

디지털플랫폼정부를 위한 기반기술(B·A·N·D) 이용 활성화 방안

제9호 2022.12.28.

목차

- Ⅰ 디지털플랫폼정부 기반기술(B·A·N·D) · 8
- Ⅱ 주요 기반 기술(B·A·N·D) 활용방안 · 12
- Ⅲ 주요 기반기술(B·A·N·D) 구현 방향 · 28



IT & Future Strategy(IF Strategy) 보고서는

21세기 한국사회의 주요 패러다임 변화를 분석하고 이를 토대로 미래 지능화 시대의 주요 이슈를 전망, IT를 통한 해결방안을 모색하기 위해 한국지능정보사회진흥원(NIA)에서 기획, 발간하는 보고서입니다.

IF Strategy는

미래의 ‘만약을 대비한 전략’을 담은 보고서를 의미합니다.

NIA의 승인 없이 본 보고서의 무단전재나 복제를 금하며, 인용하실 때는 반드시 NIA, 「IT & Future Strategy 보고서」라고 밝혀주시기 바랍니다.

보고서 내용에 대한 문의나 제안은 아래 연락처로 해 주시기 바랍니다.

발행인

황 증 성

작 성

한국지능정보사회진흥원(NIA) 정책본부 AI·미래전략센터
윤창희 수석연구원(053-230-1292, yunch@nia.or.kr)

보고서 온라인 서비스

www.nia.or.kr

◇ 디지털플랫폼정부 개념 및 필요성

- (정의) 모든 데이터가 연결되는 디지털플랫폼 위에서 국민, 기업, 정부가 함께 사회문제를 해결하고 새로운 가치를 창출하는 정부
 - 민간 주도로 수요가 높은 데이터의 플랫폼 보유 및 기술 제공을 통한 새로운 가치를 발굴하는 구도의 디지털플랫폼 비즈니스 장 마련
 - 디지털플랫폼은 비즈니스를 활성화하기 위해 3가지 구성요소를 중심으로 '네트워크 효과'를 창출
 - ※ 3요소 : 플랫폼 제공자, 플랫폼 참여자, 사용자(고객)으로 구성
- (필요성) 디지털 전환의 성공 여부가 향후 국가경쟁력의 핵심으로 부상
 - 범정부 기관 간 연계·활용되는 데이터 기반의 과학적 국정 실현을 위해서 다기관 연계 디지털플랫폼 기반으로의 환경 전환 필요
- 디지털플랫폼정부 기반기술(B·A·N·D)

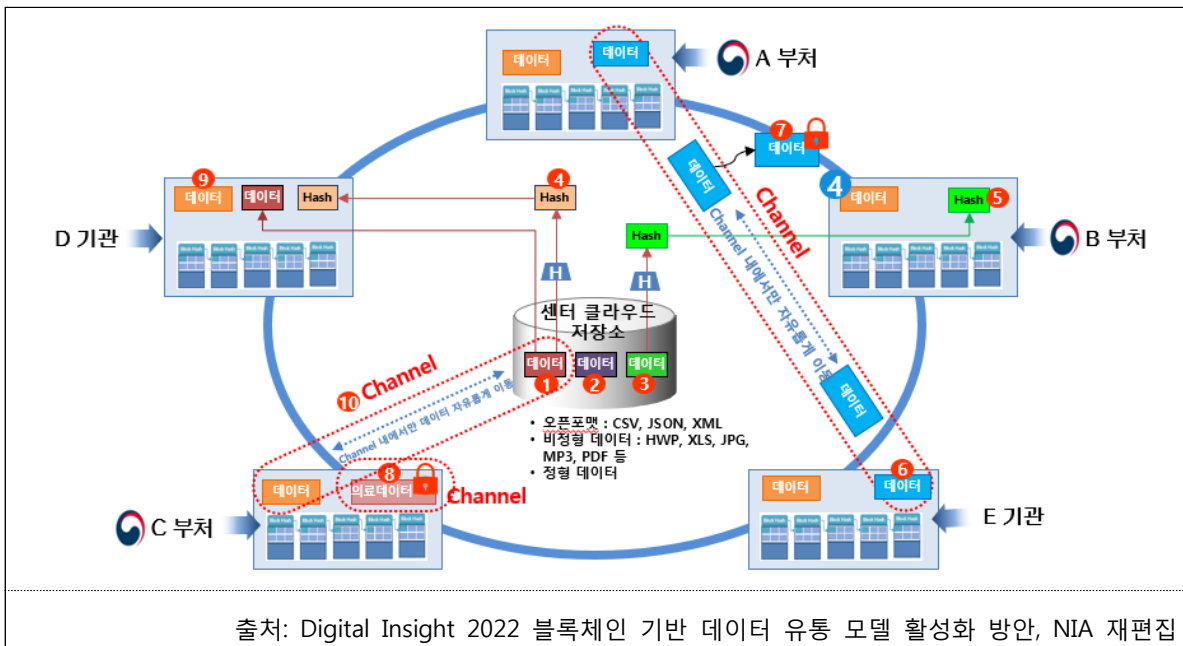
B·A·N·D 기술	구성요소
Blockchain (B)	<ul style="list-style-type: none"> • 탈중앙화 유통시스템 • Channel/메타데이터/Hash 암호기술 • 위변조 방지/투명성·익명성·기밀성·비가역성·무결성 보장 • 전자서명/이력추적
AI (A)	<ul style="list-style-type: none"> • DevOps 환경구축 • AI 플랫폼 구성요소(데이터, 컴퓨팅, 알고리즘) • 빅데이터 통합 플랫폼/인공지능 플랫폼 • AI와 Blockchain 연계 서비스
Network (N)	<ul style="list-style-type: none"> • 다부처/다기관 연계 • Open API 지원 • 클라우드 네이티브 컴퓨팅 • 클라우드 연계(public/ private/ hybrid, 정부 클라우드) • 다중/edge 클라우드 지원, MEC 컴퓨팅 • AI 프레임워크 지원/컨테이너(도커 및 쿠버네티스) 기반
Data (D)	<ul style="list-style-type: none"> • 다부처/다기관 데이터 • 정형/비정형 빅데이터 • 데이터 저장방식(data meshes/ pipelines/ ponds/ lakes) • 모든 데이터를 연결하기 위한 기술 • AI 학습 데이터/빅데이터 통합 플랫폼(수집/정제/가공)

◇ 디지털플랫폼 기반 기술(B·A·N·D) 분석

[블록체인(B)]

- 데이터 패러다임 변화에 따른 탈중앙화 유통·관리 시스템으로 전환
 - 최근 데이터가 성장과 경제를 견인하는 ‘데이터 경제’를 주도하면서 데이터의 안전한 저장·관리에서 거래·유통의 활성화로 전환 중
- 디지털플랫폼 정부의 블록체인 기반 공공 데이터 유통 모델 구축
 - (하이브리드) 다기관을 대규모 공공데이터 가운데 보안데이터는 블록체인 기술을 병행 적용하는 하이브리드 데이터 유통체계 필요

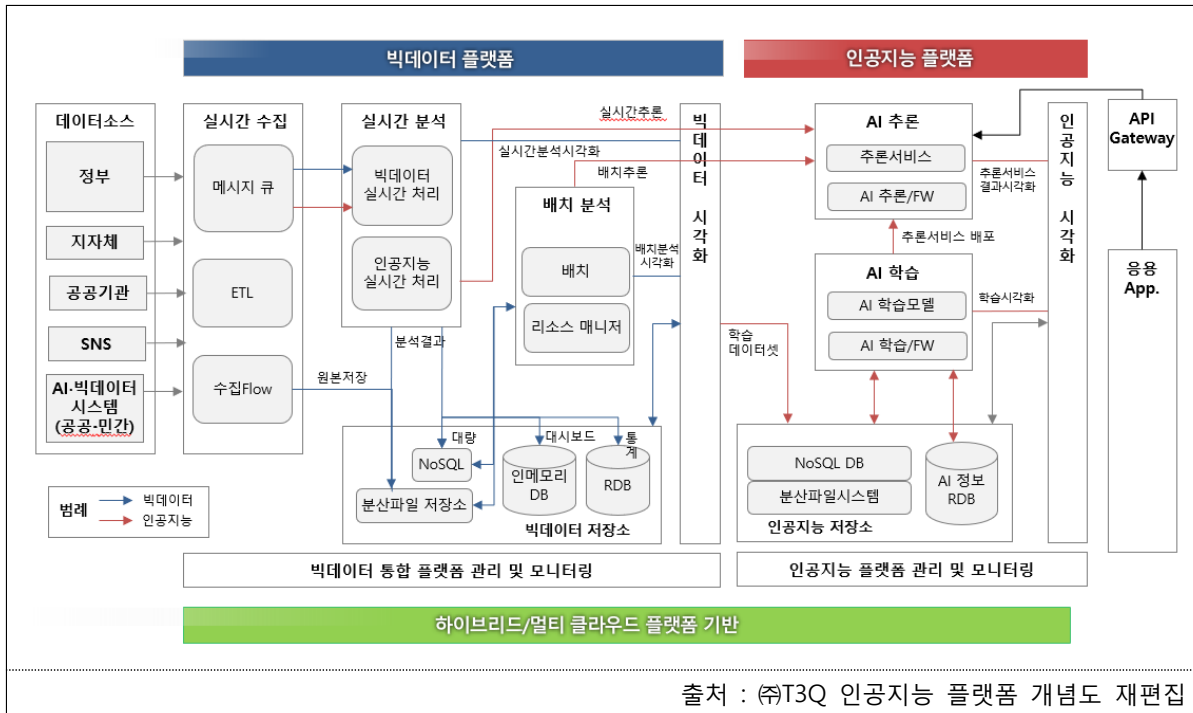
<다기관의 다양한 데이터 특성을 고려한 블록체인 기반 유통 모델(안)>



[인공지능(A)]

- 다기관 수집 데이터로부터 대규모 훈련을 수행하기 위한 클라우드 기반의 AI 플랫폼 구축
- 인공지능 서비스 구현을 위한 주요 구성요소
 - (구성요소) 인공지능 플랫폼은 수요-공급자 간 인공지능 서비스, 정보 교류의 활성화 공간으로 핵심요소는 데이터, 알고리즘, 컴퓨팅 자원

<빅데이터·인공지능 활용·연계 통합 플랫폼(안)>



[네트워크(N)]

- 다기관 대량의 데이터를 손쉽게 접근·활용할 수 있는 클라우드 구축
 - 대규모 공공데이터를 센터 클라우드와 전문영역별 엣지 및 공공기관 클라우드 영역으로 나누어 분야별 특화된 데이터 네트워크 체계 구축
 - 고도화된 소프트웨어(SDI) 기반의 센터 클라우드를 중심으로 전문 분야별 개방형 데이터 제공 및 AI 융복합 서비스 창출
 - 다기관으로의 확장성을 위해 특정 벤더에 의존하지 않는 오픈소스 기반의 클라우드-네이티브 컴퓨팅 환경의 하이브리드·멀티 클라우드 구축
- 개방형 개발모델 지향(Open API)
 - 민·관·학·연 간 상호 연계를 통한 디지털플랫폼정부 서비스 개발 고도화 및 시너지 창출을 위해 직접 응용 프로그램과 서비스를 개발할 수 있도록 공개된 API 제공
 - Git-Hub, paperswithcode와 같이 누구나 참여하여 서비스 개발·협력할 수 있는 상호 '네트워크 효과'를 창출할 수 있는 모델 구축

[데이터(D)]

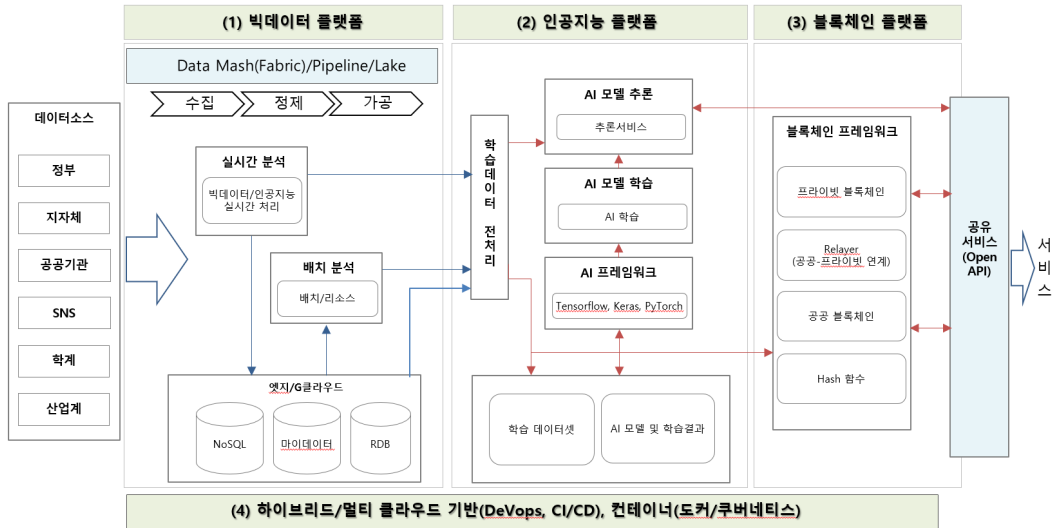
- 전방위적인 디지털 전환에 대비한 개방형 국가 데이터 연계 체계 구축
 - 디지털플랫폼정부뿐만 아니라 다양한 산업 분야들에 인공지능 기반의 융복합 서비스를 창출하기 위해서는 민·관 데이터 연계 필요
 - 누구에게나 개방되어 공정하게 접근 가능하고, 제공-중계-사용자 간에 데이터 활용을 유연할 수 있는 플랫폼 조성
- 다양한 형태의 데이터 간 유통 및 품질관리 방안 마련
 - (일반 데이터) 공공영역의 기관별로 보유하고 있는 정형·비정형·반정형 데이터 등 다양한 데이터 간의 원활한 유통체계 마련이 중요
 - 이를 위해 공공영역의 비정형 데이터의 유통표준 및 메타데이터, 품질관리 방안 마련 중요
 - (AI 학습데이터) 지도학습을 위해 정답값(Ground Truth)에 대한 라벨링 작업이 필요
- 다양한 분야의 기구축 데이터 저장·관리 체계 마련
 - (Mashes) AI 통합플랫폼과 같이 축적된 데이터를 사용자 관점에서 융합·연결하는 Data Mesh 구축
 - (Pipelines) 다양한 기관에서 수집한 데이터를 분석하기 위해 Data lake 혹은 Data Warehouse같은 저장소로 이동시켜 관리
 - (Lakes) 데이터 연못(Ponds), 호수(Lakes)와 같은 공간에 저장된 방대한 데이터와 항상 연결하여 정제·분석 가능한 도구 지원
- 마이데이터
 - 개인정보 활용은 개인의 동의에 기반한 기존 이종 데이터의 결합으로 새로운 서비스 창출 가능

◇ 주요 기반기술(B·A·N·D) 구현방향

○ B·A·N·D 기반의 통합플랫폼 구축방향

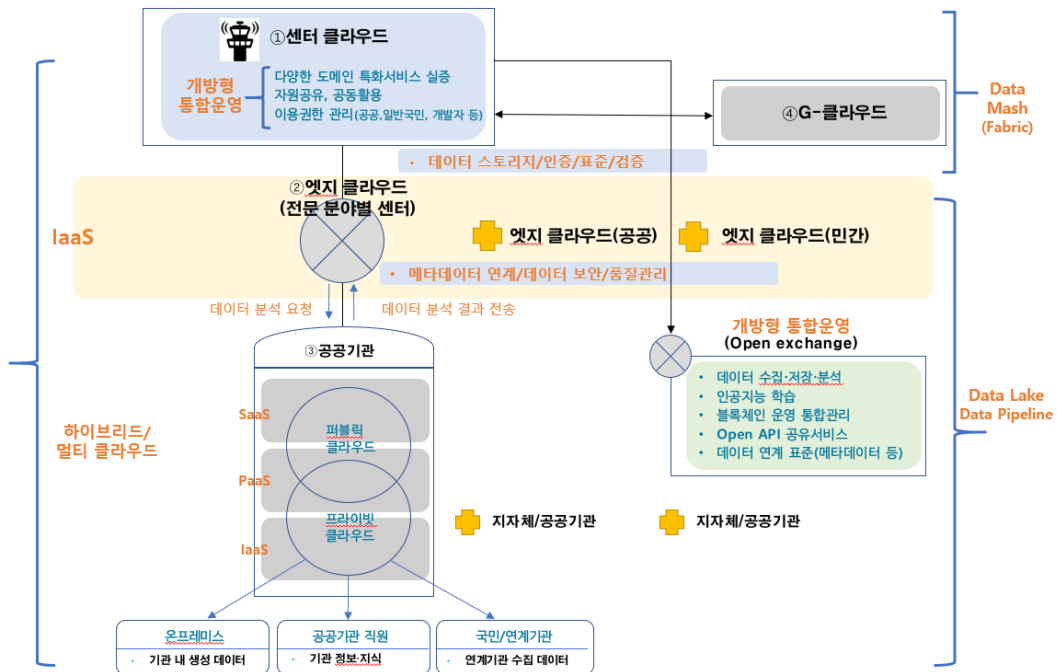
- 빅데이터·인공지능 플랫폼, 블록체인 기술 적용 및 Open API 서비스로 확장성·유연성을 높이는 통합 플랫폼 구축

<B·A·N·D 기반 통합플랫폼(안)>



- (클라우드 기반) 공공부문의 다기관 시스템 간의 원활한 연계를 위한 구성 방안 마련

<공공부문 연계를 위한 계층적 클라우드 구성(안)>



I

디지털플랫폼정부 기반기술(B·A·N·D)

1 디지털플랫폼정부 개념

□ 디지털플랫폼정부 정의

- 모든 데이터가 연결되는 디지털플랫폼 위에서 국민, 기업, 정부가 함께 사회문제를 해결하고 새로운 가치를 창출하는 정부[1]
 - 정부가 일방적인 공급자로서 디지털 서비스를 제공하는 방식에서 벗어나 민간과 협업하여 새로운 공공서비스를 창출하는 모델
- 클라우드 기반의 빅데이터를 수집·정제·가공하여 AI 모델 생성 후 수요-공급자 간 상호 지속적인 서비스 개발이 가능한 선순환 환경 조성
 - 필요시 민간 주도로 수요가 높은 데이터의 플랫폼 보유 및 기술 제공을 통한 새로운 가치를 발굴하는 구도의 디지털플랫폼 비즈니스 장 마련
- 디지털플랫폼은 비즈니스를 활성화하기 위해 3가지 구성요소를 중심으로 '네트워크 효과'¹⁾를 창출

※ 디지털플랫폼 정의 : ICT 관점에서 컴퓨터 운영체제나 어떤 소프트웨어가 제공하는 공간 위에서 상호 간 서비스, 정보 등을 교류할 수 있는 장[2]

- ① 플랫폼 제공자 : 플랫폼 참여자들에게 활동하는 핵심자원(클라우드 인프라, 인공지능 도구, 데이터 등)을 제공[3]
- ② 플랫폼 참여자 : 제품 또는 서비스를 플랫폼 생태계 상에 제공(정부, 기업, 학교, 개인 등)하는 제3의 참여자로 플랫폼에 다양한 가치 부여[4]
- ③ 사용자(고객) : 플랫폼 상에 거래되는 서비스·제품을 구매하고 피드백하는 일련의 과정을 통해 새로운 데이터를 생산하며 네트워크 효과를 창출[5]

1) 네트워크 효과란 특정 상품에 대한 수요가 다른 사람들의 수요에 의해 영향을 받는 것으로, 상품, 서비스의 수요가 주위에 서로 영향을 미쳐 점차 그 상품을 선택하는 사람들이 증가하는 효과(위키백과)

2 디지털플랫폼정부 필요성

□ 디지털 전환의 성공 여부가 향후 국가경쟁력의 핵심으로 부상

- 초연결 사회 도래에 따라 디지털 전환의 성공여부가 국가 명운을 결정할 변곡점이라 인식하고 세계 각국은 디지털 총력전 전개 양상
- 범정부 기관 간 연계·활용되는 데이터 기반의 과학적 국정 실현을 위해서 다기관 연계 디지털플랫폼 기반으로의 환경 전환 필요
 - (협업) 코로나19, 글로벌 공급망 위기 등 국내·외적인 위기·재난 발생 시 정부 개별 부서·분야의 산발적 대응에는 한계
 - (기술) AI, 빅데이터, 클라우드, IoT 등 신기술이 급속히 발전하고 모든 데이터와 사물이 플랫폼으로 연결되는 초연결 사회 본격 도래
 - (사회) AI 기반 신기술을 통한 행정효율 및 국민편익 증진을 위한 정부의 노력이 요구되며 신기술이 사회 모든 분야의 기반으로 확대
 - (정부) 다양한 기관에서 수집된 데이터 분석을 통한 위험탐지 및 정부 운영의 효율성을 높일 수 있는 인공지능 플랫폼 환경 마련 필요

□ 서비스가 양방향에서 플랫폼 기반의 상호작용 구조로 전환

- 과거 정부 서비스 형태는 수요-공급자 간에 플랫폼이 아닌 양방향 형태의 전통적인 가치사슬 모델이었으나[6]
- 디지털플랫폼정부에서 민·관이 데이터, 알고리즘, 컴퓨팅 자원을 공유 하면서 구성원 간에 상호 보완적 관계를 형성하며 새로운 가치 창출 필요

3 디지털플랫폼정부를 위한 기반기술(B·A·N·D)

□ B·A·N·D : 블록체인(B), 인공지능(A), 네트워크(N), 데이터(D)

○ 디지털플랫폼정부 서비스 구현을 위해서는 다기관 연계가 필요하고 기술적으로 복잡·다단한 구조를 가지고 있어 주요 기반기술을 B·A·N·D로 정리

- 대규모 범정부 빅데이터(D)와 인공지능(A)이 결합한 새로운 서비스 개발을 위해 클라우드(N) 기반으로 데이터의 저장·연계 필요

※ 네트워크(N)는 다양한 기관의 대규모 데이터 활용이 비교적 쉽고 무한한 확장성이 강점이 클라우드 기술을 컨테이너 기반으로 정리

- 공공영역의 다양한 데이터를 안전하게 유통시켜 맞춤형 서비스를 선제적으로 제공하기 위하여 블록체인(B) 기반 데이터 관리 중요

<디지털플랫폼정부를 위한 B·A·N·D 요소기술>

B·A·N·D 기술	구성요소
Blockchain (B)	<ul style="list-style-type: none"> 탈중앙화 유통시스템 Channel/메타데이터 Hash 암호기술 위변조 방지/투명성·익명성·기밀성·비가역성·무결성 보장 전자서명/이력추적
AI (A)	<ul style="list-style-type: none"> DevOps 환경구축 AI 플랫폼 구성요소(데이터, 컴퓨팅, 알고리즘) 빅데이터 통합 플랫폼 인공지능 플랫폼 AI와 Blockchain 연계 서비스
Network (N)	<ul style="list-style-type: none"> 다부처/다기관 연계 Open API 지원 클라우드 네이티브 컴퓨팅 클라우드 연계(public/ private/ hybrid, 정부 클라우드) 다중/edge 클라우드 지원, multi-access edge computing AI 프레임워크 지원 및 컨테이너(도커 및 쿠버네티스) 기반
Data (D)	<ul style="list-style-type: none"> 다부처/다기관 데이터 정형/비정형 빅데이터 데이터 저장방식(data meshes/ pipelines/ ponds/ lakes) 모든 데이터를 연결하기 위한 기술 AI 학습 데이터 및 빅데이터 통합 플랫폼(수집/정제/가공)

○ **블록체인(B)**

- 민·관·학 기관들의 노드화를 통해 블록체인 네트워크에 참여하는 방식의 데이터 통합 모델 구현

※ 통합된 데이터를 모두 블록체인에 저장하는 것은 현실적으로 어렵기 때문에 공공 목적의 일부 데이터만 블록체인에 저장하여 활용하는 하이브리드 형태로 운영

○ **인공지능(A)**

- 민·관·학 분야의 다양한 학습데이터, 초거대 AI 플랫폼을 활용한 딥러닝 기반의 AI 모델 개발, 훈련 및 추론 과정을 통해 서비스 제공
- 인공지능의 자연어 처리, STT, TTS, 비전, 이미지·동영상 객체 식별 등 다양한 기술의 다기관 간 연계·적용을 통한 새로운 서비스 개발

※ 서비스 유연성 확보를 위해 Open API를 이용한 AI 융합 서비스(X+AI) 개념으로 기능 추가

○ **네트워크(N)**

- (클라우드) 대량의 데이터를 손쉽게 접근·활용할 수 있는 클라우드-네이티브 컴퓨팅 환경의 하이브리드·멀티 클라우드 환경 구축

※ 인공지능 모델 생성·갱신·훈련 및 추론 등이 효율적으로 운영될 수 있도록 코어-엣지-종단 간 다양한 분야에 특화된 융합형 클라우드 인프라 구축

- (인프라) 컨테이너(도커 및 쿠버네티스) 기반으로 클라우드를 구축하여 향후 디지털플랫폼 정부 시스템과의 호환성 제고

※ 클라우드 연계(public/private/hybrid, G클라우드), edge, GPU 지원 등 IaaS 인프라 조성

○ **데이터(D)**

- 기존 산업 데이터와 결합, 전방위적인 디지털 전환에 대비할 수 있도록 국가 차원에서 민·관·학 데이터 연계를 통한 플랫폼 구축

- 다부처·다기관의 정형·비정형 빅데이터를 활용하기 때문에 데이터 저장방식(data meshes/piplines/ponds/lakes) 및 AI 학습데이터와 같은 데이터 가공 방향 수립 필요

II

주요 기반 기술(B·A·N·D) 활용방안

1 블록체인 (B)

□ 데이터 패러다임 변화에 따른 탈중앙화 유통·관리 시스템으로 전환

- 최근 데이터가 성장과 경제를 견인하는 ‘데이터 경제’를 주도하면서 데이터의 안전한 저장·관리에서 거래·유통의 활성화로 전환됨에 따라
 - 데이터를 중앙에서 소유·통제하는 환경에서 개인으로 권한을 강화하여 새로운 가치를 창출하는 패러다임으로 변화 중
- 블록체인은 탈중앙화 기술로 해시(Hash) 암호기술을 기반으로 네트워크에 데이터를 투명하게 기록하고 다수의 컴퓨터에 복제·저장하는 구조로
- 정부·지자체·공공기관 등 다기관의 민감 데이터 유통이 필요한 디지털 플랫폼 정부 서비스의 데이터 특성에 맞는 세부 설계 과정을 통한 통합 모델 구축 가능

□ 디지털플랫폼정부의 블록체인 기반 공공 데이터 유통 모델 구축

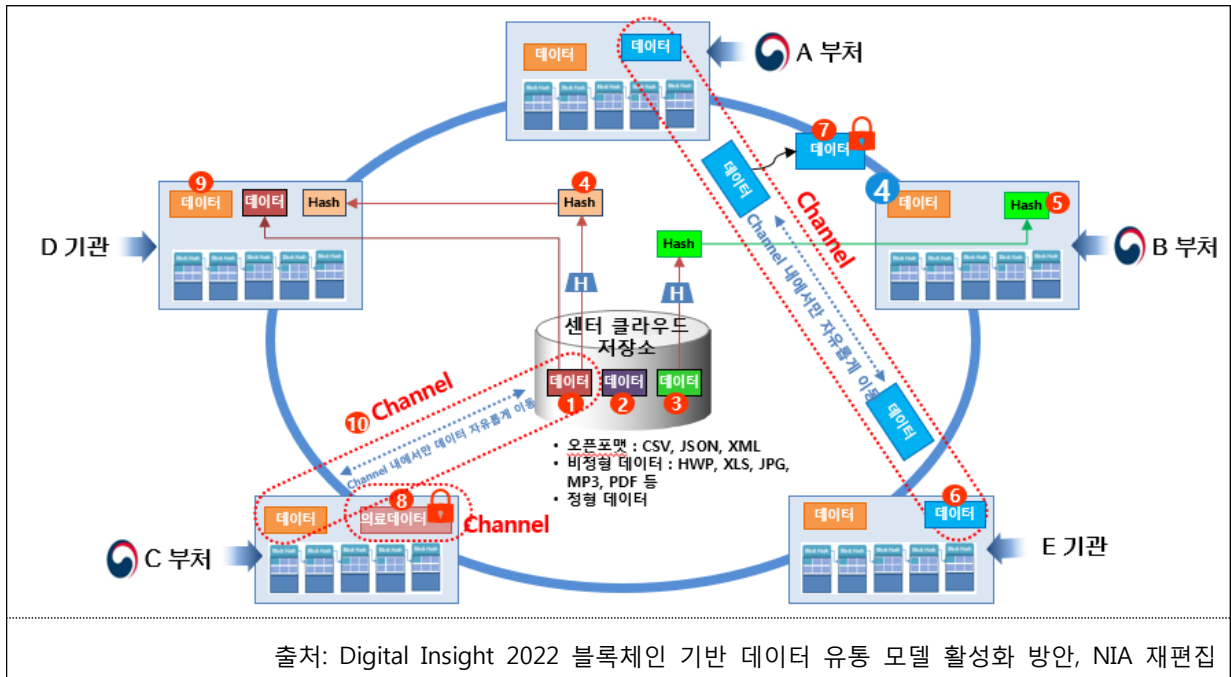
- (하이브리드) 다기관의 대규모 공공데이터 가운데 보안데이터는 블록체인 기술을 병행 적용하는 하이브리드 데이터 유통체계 구축
 - 데이터 용량·처리속도, 통제·기밀성 등의 이슈로 대부분의 데이터를 블록체인에 저장하는 것은 현실적으로 곤란
 - 따라서 인증이 필요한 일부 보안데이터만 제한적으로 블록체인에 저장하여 운영하는 하이브리드 형태로 플랫폼을 구축·운영하는 모델 구축
- ※ 대민 서비스 가운데 실시간 관계기관 공유·투명성 및 무결성이 요구되는 데이터는 블록체인에 직접 저장하여 활용(데이터 처리속도, 용량에 따라 유연하게 설계)

<하이브리드 블록체인 데이터 저장 체계>

저장소	주요내용
블록체인	<ul style="list-style-type: none"> 기밀성은 필요 없지만 투명성과 무결성이 필요한 주요 데이터에 한해 블록체인에 저장
데이터베이스	<ul style="list-style-type: none"> 기밀성이 필요하지만 데이터의 수정, 삭제가 요구되는 데이터를 블록체인 대신 기존 데이터베이스에 저장
해시값 블록체인 저장	<ul style="list-style-type: none"> 기밀성 및 데이터 수정, 삭제뿐만 아니라 무결성도 요구되는 경우, 데이터는 원본 데이터베이스에 저장하고 데이터의 해시값만 블록체인에 저장하여 무결성 검증

- 데이터 특성에 맞추어 디지털플랫폼정부 관계기관에서 저장·생성되는 데이터를 센터-엣지-기관별로 저장하고 필요시 연계
 - 공공영역에서 적용 가능한 데이터 유통 모델을 센터 클라우드로의 통합 및 특정 부처·기관 간의 데이터 채널 구축, 데이터-해시값 노드 저장 방안 고려
 - 데이터 저장소와 블록체인 적용이 필요한 데이터 특성에 따라 저장소와의 연계 여부를 결정하여 세부 설계

<다기관의 데이터 특성을 고려한 블록체인 기반 유통 모델(안)>



- ①, ②, ③ 데이터 : 모든 데이터는 기본적으로 센터 클라우드 통합하여 저장
- ④ Hash : ①번 데이터는 별도로 D 기관 노드에 저장되며 동시에 Hash값도 저장됨 (무결성 및 기밀성 필요)
- ⑤ Hash : ③번 데이터의 Hash값만 B 부처 노드에 저장됨(무결성 필요)
- ⑥ 데이터 : 엣지-기관 또는 센터-엣지 클라우드 노드 간 채널을 통해 두 노드에게만 공개되는 데이터(기밀성 필요)
- ⑦ 데이터 : ⑥번 데이터는 A 부처와 E 기관 외에 다른 노드에게는 공개되지 않음
- ⑧ 의료데이터 : ⑧번 데이터는 C 부처에 한정하여 저장(단독채널)
- ⑨ 데이터 : ⑨번 데이터는 참여 노드 모두에게 투명하게 공개
- ⑩ 채널 : 특정 기관에 한해서 공공 데이터의 기밀성이 요구되면 특정 기관 간에만 유통될 수 있도록 블록체인 Channel 운영

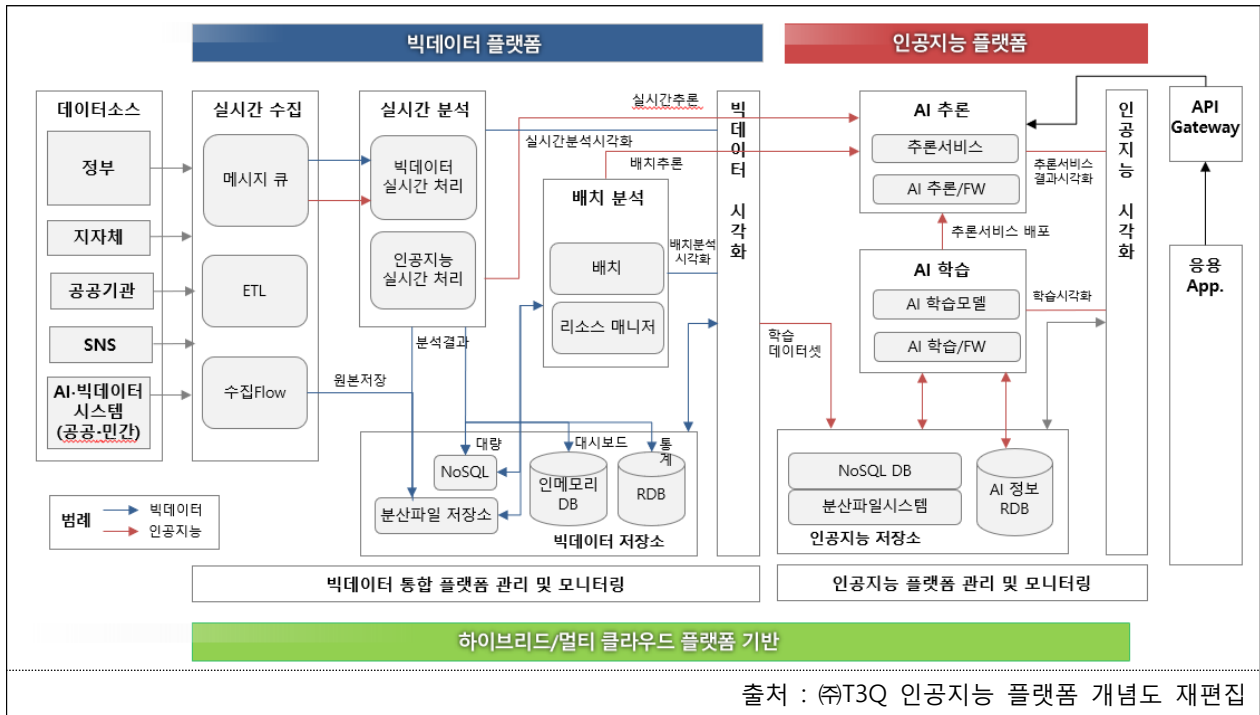
2 인공지능 (AI)

□ 인공지능 융합 서비스 환경 구축

- 대기관 수집 데이터로부터 대규모 훈련을 수행하기 위한 클라우드 기반의 AI 플랫폼 구축
 - 인공지능 서비스는 대규모 데이터 기반의 AI 모델 생성·갱신·훈련 및 추론 등의 과정을 효율적으로 연계시킬 수 있는 체계 중요
 - ※ 디지털플랫폼정부의 인공지능 서비스 : 정부·민간·지자체·공공기관·연구기관 등 대기관의 데이터를 활용하고, AI를 접목해 그 과정에서 다양한 서비스를 창출하는 개념
- 인공지능은 원천 데이터를 수집·가공하여 모델에 적용·구현 후 서비스 개발을 통해 최종 가공 데이터를 제공하는 공급자 역할 수행
 - 데이터 품질관리를 위한 유효성 검증, 가공, 클린징 등의 작업을 자동화 툴을 병행 사용하여 인공지능 데이터 활용도 제고
 - ※ 데이터 품질관리, 데이터 가공, 데이터 클린징(Cleansing) 등 데이터를 활용한 인공지능 서비스 개발 단계에서 활용하는 AI 기술 지원
- 인공지능 서비스 구현을 위한 주요 구성요소
 - (구성요소) 인공지능 플랫폼은 수요-공급자 간 인공지능 서비스, 정보 교류의 활성화 공간으로 핵심요소는 데이터, 알고리즘, 컴퓨팅 자원[7]
 - ※ 한곳에서 데이터 로딩, 모델훈련, 검증(파일럿 서비스)까지 가능한 원스톱 파이프라인을 구축하여 머신러닝, 딥러닝 개발을 위한 자동화된 툴을 통합 제공하도록 구성 필요
- ① (데이터) 다부처·대기관의 정형·비정형 빅데이터를 활용하는 만큼 데이터 저장방식(data meshes/piplines/ponds/lakes) 및 AI 학습 데이터와 같은 데이터 가공 방향 수립 필요
 - AI 학습데이터, 공공데이터를 Migration하여 다양한 분석이 가능한 서비스로 연계

- 신뢰도 높은 AI 모델을 위해 데이터 유효성과 편향성 검증, 데이터 보호기능 지원
- ② (컴퓨팅) 데이터 저장·학습을 수행하는 코어 영역, 학습·추론하는 엣지 클라우드, 최종 사용자로 구분하여 플랫폼 서비스 제공
 - 컨테이너(도커 및 쿠버네티스) 기반으로 클라우드를 구축하여 향후 디지털플랫폼 정부 시스템의 확장성 제고
 - ※ 인공지능 개발에 필요한 GPU, TPU, FPGA 컴퓨팅 자원 및 Tensorflow, Pytorch 프레임워크 등 환경 제공
 - 향후 확장성 있는 플랫폼으로 진화하기 위해 오픈소스 기반 클라우드 연계(public/private/hybrid, G클라우드), 다중/edge 클라우드 지원, GPU 기반 환경 등 IaaS 인프라 조성 전략 필요
- ③ (알고리즘) 인공지능 기술 고도화 및 지속 가능한 관리·운영이 수월한 AI 알고리즘 개발환경 조성[8]
 - 개발자들의 유연한 알고리즘 개발 플랫폼 제공을 위한 DevOps 환경 구축
 - AI/ML 기반 이미지·영상분석, 음성인식·변환, NLP 등 다양한 AI 기술개발이 가능한 알고리즘 등
 - AI/ML 적용 가능한 데이터 제공·활용·분석 및 DW, NoSQL 등 다양한 데이터 분석이 가능한 자동화 도구 등
 - ※ AI/ML 개발을 위해 워크플로우(데이터 전처리→빌드→학습&튜닝→적용&관리) 전과정 지원
- 상기 인공지능 특성을 고려한 융합 서비스 구축을 위해 빅데이터 기반의 인공지능 통합플랫폼 개념(안) 제시

<빅데이터·인공지능 활용·연계 통합 플랫폼(안)>



※ 대용량의 데이터 처리 및 지능형 분석 기술 적용을 위한 실시간 처리 기반 통합플랫폼·인공지능 플랫폼

빅데이터 플랫폼

- 데이터 소스 기관으로부터 수집한 데이터를 수집, 분석, 저장 및 시각화를 구현하여 유용한 정보를 통합적으로 관리
- 데이터 현황 분석(건수·응답·트래픽·오류 등), 장애 및 이상 징후 탐지 등 작동 구간의 모든 시스템에 대한 각각의 정보를 실시간 수집, 분석하여 디지털플랫폼 통합 모니터링 제공
- 사용자 주도 분석 영역을 구분하여 AI 플랫폼의 딥러닝 학습, 예측 작업 제공

※ 사용자 책임 범위에 대한 데이터의 기술·제도적 세부 규정을 수립하여 데이터 접근 권한 관리 강화

인공지능 플랫폼

- 인간의 학습능력, 추론능력, 지각능력, 자연어 이해능력 등을 구현하여 디지털플랫폼 서비스 목적에 적합한 이상 징후 예측 자동화 및 기법 개발 지원

- AI 플랫폼에서 제공하는 분석 통합 개발 및 운영 환경은 데이터 생성, 모델설계, 모델학습, 모델테스트, 시각화, 빅데이터 연계 처리, 추론에 적용이 가능하고
- 이러한 일련의 AI 플랫폼 운영 프로세스는 향후 디지털플랫폼 정부 서비스 및 관계기관 간 연계를 통한 서비스 자동화 및 기법에 활용

하이브리드/멀티 클라우드 플랫폼

- 빅데이터·인공지능 플랫폼을 클라우드 기반으로 구성하여 긴밀하게 연계·작동될 수 있도록 통합적으로 개발하여 관리
 - 실시간으로 수집·분석한 데이터를 NoSQL, 분산파일 저장소 및 인 메모리 DB, RDB 등 빅데이터, 인공지능 저장소에 저장하되
 - 개인정보 등 민감데이터는 IaaS 인프라 및 블록체인 기술을 적용하여 분산 관리하는 하이브리드/멀티클라우드 방식 적용
- ※ 향후 블록체인 기술 및 공유 서비스(Open API)와의 연동을 위해 다양한 기관들의 플랫폼과 연계 가능한 유연성·확장성이 강점인 클라우드 레이어 구성 필요

3 네트워크 (N)

- 다기관 대량의 데이터를 손쉽게 접근·활용할 수 있는 클라우드 구축
 - 대규모 공공데이터를 센터 클라우드와 전문영역별 엣지 및 공공기관 클라우드 영역으로 나누어 분야별 특화된 데이터 네트워크 체계 구축
 - 대규모 데이터를 수집·분석하여 AI 훈련·추론을 주도하는 센터와 메타 데이터 연계, 품질관리를 지원하는 엣지 클라우드 및 공공기관으로 구분
 - ※ 고도화된 소프트웨어(SDI)²⁾ 기반의 개발형 통합운명을 지향하는 센터 클라우드를 중심으로 전문 분야별 개방형 데이터 제공 및 AI 융복합 서비스 창출
 - 센터-엣지-기관 간 클라우드의 원활한 협력에 따른 데이터 중심 인 프라 측면에서 구조화
 - 대규모 공공기관이 보유하고 있는 방대한 데이터를 Data Mashs, Pipelines, Lakes 체계로 나누어 데이터 전달·교환·분석 등을 수행할 수 있는 체계 마련[데이터(D)] 참조
 - 다기관으로의 확장성을 위해 특정 벤더에 의존하지 않는 오픈소스 기반의 클라우드-네이티브* 컴퓨팅 환경의 하이브리드·멀티 클라우드 구축
 - 대규모 데이터 기반의 인공지능 모델 생성·갱신·훈련 및 추론 등의 과정을 효율적으로 연계시킬 수 있는 체계
 - * 가상머신에서 벗어나 가볍고 유연한 컨테이너를 중심으로 전환하여 활용하는 클라우드 네이티브 컴퓨팅 패러다임으로 확산 추세[9]

2) SDI(Software-Defined Infra) : IoT(Internet of things)와 클라우드가 서로 연동되도록 5G 통신 및 소프트웨어-정의 네트워킹(Software-Defined Networking: SDN)과 네트워크-기능-가상화(Network Function Virtualization: NFV) 기술이 지원하는 유연성있는 프로그램 구현이 가능한 소프트웨어-정의 인프라를 의미함

<클라우드 네이티브 서비스 개요>

구분	세부내용[10]
정의	<ul style="list-style-type: none"> 클라우드의 장점을 최대한으로 활용할 수 있도록 애플리케이션을 구축하고 실행하는 방식으로 퍼블릭, 프라이빗, 하이브리드 클라우드 환경에서 확장성 있게 서비스 개발 및 운영 가능 컨테이너, 서비스 메시, 마이크로서비스, 불변의 인프라스트럭처, 선언적 API 등의 접근 방식이 있고, 회복성, 관리 편의성, 가시성 추구하는 서비스
4대요소	<ol style="list-style-type: none"> ① DevOps : 애플리케이션 개발-운영 간의 협업 프로세스를 자동화하는 것으로 결과적으로 애플리케이션의 개발과 개선 속도를 향상시킴 ② CI/CD : CI(Continuous Integration)은 개발자가 작업한 코드를 자동으로 테스트하고 통과 후 코드를 통합하여 저장. CD(Continuous Deployment)는 작업코드, 변경사항을 테스트 후 리포지토리에 업로드되고 서비스 배포, 반출까지 자동화 ③ MicroService : 애플리케이션을 구성하는 서비스들을 독립적인 작은 단위로 분해하여 구축하고 각 구성 요소들을 네트워크로 통신하는 아키텍처로 서비스 안정성과 확장성(scaling)을 지원 ④ 컨테이너 기반 인프라 : 가상화 기술로, 시스템 가상화가 아닌 애플리케이션을 구동할 수 있는 컴퓨팅 작업을 패키징하여 가상화함

□ 개방형 개발모델 지향(Open API)

- 민·관·학·연 간 상호 연계를 통한 디지털플랫폼 정부 서비스 개발 고도화 및 시너지 창출을 위해 직접 응용 프로그램과 서비스를 개발할 수 있도록 공개된 API 제공

- SOAP*, REST**와 같이 정부 공유서비스 개발, 인증 가이드[11] 등에 따른 방식과 병행하여 다양한 서비스 개발 유연성 확대

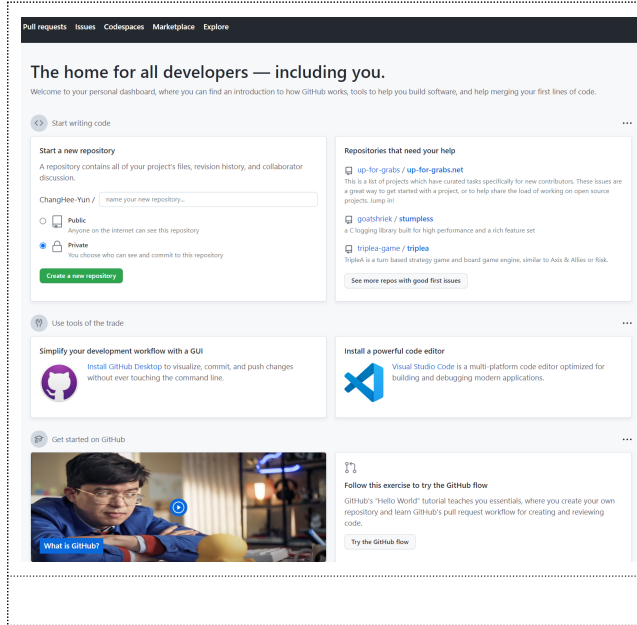
* SOAP : 서비스명, 서비스 URL, 서비스오퍼레이션 정의

** REST : 서비스명, 서비스 URI로 공유서비스 정의

- Git-Hub, paperswithcode와 같이 누구나 참여하여 서비스 개발·협력할 수 있는 상호 '네트워크 효과'를 창출할 수 있는 모델 구축 필요

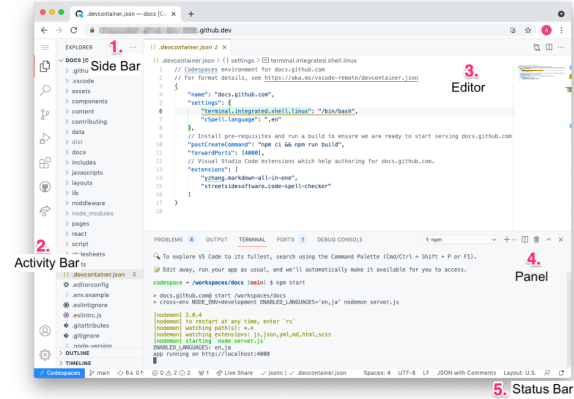
※ 현재 민간·공공 분야 각각(Silo) 다양한 형태로 데이터, 플랫폼 환경을 구축하여 다양한 요소 기술 연계가 필요한 디지털플랫폼 서비스 개발에 한계로 작용할 수 있음

<GitHub 상호협력 개발 환경(예시)>



Working in a codespace in the browser

Using Codespaces in the browser provides you with a fully featured development experience. You can edit code, debug, use Git commands, and run your application.



출처 : GitHub(<https://docs.github.com/en/codespaces>)

4 데이터 (D)

□ 전방위적인 디지털 전환에 대비한 개방형 국가 데이터 연계 체계 구축

- 디지털플랫폼 정부뿐만 아니라 다양한 산업 분야들에 인공지능 기반의 융복합 서비스를 창출하기 위해서는 민·관 데이터 연계 필요
 - 누구에게나 개방되어 공정하게 접근 가능하고, 제공-중계-사용자 간에 데이터 활용을 유연할 수 있는 플랫폼 조성
- ※ 코로나19, 글로벌 공급망 위기 등 동시다발적인 국가 위기·재난에 효율적으로 대응하기 위해서도 국가 차원의 데이터 허브 필요

□ 다양한 형태의 데이터 간 유통 및 품질관리 방안 마련

- (일반 데이터) 공공영역의 기관별로 보유하고 있는 정형·비정형·반정형 데이터 등 다양한 데이터 간의 원활한 유통체계 마련이 중요
 - 민·관 데이터 유형을 보면 일부 데이터를 제외하고 대부분 텍스트, 문서, 이미지, 동영상, 음성 등의 비정형 데이터 다수
 - ※ 전세계 데이터의 90%가 최근 2년간 생성되었고 그중에 80%는 비정형 데이터이며 정형데이터의 2배 속도로 증가되고 있는 추세(IBM연구소)
 - 특히 급속히 증가하고 유의미한 내용이 많은 비정형 데이터를 우선 개방·연계하고 관리할 수 있는 방안 마련 필요
- 이를 위해 공공영역의 비정형 데이터의 유통표준 및 메타데이터, 품질관리 방안 마련 중요[12]
 - 기관별로 다양한 형태로 구축한 데이터의 유통을 위한 데이터 표준화 작업 선행
 - 메타데이터는 비정형데이터에 직·간접적 연관 정보를 제공할 수 있어 다기관 데이터의 관계, 검색, 관리 등에 용이

- 메타데이터 품질관리를 위해 품질관리 수준 평가 지표개발 등 고품질 유지관리 세부 방안 수립

※ 메타데이터 유형은 기술메타, 설명메타, 관리메타 등 10개 유형으로 선정 가능하며, 관리항목도 Dublin Core, MPEG-7, DCAT, Schema.org 등 국내외 선진사례를 분석하여 활용 가능[13]

- (AI 학습데이터) 지도학습을 위해 정답값(Ground Truth)에 대한 라벨링 작업이 필요

- 인공지능 모델에 따라 필요한 정답에 맞도록 적절한 데이터 라벨링 과정이 필요

- 정부·공공·지자체·학계 등 대민 서비스를 중심으로 정형·비정형 데이터를 선정하여 AI 서비스에 활용할 수 있도록 구축

※ 과기정통부는 NIA(AIHub.or.kr) 허브를 통해 AI 기술, 제품, 서비스 발전 및 지능형 제품·서비스, 챗봇 등 다양한 분야에서 영리적·비영리적 연구·개발 목적으로 구축[14]

□ 다양한 분야의 기구축 데이터 저장·관리 체계 마련

- 그간 구축한 정부·학계·민간 데이터 등을 체계적으로 통합·관리할 수 있는 플랫폼 구축

- 센터 클라우드를 중심으로 지속성·보안성·접근성을 높일 수 있는 데이터 통합환경 지원

※ 데이터 특성에 따라 단계적으로 중앙-엣지-사용자 계층에 따라 클라우드 구성을 구분하여 설계

- (Mashes) AI 통합플랫폼과 같이 축적된 데이터를 사용자 관점에서 융합·연결하는 Data Mesh 구축[15]

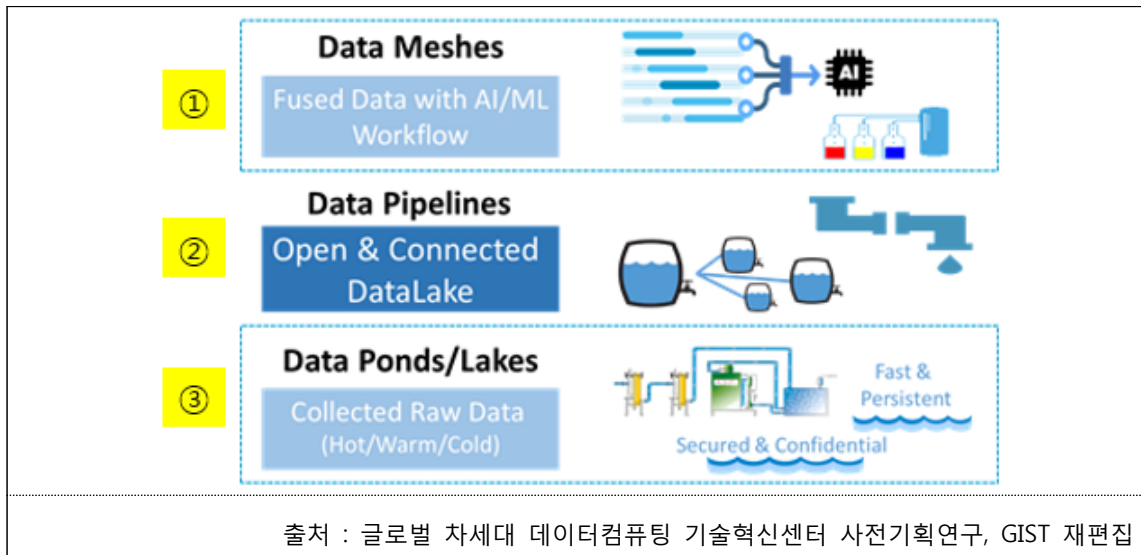
※ 향후 딥러닝, 추론 기능을 지원하는 인공지능 기술을 지원할 수 있는 기능 추가

- (Pipelines) 다양한 기관에서 수집한 데이터를 분석하기 위해 Data lake 혹은 Data Warehouse같은 저장소로 이동시켜 관리[16]

- (Lakes) 데이터 연못(Ponds), 호수(Lakes)와 같은 공간에 저장된 방대한 데이터와 항상 연결하여 정제·분석 가능한 도구 지원[17]

- 기관-엣지-센터 클라우드 간 Mash, Pipeline-Lake 기능이 원활히 작동할 수 있도록 플랫폼 레이어 간 연계표준, 데이터 표준체계 수립

<데이터 중심 통합 데이터 센터 (예시)>[18]



- ① Data Meshes(Fabric) : 복잡한 대규모 조직의 데이터플랫폼 구현을 위한 아키텍처로서, 탈중앙화로 분산화하여 데이터 처리[붙임1] 참조
 - 단일 플랫폼의 중앙집중 데이터 처리방식은 정확하게 어디에 무슨 데이터가 있는지 알기 어렵고 병목현상으로 많은 시간 소요

※ 클라우드, 온프레미스 데이터 검색을 위한 메타데이터 및 AI에 중점을 두면서, 저장 장소와 관계 없이 데이터의 접근성 및 확장 가능성을 높이는 Data Fabric 개념[19]을 포함

- ▣ 대규모 다기관 연계가 필요한 디지털플랫폼 환경에서 전문 영역별 분산된 도메인 간 데이터가 원활히 처리될 수 있도록 세부 설계 필요

- ② Data Pipelines : 데이터 파이프라인 아키텍처로부터 원시데이터에 대한 ETL* 기능 수행[붙임2] 참조

* E-T-L 개념 : (Extract) 데이터 소스로부터 원시데이터를 추출, (Transform) 데이터를 특정 형식으로 변환, (Load) 시스템에 데이터를 로딩하는 것

- 데이터 소스, 목적지, 데이터 흐름 저장소, 처리, 워크플로우, 모니터링 7가지 기능을 바탕으로 기관 간 데이터 유통 관리

- ▶ 기관-엣지-센터 간 하나의 파이프라인에서 구성요소 간 기능이 작동할 수 있도록 세부 설계 필요

③ Data Lakes

- 확장 가능하고 안전한 플랫폼을 제공하며 온프레미스, 클라우드, 엣지 등을 가리지 않고 모든 시스템의 데이터를 속도제한 없이 수집 가능
- 대규모의 다양한 데이터를 고품질로 저장하고, 실시간 또는 일괄 처리로 데이터를 처리·분석할 수 있는 중앙집중식 저장소[20]

※ 데이터 레이크는 SQL, Python, R 등의 언어와 타사 데이터 또는 분석 애플리케이션을 사용하여 분석하고 필요시 데이터 정제작업을 수행[21]

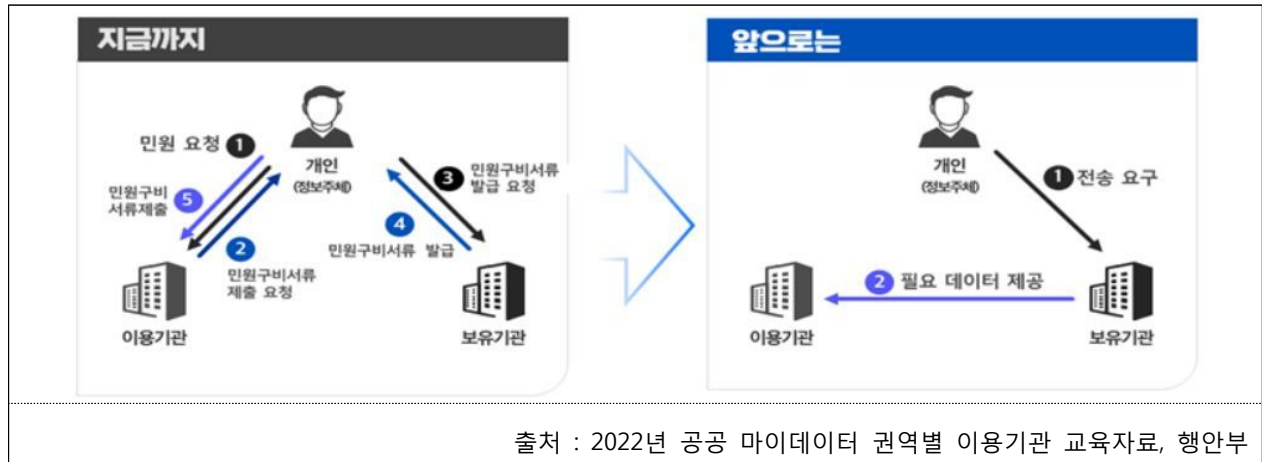
- ▶ 기관-엣지-센터 저장소 간에 Connected Data 저장소의 개념으로 논리적 중앙집중 저장소 구현 필요

□ 마이데이터

- 마이데이터를 통해 이를 신뢰하고 안전하게 제공할 수 있는 제도적 환경이 조성되고 있는 상황
 - 데이터 3법, 민원처리법, 전자정부법 등의 시행으로 안전한 데이터 거래·활용 기반이 조성되고, 데이터 주권 이슈 대응 중[22]
 - 행안부에서는 공공 마이데이터*를 도입하여 국가, 기업이 보유하고 있는 국민 본인에 관한 정보를 적극적, 능동적으로 관리하고 활용할 수 있는 새로운 데이터 활용 체계 마련하여 운영[23]

* 공공 마이데이터 서비스 : 공공-행정기관에서 보유하고 있는 개인 행정정보를 정보주체인 국민이 본인 또는 본인이 지정한 제3자에게 제공하도록 요구할 수 있는 제도

<공공 마이데이터 활용 개선 방향>



- 마이데이터를 활용한 디지털플랫폼 정부 서비스는 과거와 다른 혁신 서비스를 구현할 수 있는 촉매제 역할을 할 것으로 전망
 - 개인의 동의에 기반한 개인정보 활용으로 기존 이종 데이터의 결합으로 새로운 서비스 창출 가능
 - 예를들면, 1-2-3차 병원 간 이동 진료시 의료기록 공유로 일관된 병력 추적으로 보다 정확한 진료 가능 및 자료 이관 불편 해소

참고1 **데이터 유형(정형/비정형/반정형)[24]**

<p>○ 정형 데이터</p>	<p>미리 정해놓은 형식과 스키마(테이블, 컬럼) 구조에 따라 데이터를 검색, 수집, 활용 등이 가능</p> <p style="text-align: center;"><정형데이터 구조 및 예시></p> <div style="display: flex; justify-content: space-around; align-items: center;"> <table border="1" style="border-collapse: collapse; text-align: center;"> <caption>스키마에 의해 정의된 컬럼</caption> <thead> <tr> <th>컬럼1</th> <th>컬럼2</th> <th>컬럼3</th> <th>컬럼4</th> </tr> </thead> <tbody> <tr> <td>데이터1</td> <td>데이터2</td> <td>데이터3</td> <td>데이터4</td> </tr> <tr> <td>데이터5</td> <td>데이터6</td> <td>데이터7</td> <td>데이터8</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table> <div style="font-size: 2em; margin: 0 10px;">}</div> <div style="font-size: 0.8em;"> <p>컬럼에 의해 정의된 데이터</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <caption>정형데이터의 예</caption> <thead> <tr> <th></th> <th>Sepal.Length</th> <th>Sepal.Width</th> <th>Petal.Length</th> <th>Petal.Width</th> <th>Species</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>5.1</td> <td>3.5</td> <td>1.4</td> <td></td> <td>setosa</td> </tr> <tr> <td>2</td> <td>4.9</td> <td>3.8</td> <td>1.4</td> <td></td> <td>setosa</td> </tr> <tr> <td>3</td> <td>4.7</td> <td>3.2</td> <td>1.3</td> <td></td> <td>setosa</td> </tr> <tr> <td>4</td> <td>4.6</td> <td>3.1</td> <td>1.5</td> <td></td> <td>setosa</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table> </div> </div>	컬럼1	컬럼2	컬럼3	컬럼4	데이터1	데이터2	데이터3	데이터4	데이터5	데이터6	데이터7	데이터8		Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	1	5.1	3.5	1.4		setosa	2	4.9	3.8	1.4		setosa	3	4.7	3.2	1.3		setosa	4	4.6	3.1	1.5		setosa
컬럼1	컬럼2	컬럼3	컬럼4																																																		
데이터1	데이터2	데이터3	데이터4																																																		
데이터5	데이터6	데이터7	데이터8																																																		
...																																																		
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species																																																
1	5.1	3.5	1.4		setosa																																																
2	4.9	3.8	1.4		setosa																																																
3	4.7	3.2	1.3		setosa																																																
4	4.6	3.1	1.5		setosa																																																
...																																																
<p>○ 비정형 데이터</p>	<p>정의된 형태가 없이 동영상, 오디오, 사진, 보고서, 메일 등 구조가 다양하며 데이터 관리를 위해 메타데이터, 콘텐츠 자체를 저장하는 원본 데이터로 구분</p> <div style="text-align: center;"> </div>																																																				
<p>○ 반정형 데이터</p>	<p>데이터의 구조 정보가 파일 형식이며, 데이터의 형식과 구조가 변경될 수 있는 데이터로서 스키마, 메타데이터 특성을 가지고 있어 주로 XML, HTML, JSON 파일 형태로 저장</p> <div style="text-align: center;"> </div>																																																				

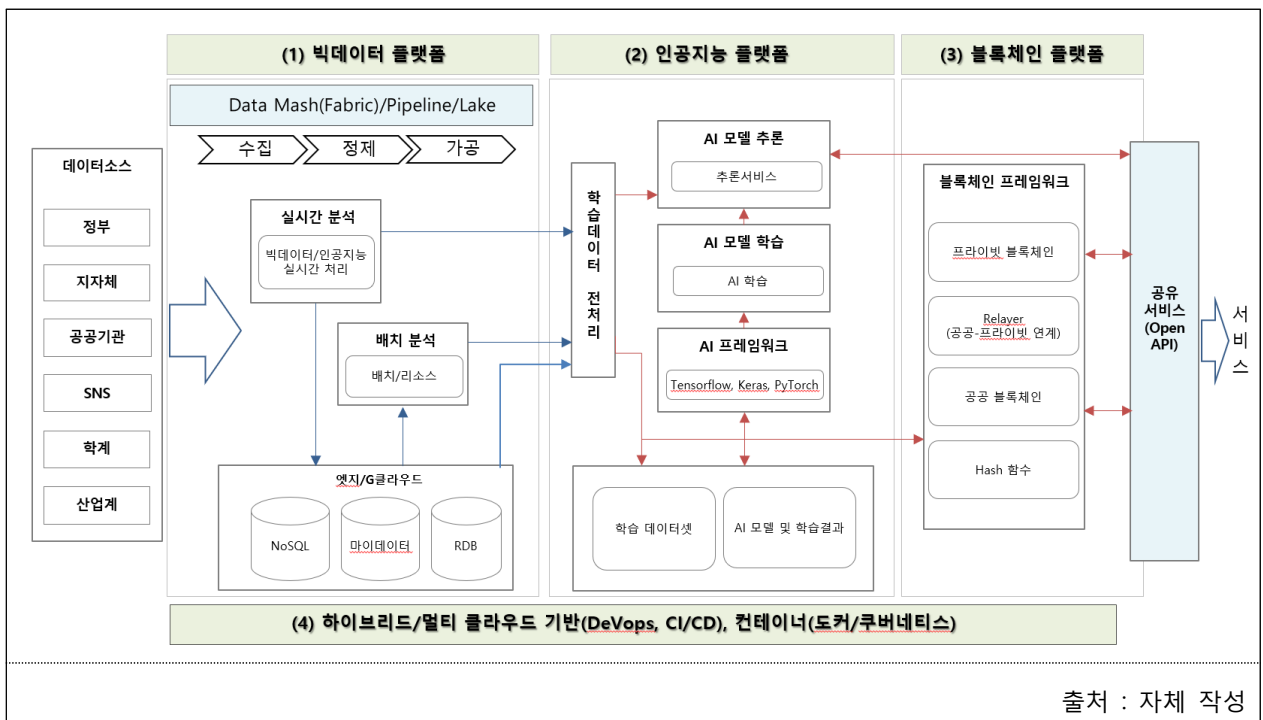
III

주요 기반기술(B·A·N·D) 구현 방향

1 B·A·N·D 기반의 통합플랫폼 구축방향

- (B·A·N·D 기반) 빅데이터·인공지능 플랫폼, 블록체인 기술 적용 및 Open API 서비스로 확장성·유연성을 높이는 통합 플랫폼 구축

<B·A·N·D 기반 통합플랫폼(안)>



출처 : 자체 작성

(1) (빅데이터 플랫폼) Mash·Pipeline·Lake 기술을 적용하여 외부 데이터소스(정부·지자체·공공·민간)에서 수집된 데이터의 수집·정제·가공 과정을 체계적으로 수행

※ 외부 빅데이터의 실시간·배치 분석을 통해 정형·비정형 및 마이데이터를 빅데이터 저장소에 체계적으로 관리함으로써 인공지능에 활용할 수 있는 학습데이터 전처리 수행

(2) (인공지능 플랫폼) 전처리된 학습데이터를 활용하여 AI 추론, AI 학습 과정을 통해 모델의 정확도 검증 이후 서비스 반출

※ Tensorflow, Keras, PyTorch 등 AI 프레임워크를 통한 학습데이터 활용 및 AI 모델학습·추론 과정을 Open API 공유서비스를 통해 개발의 네트워크 효과 창출 환경 조성

(3) (블록체인 플랫폼) 민·관 영역의 다양한 대규모 데이터 가운데 일부 민감 데이터에 제한적으로 블록체인을 적용하여 데이터 활용도 제고

※ 코어-엣지-기관 간의 대규모 데이터의 탈중앙화 데이터 유통체계에 블록체인 인증 기술을 적용하여 공공영역뿐만 아니라 민간의 데이터도 결합하여 과거에 기대하기 힘들었던 다양한 국민 체감형 서비스 창출 가능

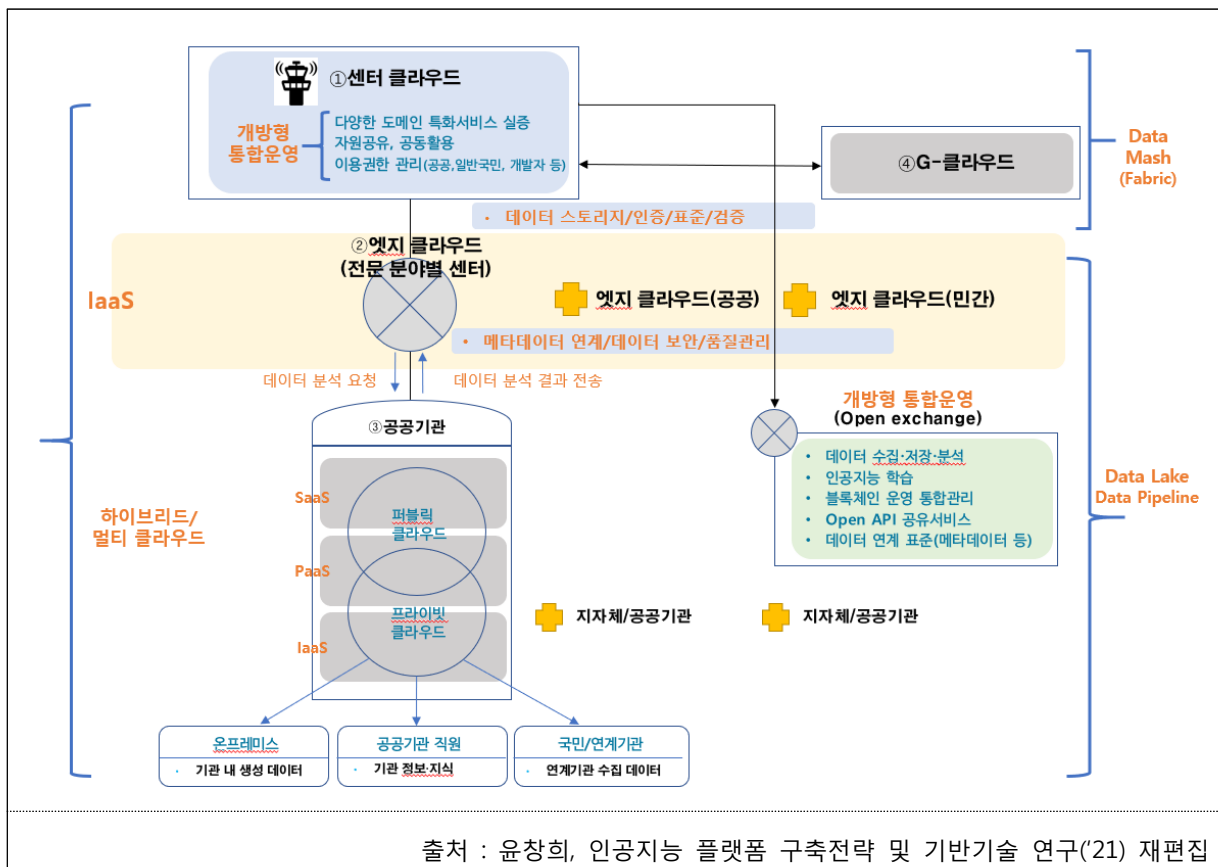
(4) (클라우드 기반) 비공개 데이터를 다루는 정부 데이터 특성상 '데이터 보안' 이슈 해결방식의 인프라 설계

- 민감정보는 디지털플랫폼 정부 온프레미스(IaaS 인프라) 또는 블록체인 활용하여 구축하고 기타 PaaS/SaaS 인프라 기능은 퍼블릭 클라우드를 활용하는 하이브리드 또는 멀티클라우드 방식 구현

※ 다양한 장치, 서비스 결합, 소통수준, 업무처리 가능 수준, 플랫폼 연계 등에 주요 기능 제시

- 공공부문의 다기관 시스템 간의 원활한 연계를 위한 구성 방안 마련

<공공부문 연계를 위한 계층적 클라우드 구성(안)>[25]



① 센터 클라우드(디지털플랫폼 정부)

- 디지털플랫폼 정부의 근간이 되는 플랫폼으로서 공공부문 각 도메인 별로 운영되는 엣지 클라우드 및 G-클라우드 허브 역할 수행하며 대국민 서비스 제공
- 데이터 메쉬(Mash) 기반의 빅데이터 통합-AI-블록체인 플랫폼 간 통합 운영 역할 포함

※ 디지털플랫폼 정부 서비스에 필요한 데이터 스토리지 · 표준화 · 품질검증, Open API 연계, 블록체인 인증, 플랫폼 연계 등 수행

② 엣지 클라우드

- 공공분야 전문 도메인(정부·지자체·공공기관)별로 엣지 역할을 수행할 기관을 지정하여 엣지-관계기관 간 운영에 필요한 자원관리

※ 데이터 파이프라인(Pipeline), 데이터 레이크(Lake) 역할을 동시에 수행하는 영역으로써 데이터 분석 요구 · 결과 전송 및 메타데이터 설계, 데이터 보안, 데이터 품질관리 등 수행

③ 공공기관

- 대국민 서비스에 필요한 데이터를 대규모 데이터를 실제 보유하고 있는 기관으로서 기관 내 생성데이터, 연계기관 수집데이터 등을 관리

※ 데이터 레이크(Lake) 영역에 해당되며 기관 내부 데이터·기능에 적합한 SaaS, PaaS, IaaS 인프라별 퍼블릭·프라이빗 클라우드 구성 방향 결정

④ G-클라우드

- 디지털플랫폼정부 서비스 제공을 위해 필요한 정부 관계부처 클라우드 센터와 연계

※ 데이터 메쉬(Mash)를 지원하는 영역으로서 필요시 산하기관 데이터의 중요도에 따라 연계 범위를 체계적으로 구분하여 관리 필요

2 정책적 제언

□ 서비스 수요가 많은 분야부터 단계적 추진 필요

- 디지털플랫폼 정부 구현은 범정부 기관의 대규모 연계가 필요한 만큼 우선 적용할 서비스 중심으로 단기·중기·장기 단계별로 나누어 추진
 - 국민의 서비스 이용률이 집중된 기관의 데이터를 중심으로 국민 체감형 서비스를 개발하여 단계적 시행
 - 대국민 민원 서비스뿐만 아니라 과학적 국정운영의 근간을 마련하기 위한 관계기관 간 데이터 연계방안은 별도로 단계별 전략 수립

□ 플랫폼, 데이터 간 연계를 위한 세부 설계 필요

- 디지털플랫폼 정부는 범정부 대규모 기관 간 기반 기술(B·A·N·D)에 대한 연계가 중요한 만큼 분야별 세부 실행계획 도출
- (블록체인) 공공영역 연계 데이터 범위에 따라 인증을 수행하는 전담 기관 또는 센터 클라우드 기능추가 등에 대한 세부 방안 도출
 - ※ 공공분야에 적용할 블록체인 데이터 저장소, 암호기술, 관계기관 노드 간 연계 이슈 도출
- (인공지능) 빅데이터 통합플랫폼과의 연계를 바탕으로 Open API을 이용한 참여자 중심의 AI 서비스 창출 비즈니스 모델 도출
 - ※ 공공은 공급·관리 중심으로 운영하고 민간이 주도하여 AI 서비스 개발 및 새로운 가치를 창출할 수 있는 선순환 플랫폼 구축
- (네트워크) 범정부 관계기관 간의 연계를 수행하는 핵심 인프라인 만큼 기관별 클라우드 연계 이슈 전수 조사 및 개선방안 사전 도출 필요
 - ※ 센터-엣지-기관별 그룹화 방안, 허브기관 도출 등 데이터 분산 계획 마련
- (데이터) 연계 기관 간 데이터 Migration에 필요한 데이터 연계 범위·수준 및 기술·제도적 이슈 도출
 - ※ 대국민 서비스 우선 순위에 따라 부처, 공공, 지자체 간 협의체 구성하여 합의 도출

□ 서비스 지속성을 위한 상시 지원 환경 구축 필요

- 서비스 구현 이후 발생 가능한 다양한 기술, 제도, 유지보수 등의 이슈를 전담하여 해결할 수 있는 기관 지정 또는 신설 필요
 - 정부·지자체·공공기관 등 여러 기관 간 연계를 통해 서비스가 구현됨으로 이슈 해결을 위한 컨트롤 타워 역할이 중요
- 클라우드 세부 구성 방안 배포 및 상시 온·오프라인 지원체계 구축
 - 일원화된 통합플랫폼의 센터, 엣지, 기관별로 연계된 다양한 시스템을 지원할 수 있는 원격지원 시스템 등

□ 참여기관 인센티브 방안 마련 필요

- 디지털플랫폼 정부에 선도적으로 참여하여 기반기술(B·A·N·D)을 활용해 서비스하는 기관을 대상으로 인센티브 부여 방안 마련 필요
 - 참여기관 플랫폼, 데이터 기반 기술 연계 및 향후 운영·유지 등을 위한 복잡한 세부 업무를 직접 수행해야 할 업무가 과중한 만큼
 - 적극적인 참여 유도를 위해 이에 상응하는 기관 또는 개인에 대한 인센티브 방안 마련 필요

붙임1 데이터 메쉬 플랫폼 개념도

<p>개요</p>	<ul style="list-style-type: none"> ○ 메쉬 개념 - 플랫폼(Data Infrastructure-as-a-Platform) 상에 데이터를 도메인별(Domain 1~6)로 분산하여 데이터를 처리하는 모델
<p>주요 내용</p>	<div data-bbox="295 624 1442 1354" style="border: 1px solid black; padding: 10px; margin-bottom: 10px;"> </div> <ul style="list-style-type: none"> ① Data Infrastructure-as-a-Platform : 기존 데이터웨어하우스 시스템으로부터 데이터를 도메인별로 적절히 분배하여 관리하는 플랫폼 ② Domain(1~6) : 데이터 특성에 따라 주제별로 분산하여 데이터를 추출·변환·로딩 과정을 처리할 수 있도록 별도로 운영하는 시스템 ③ E-T-L : (Extract) 데이터 소스로부터 원시데이터를 추출, (Transform) 데이터를 특정 형식으로 변환, (Loads) 시스템에 데이터를 로딩하는 것을 의미 ④ 데이터웨어하우스 : ETL 파이프라인을 통해 기존 시스템(Legacy)으로부터 데이터를 마이그레이션하여 분석 수행

출처 : MONTE CARLO Data Mash 데이터플랫폼 개념도 재편집(2022.12)

붙임2 데이터 파이프라인 구성요소

<p>개요</p>	<p>○ 데이터 파이프라인 개념 : 데이터 소스, 목적지, 데이터 흐름 저장소, 처리, 워크플로우, 모니터링 7가지 기능을 바탕으로 데이터 파이프라인 아키텍처로부터 원시데이터에 대한 ETL 기능 수행</p>
<p>주요 내용</p>	<div data-bbox="300 546 1437 1113" style="border: 1px solid black; padding: 10px; text-align: center;"> <p>DATA PIPELINE COMPONENTS</p> </div> <p>① 소스(Origin) : 파이프라인의 데이터 입력 지점으로 데이터 소스와 저장 시스템(데이터 웨어하우스/레이크/레이크 하우스) 등을 의미</p> <p>② 목적지(Destination) : 데이터가 전송되는 마지막 위치로서 목적지에서 데이터 분석, 데이터 레이크웨어하우스로 이동하여 저장 가능</p> <p>③ 데이터 흐름(Dataflow) : Origin에서 Destination까지 데이터의 이동을 의미하는 것으로 이동하면서 발생한 변화 및 저장을 포함</p> <p>④ 저장소(Storage) : 파이프라인을 통과하면서 데이터가 보존되는 시스템을 의미하는 것으로 데이터 양, 시스템 환경 등에 따라 결정</p> <p>⑤ 처리(Processing) : 처리는 소스로부터 데이터를 수집하고, 저장하고, 변환하고, 목적지에 전달하기 위한 활동</p> <p>⑥ 워크플로우(Workflow) : 워크플로우는 프로세스(작업)의 연속과 작업 task 수행을 위한 그들간의 종속성을 정의하는 것</p> <p>⑦ 모니터링(Monitoring) : 파이프라인 단계 상태를 체크하는 것으로, 데이터 증가함에 따라 데이터 정확도, 일관성, 유실 여부 등을 확인함</p>

출처 : Altexsoft Data Pipeline Component 재편집(2022.12)



참고 자료

- [1] 윤석열 정부의 디지털 혁신 드라이브, DPG 디지털플랫폼정부 시리즈, p.10, 2022
- [2] S. Choi, "The great dictionary of industrial safety," Gold, May 2004. (Oejin Park "The concept and introduction strategy of artificial intelligence platform," 4p, Re-quoted)
- [3] 디지털플랫폼 정부의 개념과 특징, IT&Future Strategy NIA, p.11, 2022
- [4] 김준연, "디지털플랫폼과 생태계 혁신", SPRI, p.36, 2022
- [5] 윤창희외, "공공분야 인공지능 플랫폼 분석 연구", 한국통신학회, p.1994-1995, 2021
- [6] AI미래전략센터, "디지털플랫폼 정부의 개념과 특징", NIA, p.7-9, 2022
- [7] 윤창희, "공공분야 인공지능 플랫폼 분석 연구", 한국통신학회, p.2001-2002, 2022
- [8] 윤창희, "글로벌 선도기업 인공지능 플랫폼 분석 연구", 2021
- [9] 임철홍 외, "클라우드 컴퓨팅 효과를 최대화하는 클라우드 네이티브", 한국정보처리학회, p.15-16, 2021
- [10] <https://cloudmt.co.kr/?p=3927>
- [11] 공유서비스(OpenAPI)개발 가이드라인 및 기술표준적용 가이드라인(행안부 디지털 정부국 정보화 표준/지침)
- [12] 윤창희 외, "인공지능 데이터 품질검증 기술 및 오픈소프 프레임워크 분석연구", 한국멀티미디어학회, p.1403-1405, 2021
- [13] 공공기관 비정형데이터 품질관리 모델 개발사업 비정형데이터 관리방안, 행안부, 2021
- [14] <https://aihub.or.kr/intrcn/guid/usagepolicy.do?currMenu=151&topMenu=105>
- [15] <https://www.montecarlo.com/blog-what-is-a-data-mesh-and-how-not-to-mesh-it-up/>
- [16] <https://www.ibm.com/kr-ko/topics/data-pipeline#>:
- [17] Fatemeh Nargesian et al, "Data Lake Management: Challenges and Opportunities", p.1987-1988, 2019
- [18] 김남곤 외, "글로벌 차세대 데이터컴퓨팅 기술혁신센터 사전기획연구", GIST, p.16-17, 2018
- [19] Mikhtoniuk, Sergii, and Ozge Nilay Yalcin. "Open Data Fabric: A Decentralized Data Exchange and Transformation Protocol With Complete Reproducibility and Provenance." arXiv preprint arXiv:2111.06364. p.6-9, 2021

[20] <https://cloud.google.com/learn/what-is-a-data-lake?hl=ko#:~:text=>

[21] <https://www.grazitti.com/blog/data-lake-vs-data-warehouse-which-one-should-you-go-for/>

[22] https://spri.kr/posts/view/23140?code=industry_trend

[23] 2022년 공공 마이데이터 권역별 이용기관 교육자료, 행안부, 2022

[24] http://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=175128-2

[25] 윤창희, 인공지능 플랫폼 구축전략 및 기반기술 연구, p.48, 2021

IT & Future Strategy 보고서

- 제1호(2022.4.18.) 「3년간 우리나라 D.N.A 분야 혁신기업 현황 분석」
- 제2호(2022.5.12.) 「디지털 전환 시대, NIA가 전망한 환경 변화 13대 이슈」
- 제3호(2022.7.25.) 「주요 국가 AI 거버넌스 분석」
- 제4호(2022.8.9.) 「주요 국가 AI 전략 분석」
- 제5호(2022.9.8.) 「디지털플랫폼 정부 서비스를 위한 인공지능(AI) 가상비서 발전 방향」
- 제6호(2022.9.8.) 「글로벌 6대 빅테크 기업의 M&A 현황 및 주요 트렌드 분석」
- 제7호(2022.10.12.) 「디지털 플랫폼 정부의 개념과 특징」
- 제8호(2022.12.5.) 「주요 국가 AI 거버넌스 분석(하)」
- 제9호(2022.12.28.) 「디지털플랫폼정부를 위한 기반기술(B·A·N·D) 이용 활성화 방안」

1. 본 보고서는 방송통신발전기금으로 수행한 정보통신·방송 연구개발 사업의 결과물이므로, 보고서의 내용을 발표할 때는 반드시 과학기술정보통신부 정보통신·방송 연구개발 사업의 연구결과임을 밝혀야 합니다.
2. 본 보고서 내용의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를「한국지능정보사회진흥원(NIA)」이라고 밝혀 주시기 바랍니다.
3. 본 보고서의 내용은 한국지능정보사회진흥원(NIA)의 공식 견해와 다를 수 있습니다.

비매품/무료

93350



9 788984 837447

ISBN 978-89-8483-744-7