|---|--|---------------------|--|--------------|-------|
| 업무유형 A | 범죄다발지 분석 (Hot-Spot) | 범죄예측 | 특정지역/시간대별 범죄 빈발지역(Hot-Spot) 및 밀집정도 분석을 통해 사전 범죄발생 예측 분석 수행 | 불규칙 (개인적 역량에 의해 발생) | 경찰청/지방경찰청/지구대 (범죄수사관: 형사, 경찰) | 지리적정보 파일링시스템 | 정책 수립 |
| 업무유형 B | 연쇄범죄자 거주지예측분석 | 범죄예측 | 연쇄적 범죄의 발생지점 중심으로 근접성 영역 추출, 범인 거주 및 주 활동영역 도출 등의 패턴 분석을 통한 연쇄범죄자의 거주지예측 분석 수행 | 불규칙 (특수사건 발생시) | 경찰청/지방경찰청/지구대 (범죄수사관: 형사, 경찰) / 과학수사센터 (범죄분석팀) | 지리적정보 파일링시스템 | |
| 분석방향 | | - 9대 범죄 및 상륙률 등 특별법 134종에 대한 범죄분석 예측분석 업무관련은 모두 유사하며, 기본적인 업무처리 유형 A,B를 기반으로 수행일자/이웃정보/RDR/발생주기/수행내용 등의 분석 수행함 ① 업무유형A 분석범위: 범죄다발지 분석의 경우, 9대 범죄 유형 중 '밀집정도' 유형 기준 분석 ② 업무유형B 분석범위: 연쇄범죄자 거점예측 분석의 경우, 9대 범죄 유형 중 '상범죄자' 유형 기준 분석 | | | | | |

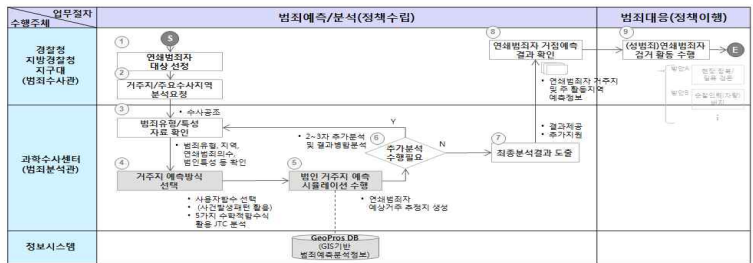
○ 범죄다발지 분석(Hot-Spot) : ① 침입절도 유형

- 범죄수사부서에서는 특정지역/시간대별 범죄 빈발지역(Hot-Spot) 및 밀집 정도 분석을 통해 도출된 분석결과를 목검문이나 순찰차량/인력 배치 등의 범죄대응 업무에 활용하는 프로세스임



○ 연쇄범죄자 거주지예측분석 : ② 연쇄 성범죄자 유형

- 범죄분석관은 선정된 연쇄범죄자의 연쇄적 범죄 발생지점을 중심으로 범죄 발생 지점간 근접 정도 분석을 통한 근접 영역 추출, 범인거주 및 주 활동영역 도출 등의 공간적 패턴분석을 통한 연쇄범죄자의 거점예측분석 및 범죄수사관의 검거 활동을 지원하는 프로세스임



(라) 범죄예측 정책업무 이용정보 현황

- 경찰청에서는 범죄발생예측분석을 위하여 경찰청 내부시스템 중 4개의 개별시스템 DB정보를 연계하여 활용하고 있으며, 각기 보유하고 있는 정보특성을 범죄분석 변이요소로 이용함

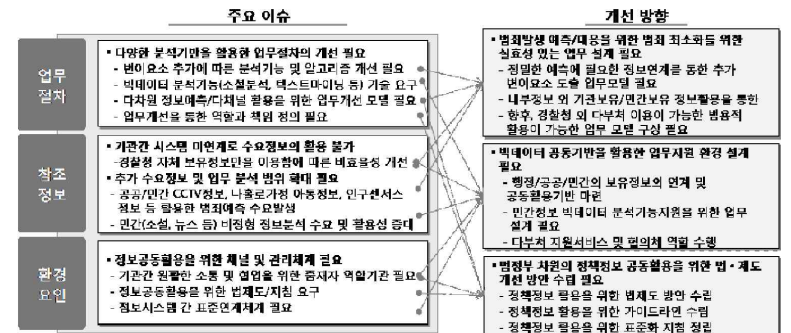
| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|-----------|--------------------|-----------|------------------------|----|------------------------------------|
| 범죄사건정보 | 사건분석 피의자분석 | 경찰청 | 형사사법정보 시스템(KICS) | 정형 | 현재 GeoPros 연계 및 요약DB 생성활용 |
| 수법원지정보 | 범행수법분석 | 경찰청 | 수사종합검색 시스템(CRIFISS) | 정형 | |
| 주민전과정보 | 수사대상자 (주민/전과)분석 | 경찰청 | 지문자동검색 시스템(AFIS) | 정형 | |
| 사건분석정보 | 범인거점 추정지표 | 경찰청 | 과학적범죄 분석시스템(SCAS) | 정형 | |
| 공간(GIS)정보 | 지도표기/활용 | 국토 해양부 | 국가공간정보 유통시스템 | 정형 | |

(마) 국내/해외 선진사례분석

| 구분 | 사례 | 시사점 |
|--------|--|---|
| 국내 /미국 | • 인천시 송도, 미국 멤피스시 등 실시간 범죄 감시 센터(ICO) 운용 사례 | <ul style="list-style-type: none"> • 사회안전과 범죄 예방활동을 종합적으로 고려한 환경적 변화 제시 • 범죄유형별 패턴분석 및 예측 알고리즘 등 다양한 분석 기술의 필요성 대두 • 빅데이터 기반 과학적인 범죄예측 및 대응전략을 통해 제한된 경찰인력으로 광범위한 영역의 순찰, 범죄 예방활동 가능 |
| 미국 | • CIA 빅데이터 해석을 통한 예측분석기술 확보 주력 사례 | |
| 미국 | • 샌프란시스코, 범죄예방시스템 이용 안전지역 사회구축 사례 | |
| 미국 | • 캘리포니아 산타크루스 경찰국(SCPD)의 범죄 예측 프로그램 추진 사례 | |
| 영국 | • 영국 내무성의 신뢰성 높은 '범죄지도 (Crime Mapping)' 제작 추진 사례 | |

(바) 이슈 및 개선방향 도출

- 범죄발생 최소화를 위한 미래 모델 수립 시 해결되어야 하는 주요 이슈 리스트를 도출하고 개선 방향을 종합함



(2) 재난-재해 정책 활용과제 현황 분석

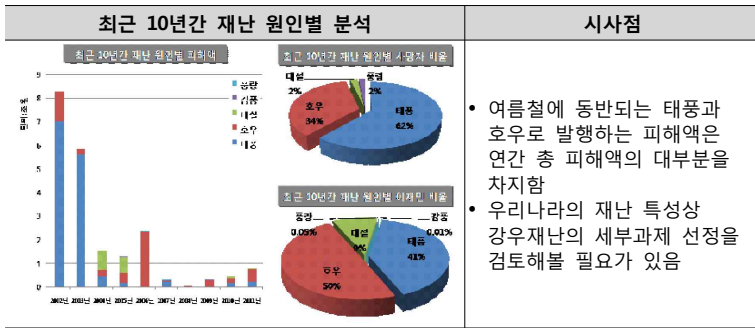
(가) 목적 및 범위

- 재난 전조감지 체계 이해관계자들의 현행 업무절차 및 참조정보, 요구사항, 정보 수요 등의 상세 현황을 분석하고, 재난 전조감지 목표모델 설계를 위한 개선방향을 도출함

(나) 업무 현황분석

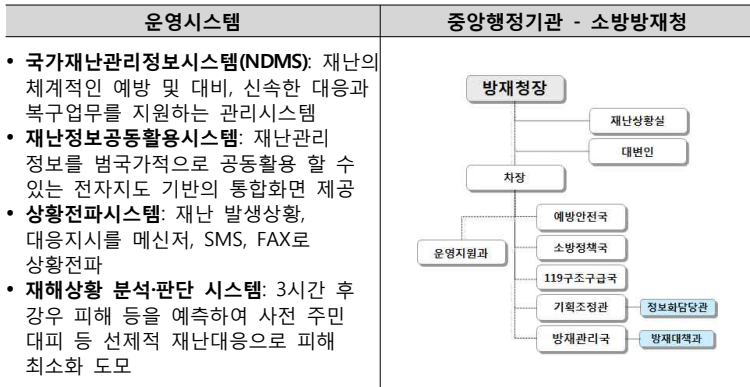
○ 재난 유형별 분석

- 우리나라 최근 10년간 재난 발생 패턴 및 피해액, 사망자 및 이재민의 통계를 기반으로 파급력이 큰 재난 유형을 분석함



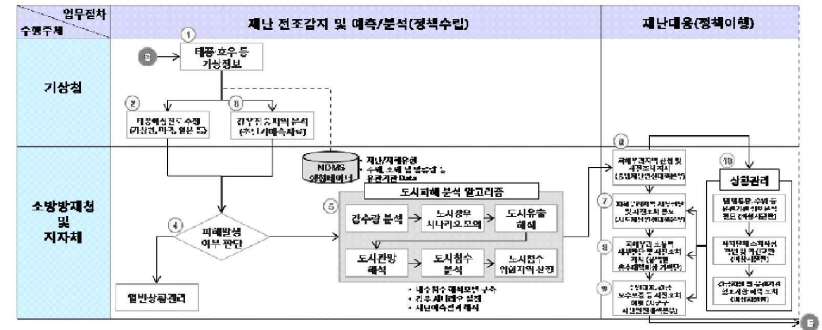
○ 현행 업무 수행체계 분석

- 소방방재청에서 재난-재해 통합관리체계로 국가재난관리정보시스템(NDMS)을 운영 중에 있으나 전조감지-선제적 예측보다는 사후대응에 초점이 맞춰져있음

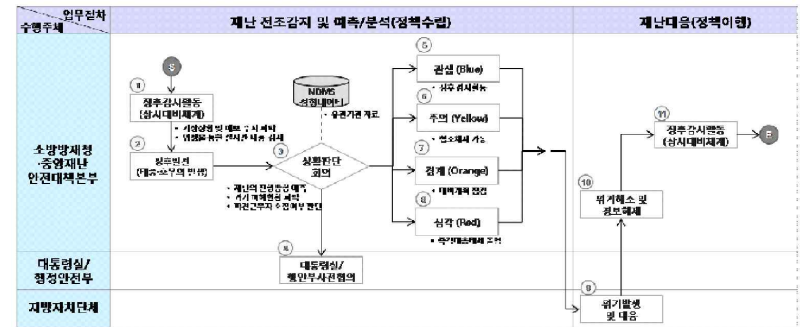


(다) 재난 전조감지 정책업무 프로세스 MAP

○ 강우재난 분석 및 판단



○ 태풍·호우 위기경보 발령



(라) 해외 선진사례 분석

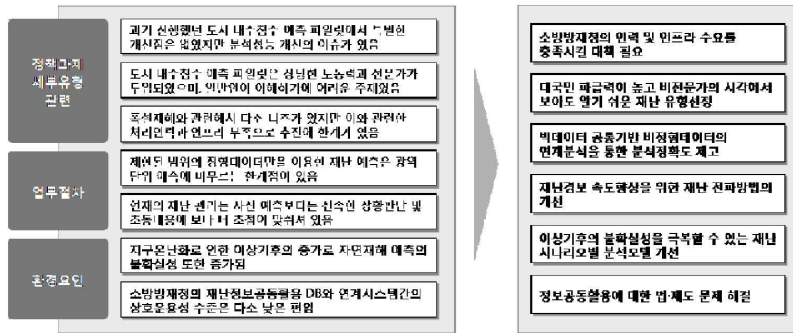
| 국가 | 명칭 | 시사점 |
|----|---------------------------------|---|
| 미국 | 건전성 모니터링을 기반으로 한 재난전조 관리 | <ul style="list-style-type: none"> • 센서를 이용한 건전성 모니터링 • 기반시설 정부수집국의 자산관리 시스템과 리스크 분석센터간 협력관계를 통해 국가 위험에 대응 |
| 일본 | 토사, 지진재해 등의 자연재해분야에서의 전조현상정보 활용 | <ul style="list-style-type: none"> • 전조현상의 발생과 시간적 관계를 참고하여 전조현상을 피난권고 등의 기준으로 활용 • 자율방재단이나 주민 참여의 전조감지 및 상황전파 매뉴얼 수립 |

(마) 정책지원 이용정보 현황

| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|---------------------|--------------|-----------|------------------------|----|-----------|
| 기상정보 | 기상정보분석 | 기상청 | 종합기상정보시스템 | 정형 | 연계 (NDMS) |
| 지자체 수위, 우량, 댐 정보 | 댐수위 및 방류량 분석 | 한국수자원공사 | 지자체 수공간 수문관측자료 공동활용시스템 | 정형 | |
| 소통상황, 도로정보, 기상정보 | 도로교통 상황분석 | 한국도로공사 | 고속도로 교통관리시스템 | 정형 | |
| 산사태 위험지구 | 산사태 위험지역 분석 | 산림청 | 산사태 위험지 관리시스템 | 정형 | |
| 한강-중랑천-하상도로 침수수위 정보 | 교통통제구간 분석 | 국토해양부/경찰청 | 수자원관리종합정보시스템/도시교통정보시스템 | 정형 | |

(바) 이슈 및 개선방향 도출

- 재난 전조감지를 위한 미래 모델 수립 시 해결되어야 하는 주요 이슈 리스트를 도출하고 개선 방향을 종합함



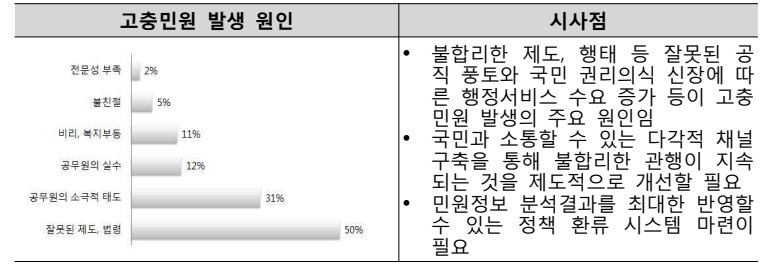
(3) 민원환류 정책 활용과제 현황분석

(가) 목적 및 범위

- 민원정책 환류 이해관계자들의 현행 업무절차 및 참조정보, 향후 요구사항 및 정보 수요 등의 상세 현황 분석을 기반으로 민원 데이터 분석 및 환류를 통한 고충민원 감소를 위한 개선방향 및 미래 모델을 도출함

(나) 업무 현황분석

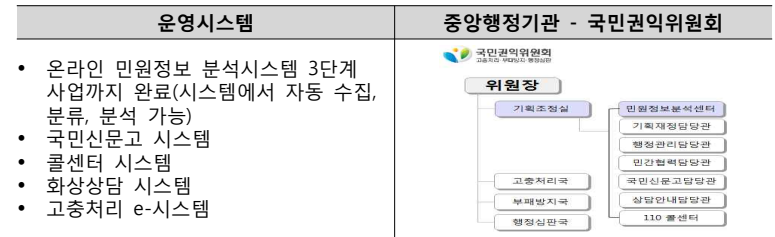
- 고충민원 발생원인 분석
 - 고충민원 발생 원인을 조사한 결과, 잘못된 제도 및 법령이 주 원인으로 파악됨에 따라, 본 사업은 민원정보 분석을 통한 정책의 환류시스템을 통해 잘못된 제도나 법령으로 인한 민원이 반복적으로 발생하는 악순환을 방지하고자 함



- 불합리한 제도, 행태 등 잘못된 공직 풍토와 국민 권리 의식 신장에 따른 행정서비스 수요 증가 등이 고충민원 발생의 주요 원인임
- 국민과 소통할 수 있는 다각적 채널 구축을 통해 불합리한 관행이 지속되는 것을 제도적으로 개선할 필요
- 민원정보 분석결과를 최대한 반영할 수 있는 정책 환류 시스템 마련이 필요

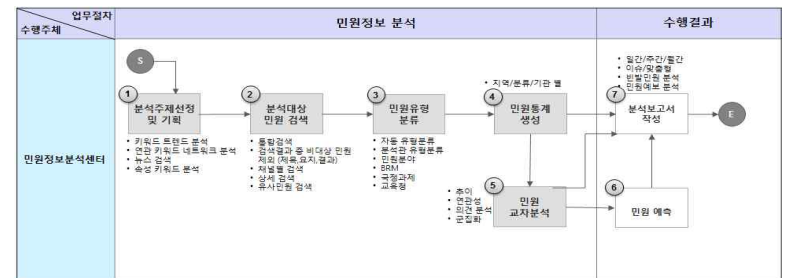
○ 현행 업무 수행체계 분석

- 국민권익위원회 기획조정실 소속 민원정보 분석센터에서 민원정보 분석시스템을 포함한 각종 시스템을 활용하여 민원 정보 수집, 민원 동향, 이슈 분석 등 각종 민원정보 분석 분석결과를 관련 부처 및 기관에 제공, 활용 실태 확인

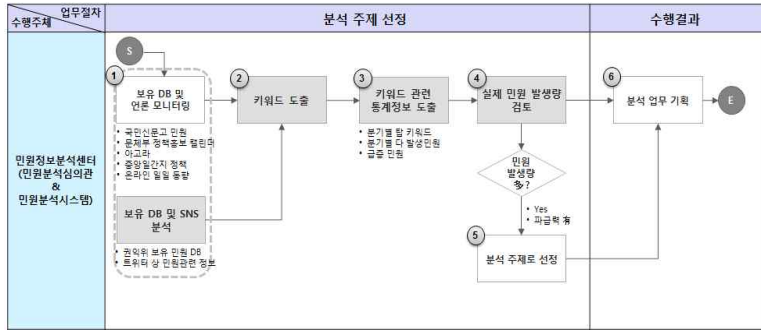


(다) 민원 분석/환류 정책업무 프로세스 분석

- 민원정보 분석/환류 업무는 민원정보분석센터의 민원분석관을 통해 분석시스템, 대외 정보 등을 기반으로 다양한 분석을 수행하고, 수행 결과물은 일일, 주간, 월간, 이슈 및 맞춤형 민원 보고서 형태로 제공
- 민원정보 분석 업무



○ 분석주제 선정 업무



(라) 정책지원 참조정보 현황

○ 정책지원 현 이용정보 현황

- 민원 분석/환류 정책담당 부서(국민권익위원회 민원정보 분석센터)에서는 현재 광범위한 민원 관련 정보를 보유·수집하고 있으며 각종 정책정보 및 통계정보를 분석 과정에서 참고하여 유의미한 분석결과를 도출함

| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|--------------|-----------------|-----------|------------|-----|------|
| 온라인 정책정보 | 발생가능 | 문화체육관광부 외 | 공감코리아 포털 | 비정형 | 조회 |
| 정책캘린더 | 민원 예측 시 사용 | 기타 | - | 비정형 | 조회 |
| 인구통계 | 민원분석 시 참조 | 통계청 | 국가 통계 포털 | 정형 | 수집 |
| 국민신문고 민원 | 민원관련 DB 구축 및 분석 | 국민권익위원회 | 민원정보 분석시스템 | 비정형 | 수집 |
| 110 콜센터 상담내용 | | 국민권익위원회 | | 비정형 | 수집 |
| 부처민원게시판 민원정보 | | 해당부처 | | 비정형 | 연계 |
| 지자체 민원정보 | | 지자체 | | 비정형 | 연계 |
| SNS 민원관련정보 | | 민간 | | 비정형 | 수집 |

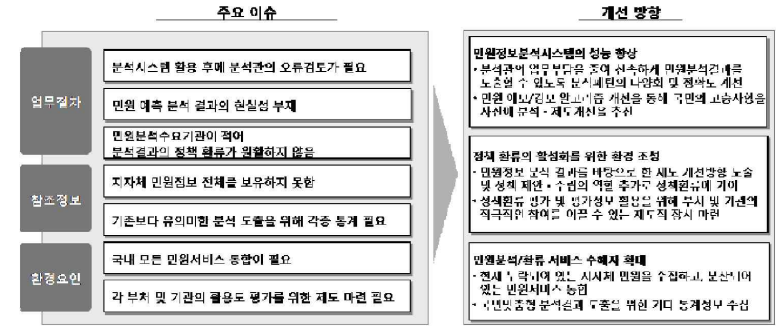
(마) 국내 및 해외 선진사례

○ 민원 분석/환류 관련 국내 및 해외 빅데이터 활용 선진사례 분석을 통해 시사점을 도출함

| 구분 | 모델 | 내용 | 시사점 |
|----|---------------------------------|--|--|
| 국내 | 한국도로공사, 고객 목소리 분석시스템을 통한 서비스 혁신 | <ul style="list-style-type: none"> 도로공사 고객의 목소리(VOC, Voice of Customer) 분석을 통한 개선 활동 콜 상담서비스, 민원관리 시스템, 채팅 상담 시스템을 고도화된 언어처리 기법으로 분석하여 고객만족도에 도움이 될 수 있는 지표와 이슈 도출 분석 주제를 기반으로 키워드와 도표를 추출하여 유형별, 시간별 분석 실시 연관도와 주이 분석, 토픽간의 연관성을 따르고 쉽게 확인 가능 | <ul style="list-style-type: none"> 국민의 의사를 반영한 정책 마련 및 분석 시스템으로 발전 가능 일관된 질서 경로가 다양해짐에 따라 데이터를 한곳에 통합하여 의미 분석하는 기술 필요 고객의 불만을 정확히 적극적으로 반영 가능 |
| 해외 | 뉴욕 시각유스시, 데이터 분석기반으로 스마트 시티 구현 | <ul style="list-style-type: none"> 낙후지역 발전에 필요한 요소로 효율적으로 제공하기 위한 시스템 마련 목적 새로운 도시계획 및 설계 시 주민들에게 필요한 사항을 적극 고려함으로써 과학적인 정책 수립을 위한 기반 마련 | <ul style="list-style-type: none"> 관련성이 낮아 보이는 데이터 분석을 통해 지역 특성에 맞는 새로운 도시 개발 추진 주거 지역의 불평등과 교육 및 직업훈련 부의과 관련되어 있다는 사실로 데이터를 통해 분석하고 해결방안 제시 |

(바) 이슈 및 개선방향 도출

- 민원 분석/환류 정책과제의 현황분석을 통해 주요 이슈 리스트를 도출하고 개선방향 종합함



(4) 교통사고 개선 정책 활용과제 현황분석

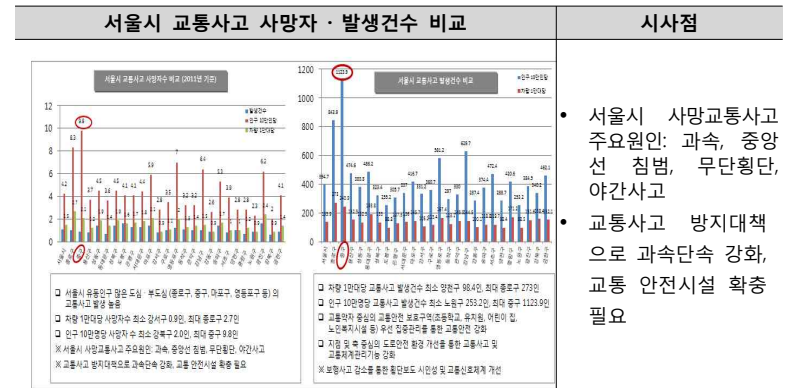
(가) 목적 및 범위

- 교통안전 관련자료 및 주민참여 개선의견 등을 이용하여 분석모델을 구축하고, 이를 정책에 활용하여 교통사고 발생 최소화 업무모델 설계를 위한 현황분석을 수행함

(나) 업무 현황분석

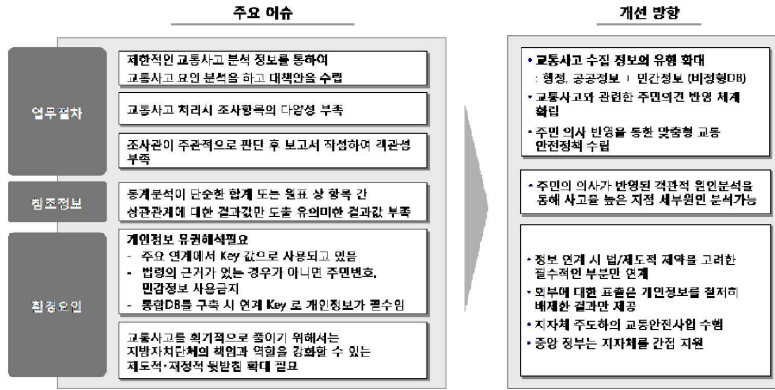
○ 서울시 자치구별 교통사고 현황

- 서울시 25개 자치구별 교통사고 현황을 분석한 결과 특별·광역시의 경우 교통사고 치사를 보다 교통사고 발생률이 전국 평균 대비 더 크게 분석되므로 도시지역은 교통사고 발생 감소에 대한 중점 관리가 필요함



- 서울시 사망교통사고 주요원인: 과속, 중앙선 침범, 무단횡단, 야간사고
- 교통사고 방지대책으로 과속단속 강화, 교통 안전시설 확충 필요

○ 교통사고 감소를 위한 미래 모델 수립 시 해결되어야 하는 주요 이슈 리스트를 도출하고 개선 방향을 종합함



(5) 국가미래전략 수립 정책 활용과제 현황 분석목표 모델 수립

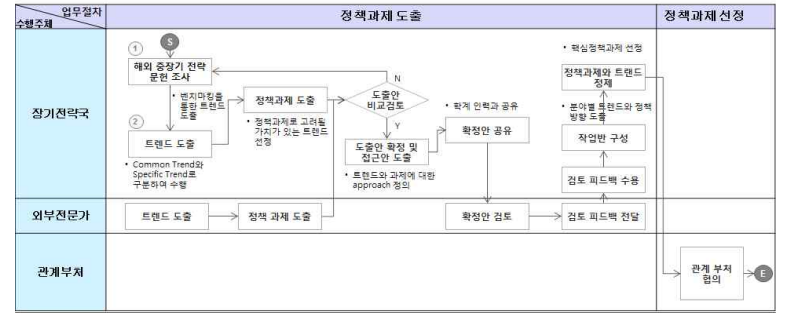
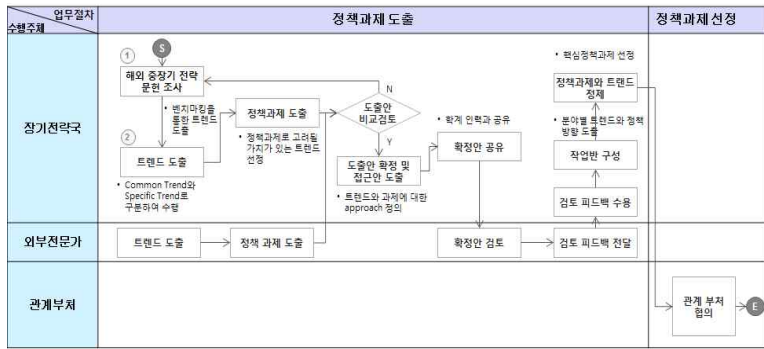
(가) 목적 및 범위

○ 다양한 정보분석 기반의 국가 미래전략 및 위기대응 전략수립 업무 현황 및 해외 유사사례를 분석을 통해 시사점 및 방향성을 도출함

(나) 업무절차

○ 업무수행 프로세스

- 기획재정부의 장기전략국은 글로벌 외부 전문가 분석을 참조하여 정책과제를 도출하고, 핵심정책과제 선정을 위한 approach 정의와 학계전문가의 피드백을 수렴하여 관계부처와 협의함



(다) 환경요인

○ 해외 선진사례분석


- 싱가포르의 환경탐색(Horizon Scanning)을 통해 획득된 데이터의 분류 및 관계분석등을 수행하는 데이터분석 실험센터(REC, RAHS Experimentation Centre)를 통해 국·지역 영향을 미칠 수 있는 잠재적 위험요소와 불확실성 요소를 탐색하여 이마징 이슈를 분석 및 대응 방안 마련함

| 구분 | 모델 | 내용 |
|------|------------|---|
| 싱가포르 | 싱가포르의 RAHS | <ul style="list-style-type: none"> · What: <ul style="list-style-type: none"> - 국가안보조정서사무국(NSCS)은 국가안전을 위협하는 요소에 대한 평가와 주변 환경변화를 탐지하여 새로운 기회를 발굴하는 RAHS(Risk Assessment Horizon Scanning) 프로그램 운영 · Result: <ul style="list-style-type: none"> - 해안 안전 확립을 위한 해상 상황 인식(Maritime Situational Awareness)을 통한 해상 테러, 해안침투 등 해안 영역에서의 안전 확보 및 조류독감 확산에 대한 미래 시나리오 연구와 대응방안 마련 |


- 영국에서는 다양한 주제와 이슈들의 연관분석을 통해 10~15년 이후 미래를 전망하는 미래신성장동력 발굴 프로젝트(Future Projects), 정부의 미래전략 및 정책 개발을 위해 필요한 미래 분석 가이드 및 방법론을 제시하는 미래연구역량강화 및 기법개발(Training and tools), STEEP(정치, 경제, 사회, 환경, 기술)별로 50년 이후의 정책 이슈를 포괄하는 연구 및 정책 보고서 검색 서비스를 제공하는 시그마 스캔(Sigma Scan)등을 통해 미래전략 수립 및 위기 대응 방안 수립함

| 구분 | 모델 | 내용 |
|----|---------------------------------------|---|
| 영국 | The Foresight Horizon Scanning Centre | <ul style="list-style-type: none"> · What: <ul style="list-style-type: none"> - 최신 과학이론과 데이터 등 증거기반의 정책분석 서비스를 제공하고 정부의 혁신적 전략 및 중장기 미래 정책개발 지원 · Result: <ul style="list-style-type: none"> - 영국인의 비만 문제가 사회를 위협하는 중요한 사항으로 판단하고, 과학적 근거를 기반으로 비만대책을 수립(Tackling Obesities : future choices) - 사면이 바다인 영국의 지정학적 특성으로 인해 해수면 상승으로 인한 위험 등 발생 가능한 잠재적 위험을 파악하여 30~100년 이후를 대비한 위험관리 프로젝트(Future Flooding)를 추진 - 글로벌 차원으로 확산되는 동식물 및 인간 전염병 확산에 대한 문제를 효과적으로 대응하기 대응책을 제시 |

- 미국의 경우 국립과학재단(National Science Foundation, NSF), 국립보건원(National Institutes of Health, NIH), 국방부(Department of Defense, DoD), 고등방위연구계획국(Defence Advanced Research Projects Agency, DARPA), 에너지부(Department of Energy, DoE), 지질조사원(Geological Survey, USGS)과 같은 유전자 연구 및 의료, 교육, 지구과학 및 국방분야 등 빅데이터 활용 효과가 뛰어난 분야의 기관들이 참여하여 수행하였음

| 구분 | 모델 | 내용 |
|----|--|--|
| 미국 | 빅데이터 추진 전략(Big Data Initiative)  | 국립보건원(NIH): - 빅데이터 과학, 공학 핵심 기술 개발 및 계층 프로젝트 데이터의 아마존 무료 공개 - 국립보건원과 국립과학재단은 대용량의 다양한 데이터로부터 유용한 정보를 추출, 가시화, 분석, 관리하는 핵심 과학 및 기술을 발전시키기 위해, 전 세계 1000개 유전체 프로젝트(1000 Genomes Project)에서 생성된 무려 200테라바이트 용량의 세계 최대 인류 유전자 변형 데이터 세트를 아마존 웹 서비스(AWS) 클라우드에서 무료로 공개 - 매우 큰 용량의 데이터인 관계로 일부 연구자들만 사용할 수 있었으나, 향후 무료로 사용할 수 있도록 추진하고 사용한 만큼의 컴퓨팅 서비스에 대한 비용만 지불 가능 |

- EU의 경우 복잡계 과학자들과의 협업을 통해 빅데이터 프로젝트를 수행함

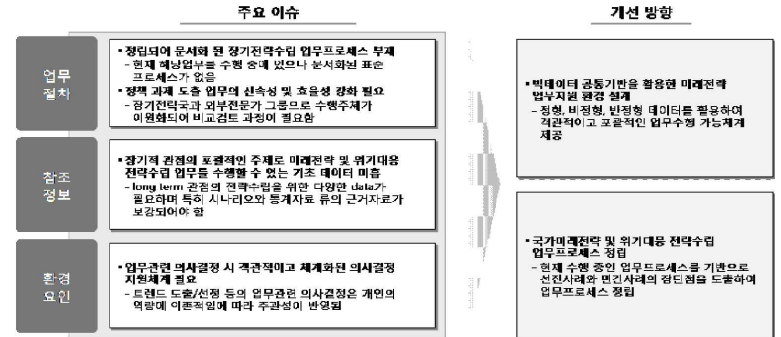
| 구분 | 모델 | 내용 |
|----|---|--|
| EU |  FuturICT와 iKnow 프로젝트 | FutureICT: - 지구 신경망 시스템, 전지구 시뮬레이터와 글로벌 참여 플랫폼을 구현하여 세계 변화의 방향과 새로운 지식 탐구 - 사회과학자와 함께 대용량의 데이터를 처리하는데 익숙한 자연과학자, 공학자, 컴퓨터과학자, 물리학자, 복잡계 과학자들의 협업을 통해 복잡계 과학(Complex System Science) 바탕으로 추진 iKnow(interconnect Knowledge) 프로젝트 - 자연재난, 테러와 참여와 소통의 web2.0 인적활동 및 글로벌 경제 위기 등 불확실성 속에서 예측하지 못한 사건에 대한 미래준비를 위한 프로젝트 - 전세계의 약신호(weak signal)와 와일드 카드(wild cards)를 포착하기 위해 데이터 분석 기반의 horizon scanning을 활용 - 체코, 핀란드, 독일, 이스라엘, 영국과 같이 국가단위의 환경스캐닝을 통해 미래의 핵심 이슈 제시 - 전세계로부터 포착한 약한신호와 와일드카드를 분석하여, 44개의 정책 브리프 제시 - 유럽전역, 과학기술 컨퍼런스 및 환경스캐닝을 통한 주요 이슈로부터 66명의 전문가들의 통찰력 보고서 제공 |

(라) 요구사항

- 인터뷰
 - 통계정보와 추정정보를 활용하여 미래예측 업무를 수행하고 있으나, 업무범위 및 예측기간의 확장을 위해 업무 시나리오와 통계자료와 같은 근거자료가 더 필요함

(마) 개선방향 도출

- 현황분석 결과를 업무절차, 참조정보, 환경요인 관점으로 이슈를 도출하고 개선을 위한 개선 방향을 도출함
 - 빅데이터를 활용한 공통기반을 활용한 업무지원 환경 설계이 필요함
 - 국가미래전략 및 위기대응 전략수립 업무프로세스 정립 필요



라) 학술데이터 공유활용 현황

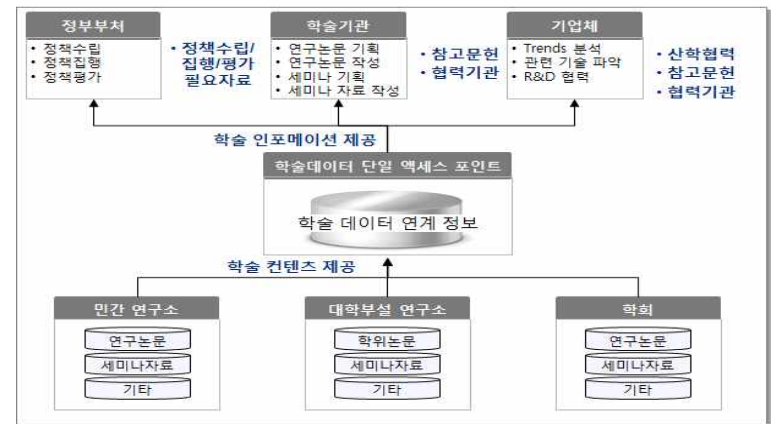
(1) 목적 및 범위

- 학술데이터의 공유 및 활용을 위해 관련 현황분석과 해외사례를 고찰하고, 공유 및 활용 모델을 정의하여 모델 별 평가를 통해 이행 과제를 정의함

(2) 학술데이터 공유 및 활용 정의

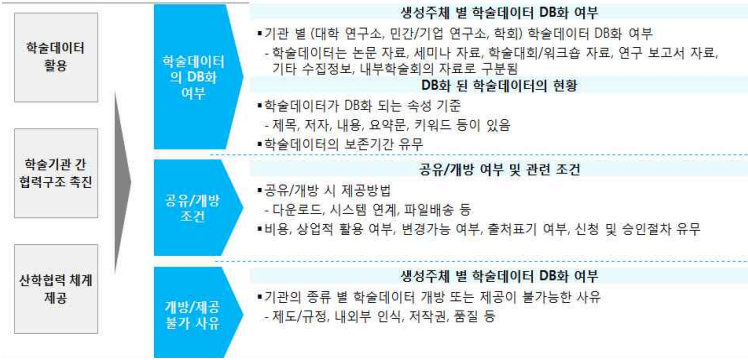
- 학술데이터 공유/활용의 정의는 다양한 학술데이터를 단일 액세스 포인트를 활용하여 정부부처, 학술기관, 기업체가 활용하도록 하는 것을 의미함

학술데이터 공유/활용 정의 개념도



(3) 학술데이터 공유 및 활용 관점

- 학술데이터의 공유 및 활용은 학술데이터 활용, 학술기관간 협력구조 촉진, 산학협력체계 제공으로 구성되어 있음. 본 프로젝트에서는 학술데이터의 DB화 여부, 공유/개방 조건, 개방/제공 불가 사유를 통해 학술데이터의 공유 및 활용 현황을 분석하였음



(4) 인터뷰 결과

- 3대 분야 5대 정책과제 담당 기관에서는 학술데이터 필요시 산하 연구기관이나 국가 통계지표를 활용하고 있는 것으로 나타남

(5) 시사점 도출

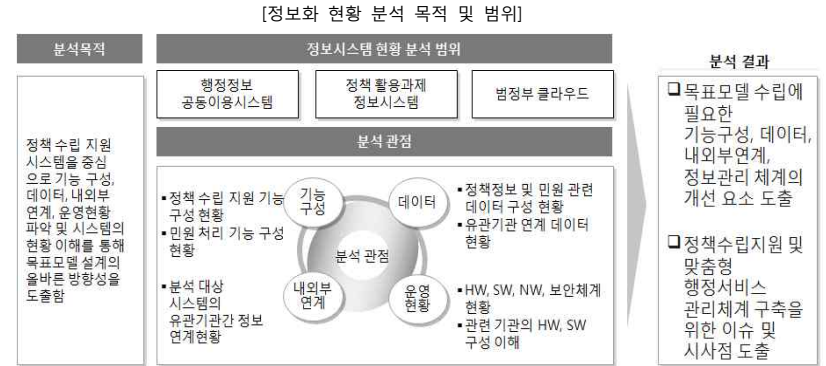
- 학술데이터 공유/개방을 위해 표준화에 의한 DB화 및 제공체계가 필요함

| 분석관점 | 현황분석 결과 | 시사점 |
|---------------------|--|--|
| 학술데이터의 DB화 여부 | <ul style="list-style-type: none"> 조사 대상 중 약 97%에 해당하는 학술데이터가 DB화 됨 논문자료 위주의 학술데이터임 다양한 속성기준으로 DB화 됨 대부분 다운로드 형식으로 제공 중임 | <ul style="list-style-type: none"> 논문 외 학술데이터에 대한 DB화 방안 필요 DB속성기준 표준화 방안 필요 제공방안의 표준화 필요 |
| 학술데이터의 5가지 공유/개방 조건 | <ul style="list-style-type: none"> 학술기관 별 silo 형태의 관리 실시 및 유료화 기준 상이함 상업적 사용이나 데이터의 변경이 불가능함 학술데이터 공유/개방을 담당하는 내부절차가 없음 | <ul style="list-style-type: none"> 유료 학술데이터의 공유방안 필요 학술데이터 공유/개방 표준절차 및 활용 가이드라인 필요 |
| 학술데이터의 개방/제공 불가 사유 | <ul style="list-style-type: none"> 개방/공유 관련 선례가 없어, 해당절차, 저작권확보, 데이터 품질에 대한 열려가 있음 개방/공유 시 자체서비스 비활성화 및 현 IT시스템의 수용가능여부 확인필요 | <ul style="list-style-type: none"> 저작권 관리체계 필요 공유 시스템 표준화 필요 |

3) 빅데이터 공통기반 및 활용과제 업무 정보화 현황분석

가) 정보화 현황분석 개요

- 빅데이터 공통기반 및 활용과제 방안을 설계하기 위해 검토한 행정정보공동이용시스템, 범죄발생최소화, 교통사고 감소 등 정책정보활용과제 관련 시스템 그리고 범정부 클라우드 현황을 기능구성, 데이터, 내외부 연계, 운영현황 관점에서 분석함



나) 행정정보공동이용시스템 분석

(1) 시스템 개요

행정정보공동이용시스템은 행정정보조회서비스, 행정정보유통서비스, 전자민원 서류관리시스템 등 주요 서비스로 구성되며 '13.01 현재, 135여종의 행정정보를 438개 기관에서 활용 중이며 '12.03 현재 361개 기관, 884개 시스템과 연계하여 1,421종 행정정보를 유통 중인

[시스템 개요]

| | | |
|---------------|--|---|
| 개요 | <ul style="list-style-type: none"> '13.01 현재, 135여종의 행정정보를 행정정보공동이용센터를 통해 438개 기관에서 활용 기관 간 정보 조회 및 전달기능을 넘어 범정부 정보유통 허브로서 진화중임 '08 수주자 맞춤형 행정정보 공동이용체계 구축을 위한 BPR/ISP 및 고도화를 통해 정보중추 및 기관 확대 주요서비스: 행정정보조회서비스, 행정정보유통서비스, 전자민원서류관리서비스 | <ul style="list-style-type: none"> 행정정보공동이용시스템은 '13.01 현재 135여종의 행정정보를 438개 기관에서 활용하도록 지원해주는 서비스이며, 행정정보조회 서비스, 행정정보유통서비스, 전자민원서류관리서비스 등 3대 서비스로 구성됨 '12.03 현재 361개 기관 884개 시스템을 연계하여 1,421종 행정정보를 유통 중인 빅데이터 공통기반 구축시 별도의 비용으로 연계기반 구축 없이 행정정보유통 서비스 연계 기반을 그대로 활용할 수 있는 장점이 있음 |
| 행정정보 조회서비스 | <ul style="list-style-type: none"> 민원·사무처리에 필요한 구비서류를 민원인이나 보유기관에서 제출 받지 않아도 담당자가 전산상으로 정보를 조회·확인하여 업무를 처리하는 서비스 135종의 행정정보를 행정정보 공동이용센터를 통해 438개 이용기관에서 활용 | |
| 행정정보 유통서비스 | <ul style="list-style-type: none"> 행정·공공·금융기관 간에 필요한 행정정보를 암호화하여 안전하게 유통해 주는 서비스로, 사회복지통합관리망 등 국가적인 과제를 효율적으로 지원하는 서비스 '12.03 현재, 361개 기관 884개 시스템 연계하여, 1,421종 행정정보 유통 중 복지급여수급자, 장애인 복지카드, 여권만료예정자 자료 중계처리 사실여부(Y/N) 확인서비스 및 실시간 유통 서비스 제공 | |
| 전자민원 서류관리 시스템 | <ul style="list-style-type: none"> 민원인이 인터넷으로 민원신청 시 온라인으로 직접 작성 제출한 전자화 서류와 민원서류 발급기관의 처리결과 등을 전자문서화하여 등록, 보관, 조회할 수 있도록 무방문 온라인 완결처리를 지원하는 서비스 민원24, 위택스(행안부 지방세 신고납부 포함), 농촌진흥청 전자민원포털 등과 연계, 총 1,800개 사무에서 활용 | |

*1 Source : 전자민원서류 관리화단 ISP, 2012; 행정안전부 / 빅데이터 활용기관 마련 및 활용을 위한 BPR/ISP, 2012; 행정안전부

(2) 시스템 구성

행정정보공동이용시스템은 정보공유서비스와 운영관리서비스로 나뉘어지며 정보 공유서비스에는 정보조회서비스, 정보유통서비스, 맞춤형정보유통서비스, 정보유통 기반이 있고, 운영관리서비스에는 통합모니터링시스템 등 운영관리시스템, 공동 이용사무관리시스템, 신원정보관리시스템 등이 있음

(3) 데이터 구성

행정정보공동이용시스템은 정보유통허브포털, 정보유통시스템, 맞춤형서류정보 열람체계, 정보유통허브시스템, 전자민원서류관리시스템에 대한 데이터베이스를 구비하고 있으며 이중, 정보유통허브포털의 정보송수신, 현황통계 및 정보소재 안내 DB를 추후 빅데이터 공통 기반과의 연계시 참조할 필요가 있을 것으로 판단됨

[데이터베이스 구성도]



(4) 하드웨어 구성

서울청사에 위치한 공공 및 금융기관 중계 및 연계서버를 제외한 행정정보공동 이용시스템의 DB서버, 응용서버, WEB서버, WAS서버는 정부통합전산(광주)센터에 위치하고 있으며 이중화되어 무중단 서비스를 제공할 수 있도록 구축 되어 있음

(5) 소프트웨어 구성

행정정보공동이용시스템(광주센터)에 주요서버 및 관련 S/W가 이중화되어 구성 되어 있으며 대부분의 주요서버에 로그수집 Agent 및 네트워크 모니터링 S/W 등이 설치되어 장애탐지에 유연하게 대처할 수 있도록 S/W기반이 구성되어 있음

(6) 네트워크 구성

행정정보공동이용시스템이 위치하는 정부통합전산센터(광주센터)의 라우터, 백본 스위치, L4스위치 등 주요 네트워크 장비와 방화벽, IPS 등 주요 네트워크 보안 장비가 대부분 이중화 구성되어 있어 가용성 및 보안성이 확보되어 있음

(7) 연계구성

행정정보공동이용시스템의 EAI 연계 기반은 이용기관과 제공기관 송수신연계서버에 BizMaster Remote Adapter 및 센터의 EAI API, BizMaster Server(H/A 구성)로 구성됨

ESB방식은 이용기관에 개방형표준연계API, 행정정보공동이용시스템의 Indigo ESB 서버, 보유기관 송수신연계서버에 ESB Agent가 상호 연계하도록 구성됨

(8) 보안구성

행정정보공동이용 정보시스템 요소별 보안구성은 통합보안관제, 내부정보유출 모니터링 등 모니터링, 단말기 해킹 톨 차단 등의 PC보안, 침입차단 및 침입방지 등의 네트워크보안, SecureOS 기반의 서버보안, DB보안, 증적관리, 유통정보 암호화 등의 응용시스템보안 등으로 구성되며 빅 데이터 공통기반 데이터 설계시 보안체계 준용을 고려해 볼 수 있음

[행정정보공동이용센터 정보보호체계]



다) 정책정보활용과제 정보시스템 분석

(1) 분석 개요

정책 활용과제 정보시스템 현황분석은 활용과제 업무분석에서 도출된 유관기관 정보시스템을 대상으로 해당 정보시스템의 기능, 데이터, 연계 현황에 대한 이슈/시사점, 정보화 요구사항을 도출하여 공통기반 설계를 위한 방향성을 도출하기 위해 수행하였으며 분석대상 및 수행방법, 분석관점은 다음과 같음

| 구분 | | 설명 |
|------|-----------------|--|
| 분석대상 | 범죄예측 활용과제 정보시스템 | <ul style="list-style-type: none"> 형사사법정보시스템(KICS) 지리적 프로파일링시스템(GeoPros) |

| | | |
|--|--------------------|--|
| | | <ul style="list-style-type: none"> 수사종합검색시스템(CRIFSS) 과학적범죄분석시스템(SCAS) 지문자동검색시스템(AFIS) |
| | 재난대응 활용과제 정보시스템 | <ul style="list-style-type: none"> 국가재난정보센터 재난관리정보DB센터 재난정보공동활용시스템 재난영상정보(CCTV) 통합·연계시스템 재난상황분석·판단연계시스템 |
| | 민원환류 활용과제 정보시스템 | <ul style="list-style-type: none"> 온라인민원정보분석시스템 정부민원안내콜센터 국민신문고 |
| | 교통사고 개선 활용과제 정보시스템 | <ul style="list-style-type: none"> 교통안전정보관리체계(TAAS) 교통경찰업무관리시스템 |
| | 수행방법 | <ul style="list-style-type: none"> 각종 문헌자료 및 홈페이지를 통한 정보시스템 시사점 자료 수집 인터뷰 및 정보화 시사점 조사 |
| | 분석관점 | <ul style="list-style-type: none"> 정보시스템 기능 파악 정보시스템 데이터 파악 정보시스템 연계 현황 파악 |

(2) 범죄 예측 활용과제 정보시스템

경찰청 과학수사 전산시스템은 주민·범죄경력·수배·지문정보 등 각종 수사자료를 처리하고, 원활한 업무를 처리하기 위해 구축·운영 중에 있으며, 경찰청 과학수사 센터 인터뷰를 통해 범죄 예측 활용과 관련된 시스템이 5개인 것으로 파악되었음

○ 형사사법정보시스템포털(KICS)

- 형사사법정보시스템포털은 각 수사시스템(사건수사시스템 등)의 관문역할을 수행하는 시스템으로 수사지식공유, 범죄통계, 사건관리 기능으로 사용하고 있으며, 수사도우미, 수사자료실, 수사공조, 참여마당, 피해자지원, 관리자 기능 등을 제공하고 있음

○ 지리적 프로파일링시스템(GeoPros)

- 지리정보시스템(GIS)의 공간분석 기능을 이용 주요 범죄 발생현황은 물론, 특정 지역내 범죄 다발지역을 분석 하여 목검문 등 다양한 범죄예방 및 검

거에 활용하고, 각종 수사전산 시스템과 연계를 통해 특정 지역내 동종 전과, 유사전과자 등 우선 수사대상군을 추출 범죄수사에 활용, 연쇄범죄자의 행동 패턴을 분석하는 시스템이며, 주요 제공 기능으로 위치찾기, 범죄다발지분석, 사건분석, 피의자분석, 수사 대상자분석, 통계 등이 제공되고 있음

○ 수사종합검색시스템(CRIFSS)

- 과거 단일망으로 운영되던 「범죄수법영상전산시스템」을 중심으로 변사·조직폭력영상시스템을 통합하고, 교통면허사진·주민·수용자·AFIS지문·수배자료 등을 연계하여 일선 현장에서 필요한 수사자료를 종합·효율적으로 제공하는 시스템이며, 주요 기능으로 통합, 수법, 조폭, 변사, 차적 등이 제공되고 있음

○ 과학적범죄분석시스템(SCAS)

- 범죄분석자료 관리시스템, 범죄현장자료 관리시스템, 기존 시스템과 데이터 통합 및 연계 그리고 범죄 분석 시스템의 구축을 통하여 유사 범죄 및 미해결 사건의 해결에 대한 정보를 제공에 목적을 두며, 또한 기존의 개별 시스템에 산재되어 있는 데이터를 통합하여 다차원적 조건 검색을 통한 신속 정확한 수사 자료 검색지원을 제공하는 시스템이며, 주요 기능으로 범죄분석자료 관리, 발생사건분석, 현장감식자료, 흔적증거관리, 통계자료 등이 제공되고 있음

○ 지문자동검색시스템(AFIS)

- 전국 읍·면·동사무소에서 작성한 주민등록발급신청서와 내·외국인 범죄자의 십지문원지 등을 디지털 이미지로 전환하고 지문관련 데이터베이스를 구축하여, 이를 기반으로 신원불상자·인적도용자·변사자·범죄현장 유류지문의 신원확인 등 효율적인 수사업무 지원을 목적으로 개발된 시스템이며, 주요 기능으로 지문검색, 사건관리, 사건처리현황, AFIS자료현황, 매칭현황 등이 제공되고 있음

(3) 재난대응 활용과제 정보시스템

국가재난관리정보시스템은 예방·대비·대응·복구 등 재난관리 단계별 업무를 지원하는 전국단위 종합정보시스템으로 재난감지, 상황전파, 정보수집 및 분석, 대국민 서비스를 제공하고 있음

○ 국가재난정보센터

- 태풍/호우 등 국민행동요령 플래시 애니메이션, 민방위 교육일정·대원정보, 사이버 방재교육 등의 정보를 종합적으로 제공하는 대국민 재난정보전문 포털 사이트로 재난관리정보DB센터, 재난정보공동활용센터, 재난관리업무지원시스템, 시군구 재난관리시스템에 있는 재난관련 정보를 통합하여 대국민에게 제공함

○ 재난관리정보 DB센터

- 재난관리 정책수립 및 의사결정을 지원하기 위한 기반 구축 및 분석·예측 서비스 제공을 목적으로 업무시스템별로 분산·관리되고 있는 재난정보(662종)의 표준화 및 통합관리, 축적된 재난정보를 바탕으로 과학적인 분석·예측(308개 서비스)을 통한 의사결정 지원, 예방안전 분야, 소방정책분야, 방재관리 분야, 재난상황 분야의 통계분석 서비스, 태풍 피해 예측 및 지역별 주간 안전사고 위험지수 산출 등 예측 서비스를 제공하고 있음

○ 재난정보공동활용시스템

- 42개 재난관리책임기관별로 보유·관리하고 있는 재난관리 정보(186종)를 연계하여 범국가적으로 공동 활용 할 수 있도록 전자지도 기반의 통합화면을 제공하며, 유관 기관에 GIS정보를 별도 연계서버 설치 없이 웹으로 서비스하는 기능을 구축하여 제공하고 있음

○ 재난영상정보(CCTV) 통합·연계시스템

- 지자체별로 다양한 형태의 재난영상정보(CCTV)를 개별적이고 독자적으로 관리 및 운영해 오던 것을 통합 모니터링 할 수 있는 체계를 구축하고, 영상정보와 현장상황 정보를 연계하여 실시간 제공하는 전자상황판 서비스와 지능형 영상감지기술을 적용한 재난유형별 자동 위험 알림 기능으로 실시간 탐지를 제공하고 있음

○ 재난상황분석·판단연계시스템

- 기후변화 영향으로 국지적·돌발성 집중호우가 빈발하면서 강우 패턴과 피해 양상이 과거와 다른 형태로 발생됨에 따라 기상정보·재해정보를 활용한 위험요인 사전 종합·분석 후 신속한 선제대응을 위한 의사결정 지원체계 지원을 위해 추진되었으며, 재난상황분석·판단연계시스템은 3시간 후 강우, 하천범람, 도시침수, 토석류 피해, 연안피해를 예측하여 사전 주민 대피 등 선제적 재난대응으로 피해 최소화를 도모함

(4) 민원 환류 활용과제 정보시스템

온라인 민원정보 분석시스템은 국민신문고, 110콜센터, 국가대표포털, 행정기관 및 지자체 홈페이지의 민원관련 정보의 통합DB화를 통하여 민원에 대한 분류·분석·통계 생성 등의 민원정책을 지원하는 시스템임

○ 온라인 민원정보 분석 시스템

- 온라인 민원정보 분석시스템의 주요 사용자는 권익위원회의 민원분석관과 공동 활용기관 사용자이며, 시스템을 통하여 다양한 민원정보 분석 보고서를 작성하여 개별기관에 제공하며, 웹에 있는 민원정보를 수집하기 위하여 민원정보 수집 서버(웹크롤러)와 신문고 민원DB 및 110콜센터 DB 연계서버를 통해 데이터를 수집함
- 수집대상 시스템 및 데이터의 특성을 고려하여 수집주기를 선택적으로 적용하였으며, 수집된 데이터를 활용하여 민원에 대한 예보 및 경보체계를 구현하기 위해 수집된 민원데이터 매트구성 및 예측 및 조기경보를 위한 알고리즘 모델링 등의 작업을 수행함

| 수집 대상 | 주요내용 | 수집 주기 |
|---------------------------|--|-----------|
| 국민신문고 민원, 국민제안(지자체 민원 포함) | ■민원제목, 내용, 날짜, 접수요지, 처리 결과, 기관명 등 | 실시간 |
| 110 콜센터 상담 | ■상담내용, 날짜, 지역, 처리기관명 등 | 실시간 |
| 공감코리아 게시판(www.korea.kr) | ■정책정보, 보도자료 | 4-6시간 |
| 정부기관 홈페이지 게시판 | ■29개 중앙행정기관 민원게시판 민원 | 6-8시간 |
| 지자체 홈페이지 게시판 | ■광역시도 민원게시판 민원 | 6-8시간 |
| 교육청 홈페이지 게시판 | ■194개 교육청/지청 민원게시판 민원 | 6-8시간 |
| 교육청 신문고 민원 DB | ■교육청 국민신문고 민원DB | 실시간 |
| 지자체 홈페이지 게시판 | ■기초지자체 민원게시판 민원 | 6-8시간 |
| 예산낭비신고센터 민원 | ■예산낭비신고센터 DB 연계 | 실시간 |
| 정책포럼/전자공청회 데이터 | ■정책포럼 발제글 및 댓글 연계 ■전자공청회 글 및 댓글 연계 | 실시간 |
| 다출처 민원 데이터 수집 및 연계 | ■민원 관련 키워드와 다출처 데이터 분석 연계 ■다출처 키워드 분석 결과 조회 | 2시간 / 실시간 |

○ 정부민원안내콜센터

- 단일 전화번호체제로 다수의 정부기관의 전화민원을 통합적으로 응대 및

처리하기 위한 정부민원안내콜센터를 운영하고 있고 있으며, 다음과 같은 서비스를 제공함

- 전국 단일 전화번호체계로 시내요금만 부담하는 민원안내 서비스 제공
- 단순 문의사항에 대해서도 콜센터 상담원이 직접 상담·안내 서비스 제공
- 생활·고충 민원에 대한 기존 민원창구와 네트워크를 통한 서비스 제공
- 노약자 등이 불편한 ARS 다단계 상담이 아닌, 상담원 직접응대 서비스 제공

○ 국민신문고

- 국가권익위원회에서는 행정기관 등의 불합리한 정책 등으로 국민의 권리 및 이익이 침해되지 않도록 제도나 관행을 고발하고, 각종 정책에 의사결정과정에 의견을 제시할 수 있는 체계를 구성하여 운영하고 있으며, 다음과 같은 서비스를 제공함
 - 일원화된 온라인 창구를 통해 민원·국민제안을 처리하고, 정책토론헤를 위한 정책토론, 시민클럽시스템 구축·운영
 - 공무원제안규정에 의한 공무원제안시스템과 제도개선과제의 효율적 관리를 위한 통계 및 제도개선시스템 구축·운영
 - 모든 중앙부처·지자체 및 주요 공공기관 통합·연계 서비스 제공
 - 범정부 민원·정책Q&A 서비스 제공
 - 범정부 예산낭비신고센터 통합 구축·운영

(5) 교통사고 개선 활용과제 정보시스템

○ 교통안전정보관리체계(TAAS)

- 교통안전정보관리체계는 경찰청, 손해보험협회, 공제조합으로부터 수집된 교통사고자료를 통합하여 도로교통사고 분석 정보 및 기초통계자료를 제공함
- 교통안전정보관리체계 내 웹서비스 시스템은 교통사고 정보를 GIS기반으로 제공하는 "교통사고 GIS검색 시스템"을 통해 교통사고주제도 서비스, 교통사고 사례, 사망사고, 지역별/경찰서별/노선별 사고 검색, 심층 분석 서비스를 제공하며, GIS 기반의 정보 시각화를 통하여 사용자 편의성을 강화하였으며, 경찰청에서 입력한 교통사고DB를 기반으로 다양한 관점별 통계정보를 제공함
- 또한 교통사고 통합DB 구성을 위하여 경찰청, 손해보험협회, 공제조합으로부터 수집된 데이터에 대한 정제작업을 수행하고, 정제작업이 완료된 원천자료를 이용하여 통계생성 및 분석을 수행하기 위한 Data Mart를 구성함

○ 교통경찰업무관리시스템

- 교통경찰업무관리시스템은 교통경찰전산업무 중 운전면허·행정처분, 교통단속, 교통사고, 자동차운전학원, 교통업무정기통계, 민원접수업무를 Web 환경으로 전부 통합 운영하고 있으며, 교통사고와 관련된 데이터(교통사고 접수 대장)를 입력할 수 있는 기능이 제공되고 있음
- 교통경찰업무관리시스템의 DB는 교통사고기본정보, 현장상황정보, 사고유발원인 정보, 사상자정보, 사고관련차량정보, 해당사고와 직결된 사상자 행동유형 정보로 구성되어 있음

라) 범정부 클라우드 현황 분석

(1) 시스템 개요

- G-클라우드는 정부통합전산센터의 공동 활용형 정보자원을 필요한 만큼 신속하게 제공하는 기술 및 서비스로 정의되며 주요 추진전략으로는 서비스 신속화, 자원 공유화, 관리체계 자동화와 기술 표준화 등이 있음

| [시스템 개요] | |
|-------------------------------------|---|
| 개요 | <ul style="list-style-type: none"> · 중앙 행정기관의 스마트 전자정부 서비스를 위해 정부통합전산센터의 공동 활용형 정보자원을 필요한 만큼 신속하게 제공하는 기술 및 서비스 |
| Keyword | 추진전략 |
| 서비스 신속화 (Service Agility) | <ul style="list-style-type: none"> · 단순화·표준화 기반의 계층형 아키텍처(Layered Architecture) 구현 · 국가정보통신망 내 G-클라우드 서비스의 온라인 제공 |
| 자원 공유화 (Resource Sharing) | <ul style="list-style-type: none"> · 가상화 된 정보자원을 기반으로 협업 행정을 위한 공통 플랫폼 제공 · 부처 공통 업무를 대상으로 멀티테넌시 기반 서비스 추진 |
| 관리체계 자동화 (Management Automation) | <ul style="list-style-type: none"> · 통합센터 기존 운영관리체계를 클라우드 기반으로 고도화 전환 · Ready-Made 서비스 제공을 위한 클라우드 관리 도구 도입 |
| 기술 표준화 (Technology Standardization) | <ul style="list-style-type: none"> · 저비용 고효율 구조의 범용 인프라(x86-공개 SW) 도입 장려 · 공공분야 표준형 인프라 구조 설계 및 맞춤형 주문 제작 |

(2) 범정부 클라우드 추진현황

- '09.12월, 범정부 클라우드 활성화 추진계획 발표를 기점으로, 관계부처 합동 클라우드 종합계획수립('10년~), 중앙부처정보 자원통합추진('08~'11년), 공공 부문 클라우드 활용범위 확대('12년~)에 이르기까지 범정부적으로 지속적으로 추진중에 있음

[법정부 클라우드 추진현황]

| 구분 | 구축내용 |
|------------------------------|--|
| 관계부처 합동 클라우드 종합계획 수립 및 지속 협력 | <ul style="list-style-type: none"> 클라우드 산업발전을 위한 '법정부 클라우드 활성화 추진계획' 발표('09.12월) '클라우드 확산 및 경쟁력 강화 전략' 경제정책조정회의 보고('11.5월) 종합계획 근거 부처별 과제 추진, 분기별 클라우드 정책협의회 개최, 추진상황 상호 점검, 현안검토, 협력과제 도출('10년~) 행안부(정부통합전산센터)·지경부(NIPA,ETRI) 간 MOU 체결, 신기술 개발·적용 협력('11년~) |
| 정부 클라우드 추진방향 마련 | <ul style="list-style-type: none"> 정부의 클라우드 도입을 위한 '클라우드 기반의 법정부 IT 거버넌스 추진계획' 수립 및 정보화전략위원회 보고('11.6월) |
| 중앙부처 정보자원 통합의 성공적 추진 | <ul style="list-style-type: none"> 중앙부처 정보자원 통합에 클라우드 기술 도입 및 구축비용 30% 절감 '08~'11년 정부통합전산센터가 관리하는 부처 업무시스템 328개 HW 통합 구축으로 개별 구축 대비 30%(1,023억원) 절감 정부통합전산센터에 자원 할당·회수 자동화 기능을 갖춘 'G-클라우드' 인프라 서비스 및 공통 개발 플랫폼서비스 시범 구축 |
| 장소 제약 없는 클라우드 사무환경 시범도입 | <ul style="list-style-type: none"> 개인 PC의 HW-SW 자료를 모두 통합·사용하는 클라우드 PC 방식 사무환경 행안부 시범 적용('11.12월~) 표준 기술모델, 기능요건, 고려사항 등을 정의한 '행정기관 클라우드 사무환경 도입 가이드라인' 마련('12.5월) |
| 공공부문 클라우드 활용범위 확대 | <ul style="list-style-type: none"> 부처별로 고유 업무에 대한 클라우드 활용·검토·추진 본격화('12년~) ※ 스마트교육·디지털교과서(교과부), 디지털콘텐츠 활성화(문화부) |

(3) 법정부 클라우드 추진방향

법정부 클라우드는 '정부 클라우드의 World Best Practice'를 비전으로 하고 정보화 신규투자 여력확보, 일 잘하는 정부구현, 공공서비스 품질제고, 클라우드 분야 경쟁력 강화를 목표로 추진되고 있음

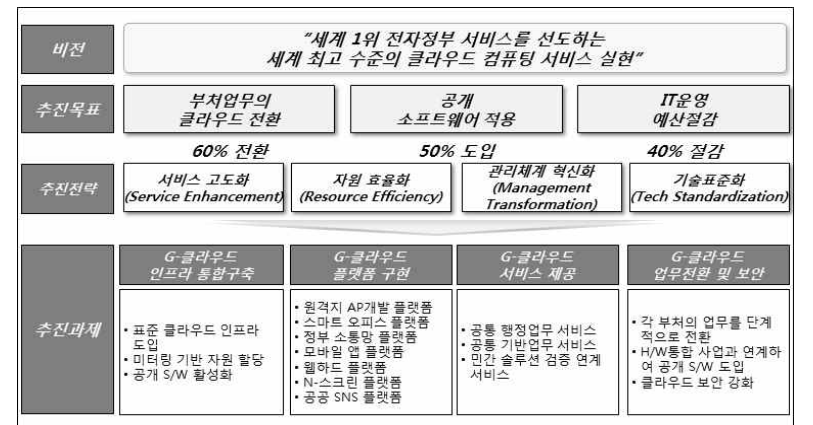
[법정부 클라우드 추진방향]



(4) G-클라우드 추진방향

- 정부통합전산센터가 추진중인 G-클라우드는 법정부 클라우드 추진방향에 기초하여 부처업무의 클라우드 전환, 공개 소프트웨어 적용, IT 운영예산 절감을 목표로 G-클라우드 인프라 통합구축, G-클라우드 플랫폼 구현, G-클라우드 서비스 제공, G-클라우드 업무전환 및 보안을 주요 과제로 하여 추진중임

[G-클라우드 비전 및 목표]



- G-클라우드드는 정부부처의 인프라, 플랫폼, 서비스를 클라우드 기반으로 조성하고 이를 공통적으로 활용하여 체계적이고 안전하게 관리하는 것을 목표모델로 추구하고 있음

[G-클라우드 목표모델]



- '12년까지 PaaS, IaaS를 구축완료하고 '13년 현재, SaaS 구축, 클라우드 보안체계 강화, IT서비스 관리 정립, 중소기업서비스 지원, 클라우드 재해복구체계 구축 등을 수행중이며 향후 공동 SaaS 확대, 공공-민간 협업체계 조성, 글로벌 IT HUB 등을 과업을 수행하는 방향으로 G-클라우드 인프라 제공 로드맵을 구성하고 있음

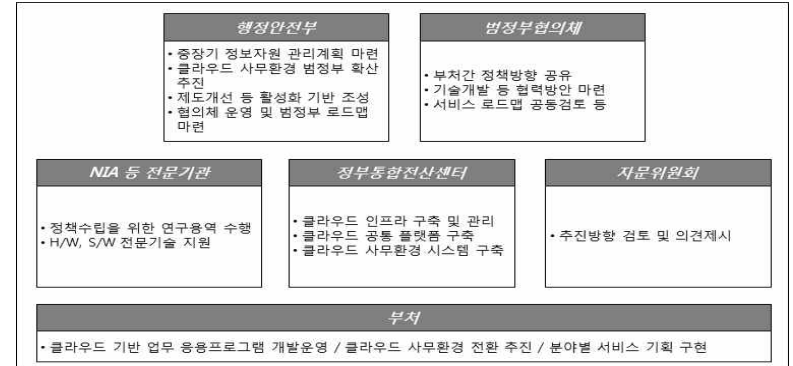
[G-클라우드 인프라 제공 로드맵]



(5) G-클라우드 추진체계

- G-클라우드 추진체계는 범정부 클라우드 로드맵 등을 마련하는 행정안전부의 주도하에 정부통합전산센터가 인프라 구축 및 관리, 공동 플랫폼 구축, 사무환경 시스템 구축 등의 업무를 수행하며 범정부협의체를 통하여 부처간 정책방향 공유, 기술개발 등 협력방안 마련 등의 협의를 수행하도록 구성되어 있음

[G-클라우드 추진체계]



(6) G-클라우드 업무적용 기준

- G-클라우드의 인프라 환경은 WEB/WAS 등 중소형 업무에 대해서는 x86기반 저가형 블레이드 서버, Linux/Windows OS 기반으로 구성하고 DB 등 대형업무에 대해서는 UNIX 기반 고가서버를 이용하여 구성하는 것을 원칙으로 함

[G-클라우드 업무적용 기준]



(7) 인프라 운영현황

○ 서비스 특징 및 범위

- '13년 현재, 주로 제공되는 G-클라우드 서비스는 개발환경 제공을 위한 PaaS와 가상머신, 방화벽 등의 인프라 기반을 제공하는 IaaS이며 서비스 제공시간 단축, 시스템 가용성 향상, 시스템 확장성 보장, 고객기관 최적 S/W stack 제공이 특징임

[서비스 특징 및 범위]

| 서비스 특징 | 서비스 범위 |
|--|---|
| 1. 서비스 제공시간 단축 <ul style="list-style-type: none"> • H/W 사양, S/W 버전 별 템플릿 제공 • 서버 신청/승인/생성 1일 내 가능 | 1. Platform As A Service <ul style="list-style-type: none"> • 개발 Language 제공 • 개발환경을 위한 시스템 SW 스택 제공 • 실행환경 Run Time 모듈 제공 • 개발/테스트용 템플릿 이미지 제공 • 전자정부 표준 프레임워크 제공 |
| 2. 시스템 가용성 향상 <ul style="list-style-type: none"> • 서버의 논리적, 물리적 클러스터 구성 • 서버 장애 시 자동 Fail-Over 기능 | 2. Infrastructure As A Service <ul style="list-style-type: none"> • 가상머신(Virtual Machine) 제공 • 공개 및 국산S/W Stack 제공(OS,WEB,WAS,DBMS) • 데이터 저장용 스토리지 및 VTL 기반 백업 제공 • 방화벽, 로드밸런싱(L4) 인프라 제공 • 기관별 독립 Virtual Network 제공 |
| 3. 시스템 확장성 보장 <ul style="list-style-type: none"> • 서버 폭주 시 서버(가상머신) 복제 및 수평확장 인프라 제공 | |
| 4. 고객기관 최적 S/W stack 제공 <ul style="list-style-type: none"> • 다양한 고객기관 요구사항 수용이 가능한 서비스 카탈로그 제공 | |

○ x86, UNIX 기반 G-클라우드 제공 인프라

- G-클라우드의 인프라는 가상화 구성 및 자동화 구성을 기본으로 하며 x86환경을 우선 적용하고 1등급 업무시스템 및 2등급 DB시스템에 대해서는 UNIX 환경으로 이관하는 것을 원칙으로 함

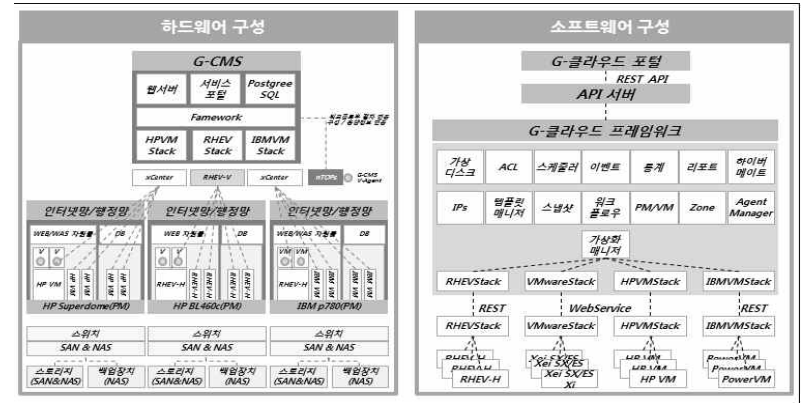
[x86, UNIX 기반 G-클라우드 인프라]

| 인프라 구성 | 적용방안 | | | | | | | | | | | | | | | |
|---|---|----------|--------|-------|-------|--------|--------|--------|--------|---------------|---------------|----------|--------|-------|--------|---|
| 시스템 구성 <table border="1"> <tr> <td>공개 S/W</td> <td>상용 S/W</td> <td>상용 S/W</td> </tr> <tr> <td>X86서버</td> <td>X86서버</td> <td>UNIX서버</td> </tr> <tr> <td>하이퍼바이저</td> <td>하이퍼바이저</td> <td>하이퍼바이저</td> </tr> <tr> <td>Linux/Windows</td> <td>Linux/Windows</td> <td>UNIX O/S</td> </tr> <tr> <td>X86 환경</td> <td>X86환경</td> <td>UNIX환경</td> </tr> </table> | 공개 S/W | 상용 S/W | 상용 S/W | X86서버 | X86서버 | UNIX서버 | 하이퍼바이저 | 하이퍼바이저 | 하이퍼바이저 | Linux/Windows | Linux/Windows | UNIX O/S | X86 환경 | X86환경 | UNIX환경 | 신규시스템 적용원칙 <ul style="list-style-type: none"> • 클라우드 HW 인프라 우선 적용 • X86, UNIX 환경 중, x86 환경 우선 적용 • 긴급 자원 요청 시 x86 환경 적용 |
| 공개 S/W | 상용 S/W | 상용 S/W | | | | | | | | | | | | | | |
| X86서버 | X86서버 | UNIX서버 | | | | | | | | | | | | | | |
| 하이퍼바이저 | 하이퍼바이저 | 하이퍼바이저 | | | | | | | | | | | | | | |
| Linux/Windows | Linux/Windows | UNIX O/S | | | | | | | | | | | | | | |
| X86 환경 | X86환경 | UNIX환경 | | | | | | | | | | | | | | |
| 가상화 구성 <ul style="list-style-type: none"> • 인프라 구성 표준화 및 등적 자원 공유화 | 노후대체제 적용원칙 <ul style="list-style-type: none"> • 3등급 업무 시스템 x86 환경으로 이관 • 2등급 업무 중 WEB/WAS시스템을 x86환경으로 이관 • 1등급 업무 시스템 및 2등급 DB시스템은 UNIX 환경으로 이관 | | | | | | | | | | | | | | | |
| 자동화 구성 <ul style="list-style-type: none"> • 미터링/자원할당 및 회수자동화 구성 | | | | | | | | | | | | | | | | |

○ G-클라우드 관리환경

- G-클라우드의 관리환경은 G-CMS(Cloud Management System)에서 REST API 방식으로 x86 기반의 REHV Stack, Vmware stack, UNIX 기반의 HP VM stack, IBM stack 을 통합관리하며 대쉬보드인 nTOPS와 실시간 연계하여 사용자에게 신속하고 정확하게 클라우드 자원 현황에 대한 정보를 제공함

[G-클라우드 관리환경]



○ 원격지 AP 개발 플랫폼 및 보안

- G-클라우드는 원격지AP개발 플랫폼을 통해 센터에 개발환경 제공을 요청하고 센터는 기술컴포넌트, 멀티테넌시 개발 가이드, PaaS 이미지 구성 및 제공 등의 업무를 수행함. G-클라우드는 8방어 4분석의 인증된 보안체계를 구성하여 클라우드 자원에 대하여 보안을 적용하고 있음.

○ 주요 이슈 및 현황(인터뷰 결과)

- 인터뷰 결과, 정부통합전산센터로 '13년 신규 입주되는 빅데이터 공통기반은 G-클라우드 적용대상이 아니며 시스템 설계 시 공공기관-민간과의 데이터 연계 협의, 권한관리체계 구성, 시스템 확장성 등을 고려하여 시스템을 구축해야 하는 것으로 파악됨

[G-클라우드 관리환경]

| 구분 | 질의사항 | 주요 현황 및 이슈 |
|------|----------------|---|
| 일반사항 | G-클라우드 추진현황 | • IaaS, PaaS, DaaS를 구축 운영중이며SaaS는 계획 수립 단계임 |
| | 빅데이터 분석 플랫폼 구축 | • 47개 부처 2만여 서버에 대하여 장애 및 사이버 침해사고 발생시 AP에 대한 정탐, 해킹추적에 대한 성능지표 추적탐지를 위해 1일 기준 DVD 800장 |

| 구분 | 질의사항 | 주요 현황 및 이슈 |
|--------------|---------------------|---|
| | | 분량의 반정형 로그데이터를 map-reduce 기술을 활용하여 분석할 수 있는 빅데이터 분석 플랫폼을 구축하는 계획을 수립중임 |
| 범정부 클라우드 활용 | G-클라우드 전환시 효과 | <ul style="list-style-type: none"> 자원할당 용이성, 신규 구축시 구축/운영비용 절감, 운영안정성 확보 |
| | G-클라우드 서비스 이용 가능 여부 | <ul style="list-style-type: none"> 현재 센터내 입주되어 운영중인 장비에 대하여 x86 기반으로 클라우드 플랫폼 통합 운영중이나 신규 입주시스템은 서비스 대상이 아님 신규 입주시스템의 경우, 주관부처에서 예산 수립을 담당하고 센터는 H/W만 위임발주 함 네트워크 구성의 경우, 센터 백본을 제외한 내부 네트워크 자원산정은 입주 부처에서 해야 함 |
| 빅데이터 공통기반 마련 | 빅데이터 개발환경 | <ul style="list-style-type: none"> 빅데이터 개발환경은 현재 표준화되어 있지 않았으며 신규 구축되는 범정부 빅데이터 공통기반의 경우 센터의 클라우드 서비스 대상이 아니므로 자체 개발이 가능함 |
| | 클라우드 기반 구축시 고려사항 | <ul style="list-style-type: none"> 센터 입주 시, 보안은 센터의 8방어 4분석 방어체계(e-ANSISUNG) 기준을 수용해야 함 민간정보 연계 시 법적 근거를 확보해야 함 정보 보유기관과의 협의, 제도개선 등이 선행되어야 하고 접근권한관리, 행정망과 고립망간 데이터 연계, 시스템 확장성을 고려한 Scale-out 방식 H/W 설계 등도 고려되어야 함 |

4) 법제도 현황 분석

가) 개요

(1) 법제도 개선 추진 목적

- 빅데이터 등 대규모 정보자원의 효율적 활용을 위한 법제도적 기반을 마련하기 위하여, 현행법령 상의 장애요인을 분석·식별하고, 정책추진을 위한 근거를 마련하는 것을 목적으로 함

법제도 개선의 추진 목적

| | |
|-----------------------|---|
| 현행 법령들과의 모순과 충돌을 극복 | <ul style="list-style-type: none"> 빅데이터 활용과 관련된 정보자원의 관리·활용에 대한 정보화 관련 법령의 현황 및 문제점 분석 행정조사기본법, 저작권법, 국유재산법, 민원사무처리에 관한 법률 등 관계 법령에 대한 문제점 분석 및 시사점 도출 빅데이터 마스터 플랜상의 6대 분야 16개 과제에 대한 개별법령 등을 분석하여 빅데이터 활용가능성 검토 |
| 빅데이터 활용을 위한 정책추진 근거마련 | <ul style="list-style-type: none"> 행정기관 등이 수집·보유하는 대규모 정보자원을 활용하기 위한 법제도적 근거를 마련 빅데이터 활용을 위한 추진체계, 활용절차, 활용정보의 범위와 한계 등의 점검 등 실용성 있는 제도적 기반 마련 |
| 개인정보의 안정적 보호와 활용방안 도출 | <ul style="list-style-type: none"> 개인정보 보호법 등을 분석하여, 빅데이터 활용 과정에서 발생할 수 있는 법제도적 이슈와 시사점 도출 민간정보, 비민간정보, 식별정보, 결합정보 등 개인정보의 유형과 속성을 기준으로 개인정보의 안정적 보호를 도모하고, 공정하고 적법하게 개인정보를 활용할 수 있는 방안을 모색 |

(2) 분석범위 및 방법

- 빅데이터 활용을 위한 관련 법령의 분석범위·유형·방법을 제시하여, 법제도 수립 절차의 객관성과 전문성을 확보

분석범위 및 분석방법 도출

| | | |
|------|---|---|
| 분석목적 | <ul style="list-style-type: none"> 국가차원의 빅데이터 활용을 위한 법제도 개선방안을 도출 현행법령의 모순과 충돌을 극복하고, 실용성 있는 제도적 기반마련 | |
| 분석범위 | 정보화 관련 법령의 분석 | <ul style="list-style-type: none"> 정보사회, 지식기반사회, 빅데이터 활용의 현법적 의미 분석 국가정보화기본법, 전자정부법 등 정보화 법령의 현황 및 한계점 분석 |
| | 행정업무 근거법령과 정책 관련 법령 분석 | <ul style="list-style-type: none"> 행정조사기본법, 민원사무처리에 관한 법률, 행정업무의 효율적 운영에 관한 규정 등 빅데이터 활용과 관련된 행정권 발동의 수권 법령 분석 저작권법, 국유재산법, 정부조직법 등 빅데이터 활용과 관련된 정책 근거 법령 분석 개인정보 보호법, 전자정부법 상의 행정정보 공동이용 및 연계 관련 규정 |
| | 정책과제 관련 법령분석 | <ul style="list-style-type: none"> 빅데이터 마스터플랜상의 6대 분야 16개 과제관련 법령분석 |
| 분석방법 | <ul style="list-style-type: none"> 정보화 및 빅데이터 등과 관련된 논문과 문헌 등을 참고 분석하여, 법제도 개선방안 수립절차의 객관성과 전문성을 확보하고, 빅데이터 활용을 위한 제도적 기반의 타당성을 확보 정보화 법령 및 유사 입법례 등을 분석하여, 실효성있고 실행용이한 법제도 개선방안 도출 개인정보의 보호와 관련된 논문과 문헌, 보고서 등을 참고하여, 안정적인 개인정보 보호체계 구축 | |

나) 정보화법령 분석

(1) 헌법상 정보화사회의 의미

- 헌법규정은 정치적·이념적·프로그램적 성격을 지니는 것으로 평가되며, 법률의 위헌여부, 공권력의 행사 또는 불행사로 인한 기본권 침해 등을 판단하기 위한 심사기준 및 재판기준으로 활용됨
- 헌법은 기본권 실현을 위하여 다양한 의무를 국가에 부과하고 있으나, 헌법 규정 자체는 정책집행의 직접적 근거로 기능하지 못하며, 국회를 통한 법률의 형식으로 비로소 행정작용의 법적 근거(행정의 법률유보의 원리)로 기능하게 됨
- 국가정보화와 전자정부는 국가의 최종적 목표인 기본권 실현을 위한 수단으로서 기능하고 있음

(2) 국가정보화기본법

- 국가정보화기본법은 기본이념, 방향, 원칙과 정책실행에 필요한 추진체계 등에 규율하고 있으며, 방침적·선언적 효력만을 가질 뿐, 행정기관 등이 보유하는 정보와 데이터를 연계·융합하기 위한 법적 근거로 기능하지 못하는 한계를 지님
- 국가정보화기본법상의 정보 공유·유통·관리·활용 등에 대한 규정이 존재하나, 문리적 해석상 지식정보(자료 등)에 범위가 한정되어 있음

(3) 전자정부법

- 전자정부법은 전자정부의 도입 및 촉진과 행정업무의 전자화와 이와 관련한 법적 효력 및 행정정보 공동이용 등에 대하여 규율하고 있으나, ICT분야의 최신 기술 동향인 빅데이터, 클라우드 서비스에 대한 규정은 충분히 담아내지 못한 상태

다) 행정업무 관련 법령 분석

(1) 행정조사기본법

- 행정조사기본법은 국가정책실현을 위한 자료조사 및 정보수집에 적용되는 기본법이나, 최소수집, 목적구속, 예방중심, 중복조사금지 등 빅데이터의 일반적 속성과 배치되는 법원리가 적용
- 따라서, 법제도 개선 시, 동법의 적용을 배제시키거나 우선 적용될 수 있는 특별법적 효력을 부여하여, 법률간의 모순과 상충을 해소시킬 필요
- 정책결정에 활용하기 위한 민간에서 작성되는 정보에 대한 수집, 저장 등에 대한 법적 근거가 없음

(2) 정부조직법

- 정부조직법상 행정기관간 독점적·배타적인 권한부여로 인해 협업과 공유가 어렵다는 특징. 따라서, 대규모 정보자원의 공유활동 등에 대한 일반적 근거와 세부절차 등을 정비하여, 빅데이터 활용기반을 구축할 필요
- 정보자산의 재산권적 측면에서는 기획재정부가 총괄청으로 컨트롤타워역할을 하므로, 행정안전부 중심의 추진체계를 새로 정립할 필요

(3) 저작권법

- 저작물인 정보자원을 빅데이터에 분석하는 경우, 개별적 사용승인을 얻는 것은 사실상 불가능하므로, 공정이용법리와 같이 공익적 정책목적을 수행하기 위해 정당한 이용으로 간주할 수 있는 법적 근거 필요

(4) 개인정보보호법

- 개인정보의 익명화, 식별불가능화, 분리를 통한 이용 등 비교적 활용이 용이한 제도적 근거를 두고, 빅데이터 활용 시 특정개인에 대한 분석이 수반되는 경우에는 개별적, 구체적으로 대상과 범위를 특정하여 입법화

라) 정책과제 관련 법령 분석

(1) 정책과제별 분석대상정보 및 공유활용가능성

- 정책과제별로 분석대상 정보를 나열하고 정보의 활용가능성을 검토

| 정책분야 | 대상과제 | 분석대상 정보 | 정보의 활용가능성 |
|-----------------|-----------------|----------------------------|--|
| 범죄발생 최소화 | 범죄경력정보, 수사경력정보 | 범죄정보 | 범죄정보 등은 재판, 수사, 형집행 등의 목적으로 최소한도로 이용 |
| | | 신고정보, 주민신고정보 | 신고자 보호를 위해 가명허용, 신고인에 대한 공개 및 보도 금지 |
| | | 공간정보 | 국민들이 쉽게 접근하고 이용하도록 정보의 보급 및 활성화 |
| | 인구 등 통계정보 | 통계정보는 공공재로서 널리 이용, 보급되어야 함 | |
| 예측기반 자연재해 조기감지 | 기상정보 | 기상정보 | 기상정보는 보급 및 이용활성화 대상 |
| | | 재난정보 | 정부는 재난예방을 위하여 필요한 정보를 수집·보유·공동이용할 수 있음 |
| 국민복지 | 민원정책 환류시스템 | 민원정보 | 민원은 개인의 권리·의무와 관계된 정보이며, 개인정보로서 취급 |
| | | 복지목적의 정보연계 | 추상적인 복지를 목표로 개인정보 취합불가, 별도의 법적 근거 필요 |
| | 복지 수요공급 맞춤형 서비스 | 민원정보 | 상위 동일 |
| | | 맞춤형 서비스의 개념설정 | 국가급부행정의 작동을 위하여 현행법상으로는 신형개념 필요 |
| 일자리 분석예측 고용정책지원 | 고용정보, 노동정보 | 고용정책시행 | 고용정책시행을 위하여 국가는 필요한 정보를 수집하고 조사·공표 |
| | | 노동정보 | 구직, 구인을 신속히 연결하기 위한 직업 현황 및 전망정보를 수집관리 |

| | | | | |
|------------------------|-----------------------------|--|---|--|
| | | | 경제, 산업 동향과 전망 등의 인력수급 정보를 작성 및 공표 | |
| | | 관련정보의 연계 취합 | 고용안정정보망 등의 구축 및 운영 및 정보시스템의 연계 | |
| 국가경제 | 과세데이터분석 탈세방지 | 납세자, 과세현황, 납부이력 | 국세기본법상 중복조사 금지원칙의 문제 발생 우려 | |
| | | 소비·지출관련 비정형데이터 | 개인정보의 관련성 및 민감도에 따라 활용가능성의 차이발생 | |
| | 과세정보에 대한 통계자료 | 납세자의 과세정보를 직·간접적으로 확인하거나 추정할 수 없도록 작성 | | |
| | 경제데이터분석기 반 정책수립 | 무역통상관련 정보 | 지자체나 기업 등에 정보제공, 정보유통·교환체계 구축 | |
| | | 무역거래 관련 통계 | 무역 관련 정보 및 통계 수집, 유통촉진 및 공급 활용 | |
| 물가정보 | | 사업자의 원가 및 영업비밀을 보장하는 한도 내에서 이용 | | |
| 생활만족도, 행복지수 등 | 민간기구에서 조사·작성·발간하는 정보로서 이용가능 | | | |
| 국가 인프라 | 주민참여 교통사고 감소체계 | 교통정보 | 교통조사실시가능, DB구축 및 정기적 발행·공표, 정책자료로 활용 | |
| | | 보험정보 | 개인정보 누설 및 목적외사용 금지, 식별불가능한 통계자료는 이용가능 | |
| | | 의료정보, 진료기록 | 의료법상 열람교부사유를 14가지로 한정하여 열거 | |
| | 실시간 네트워크 재난관리 | 네트워크 사용량 | 정보통신망의 안정성·신뢰성 제고를 위하여 기술·기기에 대한 정보관리 | |
| | | 장비동작정보, 기술및 시장정보 | 정보통신망 관련기술을 체계적·종합적으로 관리·보급, DB구축·제공 | |
| | | 전력공급정보 | 지능형 전력망 관리를 위한 통계를 작성·관리·공개 | |
| | | 침해정보, 사고조사정보 | 침해사고의 원인 분석 및 대책마련 목적 내에서만 사용, 분석 후 파기 | |
| | 산업지원 | 자영업자 창업실패 예방 | 중소기업, 소상공인, 전통시장, 벤처, 등의 경영·금융 등 창업관련 정보 | 각 분야별 창업지원 및 육성을 위한 법령 등에서는 지원정책을 위한 실태조사, 통계작성, 관련자료 등을 수집·분석하여 제공 |
| | | | 산업기술정보 | 기술·자원활용 극대화와 기술혁신 촉진을 위하여 관련정보를 수집·분석 |
| | | 제조공정 장애예측 | 부품 및 소재 정보 | 부품, 소재, 생산설비 등의 정보를 분석·가공하여 DB구축 및 통계작성 |
| 자원수급 및 공급정보 | | | 산업 수요·공급에 대한 실태조사와 통계작성 및 제공 | |
| 수급전망기반 농수산물 생산관리 | | 농수산물생산정보 | 농수산물의 재고물량, 소비동향, 해외시장 정보등을 조사·분석하여 공표 | |
| | | 농수산물유통정보 | 국가는 유통정보시스템을 구축·운영하거나 지원 | |
| | | 농수산물품질정보 | 생산, 유통, 소비 등에 대한 통계조사 실시 | |
| | | 물류표준화 및 이력정보 | 물류표준화 및 이력추적관리를 통해 품질관리 시행 | |
| | | 농수산 관련 통상 및 국제정보 | 공공기관을 통한 수집, 공개 및 통계작성 등 | |
| 과학기술 | | 국가기후 위험요소 선제적 대응 | 기상정보 | 공공재로서 기상정보의 예보, 특보, 공고·발표 등 이용·보급 촉진 |
| | 기상정보의 연계 및 공동활용 | | 각종 기상정보를 연계 공동활용하고 정보시스템 구축·운영, 국제협력 추진 | |
| | 기상관측자료 | | 전국적 관측망구성, 상호교환 및 공동활용 | |
| | 유전자 의료정보 분석 국민건강증진 | 유전자정보 | 염색체, 혈액, 세포, DNA, RNA등의 정보는 연구 및 치료목적으로 사용 | |
| | | 의료정보 및 진료기록 | 의료법상 14가지 법정 사유에 의해서만 교부·열람 | |
| 위성영상을 통한 재난대응 | 보건의료 통계 | 보건의료에 대한 통계등을 작성하여 이용 및 보급 | | |
| | 위성정보 | 개인의 사생활을 보호하는 범위 내에서 위성정보를 보급·활용 촉진 | | |
| | 기상위성정보 | 기상위성 관측망을 구축·운영하고 관련정보를 수집·활용 | | |
| | 재난정보 | 정부는 재난예방을 위하여 필요한 정보를 수집·보유·공동·이용할 수 있음 | | |

(2) 활용가능성에 따른 유형별 분류

○ 빅데이터 활용가능성이 높은 유형의 정보와 관련 법제

| 정보활용 유형 및 특징 | 대상유형 | 관련법제 | 제약내용 |
|---|---------------------------|-----------------------------------|--|
| 활용가능성이 높은정보 <small>(관련 법령에서 공개, 공표, 발표하거나 이용 및 보급을 촉진하는 정보유형)</small> | 공간정보 | 국가공간정보에 관한 법률, 공간정보산업진흥법 | 법령에서 수집·작성된 정보를 공개, 공표, 발표 하는 등 이용·촉진 및 보급·활성화를 규율하고 있는 정보 |
| | 인구, 거주, 경제 등 통계정보 | 통계법 | |
| | 기상, 기후, 관측 정보 | 기상법, 기상산업진흥법, 기상관측표준화법 | |
| | 태풍, 홍수, 붕괴, 폭발, 오염 등 재난정보 | 재난 및 안전관리 기본법, 자연재해대책법 | |
| | 고용, 노동, 인력, 직업 등 고용정보 | 고용정책기본법 | |
| | 무역, 통상, 거래 등 경제정보 | 대외무역법, 무역거래기반조성에 관한 법률 | |
| | 생활만족도, 행복지수 등의 지표 | 통계법외 민간조사기관에서 공표 및 발표 | |
| | 교통정보 | 국가통합교통체계효율화법 | |
| | 보건의료통계 | 보건의료법 | |
| | 네트워크, 장비 등 통신망 정보 | 정보통신기반 보호법, 정보통신망법 | |
| | 기회, 기술, 시장 등 시장정보 | 정보통신산업진흥법 | |
| | 전력량, 공급 및 수요 등의 정보 | 지능형전력망의 구축 및 이용촉진에 관한 법률 | |
| | 부품, 소재, 자원수급, 공급 정보 | 부품·소재기업 등의 육성에 관한 특별조치법, 산업발전법 | |
| | 산업기술혁신 통계 및 동향정보 | 산업기술혁신촉진법 | |
| | 중소기업 창업정보 | 중소기업기본법 | |
| | 소상공인 등 창업정보 | 소기업 및 소상공인 지원을 위한 특별조치법 | |
| | 전통시장 등 창업정보 | 전통시장 및 상점가 육성을 위한 특별법 | |
| | 벤처기업 등 창업정보 | 벤처기업육성에 관한 특별조치법 | |
| | 여성기업 등 창업정보 | 여성기업지원에 관한 법률 | |
| | 장애인 기업 등 창업정보 | 장애인기업활동촉진법 | |
| 1인 기업 등 창업정보 | 1인 창조기업 육성에 관한 법률 | | |
| 농수산 관련 일반 현황 정보 | 농어업·농어촌 및 식품산업기본법 | | |
| 농수산 유통 및 통상 정보 | 농수산물 유통 및 가격안정에 관한 법률 | | |
| 농어업 경영정보 | 농어업 경영체 육성 및 지원에 관한 법률 | | |
| 식품의 품질 및 수급관련 정보 | 식품산업기본법 | | |
| 농수산 물류, 이력, 품질 정보 | 농산물 품질관리법 | | |
| 수산업 유통, 어장, 수출, 통상, DB정보 | 수산업법, 수산자원관리법 | | |
| 위성영상, 음성, 데이터, 등 위성정보 | 우주개발진흥법, 기상법 | | |

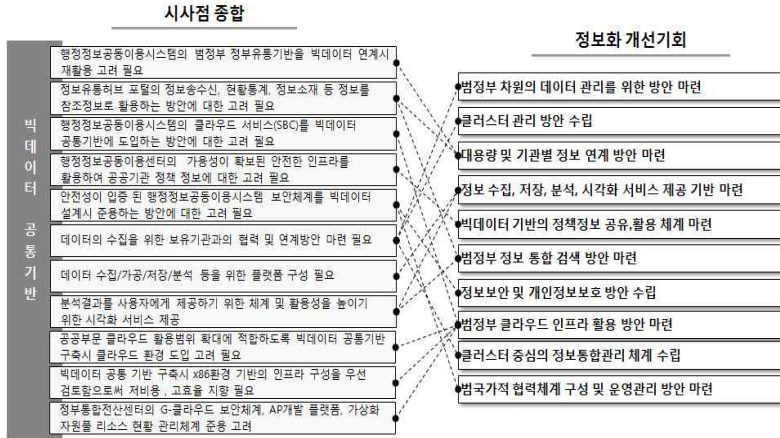
○ 빅데이터 활용이 어려운 유형과 관련 법제

| 정보활용 유형 및 특징 | 대상유형 | 관련법제 | 제약내용 |
|--|--------------------|--|--|
| 법제도개선을통해 빅데이터 활용이 가능한정보 <small>(관련 법령에서 목적외 사용을 금지하거나 별도의 법적 근거를 요하는 정보)</small> | 물가정보 | 물가안정에 관한 법률 | 물가안정정책 목적외로만 사용, 특정사업 자에 대한 원가 및 영업비밀 등을 보호 |
| | 산업기술정보 | 산업기술의 유출방지 및 보호에 관한 법률 | 산업기술의 부정확한 유출방지 및 경쟁력 강화 |
| | 신고정보, 주민신고정보 | 공익신고자보호법, 특정범죄신고자 보호법 부패방지 및 국민권익위원 회의 설치와 운영에 관한 법률 | 신고인 정보보호, 공개/보도금지, 가명사용허용, 신변보호조치 등 |
| | 민원정보 | 민원사무처리에 관한 법률 | 민원은 국민의 권리의무와 관계된 사항으로서 개인정보에 해당 |
| | 보험정보 | 보험업법 | 개인정보 누설 금지 및 목적외 사용금지 |
| 빅데이터활용을위한법제 도개선이아려운정보 <small>(관련 법령에서 특정목적에만 사용하도록 하거나 제한적으로 열거하고 있는 경우)</small> | 통신망 침해, 사고정보 | 정보통신망 이용촉진 및 정보보호에 관한 법률 | 침해원인분석 및 예방목적으로 사용, 분석 후 파기의무 |
| | 납세자, 과세정보, 납부이력 | 국세기본법 | 중복조사금지원칙의 위반가능성 |
| | 의료정보 및 진료기록 | 의료법 | 열람·교부에 대한 14가지의 제한적 법정사유 |
| | 범죄경력 및 수사경력정보 | 형의 집행 및 실효에 관한 법률 | 수사, 재판, 형집행을 위해서만 이용 |

5) 정보화 현황분석 종합

- 행정정보공동이용센터, 정책활용과제 관련 시스템, 범정부 클라우드 현황 분석을 통해 도출된 시사점을 종합하여 목표모델 수립을 위한 정보화 개선기회를 도출함
- 주요 정보화 개선기회는 행정 및 공공데이터의 연계는 기존 행정정보공동이용시스템을 활용하고, 빅데이터 기반의 정책정보 공유,활용체계를 만드는 것임

[정보화 현황분석 종합]

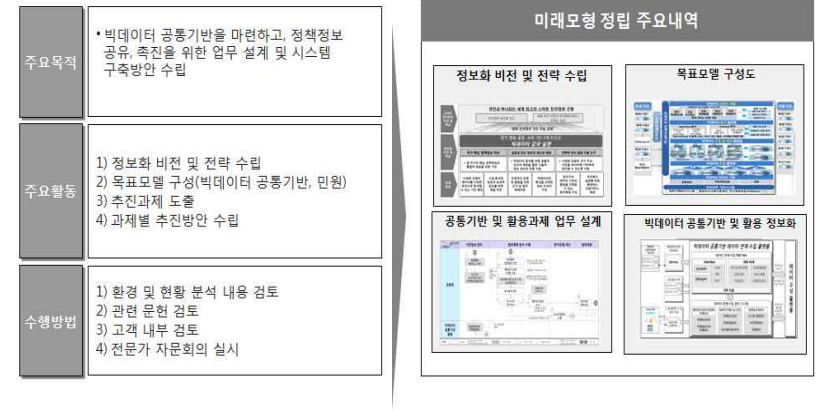


나. 미래모형 정립

1) 미래모형 정립 개요

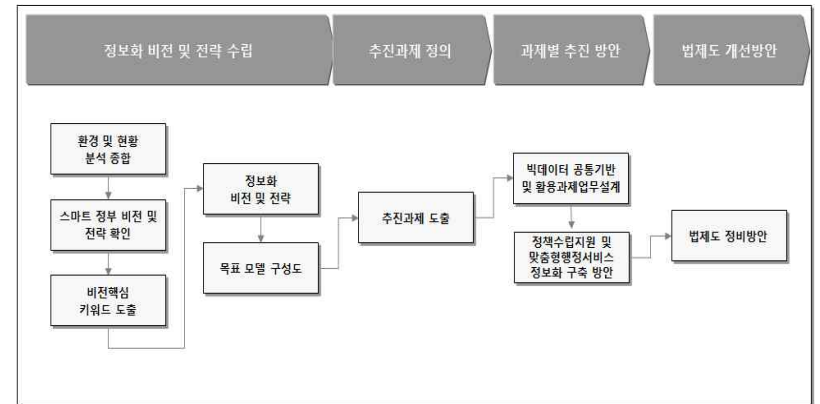
- 빅데이터 공통기반 마련 및 활용 미래모형 정립은 현황분석을 통해 도출된 설계 방향에 따른 전략적 실행을 위해 정보화 비전과 전략, 추진과제, 과제별 추진방안 그리고 법제도 개선방안을 마련하는 것임

[빅데이터 공통기반 마련 및 활용 미래모형정립 목적 및 주요 내역]



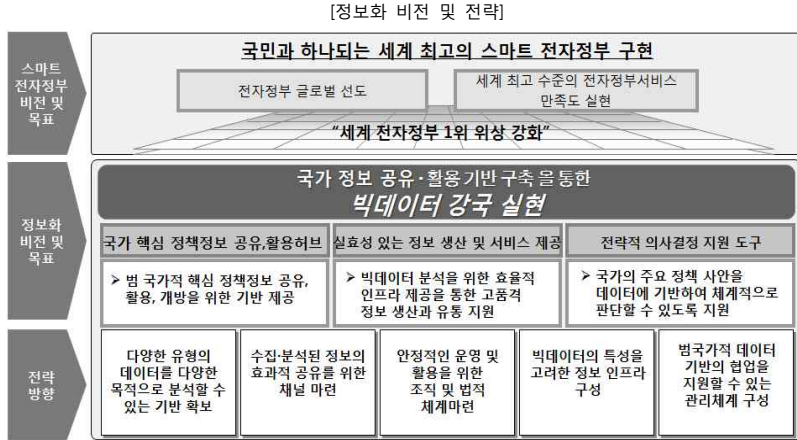
- 미래모형은 환경 및 현황분석을 통해 도출된 추진방향을 토대로 정보화 비전 및 전략을 수립하고, 추진과제를 정하여 과제별 추진방안을 수립하고, 원활한 시스템 운영을 위한 법제도 개선방안을 제시함

[빅데이터 공통기반 마련 및 활용 미래모형 정립 Framework]



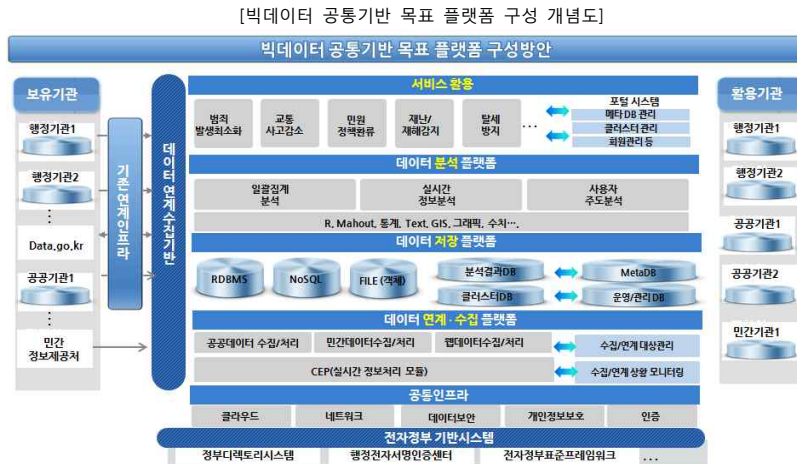
2) 비전 및 전략

- 빅데이터 공통기반 및 활용과제와 수혜자 맞춤형 행정서비스를 구축함으로써 추구하는 비전은 “빅데이터 강국실현”이며, 이를 달성하기 위한 3대목표와 5대전력방향을 제시함



3) 빅데이터 공통기반 및 활용 목표 플랫폼

- 빅데이터 공통기반 및 활용과제 구성을 위한 목표 플랫폼은 활용기관의 이용편의성을 감안하여, 데이터 연계/수집, 저장, 분석 플랫폼과 공통인프라를 구축함

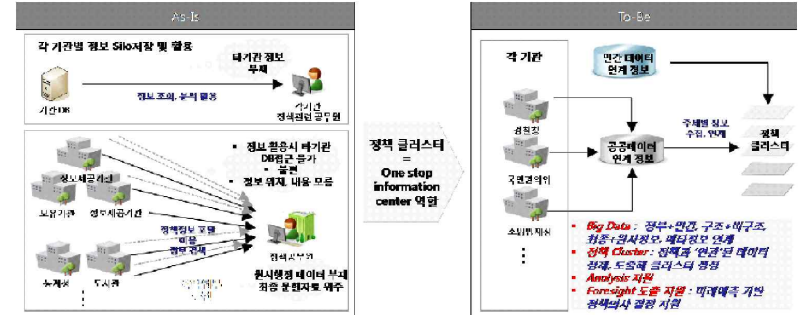


4) 빅데이터 공통기반 및 활용과제 업무 설계

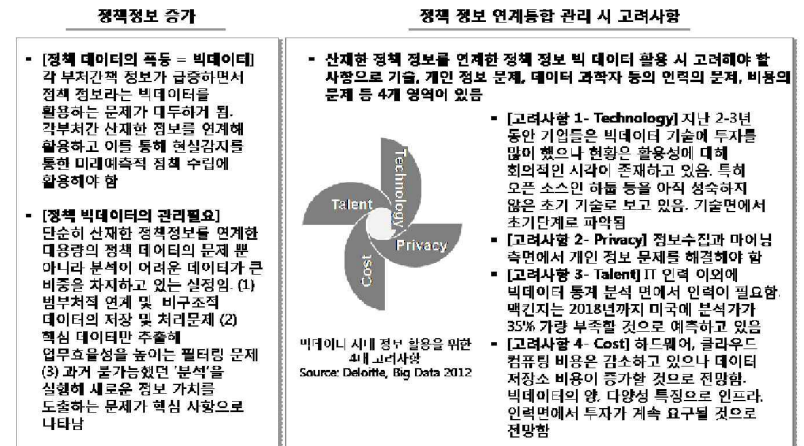
- 가) 정책정보지원 클러스터의 필요성

(1) 정책정보 증가 및 연계 통합 관리 필요

현황분석 결과에서 얻어진 시사점에서 범정부 분류체계에서 정보 활용도가 미흡한 점을 보완하고 민간 정보까지 포함해 정책 주제별로 연관성 있는 정보를 수집 연계한 정책 클러스터가 필요하며 이를 통해 정책 정보 신속한 검색 강화, 맞춤형 정보 제공과 활용이 가능함



- 정부 부처에 산재한 정책 정보가 증가하고 있는 현황에서 흩어져 있는 정보 소스들을 연계, 통합해 정보에 대한 검색, 접근을 신속하게 할 수 있도록 정책 클러스터라는 개념이 필요함



(2) 민간 데이터 활용 필요

현황분정부 시스템 외부에서 증가하는 민간 데이터를 포함한 Big Data(비구조적 데이터 등)에서 정보를 수집해 활용할 수 있도록 정책 클러스터의 개념이 필요함

○ 민간데이터 증가

- [민간 데이터의 증가 = 빅데이터] 모바일, 온라인 웹, SNS서비스 등 정부 시스템 외부에서 민간데이터가 폭발하고 있음. IDC는 2011년 생성된 정보의 양이 1.9제타바이트에서 향후 5년내 약 9배까지 증가할 것으로 예측했음. 민간 데이터를 활용해 현실에 적합한 정책 개발 및 현재보다는 미래예측에 초점을 맞춘 정책개발이 시급함

- [민간 데이터 활용 필요] (1)정부 시스템 외부의 비구조적 데이터의 폭등 중에서 이를 정책과 연관된 데이터로 검색·추출·수집하는 문제 (2)실시간 분석 (3)분석을 통해 미래예측적 시나리오를 만들고 문제의 사후 대응 해결방식에서 사전 대응 방식으로 해결하는 문제가 핵심이 됨

○ 민간데이터 활용 시 고려사항

| 민간데이터 증가 | 민간데이터 활용 시 고려사항 |
|--|---|
| <ul style="list-style-type: none"> [민간 데이터의 증가 = 빅데이터] 모바일, 온라인 웹, SNS서비스 등 정부 시스템 외부에서 민간데이터가 폭발하고 있음. IDC는 2011년 생성된 정보의 양이 1.9제타바이트에서 향후 5년내 약 9배까지 증가할 것으로 예측했음. 민간 데이터를 활용해 현실에 적합한 정책 개발 및 현재보다는 미래예측에 초점을 맞춘 정책개발이 시급함 [민간 데이터 활용 필요] (1)정부 시스템 외부의 비구조적 데이터의 폭등 중에서 이를 정책과 연관된 데이터로 검색·추출·수집하는 문제 (2)실시간 분석 (3)분석을 통해 미래예측적 시나리오를 만들고 문제의 사후 대응 해결방식에서 사전 대응 방식으로 해결하는 문제가 핵심이 됨 | <ul style="list-style-type: none"> 정책정보로서 정부 시스템 외부의 민간 데이터를 활용하기 위해서는 비구조적 데이터의 추출 및 수집, 실시간 분석의 문제, 활용 면에서 사후 대응적 문제해결이 아닌 미래예측력을 통한 사전대응적 문제해결이 중요한 핵심 사항임 <ul style="list-style-type: none"> [핵심 1- 비구조적 데이터] 모바일, 온라인, SNS 등 시스템의 정형데이터가 아닌 비정형 데이터가 폭등하고 있으며 이 대용량 정보 중에서 가치 있는 정보를 추출하는 새로운 정보관리 체계가 필요함 [핵심 2- 실시간 분석] 분석을 어떻게 하느냐가 빅데이터 전략의 핵심사항으로 부상하고 있음. 기존의 결과분석(history 분석)에서 넘어선 실시간 분석과 분석을 통한 전망, 미래예측 가능한 분석으로 초점이 옮겨가고 있음 [핵심 3- 사전대응] 활용은 사후대응의 고도화도 중요하겠지만 사전대응의 단계로 업그레이드해야 할 것으로 판단됨. 미래예측적 사고를 가지고 전조감지, 모니터링을 통한 사전대응적 해결방안을 마련해야 함 |

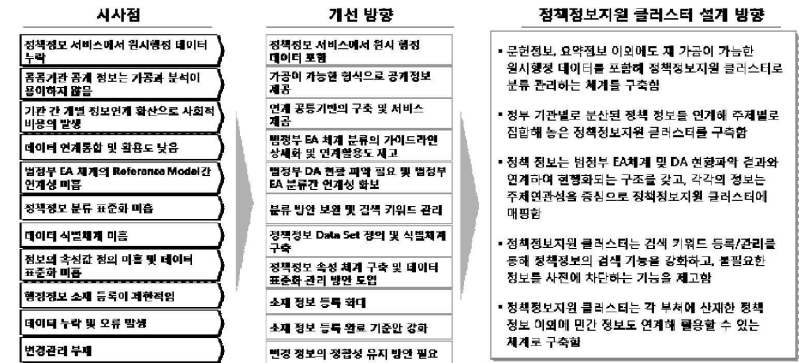
○ 고려사항

- 정보 개방, 공유를 위한 법적인 근거가 마련되어야 함
- 클러스터를 형성하기 위한 DA 체계 확립, 데이터 표준화, 데이터 연계가 필요함
- 소재등록의 정확성 제고 및 데이터 연계 활용 방안이 확립되어야 함

나) 정책정보지원 클러스터 설계

(1) 정책정보지원 클러스터 설계 방향

정책정보 지원 현황 분석에서 도출한 수집 및 활용체계, 분류체계, 소재정보 관련 시사점과 개선방향을 참조하여 다음과 같이 정책정보지원 클러스터의 설계 방향을 설정함



(2) 정책정보지원 클러스터 개요

○ 정책정보지원 클러스터의 필요성, 용도 및 목적은 다음과 같음

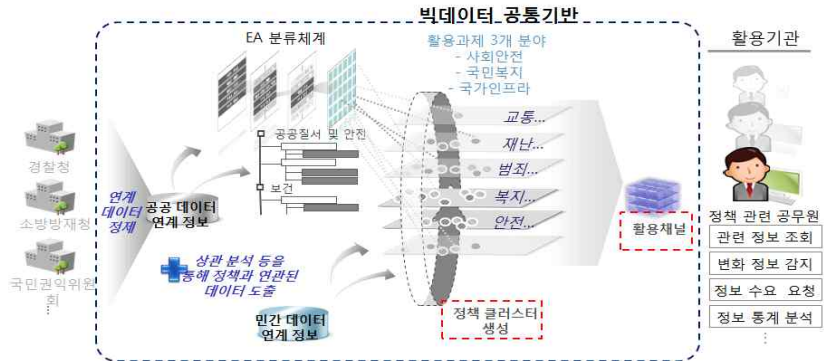
| | | |
|-----|----------------------|--|
| 필요성 | 정책정보 증가 및 연계통합 관리 필요 | 정부 부처에 산재한 정책 정보가 증가하고 있는 현황에서 핵심 데이터 연계 통합 및 필요도 제고를 위해 Data 집합의 개념이 필요함 |
| 용도 | 민간데이터 활용 필요 | 정부 시스템 외부에서 증가하는 민간데이터를 포함한 Big Data를 정책 업무에 활용하기 위해 민간 데이터의 접근 단위가 중요함 |
| | 범정부 정책 정보 활용 체계 정립 | 범정부 분류 체계에 의한 분류 정보가 데이터 정합성 통일성, 검색 미흡 등 활용 면에서 미흡한 점을 보완하기 위해 정책 주제별로 관계성이 높은 범정부 정책 정보를 연계해 클러스터로 제공함 |
| | 민간정보 포함 활용 체계 정립 | 민간의 빅데이터까지 포함해 정책 주제별로 핵심 데이터를 추출, 수집, 연계해 클러스터로 제공하고 빅데이터 공통기반이 제공하는 서비스 기능을 활용해 정책 업무 효율성을 제고함 |
| 목적 | 정보 활용도 제고 (신속성, 효율성) | 공공 및 민간 정보로 구성된 빅데이터에서 데이터 연계, 통합 및 추출을 통해 맞춤형 정책정보지원 체계를 제공함으로써 정책수립, 집행 등의 활동을 효율적으로 지원함 |
| | 정보의 신규기저 창출 | 정책 정보를 주제별로 연관된 형태로 제공함으로써 단독 정보를 통해 얻을 수 없었던 융합된 정보 가치를 신규기저로 창출하고, 이를 통해 정책 수립 등 정책관련 의사결정 업무를 효율적으로 지원함 |

○ 정책정보지원 클러스터의 정의

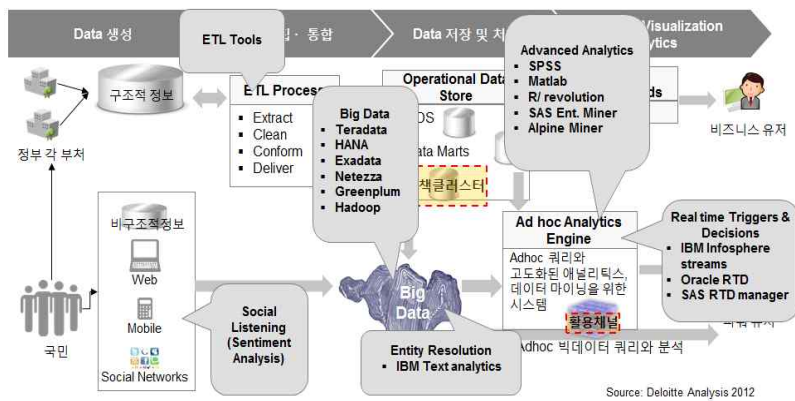
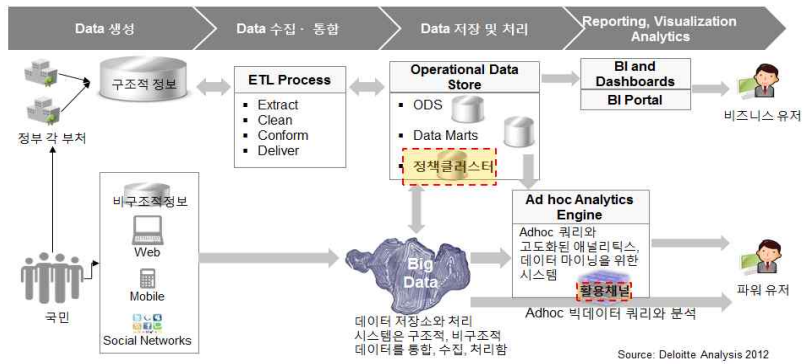
- (업무관점) 정책정보 및 민간데이터를 대상으로 정책 업무 주제와 관계성 있는 데이터를 연계수집한 '데이터의 논리적 집합체' 를 의미
- (DW와 Big Data 통합 프로세스 관점) 구조적 정보 DB에서 ETL을 통해 수집한 정보와 빅데이터에서 수집한 정보의 주제별 데이터 연계 정보

※ BRM 분류체계와의 주요 차이점 : 분류체계에서는 하나의 데이터 개체가 한 개의 분류명에 포함되지만, 정책정보지원 클러스터에서는 하나의 데이터 개체가 활용되는 주제에 따라 복수개의 클러스터에 포함될 수 있음(공급자에서 사용자관점으로)

[정책정보지원 클러스터의 정의 (업무 관점)]



[정책정보지원 클러스터의 정의 (DW와 Big Data 통합 프로세스 관점)]



(3) 정책정보지원 클러스터의 특징 및 구성요소

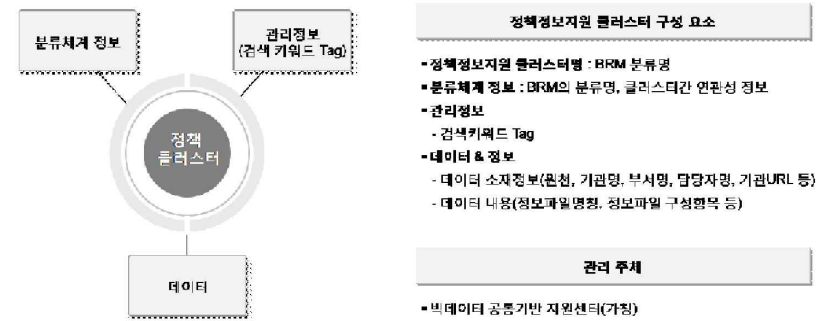
- 정책정보지원 클러스터의 특징은 범정부 Data 통합연계를 통해 소재정보 제공 및 고도화된 검색기술을 적용한 신속한 검색과 효율성을 제고하는 특징이 있음

[정책정보지원 클러스터의 특징]

| | |
|-----------------------------|--|
| 정책 기준으로 Data 연계 · 추출 | <ul style="list-style-type: none"> · 정책 클러스터는 BRM 분류체계에서 '정책 가능' 명을 준용해 클러스터의 이름을 부여하지만 클러스터에 속한 데이터들은 DB간 연관성과 융점성을 기준으로 수집, 통합, 연계에 생성됨(구조적, 비구조적 데이터 모두 포함) · 분류체계의 MECE한 특성을 갖지 않으므로 한 클러스터에 속한 DB가 다른 클러스터에 속할 수 있음 |
| Data 위치 정보 제공 | <ul style="list-style-type: none"> · Data의 유관부처 DB, 원천소스 위치, 접근 정보를 보유함 · 정부 기관 곳곳에 산재한 DB를 중에서 필요한 데이터를 선택할 수 있도록 위치정보를 제공함으로써 신속하게 정보를 수집하고 분석에 착수할 수 있게 지원함 |
| 경과된 검색 가능 제공 | <ul style="list-style-type: none"> · 정책과 연관되는 Data는 클러스터에 모두 포함가능하며 Data 정책 기능을 포함한 지능형 검색 기능 강화 및 가용성 Data를 명확히 규정 · 클러스터를 구성하고 있는 DB들은 각각 고유속성과 참조속성을 가지며 속성수준에서 서로 연관성과 높은 융점성을 가져 주제와 무관한 데이터에 불필요하게 노출되는 확률을 최소화함 |
| 변경 관리 (생성-운영) | <ul style="list-style-type: none"> · 각 부처의 원천 Data가 변경됨에 따라 정책 클러스터의 여러 데이터를 변경하는 라이프 사이클 관리가 필요함 · 정책 업무 주제에 따라 생성 후 정보내용의 변경이 발생하면 변경관리가 필요함 |

- 정책정보지원 클러스터의 구성 요소는 각 부처의 개별 DB에서 추출한 데이터 소재 정보 및 내용정보와 민간 데이터에서 수집된 정보 및 검색키워드 Tag(관리정보)와 분류체계 정보를 포함함

[정책정보지원 클러스터의 구성 요소]



(4) 정책정보지원 클러스터 후보군 도출

- 정책업무와 정보관리를 통합적으로 일치시켜 관리하기 위해 기능연계모델(BRM) 분류 체계를 사용해 정책 정보를 관리하며 정책정보지원 클러스터명은 BRM 3레벨 '대기능'명을 준용하며 정책정보지원 클러스터에 포함되는 정보는 다음과 같은 방안으로 도출함

| | | |
|-----------------|---------------|---|
| 정책 클러스터 분류 체계 | BRM 체계 준용 | <ul style="list-style-type: none"> 클러스터 분류 체계는 BRM 체계를 준용하며 클러스터 명은 3레벨 '대기능'명을 사용함. BRM 1, 2레벨을 준용해 상위 개념으로 묶으므로 클러스터는 3단계로 구성됨 BRM 1단계: ex. 공공질서 및 안전 2단계: ex. 경찰 3단계: ex. 치안활동 |
| 클러스터 도출 기준 및 방안 | 정보식별 (What) | <ul style="list-style-type: none"> 문헌/오야정보, 원시행정DB, 원시대용량데이터, 민간 데이터별로 정보 개체를 식별(따라)한 후 클러스터에 수집할 것인지를 정함. DB의 경우 DB전체 또는 관련데이터에서 필요 속성값을 선별하고 민간데이터는 빅데이터 기술을 통해 필요한 정보 개체를 생성해야 함 |
| | 정보 가치평가 (Why) | <ul style="list-style-type: none"> 정보가 식별된 후 정보적 유용성, 법적, 재정적, 사회적, 역사적, 활용적, 품질적 가치를 평가해 클러스터에 수집할 가치가 있는 정보 개체를 도출함 |
| | 방안 (How) | <ul style="list-style-type: none"> 개체식별 → 데이터평가/선별 → 수집/획득 → 분류/기술(메타데이터) → 클러스터 생성 → 활용의 프로세스를 통해 클러스터를 도출하고 관리를 수행함 |
| 클러스터 후보군 도출 | 후보군 도출 | <ul style="list-style-type: none"> 3개 분야(사회안전, 국민복지, 국가인프라)를 대상으로 정책의 중요도와 파급효과를 고려해 클러스터 후보군 명을 선정하고 위에 제시한 정보 식별과 가치평가의 프로세스를 통해 클러스터 후보군에 수집되는 데이터를 선별함 |

○ 정책정보지원 클러스터 후보군 도출은 3개 분야(사회안전, 국민복지, 국가인프라)를 대상으로 정책의 중요도, 시급성, 파급효과를 고려해 최종적으로 30개를 정함

클러스터명 예비후보군(123개)

| |
|---|
| BRM 3개 정책분야 |
| 공공질서 및 안전 (24개) |
| 사회복지 (51개) |
| 수송 및 교통 (18개) |
| 일반공공행정(30개) |
| 방법 |
| <ul style="list-style-type: none"> BRM의 3개 정책분야를 대상으로 정보 현황을 검토한 후 선별기준에 의거 최종 후보군을 도출 |
| 목적 |
| <ul style="list-style-type: none"> 정보 가치가 있는 정책 클러스터 도출을 통한 정책 의사결정 지원 체계 마련 |

123개
예비후보군 중
93개 제외

“최종적으로
30개의 클러스터
후보군 도출”

후보군 선별기준

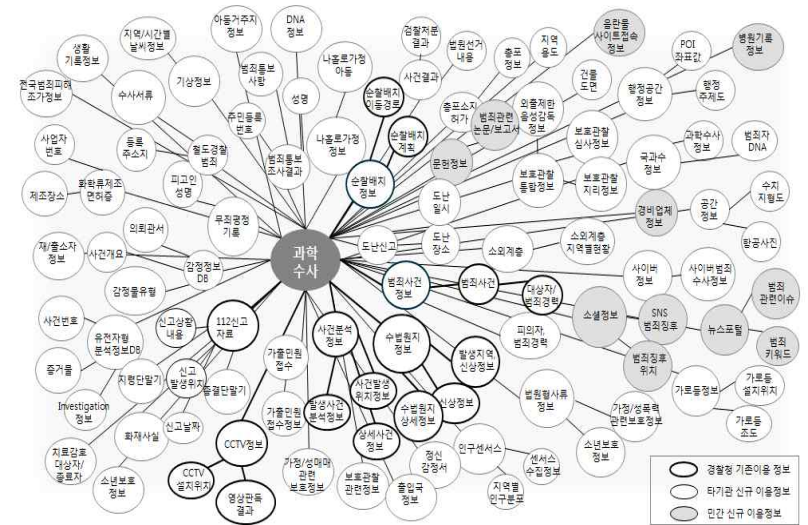
- ✓ 정책의 중요도 및 시급성
- ✓ 파급효과
- ✓ 정보 예상 활용 빈도
- ✓ 예산안 편성방향(국정원리, 국정목표)

클러스터명 최종 후보군 (30개)

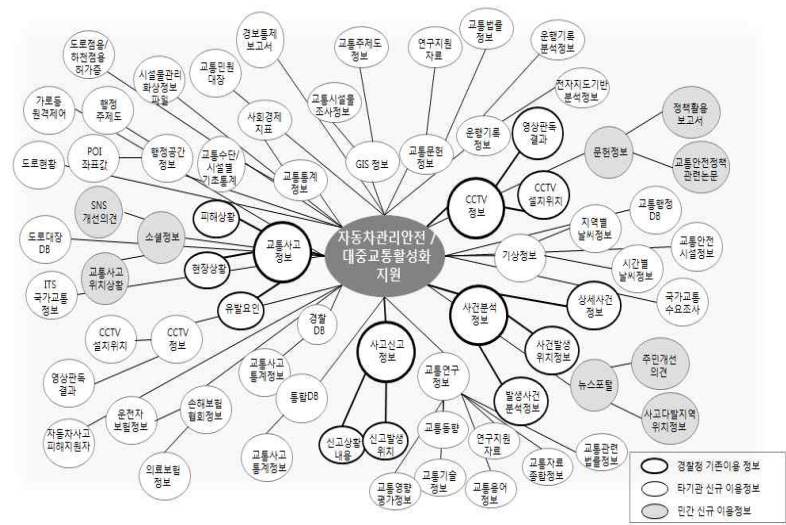
| | |
|-------------|-----------------------|
| 과학수사 | 노인생활안정 |
| 경찰활동 | 청소년정책 |
| 고정활동 | 직업능력개발 |
| 재난상황 관리 | 보육지원강화 |
| 재난예방관리 | 여성정책 및 인력개발 |
| 에너지 수급 안정 | 저출산대응 및 인구지원정책 |
| 가상예보 | 국가적 중장기과제 추진 지원 |
| 금융정보 분석 | 주택금융지원 |
| 국민연금제도 운영 | 주택정책 |
| 기초생활보장 급여지원 | 전력산업경쟁력 강화 및 전력 수급 안정 |
| 고용정책 | 산업재해 예방 |
| 고용평등실현 | 건설교통정보화 |
| 근로자복지증진 | 자율재관리안전 및 대중교통활성화지원 |
| 근로조건보호 | 고중민원처리 및 행정제도개선 |
| 장애인가용증진 | |
| 보건의료서비스 지원 | |

○ 클러스터 후보군에 정보를 선별해 연계 수집하는 방법은 현행 행정정보 소재등록 24,240건 정보를 대상으로 업무 주제별로 필터링해서 그 정보 내용을 선별 연계하는 방법을 택하나 장기적으로는 정보 수요에 따라 정보 보유기관과 협약을 맺고 정보분석가(Data Scientist) 등의 조직이 정보를 선별 연계하는 체제로 운영해야 할 것임. 3대 분야에서 활용과제와 연관된 정책정보지원 클러스터명을 BRM을 준용해 선정하고 이에 관련되는 정보를 연계해 예로 들면 다음과 같음

[정책정보지원 클러스터 '과학수사'의 정보]



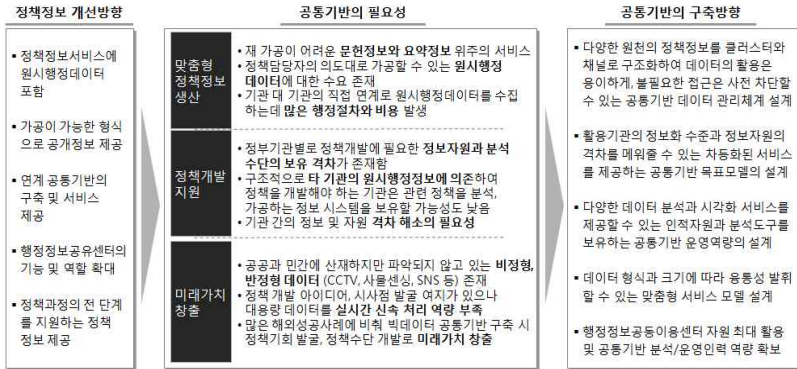
[정책정보지원 클러스터 '자동차관리안전 및 대중교통활성화 지원'의 정보]



다) 빅데이터 공통기반 업무 설계

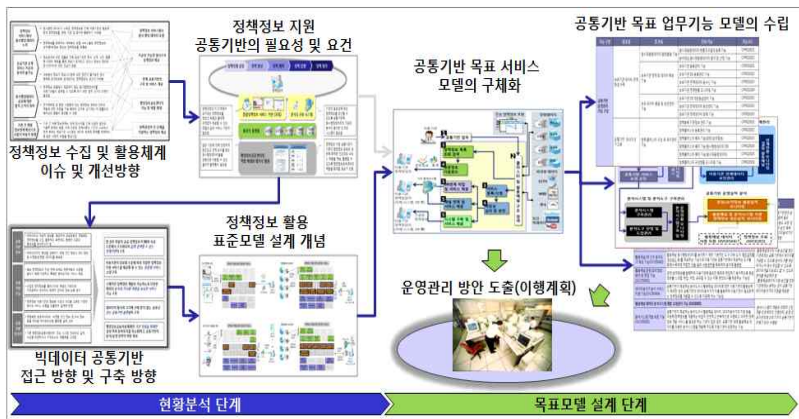
(1) 빅데이터 공통기반 목표모델 수립 개요

- 빅데이터 공통기반 현황분석을 통해 현행 정책정보 수집 및 활용 체계의 개선 방향과 공통기반의 필요성을 도출하였고, 공통기반의 구축방향을 설정함으로써 공통기반이 제공해야 하는 서비스와 보유해야 하는 업무기능을 규정하기 위한 토대를 마련하였음



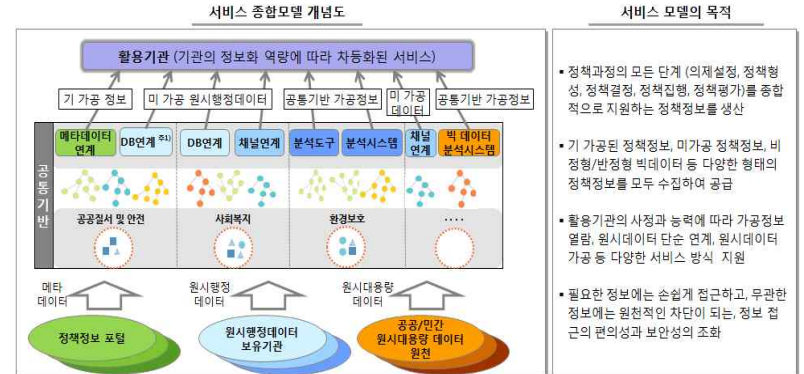
- 현황분석 단계에서 도출한 공통기반의 필요성 및 요건과 정책정보활용 표준모델 설계 개념으로부터 목표 서비스 모델을 구체화하였고 그로부터 업무기능 요소를 추출하여 분류, 체계화, 관계성 정립, 개념 정의, 업무규칙 도출 등의 과정을 거쳐 빅데이터 공통기반의 목표 기능모델을 수립하였음

[빅데이터 공통기반 목표 모델의 수립 절차 및 방향]



(2) 빅데이터 공통기반 목표 서비스 모델 수립

- 활용기관의 정책정보 수요를 최대한 충족시킬 수 있도록 기 가공된 정책정보를 선별하여 맞춤형으로 서비스하고, 활용기관이 원하는 정책정보로 가공할 수 있도록 원시행정데이터를 공급하며, 더 나아가 분석도구 및 분석시스템까지 제공해 주어 공공 및 민간에서 수집한 빅데이터를 분석하여 정책개발 이슈나 테마를 발굴해 주는 종합 서비스 모델로 설계함



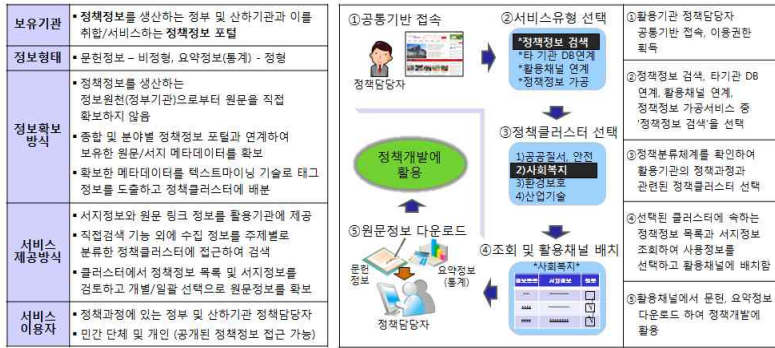
- 공통기반의 서비스 모델은 활용기관의 서비스 수요 및 수용 능력에 따라 5가지 모델로 분화되며 각 모델의 특성은 다음과 같은 항목들에 의해 유사성과 차이점을 비교할 수 있음

[공통기반 5개 목표 서비스 모델의 특성 비교]

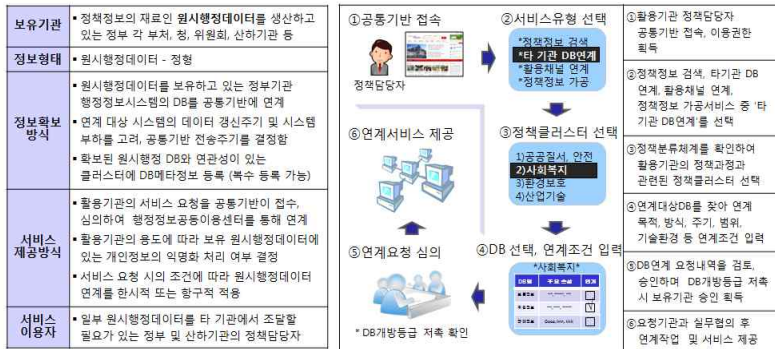
| 서비스 모델 | 서비스 개요 | 정보수집 원천 | 수집정보 유형 | 수집정보 특성 | 정책클러스터 및 활용채널 관련 특성 | 데이터 분석지원 | 제공정보 유형 | 제공정보 특성 |
|--------------------------|--------------------------------|--------------------|-----------------------|--------------------------|--|----------------------|-----------------------|-----------------------|
| 문헌정보 기반 정책정보 조회 서비스 | 기 가공 정책정보의 데이터베이스 조회 및 원문 열람 | 중합 및 분야별 정책정보 포털 | • 문헌정보 • 요약정보 (통계) | • 가공 • 비정형 • 정형 | • 텍스트마이닝을 통한 정보의 클러스터 자동 배분 • 정보의 채널 수동 배분 | 미 지원 | • 문헌정보 • 요약정보 (통계) | • 가공 • 비정형 • 정형 |
| 원시행정데이터 DB연계 서비스 | 활용기관 시스템에 필요한 일부 타 기관 DB의 연계 | 원시행정 데이터 생산 정보시스템 | 원시행정 데이터 | • 미 가공 • 정형 | • 클러스터 검색 및 DB 선택 • 활용채널 생성 없음 • 연계 전 DB 개방등급 협의 | DB 연계 | 원시행정 데이터 | • 미 가공 • 정형 |
| 원시행정데이터 활용채널 연계 서비스 | 활용기관 시스템에 필요한 다수의 타 기관 데이터 연계 | 원시행정 데이터 생산 정보시스템 | 원시행정 데이터 | • 미 가공 • 정형 | • 클러스터 검색, 채널 생성 • 활용채널 DB 구축 후 연계 • 연계 전 DB 개방등급 협의 | • 채널 설정 • 채널 연계 | 원시행정 데이터 | • 미 가공 • 정형 |
| 원시행정데이터 활용채널 정책 개발 서비스 | 활용기관에 필요한 다수의 타 기관 데이터를 분석, 가공 | 원시행정 데이터 생산 정보시스템 | 원시행정 데이터 | • 미 가공 • 정형 | • 클러스터 검색, 채널 생성 • 활용채널 DB 구축 후 이용 • 이용 전 DB 개방등급 협의 | • 채널 설정 • 데이터 분석 | • 문헌정보 • 요약정보 (통계) | • 가공 • 비정형 • 정형 |
| 공공 및 민간 빅데이터 활용 정책개발 서비스 | 공공 및 민간의 빅 데이터로부터 정책정보 분석, 가공 | 원시대용량 데이터 생산 정보시스템 | 원시대용량 데이터 | • 미 가공 • 비정형 • 반정형 | • 가공 전 정보 주제영역을 판독하기 어려움 • 클러스터 및 채널 무관 | • 데이터 수집 • 데이터 분석 | • 문헌정보 • 요약정보 (통계) | • 가공 • 비정형 • 정형 |

- 공통기반 서비스 모델은 5단계의 정책과정 각각에 전부 또는 일부가 적용될 수 있으며 각 모델별 서비스 구성과 서비스 절차는 다음과 같음

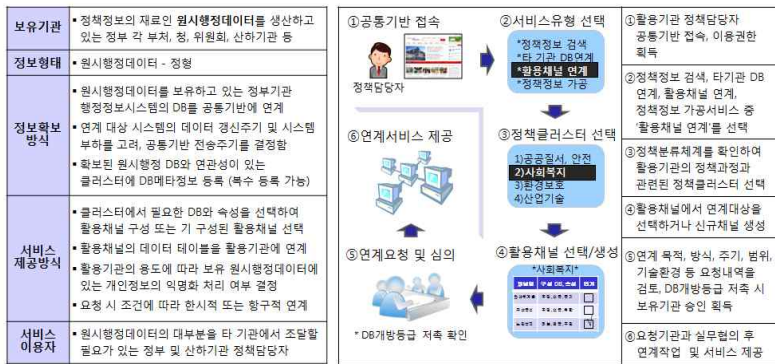
[서비스모델 1 - 문헌정보 기반의 정책정보 조회 서비스 모델]



[서비스모델 2 - 원시행정데이터 DB의 연계 서비스 모델]



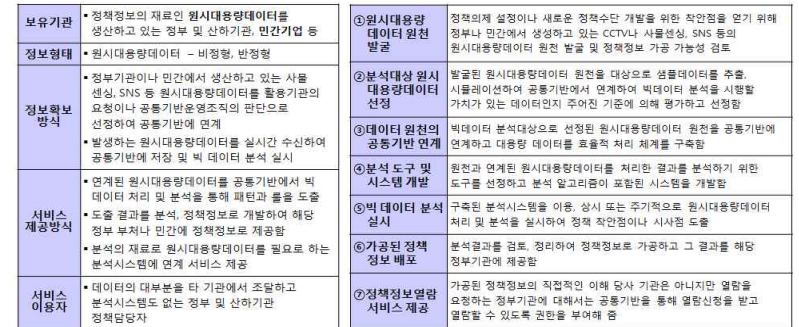
[서비스모델 3 - 원시행정데이터 활용채널 연계 서비스 모델]



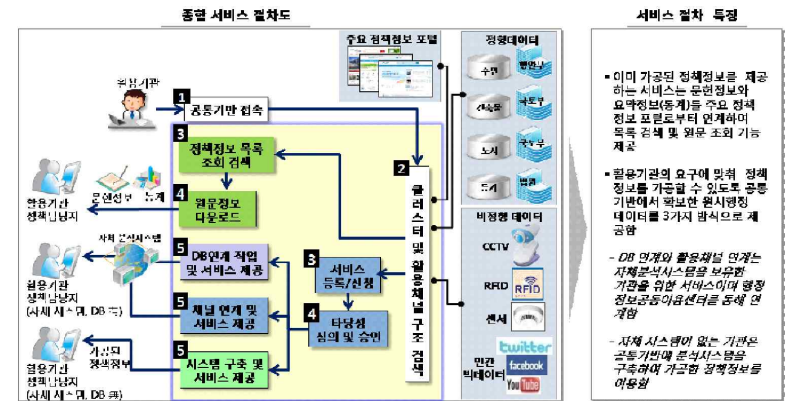
[서비스모델 4 - 원시행정데이터 활용채널 정책개발 서비스 모델]



[서비스모델 5 - 공공 및 민간 빅데이터 활용 정책개발 서비스 모델]



○ 5가지 공동기반 서비스 모델을 종합하여 구성한 서비스 절차도는 다음과 같은

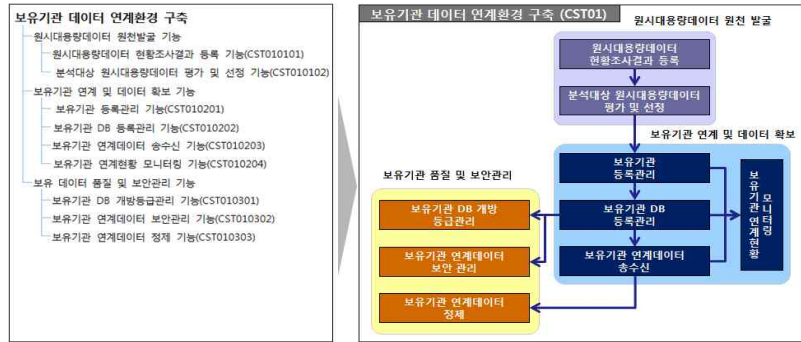


(3) 빅데이터 공통기반 목표 업무기능 모델 설계

공통기반 업무기능은 '서비스 환경 구축' 업무와 '서비스' 업무로 구분하여 도출하였으며 업무기능간의 선후관계, 의존관계는 다음과 같음

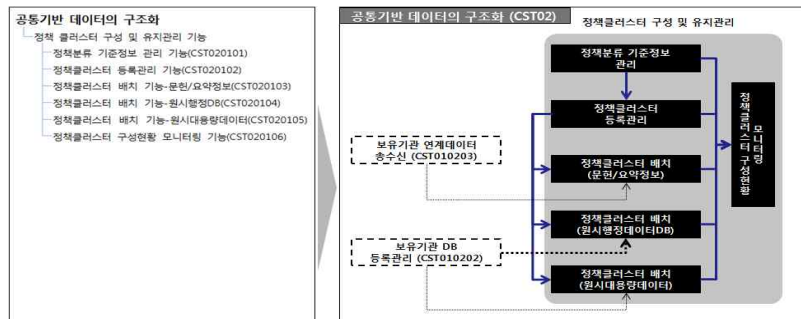
○ 서비스 환경 구축 업무(1/3) : 보유기관 데이터 연계 환경 구축

원시데이터의 원천 정보를 발굴하여 등록하고, 등록된 데이터를 기준으로 보유기관과 협의하여 확보 및 데이터의 품질을 관리하는 업무



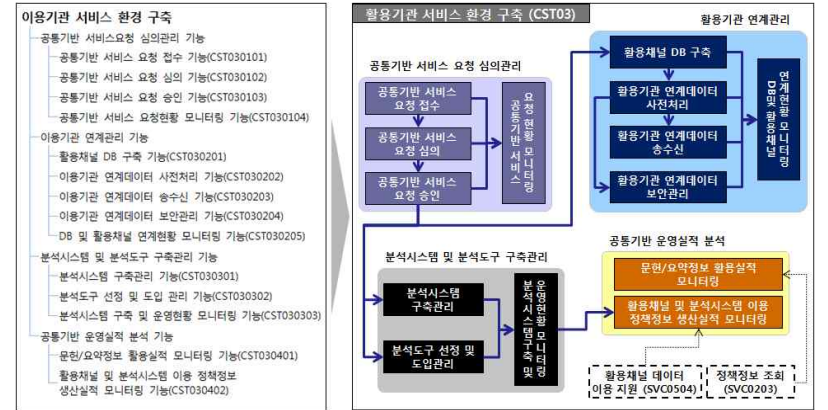
○ 서비스 환경 구축 업무(2/3) : 공통기반 데이터의 구조화

등록된 데이터를 정책 분류기준에 따라 관리하고, 데이터 유형 및 논리적 연계성을 검토를 통해 정책정보지원 클러스터와의 관계를 정의하여 구조화하는 업무



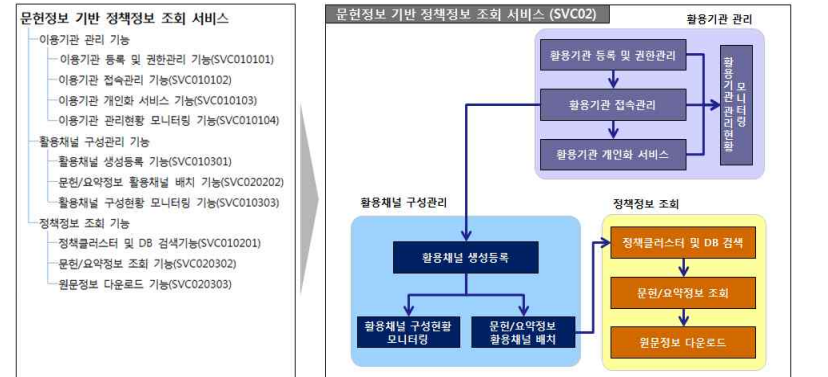
○ 서비스 환경 구축 업무(3/3) : 이용기관 서비스 환경 구축

이용기관에서 필요로 하는 서비스 기능을 심의하여 분석 기반을 확보·관리하고, 이용기관의 데이터 활용과 분석기반 활용에 따른 공통기반 운영 실적을 분석함



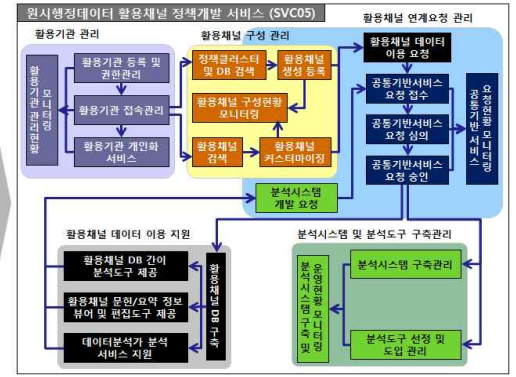
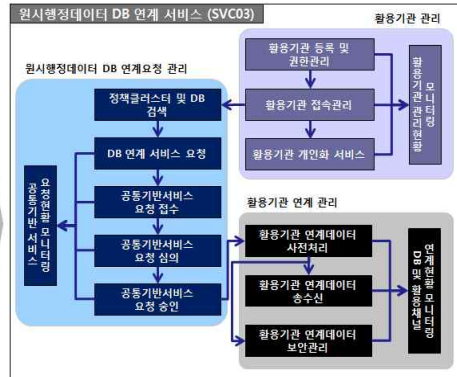
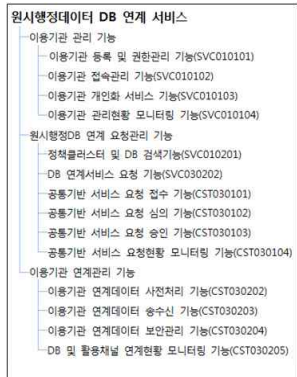
○ 서비스 업무(1/4) : 문헌정보 기반 정책정보 조회 서비스

문헌정보에 대한 이용기관의 수요를 활용 채널을 통해 반영하고, 정책정보지원 클러스터, 문헌/요약 정보 조회 및 다운로드 등의 기능을 제공하는 서비스



○ 서비스 업무(2/4) : 원시행정데이터 DB 연계 서비스

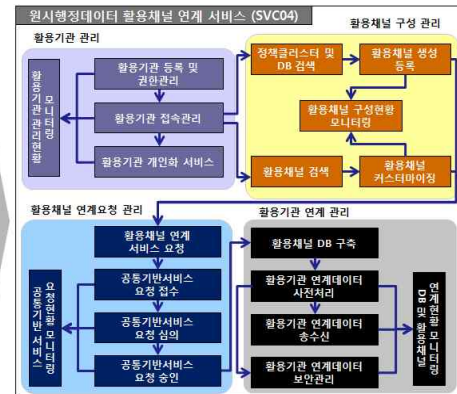
원시행정데이터에 대한 이용(활용)기관의 수요를 관리하고, 연계 요청 정보에 대한 서비스의 심의 및 승인처리를 통해 이용기관에서 데이터를 확보할 수 있도록 데이터를 연계하여 제공하는 서비스



○ 빅데이터 공통기반 업무기능 모델의 프로세스 흐름도

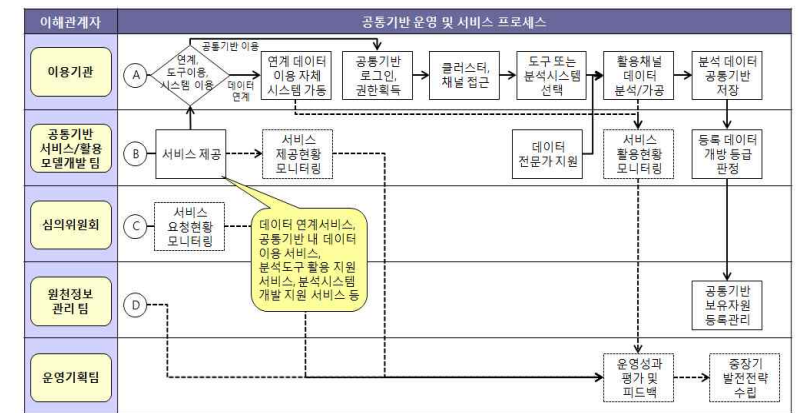
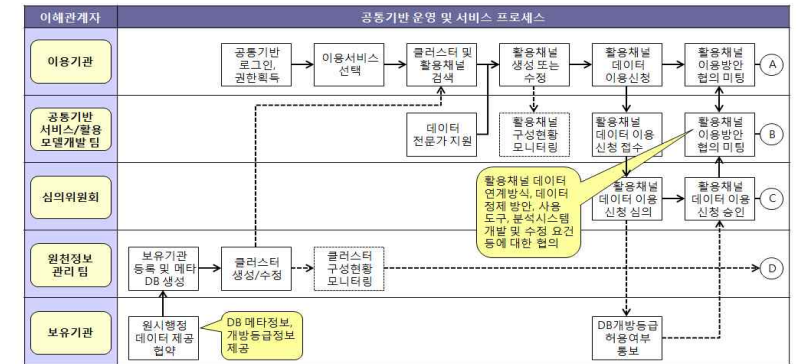
○ 서비스 업무(3/4) : 원시행정데이터 활용채널 연계 서비스

활용 채널을 통해 원시행정데이터에 대한 연계 수요를 관리하고, 신규 연계 서비스 등에 대한 상태를 관리하여 요청접수부터 연계 정보 제공까지 수행하는 서비스



○ 서비스 업무(4/4) : 원시행정데이터 활용채널 정책개발 서비스

원시행정데이터에 대한 이용(활용)기관의 연계 요청 정보뿐만 아니라 분석 서비스에 대한 수요도 접수하여 승인을 통해 활용 채널로 제공하는 서비스(분석도구 선정 및 도입과 분석 시스템의 구축도 관리하여 운영현황을 모니터링 함)



라) 정책 활용과제 업무 설계

(1) 범죄예측 정책 활용과제 목표 모델 수립

(가) 정책활용과제 정의서

- 빅데이터 공통기반을 활용하여 '범죄발생 장소·시간 예측을 통한 범죄 발생 최소화' 과제를 수행하는 업무 프로세스를 설계하고, 참조정보를 도출함

| | | | |
|-------|---|--|--|
| 과제명 | • 범죄발생 장소·시간 예측을 통한 범죄발생 최소화 | 담당부서 | • 경찰청 과학수사센터 - 범죄활동고역계(범죄활동분석팀) |
| 과제 개요 | • 빅데이터 공통기반을 이용하여 범죄발생 장소·시간 예측의 정확도 향상을 통해 범죄 발생 최소화를 지원 | 정확도 | 이제 실성 ● 상제 형성 ● 상제 분석 ● 상제 실행 ● 상제 평가 ● |
| 과제 내용 | <ul style="list-style-type: none"> • 정형 데이터와 비정형 데이터의 다양한 참조정보 확대와 분석모델 개선을 통한 예측정밀도 향상 및 가변적 범죄예측업무모델 구현 × 정형 데이터 : 경찰청 내부 지역별, 시간별 과거 범죄발생 내역, 순찰경로 등 정보와 나열로그인 이동, 날씨, CCTV위치, 가로등 주소, 공간형질정보(여/여/구/구역, 지역상권분포) 등 경찰청의 중앙부처/자치단체 등의 보유정보 비정형 데이터 : 실시간 SNS 범죄징후 탐지, 뉴스 등의 최신 범죄이슈 및 시기에 범죄 트렌드 분석 등 × 참조정보 선정 시 연계 정보간 동일 시계열 조건과 상관분석 결과로 의사결정 및 모델 개선 시 시뮬레이션 기반의 분석 사용 • 실시간 범죄징후 감지를 통한 범죄 사전 대응으로 범죄발생 최소화 • 신고정보 기반의 피동적 대응에서 범죄 징후 정보를 능동적으로 도출하여 사전 대응 | | |
| 개선 방향 | 원활분석 개선방안 | 업무체계 관점 | |
| | <ul style="list-style-type: none"> • 예측정밀도 향상을 위한 정형/무형/민간 보유정보 연계 활용 • 다양한 데이터 소스 적용이 가능한 분석모델 개선 방안 마련 • 빅데이터 공통기반의 분석지능을 활용한 업무 설계 구현 • SNS, 포털뉴스 등 민간 비정형정보의 빅데이터 분석이 가능한 플랫폼 도입 • 향후 업무실현성이 높은 범죄예측 결과정보의 공동이용 고려 | <ul style="list-style-type: none"> • 다부처 보유의 다양한 정형/비정형 참조정보 연계 및 확대를 통한 주기별 범죄트렌드 감지, 도출을 통한 내이요인 활용 • 범죄상황, 유형새변화를 통해 다양한 예측 가능한 도출구성 및 개선 운영체계 마련 • 실시간 범죄징후정보 트라거링의 정보분석을 통한 범죄발생 사전탐지 및 사전대응 시간 확보 | |

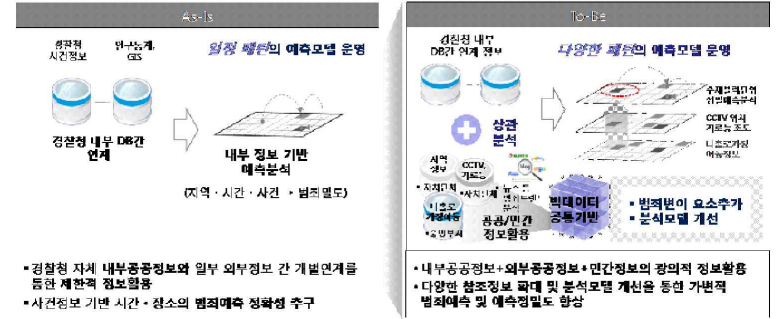
(나) 설계 방안

- 빅데이터 공통기반을 통한 범죄발생 최소화를 달성하기 위해 범죄예측 정확도 향상과 범죄 징후 탐지를 To-Be 방향으로 하여 각각에 대한 Approach 방안을 작성함

| To-Be 방향 | Approach 방안 | 빅데이터 SV가 관점 | 내용 | 기대효과 | 예시 |
|-------------|--------------------------|-------------------------------------|--|--|----|
| 범죄예측 정밀도 향상 | ① 참조정보 확대를 통한 예측모델 개선 | • Variety • Volume | <ul style="list-style-type: none"> • 참조정보 확대 - 종류 + 비정형 - 정보량(▲) • 비정형정보 분석(비 데이터기울) 시군/주기별 범죄 트렌드 도출 및 분야 요인 추가 | <ol style="list-style-type: none"> 1) 기존 정보의 가용성 향상 2) 신(성)소의 발견여야 작성 | |
| | ② 다양한 예측 모델 운영체계 제공 | • Variety • Volume • Velocity | <ul style="list-style-type: none"> • 예측대상제 대한 세그먼트이션 • 범죄상황/유형 세분화 통해 예측 가능한 모델구성 및 관리 체계 마련 | <ol style="list-style-type: none"> 1) 맞춤형 보편적 관측 통한 유출 감시 향상 2) 다양한 분석대 명칭의 사용자대 맞춤 사용 | |
| 범죄징후 탐지 | ③ 범죄 징후 수집분석을 통해 대응시간 확보 | • Variety • Velocity | <ul style="list-style-type: none"> • 범죄정보 Triggering • 실시간 범죄콜링 정보 분석 • 112신고정보와 탐지정보 결합분석 | <ol style="list-style-type: none"> 1) 실시간 범죄정보 탐지 및 대응 2) 정밀한 분석을 통한 위험도 판별/병치 대응 | |

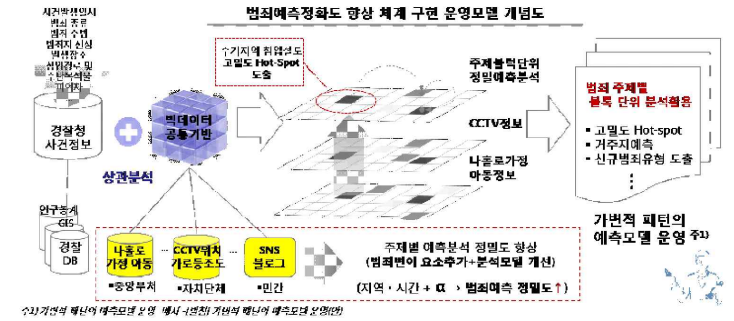
(다) As-Is & To-Be

- 공공기관 보유정보 및 민간 뉴스포털 등 범죄발생 예측성 향상에 필요한 변이 요인을 빅데이터 기반환경 이용 및 업무에 적용함으로써 보다 정밀도가 향상된 범죄발생예측을 통한 국민 신뢰성 향상 및 국가차원의 정책 역량을 강화함

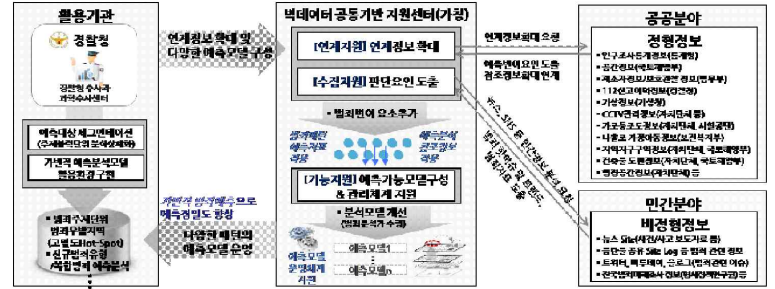


(라) 운영모델 구현방안

- 기존 경찰청 내 정보에 의존했던 분석방식을 빅데이터 공통기반 지원을 통해 다양한 참조정보 확대 및 가변적 패턴의 범죄예측모델 운영환경 지원으로 고밀도의 범죄발생 예측도를 향상함

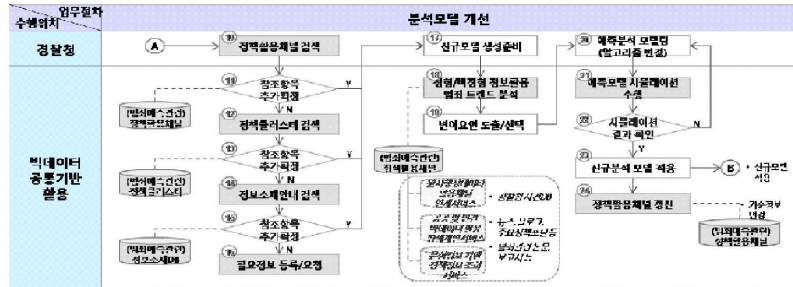
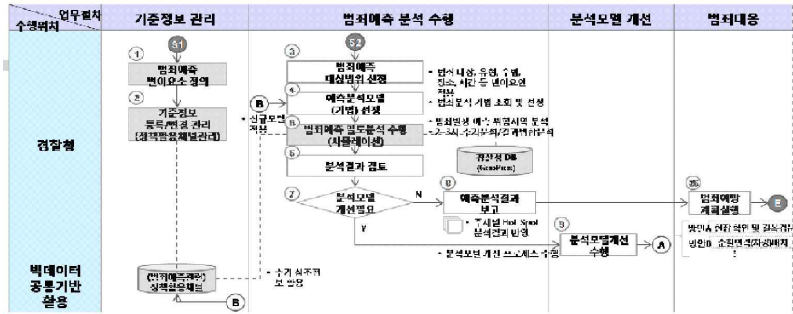


- 예측대상의 분석상세화를 통한 범죄예측 정밀도 향상을 위해 다양한 공공/민간분야의 정형/비정형 정보를 연계 활용하여, 가변적 패턴의 예측모델 운용이 가능한 지원환경을 빅데이터 공통기반을 통해 제공하는 모델로 설계함



(가) 범죤예측을 통한 범죤발생 최소화 업무 프로세스

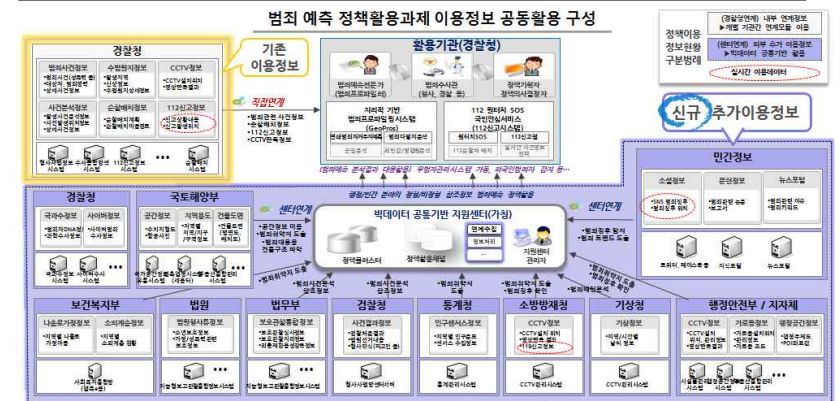
- 범죤발생 예측 정확도 향상을 위한 업무 프로세스는 기준정보관리, 범죤예측 분석수행, 분석모델 개선, 범죤대응으로 구성함



(바) 이용정보 현황 및 흐름

| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|---------------|-------------------|-----------------------------------|--------------------|------------|-----------|
| 인구조사통계정보 | 인구분포 도출 | 통계청 | 통계시스템 | 정형 | 수집 |
| 공간지도정보 | 지역관망도 (공간정보활용) | 국토해양부 국립지리정보원 | 국가공간정보 유통시스템 | 정형 | 연계 |
| 순찰배치정보 | 지역/시간별 순찰배치 확인 | 경찰청 | 순찰배치 시스템(IDS) | 정형 | 연계 |
| CCTV정보 | 범죤취약지 도출 | 시군구(230) 시설관리공단, 소방방재청, 경찰청 | 시설물 관리시스템 | 정형 /비정형 | 연계 |
| 가동정보 | 범죤취약지 도출 | 시군구 시설관리공단 | 시설물 관리시스템 | 정형 | 연계 |
| 나홀로가정 이동정보 | 범죤취약지 도출 | 보건복지부, 지자체 | 사회복지통합망 (e-행복) | 정형 | 연계 /수집 |
| 기상정보 | 날씨/범죤 패턴 상관 분석 | 기상청 | 기상청시스템 | 정형 | API 연계 |
| 지역용도지구정보 | 범죤취약지 도출 | 국토해양부, 지자체 | 세종터,부동산통합 관리시스템 | 정형 | 연계 |
| 행정공간정보 | 범죤취약지 도출 | 행정안전부 | 공간정보시스템 | 정형 | 연계/API |
| 치료감호 | 범죤예측 참조정보 | 법무부 | 치료감호 | 정형 | 연계 |

| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|------------------|---------------------------------|-------------------|-------------------|------------|-----------|
| 대상자/종료자 관련 정보 | (상관분석) | (치료감호소) | 정보시스템 | | |
| 소년보호정보 | 범죤예측 참조정보 (위험지수 측정) | 법무부 (소년원) | 소년보호시스템 | 정형 | 연계 |
| 가정/성매매관련 보호정보 | 범죤예측 참조정보 (상관분석) | 법원 | 가정/성매매관련 보호시스템 | 정형 | 연계 |
| 사건정보 | 범죤예측 참조정보 (상관분석) | 법원 | 형사사법정보망재 판사시스템 | 정형 | 연계 |
| 재/출소자정보 | 범죤예측 참조정보 (상관분석) | 법무부 (교정국) | 교정종합 정보시스템 | 정형 | 연계 |
| 정신감정서 | 범죤예측 참조정보 (상관분석) | 법무부 (보호국) | - | 비정형 | 수집 |
| 보호관찰관련정보 | 범죤예측 참조정보 (상관분석) | 법부부 (보호관찰소) | 보호관찰조사 통합정보시스템 | 정형 | 연계 |
| 출입국정보 | 범죤예측 참조정보 (상관분석) | 법무부 (출입국관리국) | 출입국관리 시스템 | 정형 | 연계 |
| DNA정보 | 범죤예측 참조정보 (상관분석) | 경찰 | - | 정형 | 수집 |
| 성범죄자 신상공개정보 | 범죤예측 참조정보 (상관분석) | 여성가족부 | - | 비정형 | 수집 |
| 아동거주지정보 | 범죤통계지표 | 교육과학기술부 | - | 정형 | 수집 |
| 생활기록정보 | 범죤예측 참조정보 (상관분석) | 교육과학기술부 | - | 정형/비정형 | 수집 /연계 |
| 전국범죤피해조사정 보 | 범죤통계지표 | 한국형사 정책연구원 | - | 비정형 | 수집 |
| 생명보험정보 | 범죤예측 참조정보 (범죄관련자 사전응우 표적) | 생명보험협회 | - | 비정형 /정형 | 수집 |
| 뉴스, 문헌정보 | 범죤트렌드 분석 범죤패턴 분석 범죤위험지표 | 인터넷뉴스포털 문헌정보포털 | - | 비정형 /정형 | 수집 |
| 경비업체정보 | 범죤예측 참조정보 (상관분석) | 민간경비업체 | - | 비정형 | 수집 |
| 음란물Site 접속기록 | 범죤예측 참조정보 (상관분석) | 민간Site업체 | - | 정형 | 수집 |
| 병원기록정보 | 범죤돌발행위 분석 참조정보 | 병원 | - | 비정형 /정형 | 수집 |



정확함/과제 정의서

| | | | | | | | |
|-------|---|-------|---------------|---|----|----|----|
| 과제 정의 | • 재난-재해 전조감지 대응 프로세스 설계 | 담당 부서 | • 소방방재청 방재대책과 | | | | |
| 과제 개요 | • 정형·비정형 데이터에 기반한 재난발생 사전 예측으로 위험지 역을 파악하고 신속한 경보발령 및 대응으로 국민피해 최소화 | 장비 단계 | 이차 | 삼차 | 심차 | 상차 | 심차 |
| 과제 내용 | • 소방방재청 대응경정형정보와 빅데이터 공동기반 비정형정보의 연계 분석을 통한 감응재난 전조모니터링 및 감지 ※ 정형 데이터 : 재난제어이력정보, 기상정보, 침수정보, 지역구간정보 등 비정형 데이터 : 행정정보(행정안전부 주민정보), SNS정보(전조감지, 발생위치), 지자체 보유 교통수위 및 저유량 정보 등 • 빅데이터 기반의 전조감지 및 응급분석을 통한 위험도 예측으로 신속 정확한 재난/재해 예방 서비스 제공 - 점입, 비정형 데이터의 연계분석을 통해 이전에는 예측되지 못한 전조를 감지하고 예경보체계의 신뢰도 향상 | | | | | | |
| 개선 방향 | 현황분석 개선방향 | | | 업무일계 관점 | | | |
| | • 빅데이터 공동기반 비정형정보 연계 분석을 통한 분석정확도 제고 • 이상기후에 취약지역을 파악할 수 있는 재난유형별 분석모델 개선 • 재난경보 속도향상을 위한 재난 선방범역 개선 | | | • 빅데이터 공동기반의 비정형데이터를 활용한 감지정확도 향상 분석 및 예측을 통해 감응재난이 분석모델을 개선할 수 있는 지원체계 마련 • 빅데이터 공동기반의 추가정보를 활용하여 재난 진피대상을 확보하고, 재난 경보선에서 속도 및 정확도도 향상될 수 있는 범용을 모색함 | | | |

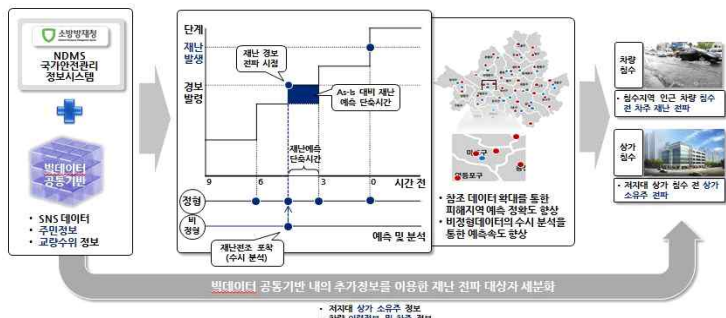
(나) 설계방향

- 빅데이터 3V 관점에서 재난경보 속도향상 측면과 전파대상의 세분화 측면 방안, 기대효과를 고려하여 재난피해 최소화를 위한 To-Be 모델 설계의 방향성을 정의함

| To-Be 방향 | Approach 방안 | 빅데이터 3V 관점 | 내용 | 기대효과 | 예시 |
|-----------|----------------------------------|-------------------------|--|---|------------------------|
| 재난경보 속도향상 | ① 실시간 정보 획득을 통한 예측모델 개선 | • Variety • Volume | • 전조감지 확대 - 수위, 제과제, 침몰, 비정형 공유정보 및 SNS 데이터 - 실시간 모니터링 • 빅데이터 공유정보를 통해 상황인식 및 의사결정 효율성 극대화 가능 | 1) 보다 신뢰성 높은 기상정보 제공 가능 2) 예측모델의 개선 지원 | 경청, 경청+비정형 |
| | ② 비정형데이터 전조감지 분석을 통해 피해지역 적시성 향상 | • Variety • Velocity | • 정형데이터를 포함한 기존에 중기적 지니 예측정보, 빅데이터 공동기반의 비정형데이터를 활용한 감지정확도 향상 • SNS 데이터 비정형 정보에서 유익한 실시간 전조감지 분석 | 1) 감응재난의 발생 및 예측모델 개선 지원 2) 재난 예측속도 및 경보정확도 향상 | NDMS GIS, SNS, 전조감지 |
| 전파대상의 세분화 | ③ 전파대상의 세분화를 통한 맞춤형 전파 체계 수립 | • Variety • Velocity | • 빅데이터 공동기반이 추가정보를 활용하여 세부 대상을 세분화함 • 예: 침수 지역을 지반으로 특성(내장지(전압, 수위, 침수) 등)에 따라 세부 대응 가능 | 1) 재난 전이속도 및 대응 정확도 향상 2) 재난 피해 최소화 지원 | NDMS, GIS, SNS, 전조감지 |

(다) 운영모델 구현방안

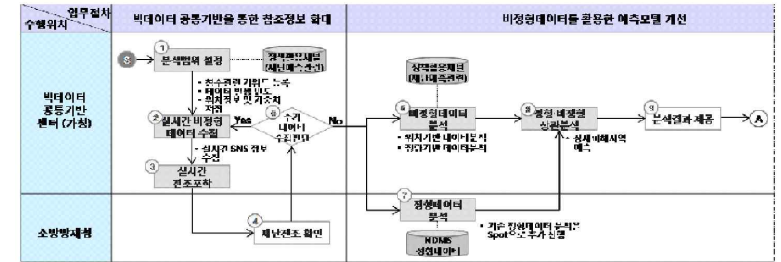
- 빅데이터 공동기반의 추가적인 비정형정보를 재난 전조감지 및 분석 업무와 재난 전파시의 대상 세분화에 활용하는 향후 운영모델을 수립함



(라) 재난 전조감지 업무 프로세스 MAP

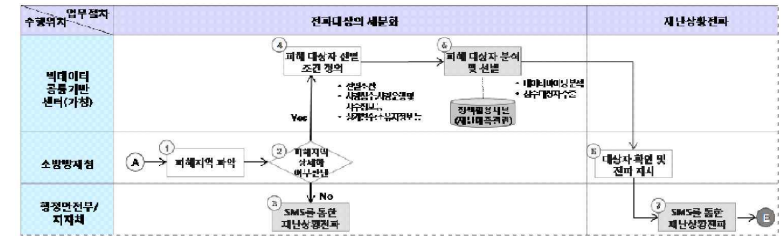
- 업무유형 A: 비정형 데이터를 활용한 예측모델 개선

- 재난경보 전파속도의 향상을 통해 보다 향상된 재난-재해 전조감지 가능



- 업무유형 B: 전파대상의 세분화

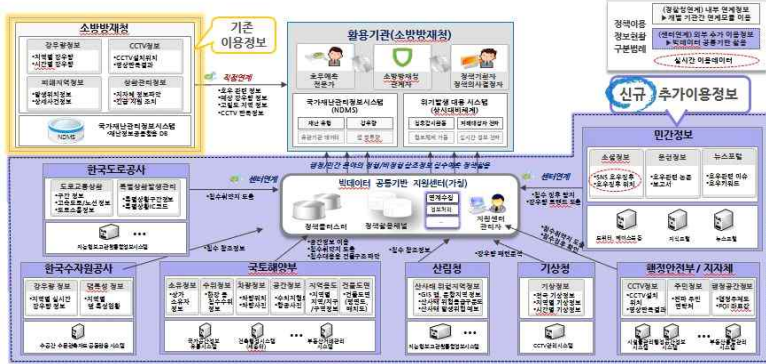
- 세부 전파지역 정보를 활용한 전파대상자 분석으로 맞춤형 재난전파체계를 구축



(마) 이용정보 현황 및 흐름

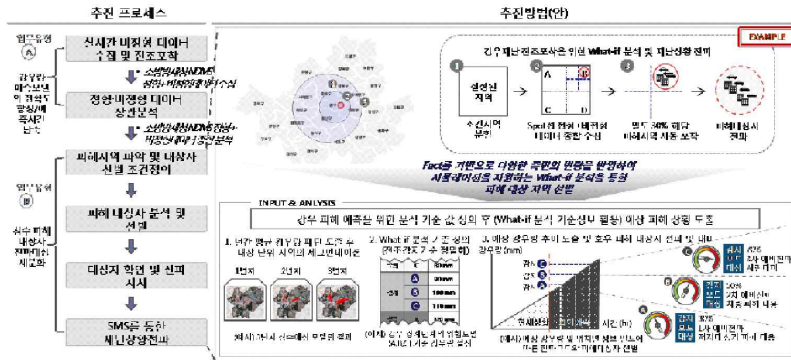
- 소방방재청의 재해상황분석-판단 시스템의 현 활용정보를 중심으로 주요 이용 정보 구성 및 정보흐름을 정리함

| 정보명 | 수요정보(항목) | 용도 | 보유기관 | 관련시스템 | 유형 | 수집 방법 |
|---------------------|---|------------------|-------------|-----------------------|-------|--------|
| 기상정보 | • AWS™ 관측지점, 예보정보, 대강정보, 지진정보, 주간예상, 기상정보, 산악정보, 내륙지역정보, 기상정보, 국가기상 | 기상정보보급 | 기상청 | 종합기상 정보시스템 | 실형 | NDMS |
| 시차제 수위, 무강, 수문 등 정보 | • 무강제수위, 수문제수위, 시차제, 시차제, 시차제, 수문제수위, 수문제수위, 수문제수위 | 대수위 및 방류량 분석 | 한국수자원공사 | 시차제 수문제수위 방류량 분석 | 정형 | NDMS |
| 소통정보, 도로정보 | • IC카드, 유계소통, 특별시차제수위, 구강정보, 고속도로 정보, 고속도로, 도로수문제수위, 도로 CCTV 영상정보 등 | 도로교통 상황분석 | 한국도로공사 | 고속도로 교통관리시스템 | 점입 | NDMS |
| 민상-중간형 영상도로 위성영상 | • GIS, 운영지역, 사면영상정보, 산사태위험지구, 산사태 영상정보, 산사태 위험지구, 산사태 위험지구 | 산사태 위험지역 분석 | 산림청 | 산사태 위험지 관리시스템 | 실형 | NDMS |
| SNS 정보 | • 재난유형 및 위치정보 포함 SNS 데이터 (Twitter, Me2day) | 고도, 재-구간 분석 | 국도대청부 (경시청) | 수자원기술통 정보시스템, 도시정보시스템 | 점입 | NDMS |
| 주민정보 | • 전파대상 주민 연락처 및 소재정보 | 재난 상황 및 민생취약지 도출 | 민간 | 빅데이터 공동기반 | 비정형 | API 수집 |
| 차량정보 | • 차주정보, 차량등록번호 (연간 CCTV 촬영) | 현미경상 소재 정보 | 행정안전부 | 주민등록정보 시스템 | 점입 | 연계 |
| 지자체 수위정보 | • 뚝수위정보, 하천미리관리정보 | 차량소재지 도출 | 국도대청부 / 지자체 | 차량소재지 정보시스템 | 정형/실형 | 연계/수집 |
| 민통선능기정보 | • 민통선 및 상가 소유주 연락처 정보 | 시차제 운영 지역 수위정보 | 국도대청부 | 하천관리지리 정보시스템 | 점입 | 연계 |



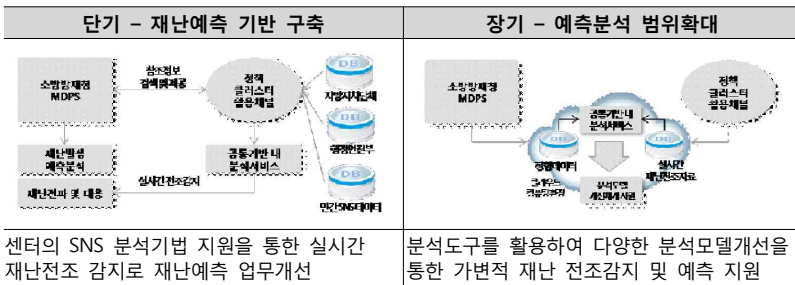
(바) 가우재난 전조감지 예측모델 운영(안)

- 빅데이터 공통기반을 활용하여 지역 및 차량에 대해 침수 위험성에 대한 밀도에 대한 What-if 분석을 통하여 피해 대상자 선별 조건을 정의함



(사) 추진 로드맵

- 재난 전조감지 활용과제의 향후 방안을 위해 단계별 이행계획을 수립함



센터의 SNS 분석기법 지원을 통한 실시간 재난전조 감지로 재난예측 업무개선

분석도구를 활용하여 다양한 분석모델개선을 통한 가변적 재난 전조감지 및 예측 지원

(아) 기대효과

- 정성적 기대효과

| 수요자 측면 (소방방재청) | 수혜자 측면 |
|--|--|
| <ul style="list-style-type: none"> • 전조감지: 실시간 전조포착을 통해 보다 신속한 상황판단 가능 • 재난예측: 정형·비정형 데이터를 통해 기반으로 What-if 분석을 통해 예측속도·정확도 향상 • 재난전파: 세분화된 전파대상에게 효과적인 재난 전파 및 대응 유도 가능 | <ul style="list-style-type: none"> • 유과기관: 지역별/시간별 재난피해 예측 지역의 분석결과정보 등 정책지원정보를 근거자료로 활용하여 각 기관별 담당 정책수립 시 활용 • 국민: 지역/시간대별 재난피해예측을 통한 피해 사전대응 및 국민 안전성 보장 |

- 정량적 기대효과

| 이행과제 | 성과 지표명 | 기대효과내용 | 산출 방식 | 산출 금액(년간액 원) | | | | |
|-------------------------|--------------------|---|---|---|-------|-------|-------|-------|
| | | | | 2014년 | 2015년 | 2016년 | 2017년 | 2018년 |
| 침수 예측 시간 단축 및 대상자 집중 전파 | 차량 및 저지대 상가 피해액 감소 | <ul style="list-style-type: none"> • 차량 침수 피해 예측시간 단축 및 대상자 집중 전파에 따른 피해액 감소효과 • 777만건 X 700억원/건 X 50%* X 20%* = 5.44억 원 감소 | 연간 침수 피해 차량 전수 X 연간 1대 차량당 피해액 X 대응시간 단축 및 대상자 집중 전파에 따른 피해액 감소효과 | - | 5.44 | 5.44 | 5.44 | 5.44 |
| | | | | - | 40.4 | 40.4 | 40.4 | 40.4 |
| 합계 | | | | 45.84 | 45.84 | 45.84 | 45.84 | |
| 전제사항 | | <ul style="list-style-type: none"> • 침수 위험 3대의 차량을 1대가 소용한다는 가정 • 주요 대상자 인원 비율은 50%로 가정 • 주요 차량 침수 위험 대상자 집중 전파 체계 후 대차 확대 비율은 20%로 가정 (즉, 기존 정보 체계하에서 대차인원 차량은 100으로 하면 120 수준으로 확대할 수 있음) • 주요 상가 침수 위험 대상자 집중 전파 체계 후 대차 비율은 10%로 가정 (빠른 시간 안에 가변 소자를 통을 정기는 비율은 10%로 나머지 90%는 기타 무결간대 대체시 갖는 부분으로 고려) | 관련 Factor | <ul style="list-style-type: none"> • 주요 연간 자동차 침수는 총 777건 • 주요 차량 당연도 피해액은 700만 원 • 주요 침수로 인한 인원이 위치한 저지대 상가 피해액은 808억 원 (서울 저지대 상가 피해액, 전국/서울 인구 밀도를 이용하여 전국 저지대 상가 피해액 추후 2011년 기준, 비 주거용, 명세 소상용만 피해사양명 고려) | | | | |

(3) 민원환류 정책 활용과제 목표 모델 수립

(가) 정책활용과제 정의서

- '민원데이터 분석을 통한 정책의 환류시스템 마련'의 실효성 있는 업무 설계를 하기 위해 현 국민권익위원회 민원 분석/환류업무를 중심으로 향후, 빅데이터 공통기반을 활용한 정책활용과제를 설계함

| 과제명 | 민원 데이터 분석을 통한 정책의 환류 시스템 마련 | 담당 부서 | <ul style="list-style-type: none"> • 국민권익위원회 - 민원정보분석센터 | | | | |
|-------|---|-------|---|--|-------|-------|-------|
| 과제 개요 | 빅데이터 공통기반을 활용하여 민원정보 수집범위를 확대하고 정확하고 유의미한 분석 결과를 도출하고, 원활한 정책 환류를 지원 | 정책 단계 | 의제 설정 | 정책 형성 | 정책 수립 | 정책 집행 | 정책 평가 |
| 과제 내용 | <ul style="list-style-type: none"> • 정형 데이터와 비정형 데이터를 포함한 다층적 민원 관련 DB를 다각도로 분석함으로써 정부 부처 및 지자체에서 이를 반영한 제도 개선과 정책 수립을 할 수 있도록 이용 정보 범위 및 민원정보분석시스템 이용 기관 확대, 분석연진의 정확도 및 활용도 개선 뿐만 아니라 정책 환류 과정에서 국민권익위원회의 역할 증대 추진 | | | | | | |
| 개선 방향 | 현행분석 개선방향 | | | 업무설계 관점 | | | |
| | <ul style="list-style-type: none"> • 민원정보 분석 시 활용하는 참조 정보 수집범위 확대 필요 • 분석시스템의 분석패턴 다양화 및 수집 채널별 가중치 적용으로 실질적인 국민의 민원 수요 분석 • 민원정보 분석결과와 적극적인 정책반영 추진 • 각 부처 혹은 기관의 민원관련 정책의 평가 활성화 | | | <ul style="list-style-type: none"> • 부처나 기관의 요청 유무와는 상관없이 민원정보를 분석시스템을 활용하여 각종 통계정보와 교차 분석하여 사회 전반적인 취약점·개선점 도출 • 국민 수요를 충족시키는 신규 정책 기획 및 제안 • 부처나 기관의 민원관련정책 평가 및 평가결과 피드백 | | | |

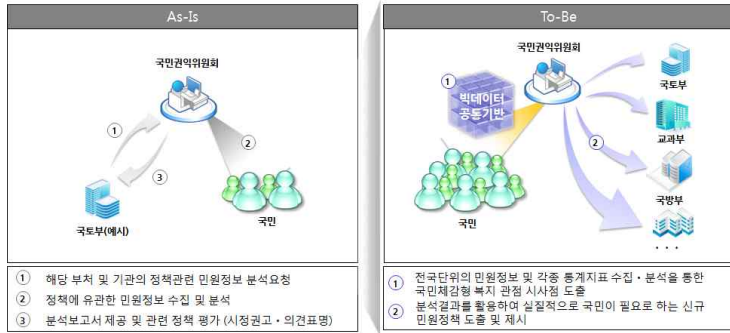
(나) 설계방향

- 빅데이터 3V 관점에서 재난정보 속도향상 측면과 전파대상의 세분화 측면 방안, 기대효과를 고려하여 재난피해 최소화를 위한 To-Be 모델 설계의 방향성을 정의함

| To-Be 방향 | Approach 방안 | 빅데이터 3V ¹⁾ 관점 | 내용 | 기대효과 | 예시 |
|----------------------------|--|--------------------------|---|--|----|
| 고충민원 감소 신규 민원정책 도출체계 구축 | ① 정보의 확대를 통한 국민 복지 수혜자 확대 | • Variety • Volume | • 참조정보 확대 - 서울·광역자치단체 민원 - 민원24의 민원발급 현황 및 통계정보 - 통계청의 각종 통계정보 | 1) 기존 정보의 가용성 향상 2) 분석결과와 다양성 및 실효성 향상 | |
| | ② 분석시스템의 개선성을 통한 적시성 있는 국민제감형 정책 도출가능 구현 | • Variety • Velocity | • 분석시스템의 성능 향상으로 분석관의 업무 부담 감소 및 신속한 분석결과 도출 • 정보 분석 패턴 다양화를 통한 국민의 잠재 니즈 도출 | 1) 분석 정확도 개선으로 업무 처리속도 및 실효성 향상 2) Customizing 된 분석결과 및 정책관리 도출 | |

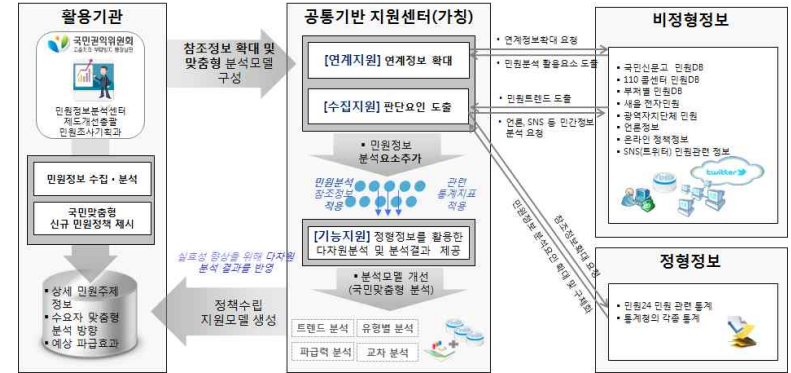
(다) As-Is & To-Be

- 빅데이터 공동기반에서 제공하는 전국 단위의 민원정보와 각종 통계지표를 분석시스템을 활용하여 국민 복지 향상에 기여할 수 있는 분석 결과와 민원 정책관련 시사점을 도출하고 이를 활용하여 국민들이 실질적으로 필요로 하는 정책을 제시함



○ 민원환류체계 구현 운영모델

- 빅데이터 공동기반 지원요소를 활용하여 정책동향과 SNS정보를 분석하고, 민원수요를 파악하여, 분석주제 및 방향 설정, 민원정보와 통계지표정보를 교차분석하며, 분석결과를 반영한 실효성 있는 민원정책 수립을 위한 지원함



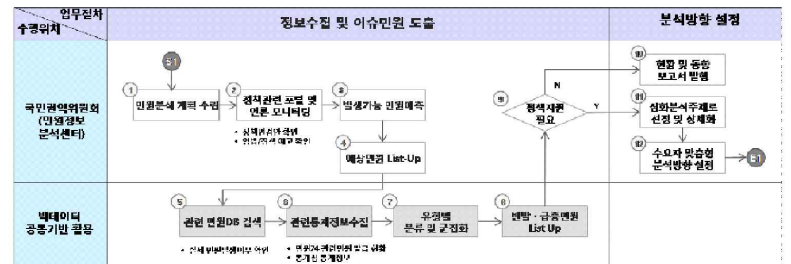
(라) 업무 프로세스 Map

○ 업무 프로세스 유형

- 민원환류 정책과제의 업무 프로세스를 민원정보 및 관련 기타정보의 수집과 정제 측면, 분석과 활용 측면으로 나누어 각 유형별 업무의 수행내용과 방식, 발생주기, 주무부서를 선정하고 업무 프로세스 설계 방향을 도출함

| 유형 | 업무 | 방식 | 수행내용 | 발생주기 | 주무부서 |
|---------|----------------------|-----------------|---|---------------------|--|
| 업무 유형 A | 민원관련정보 수집 및 분석방향 설정 | 정보 수집 정제 | 국민신문고, 110콜센터, 서울·광역자치단체 홈페이지, SNS를 통해 접수된 각종 민원정보와 민원24 통계청의 관련 통계정보, 다양한 언론 대외정보를 빅데이터 공동기반을 통해 수집·통합하고 이를 활용하여 실질적인 국민의 요구사항을 파악할 수 있는 분석주제 선정 | 주기적 (지속적인 분석과제 발굴) | • 국민권익위원회 (민원정보분석센터) |
| 업무 유형 B | 민원정보 분석 및 신규 민원정책 제시 | 정보 분석, 분석 결과 반영 | 민원정보분석센터의 민원정보 분석결과를 바탕으로 제도개선총괄담당부와 민원조사기획과의 업무공조를 통해 도출된 결과에서 시사하는 개선점, 보완점을 파악하여 과제 발굴 및 고충민원 발생을 최소화할 수 있는 신규 민원정책 제안 및 평가 | 불규칙 (신규정책 필요시 업무공조) | • 국민권익위원회 (민원정보분석센터, 민원조사기획과, 제도개선총괄담당관) |

○ 업무유형A(민원관련정보 수집 및 분석방향 설정)



○ 업무유형B(민원정보 분석 및 신규 민원정책 제시)

(4) 교통사고 개선 정책 활용과제 목표 모델 수립

(가) 정책활용과제 정의서

- '주민참여형 교통사고 감소체계 구축'의 실효성 있는 업무 설계를 하기 위해 전문가 인터뷰와 문헌자료를 중심으로 정책활용과제를 설계함

| | | | | | | | | | | | | | |
|-------|---|---|---|-------|-------|-------|-------|-------|---|---|---|---|---|
| 과제명 | 주민참여형 교통사고 감소체계 구축 | 담당 부서 | 국토해양부 첨단도로화결과 국토해양부 중합교통정책과 | | | | | | | | | | |
| 과제 개요 | <ul style="list-style-type: none"> • 필요성: 지능형 교통체계 계획수립, 자율비상 재동장치 개발 등 교통사고 감소를 위해 노력 중이나 아직 OECD 평균수준의 2배가 넘는 수준 • 목표: 교통안전 관련자료 및 주민참여 개선의견 등을 이용하여 분석모델을 구축하고 이를 정책에 활용하여 교통사고 발생 최소화 | 정책 단계 | <table border="1"> <tr> <td>의제 설정</td> <td>정책 형성</td> <td>정책 결정</td> <td>정책 집행</td> <td>정책 평가</td> </tr> <tr> <td>●</td> <td>●</td> <td>●</td> <td>●</td> <td>●</td> </tr> </table> | 의제 설정 | 정책 형성 | 정책 결정 | 정책 집행 | 정책 평가 | ● | ● | ● | ● | ● |
| 의제 설정 | 정책 형성 | 정책 결정 | 정책 집행 | 정책 평가 | | | | | | | | | |
| ● | ● | ● | ● | ● | | | | | | | | | |
| 과제 내용 | <ul style="list-style-type: none"> • 정형 데이터와 비정형 데이터를 결합하여 신호등, 과속 방지턱 설치 등의 개선방안 도출 및 교통사고 예측서비스 제공 ※ 정형 데이터: 자동차보험, 차량등록, GIS, 지능형 교통정보, 의료정보 등 비정형 데이터: 주민 개선의견, 게시판, 교통방송 등의 교통사고 위치 등 • 정형데이터를 통해 발생지점별 사고규모, 피해액 등을 도출하고 주민들의 교통사고 개선의견을 반영하여 사고유형·시간대별 교통사고 발생지점 발견 및 개선 | | | | | | | | | | | | |
| 개선 방향 | 현황분석 개선방향 <ul style="list-style-type: none"> • 경찰 DB를 중심으로 여러 부처 시스템에서 수집 및 추출되는 자료의 호환성 강화 • 자료의 객관화 및 표준화, 다양화 시킴 • 빅데이터 기술을 사용하여 분석시스템의 실시간 분석과 대응방 분석을 가능하게 함 • 교통사고 분석 전문가 - 대용량 정보에 대한 분석 환경 마련 | 업무설계 관점 <ul style="list-style-type: none"> • 다부처 보유의 다양한 정형/비정형 참조정보 연계 및 확대를 통하여 교통사고 위험지역과 사고요인 분석의 정확도 향상 • 실시간으로 교통상황, 사고요인을 담은 비정형 데이터를 다양한 장구에서 수집하고, 빅데이터 공통기반 분석기술을 통한 고도화 분석 후 운전자에게 교통정보를 제공 가능한 운영체계 마련 | | | | | | | | | | | |

(나) 설계방향

- 빅데이터 3V 관점에서 재난경보 속도향상 측면과 전파대상의 세분화 측면 방안, 기대효과를 고려하여 재난피해 최소화를 위한 To-Be 모델 설계의 방향성을 정의함

| To-Be 방향 | Approach 방안 | 빅데이터 3V 관점 | 내용 | 기대효과 | 메시 |
|---------------------|--------------------------|------------------|---|---|--|
| 도로운영 환경 실시간 연대 | 소셜정보(트위터 등) 등 도로운영 환경가이드 | Variety Velocity | SNS 기반의 위험정보 수집을 통한 VMS 교통상황을 안내하여 운전자들이 적절한 행동을 취할 수 있도록 교통정보를 제공하여 안전과 편의를 도모 | 1) 실시간 위험요인 조기감지를 통한 사고예방 2) 도로소통 운영의 안전성과 효율성을 증진 | 안개 결빙 지역 SNS를 통한 제보 VMS 안내를 통한 교통사고 예방 |
| 주민이 원하는 교통안전 시설물 설치 | 주민의사 수집창구 다양화 | Variety | SNS 기반의 지역 주민의사를 반영한 교통안전 시설물 설치 등 새로운 방식의 교통안전 개선 방안 마련 | 1) 주민체감형 교통안전시설 설치 등으로 교통사고 위험요소 제거 2) 또저대비 높은 교통안전 개선효과 | 사고 위험지역 위도를 통해 민원신청 주민 개선의견을 반영한 안전시설물 설치 |

(다) As-Is & To-Be > 주민이 원하는 교통안전시설물 설치 모델

○ As-Is

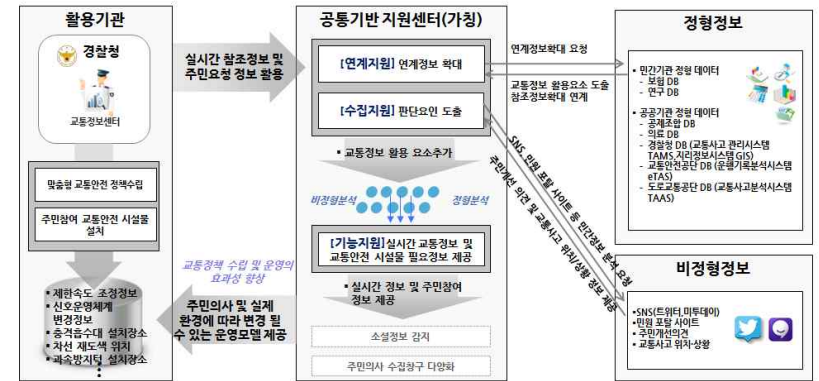
- 경찰청 자체 교통사고 분석시스템인 TAMS를 이용하여 교통사고 이력 등 경찰청 내부정보 중심의 교통안전정책을 수립함
- 불명확한 수요대상자와 한정된 데이터 분석으로 인해 비효율적인 교통안전정책 수립과 안전시설물을 설치함

○ To-Be

- SNS 및 소셜미디어 등을 통한 지역 주민의사 수집 창구를 확대하여 주민 체감형 교통안전시설물을 설치함

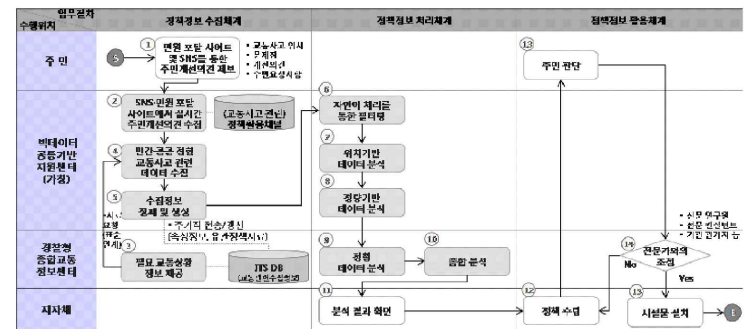


(라) 교통사고 감소 운영모델



(마) 업무 프로세스 맵

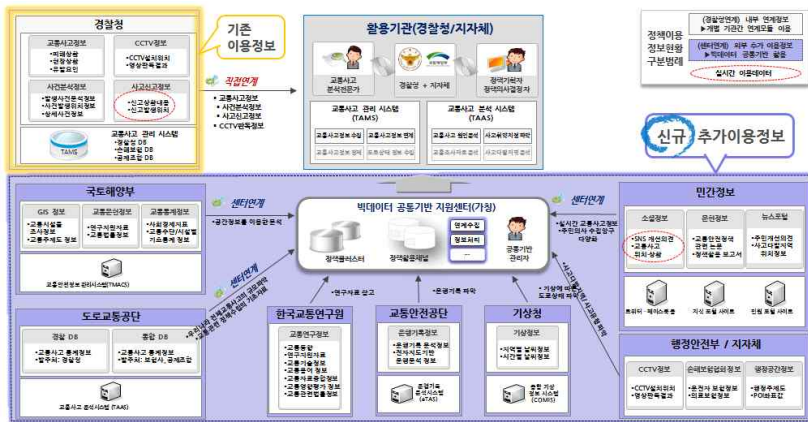
- 주민이 원하는 교통안전시설물 설치 운영모델 프로세스는 정책정보 수집, 처리, 활용체계로 나뉘며 SNS를 활용하여 수집된 주민의견을 반영하여 주민체감형 교통안전정책을 수립하는 프로세스임



(바) 교통사고 분석을 위한 이용정보 현황

- 경찰청에서 현재 이용하고 있는 교통사고정보는 손해보험사, 공제조합, 경찰청내 자체 DB를 이용하여 교통사고 원인분석을 하고 있음

| 정보명 | 용도 | 보유기관 | 관련시스템 | 유형 | 수집방법 |
|----------------|---------------------|-------------------------------------|------------------|----------|-----------|
| 도로교통정보 | 교통상황 파악 | 경찰청 종합 교통정보센터 | 지능형교통체계 (ITS) | 정형 | 수집 (의일배차) |
| CCTV 영상 | 상세 현황 확인 | 경찰청 종합 교통정보센터 (도시지역) 국토해양부 (고속도로국도) | 비정형 | 실시간 연계 | |
| 돌발상황 발생정보 | 돌발 상황 대처 | 국토해양부/ 지자체 | 정형 | 상황발생시 연계 | |
| 공간정보 (GIS) | 지도표기/활용 | 국토해양부 | 국가공간정보 유통시스템 | 정형 | 요청시 수집 |
| SNS 포털사이트 민원정보 | 교통상황 파악 | 경찰청 종합 교통정보센터 | 빅데이터 공통기반 | 비정형 | 실시간 수집 |
| 교통통제정보 | 교통안전 정책활용 | 국토해양부 | 지능형 교통체계(ITS) | 정형 | 실시간 연계 |
| 연구자료 | 교통사고요인 분석/교통안전 정책활용 | 교통정보연구원 | - | 정형 | 실시간 연계 |
| 사고 이력정보 | 교통사고 요인 분석 | 손해보험사 | - | 정형 | 실시간연계 |
| 기상정보 | 기상 파악 | 기상청 | COMIS | 정형 | 실시간 연계 |
| 교통사고관리정보 | 교통사고 요인 분석 | 경찰청 | 교통사고관리 시스템(TAMS) | 정형 | 요청시 연계 |
| 도로상대 정보 | 교통안전 정책활용 | 국토해양부 | 지능형교통체계 (ITS) | 정형 | 실시간 연계 |
| 도로관리 정보 | | | | | |
| 차량검지 정보 | | | | | |



(사) 추진 로드맵

- 교통사고 감소 정책활용과제의 향후 활용방안은 경찰청의 업무와 관련 조직, 관련기관간 협업, 빅데이터 공통기반의 원활한 환경구축, 정보공동활용 관련 법제도 개선 등의 사안을 고려하여, 단계별 이행계획(안)을 수립하여 추진하는 것이 중요함



(아) 기대효과

- 정성적 기대효과

| 구분 | 기대효과 | |
|--------|-----------|---|
| 수요자 측면 | 지자체 관점 | ✓ 주요교통사고 저점별 개선방안 마련, 신호등, 과속방지턱 설치 등의 교통사고 위험요소 제거를 위한 현실적 효과성이 높은 우선정책대상 식별 가능 |
| | 경찰청 관점 | ✓ 기존 분석 범위에 비정형 형식의 데이터가 추가라 혁신 분석을 통한 정책 합류 구현 |
| | 교통관리센터 관점 | ✓ 연계 결정구간, 도로개소, 장애물 등 위험요소를 실시간 감지하여 운전자에게 교통정보를 제공하고 돌발상황 발생시 신속하게 처리함으로써 2차 사고를 예방 |
| 수혜자 측면 | 국민 | ✓ 주민 요청 등의 의견수렴 및 교통안전 개선 사항의 정책반영으로 실질적 교통사고 감소 |

- 정량적 기대효과

- 주민개선의견 반영을 통한 교통안전시설물 설치를 통해 `15년 약 7.1억 원의 절감편익이 예상되며, 4년간(`15년~`18년) 년평균 약 12.4억 원의 절감편익의 정량적 효과가 기대됨

| 이행과제 | 성과 지표명 | 기대효과 내용 | 산출 방식 | 산출금액(년간/억 원) | | | | |
|------------------------|-------------------------------|---|--|---|---|---|---|-------|
| | | | | 2014년 | 2015년 | 2016년 | 2017년 | 2018년 |
| 주민개선의견 교통사고 감소체계 구축 방안 | 교통사고 감소에 따른 절감편익 | 주민개선의견 반영을 통한 교통안전시설물 설치에 따른 교통사고 발생률 감소에 따른 절감편익 | $E = \alpha(\beta_0 - \beta_1) \cdot Y$ $\beta_0 = \text{교통사고 절감 편익 비율} = 36.723 \text{ (천원/2인)}$ $\beta_1 = \text{시설물 기준 해당 안전시설물 설치 전 유 교통사고 발생률}$ $Y = \text{교통안전시설물 우선순위 반영 효과 반영비율}$ <p>교통사고 감소비용: $36,723 \times ((2,664 - 2,086) / 3) = 7,075,298,000 \text{ 원}$</p> | 교통안전 시설물 우선순위 반영 효과(10%) 70.8억(10.10%) = 7.1억 | 교통안전 시설물 우선순위 반영 효과(15%) 70.8억(10.15%) = 10.6억 | 교통안전 시설물 우선순위 반영 효과(20%) 70.8억(10.20%) = 14.2억 | 교통안전 시설물 우선순위 반영 효과(25%) 70.8억(10.25%) = 17.7억 | |
| 합계 | | | | 7.1 | 10.6 | 14.2 | 17.7 | |
| 전제사항 | '교통안전시설물 우선순위 반영효과를 10%라고 가정함 | | | 관련 Factor *2011년 교통사고 건수: 221,711 건 *2011년 교통사고 사망자: 5,229 명 *2011년 교통사고 부상자: 341,391명 *자랑금액비율: 2,215,900,000/1000000 *인명피해비용: 2,064,196,700,000원 *사상자피해비용: 2,706,802,804,000원 (부상자) 354,982,122,000원(사망자) 교통사고 건당 발생비용: 36,723,000원 | | | | |

(5) 국가미래전략 수립 정책 활용과제 목표 모델 수립

(가) 정책 활용과제 정의서

- 미래 이슈 분석기반 마련과 운영을 위해 현 기획재정부 장기전략국의 미래이슈 탐색업무를 기반으로 국가 미래전략 및 위기대응 전략 수립업무 모델을 설계함

| | | | | | | | | | | | | | |
|-------|--|---|---|-------|-------|-------|-------|-------|---|---|---|---|---|
| 과제명 | • 국가 미래전략 및 위기대응 전략 수립 | 담당 부서 | • 기획재정부 - 장기전략국 | | | | | | | | | | |
| 과제 개요 | • 정부 보유 데이터, 국토 자원 및 사회기반시설의 사물센싱 데이터 등 다양한 정보를 분석하여 국가 미래전략 및 위기대응 전략을 수립할 수 있는 미래 이슈 분석기반을 마련 | 정책 단계 | <table border="1"> <tr> <td>의제 설정</td> <td>정책 형성</td> <td>정책 결정</td> <td>정책 집행</td> <td>정책 평가</td> </tr> <tr> <td>●</td> <td>●</td> <td>○</td> <td>○</td> <td>○</td> </tr> </table> | 의제 설정 | 정책 형성 | 정책 결정 | 정책 집행 | 정책 평가 | ● | ● | ○ | ○ | ○ |
| 의제 설정 | 정책 형성 | 정책 결정 | 정책 집행 | 정책 평가 | | | | | | | | | |
| ● | ● | ○ | ○ | ○ | | | | | | | | | |
| 과제 내용 | • 국가의 미래 이슈와 잠재적 위기를 예측할 수 있는 빅데이터 기반 Horizon Scanning Center 도입 ※ 정형 데이터 : 인구 통계, 경제 지표 등 비정형 데이터 : SNS, 뉴스기사, 학술데이터 등 • 미래 예측 정보 등 미래 이슈 정보를 수집/분석하고, 정책 방향성 제시를 위한 미래 키워드 도출 | | | | | | | | | | | | |
| 개선 방향 | 업무절차 관점 | 참조정보 및 환경요인 관점 | | | | | | | | | | | |
| | • 국가미래전략 및 위기대응 전략수립 업무프로세스 정립 필요 - 현재 수행 중인 업무프로세스를 기반으로 선진사례와 민간사례의 장단점을 도출하여 업무프로세스 정립 | • 이슈도출을 위한 트렌드 자료와 분석을 위한 기초자료 활용강화 - 통계자료, SNS 분석자료, 학술데이터 등의 정형, 비정형, 반정형 데이터를 활용하여 업무수행의 신속성과 객관성을 강화 | | | | | | | | | | | |

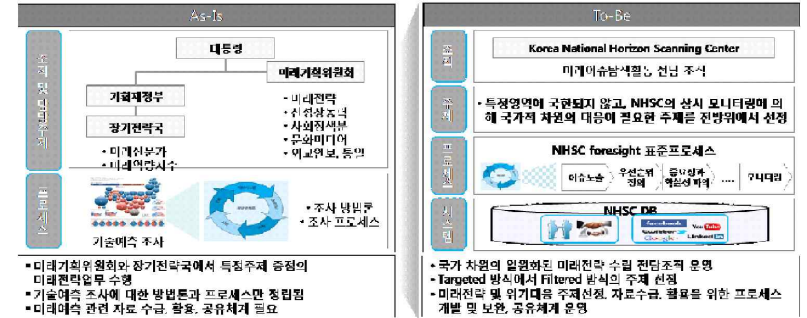
(나) 설계방향

- 국가 미래전략 및 위기대응 전략 수립업무를 수행 할 수 있는 Horizon Scanning Center 도입 및 업무프로세스를 정립함

| To-Be 방향 | Approach 방안 | 빅데이터 AV의 관점 | 내용 | 기대효과 | 예시 |
|-----------------------------------|---|---|---|--|----|
| 이슈도출을 위한 트렌드 자료와 분석을 위한 기초자료 활용강화 | Horizon Scanning Center 도입을 통한 자유수급체계 강화 및 업무프로세스 파생됨 | <ul style="list-style-type: none"> Variety Velocity Volume | <ul style="list-style-type: none"> 기초정보의 빅데이터 분석으로 미래 키워드 도출 참소정보 확대 <ul style="list-style-type: none"> - 종류 : 정형 + 비정형 - 정보량(▲) | 1) 기존에 머물렀던 기초 데이터 보강 2) 의사결정을 위한 과학적이고 체계화된 의사결정 지원체계 제공 | |
| 국가미래전략 및 위기대응 전략수립 업무프로세스 재정립 | | <ul style="list-style-type: none"> Variety Velocity | <ul style="list-style-type: none"> 미래연구 정의 및 미래연구 방법론의 고려사항에 따른 프로세스 설계 원리수립 - 선진사례의 Lessons & Learned를 현재 수행중인 장기전략수립업무 프로세스에 반영 | 1) 장기전략수립 업무프로세스 재정립을 통한 정책 과제 도출 업무의 신속성 및 정확성 개선 | |

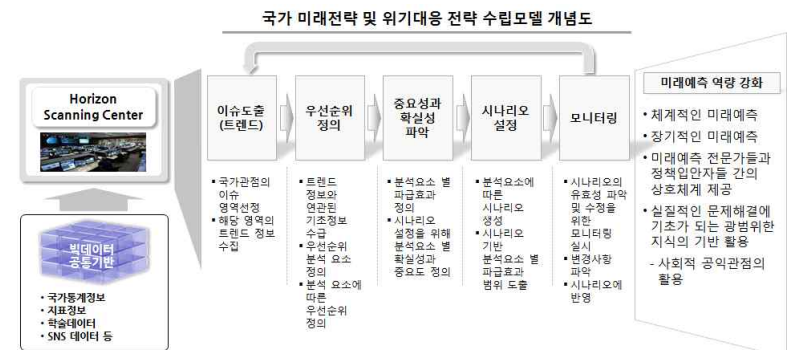
(다) As-Is & To-Be

- 국가 미래전략 및 위기대응 전략 수립업무를 수행 할 수 있는 Horizon Scanning Center 도입 및 업무프로세스를 정립함

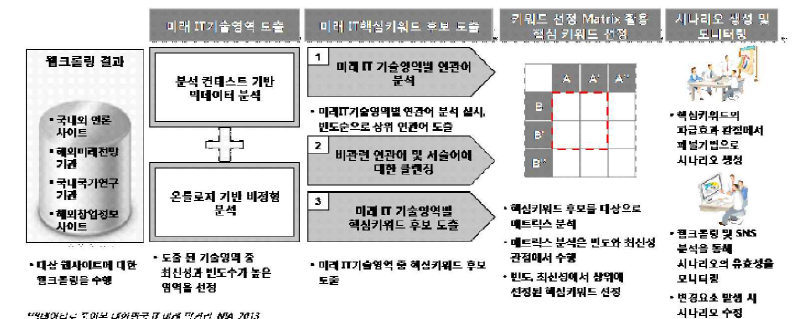


(라) 운영 모델 구현 방안

- 빅데이터 공동기반을 통해 미래에 발생 가능한 잠재적인 사건에 대한 정보를 수집하고, 이슈 도출을 통해 미래대응을 하는 것임



- 국가 미래전략 및 위기대응 전략 수립모델의 개념을 IT영역에 적용하여 이슈도출에서 모니터링까지의 과정 예시



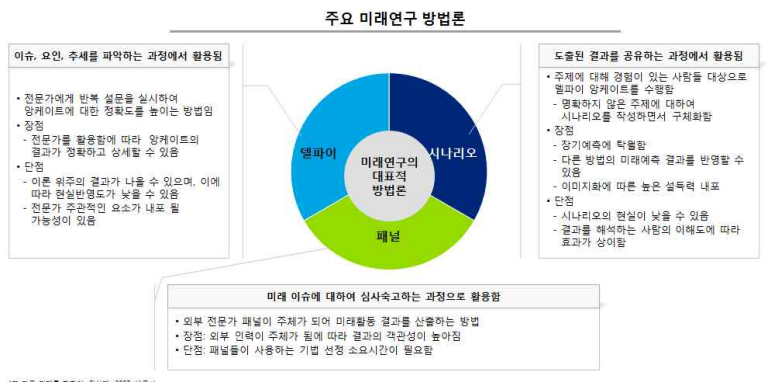
“인생이라는 도마에 대한민국 IT 내세워서” NHA, 2013

- Horizon Scanning Center의 3가지 운영형태 중 국가의 미래이슈 탐색 능력을 향상하고 범위가 범용적인 Korean Horizon Scanning Center 모델을 권고함

| Horizon Scanning Center 형태 | | |
|---|---|---|
| 국가의 미래 역경준대를 위한 Horizon Scanning 방안 | | |
| <ul style="list-style-type: none"> • 미래연구 수요 시스템 구축과 정치적 지지, 국회의 적극적인 참여 및 활성화와 같은 지원 촉진 • 짧은 정책 주기로 인한 단기적 정확적 미래연구에서의 탈피 • 공무원과 정치인의 교육 및 훈련 강화, 전문 네트워크 구축 및 지식 공유 활성화 | | |
| Horizon Scanning Korean Project | Korean Horizon Scanning Center | Horizon Scanning Center for Security (Tech.) |
| <ul style="list-style-type: none"> • 신진사채 - 네덜란드 Horizon Scanning Project • 목적 - 공공 정책 의제와 관련된 한국 전반의 미래 트렌드 및 이슈를 탐지하고 평가하여 정부의 활용방안과 의회 및 정부, 대중들의 미래이슈 탐색 접근법 연구 • 정책 영역 - 모든 공공 정책 분야를 다룸 • 방법 - 프로젝트 형태임에 따라 임시적으로 운영 - 소규모의 내부 프로젝트 팀이나 외부 조직과 연합하여 운영 | <ul style="list-style-type: none"> • 신진사채 - 영국 Foresight Programme • 목적 - 정부 행정조직의 미래이슈탐색 능력 향상을 위한 필수적인 미래이슈 방법, 절차, 전략적인 지원 제공 • 정책 영역 - 모든 공공 정책 분야를 다룸 • 방법 - 정부조직에 내부에 영장력 있는 미래이슈탐색을 위한 조직을 구성 - 기존의 조직 예를 들면 기획재정부나 교육과학기술부 내에 미래이슈탐색 조직 운영 | <ul style="list-style-type: none"> • 신진사채 - 싱가포르의 Risk Assessment and Horizon Scanning Programme • 목적 - 안보와 미래과학기술과 관련된 위협요인/기회요인을 탐색/분석하여 관련 정부 조직 및 이해관계자에게 제공 • 정책 영역 - 안보와 과학기술에 한정 • 방법 - 총리실 내 국가안정보장위원회나 국가과학기술위원회 내에 미래이슈탐색 센터 운영 |

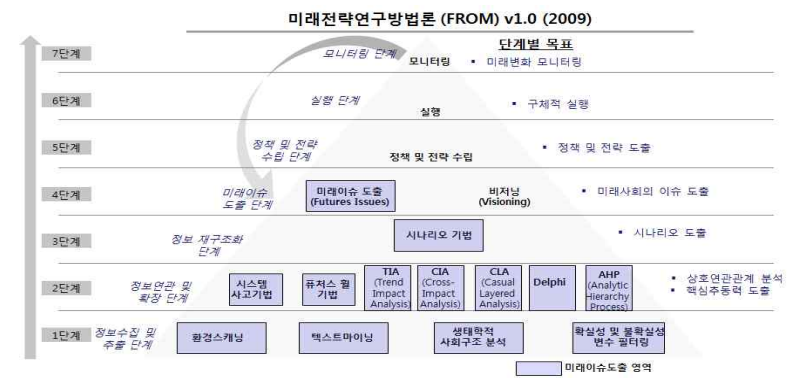
국회과학기술위원회 국정감사 국정질문 제 2011-0024호 국정감사답변문답

- 미래예측에 주로 활용되는 기법은 델파이, 시나리오, 패널 기법이며 그 특성과 장/단점은 다음과 같음



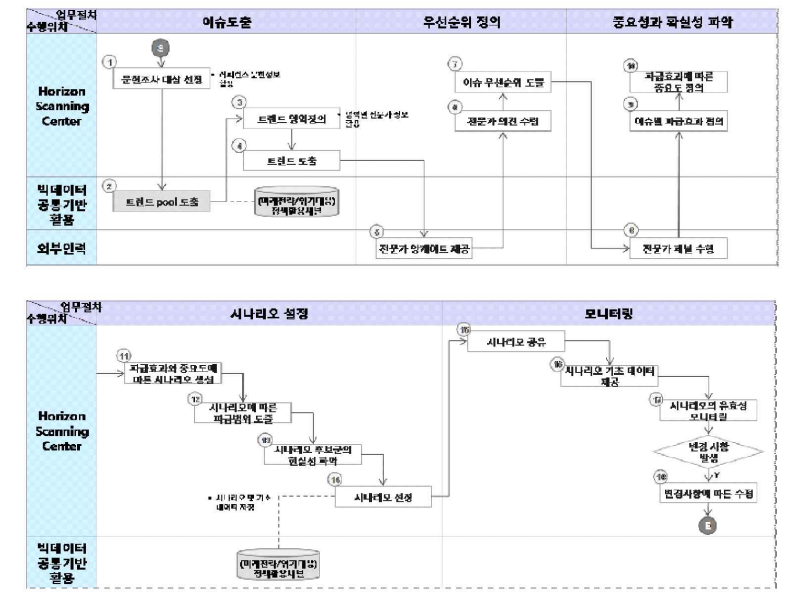
*본 미션을 참조, 김성태, 2007, 본문서

- 한국정보화 진흥원이 제시한 미래전략연구 방법론은 7단계 업무를 통해 전략 수립 업무를 수행하도록 함



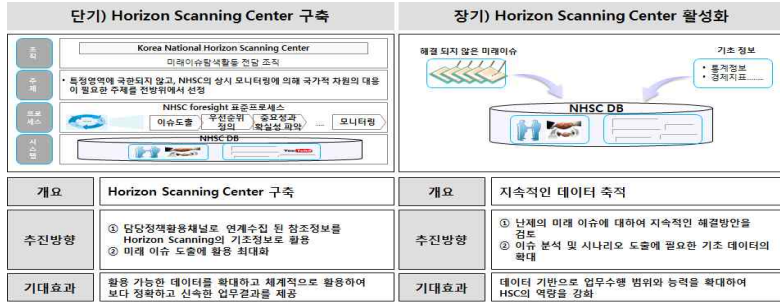
(마) 업무 프로세스 Map

- 미래전략 및 위기대응 전략 수립업무는 이슈도출, 우선순위 정의, 중요성과 확실성 파악, 시나리오 설정, 모니터링으로 구성되어 수행함



(바) 추진 로드맵

- Horizon Scanning Center를 구축하여 미래 이슈 도출 업무를 담당하도록 하고 빅데이터 공통기반에 지속적인 데이터의 축적과 고도화를 통해 운영모델 개선함



(사) 기대효과

- 국가 미래 전략 수립 업무에 따른 기대효과는 위기대응 및 기회 포착 역량 함양, 전문가 네트워크 형성 및 국격 제고에 있음

| 구분 | 내용 |
|---|---|
| <p>잠재적 위기과 기회에 대하여 대응 또는 포착 할 수 있는 역량 함양</p> | <ul style="list-style-type: none"> ✓ 발생 가능성이 높은 위기에 대한 선제적 대응을 통해 안정적인 국정운영 및 민생수준을 보장 ✓ 현실적으로 해결 불가능한 미래 이슈에 대한 DB화를 통하여 지속적 해결방안 모색 ✓ 잠재적 기회의 파급효과를 국가발전에서 분석하여 기업의 경쟁력 강화 |
| <p>전문가 네트워크 형성</p> | <ul style="list-style-type: none"> ✓ 분야별 전문가 DB 구축으로 전문가 참여도 향상하고, 업무결과 수준 향상 ✓ 전문가 간 네트워크 형성을 통해 영역 내외에서 시너지 효과 유도 |
| <p>해외 정부기관과의 공동 대응을 통한 국격 제고</p> | <ul style="list-style-type: none"> ✓ 위기나 기회의 파급범위가 글로벌인 경우 공동 대응을 통해 한국의 위상을 높임 ✓ 위기에 대한 부담과 기회에 대한 이익을 타 정부기관과 분배하여, 위기 감소와 기회 획득을 실현 |

마) 학술데이터 공유-활용 모델 수립

(1) 개요

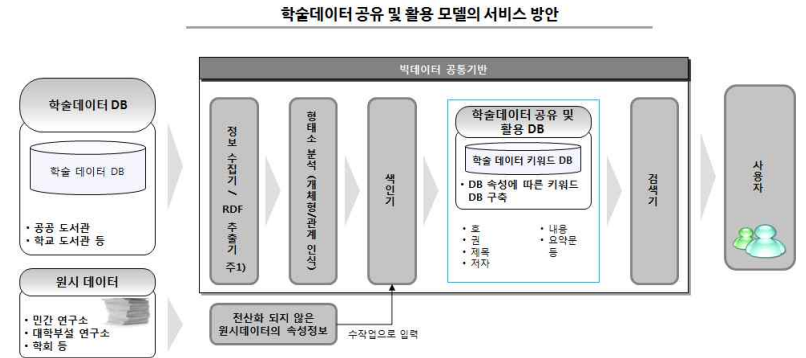
- 학술데이터의 공유 및 활용 저해요인을 고려하여 단기적, 중기적, 장기적 관점의 공유 및 활용 모델을 수립함

학술데이터 공유-활용 과제 정의서

| 과제명 | 과제개요 | 정책단계 | 의제 선정 | 정책 우선 | 정책 우선 | 정책 심화 | 정책 평가 |
|--|--|---|-------|-------|-------|-------|-------|
| <p>■ 학술데이터 공유 및 활용 모델 수립</p> | <p>■ 학술데이터 생성, 관리하고 있는 학술데이터의 공유 및 활용을 활성화 하기 위한 학술데이터 공유 및 활용 모델을 수립</p> | <p>정책 단계</p> <p>● ● ● ● ●</p> | ● | ● | ● | ● | ● |
| <p>■ 학술데이터의 공유 및 개방 조건과 개방 및 제공불가 사유에서 도출한 범형성과 국내/외 참조사례에서 도출된 범형성이 적용된 학술데이터 공유 및 활용 모델 수립</p> <p>- 학술기관 내 정책을 포함한 관련 정책과 저작권 문제가 고려된 공유 및 활용 모델</p> <p>- 학술데이터 뿐만 아니라 학술자료 저자 정보 활용을 통한 협업적인 모델</p> | ● | | ● | ● | ● | ● | |
| <p>■ 학술데이터의 공유 및 개방 조건과 개방 및 제공불가 사유에서 도출한 범형성과 국내/외 참조사례에서 도출된 범형성이 적용된 학술데이터 공유 및 활용 모델 수립</p> <p>- 학술기관 내 정책을 포함한 관련 정책과 저작권 문제가 고려된 공유 및 활용 모델</p> <p>- 학술데이터 뿐만 아니라 학술자료 저자 정보 활용을 통한 협업적인 모델</p> | ● | | ● | ● | ● | ● | |
| <p>개선 방향</p> <p>■ 학술데이터의 meta DB를 활용하여 비용효율성과 관리효율성 향상</p> <p>- 학술데이터의 속성정보를 관리하는 학술데이터 portal 도입</p> | <p>■ 학술데이터 저작권 비용과 권리성 고려</p> <p>- 저작권 획득을 및 수급을 통해 사용자에게 편리한 공유/활용 서비스 제공</p> | <p>■ 학술데이터의 장벽을 통해 범접 및 자발적인 학술데이터 공유/활용 유도</p> <p>- 학술데이터의 소셜 네트워크 기능을 제공하여 협업적인 및 다양한 전문가 정보 제공</p> | | | | | |

(2) 학술데이터 공유 및 활용 모델의 서비스 방안

- 빅데이터 공통기반을 활용하여 외부 학술데이터 DB의 속성정보를 키워드화 하여 색인하고 검색기능을 통해 서비스를 제공하며, DB화 되지 않은 학술데이터의 속성정보도 수작업으로 입력하여 "학술데이터 공유 및 활용 DB"에 저장함



주1) Resource Description Framework의 약자로서 제목, 저자, 최종 수정일, 저작권과 같은 문서에 관한 메타 데이터를 xml 기반으로 표현함

(3) 학술데이터 공유 및 활용 모델 정의

- 학술데이터 공유 및 활용 모델은 학술자료를 수급하여 제공하는 모델, 학술데이터의 meta 정보를 제공하는 모델, 전문가 portal을 통해 학술데이터 저자를 주축으로 하는 모델이 있음

학술데이터 공유 및 활용 모델

| 학술데이터 meta DB 기반 공유 및 활용 Portal | 학술데이터 DB 기반 공유 및 활용 Portal | 학술데이터 및 학술자료 저자 DB Portal |
|--|--|---|
| <p>정의</p> <ul style="list-style-type: none"> - 온라인 학술데이터 portal을 구축하고 학술데이터의 제목, 저자, 요약정보, 실제 위치정보와 같은 meta 정보를 제공함 <p>목적</p> <ul style="list-style-type: none"> - 비용효율적인 학술데이터 정보를 제공함 <p>장점</p> <ul style="list-style-type: none"> - 소규모의 운영 시스템으로 운영이 가능함 - 실제 데이터를 수급하는 것이 아니라 학술자료에 대한 정보만 수급하여 제공함에 따라 자료 수급이 신속하게 이루어 질 수 있음 <p>단점</p> <ul style="list-style-type: none"> - 학술자료 볼 저작권 내역이 상이하여 사용자의 사용성이 낮을 수 있음 - 온라인 도서관 웹사이트나 Google과 같은 온라인 검색사이트와 기능이 중첩 될 수 있음 | <p>정의</p> <ul style="list-style-type: none"> - 온라인 학술데이터 portal을 구축하고 학술데이터 저장 및 제공과 같은 관리/서비스 기능을 구현함 <p>목적</p> <ul style="list-style-type: none"> - 실제 학술데이터 파일을 보유함에 따라 공유 및 활용에 있어서 체계적이고 신속한 서비스를 제공함 <p>장점</p> <ul style="list-style-type: none"> - 실제 학술데이터 파일을 보유하고 있어서 정제이나 저작권과 공유/활용 제어 문제를 해결 할 수 있음 - 다양한 서비스 제공이 가능함 <p>단점</p> <ul style="list-style-type: none"> - 대규모의 portal 운영 시스템이 필요함 - 학술데이터의 수급, 저작권 획득 관리에 상대적으로 많은 인력이 필요함 | <p>정의</p> <ul style="list-style-type: none"> - 학술자료 저자 정보를 제공하고, 저자 간 네트워크 형성을 유도하여 저자 스스로 학술 데이터를 공유 하도록 함 <p>목적</p> <ul style="list-style-type: none"> - 학술자료 저자 중심의 온라인 Portal을 통해 학술데이터 뿐만 아니라 전문가 정보를 제공함 <p>장점</p> <ul style="list-style-type: none"> - 전문가 간 네트워크 형성을 통해 학술데이터 뿐만 아니라 전문지식 공유가 가능함 - 학술자료 저자에 의한 학술데이터 조달을 통해 수급이나 저작권 문제를 사전에 예방함 <p>단점</p> <ul style="list-style-type: none"> - 대규모의 portal 운영 시스템이 필요함 - 활성화를 위한 지속적인 관리방안이 필요함 |

(4) 학술데이터 공유 및 활용 모델 평가

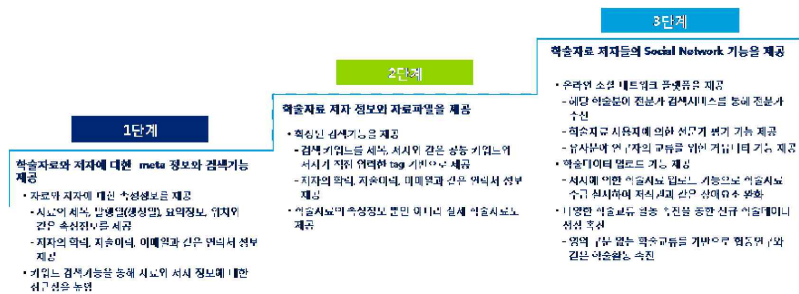
- 학술데이터 공유 및 활용 모델에 대한 평가는 비용, 효과, 장애요소 관점에서 실시하여, 비용 효율적이며 구축/운영에 따른 파급효과를 장애요소 없이 원활히 발생할 수 있는 모델을 선정하도록 함

학술데이터 공유 및 활용의 비용대비 효과 분석

| | 학술데이터 meta DB 기반 공유 및 활용 Portal | 학술데이터 DB 기반 공유 및 활용 Portal | 학술데이터 및 학술자료 저자 DB Portal |
|------|---|---|--|
| 비용 | <ul style="list-style-type: none"> 학술데이터 공유 활용 모델 별 구현을 위한 시스템 및 DB 설계 등의 금전적 비용 | <ul style="list-style-type: none"> meta DB를 확장함에 따라 시스템 성능, 저장공간, 저작권 비용이 낮음 | <ul style="list-style-type: none"> 실제 학술데이터를 제공하기 위해 저장 및 검색 능력을 제공 할 수 있는 시스템이 필요하고, 데이터 수급을 위한 저작권 비용이 발생함 |
| 효과 | <ul style="list-style-type: none"> 학술데이터 공유 활용 모델 별 구현 후의 관련 데이터 인용 및 활용 건수 향상 등의 효과 | <ul style="list-style-type: none"> 학술데이터 정보를 관리/제공하여 gateway 역할을 수행함 실제 파생은 제공하지 않음 | <ul style="list-style-type: none"> 학술데이터 속성정보 뿐만 아니라 실제 학술데이터도 제공함 학술자료의 전이가 이루어지기 때문에 사물적인 학술데이터 수급이 가능하고, 원본가 정보에 기반한 열람요구를 기대할 수 있음 |
| 장래요구 | <ul style="list-style-type: none"> 학술데이터 공유 활용 모델 별 구현을 위한 용이성 | <ul style="list-style-type: none"> 학술데이터 속성 정보만 수급/제공함에 따라 장애요소가 없음 | <ul style="list-style-type: none"> 학술자료의 전이가 속건할 수 있는 지속적인 관리방안이 필요함 |

(5) 학술데이터의 공유 및 활용 로드맵

- 학술데이터의 공유 및 활용을 위해 단기적으로는 비용 및 장애요소 관점에서 우수한 모델을 도입하고 지속적인 고도화를 통해 효과관점의 장기 모델을 운영하도록 함



(6) 기대효과

- 학술데이터의 공유 및 활용 과제의 정성적 기대효과를 수요자 측면과 수혜자 측면으로 도출함

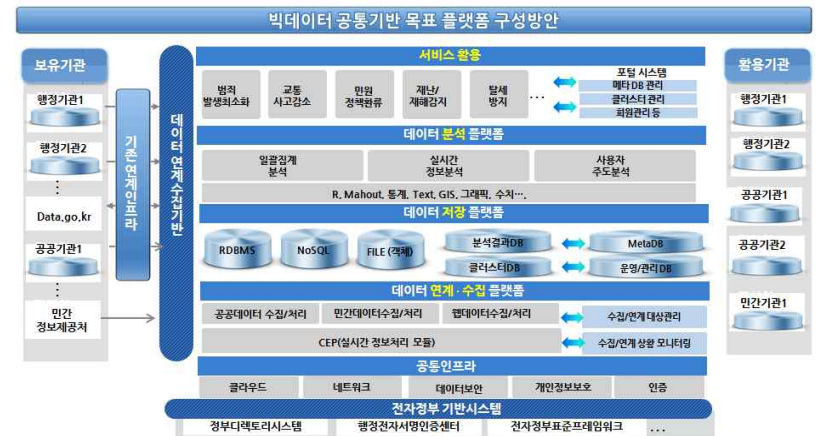
| 구분 | 기대효과 | |
|--------|-----------------|---|
| 수요자 측면 | 학술데이터 관리체계 향상 | ✓ 국가적 전자정보 콘텐츠의 일부분인 학술데이터 관리체계를 통하여 외부 학술데이터 DB와 일치 학술데이터에 대한 관리를 수행할 수 있음 |
| | 정책 정보의 품질향상 | ✓ 정책 정보 생성 시 학술데이터를 보다 신속하고 다양하게 참조하여 정책 정보의 품질을 향상 할 수 있음 |
| | 국내외 정보 협력체계 형성 | ✓ 학술데이터 관리체계를 활용하여 학술저자나 학술기관 간 협력체계를 지원할 수 있으며, 이를 통한 다양한 협력체계를 촉진 할 수 있음 |
| 수혜자 측면 | 학술데이터 활용 편의성 향상 | ✓ 학술 데이터 상이한 관리주체를 민원사항에 따라 한 곳에서 모든 학술데이터를 열람할 수 있음 |
| | 학술데이터의 다양화 | ✓ 서미나 자료와 같이 기존에 관리되지 않았던 학술데이터를 포함하여 다양한 학술데이터에 대한 접근이 가능해짐 |

5) 빅데이터 공통기반 및 활용과제 정보화 구축 방안

가) 정보화 구축방안 개요

- 빅데이터 공통기반 및 활용과제 구성을 위한 목표 플랫폼은 활용기관의 이용편의성을 감안하여, 데이터 연계/수집, 저장, 분석 플랫폼과 공통인프라를 구축함

[빅데이터 공통기반 목표 플랫폼 구성 개념도]



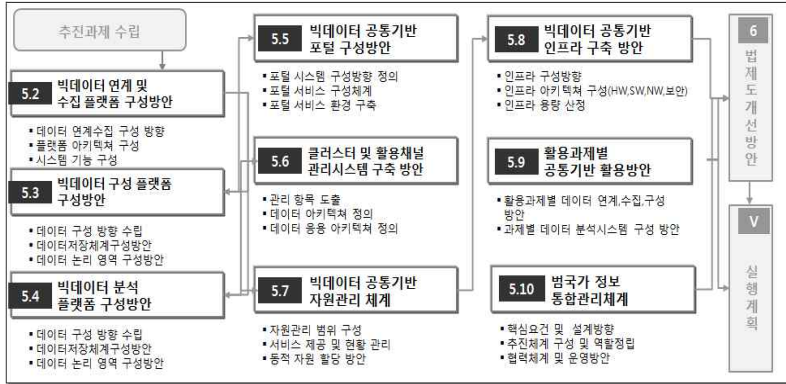
- 정책담당자의 정책수립 편의성 및 효율성을 제고하기 위해서 범국가적으로 데이터를 공동으로 활용할 수 있는 공통기반 및 정책활용과제 지원을 위한 시스템 구축방안을 수립함

[정보화 구축 목적 및 방법]

| | |
|-------|---|
| 목적 | <ul style="list-style-type: none"> 정책담당자의 정책수립 편의성 및 효율성을 제고하기 위해서 범국가적으로 데이터를 공동으로 활용할 수 있는 공통기반을 구축하고 정책활용과제 지원을 위한 시스템 구축 방안을 수립함 |
| 방법 | <ul style="list-style-type: none"> 빅데이터 공통기반 구축 방안 수립 빅데이터 공통기반은 데이터 연계 수집, 데이터 구성, 데이터 분석 단계별 시스템 구성 방안 수립 빅데이터 공통기반 및 정책활용과제 업무 설계에서 도출된 업무 설계(BPR) 내용을 바탕으로 시스템 구축방안 빅데이터 공통기반 플랫폼 설계를 위한 전문가 자문회의, 고객 검토된 내용 반영 |
| 수립 절차 | |

- 빅데이터 공통기반 및 활용방안을 위한 정보화 구축방안은 현황분석을 통해 도출된 추진과제를 중심으로 설계함
- 즉, 빅데이터 연계수집, 구성, 분석, 포털구성, 클러스터 및 활용채널 등 플랫폼 아키텍처를 설계하고, 인프라 구축 및 범국가 정보통합관리체계를 제시함

[미래모형 정립 Framework]

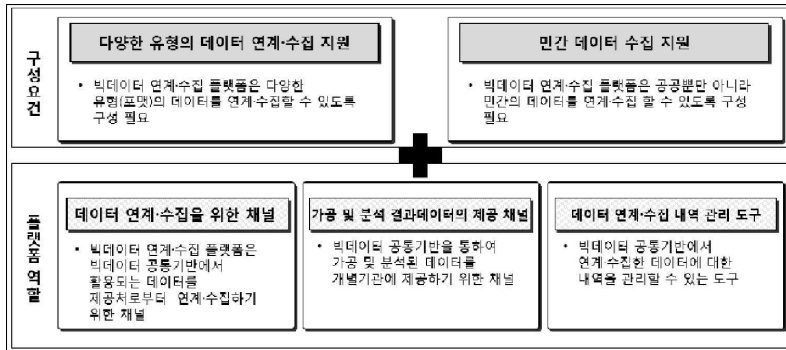


나) 빅데이터 연계-수집 플랫폼 구성 방안

(1) 빅데이터 연계-수집 플랫폼 구성 요건 및 역할

빅데이터 연계-수집 플랫폼의 역할은 빅데이터 공통기반에서 분석에 활용되는 데이터를 연계-수집할 수 있는 인프라이며, 연계-수집 대상은 행정-공공기관 및 민간을 포함하고 있음

[빅데이터 연계-수집 플랫폼 구성 요건 및 역할]



(2) 빅데이터 연계-수집을 위한 기술 요소 분석

빅데이터 공통기반은 다양한 환경에서 데이터를 연계-수집함에 따라 ESB 방식의 인프라 구성이 효과적일 것으로 판단되며, 행정정보공동이용시스템이 ESB방식으로 구성되어 있기 때문에 기술적 호환성도 높으며, 대용량 데이터 처리에 있어 유리함. ESB의 경우 초기 구축 비용이 높음에 따라 구성요소에 대한 단계적 구축을 통한 비용적 부담 해소 가능

[빅데이터 연계-수집 플랫폼 기술 요소 분석]

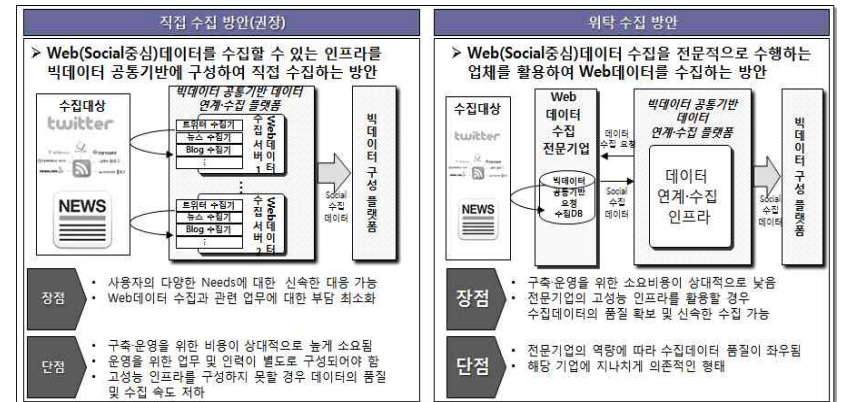
| 구분 | EAI(Point-to-Point) | EAI(Hub and Spoke) | ESB(Bus) |
|----|--|---|---|
| 특징 | <ul style="list-style-type: none"> 단순한 구조와 밀접한 연관성이 있는 어플리케이션 고장점 연결을 갖는 방식 복수 어플리케이션간 1:1 통합 어플리케이션간 연결 구조로 링크마다 연계로직과 I/F 관리됨 | <ul style="list-style-type: none"> 미들웨어(Hub)를 이용하여 비즈니스 로직 중심으로 어플리케이션을 연계하는 방식 단일 플랫폼인 허브시스템을 통해 어플리케이션을 통합하는 중앙집중식 연결(1:N) 비즈니스 로직을 중심으로 기업내 어플리케이션을 연계 호환성은 높아지지만 구축 비용이 높음 병다 중속적인 기술 사용 | <ul style="list-style-type: none"> 미들웨어(Bus)를 이용한 서비스 중심의 시스템 연계 방식으로 SOA 사상과 더불어 ESB라고 하는 경향이 강함 EAI에 따른 I/F와 SOA상상을 접목한 형태 서비스 중심의 프로세스 통합 표준기술(Web서비스, XML 등) 채용으로 이기종에 대한 표준통합 환경을 제공 일의 프로토콜 지원 버스(생산구조) 형태의 느슨한 결합 재사용 가능한 컴포넌트를 조합하는 방식 |
| 장점 | <ul style="list-style-type: none"> 복잡하지 않은 환경에서 단순한 연동일 때 빠른 적용 가능 | <ul style="list-style-type: none"> 신규 어플리케이션 도입시 확장 용이 개발 및 유지보수와 생산성/면의성 증대 | <ul style="list-style-type: none"> 표준기술 사용 및 서비스 단위 재사용으로 통합 비용 절감 버스 형태의 연결 구조로 확장성/유연성이 높음 행정정보공동이용시스템이 ESB방식을 중심으로 데이터를 연계함 |
| 단점 | <ul style="list-style-type: none"> 시스템의 독립성 보장이 어려움 시스템 수가 증가할수록 유지보수 비용이 급격히 증가함 확장성/유연성이 떨어짐 연계 로직의 재사용 불가 | <ul style="list-style-type: none"> 시스템 추가 확장이 EAI 제품의 Adaptor를 추가 구입하거나 지속적인 비용이 투입됨 중앙 Hub 강에서 전체 시스템에 영향을 미침(Single Point of Failure) | <ul style="list-style-type: none"> 초기 구축 비용이 높음 |

(3) 웹데이터 수집 방안

(가) 직접 수집 Vs 위탁 수집

- 데이터 공통기반에서 Social Web 데이터를 확보하는 방안은 인프라를 자체적으로 구성하여 직접 수집하는 방안과, Web 데이터 수집을 전문적으로 대행하는 업체를 활용하는 방안으로 구분될 수 있음

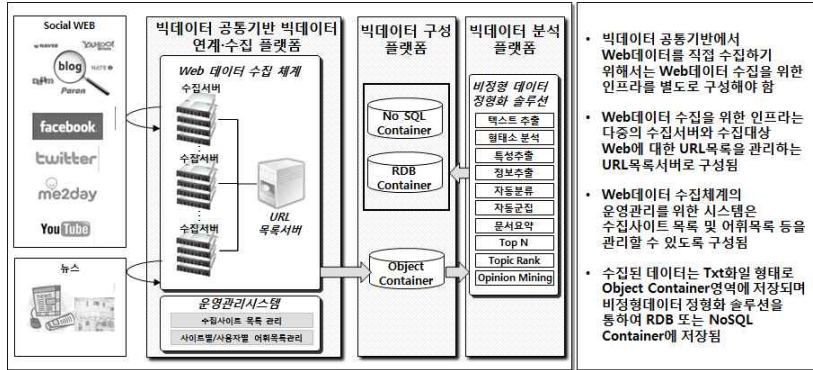
[직접 수집 Vs 위탁 수집]



(나) 직접 수집을 위한 아키텍처

- 빅데이터 공통기반에서 Web데이터를 직접 수집하여 활용할 경우 Web데이터 수집을 위한 별도의 인프라 구성 및 이를 운영관리하기 위한 시스템이 도입되어야 함

[Web데이터 직접 수집을 위한 아키텍처]

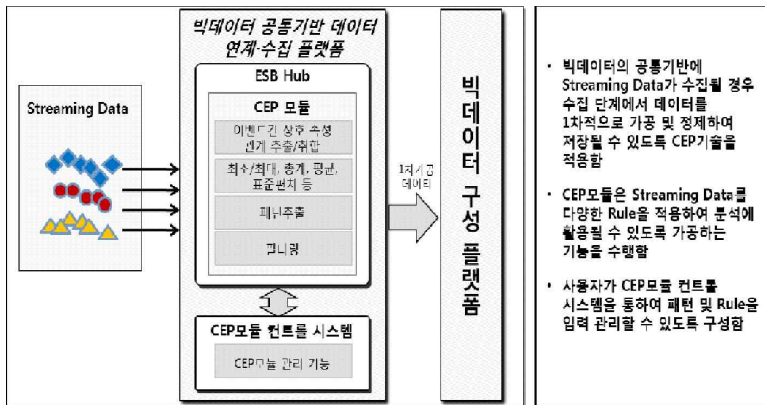


- 빅데이터 공통기반에서 Web데이터를 직접 수집하기 위해서는 Web데이터 수집을 위한 인프라를 별도로 구성해야 함
- Web데이터 수집을 위한 인프라는 다중의 수집서버와 수집대상 Web에 대한 URL목록을 관리하는 URL목록서버로 구성됨
- Web데이터 수집체계의 운영관리를 위한 시스템은 수집사이트 목록 및 어휘목록 등을 관리할 수 있도록 구성됨
- 수집된 데이터는 Txt화일 형태로 Object Container영역에 저장되며 비정형데이터 정형화 솔루션을 통하여 RDB 또는 NoSQL Container에 저장됨

(4) Streaming Data 처리 인프라 구성

- 빅데이터 공통기반에 수집되는 데이터가 Streaming Data일 경우 해당 데이터를 분석에 효과적으로 활용될 수 있도록 1차 가공을 할 수 있는 CEP(Complex Event Processing)모듈을 빅데이터 연계·수집 플랫폼에 적용함

[Streaming Data 처리 인프라 구성]

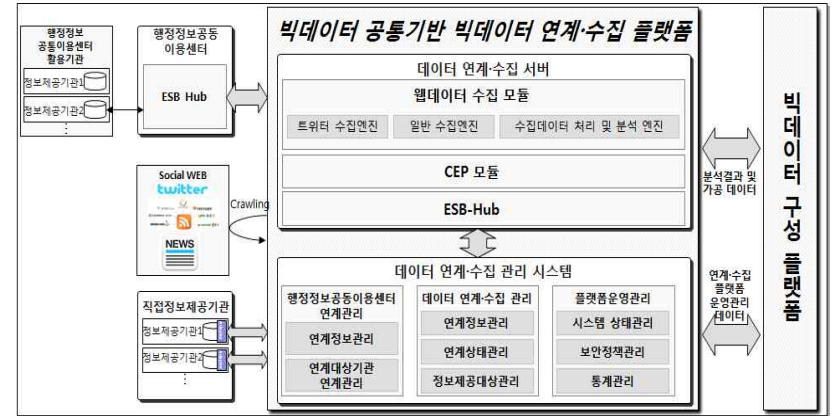


- 빅데이터의 공통기반에 Streaming Data가 수집될 경우 수집 단계에서 데이터를 1차적으로 가공 및 정제하여 저장될 수 있도록 CEP기술을 적용함
- CEP모듈은 Streaming Data를 다양한 Rule을 적용하여 분석에 활용될 수 있도록 가공하는 기능을 수행함
- 사용자가 CEP모듈 컨트롤 시스템을 통하여 패턴 및 Rule을 입력 관리할 수 있도록 구성함

(5) 빅데이터 연계·수집 플랫폼 아키텍처

- 빅데이터 공통기반 연계·수집 플랫폼은 웹데이터 수집 모듈, CEP모듈, ESB-Hub, 데이터 연계·수집 관리시스템으로 구성됨

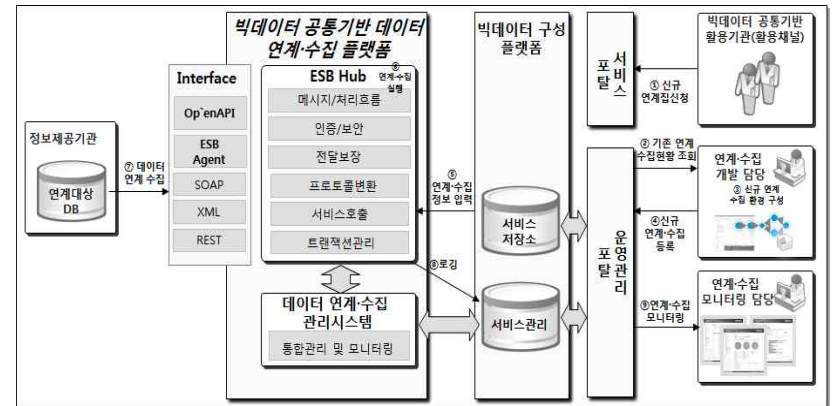
[빅데이터 연계·수집 플랫폼 아키텍처]



(6) 연계·수집 운영 모델

- 빅데이터 공통기반 활용기관(활용채널)에서 신규로 데이터 연계·수집을 요청할 경우 아래와 같은 절차를 통하여 운영됨

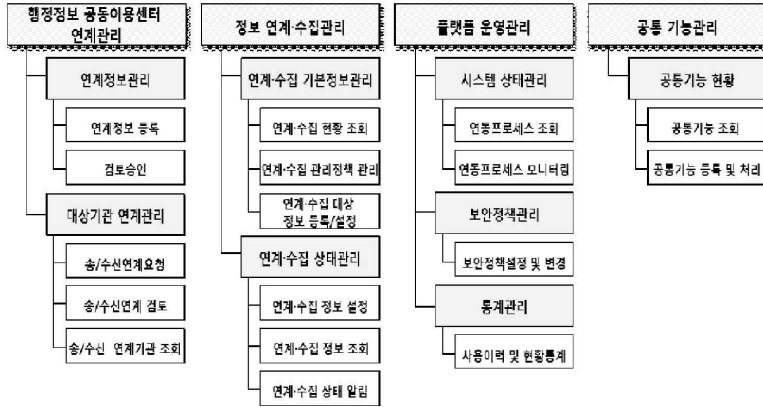
[빅데이터 연계·수집 운영모델]



(7) 빅데이터 연계·수집 관리시스템 기능 구성

- 데이터 연계·수집 관리시스템의 기능은 행정정보공동이용센터 연계관리, 정보 연계·수집 관리, 플랫폼 운영관리, 공통기능 관리 등으로 구성됨

[빅데이터 연계·수집 관리시스템 기능 구성]

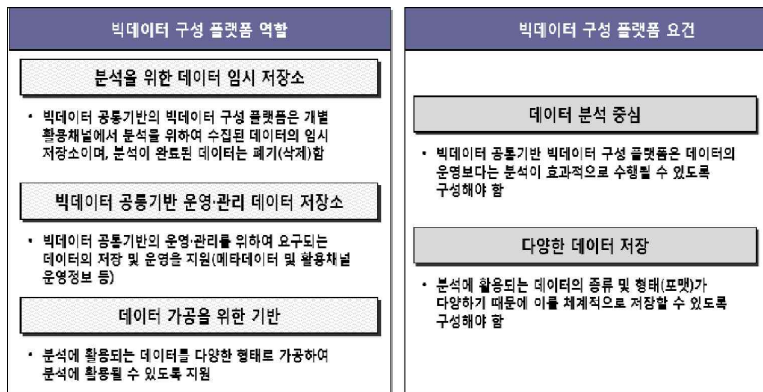


다) 빅데이터 구성 플랫폼 구축 방안

(1) 빅데이터 구성 플랫폼 역할 및 요건

- 빅데이터 공통기반의 데이터 저장은 수집된 데이터의 유형 및 분석의 형태를 고려하여 다양한 방식으로 저장할 수 있는 체계로 구성함

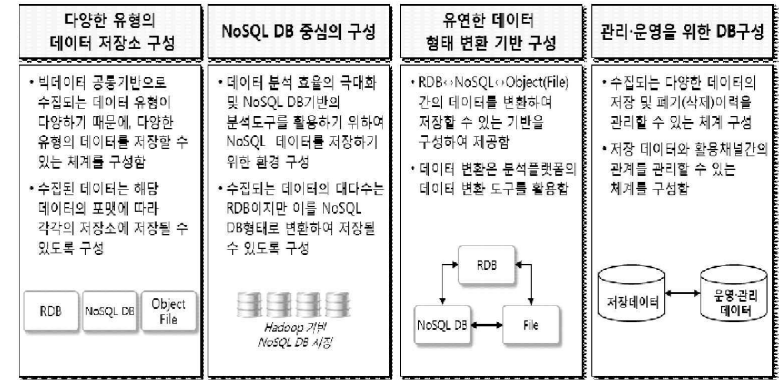
[빅데이터 구성 플랫폼 역할 및 요건]



(2) 빅데이터 구성 플랫폼 구축 방향성

- 빅데이터 공통기반 빅데이터 구성 플랫폼 구축을 위한 핵심 방향은 ① 다양한 유형의 데이터 저장소 구성 ② NoSQL DB 중심의 구성 ③ 유연한 데이터 형태 변환 기반 구성 ④ 관리·운영을 위한 DB구성임

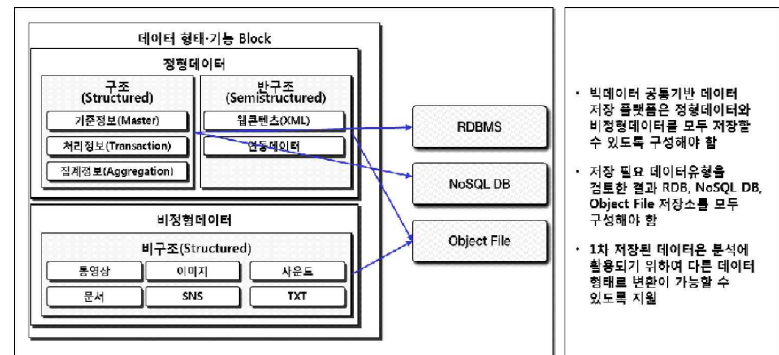
[빅데이터 구성 플랫폼 구성 방향성]



(3) 저장 필요 데이터 및 저장 방식

- 빅데이터 공통기반의 빅데이터 구성 플랫폼은 수집되는 다양한 유형의 데이터를 저장하기 위하여 RDB, NoSQL DB, Object File 저장소가 각각 필요함

[저장 필요 데이터 및 저장 방식]



(4) 분석을 위한 데이터 저장 체계 비교

- 빅데이터 분석을 위한 데이터 저장방식은 크게 RDB CLUSTER, MEMORY CACHE, NoSQL방식이 있으며, 이중 가장 효율적인 방식은 NoSQL방식임

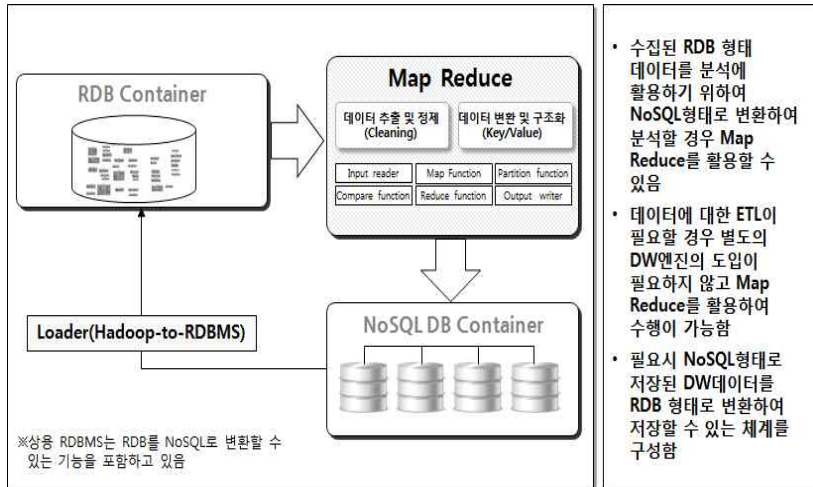
[분석을 위한 데이터 저장체계 비교]

| 유형 | Sample Solutions | 개요 | 장점 | 단점 |
|--------------|-----------------------------------|------------------------------------|---|--|
| RDB CLUSTER | OracleRAC, MySQL cluster | 분산 확장과 안정성을 강화한 RDB솔루션 | <ul style="list-style-type: none"> 최소한의 개별 비용으로 도입이 가능 비용 예측이 가능 상용화된 지원이 가능 | <ul style="list-style-type: none"> 확장 비용이 상대적으로 높으며, 적용 및 관리에 고급 기술이 필요함 읽기 성능만 향상 |
| MEMORY CACHE | Memcached, Coherence, Santa | RDB 앞단에 메모리 캐시를 두고 읽기 또는 읽기/쓰기를 캐싱 | <ul style="list-style-type: none"> 가격대비 효율성이 disk보다 높음 읽기/쓰기 성능을 크게 향상 가능 읽기/쓰기 수평 확장 가능 | <ul style="list-style-type: none"> 고가의 메모리 비용 메모리 크기에 제약이 있음 RDB를 저장소로 사용하지만 장애시 일부 데이터 손실 가능 |
| NoSQL | BigTable, Dynamo, Cassandra, etc. | 디스크 기반의 수평 확장이 뛰어나 데이터 저장소 | <ul style="list-style-type: none"> 읽기/쓰기 성능 및 확장성이 뛰어나 일반 저가 장비의 사용 가능 | <ul style="list-style-type: none"> Persistence Layer에 대한 신규 개발 필요 상용화된 지원이 어려울 수 있으나 최근 RDBMS 업체들이 NoSQL관련 제품을 출시하고 있음 |

(5) 데이터 형태 변환(RDB to NoSQL)

- 빅데이터 공통기반에서는 RDB형태 데이터의 정제작업을 위하여 MapReduce를 활용할 수 있으며, 별도의 DW 솔루션은 구성하지 않음

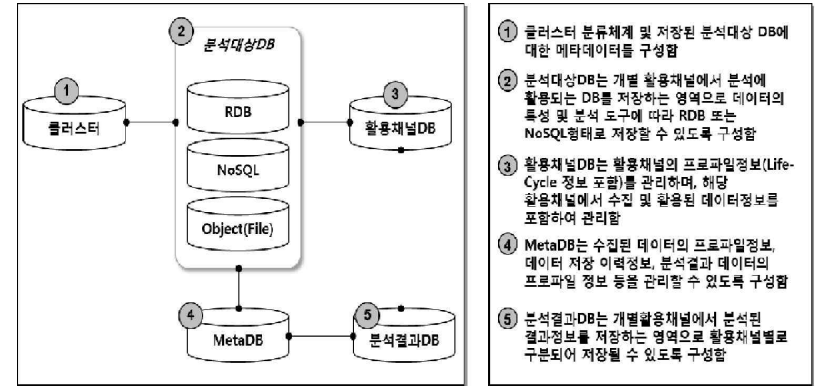
[데이터 형태 변환(RDB to NoSQL)]



(6) 관리·운영을 위한 DB 구성

빅데이터 공통기반의 물리적 데이터 구분은 분석대상 DB, 분석결과 DB, 활용채널 DB, MetaDB로 구분되며, 개별 DB간의 관계를 관리 할 수 있도록 구성함

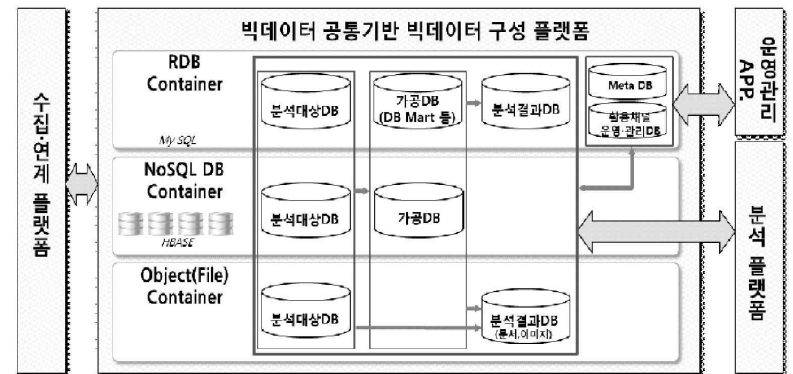
[관리·운영을 위한 주요 데이터]



(7) 빅데이터 구성 플랫폼 아키텍처

- 빅데이터 공통기반의 빅데이터 구성 플랫폼은 물리적으로 RDB Container, NoSQL Container, File Container로 구성되어 있으며, 논리DB의 물리적 속성에 따라 각각의 Container 영역에 구성되어 운영됨

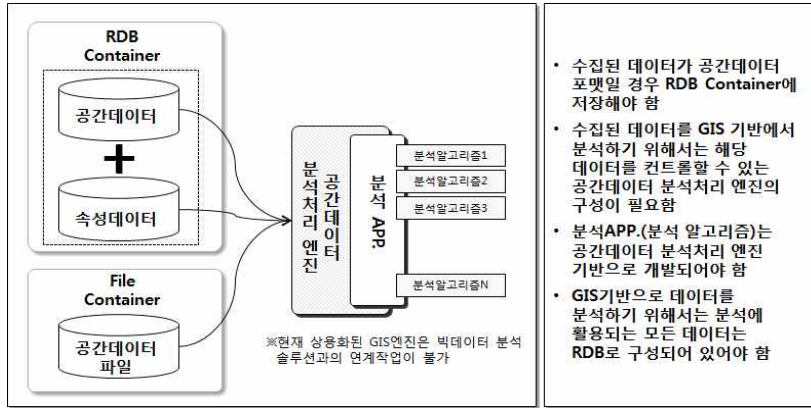
[빅데이터 구성 플랫폼 아키텍처]



(8) GIS데이터 분석 지원 방안

빅데이터 공통기반 빅데이터 구성 플랫폼에 저장된 데이터가 GIS데이터일 경우 RDB컨테이너에 저장되어야 하며, GIS 기반 분석을 위해서는 별도의 공간데이터 분석처리 엔진 도입이 필요함

[GIS데이터 분석 지원 방안]



라) 빅데이터 분석 플랫폼 구성 방안

다양한 대용량의 범국가적 데이터를 분석할 수 있는 기술 조사를 통해 빅데이터 시스템 구축 방향을 도출하고 향후 시스템을 구축하기 위한 설계방안을 수립함

(1) 빅데이터 분석 개요

추적, 활용, 확산 및 공유 단계를 거친 데이터 패러다임이 IT 기술의 발전과 모바일 환경 확대 등에 따른 정보 양의 폭발적인 증가에 따라 대량 데이터의 분석과 추론을 통한 가치 창출의 패러다임으로 변화되고 있음

빅데이터 분석 플랫폼을 통해 범국가적 데이터를 분석하여 제공하는 정보에 대한 기본적인 선정기준은 정책 수립단계별 정책 수립에 필요한 데이터 및 분석 결과를 지원하기 위한 정보를 기준으로 함

(2) 빅데이터 분석 기술

빅데이터 분석 기술 대부분의 분석기법들은 통계학과 전산학, 특히 기계학습/데이터 마이닝 분야에서 이미 사용되던 기법들이며, 분석기법들의 알고리즘을 대규모 데이터 처리에 맞도록 개선하여 빅 데이터 처리에 적용시키고 있음

최근 소셜미디어 등 비정형 데이터의 증가로 인해, 분석기법들 중에서 텍스트/오피니언 마이닝, 소셜 네트워크 분석 등이 주목을 받고 있음

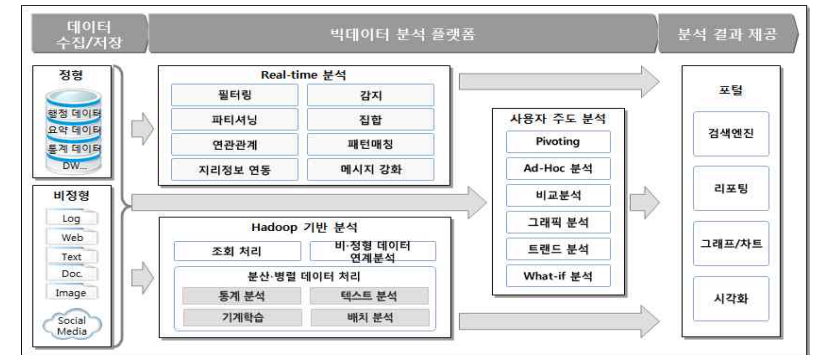
[주요 빅데이터 분석 기술]

| 기술 | 설명 |
|-------------------------------------|---|
| 데이터 마이닝 | 대용량의 데이터베이스에 있는 데이터로부터 패턴인식, 통계적 기법, 인공지능 기법 등을 이용하여 숨겨져 있는 데이터간의 상호 관련성 및 유용한 정보를 추출 |
| 텍스트 마이닝 | 텍스트 마이닝은 텍스트 기반의 데이터로부터 새로운 정보를 발견할 수 있도록 정보 검색, 추출, 체계화, 분석을 모두 아우르는 Text-processing 기술 및 처리 과정 |
| Opinion 마이닝 (or Sentiment analysis) | 텍스트 마이닝의 관련 분야로는 오피니언 마이닝, 혹은 평판 분석(Sentiment Analysis)라고 불리는 기술로 소셜미디어 등의 정형/비정형 텍스트의 긍정(Positive), 부정(Negative), 중립(Neutral)의 선호도(작자의 의견이나 감정 등)를 판별하는 기술 |
| 소셜웹 이슈분석 | 대용량 소셜 미디어를 언어분석 기반 정보추출을 통해 이슈를 탐지하고, 시간의 경과에 따라 유통되는 이슈의 전재과정을 모니터링하고 향후 추이를 분석하는 기술 |

(3) 빅데이터 분석 플랫폼 구축 방안

향후 빅데이터 비즈니스 요구에 따라 유연하게 대응하기 위해 빅데이터 분석 플랫폼은 크게 Real-time 분석, Hadoop 기반 분석, 사용자 주도 분석 영역으로 구분하여 구성함

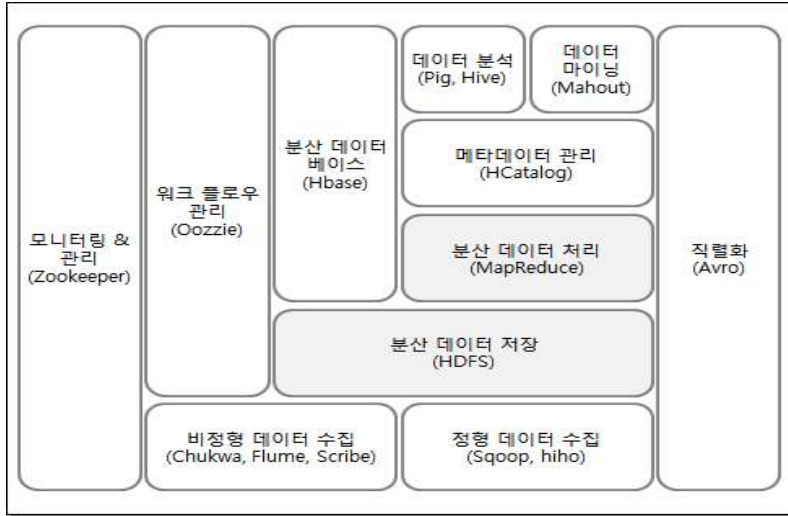
[빅데이터 분석 플랫폼 구성도]



(가) Hadoop 기반 분석

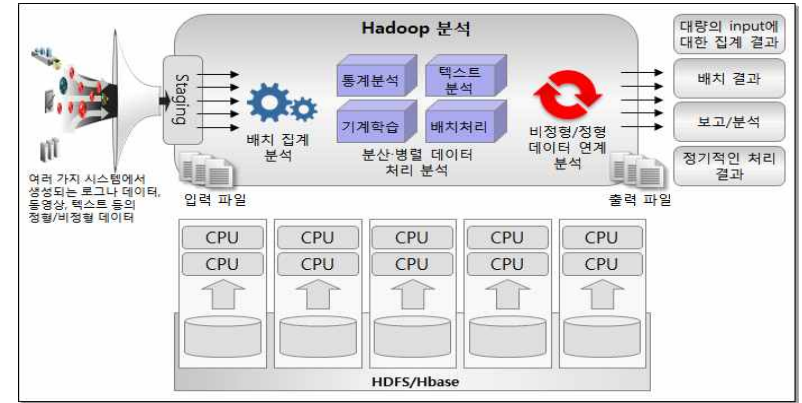
- 하둡은 구글의 분산파일시스템과 맵리듀스의 오픈소스 구현체로 대규모의 비정형 데이터 분석을 배치로 처리하는데 주로 사용되고 있으며, 비정형 분석, 데이터의 실시간 분석, 정형 데이터 처리, 다양한 분석 알고리즘, Workflow, 시각화 같은 기술들을 다양한 오픈소스 프로젝트의 형태로 개발되어서 바로 활용 가능한 수준으로 제공하는 하둡 에코시스템이 있음

[하둡 에코시스템 구성도]



- 하둡 기반 분석은 여러 가지 시스템에서 생성되는 로그나 데이터, 동영상, 텍스트 등 정형/비정형 데이터를 활용하여 배치 집계 분석, 분산·병렬 데이터 처리, 비정형/정형 데이터 연계 분석을 수행하도록 구성하며, 주요 기능은 다음과 같음
 - 대량의 데이터에 대한 분산 병렬 처리
 - 배열 및 행렬로 표현된 데이터에 대한 통계 분석
 - 소셜 데이터나 반정형 형태의 로그 등에서 원하는 데이터를 추출하기 위한 텍스트 분석
 - 숨겨져 있는 데이터간의 상호 관련성 및 유용한 정보 추출을 위한 기계학습
 - 비정형 데이터와 정형 데이터의 연계 분석

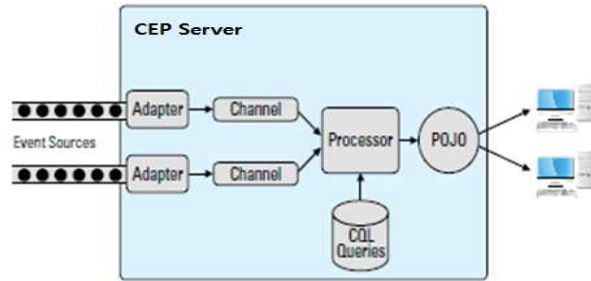
[하둡 기반 분석 개념도]



(나) 실시간(Real-time) 분석

- 실시간 분석을 지원하는 스트림 컴퓨팅(Stream Computing)은 기존의 축적된 과거 데이터에 대한 분석뿐만 아니라 실시간으로 흘러 들어오는 다양한 데이터를 분석하는 방법으로 대용량 데이터에 대한 이동 중(in-motion) 분석 환경을 제공함
- 스트림 컴퓨팅 기술 중 하나인 CEP(Complex Event Processing)는 여러 이벤트 소스로부터 발생한 이벤트를 대상으로 실시간으로 의미 있는 데이터를 추출하여 대응되는 액션을 수행하는 것으로 Adapter, Channel, Processor, Beans(POJO) 컴포넌트로 구성되며, 컴포넌트 기능은 다음과 같음
 - Adapter : 인바운드 및 아웃바운드 데이터 인터페이스로 이벤트 소스로부터 스트림 데이터를 수신, 가공, 내부 이벤트 타입으로 변환을 담당하고 생성한 이벤트를 연결된 컴포넌트로 전달함
 - Channel : 채널은 상하위 컴포넌트가 비동기적으로 동작할 수 있도록 하는 쓰레드 풀과 관련된 큐그림에서는 아답터에서 수신한 이벤트를 프로세스에 전달하는 역할
 - Processor : 프로세스는 CEP 엔진의 인스턴스로 CQL 쿼리를 담당하는 부분으로 CQL 쿼리는 필터링, 집합 함수(Aggregation), 패턴 매칭, 조인(스트림간, 스트림과 DB, 스트림과 캐쉬 등)등을 지원
 - Beans(POJO) : POJO는 프로세스에서 걸러진 의미 있는 데이터를 외부 서비스와 연동하기 위한 비즈니스 로직을 구현하기 위해 사용

[CEP 흐름도]



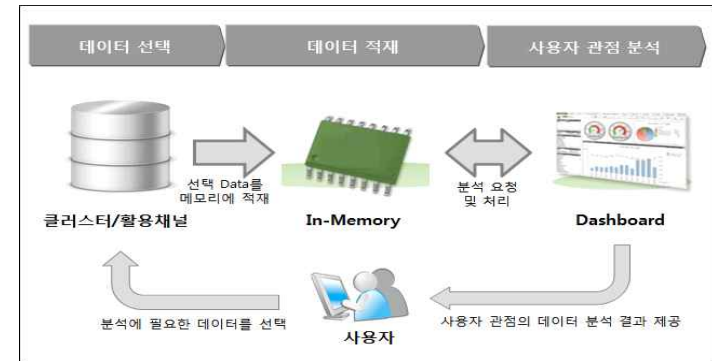
- CEP에서 의미 있는 이벤트를 걸러내는 역할은 Processor가 담당하며, Processor는 질의 언어로 CQL(Continuous Query Language)를 사용하여 센서 장치, 시스템 등에서 발생하는 지속적인 대량 스트림 데이터에서 의미 있는 데이터를 추출할 수 있는 다음과 같은 기능으로 구성됨

| 기능 | 설명 |
|----------------------|---|
| 이벤트 필터링 | 조건문을 통해 원치 않는 데이터를 걸러냄 |
| 신규 이벤트 감지(중복 데이터 방지) | 주로 데이터의 소스가 센서 등 기계장치이므로 일정 간격으로 발생한 데이터는 동일한 데이터 또는 값의 차이가 크지 않은 의미적으로 동일한 데이터가 연계 되므로 이러한 중복이 많은 데이터 중에서 의미 있는 새로운 이벤트 데이터만을 추출 |
| 이벤트 파티셔닝 | 이벤트 소스에서 오는 데이터가 주식정보와 같이 여러 거래 종목이 동일 소스에서 같이 섞여 입력 될 경우 파티셔닝을 통해 종목 단위로 분리, 질의 |
| 이벤트 집합 질의 | 타임 윈도우 등을 이용해 최근 1분, 최근 한시간과 같이 구간 단위로 이벤트 데이터에 대한 실시간 집합 연산을 수행 |
| 이벤트 메시지 강화 | FROM에 오는 대상은 스트림 이벤트 소스는 물론, DB 테이블 coherence in-memory 캐쉬 데이터가 될 수 있으며, 이런 이벤트 소스와 DB, 캐쉬 데이터 정보와 조합하여 이벤트 데이터를 좀더 의미있는 데이터로 변환 |
| 이벤트 연관관계 설정 | 이벤트 소스가 둘 이상인 경우, 각 소스로 서로 다른 유형의 정보가 오는 경우 둘을 조합하여 데이터를 생성 |
| 분실 이벤트 감지 | CQL 패턴 매칭을 이용하여 분실 이벤트 감지에도 사용되는데 특정 패턴이 일정시간 나타나지 않는 경우 이벤트가 분실된 것으로 판단 |
| GPS와 지리 정보 연동 | GPS 정보를 지속적으로 수신 받는다고 하면, GPS 정보는 지도 정보와 연동해 활용할 수 있으며, 지속적인 수신 GPS 정보를 지리 정보와 연계하여 특정 지역 안에 들어왔다고 하는 의미 있는 이벤트만 걸러냄 |

(다) 사용자 주도 분석

- 전통적인 BI 데이터 접근방식은 데이터가 변경될 때마다 분석을 위한 모든 프로세스가 다시 진행되어야 하나, 사용자 주도 데이터 접근 방식은 변경 대상 데이터에 대한 선택/제거만으로 실시간 연관 검색 및 분석 수행이 가능하며, 연계/수집/저장된 데이터에서 분석에 필요한 데이터를 선택하면 ETL 기능을 통해 메모리에 해당 데이터를 적재하고, 사용자의 관점에 따라 다양하게 분석을 수행할 수 있도록 기능을 지원함

[사용자 주도 분석 개념도]



○ 데이터 적재 방안

- 별도 RDBMS와 같은 별도의 포맷으로 데이터를 저장하지 않고 데이터를 빠르게 재혼합/재결합하여 변화에 빠르게 대응할 수 있도록 모든 데이터를 메모리에 적재하기 위해 다음과 같은 기능을 지원함
 - . 실시간으로 변화하는 데이터에 대해 정해진 주기로 변화된 데이터만 적재
 - . 서로 다른 다양한 데이터 구조를 하나의 모델로 생성
 - . 최소한의 메모리 사용을 위해 압축 데이터 형태로 메모리에 적재
 - . 별도의 RDBMS가 아닌 메모리에 적재되는 Associative Data Model 기반
 - . 이중 DB 데이터들을 개별 테이블의 데이터로 적재

○ Dashboard 구현 방안

- 사용자가 저작물을 통해 메모리에 적재된 데이터를 쉽고 다양하게 Dashboard로 구현할 수 있도록 일반적이고 표준화된 명령어/메뉴/툴바를

제공하고, 웹 브라우저를 통한 AJAX 환경 등을 제공함

- 일반적이고 표준화된 명령어/메뉴/툴바 제공
- 사용자에게 친숙한 툴 환경 제공
- 빠른 개발을 위한 단축키 명령어 제공
- 다양한 뷰(웹, 모바일 등) 환경에서 일관성 있는 UI 제공
- 웹 브라우저를 통한 AJAX 환경 제공
- 다양한 Rich web client 환경 제공

○ 사용자 주도 분석 주요 기능

- 사용자들이 중요 비즈니스 현안에 대해 추측하지 않고 정확하게 판단할 수 있도록 사용자 중심의 시각으로 데이터를 분석할 수 있는 다음과 같은 기능을 제공함

| 기능 | 설명 |
|-----------------|---|
| Drilling | · 분석에 필요한 모든 데이터를 메모리에 적재하고 데이터들 사이의 연관(Association)을 자동으로 관리하여, Drill Up/Down/Across/Through 분석을 제한 없이 유연하게 처리 |
| Pivoting | · Pivot Table은 디멘전과 메저 양방향의 Pivoting을 지원하므로 다양한 관점에서 데이터를 분석하도록 지원 · 다양한 Mini Chart, Sparkline 등의 차트를 Pivot Table 내에 삽입하여 분석을 확장 |
| Ad-Hoc 분석 | · 사용자가 새로운 관점(Dimension)으로 분석을 하거나 새로운 Measure를 추가하고자 하는 경우 언제나 측색에서 기존 분석을 확장할 수 있도록 지원 |
| Multi-Fact Join | · Link Table 또는 Concatenate를 통해 데이터 모델링 · OLAP 객체에서 Multi Fact Join |
| 비교분석 | · 선택한 값들과 레이아웃을 그대로 저장하고 있으므로 특정 상태에 대한 비교 · 특정 상태의 연산이 수행된 OLAP 객체를 Detach하면 그 객체는 다른 선택의 영향을 받지 않고 상태 유지 · 선택된 집합, 또는 전체 데이터 집합을 대상으로 또 다른 집합을 정의하여 비교 분석 · 그룹의 정의와 그룹 상태의 상속을 통해 그룹간 비교 |
| 그래픽 | · 사용자의 이해를 위한 최신의 그래픽을 제공 · Extension 기능으로 내장 객체를 확장하거나 외부 객체 사용 |
| 트렌드 분석 | · Time series trend를 기반으로 과거 추이를 분석하는 Backcasting과 미래 추이를 예측하는 Forecasting 지원 |
| What-if 분석 | · Fact를 기반으로 다양한 측면의 데이터를 반영하여 시뮬레이션을 지원 |

(4) 분석 플랫폼 영역별 활용 방안

빅데이터 분석 플랫폼은 크게 Hadoop 분석 영역, Real-time 분석 영역, 사용자 주도 분석 영역으로 구분되며, 분석이 필요한 업무 및 데이터에 맞게 구축되어야 함

| 분석 영역 | 역 할 | 분석 기법 | 적용 업무 | 분석 데이터 |
|--------------|---|--|--|--|
| Hadoop 기반 분석 | · 연계/수집된 정형-비정형 데이터에 대해 그룹핑, 정제 등 전처리 과정을 통해 배치(Batch)로 데이터를 분석 | · 통계분석 · 텍스트분석 · 기계학습 (Mining) 분석 · 배치분석 등 | · 연계/수집/저장된 정형-비정형 대용량 데이터를 정해진 주기에 따라 분석이 필요한 업무 | · 분석하기 위해 저장된 대용량 데이터 · 요약 데이터가 아닌 대규모 원본 데이터를 대상으로 분석 |
| Real-time 분석 | · 스트리밍(Streaming) 컴퓨팅 기술 적용을 통해 빅데이터를 수집과 동시에 분석을 수행 | · 필터링, 감지 · 파티셔닝 · 집합 (Aggregation) · 패턴매칭 · 지리정보 연동 등 | · 데이터를 보관하지 않은 상태에서 분석이 필요한 업무 · 이동 중인 데이터에서 가치 있는 Insight를 찾기 위한 업무 | · 대량으로 지속적으로 입력되는 데이터, 시간 순서가 중요한 데이터 · 전통적인 관계형 데이터 베이스에서 실시간 처리 및 분석이 불가능한 스트림 데이터 분석 |
| 사용자 주도 분석 | · 사용자들이 자신들만의 시각으로 데이터를 분석 할 수 있도록 분석 대상 데이터에 대한 선택/제거 만으로 실시간 연관 검색 및 분석을 수행 | · Pivoting · Ad-Hoc 분석 · 비교 분석 · 그래픽 분석 · 트렌드 분석 · What-if 분석 등 | · 사용자가 원하는 데이터를 선택해서 자신만의 시각으로 분석이 필요한 업무 · Fact를 기반으로 다양한 측면의 데이터를 반영하여 최적의 결과를 찾기 위한 업무 | · 시스템이 아닌 사용자 결정에 따라 선택되어지는 데이터 · 제한 없이 원하는 방식으로 데이터를 혼합 또는 결합하여 분석 |

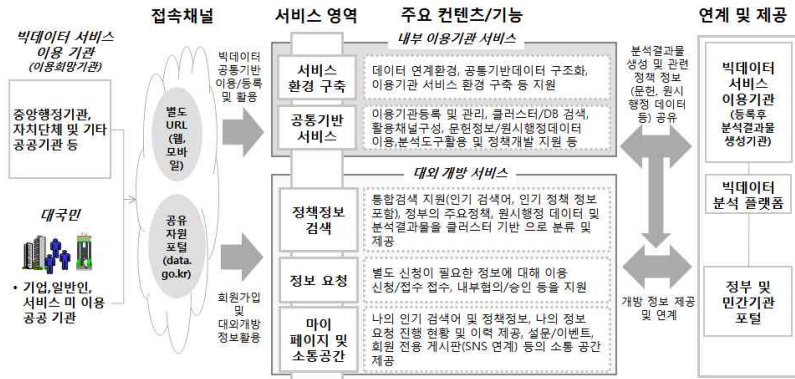
마) 빅데이터 공통기반 포털 구성 방안

빅데이터 공통기반 서비스 활용기관에게는 공통기반 활용 편의를 극대화하고 활용기관이 공통기반 활용과정에서 생성한 정보(문헌 자료, 원천데이터, 분석결과물 등)를 공통기반 서비스에 등록하지 않은 공공기관(이하 '미 이용 공공 기관')과 민간 등 대국민에게 개방하기 위한 포털 구성 방안을 제시함

(1) 목표 서비스

빅데이터 공통기반 포털은 서비스 활용기관을 대상으로 하는 내부활용기관서비스와 대국민(서비스 미 이용 공공기관, 기업, 일반인 등)에게 제공하는 대외개방서비스로 구분하여 구성함

[빅데이터 공통기반 포털 서비스 개념도]



(2) 서비스 환경 구축

데이터 보유기관의 정책관련자료(문헌정보, 원시행정데이터 등)를 클러스터 분류 체계 기반으로 구조화하고 데이터 연계, 채널 연계, 서비스 요청 심의 및 분석도구 구축 등의 이용환경을 서비스화 함으로써 활용기관의 시스템 환경에 적합한 서비스 이용환경을 사전에 구축할 수 있도록 지원하며, 주요 기능은 다음과 같음

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|-----------|---------------|-----------------------|
| 서비스 환경 구축 | 보유기관 데이터 연계환경 | 원시대용량데이터 원천발굴 기능 |
| | | 보유기관 연계 및 데이터 확보 기능 |
| | | 보유 데이터 품질 및 보안관리 기능 |
| | 공통기반 데이터 구조화 | 정책 클러스터 구성 관리 기능 |
| | | 정책 클러스터 유지관리 기능 |
| | 활용기관 데이터 연계환경 | DB 및 활용채널 연계관리 |
| | | 연계 데이터 관리 |
| | 서비스 이용 및 운영환경 | 공통기반 서비스요청 심의관리 기능 |
| | | 분석시스템 및 분석도구 구축 관리 기능 |
| | | 공통기반 운영실적 분석 기능 |

(3) 공통기반 서비스

데이터 보유기관의 정책관련자료(문헌정보, 원시행정데이터 등)를 활용하여 심도 있는 분석업무를 통한 정책개발지원이 될 수 있도록 문헌 정보, 원시행정데이터 DB 활용채널 검색/조회, 채널 구성 및 관리, 데이터 연계 및 분석개발 시스템 요청 등의 서비스를 제공하며, 주요 기능은 다음과 같음

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|----------|------------------|---------------------------|
| 공통기반 서비스 | 활용기관 공통 서비스 | 활용기관 관리 기능 |
| | | DB 및 데이터 검색기능 |
| | | 활용채널 구성 관리 기능 |
| | 문헌정보 기반 정책정보 조회 | 정책정보 조회 기능 |
| | | 문헌정보 활용채널 구성 관리 기능 |
| | 원시행정 데이터 연계 및 요청 | 원시행정데이터 DB 연계 요청/관리 기능 |
| | | 활용채널 연계 요청/관리 기능 |
| | 정책 개발 지원 | 활용채널 데이터 이용 요청/관리 기능 |
| | | 활용채널 데이터 이용 지원 기능 |
| | | 활용채널 데이터 분석시스템 개발 요청관리 기능 |

(4) 정책정보 서비스

서비스 활용기관이 공통기반 활용과정에서 생성한 행정정보, 문헌자료, 원시 행정 데이터와 분석 결과물을 한 곳으로 통합하여 제공함으로써 대국민의 업무(연구, 기획 등)와 생활에 필요한, 다양하고 고품질의 정보를 쉽게 취득할 수 있도록 다음과 같은 기능을 제공함

| 메인 메뉴 | 하위 메뉴 | 세부 기능 | |
|-------|-------------|---|--|
| 정책정보 | 행정정보 및 문헌자료 | 정부부처별 정보 프로파일(정보제목, 요약, 부처, 담당자 명, 신청기반정보 여부 등) 조회 (공통) | |
| | | 정부 부처별 원문(또는 본문) 조회, 출력 및 다운로드 (공통) | |
| | | 클러스터 분류체계 별 정보 조회 기능(공통) | |
| | | 원시행정 데이터 | 정부부처별 정보 프로파일(정보제목, 요약, 부처, 담당자 명 신청기반정보 여부 등) 조회 (공통) |
| | | | 정부 부처별 원문(또는 본문) 조회, 출력 및 다운로드 (공통) |
| | | | 클러스터 분류체계 별 정보 조회 기능(공통) |
| | | 통계 데이터 뷰어 및 변환 지원 | |

| | | |
|-------|--------------------|--|
| | 분석결과 | 정부부처별 정보 프로파일(정보제목, 요약, 부처, 담당자명, 신청기반정보 여부 등) 조회 (공통) |
| | | 정부 부처별 원문(또는 본문) 조회, 출력 및 다운로드 (공통) |
| | | 클러스터 분류체계 별 정보 조회 기능(공통) |
| | | 해당 분석결과물에 활용된 관련 정책정보 조회 |
| 정보제공자 | 정보 제공기관 지원 서비스 | |
| | 타 기관 포털과의 연계 인터페이스 | |

(5) 검색 서비스

통합검색서비스와 함께 디렉토리 및 인기 검색어 서비스를 제공함으로써 업무, 연구, 생활에 필요한 정책정보의 소재 파악 및 세부 주제별 접근과 하위부서/담당자별 정보소재 파악을 지원하기 위하여 디렉토리는 정부부처/기능과 클러스터 분류를 모두 지원할 수 있도록 구축하고, 검색결과물의 최종 모습에는 정보 프로파일을 포함하도록 다음과 같은 기능을 제공함

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|-------|---------------------|--------------------|
| 검색 | 정부부처/기능/클러스터 별 디렉토리 | 통합 디렉토리 서비스 |
| | | 개별 디렉토리 서비스 |
| | 인기 검색어 | 검색어 순위 제공 서비스 |
| | | 검색어별 관련 정보 조회 서비스 |
| | 인기 정책 | 인기 정책정보 순위 제공 |
| | | 인기 순위별 정책정보 조회 서비스 |
| 통합검색 | 통합검색 서비스 | |

(6) 정보요청 서비스

분석결과물 중 정보 제공시 내부 협의가 필요한 자료를 신청기반 정보목록으로 사전에 공개하도록 함으로써 부처별 보유 정보를 가능한 범위 내에서 최대한 공개할 수 있도록 유도하고, 통합정보공개시스템과 연계하여 법령에 따른 정보공개 절차에 대한 인지도도 높일수 있도록 다음과 같은 기능을 제공함

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|----------------------|-----------|--|
| 정보요청 | 정보공개절차 연계 | 정보공개 절차 지원 |
| | | 정보공개 관련 법령 및 절차 요약 정보 제공 |
| | 정보제공자 | 정보이용신청 접수 기능 |
| | | 정보이용 승인 기능 |
| | | 정보이용신청 접수/내부협의/승인 진행현황 관리 기능 |
| | | 정보이용신청 접수/내부협의/승인 결과 제공 기능 |
| | | 승인정보 업로드 기능 |
| | 제공정보현황 | 신청 기반 정보목록 조회 기능 |
| | | - 부처별/디렉토리(분류)별 정보 조회 - 검색 및 관련 메뉴/서비스 연계 |
| | 정보이용신청 | 정보이용신청 기능 |
| - 검색 및 관련 메뉴/서비스와 연계 | | |
| 정보이용신청 내역 확인 기능 | | |
| - 마이페이지/스마트폰/이메일로 확인 | | |
| 정보이용신청 진행현황 모니터링 기능 | | |
| - 신청접수/검토/승인결과 확인 | | |
| - 마이페이지/스마트폰/이메일로 확인 | | |
| - 승인정보 다운로드 기능 | | |

(7) 마이페이지 서비스

마이페이지 서비스를 제공함으로써 항상 필요정보에 대한 이용편의를 극대화하고 개별 이용자 요구사항에 대해 신속하고 지속적으로 대응할 수 있는 개인화 서비스 기반을 형성하게 되며, 마이페이지가 지속적인 개인화 서비스 기반 역할을 할 수 있도록 개인화 공간을 제공함과 동시에 내가 자주 찾는 정책정보, 정보이용이력 등과 같이 개인의 이용 특성을 자연스럽게 파악할 수 있는 기능을 다음과 같이 제공함

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|-------|-----------|----------------------|
| 마이페이지 | 나의 인기 검색어 | 나의 검색어 순위 제공 서비스 |
| | | 나의 검색어별 관련 정보 조회 서비스 |
| | 나의 관심 정책 | 관심 정책 등록 |
| | | 내가 자주 찾는 정책정보 |
| | | 나의 관심 정책 및 정보 조회 |
| | 정보이용현황 | 나의 정보이용 신청 내역 |

| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|-------|-------|--|
| | | 나의 정보이용신청 진행현황 모니터링 - 신청/접수, 내부협의/승인결과 확인 - 승인정보 다운로드 기능 |
| | | 정보이용이력 - 이용신청 및 다운로드 이력 - 이벤트/설문 참여 이력 - 게시판 작성 이력 |

(8) 소통공간 서비스

소통공간 서비스를 통해 이용자의 다양한 의견을 수렴할 수 있는 공간을 제공하며, 실질적 의견 수렴 효과를 위하여 회원전용 게시판을 커뮤니티로 활용할 수 있도록 다음과 같은 기능을 제공함

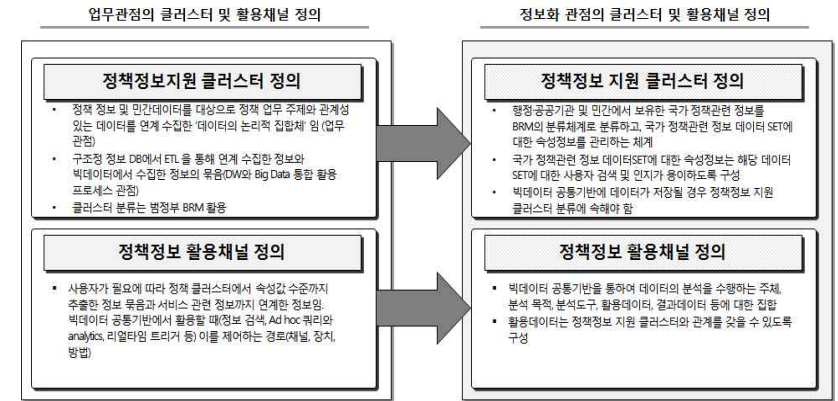
| 메인 메뉴 | 하위 메뉴 | 세부 기능 |
|---------------|------------|--|
| 소통공간 | 회원가입 | 회원정보 등록 및 확인 |
| | | 관심 분야 등록 |
| | 설문/이벤트 | 설문/이벤트 등록 및 기간 설정 |
| | | 설문/이벤트 문항 생성, 변경 및 삭제 관리 |
| | | 설문/이벤트 통계 지원 |
| | 이용안내 및 FAQ | 회원 전용 게시판 |
| | | SNS 게시판 |
| | | 자유 게시판 |
| | | 게시글 등록,수정, 삭제 기능 제공 |
| | | 포털 이용안내 일반이용 안내 - 정보이용신청 이용안내 - 회원전용 게시판 이용안내 - 회원가입 탈퇴 안내 등 |
| FAQ(자주 묻는 질문) | | |

바) 클러스터 및 활용채널 관리시스템 구축 방안

(1) 클러스터 및 활용채널 정의

업무관점에서 정의한 클러스터 및 활용채널의 정의를 정보화 아키텍처 관점에서 아래와 같이 재정의 함

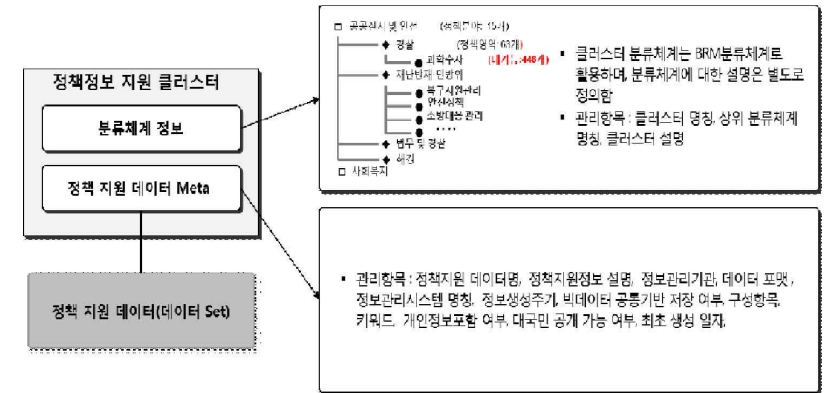
[정보화 관점의 클러스터 및 활용채널 정의]



(2) 클러스터 구성요소 및 관리항목 도출

클러스터 구성요소는 크게 분류체계정보와 정책지원정보Meta로 구분될 수 있으며, 해당 구성요소에 대한 관리항목을 아래와 같이 도출함

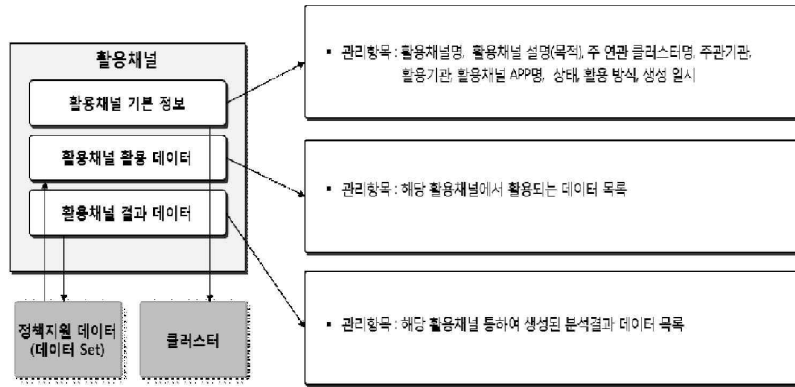
[클러스터 구성요소 및 관리항목]



(3) 활용채널 구성요소 및 관리항목 도출

빅데이터 공통기반 활용채널의 구성 요소는 활용채널 기본 정보, 활용채널 활용 데이터, 활용채널 결과 데이터로 구성되며, 각각의 관리항목을 아래와 같이 도출함

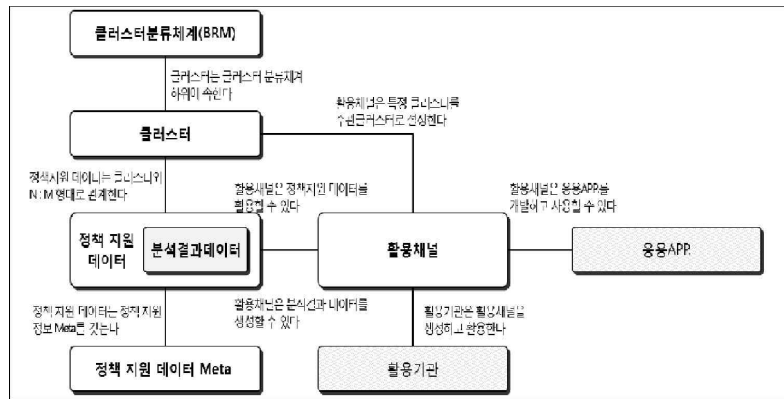
[활용채널 구성요소 및 관리항목]



(4) 주요 개체간 관계 정의

클러스터 및 활용채널과 관련된 주요개체간의 관계를 아래와 같이 정의함

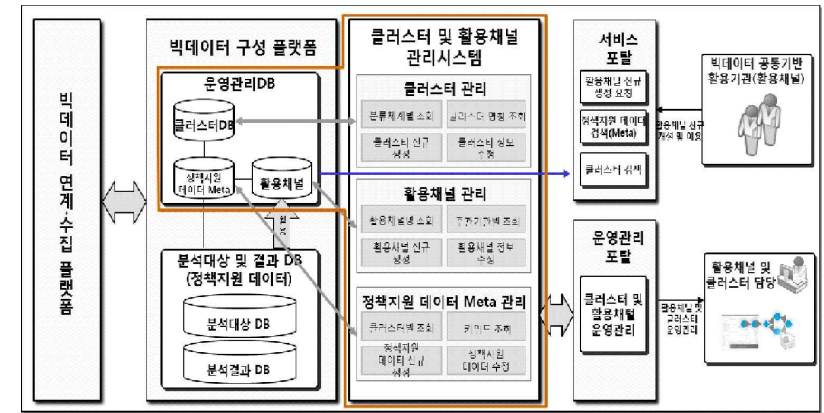
[클러스터 및 활용채널 관련 주요 개체간 관계 정의]



(5) 클러스터 및 활용채널 관리시스템 목표 아키텍처

클러스터 및 활용채널 관리시스템의 주요 구성 요소는 클러스터DB, 정책지원 데이터 Meta, 활용채널 DB, 클러스터 활용채널 관리 APP, 임

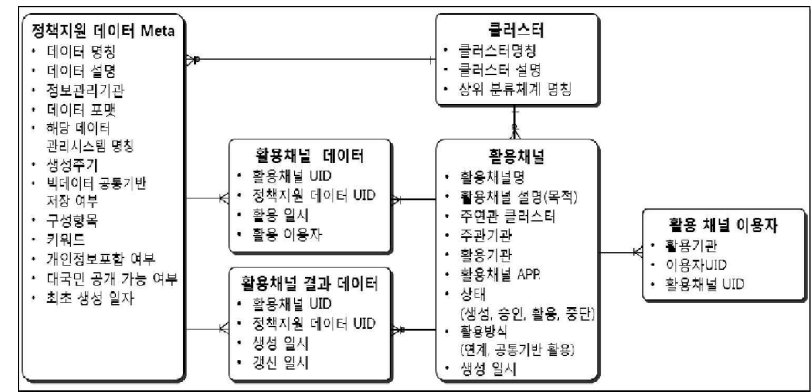
[클러스터 및 활용채널 관리시스템 목표 아키텍처]



(6) 논리데이터 모델

클러스터 및 활용채널 관리시스템의 논리데이터 엔터티는 정책지원 데이터 Meta, 활용채널 데이터, 활용채널 결과 데이터, 활용 채널 이용자로 구성됨

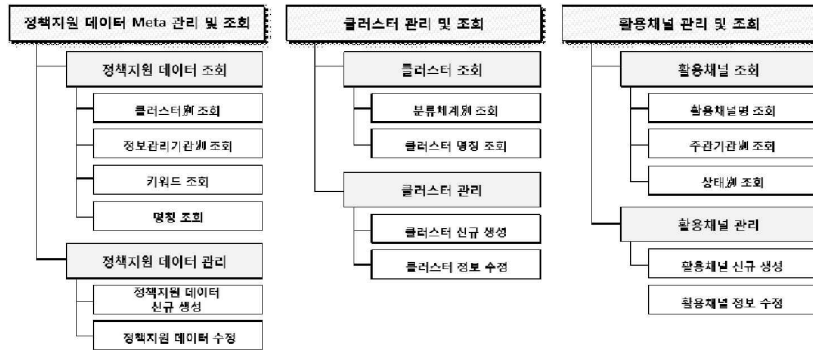
[클러스터 및 활용채널 관리시스템 논리데이터 모델]



(7) 응용기능 구성

클러스터 및 활용채널 관리시스템의 응용기능은 크게 정책지원 데이터 Meta 관리 조회, 클러스터 관리 및 조회, 활용채널 관리 및 조회로 구분됨

[클러스터 및 활용채널 응용기능 구성도]

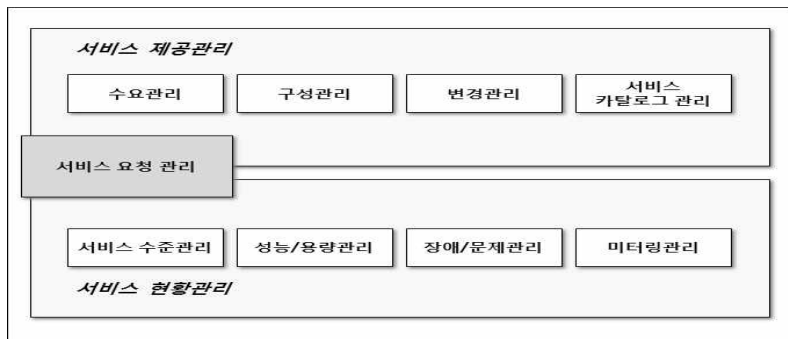


사) 빅데이터 공통기반 자원관리 체계

(1) 자원관리 범위 및 구성

클라우드 자원관리는 크게 서비스 제공관리와 현황관리의 영역으로 구분되며 제공관리는 기존 ITIL 프로세스를 반영하여 클라우드 컴퓨팅 서비스 요청관리, 서비스 카탈로그 관리, 수요관리, 변경관리, 구성관리 등으로 구성됨. 서비스 현황 관리는 서비스수준관리, 성능관리, 용량관리, 장애관리 및 미터링 관리체계로 구성 됨

[자원관리 범위 및 구성]

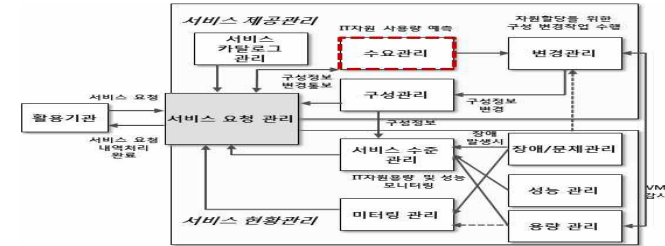


(2) 서비스 제공 및 현황관리 흐름

클라우드 서비스의 제공은 사용자의 서비스 요청관리로부터 시작되며 소요 컴퓨팅자원 예측을 위한 수요관리, 이에 기반하여 자원할당 및 변경을 위한 변경

관리, 변경작업과 구성정보간 일치성 검토를 위한 구성관리, 서비스 요청처리 및 서비스 제공을 위한 서비스 요청관리 및 서비스 수준 및 자원할당관리 등에 대한 현황관리 등이 상호 유기적으로 연계됨

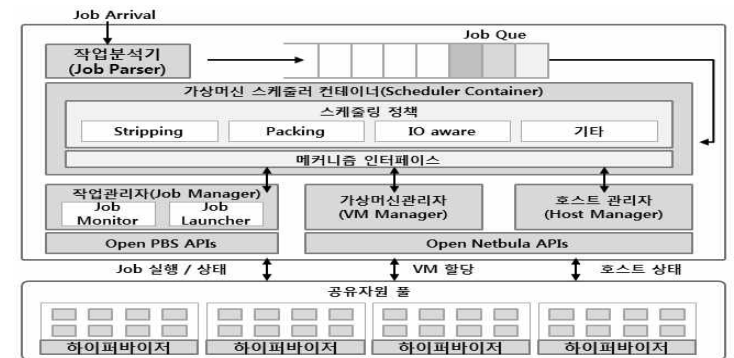
[서비스 제공 및 현황관리 흐름도]



(3) 가상 클러스터 프로비저닝 및 작업관리 프레임워크

클라우드 기반 빅데이터 공통기반의 가상 클러스터 프로비저닝 및 관리 프레임 워크 구성요소는 가상화 환경 물리서버를 관리하는 호스트관리자, 가상머신 풀을 관리하는 가상머신 관리자, 사용자가 제출한 작업 자원요청 사양에 대한 제출 명세서를 작성하는 작업분석기와 다양한 자원할당 기법을 사용하는 스케줄러 컨테이너로 구성됨

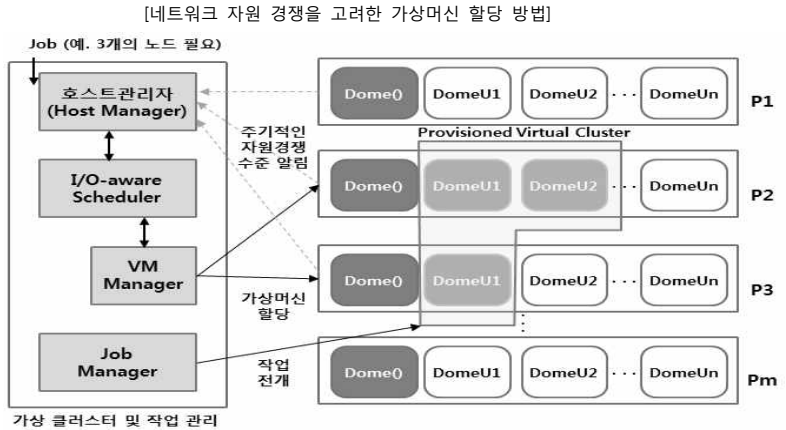
[빅데이터 공통기반 가상 클러스터 프로비저닝 및 관리 프레임워크]



(4) 네트워크 자원경쟁을 고려한 가상머신 할당

클라우드 가상환경에서 호스트관리자는 제어 도메인의 응용 프로세스의 CPU 자원 사용율과 일반 도메인 영역의 자원 여유율을 가지고 각 물리서버 별로 네트워크 경쟁수준을 관리하며 작업 스케줄러는 가상머신 할당을 위해 여유자원

을 가진 물리서버 중 네트워크 경쟁이 적은 노드들을 우선적으로 선택하여 가상 머신 구동을 요청하도록 구성함



(5) 자원관리체계 기능 구성

빅데이터 공통기반 자원관리체계의 필요 기능으로는 핵심기능과 운영기능으로 구분되며 핵심기능으로는 포털, 서비스 카탈로그, 가상화 관리, 프로비저닝, 워크플로우가 있고, 운영기능으로는 로드밸런싱, 미터링/빌링, API 관리, 보안, 백업/복구, 모니터링, SLA, 리포팅 기능 등이 있음

[자원관리체계 기능 구성(안)]

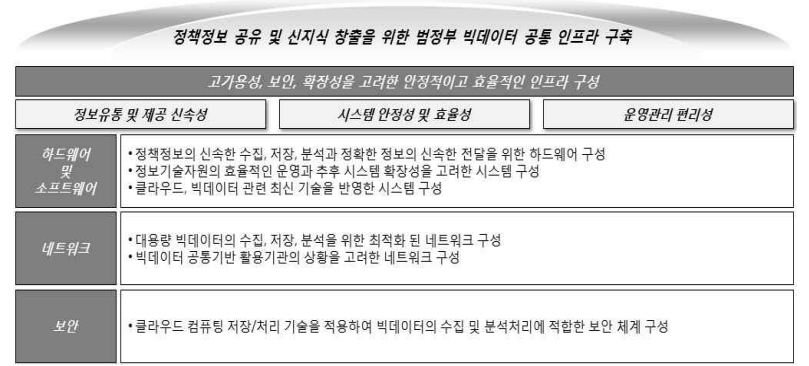
| | | | | | | | | | | |
|------|--|------------------|---------------------------------|---|----------------------------|------------------------------|----------------------------|-------------------------------|--|--|
| 정의 | • 빅데이터 공통기반 자원관리체계는 클라우드 기반 서비스를 제공하기 위한 서비스 운영도구로서 사용자 클라우드 서비스를 이용하기 위한 각종 편의 기능과 관리자가 클라우드 서비스를 운영하기 위한 각종 제어와 관리 기능을 제공하는 체계 | | | | | | | | | |
| 핵심기능 | 포털 | 서비스 카탈로그 | 가상화 관리 | 프로비저닝 | 워크플로우 | | | | | |
| | • 사용자와의 커뮤니케이션 • 각종 현황관리 | • 이용 가능한 서비스 목록 | • 가상화 자원 생성 및 회수 • 하이퍼바이저 제어 | • 가상머신의 사용자 준비 • OS, 소프트웨어 등 이미지 자동 설치 | • 사용자의 서비스 이용신청, 검토, 승인 절차 | | | | | |
| 운영기능 | 로드밸런싱 | 미터링/빌링 | API 관리 | 보안 | 백업/복구 | 모니터링 | SLA | 리포팅 | | |
| | • SW적 부하분산 • 오토 스케일링 | • 서비스 이용량 측정과 과금 | • 유관시스템과 표준연동 | • 접근 권한 관리 • 보안 | • 백업과 복구 • 복구 이미지 관리 | • 장애와 성능관리 • 가상머신, 하이퍼바이저 | • 서비스 수준관리 • 서비스 수준 리포트 | • 각종 보고서 자동생성 • 각종보고서 자동생성 | | |

아) 빅데이터 공통기반 인프라 구축 방안

(1) 인프라 구성 전략

빅데이터 공통기반 인프라 구성에 있어 정보유통 및 제공의 신속성, 시스템 안전성 및 효율성, 운영관리 편의성 등이 반영되도록 함

[인프라 구성전략]

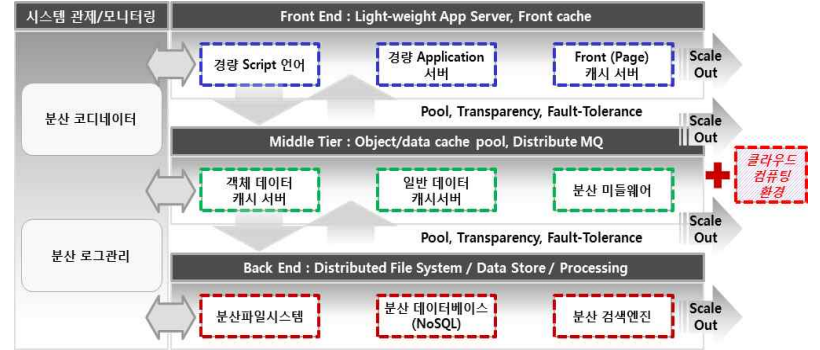


(2) 인프라 구성방안

(가) SW 아키텍처 구성

○ 일반적인 빅데이터 기반 플랫폼의 아키텍처와 Bid Data 플랫폼 요구사항들을 종합한 결과, Front End, Middle Tier, Back End 간의 투명성과 결합방지를 제공하고 수평적 확장 가능하며 비용절감 및 자원관리 효율성을 위한 클라우드 컴퓨팅 환경을 고려한 아키텍처를 구성하는 것이 타당하다고 판단됨

[빅데이터 플랫폼 SW 아키텍처 구성방향]



- 빅데이터 수집, 원본 데이터 저장, 트랜잭션 데이터 저장, 실시간 분석 플랫폼, 배치 분석 플랫폼, In-Memory 분석 플랫폼, 데이터마이닝/통계도구, 클러스터 관리 및 모니터링, 데이터 Serialization 부문별로 관련 기능이 필요함

[빅데이터 공통기반 아키텍처 필요 기능]

| 구분 | 주요 현황 및 이슈 | 오픈소스 |
|---------------------------------|---|--|
| 데이터 수집 | <ul style="list-style-type: none"> • 데이터 발생원으로부터 안정적인 저장소로 저장하는 기능 수행 | <ul style="list-style-type: none"> • Flume, Scribe, Chukwa |
| 원본데이터 저장 | <ul style="list-style-type: none"> • 수집된 데이터를 안정적으로 저장하는 저장소 • 비구조적 데이터 저장소로 주로 대용량 파일저장소 | <ul style="list-style-type: none"> • Hadoop File System • MoigileFS |
| 트랜잭션 데이터 저장 | <ul style="list-style-type: none"> • 원본 데이터를 실시간으로 저장, 조회처리를 하기 위한 저장소 • 구조적 저장소 또는 검색엔진 기술을 활용 | <ul style="list-style-type: none"> • NoSQL (Cloudera, Hbase, Cassandra) • Katta, ElasticSearch |
| 실시간 분석 플랫폼 | <ul style="list-style-type: none"> • 데이터 수집과 동시에 분석을 수행 • 복잡한 분석보다 count, sum 등 단순한 aggregation 연산 정도 수행 | <ul style="list-style-type: none"> • S4, Stom |
| 배치 분석 플랫폼 | <ul style="list-style-type: none"> • 전체 또는 부분 데이터에 대해 복잡하고 다양한 분석 수행 • 대용량 데이터 처리를 위한 분산, 병렬처리가 필요 • 단순 텍스트 분석부터 그래프 분석까지 다양한 분석 모델 지원 | <ul style="list-style-type: none"> • Hadoop MapReduce (Hive, Pig) • Giraph, GoldenOrb |
| In-Memory 분석 플랫폼(=사용자 주도분석 플랫폼) | <ul style="list-style-type: none"> • 메모리에 로딩된 데이터를 재사용하여 사용자 주도 분석이 가능하도록 함 • 여러 시스템간의 캐쉬 데이터 공유하며 개별 분석요청에 대한 유연한 처리가 가능하도록 구성함 | <ul style="list-style-type: none"> • 없음 • IBM Netezza, SAP HANA 등 |
| 데이터마이닝/통계도구 | <ul style="list-style-type: none"> • Cluster, Classification 등과 같이 데이터 마이닝을 위한 기본 알고리즘 라이브러리 및 도구 | <ul style="list-style-type: none"> • Mahout, R |
| 클러스터 관리 및 모니터링 | <ul style="list-style-type: none"> • 대부분 분산 시스템으로 구성되기 때문에 전체 클러스터에 대한 관제 및 모니터링 복잡 | <ul style="list-style-type: none"> • Zookeeper, HUE, Cloumon |
| 데이터 Serialization | <ul style="list-style-type: none"> • 이기종 플랫폼 및 다양한 종류의 솔루션을 사용하므로 데이터 전송 및 처리에 대한 표준 프레임워크 필요 | <ul style="list-style-type: none"> • Thrift, Avro, ProtoBuf |

- 빅데이터 공통기반 도입대상 주요 소프트웨어로는 R(Analytics), Mahout, Oozie, Sqoop 등이 있으며 구성요건은 다음과 같음

[빅데이터 공통기반 도입 대상 SW 구성요건]

| SW명 | 용도 | 구성요건 |
|----------------|---------------|---|
| R (Analytics) | 빅데이터 통계분석 | <ul style="list-style-type: none"> • Data analysis software • A programming language • 데이터와 관련된 입출력, 핸들링, 관리, 분석, 그래픽 등 최신의 알고리즘 및 라이브러리 제공 • 분석에 통찰을 부여할 수 있는 그래픽에 대한 강력한 지원 • Chart, Plot, MotionChart, Map 연계 등을 R에서 바로 사용 가능 |
| Mahout | 빅데이터 통계분석 | <ul style="list-style-type: none"> • 분석 작업실행 라이브러리 |
| Oozie (Flow) | Workflow 스케줄링 | <ul style="list-style-type: none"> • 하둡의 Workflow 스케줄링 : XML 기반, DAG • 코디네이터 지원 : 스케줄링/모니터링 • HTTP 인터페이스 + Command Line 인터페이스 + Web 콘솔 • 다양한 액션노드(어플)의 지원 및 제어 • ※ 도입되는 시스템과 호환성 유지 |
| Sqoop | ETL | <ul style="list-style-type: none"> • DBMS의 자료를 HDFS로 Import/Export • Command line 인터페이스 • JDBC 지원하는 모든 DBMS • RDBMS(MySQL, Oracle) + DW + NoSQL Datastore • MapReduce 프로그램 코드의 생성 • 하둡기반 시스템과의 통합 - Hive, HBase, Oozie • RDBMS와의 고성능 커넥터 지원 • Cloudera에서 개발 |
| Hive | 빅데이터 집계/분석/처리 | <ul style="list-style-type: none"> • Select, Join, Group by, Limit 등 SQL과 유사한 문법 지원 • 대용량 데이터의 분석 지원 • 테이블, 파티션 컬럼, 버킷(샘플링) 지원 • Shell : 인터랙티브 쿼리 지원 - Web & JDBC 클라이언트 제공 • ※ 도입되는 시스템과 호환성 유지 |
| Pig | 빅데이터 집계/분석/처리 | <ul style="list-style-type: none"> • 데이터-흐름 기반의 스크립트 프로그래밍 언어 • 집합, 연관배열, Tuple등의 데이터타입을 포함 • 데이터를 처리하는 하イレ벨 언어 자바 프로그램으로 복잡한 태스크를 쉽게 처리 • ※ 도입되는 시스템과 호환성 유지 |
| Map Reduce | 대량데이터 처리 | <ul style="list-style-type: none"> • 대용량 데이터를 컴퓨터 클러스터 환경에서 병렬처리 지원 • 동적 스키마 생성 • Map, Reduce Function 및 Input, 파티션, 비교, 결과 쓰기 기능 제공 |
| Elastic Search | 분산검색 엔진 | <ul style="list-style-type: none"> • 분산 검색 엔진 : 하나의 shard 가 깨졌을 때에, 다른 replica 되어 있는 곳에서 자동적으로 이동하여 보여줌으로써 SPOF 제거 |

| SW명 | 용도 | 구성요건 |
|------------|----------------|--|
| | | <ul style="list-style-type: none"> 다중 타입을 가지는 다중 Tenant(하나 이상의 인덱스를 지원, 인덱스마다 하나 이상의 타입을 지원, index level 별로 컨트롤 가능) index 구성을 다양한곳에 가능. 메모리, (분산)파일 시스템, NoSQL 등 문서 지향(미리 스키마 정의를 할 필요 없음, 색인 작업의 사용자 정의를 위해 스키마는 타입마다 정의할 수 있음) 장기간 지속을 위해 신뢰, 비동기 쓰기 가능 실시간 검색 가능 다양한 API/클라이언트 모듈 지원 |
| HBase | 분산 데이터 베이스 | <ul style="list-style-type: none"> 컬럼 구조의 저장소 HDFS와 인터페이스 제공 대용량 데이터를 안정적으로 처리 비정형 데이터의 구조적 데이터 처리 지원 (key,value) 록업 ※ 도입되는 시스템과 호환성 유지 |
| Hadoop | 빅데이터 저장용 파일시스템 | <ul style="list-style-type: none"> 리눅스의 FUSE 파일시스템에 HDFS를 Mount 대용량 데이터 저장소 쉽고 빠른 시스템 확장 지원 |
| Zookeeper | 빅데이터 관리시스템 | <ul style="list-style-type: none"> 분산 클러스터 머신과 자원의 상태관리 리더 선출 분산 Lock 상호배제 : Deadlock 방지 이벤트 처리 |
| Flume | 외부데이터 수집 | <ul style="list-style-type: none"> 다양한 다수의 서버로 부터 데이터를 안정적으로 수집해서 HDFS에 저장하는 프레임워크 로그등 대용량 실시간 데이터 수집 에이전트와 콜렉터 다중 구성 데이터를 수집해서 HDFS에 저장 및 통합관리 지원 모든 노드와 마스터가 수평적으로 확장 가능 Fault-tolerance(내고장성), 고성능 지원 Unix 기본지원, 모든 종류의 데이터, 시스템 호환 지원 동적 환경 재구성이 가능한 통합 관리 지원 Apache v2.0 라이선스, 오픈소스데이터 |
| Crawling | 민간데이터 수집 | <ul style="list-style-type: none"> SNS, Web/Blog 등의 민간 데이터 소스로부터 필요한 데이터를 안정적으로 수집 |
| 통합인증 Agent | 내부 사용자 인증 처리 | <ul style="list-style-type: none"> 전자정부 통합인증 Gateway를 활용한 인증 처리 |

○ 빅데이터 연계수집서버 관련 개발 대상 기능은 정형/비정형데이터 연계유형 등록, 비정형 데이터 연계 모듈, SNS 데이터 연계용 오픈 API 및 연계 공통관리 기능임

[빅데이터 공통기반 상세 개발 대상 기능]

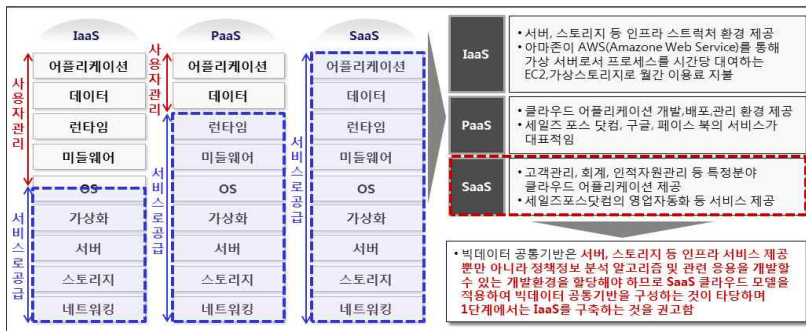
| 대기능 | 기능목록 | | 구성요건 | 비고 |
|--------------|---------------|--|--|--|
| | 중기능 | 소기능 | | |
| 연계수집 서버 | 정형 데이터 연계 | 정형 데이터 연계 유형 등록 | <ul style="list-style-type: none"> 기관별 정형 데이터 연계 처리를 위한 표준연동 방식 준용(DB Access, FTP, Sqoop, Flume API 등) | <ul style="list-style-type: none"> 보유기관 연계 |
| | 비정형 데이터 연계 | 비정형데이터 연계 대상 등록 | <ul style="list-style-type: none"> 기술유형에 따른 연계 대상 등록 | <ul style="list-style-type: none"> 보유기관 연계 |
| | | 비정형데이터 연계 모듈 | <ul style="list-style-type: none"> 비정형 데이터 연계 처리를 위한 모듈 관리(Flume Agent, Flume Collector 모듈 및 HDFS 인터페이스 API 구현) | <ul style="list-style-type: none"> 보유기관 연계 |
| | 외부 데이터 연계 | 오픈 API 처리 기준 | <ul style="list-style-type: none"> SNS 및 블로그에 대한 crawling 기능 구현 | <ul style="list-style-type: none"> SNS, 인터넷 |
| 공통관리 | 기준 정보 | 기준 정보 | <ul style="list-style-type: none"> 연계 대상 시스템 정보, 업무 로드 등 기준정보 관리 | |
| | | 송수신로그 관리 | <ul style="list-style-type: none"> 송수신 로그 및 상태정보에 대한 통합 관리 | |
| | 현황 통계 | <ul style="list-style-type: none"> 송수신 이력, 사용기관별 통계, 이용통계와 같은 현황정보 조회 | | |
| 빅데이터 클러스터 | 데이터 입수 | 데이터입수 및 품질 관리 | <ul style="list-style-type: none"> 비정형데이터에 대한 입수 및 데이터 품질확보를 위한 데이터 전체가 아닌 Garbage Filtering을 통한 데이터 수집 | <ul style="list-style-type: none"> 연계/수집 서버 |
| | 빅데이터 처리기반 | 빅데이터 파일 시스템 구성 | <ul style="list-style-type: none"> 오픈소스 기반 빅데이터 에코시스템 구성 | |
| | | 빅데이터 분산DB구성 | <ul style="list-style-type: none"> 오픈소스 기반 빅데이터 칼럼 DB 구성 | |
| 빅데이터 처리실행 | 빅데이터 집계/질의/분석 | <ul style="list-style-type: none"> 빅데이터 처리 실행을 위한 집계/질의/분석 기능 | | |
| 빅데이터 저장 | 데이터 입수/품질관리 | 데이터 입수 처리 | <ul style="list-style-type: none"> 데이터 입수 및 품질관리 | <ul style="list-style-type: none"> 빅데이터 클러스터 |
| | 데이터 저장 | 데이터 저장 및 분류 | <ul style="list-style-type: none"> 주제영역별 클러스터 데이터 체계 구성 | |
| 빅데이터 분석 클러스터 | 인증 및 권한 | 통합 인증 처리 | <ul style="list-style-type: none"> 통합인증 Gateway를 활용한 인증처리 또는 오픈 기반 인증 기능 구현 | <ul style="list-style-type: none"> 통합인증 Gateway |
| | | 권한 관리 | <ul style="list-style-type: none"> 사용자 유형별 권한관리 | |
| | 분석도구 | BI Report | <ul style="list-style-type: none"> Reporting을 통한 리포트 템플릿 구성 | |
| Analytics | | <ul style="list-style-type: none"> R 또는 Mahout 프로그램 기반 분석 환경 구성 | | |

| 대기능 | 기능목록 | | 구성요건 | 비고 |
|-------------|------------|----------------------------------|---|----|
| | 중기능 | 소기능 | | |
| 빅데이터 서비스 제공 | 워크 플로우 | OLAP | • 저장 데이터 조회 기반 구성 | |
| | | 빅데이터 처리 | • 오픈소스 기반의 빅데이터 프로세싱 흐름제어 구성 | |
| | | 시스템 공통관리 | • 시스템 환경설정, 사용자 정보관리, 장애, 배치관리 등 | |
| 모니터링 | 사용 현황관리 | • 시스템 현황, 클러스터 별, 기관별 현황 등 | | |
| | 민간/기관 정보관리 | • 이용신청 및 처리 관리 | | |
| 정보 서비스 API | 기관 서비스 | • 기관 서비스 대상 웹 UI 및 웹서비스 제공 기반 구축 | | |
| | 서비스 관리 | 서비스 등록 서비스 현황 | • 제공 서비스 정보 및 공통관리 • 서비스 이력, 사용기관별 통계 등 현황 정보 조회 | |

(나) 클라우드 환경 도입

- 빅데이터 공통기반은 정책정보 분석을 위한 서버, 스토리지 등 인프라 스트럭처 환경 제공하고 분석을 위한 관련 알고리즘과 관련 응용을 개발할 수 있는 환경을 제공해야 한다는 측면에서 3가지 클라우드 모델 중 SaaS 형태로 공통기반을 구성하는 것이 타당하며 1단계에서는 IaaS를 구축하는 것을 권고함

[빅데이터 공통기반 적용 클라우드 모델 선정]



- 공통기반은 빅데이터 구성요소에 SaaS 기반 클라우드 환경을 가미하여 조성함. 크게 서비스, 플랫폼, 인프라 영역으로 나뉘어지며 이외에 개인정보

보호 등 공통 기반 보안영역이 있음

[빅데이터 공통기반 클라우드 SW 아키텍처]



(다) 하드웨어 및 네트워크 구성방안

- 빅데이터 공통기반의 서버군은 x86기반의 블레이드 서버로 병렬 구성하여 장애 처리 및 확장이 용이하도록 구성하는 것을 원칙으로 하며 운영관리, 보안,EMS서버는 독립서버 형태로, Name Node 및 Data Node는 가상화하여 구성함

[클라우드 기반 빅데이터 공통기반 HW·NW 구성방향]

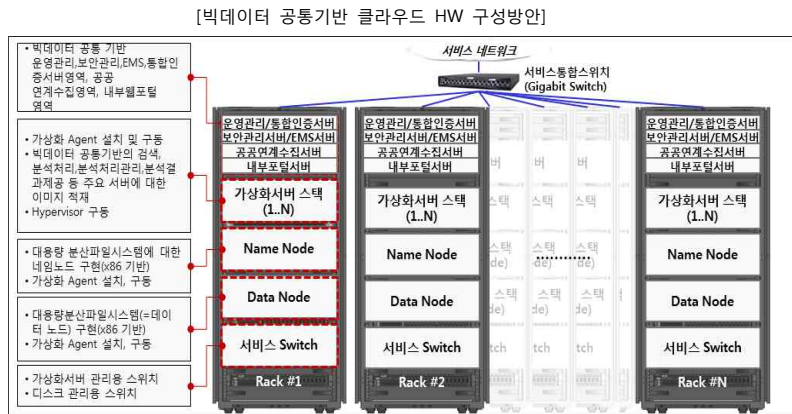
| 항목 | 설명 |
|---------|---|
| 하드웨어 구성 | <ul style="list-style-type: none"> • 빅데이터 공통기반은 관리서버와 클라우드 인프라로 구성되며 관리서버를 통하여 클라우드 인프라를 관리함 • 여러 개의 Zone으로 구성되며, 해당 Zone 안에는 VM(Virtual Machine)을 동작 시키는 서버가 존재함 • 빅데이터 공통기반의 서버 환경은 기본적으로 이중화를 고려하여 구성함 • Data Node는 향후 데이터 증가에 따른 확장이 가능하도록 구성함 • 비용 대비 효율적인 빅데이터 처리를 위해 x86 계열 서버를 적용하여 구성함 • 무장애 체계 구성을 위해 각각의 서버가 독립적으로 업무를 수행하다가 서버에 문제 발생 시 다른 서버가 업무를 대행하게 하도록 Active-Active(=Mutual Active) 형태로 구성함 |
| 운영관리 서버 | • 빅데이터 공통기반의 클라우드 관리를 위한 관리포털이 설치되며 독립서버 형태로 구성함 |
| 인증서버 | • 빅데이터 활용기관 이용자에 대한 SSO/EAM 기반 통합 인증을 처리하기 위한 서버로 독립서버 형태로 구성함 |

| 항목 | 설명 |
|-----------------------|---|
| 보안서버 | <ul style="list-style-type: none"> 바이러스, DB보안, 악성코드 보안, 키보드 보안을 위한 서버로 독립서버 형태로 구성 |
| EMS서버 | <ul style="list-style-type: none"> SMS, NMS 관리용 서버로 독립서버 형태로 구성함 |
| Name Node | <ul style="list-style-type: none"> 빅데이터 공통기반 클라우드 스택에서 가장 큰 단위의 집합구성이며 일반적으로 Data Node 50개 당 1개의 네임 노드를 배정하며 50개 이상이 되면 새로운 Name Node 를 생성함 |
| Data Node | <ul style="list-style-type: none"> Data Node는 1개의 Name Node에 대하여 구성함. 고성능 스위치와 x86기반 저비용 블레이드 서버를 사용하여 하나 이상의 클러스터를 구성함 클러스터는 하나 이상의 내장 스토리지를 탑재한 Host Machine으로 구성함 |
| 분석서버, 공공연계수집서버,내부 웹서버 | <ul style="list-style-type: none"> 분산검색엔진, 분석결과 DB서버, 분석작업 관리서버, 분석 처리서버, 분석정보 제공서버는 하나의 서버에 가상화 이미지 형태로 설치하여 운영하며, 동일한 가상이미지가 올려진 서버를 병렬로 구성하여 서버에 문제 발생 시 다른 서버가 업무를 대행하게 하도록 Active-Active(=Mutual Active) 함. 고성능 스위치와 x86기반 저비용 블레이드 서버를 사용하여 확장이 용이하도록 구성함 |
| 민간연계수집서버, 외부웹서버 | <ul style="list-style-type: none"> 민간 연계수집서버와 외부 웹서버는 DMZ Zone에 위치하며 가상화하여 구성하지 않고 각각 독립서버 형태로 구성함 |
| 네트워크 구성 | <ul style="list-style-type: none"> 인터넷 망 접근은 DMZ Zone을 통하여 접근하도록 구성함 가상화 환경에서 대용량 빅 데이터에 대한 처리에 적합하도록 구성함 백업회선을 확보하여 1개 회선 장애 발생 시에도 무중단 서비스 제공 라우팅 프로토콜을 이용한 이중 경로 확보 트래픽 분산 효과를 통한 서비스 성능 향상 VRRP(Virtual Routing Redundancy Protocol)을 통한 Fail Over 확보 정보 유통을 대비하여 원활한 서비스와 전송 안정성의 확보를 위해 대역폭 설정과 서비스 모니터링을 통한 통신 트래픽 관리를 고려해야 함 내부 네트워크는 인터넷 영역과 국가정보통신망 영역으로 구성하며 각각의 영역은 장비 및 회선 이중화 구성되어 각 경로에 대한 우회경로를 확보함 스위치는 L4스위치를 이용하여 구성하며 L2 스위치는 각 pod와 연결되는 구조로 pod 구분은 VLAN을 이용한 IP Class 제공하도록 구성하고 이중화 함 네트워크 기반 인프라에 대한 이중화 구성 및 고가용성 확보를 통해 장애에 유연하면서 대용량 데이터를 원활히 |

| 항목 | 설명 |
|----|--|
| | <ul style="list-style-type: none"> 수용할 수 있는 안정적인 네트워크 운영 환경을 구현 최적의 네트워크 구성을 위한 핵심요소는 고가용성 확보, 네트워크 부하분산, 장애 및 성능관리, 네트워크 보안, 연계망 인프라 측면에서 고려를 해야 함 가용성 확보 : 무중단 시스템의 구현을 위해 네트워크 장비 및 경로 이중화의 구성 필요 장애 발생에 대해 영향 범위 최소화 개념 적용 네트워크 부하분산 : 구성환경에 맞는 Load Balancing 알고리즘을 적용하여 최적의 업무환경 제공 주요서버 및 방화벽에 대한 부하분산을 통해 환경구성, 성능향상, 장애에 대한 투명한 처리 필요 장애 및 성능 관리 : 각각 시스템이 최적의 성능을 발휘할 수 있도록 성능 관리방안 제공 통합 모니터링 시스템을 통한 정확한 네트워크 상황 파악 및 모니터링 네트워크 보안 : 보유기관 및 활용기관의 원활한 연계를 위한 표준 프로토콜 제공 방화벽 및 침입방지시스템을 구축하고 네트워크 레벨의 보안정책 적용 연계망 인프라 : 행정정보공동이용시스템의 연계기반을 그대로 활용하고 추가 연계 기관에 대한 기반 인프라 조성 추가 연계 및 시스템 확장을 고려한 용량상정 : 네트워크 구성은 크게 4가지(웹, 내부, 활용기관연계, 행정선 연계)로 분리하여 각 영역간에는 방화벽을 통해 비인가된 접근을 제한하고 주요 서버에 대해서는 이중화 구성 및 부하분산 정책을 적용함 빅데이터 공통기반 웹 영역 : 방화벽을 통해 인터넷으로부터의 위협요소를 차단하고, 네트워크 장비의 이중화를 통해 장애 시 우회경로 확보 웹영역 및 내부서버영역에 침입방지시스템(IDS)을 구축하여 비인가된 접속 탐지 및 차단 활용기관을 위한 인터넷 기반의 웹포탈 서버가 위치하며 이중화된 인터넷 망에 연결 빅데이터 공통기반 내부 네트워크 : 서버들은 업무적으로 그룹핑하여 네트워크 대역을 분리하여 구성 주요 업무서버에 대해서는 L4 스위치를 통한 부하분산 및 이중화 구성 주요 서버에 대해서는 Workgroup switch와 이중화 연결하여 안정성을 확보함 통합관제시스템을 통해 효율적인 네트워크 인프라 관리 환경 구성. 활용기관 연계 네트워크 : 다수의 기관과 전용망 연결구성이 가능한 IP 연동기반의 네트워크 인프라 조성 행정 활용기관은 기 구축된 국가정보통신망을 통해 서비스 이용 활용기관 서비스 제공을 위한 기관연계서버가 위치하며, 내부 네트워크 구간과 연결은 2중 방화벽으로 제한 |

| 항목 | 설명 |
|----|---|
| | <ul style="list-style-type: none"> • 항공센 연계 네트워크 : 국가정보통신망으로 행정정보공동 이용센터를 통해 연결된 보유기관에 대해서는 기존의 네트워크 환경 최대한 활용 • 전용망 연결이 필요한 보유기관에 대해서는 별도의 이중화 된 연계 네트워크 환경 구성 |

- 빅데이터 공통기반은 확장 및 증설 용이성을 고려하여 운영관리서버, 통합 인증 서버, 보안서버, EMS서버 등 단독서버군, 가상화서버 스택, Name Node, Data Node 및 서비스위치를 기본 구성요소로 하여 복수개의 랙으로 구성함



- 네임노드는 데이터 노드가 40~50개 이상일 때 1개를 추가하도록 가상화하여 구성하고, 분석 관련 서버군도 가상화하여 구성함. 기타 데이터 트랜잭션이 많이 일어날 것으로 예상되는 연계수집서버(공공/민간)와 DMZ에 위치하는 분석정보 제공서버(민간)는 독립서버로 구성함

[빅데이터 공통기반 구성요소 가상화 여부]

| 구분 | 기능 | 비고 | 가상화 여부 |
|-----------|---|--|--------|
| Name Node | <ul style="list-style-type: none"> • Data Server 관리 / 모니터링 • 표준DB 및 파일시스템 메타 데이터 관리 • 메타DB에 메타데이터 저장 | <ul style="list-style-type: none"> • 데이터노드가 18~50개 이상이 되면 1개 추가 • 확장성을 고려하여 x86 기반 서버로 구성함 | ○ |
| Data Node | <ul style="list-style-type: none"> • 정형 및 비정형 빅데이터 저장 및 가공 (표준DB 포함) | <ul style="list-style-type: none"> • 네임노드 1개 당 최대 50개까지 추가 | ○ |

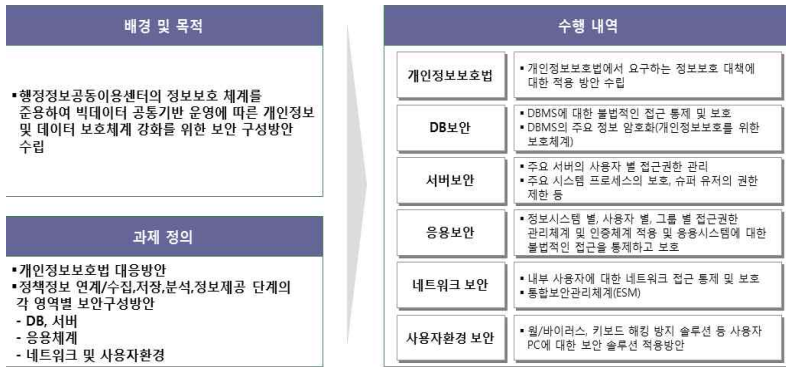
| 구분 | 기능 | 비고 | 가상화 여부 |
|---------------|---|--|--------|
| | | <ul style="list-style-type: none"> • 확장성을 고려하여 x86 기반 서버로 구성함 | |
| 연계수집서버 (공공) | <ul style="list-style-type: none"> • 공공연계수집서버는 행정정보 공동이용센터를 경유하여 대용량의 정책정보를 연계하고 분석결과를 제공하기 위한 연계 모듈 및 관련 API가 설치됨 | <ul style="list-style-type: none"> • 보유기관 연계수집서버 | X |
| 연계수집서버 (민간) | <ul style="list-style-type: none"> • SNS, 인터넷 데이터 연계수집용 서버 • Web Crawler SW가 설치 운영됨 | - | X |
| 분석정보제공 서버(민간) | <ul style="list-style-type: none"> • 민간 분석정보 제공용 웹포털 서버 | - | X |
| 분석정보제공 서버(공공) | <ul style="list-style-type: none"> • 활용기관 분석정보 제공용 웹포털서버 | - | X |
| 분산검색엔진 | <ul style="list-style-type: none"> • 수집 데이터 실시간 인덱싱, 타입별 검색처리 지원 | - | ○ |
| 분석결과 DB서버 | <ul style="list-style-type: none"> • 빅데이터 분석결과 저장 처리 | - | ○ |
| 분석작업 관리서버 | <ul style="list-style-type: none"> • 빅데이터 분석처리(실행)을 위한 작업 흐름관리서버 | - | ○ |
| 분석처리서버 | <ul style="list-style-type: none"> • 빅데이터 분석을 위한 처리 실행 서버 | - | ○ |
| 통합인증서버 | <ul style="list-style-type: none"> • GPK기반 SSO 사용자 인증 • 가상화 단말 인증 • 데이터 및 서버 접근에 대한 사용자 권한관리 | - | X |
| 보안관리서버 | <ul style="list-style-type: none"> • 바이러스, 악성코드, 키보드보안, DB보안 등 빅데이터 공통기반 보안관리 | - | X |
| EMS서버 | <ul style="list-style-type: none"> • EMS 마스터, SMS, NMS, DB/WAS Manager, EMS DBMS Reporting 모듈 • 빅데이터 공통기반 서버 모니터링 및 관리 | - | X |
| 운영관리서버 | <ul style="list-style-type: none"> • 운영 프로세스 자동화 • 자원 프로비저닝 자동화 관리 • 성능/장애 모니터링 • 미터링 : 서비스별 자원 사용량 수집 • 백업 및 보안 : VM 및 이미지 관리, 데이터 백업 및 복구, 시스템/데이터/네트워크 보안 | - | X |

(3) 정보보호체계 구성방안

(가) 개요

- 빅데이터 공통기반의 정보보호체계는 행정정보공통이용센터의 정보보호정책을 준용하여 정책정보의 연계/수집, 저장, 분석, 정보제공 단계별로 사용자환경, 응용, 서버, DB 영역의 보안 취약점에 대한 대응 및 강화 방안을 수립할 목적으로 함

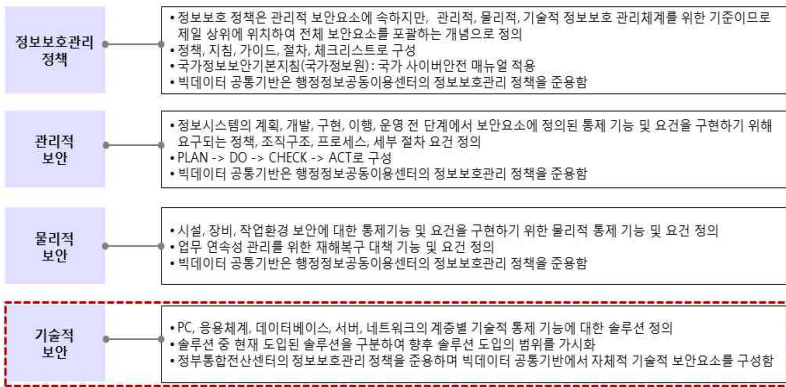
[정보보호체계 구축 방안 개요]



(나) 정보보호체계 구성방향

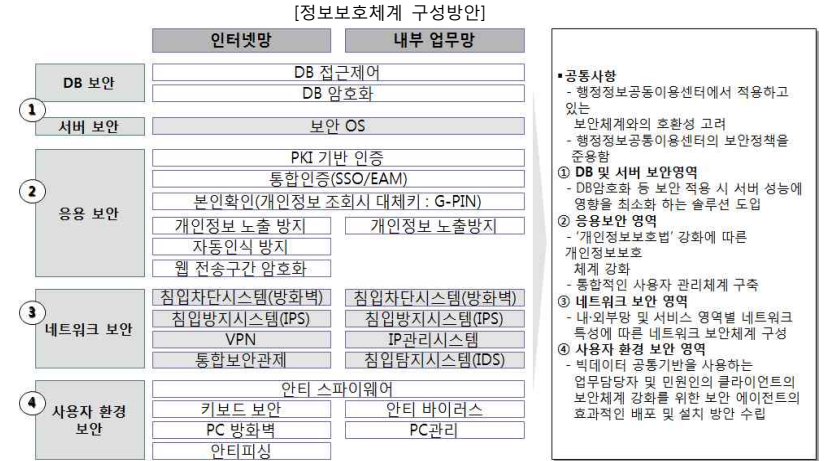
- 행정정보공통이용센터의 정보보안 체계를 준용하여 보안원칙, 기준 및 실행방안을 수립함

[정보보호체계 구성 방향]



(다) 정보보호체계 구성방안

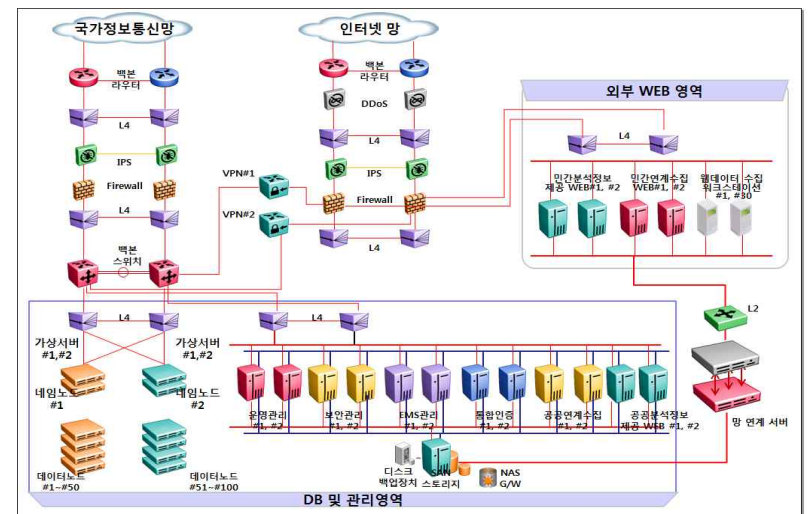
- 빅데이터 공통기반 정보보호체계 구성 시 시스템의 성능 및 서비스 제약을 발생 시킬 수 있는 요소에 대한 각 보안 영역별 고려사항을 정의함



(4) 인프라 아키텍처 구성

(가) 하드웨어 및 네트워크 구성

[하드웨어 및 네트워크 구성 개념도]



(나) 소프트웨어 구성

[소프트웨어 구성 개념도]

| 외부 WEB 영역 | DB 및 관리영역 | | | | | |
|--|--|--|--|---|---|--|
| 민간분석정보 제공 WEB #1.#2 WEB ESM EMS 서버보안 바이러스 웹구간보안 악성코드보안 키보드 보안 HA SW OS | 가상서버 #1.#2 WAS 배포관리 SW 분석개발도구 분석정보제공도움 (BI Reporting Tool) 분석 자원관리(Zookeeper) 분석처리모듈 (Pig, R, Mahout, Sqoop, Oozie 등) 분산검색엔진 바이러스 가상화 SW OS | 보안관리서버 #1.#2 WAS ESM 마스터 바이러스 마스터 서버보안 악성코드보안 HA SW OS | EMS 서버 #1.#2 WAS EMS 마스터 ESM 서버보안 백업 SW 바이러스 악성코드보안 HA SW OS | 운영서버 #1.#2 WAS 가상화자원관리 SW ESM EMS 백업 SW 바이러스 악성코드보안 서버보안 백업 SW HA SW OS | 공공연계수집서버 #1.#2 연계수집 SW 백업 SW ESM EMS 서버보안 바이러스 악성코드보안 HA SW OS | 공공분석정보 제공 WEB #1.#2 WEB ESM EMS 서버보안 바이러스 웹구간보안 악성코드보안 키보드 보안 HA SW OS |
| 민간연계수집#1.#2 연계 SW ESM EMS 서버보안 바이러스 악성코드 보안 HA SW OS | Name Node #1.#2 WAS Job Tracker 바이러스 백업 SW 가상화 SW OS | 망연계서버#1.#2 망연계 SW OS | Data Node #1.#100 WAS Map Reduce Task Tracker HBase HDFS 분석정보제공 DBMS DB 보안 바이러스 가상화 SW OS | SSO 서버 #1.#2 WAS SSO / EAM 연계수집 SW ESM EMS 서버보안 바이러스 악성코드보안 HA SW OS | | |

자) 활용과제별 공통기반 활용 방안

범국가적 데이터 공동활용을 위한 빅데이터 공통기반 구축 방안을 바탕으로 활용과제별 데이터 수집·연계·구성·분석 서비스 제공을 위한 빅데이터 공통기반 활용 방안을 제시함

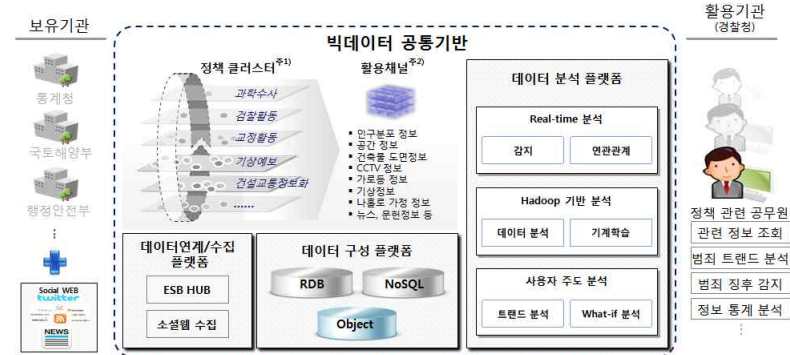
[활용과제별 공통기반 활용방안 개요]

| 활용과제 | 방법 |
|---|--|
| <ul style="list-style-type: none"> • 범죄 예측 정책활용과제 • 재난·재해 전조감지 정책활용과제 • 민원 환류 정책활용과제 • 교통사고 개선 정책활용과제 | <ul style="list-style-type: none"> • 활용과제별 BPR을 통해 데이터 수집·연계가 필요한 기관 및 시스템 파악 • 활용과제별 빅데이터 공통기반 수집·연계·구성·분석 플랫폼 활용 방안 제시 |
| 수행 절차 | |

(1) 범죄 예측 정책활용과제 구현 방안

활용기관에서 클러스터 검색/조회를 통해 범죄 예측 분석에 필요한 데이터를 선택하여 활용채널을 구성하면 데이터 연계/수집 플랫폼, 구성 플랫폼을 통해 해당 데이터를 연계/수집/저장하고 분석 목적에 맞게 기능을 지원함

[범죄 예측 정책활용과제 구현 개념도]



(가) 연계·수집·저장 대상 데이터

○ BPR을 통해 파악된 범죄 예측 정책활용과제 연계·수집·구성 대상 데이터 현황은 다음과 같음

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집방법 | 데이터구성 |
|------------------|--------------|--|------------------------|------|--------|
| 통계청 | 인구조사통계정보 | 지역의 소득수준, 주민분포, 직종, 성별, 교육수준 등의 인구센서스 수집 통계정보 | 통계시스템 | 수집 | RDB |
| 통계청 | 인구분포정보 | 지역별 소득수준, 주민분포, 직종, 성별, 교육수준 등의 인구센서스 수집 통계정보 | 통계시스템 | 수집 | RDB |
| 국토해양부 국립지리정보원 | 공간지도정보 | 수치지형도, 항공사진 | 국가공간정보 유통시스템 | 연계 | Object |
| 국토해양부 | 공간(GIS)정보 | 주제도, 발생지(법정동 주소)위치 좌표값 | 국가공간정보 유통시스템 | 연계 | RDB |
| 국토해양부, 지자체 | 지역용도지구 정보 | 지역별 지역/지구/구역 정보 - 세부지역단위 구분(1,2종 근생, 주거지역, 상가지역 등) | 세운터, 부동산통합관리 시스템 | 연계 | RDB |
| 국토해양부, 지자체 | 건축물도면정보 | 건물도면(평면도, 배치도) | 건축행정시스템 (세운터) | 연계 | Object |

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집 방법 | 데이터 구성 |
|-----------------------------|------------|---|---------------|-------|------------------|
| 시군구(230) 시설관리공단, 소방방재청, 경찰청 | CCTV정보 | CCTV설치 위치, 관리기관(자), 성능, 영상관측결과(정보, 영상) | 시설물 관리시스템 | 연계 | NoSQL DB |
| 경찰청, 소방방재청, 지자체, 도시관제센터 | CCTV정보 | 이상징후 CCTV영상기록정보, CCTV관리주체, 설치위치정보 | CCTV영상관리시스템 | 수집/연계 | NoSQL DB |
| 시군구 시설관리공단 | 가로등정보 | 가로등 설치위치, 조도, 관리기관(자) | 시설물 관리시스템 | 연계 | RDB |
| 보건복지부, 지자체 | 나홀로가정 아동정보 | 지역별 나홀로가정 아동, 소외계층 현황 분포 | 사회복지통합망(e-행복) | 연계/수집 | RDB |
| 기상청 | 기상정보 | 시간/지역별 날씨정보 | 기상청시스템 | 연계 | RDB |
| 행정안전부 | 행정공간정보 | 행정주제도, POI좌표값(상가밀집도, 단독주택/원룸, 공동 등) | 공간정보시스템 | 연계 | RDB |
| 인터넷뉴스포털, 문헌정보포털 | 뉴스, 문헌정보 | 시즌/시기별 범죄사건 뉴스, 범죄 관련 논문 및 보고서 정보 | - | 수집 | NoSQL DB, Object |
| 빅데이터 공통기반 지원센터(가칭) | 범죄징후탐지 정보 | 범죄징후 탐지 소셜 분석결과정보, 위치정보(POI값), 범죄키워드 별 결과정보 | 빅데이터 공통기반 | 연계 | RDB |

(나) 분석 플랫폼 지원 기능

- 범칙 예측 정책활용과제는 범칙예측 정확도 향상을 위한 정형/비정형 정보 활용 범칙 트렌드 분석 및 예측 모델 시뮬레이션 지원과 범칙징후 사전탐지 대응을 위한 실시간 범칙 관련 키워드 감지 및 비정형 스트림데이터의 미분석 지원이 요구됨

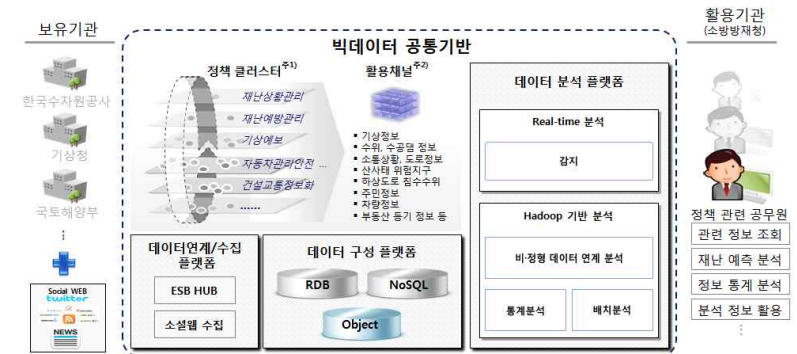
| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|--------------|-----------------------|--|-------------------------|------------|
| 범칙 예측 정확도 향상 | 정형/비정형 정보활용 범칙 트렌드 분석 | 정형/비정형 정보를 대상으로 데이터 분석, 기계학습, 트렌드 분석 기능을 통해 범칙에 대한 트렌드 분석 지원 | 사건 정보 뉴스 정보 문헌 정보 | 시기별 범칙 트렌드 |
| | 예측모델 시뮬레이션 수행 | 범칙예측 모델 개선을 지원하기 위해 다양한 변이요소 항목을 반영하여 시뮬레이션을 지원하는 기능 | 범칙예측 변이요소 정보 항목 | 시뮬레이션 결과 |

| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|----------------|------------------|------------------------|--------|--------|
| 범칙 징후 사전 탐지 대응 | 실시간 범칙관련 키워드 감지 | Real-time 분석 (감지 기능) | 소셜 데이터 | 범칙 징후 |
| | 비정형 스트림 데이터 의미분석 | Real-time 분석 (연관관계 기능) | 소셜 데이터 | 위험도 결과 |

(2) 재난· 재해 전조감지 정책활용과제 구현 방안

활용기관에서 클러스터 검색/조회를 통해 재난· 재해 전조감지 분석에 필요한 데이터를 선택하여 활용채널을 구성하면 데이터 연계/수집/구성 플랫폼을 통해 해당 데이터를 연계/수집/저장하고 분석 목적에 맞게 기능을 지원함

[재난· 재해 전조감지 정책활용과제 구현 개념도]



(가) 연계·수집·저장 대상 데이터

- BPR을 통해 파악된 재난· 재해 전조감지 정책활용과제 연계· 수집· 구성 대상 데이터 현황은 다음과 같음

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집 방법 | 데이터 구성 |
|------|------|-----------------------------------|------------|-------|--------|
| 기상청 | 기상정보 | AWS 관측자료, 해일정보, 해상예보, 지진통보, 주간해상, | 종합기상 정보시스템 | 연계 | RDB |

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집방법 | 데이터구성 |
|------------|---------------------|---|--------------------------|-------|--------------|
| | | 육상예보, 산악예보, 낙뢰관측정보, 기상정보, 국내기상등 | | | |
| 한국수자원공사 | 지자체 수위, 우량, 수공 댐 정보 | 우량관측소, 수위관측소, 시수위, 시댐정보, 시강우량, 수문기상실황감시정보, 댐특성 정보, 저수용량 정보 등 | 지자체 수공간 수문관측자료 공동활용시스템 | 연계 | RDB |
| 한국도로공사 | 소통상황, 도로정보 | IC코드, 휴게소코드, 특별상황발생관리, 구간정보, 고속도로정보, 고속도로노선, 도로소통정보, 도로 CCTV 영상정보 등 | 고속도로 교통관리시스템 | 연계 | RDB |
| 산림청 | 산사태 위험지구 | GIS맵, 운항지역, 사방댐상세정보, 산사태위험등급구분도, 산사태발생위험예보도, 산불위험지수실측 등 | 산사태 위험지 관리시스템 | 연계 | RDB |
| 국토해양부 /경찰청 | 한강·중랑천 하상도로 침수수위 정보 | 수계열 주요지점 침수수위 정보, 침수우려 취약도로 정보 등 | 수자원관리종합 정보시스템/도시 교통정보시스템 | 연계 | Object |
| 민간 | SNS 정보 | 재난유형 및 위치정보 포함 SNS 데이터 (Twitter, Me2day) | SNS | 수집 | NoSQL DB |
| 행정안전부 | 주민정보 | 전파대상 주민 연락처및 소재정보 | 주민등록정보시스템 | 연계 | RDB |
| 국토해양부 | 차량정보 | 차주정보, 차량운행정보 (인근 CCTV/활용) | 자동차관리정보 시스템 | 연계/수집 | Object |
| 국토해양부 /지자체 | 지자체 수위정보 | 홍수위험지도, 하천이력관리정보 | 하천관리지리정보시스템 | 연계 | RDB / Object |
| 법무부/대법원 | 부동산등기정보 | 부동산 소유정보및 소재정보 | 등기정보시스템 | 연계 | RDB |

(나) 분석 플랫폼 지원 기능

- 재난·재해 전조감지 정책활용과제는 재난·재해 예측 모델 개선 업무를 지원하기 위해 분석 플랫폼의 감지기능, 통계분석, 배치분석, 비·정형 데이터 연계 분석 기능의 지원이 요구됨

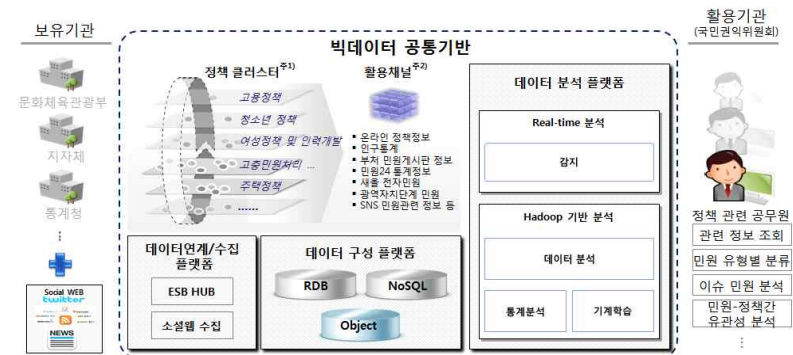
| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|----------|---|--|---------|----------------|
| 예측 모델 개선 | 실시간 전조 포착 Real-time 분석 (감지 기능) | SNS에서 수집되는 경우 재난 관련 데이터를 실시간으로 분석하여 재난·재해 발생에 대한 전조 감지 및 알림 지원 | SNS 데이터 | 재난·재해 전조 수집 정보 |

| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|--------------|------------------------------------|--|-------------------------------------|---------|
| 비정형 데이터 분석 | Hadoop분석 (통계 분석, 배치 분석) | 축적된 재난·재해 전조 자료를 기반으로 시간·지역 단위별통계 및 피해 가중 지역 도출을 위한 분석 기능 지원 | 재난·재해 전조 수집 정보 재난·재해 수집/연계 데이터 등 | 상세 피해지역 |
| 정형·비정형 연계 분석 | Hadoop분석 (비·정형 데이터 연계분석) | 관련 기관의 정형 데이터와 SNS 등을 통해 수집된 비정형 데이터를 연계하여 침수 전조 및 위험도 분석 지원 | 재난·재해 수집/연계 데이터 SNS 데이터 등 | 침수 예산지역 |

(3) 민원 환류 정책활용과제 구현 방안

활용기관에서 클러스터 검색/조회를 통해 민원 환류 분석에 필요한 데이터를 선택하여 활용채널을 구성하면 데이터 연계/수집/구성 플랫폼을 통해 해당 데이터를 연계/수집/저장하고 분석 목적에 맞게 기능을 지원함

[민원 환류 정책활용과제 구현 개념도]



(가) 연계·수집·저장 대상 데이터

- BPR을 통해 파악된 민원 환류 정책활용과제 연계·수집·구성 대상 데이터 현황은 다음과 같음

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집방법 | 데이터구성 |
|------------|--------------|---|------------|------|----------|
| 문화체육관광부 | 온라인 정책정보 | 공감코리아, 아고라, 중앙일간지정책, 온라인 일일동향 등의 정부 정책 관련 변경 및 수립 정보, 관련 보도자료 | 공감코리아 | 연계 | Object |
| | 정책캘린더 | 주요 정부정책 발표일정 및 해당부처 | - | 연계 | Object |
| 통계청 | 인구통계 | 지역별 인구수, 성별, 나이 등 통계정보 | 국가통계포털 | 연계 | RDB |
| 국민권익위원회 | 국민신문고 민원 | 민원인, 민원발생 날짜, 민원 발생지, 처리기관, 민원내용 정보 | 민원정보 분석시스템 | 연계 | RDB |
| | 110 콜센터 상담내용 | | | 연계 | Object |
| 해당부처 | 부처민원게시판 민원정보 | | | 연계 | NoSQL DB |
| 지자체 | 지자체 민원정보 | | | 연계 | NoSQL DB |
| SNS 업체 | SNS 민원관련정보 | 트위터 상 민원관련정보 | SNS | 수집 | NoSQL DB |
| 행정정보공공이용센터 | 민원24 통계정보 | 유형별 급증정형민원 시계열 통계 정보 | 민원24 | 연계 | RDB |
| 통계청 | 기타 통계정보 | 사회적 지표가 될 수 있는 통계정보 (기초생활수급자, 아동, 독거노인, 범죄발생 관련통계 등) | 통계시스템 | 연계 | RDB |
| 해당 시군구 | 새울 전자민원 | 시,군,구 별 자체 전자 민원 정보 | 새울 행정시스템 | 연계 | RDB |
| 해당시도 | 광역자치단체 민원 | 특별시, 광역시, 도 별 자체 전자 민원 정보 | 민원 행정시스템 | 연계 | RDB |

(나) 분석 플랫폼 지원 기능

- 민원 환류 정책활용과제는 민원관련 정보 분석 방향 설정을 위한 유형별 분류 및 군집화 및 이슈 민원 List-up과 민원정보 분석 및 신규 민원정책 제시를 위한 민원 주제별 분석 지원이 요구됨

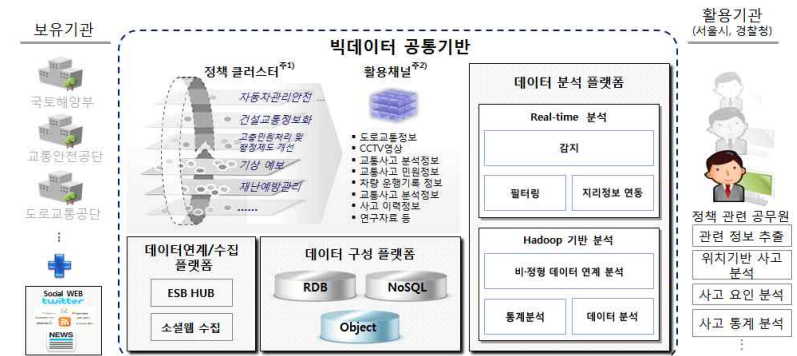
| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|----------------|--------------|----------------------------|---|-----------|
| 민원 관련 정보 분석 방향 | 유형별 분류 및 군집화 | Hadoop분석 (데이터 분석, 기계학습 기능) | 수집된 민원관련 통계정보를 분석하여 민원정보 분류기준의 상세화를 위한 분류 및 군집화 기능 지원 | 민원 유형별 분류 |

| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|------------------------|---------------|--------------------------------------|---|-------------------------|
| 설정 | 이슈 민원 List-up | Hadoop분석 (통계분석) Real-time 분석 (감지 기능) | 분류 및 군집화된 민원정보 중에 정해진 기간내 다량 발생 또는 반복적으로 이슈가 되는 민원 주제를 도출하기 위한 통계 기능 및 감지 기능 지원 | 민원정보 이슈 민원 List |
| 민원 정보 분석 및 신규 민원 정책 제시 | 민원 주제별 분석 | Hadoop분석 (데이터 분석, 기계학습 기능) | 도출된 이슈 민원을 바탕으로 민원 주제별 관련 정보 및 기존 정책 정보와 연관성 등을 분석하기 위한 데이터 분석 및 기계학습 기능 지원 | 부처별 민원 분석 결과 기존 민원 정책정보 |

(4) 교통사고 개선 정책활용과제 구현 방안

활용기관에서 클러스터 검색/조회를 통해 교통사고 개선 분석에 필요한 데이터를 선택하여 활용채널을 구성하면 데이터 연계/수집/구성 플랫폼을 통해 해당 데이터를 연계/수집/저장하고 분석 목적에 맞게 기능을 지원함

[교통사고 개선 정책활용과제 구현 개념도]



(가) 연계·수집·저장 대상 데이터

- BPR을 통해 파악된 교통사고 개선 정책활용과제 연계·수집·구성 대상 데이터 현황은 다음과 같음

| 보유기관 | 정보명 | 수요정보(항목) | 관련시스템 | 수집방법 | 데이터구성 |
|---------------|------------|---|-------------------|------|----------|
| 경찰청 | 도로 교통정보 | • 사고 잦은 지점, 주정차 허용구간 • 맞춤형 정보, 통제 정보, 교통량 (소통정보) | 지능형 교통체계(ITS) | 연계 | RDB |
| 경찰청 국토해양부 | CCTV 영상 | • 주요 도로 및 교차로에 설치되어 실시간 차량 소통상황 등을 실시간으로 전송함 • 교통사고, 안개, 결빙 등 각종 교통정보를 영상으로 촬영하여 교통정보센터에 전송함 | | 연계 | Object |
| 국토해양부/ 지자체 | 돌발상황 발생정보 | • 공사및 사고 지점 정보, 정보 등록 시간, 사고 요인 • 교통사고 상황/ 공사 진행 상황 정보 • 사고처리 현황정보, 작성자/ 정보 제공 기관 | | 연계 | RDB |
| 경찰청 | 교통사고 분석정보 | • 교통사고 피해상황, 현장상황, 유발 요인 • 현장약도, 조사자 의견 | 교통사고 관리시스템 (TAMS) | 연계 | RDB |
| 국토해양부 | 공간정보(GIS) | • 주제도, 발생지(법정동 주소)위치 좌표값 | 국가공간정보 유통시스템 | 연계 | RDB |
| SNS 포털 | 교통사고 민원정보 | • 주민개선의견 • 교통사고 위치·상황 | SNS 포털 | 수집 | Object |
| 교통안전공단 | 차량 운행기록 정보 | • 사고운행 패턴, 운행분석 종합진단 | 운행기록분석시스템 (eTAS) | 연계 | RDB |
| 도로교통공단 | 교통사고분석정보 | • 교통사고 피해상황, 현장상황, 유발 요인 | 교통사고 분석시스템 (TAAS) | 연계 | RDB |
| 교통정보연구원 | 연구자료 | • 사고 요인 연구 정보 | - | 연계 | NoSQL DB |
| 손해보험사 | 사고 이력정보 | • 사고 발생상황, 사고 발생요인 | - | 연계 | NoSQL DB |

(가) 분석 플랫폼 지원 기능

- 교통사고 개선 정책활용과제는 도로운행 환경 실시간 안내 및 교통안전 시설물 설치 분석을 위한 자연어 처리를 통한 필터링, 위치기반 데이터 분석, 정량기반 데이터 분석, 연계분석 등의 지원이 요구됨

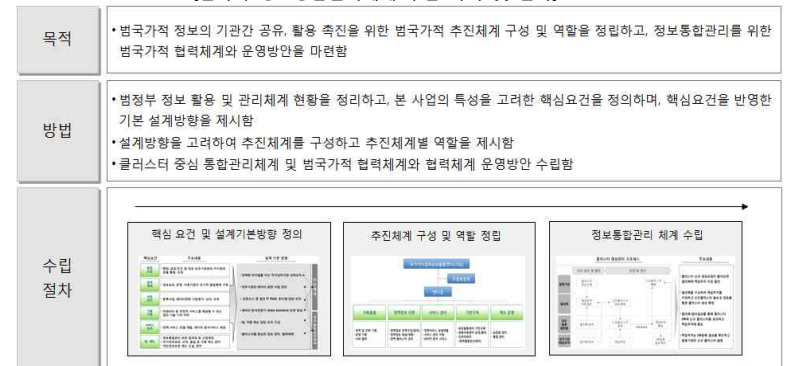
| 분석 업무 | 분석 플랫폼 활용 기능 | 빅데이터 공통기반 기능정의 | Input | Output |
|----------------|---|---|----------------------|----------------|
| 자연어 처리를 통한 필터링 | Real-time 분석 (필터링 기능, 감지 기능) | 자연어 처리를 통해 유의미한 정보를 추출하고 불용어 및 스팸 문서 등을 차단하기 위한 필터링 기능 지원 | 교통사고 관련 SNS 정보 | 교통사고 관련 정보 추출 |
| 위치 기반 데이터 분석 | Hadoop 분석 (데이터 분석) Real-time 분석 (지리정보 연동 기능) | 위치를 나타내는 정보를 가진 키워드 및 문장을 추출하기 위한 텍스트 분석 지원 | SNS 필터링 정보 지리정보 | 위치기반 사고정보 |
| 정량기반 데이터 분석 | Hadoop 분석 (통계 분석) | 문서의 발생 빈도, 위치정보, 시간정보 등을 정해진 시간 및 지역 단위별 통계 산출을 위한 기능 지원 | SNS 필터링 정보 사고정보 | 시간 및 지역 단위별 통계 |
| 연계분석 | Hadoop 분석 (비-정형 데이터 연계분석) | 교통사고 다발지역 사고요인에 대한 세밀한 분석을 위한 정형 및 비정형 정보의 연계 분석 지원 | 교통사고정보 도로정보 사고요인정보 등 | 사고요인 분석 결과 |

차) 범국가 정보통합관리체계 수립

(1) 개요

- 범국가 정보통합관리체계 수립은 범국가적 기관간 공유, 활용 촉진을 위한 추진 체계구성 및 역할을 정립하고, 정보통합관리를 위한 협력체제와 운영안을 마련함

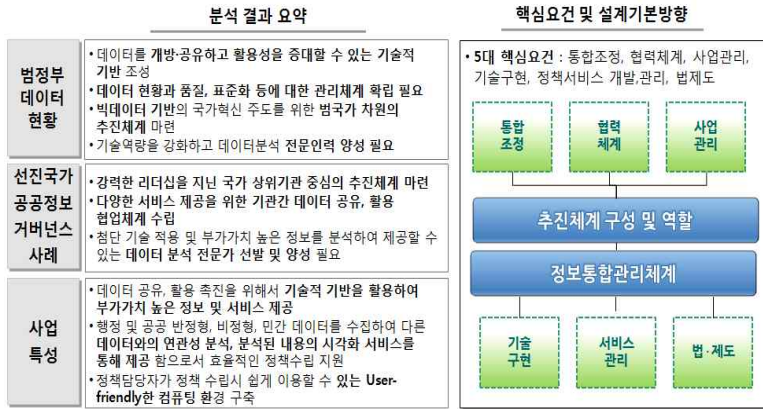
[범국가 정보통합관리체계 수립 목적 및 절차]



(2) 핵심요건 및 설계기본방향

- 범정부 데이터 현황, 선진국가 공공정보 거버넌스 사례, 본사업 특성을 분석, 검토하였으며, 그 분석결과를 바탕으로 범국가 정보통합관리체계 수립을 위한 핵심요건을 도출하고 설계기본방향을 수립함
- 설계기본방향은 추진체계구성 및 역할과 정보통합관리체계로 제시함

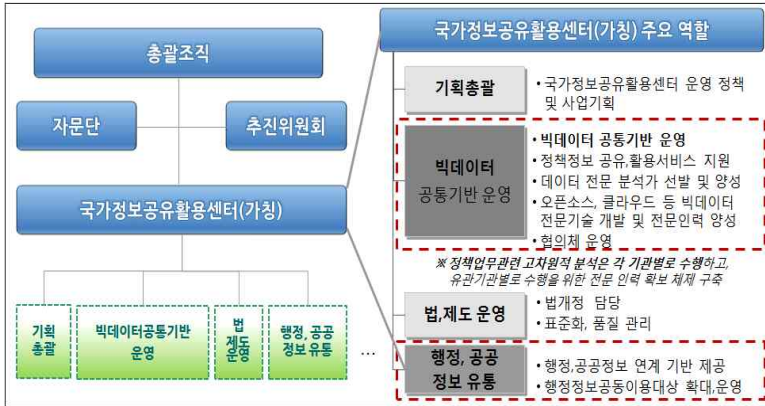
[핵심요건 및 설계기본방향]



(3) 추진체계 구성 및 역할

- 범국가 정보통합관리를 위한 추진체계는 자문단, 추진위원회, 국가정보공유활용센터로 구성하며, 국가정보공유활용센터의 주요 역할은 빅데이터 공동기반 운영과 행정, 공공 정보 유통임

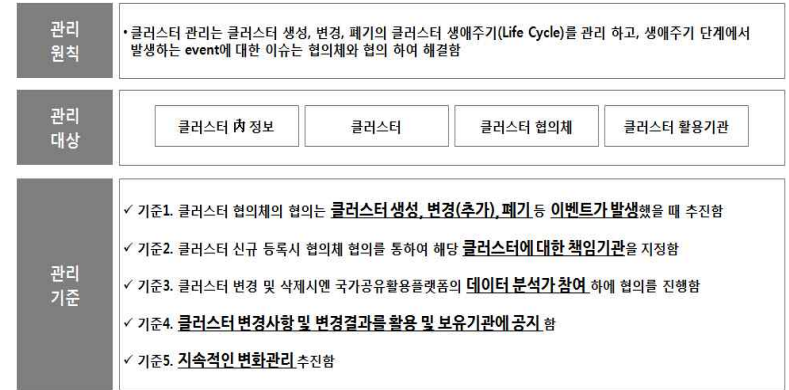
[범국가 정보통합관리 추진체계 및 주요역할]



(4) 정보통합관리체계

- 범국가 정보통합 관리는 클러스터 중심으로 관리하며, 데이터분석가, 정보 소관 부처, 국가정보공유활용센터가 협의체를 구성하여 정보를 통합관리함
- 관리 대상은 클러스터 내 정보, 클러스터, 협의체 등이며, 통합의 중심점인 클러스터 변경사항을 지속적으로 관리함으로써 정보통합관리 틀을 갖추어 감

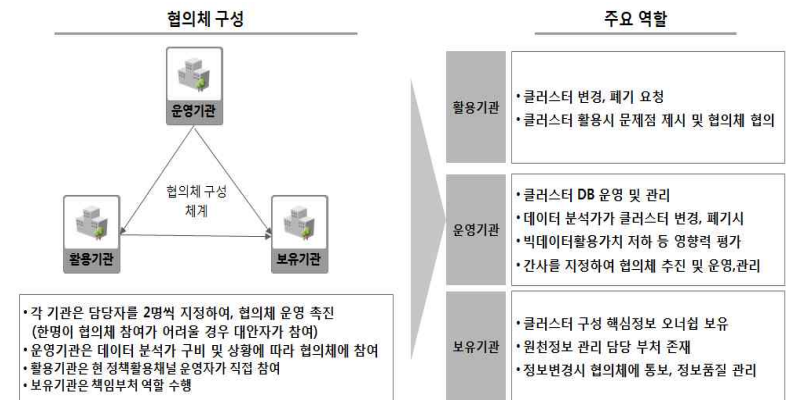
[정보통합관리 원칙 및 관리기준]



(5) 협의체 구성 및 역할

- 정보통합의 중심점인 클러스터를 잘 관리하기 위해서는 기관간 협의체가 구성되어야 하며, 협의체는 범국가 정보를 통합관리하는데 있어서 이슈를 해결하고 주요 의사결정의 수행함

[범국가 정보통합관리를 위한 협의체 구성체계]



6) 법제도 개선

가) 개요

(1) 법제도 개선이슈

- 빅데이터 활용을 위하여, 필요한 법제도적 이슈들을 점검하고 입법방식에 대한 고려사항을 도출

| | |
|---------------|---|
| 실체적 문제 | <ul style="list-style-type: none"> · 빅데이터 활용과제는 다양한 행정분야와 광범위한 행정목적 실현을 위해 이용되어야 비로서 의미를 가질 수 있으나, 다수의 법령에 걸쳐 제약사항이 존재 <ul style="list-style-type: none"> ① 행정정보 및 지식정보자원의 공유, 유통, 활용을 규정하는 정보화법령은 단순방침, 선언적 의미 ② 행정조사기본법상, 조사의 법률유보, 최소수집, 목적부합성, 중복조사 금지 등 다양한 규제원리 ③ 행정기관간 독점적, 배타적으로 부여된 행정권한으로 인해 상호간 공유, 협업이 어려움 ④ 강화된 개인정보보호법으로 인하여 단계별 동의, 제3자 제공금지, 민감정보 처리 금지 등 |
| 절차적 문제 | <ul style="list-style-type: none"> · 개별 법령별로 제약사항을 해결하고, 빅데이터 활용의 법적 근거를 마련하고 제약사항을 해결하는 것은 현실적으로 매우 어려움 · 현행 대한민국 법률 1,262개, 대통령령 1,480개 등 총 4209개 법령이 발효 중이며, 법률 중 80% 이상인 1000여개 법률이 행정법 영역에 해당(법제처 법령통계 자료('12.11.01)) · 개별제약사항이 있는 법률을 개정하는 경우, 수십~수백개의 법률이 대상이 될 수 있으므로 물리적으로 상당한 시간과 노력이 필요 |
| 고려사항 | <ul style="list-style-type: none"> · 다양한 법령상의 장애요인을 해결하고, 후속사업의 원활한 추진을 위하여 실행용이하고 신속한 입법방식에 대하여 검토 필요 |

(2) 법제도 추진방식 검토

- 입법방식별 장단점을 비교분석하여, 실행용이하고 신속하게 개정하고 정책과제를 추진할 수 있는 대안을 모색

| | 추진방식 | 장단점 분석 |
|----------------------|---|--|
| 독자법 제정 | <ul style="list-style-type: none"> · 빅데이터 활용을 위한 독립된 법률 형태의 새로운 법률안 제정 · 정보자원 관리에 대한 대표법으로 기능할 수 있으며, 유사 법률조항의 통폐합으로 입법적 효율성 증대 | <ul style="list-style-type: none"> · 장점 : 법률 개정보다 체계적으로 정비가능, 기존법 개정보다 강한 홍보효과 기대 가능 · 단점 : 독자입법에 따른 의안과정에서 상당한 시간과 노력이 필요, 법률제정 지연시 후속사업 차질우려 |
| 의안·연구 중인 법안반영 | <ul style="list-style-type: none"> · 공공데이터의 이용 및 제공에 관한 법률안(김을동 의원 대표발의안)에 빅데이터 추진 근거 신설 · 공공기관 정보자원관리에 관한 법률(연구법안) | <ul style="list-style-type: none"> · 동 법안은 국가에서 대민을 향한 데이터 공개에 초점을 맞춘 법안이므로, 행정기관 내부적 정보자원 활용과는 다른 내용과 체계로 구성됨 · 자원관리의 초점은 SW, HW, 인력 등 광범위하고 포괄적으로 다루어지므로 빅데이터와는 다른 관점에서 접근 |
| 기존법의 개정 | <ul style="list-style-type: none"> · 국가정보화기본법을 개정하여, 빅데이터 추진근거 마련 · 전자정부법을 개정하여 빅데이터 추진근거 마련 | <ul style="list-style-type: none"> · 국가정보화기본법은 해당 국가정책 분야에 대한 원칙, 방향, 이념 등에 초점이 되어 세부 절차를 다루기에 부적절 · 개정이 용이하고, 소요기간이 비교적 단기, 입법목적과 내용상 관련성이 높은 법률 |

(3) 법제도 정비방안 도출

- 전자정부법 개정방식을 중심으로 정비되어야 할 내용과 방식을 검토

| | |
|-------------------|---|
| 정비방식 | <ul style="list-style-type: none"> · 전자정부법을 개정하되, 새로운 장을 신설하고, 필요하거나 관련된 내용을 통합하여 기술 ※ 현재 전자정부법은 7개의 장과 78개의 조문으로 구성 |
| 주요정비 내용 | <ul style="list-style-type: none"> · 빅데이터 분석 및 활용근거 마련 · 기관간 정보 공유활용에 대한 책임과 권한 문제 · 개인정보 활용을 위한 보호위원회의 심의·의결 · 정책과제별 활용대상 정보와 분석결과의 공개·공표 |
| 기타 법령 개정방식 | <ul style="list-style-type: none"> · 기타 정책과제별 다수법령을 개정할 필요가 있는 경우에는 일괄정비 법안을 통하여 물리적, 시간적 제약을 최소화 ※ 유사입법례(하나의 법률로 다수의 법률을 일괄개정하는 입법례) 행정절차법의 시행에 따른 공인회계사법 등의 정비에 관한 법률, 정부부처 명칭 등의 변경에 따른 건축법 등의 정비에 관한 법률, 독점규제 및 공정거래에 관한 법률의 적용이 제외되는 부당한 공동행위 등의 정비에 관한 법률, 금융감독기구의 설치 등에 관한 법률 제정 등에 따른 공인회계사법 등의 정비에 관한 법률, 금융감독기구의 설치 등에 관한 법률제정 등에 따른 공채채 등록법시행령 등의 정비에 관한 규정 |
| 한계 | <ul style="list-style-type: none"> · 의료정보, 진료기록 등을 분석하기 위한 의료법, 범죄경력 및 수사자료를 이용하기 위한 형의 집행 및 실호 등에 관한 법률 등 일부 민감한 정보에 대해서는 정부입법안 추진과정에 필요한 부처간 협의, 공청회 등에 상당한 진통이 예상됨 |

나) 법제도 정비방안

- 빅데이터 활용을 위한 법제도 정비를 위하여 전자정부법 개정(안)을 도출

| 제7장 정책수립 지원체계 구축 < 신설 > | |
|-------------------------|--|
| 제1절 정보자원의 공유·활용 | |
| 제76조(공유·활용의 기본원칙) ① | 행정기관 등의 장은 행정정보, 공공정보 등 대규모 정보자원 등을 효율적으로 활용하여 행정업무의 효율성을 향상시키고, 국민의 복지와 편익을 증진시키기 위해 노력하여야 한다. |
| ② | 행정기관 등의 장은 정책수립을 위한 정보수집, 민원서비스의 융·복합, 신규 서비스의 창출, 융합행정의 촉진 등을 위하여 공공정보를 다른 기관과 통합하여 활용할 수 있다. |
| ③ | 행정기관 등의 장은 소관 사무와 관련하여 보유하고 있는 데이터 등의 공공정보에 대한 현황 및 관리체계를 점검하여야 하며, 다른 기관과의 공공정보를 공유·활용 있도록 적극적으로 협력하여야 한다. |
| ④ | 행정기관 등은 정보자원을 공유·활용하는 경우, 특정 개인이나 법인 등의 사생활, 영업상 비밀 등을 침해할 수 없으며, 국민의 권리를 제한하거나 의무를 부과하는 등 행정규제 목적으로 하거나 불이익한 조치를 목적으로 이용할 수 없다. |

⑤ 행정기관 등은 공유·활용하려는 정책의 내용 및 목적, 대상정보의 종류와 유형 및 분석결과 등을 국민들이 알 수 있도록 공개·공표하여야 한다.

제77조(정보자원의 공유·활용) ① 행정기관등은 정책을 수행하거나 의사결정을 하는데 필요한 정보를 수집하고 자료를 분석하기 위하여, 다른 행정기관등이 직무상 수집, 작성, 취득, 관리하고 있는 정보자원을 공유·활용할 수 있다.

② 행정기관 등은 민간에서 작성, 관리되어 공중에 공개·공표되는 정보와 자료 및 정보통신망을 통해 작성되어 일반에게 공개되는 게시글 등의 자료를 수집·분석할 수 있다.

제78조(공유·활용 대상정보) ① 제77조에 따라 행정기관 등이 공유·활용할 수 있는 정보자원의 종류는 다음 각호와 같다.

1. 「통계법」에 따른 통계정보
2. 「기상법」에 따른 기상정보
3. 「재난 및 안전관리법」에 따른 재난정보
4. 「국가공간정보에 관한 법률」에 따른 공간정보
5. 「국가통합교통체계효율화법」에 따른 교통정보
6. 「보건의료법」에 따른 보건의료통계
7. 「대외무역법」 및 「무역거래기반조성에 관한 법률」에 따라 작성되는 무역·통상 등의 경제정보
8. 기타 관련 법령에서 공개, 공표하도록 하거나 공중에 널리 이용하도록 제공·보급하도록 정하는 정보

③ 행정기관 등은 「공공기관의 정보공개에 관한 법률」에 따라 일반국민에게 공개될 것으로 예상되는 정보에 대하여 다른 행정기관의 요청이 있는 경우에 제공할 수 있다.

제79조(비공개·비밀정보의 법정이용) 행정기관등이 관련 법령에서 목적외 사용을 금지하거나, 비밀 또는 비공개 사항으로 규정된 정보를 이용하려는 경우에는 정책의 목적 및 내용, 수집·이용하고자 하는 정보의 대상 및 범위 등에 대하여 대통령령으로 정하는 바에 따라 수집·이용할 수 있다.

② 행정안전부 장관은 제1항에 따라 이용하는 대상정보와 정책목표의 부합성, 정보의 보안과 안정성 등에 대하여 심의하여야 한다.

제80조(민간정보의 수집·제한) ① 행정기관 등이 일반공중에 공개, 공표되어 작성되거나, 정보통신망을 통해 게시되는 정보등을 수집하는 경우에는 개인이나 법인 기타 단체 등을 특정하거나 식별할 수 없는 형태로 통계 및 현황 등을 작성하는 경우에 한하여 정보를 이용할 수 있다.

② 행정기관 등이 민간정보를 수집하는 경우, 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보등을 수집할 수 없으며, 통계 및 현황 등의 형태로도 이를 이용할 수 없다.

제2절 추진체계

제81조(정보자원의 총괄관리) 국가정보자원의 공유·활용에 관한 사무는 행정안전부장관이 총괄하며, 행정기관 간 정보교류와 협력에 관한 계획을 수립하여 「국가정보화기본법」 제6조에 따른 기본계획에 반영하여야 한다.

제82조(정보자원의 활용기반) ① 행정안전부장관은 행정기관 등의 정보자원 공유·활용 및 분석등을 위한 공동이용센터를 구축·운영할 수 있다.

② 행정안전부 장관은 공동이용센터를 통해 정보자원의 전송, 전달, 가공, 이용, 분석체계 지원 등 공유·활용에 필요한 시책을 마련하여 추진하도록 할 수 있다.

③ 행정기관 등은 정보자원을 공유·활용하고 하는 경우에는 특별한 사유가 없는 한 공동이용센터를 통하여 공유·활용하여야 한다.

제83조(공유·활용절차) ① 행정기관 등이 정책수립 및 의사결정 등을 위하여 다른 행정기관이 수집·보유하고 있는 정보등을 공유·활용하고자 하는 경우, 공동이용센터를 통하여 신청할 수 있다.

② 공동이용센터는 대상정보를 수집·보유하고 있는 행정기관 등에 이를 통보하여야 하며 신청을 받은 행정기관 등은 특별한 사정이 없는 한 이에 응하여야 한다.

③ 개인정보, 비밀 또는 비공개 사항으로 규정된 정보를 이용하고자 하는 행정기관 등은 행정안전부 장관의 심의를 거쳐, 추진정책의 목적 및 내용, 활용대상정보의 범위와 내용 등에 대하여 대통령령으로 근거를 마련하여야 한다.

제84조(공유·활용의 책임과 권한) ① 제공기관은 필요한 대상정보를 생성하거나, 불필요한 정보 등을 분리하는 등 가공하여 제공할 수 있다.

② 제공기관은 제1항에 따라 정보가공 등에 필요한 비용에 대하여 신청기관이 부담하도록 할 수 있다.

③ 공동이용센터는 정보의 전달, 제공 및 분석 등을 위하여 불가피한 경우, 대상정보를 일시적으로 저장할 수 있으며, 업무목적 달성한 경우에는 이를 폐기 또는 환송하여야 한다. 다만, 소재정보, 색인, 신청기관 및 제공기관 등에 대한 정보를 작성하여 보관할 수 있다.

제3절 정보의 보호 및 안전

제85조(개인정보활용의 원칙과 기준) ① 행정기관 등의 장은 공공정보 등에 개인정보가 포함되어 있는 경우, 개인정보를 분리하여 이용하거나 특정개인을 식별할 수 없는 형태로 통합·활용할 수 있다.

② 공공정보의 통합·활용에 있어 불가피하게 개인정보를 활용할 필요가 있는 경우, 정보주체에게 정당한 이익을 부여하거나 권리를 보장하는 것이 명백한 경우에 한하여, 개인정보를 통합·활용할 수 있다. 다만, 의무를 부과하거나 권익을 제한할 우려가 있는 경우에는 그러하지 아니하다.

③ 제2항에 따라 개인정보가 포함된 공공정보를 통합·활용하고자 하는 경우, 해당 행정기관의 장은 개인정보의 통합범위, 활용목적, 정보보호방안, 권리보장방안 등을 마련하여, 개인정보보호위원회의 심의·의결을 거쳐야 하며, 그 내용과 결과를 관보에 공표하여야 한다.

④ 2개 이상의 행정기관에서 공공정보를 통합·활용하여 50만명 이상의 정보주체에 관한 개인정보가 포함되는 개인정보파일을 연계하는 경우에는 개인정보 보호법 제33조에 따른 영향평가를 실시하여야 한다.

⑤ 개인정보의 통합·활용에 대한 기준과 절차, 보호위원회의 심의의결 방법 등 기타 필요한 사항은 대통령령으로 정한다.

제86조(저작물인 공공정보의 공정한 이용) 행정기관 등이 다음 각호 중 어느 하나의 사무를 추진하기 위하여 저작물인 공공정보를 활용하는 경우, 저작권법 제35조제3항에 따라 저작물을 공정하게 이용한 것으로 본다.

1. 민원사무처리에 관한 법률 제14조에 따라 복합민원으로 처리될 필요가 있는 사무
2. 행정업무의 효율적 운영에 관한 규정 제44조에 따라 융합행정으로 발굴하여 수행되는 사무
3. 행정업무의 효율화와 대국민서비스의 개선 등 공공부문의 행정 혁신을 촉진할 수 있는 사무
4. 기타 행정기관 등이 공익적인 목적을 위하여 저작권자의 정당한 이익을 해하지 않는 범위 내에서 비영리적으로 이용할 수 있는 사무

제87조(정보의 안정성 보장) 공동이용센터는 행정기관 간 정보 공유·활용시 분실, 도난, 유출, 변조 또는 훼손되지 아니하도록 필요한 조치를 시행하여야 하며, 공유·활용하려는 정보에 개인정보가 포함되어 있는 경우에는 개인정보보호법 제24조 제3항에 따라 암호화 등 안전성 확보에 필요한 조치를 하여야 한다.

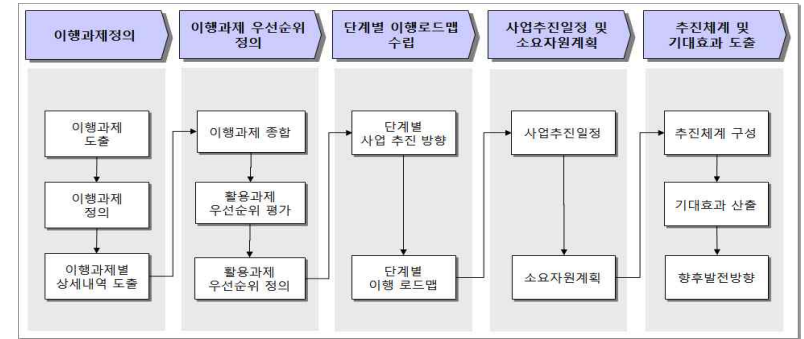
다. 이행계획 수립

1) 이행계획 수립 개요

가) 개요

미래모형 단계에서 도출된 세부개선과제를 차질 없이 이행하기 위한 과제를 도출하고, 이행과제를 실행하기 위한 소요예산 및 추진일정 그리고 추진체계와 이행과제를 달성함으로써 기대되는 효과를 정립함

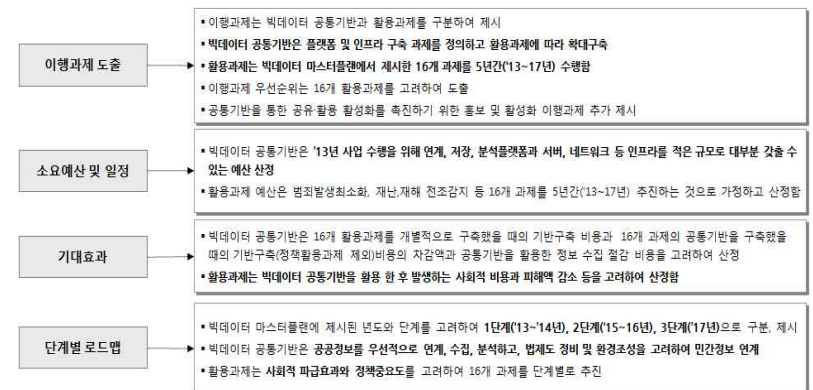
[이행계획 수립 절차]



나) 이행계획 수립 전략

이행계획수립은 과제도출, 예산, 기대효과 측면에서 다음과 같은 전략을 토대로 수립함

[이행계획 수립 전략 내용]



2) 이행과제 정의

미래모형 정립 단계에서 제시된 추진과제 및 세부 내역을 바탕으로 다음과 같은 이행과제를 도출함

| 이행과제 | 상세 내역 | 설명 |
|-------------------|-------------------------|--|
| 빅데이터 연계·수집 플랫폼 구축 | ESB-Hub구축 | <ul style="list-style-type: none"> 빅데이터 공통기반으로 데이터를 연계·수집하기 위하여 ESB-Hub를 도입 도입은 2단계로 나누어 추진하며, 단계별로 ESB-Hub 2식을 도입 |
| | 망연계 솔루션 (1식) | <ul style="list-style-type: none"> 빅데이터 공통기반의 내부망과 외부망간의 데이터 연계를 위한 망연계 솔루션을 도입 |
| | 개별기관 데이터 연계·수집(20개 대상) | <ul style="list-style-type: none"> 활용채널(과제)에서 요구되는 외부데이터를 연계·수집하기 위하여 연계환경(ESB Adaptor 설치)을 구성 |
| | 스트리밍데이터 처리 인프라 구성 | <ul style="list-style-type: none"> 수집되는 데이터가 스트리밍 데이터일 경우 해당 데이터를 선택적으로 저장 및 처리하기 위한 솔루션을 도입 |
| | 웹데이터 직접수집체계 구성 | <ul style="list-style-type: none"> 웹데이터를 직접수집하기 위한 인프라 구성을 위하여 솔루션을 도입 |
| 빅데이터 구성 플랫폼 구축 | RDB저장 소프트웨어 | <ul style="list-style-type: none"> RDB데이터를 저장하기 위한 소프트웨어를 도입 오픈소스 기반의 RDBMS도입을 원칙으로 하되, 특정데이터(GIS 등)의 특성으로 고려하여 상용DBMS도 구성 |
| | NoSQL및 Object(File) 저장소 | <ul style="list-style-type: none"> NoSQL및 Object(File)데이터를 저장하기 위하여 구성되는 인프라로, 1개의 Name노드와 18개의 Data노드로 구성된 세트(Set)을 년차별로 구분하여 도입(해당 인프라에 RDB저장 소프트웨어도 탑재됨) |
| 빅데이터 분석 플랫폼 구축 | Hadoop분석 기반 구축 | <ul style="list-style-type: none"> 연계/수집된 정형·비정형 데이터에 대해 그룹핑, 정제 등 전처리 과정을 통해 배치(Batch)로 데이터를 분석하기 위한 Hadoop기반 분석 시스템 구축 |
| | Real-time 분석 기반 구축 | <ul style="list-style-type: none"> 스트리밍(Streaming) 컴퓨팅 기술 적용을 통해 빅데이터를 수집과 동시에 분석을 수행할 수 있도록 실시간 분석 시스템 구축 |
| | 사용자 주도 분석 기반 구축 | <ul style="list-style-type: none"> 사용자들이 자신들만의 시각으로 데이터를 분석 할 수 있도록 분석 대상 데이터에 대한 선택/제거 만으로 실시간 연관 검색 및 사용자 주도형 분석을 수행할 수 있도록 분석 시스템 구축 |

| 이행과제 | 상세 내역 | 설명 |
|-----------------------|----------------------------|--|
| 빅데이터 공통기반 포털 구축 | 내부 활용기관 서비스 포털 구축 | <ul style="list-style-type: none"> 빅데이터 공통기반을 활용하여 정책정보 지원을 받는 활용기관을 대상으로 단순 정보 제공뿐만 아니라 공통기반 서비스 환경 구축을 위한 지원 서비스 제공하는 포털 구축 |
| | 대외 개방 서비스 포털 구축 | <ul style="list-style-type: none"> 연계·수집·구성된 범국가적 데이터 중 민간 기업, 일반인등에게 개방이 가능한 데이터를 기준으로 검색 및 정보 요청, 커뮤니케이션 서비스 등을 제공하는 포털 구축 |
| 클러스터 및 활용채널 관리시스템 구축 | | <ul style="list-style-type: none"> 클러스터 및 활용채널을 관리하기 위한 응용어플리케이션 구축 - 클러스터 및 활용채널 정보, 정책지원 데이터 Meta를 관리하기 위한 어플리케이션 구축 |
| 빅데이터 공통기반 인프라 구축 | 하드웨어 도입 및 설치 | <ul style="list-style-type: none"> 빅데이터 공통기반 구축을 위한 서버, 스토리지 도입 및 설치 |
| | 소프트웨어 도입 및 설치 | <ul style="list-style-type: none"> 서버, 스토리지 등을 운영 및 관리하기 위한 소프트웨어 도입 및 설치 |
| | 네트워크 도입 및 설치 | <ul style="list-style-type: none"> 네트워크를 구성하기 위한 라우터, 스위치, 방화벽, IPS 등의 도입 및 설치 |
| 빅데이터 공통기반 홍보 및 활성화 | | <ul style="list-style-type: none"> 빅데이터 공통기반의 필요성과 서비스의 장점을 홍보하고, 보유기관과 활용기관의 공통기반 참여 및 활용을 증대시키기 위한 캠페인을 기획하고 실행 |
| 활용과제별 빅데이터 공통기반 활용 구축 | 범죄발생 장소·시간 예측을 통한 범죄발생 최소화 | <ul style="list-style-type: none"> 참조정보의 확대와 분석 모델 개선을 통한 예측정밀도를 향상하고 실시간 범죄징후 감지를 통한 범죄 사전대응 체계 지원 구축 |
| | 재난·재해 전조감지를 통한 재난피해 최소화 | <ul style="list-style-type: none"> 재난을 조기에 감지하고 재해 예측 정확성 개선 및 사전 대응체계 지원 구축 |
| | 음란물 유통차단 | <ul style="list-style-type: none"> 인터넷 상의 유해 콘텐츠 종합·분석 및 등급을 분류하고, 음란물 관련 신고 데이터 분석을 통해 음란물 유통 사전 방지체계 지원 구축 |
| | 주민참여형 교통사고 감소체계 구축 | <ul style="list-style-type: none"> 트위터, 민원 게시판의 국민의견 분석, 교통사고 데이터 분석을 통해 경찰배치, 신호등·과속방지턱 설치 등 개선방안 수립을 위한 분석 지원 구축 |
| | 국가 미래전략 및 위기대응 전략 | <ul style="list-style-type: none"> 정부 보유 데이터, 국토 자원 및 사회기반시설의 사물센싱 데이터 등 다양한 정보를 활용하여 국가 |

| 이행과제 | 상세 내역 | 설명 |
|------|----------------------------|---|
| | 수립 | 미래전략 및 위기대응 전략을 수립할 수 있는 체계 지원 구축 |
| | 민원데이터 분석을 통한 정책의 환류시스템 마련 | • 온라인, 콜센터 등을 통해 축적된 민원 정보를 분석하여 국민 수요 적시 도출 및 대응할 수 있도록민원에 대한 범정부 국민소통·정책 환류 체계 지원 구축 |
| | 복지 수요·공급 매칭 맞춤형 서비스 | • 민원정보를 통한 복지 키워드 추출, 상담센터 수집 정보 등을 통한 복지 현황을 분석하여 복지 사각지대를 제거하기 위한 맞춤형 서비스 지원 구축 |
| | 일자리 현황분석·예측 | • 일자리 경력 및 수요·공급 관리와 관련된 분석모델 구축을 통해 교육훈련 프로그램 개발, 자격증 관리 등 고용정책 수립 체계 지원 구축 |
| | 자영업자 창업 실패 예방 지원 | • 업종별, 지역별, 연령별 상관데이터, 대출, 임차료, 권리금 등의 데이터를 통합 분석, 자영업 성공·실패 사례, 정부의 지원사업 정보 등의 분석 지원 구축 |
| | 실시간 네트워크 시스템 재난 관리 대응체계 마련 | • 네트워크 사용량, 장비 동작정보등의 오작동 사전예측 및 네트워크 장애차단 체계 확보 및 실시간 통신망·통신장비 모니터링을 통한 효율적인 통신망 운영 및 장애 최소화 지원 구축 |
| | 제조공정 장애예측으로 생산효율 고도화 | • 제조장비·공정에 소요되는 부품별 상태 정보 수집·분석, 중장비 시설이나 첨단 제품 생산 설비 운영에 있어 발생하는 데이터에 대한 수집·분석·예측 지원 구축 |
| | 과세 데이터분석으로 탈세 방지 | • 납세자 및 과세대상 현황, 국제납부 이력 등 정형데이터와 소비·지출 관련 비정형데이터 결합 분석, 사기성 납세 신고, 부당 환급자 등 분석 지원 구축 |
| | 데이터 분석기반 경제정책 수립 | • 정량적 지표 뿐 아니라 정성적 지표를 활용한 중장기 경제 환경 변화 전망, 지속적인 경제성장 동력 분석, 경제위기 징후 감지 등 지원 구축 |
| | 국가 기후 위험요소에 선제적 대응 | • 국내 관련 데이터와 전지구 자료교환허브를 연계·분석하여 가뭄, 식량안보, 에너지 등 국가 위험요소에 대한 선제적 대응책 마련 |
| | 위성 영상 데이터 활용으로 기후 예측 | • 위성영상 데이터와 기상, 재난 등 관련 데이터의 연계 처리·분석지원 구축 |
| | 의료데이터 분석을 | • DNA 데이터, 진료기록, 질병정보 등을 융합 분석하여 |

| 이행과제 | 상세 내역 | 설명 |
|-----------|----------------------|--|
| | 통한 국민 건강증진 | 건강 정보 진단 및 질병 발생 가능성 예측과암, 당뇨, 비만, 고혈압 등 100여개 질환의 유전 위험도 예측 지원 구축 |
| 법제도 정비 방안 | 빅데이터 활용을 위한 전자정부법 개정 | <ul style="list-style-type: none"> • 빅데이터 활용을 위한 전자정부법 개정 - 전자정부법 상에 새로운 장을 신설하여, 추진근거 마련 - 대규모 정보자원의 공유활용을 위한 추진체계 설정 - 행정기관간 정보공유활용 절차적 정비 - 활용대상정보의 범위와 유형 등을 명시 - 제공기관과 활용기관 간의 책임과 권한 등을 명시 |

3) 이행과제 우선순위 정의

가) 개요

빅데이터 공통기반 구축 부문은 연도별 연속 사업으로 일괄 추진되며, 활용과제별 빅데이터 공통기반 활용 구축 부문은 연도별로 활용과제를 구분하여 추진해야 하므로 활용과제 우선 순위를 평가하여 이행과제 로드맵에 반영함

[이행과제 우선순위]



나) 우선순위 평가 방법

정책활용과제 우선순위 평가는 평가기준에 따른 평가결과를 바탕으로 결정하며, 이 결과는 향후 단계별 이행 로드맵 작성의 기초 자료로 활용함

| 평가항목 | 개요 | 평가방법 |
|------|-------------|-----------------------|
| 시급성 | 국정 목표를 달성하기 | 빅데이터 주요 과제 추진 단계 및 차이 |

| | | |
|-----|--|---|
| | 위해 시급하게 실행해야 할 과제 | 정부 정책에 기초하여 정책 활용 과제의 시급성을 10점 척도로 평가함 - 10점 : 우선 추진 과제로 선정된 것 - 1점 : 빅데이터 주요 과제 중 시급성이 낮고 차기 정부 정책과 연관성도 낮은 과제 |
| 활용성 | 정책 클러스터의 정보를 활용할 가능성이 많다고 예상되므로 활용 기반 구축이 시급한 정도 | 빅데이터 주요 과제 및 차기정부 정책 공약과 연관성이 많아 정보 사용이 높을 것으로 예상되므로 시급성이 큰 과제에 대해 10점 척도로 평가함 |

다) 우선순위 평가 결과

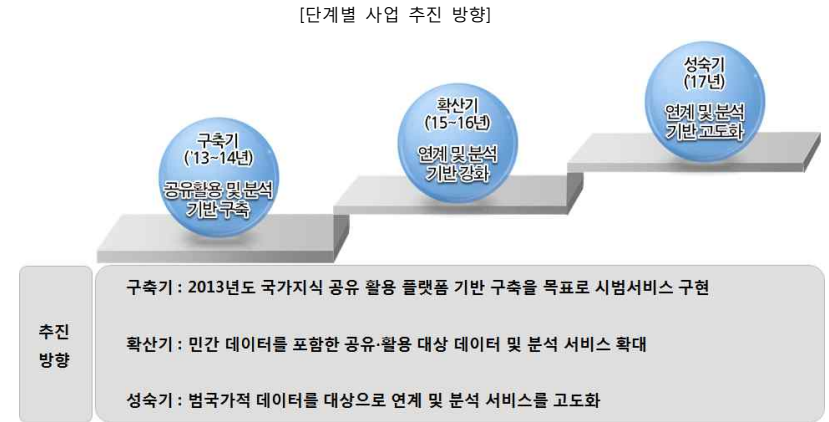
정책활용과제 우선순위 평가결과 1단계 최우선 과제는 범죄, 재난, 교통, 민원정책 환류로 분석되었음

[정책활용과제 우선순위 평가 결과]



4) 단계별 이행 로드맵

빅데이터 공통기반구축을 바탕으로 구축기에는 국가지식 공유 활용 플랫폼 기반 구축을 목표로 시범서비스 구현, 확산기에는 민간 데이터를 포함한 공유·활용 대상 데이터 및 분석 서비스 확대, 성숙기에는 범국가적 데이터를 대상으로 연계 및 분석 서비스를 고도화 방향으로 사업을 추진함



5) 소요자원 계획

○ 빅데이터 공통기반 마련 및 활용 과제 구축 총 소요예산은 5년간('13~'17년) 451억임

[이행과제별 소요예산]

| 이행과제 명 | 연도별 소요자원 [단위 : 백만원] | | | | | |
|------------------------|---------------------|-------|-------|--------|-------|--------|
| | 합계 | 2013년 | 2014년 | 2015년 | 2016년 | 2017년 |
| 빅데이터 연계·수집 플랫폼 구축 | 2,867 | 838 | 594 | 214 | 706 | 516 |
| 빅데이터 저장 플랫폼 구축 | 6,202 | 765 | 993 | 2,296 | 1,531 | 1,531 |
| 빅데이터 분석 플랫폼 구축 | 3,818 | 351 | 558 | 1,911 | 558 | 558 |
| 빅데이터 공통기반 포털 구축 | 532 | 532 | - | - | 511 | - |
| 클러스터, 활용채널 관리시스템 구축 | 196 | 196 | - | - | - | - |
| 공통인프라 구축(HW, SW, 보안 등) | 10,387 | 1,816 | 5,167 | 2,442 | 424 | 556 |
| 공통기반 홍보 및 활성화 | 97 | - | 97 | - | - | - |
| 활용과제별 공통기반 활용 구축 | 15,997 | 502 | 2,502 | 3,996 | 4,005 | 4,992 |
| 빅데이터 공통기반 유지보수 비용 | 3,499 | - | - | 449 | 1,191 | 1,877 |
| 합계 | 45,180 | 5,002 | 9,911 | 11,309 | 8,926 | 1,0031 |

6) 추진체계

가) 사업추진 체계

빅데이터 공통기반 마련 및 활용체계 구축 사업의 사업추진 체계 구성과 역할을 다음과 같이 제시함

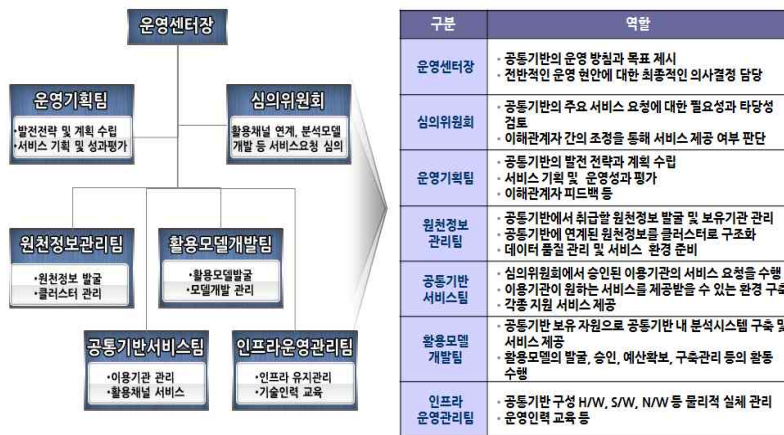
[사업 추진 체계 구성 및 역할]



나) 운영조직 체계

빅데이터 공통기반 업무설계에서 제시된 공통기반 목표 모델은 단순히 정책정보를 수집하여 제공하는 정보 서비스 모델이 아니고, 서비스를 창출하는 기반구조를 서비스 해 주는 모델임

[운영조직 체계 구성 및 역할]



7) 기대효과

가) 정성적 효과

○ 빅데이터 공통기반 마련 및 활용 과제 시스템 구축을 통한 정성적 효과는 국가, 공무원, 국민 측면에서 아래와 같이 기술함

[정성적 효과]

| 서비스 대상 | 기대효과 | 상세 내용 |
|--------|-----------------|---|
| 국가 | 민첩한 정부로 변모 | • 정부 3.0의 기본 취지인 개방, 공유, 활용에 부응하는 국민 맞춤형 정책 실현 및 공공혁신 주도 • 범죄, 질병, 재난, 금융위기 등 국민안전과 직결되는 위협에 신속한 대응 가능한 정부로 전환 |
| | 정책정보공유, 활용기반 구축 | • 기관간 정보 연계, 촉진 및 공유활용 환경조성 |
| 공무원 | 정책정보 품질 향상 | • Text, 통계, 수치 등 다양한 분석을 통해 정책보고서 가치 제고 |
| | 업무 효율성 향상 | • 다양한 정책정보 공유활용을 통한 정책 수립 업무 생산성 향상 |
| 국민 | 국민 안전성 강화 | • 범죄, 재해 발생 피해로부터 국민의 사전 대응력 강화 지원 • 안정적이고 건전한 사호 환경 구현을 통한 국민 복지 증진 기여 |

나) 정량적 효과

○ 빅데이터 공통기반 마련 및 활용 과제 시스템 구축을 통해 기대되는 정량적 효과는 '18년 1,443억원이며, 투자수익율(ROI)은 177%임

