



SW · 컴퓨팅

빅데이터

빅데이터

목차

I

표준화 개요

| | |
|--------------------------|-----|
| 1.1. 기술 개요 | 157 |
| 1.2. 중점 표준화 항목 | 159 |
| 1.3. 표준화 비전 및 기대효과 | 163 |

II

국내외 현황분석

| | |
|---------------------------|-----|
| 2.1. 연도별 주요 현황 및 이슈 | 166 |
| 2.2. 정책 현황 및 전망 | 167 |
| 2.3. 기술개발 현황 및 전망 | 168 |
| 2.4. IPR 현황 및 전망 | 172 |
| 2.5. 표준화 현황 및 전망 | 175 |
| 2.6. 오픈소스 현황 및 전망 | 182 |

III

국내외 표준화 추진전략

| | |
|--------------------------------------|-----|
| 3.1. 표준화 SWOT 분석 | 185 |
| 3.2. 중점 표준화 항목별 국내외 추진전략 | 186 |
| 3.3. 오픈소스 국내외 추진전략 | 201 |
| 3.4. 중기(3개년) 및 장기(10개년) 표준화 계획 | 202 |

| | |
|------------|-----|
| 작성위원 | 204 |
|------------|-----|

| | |
|------------|-----|
| 참고문헌 | 205 |
|------------|-----|

| | |
|----------|-----|
| 약어 | 206 |
|----------|-----|

I. 표준화 개요

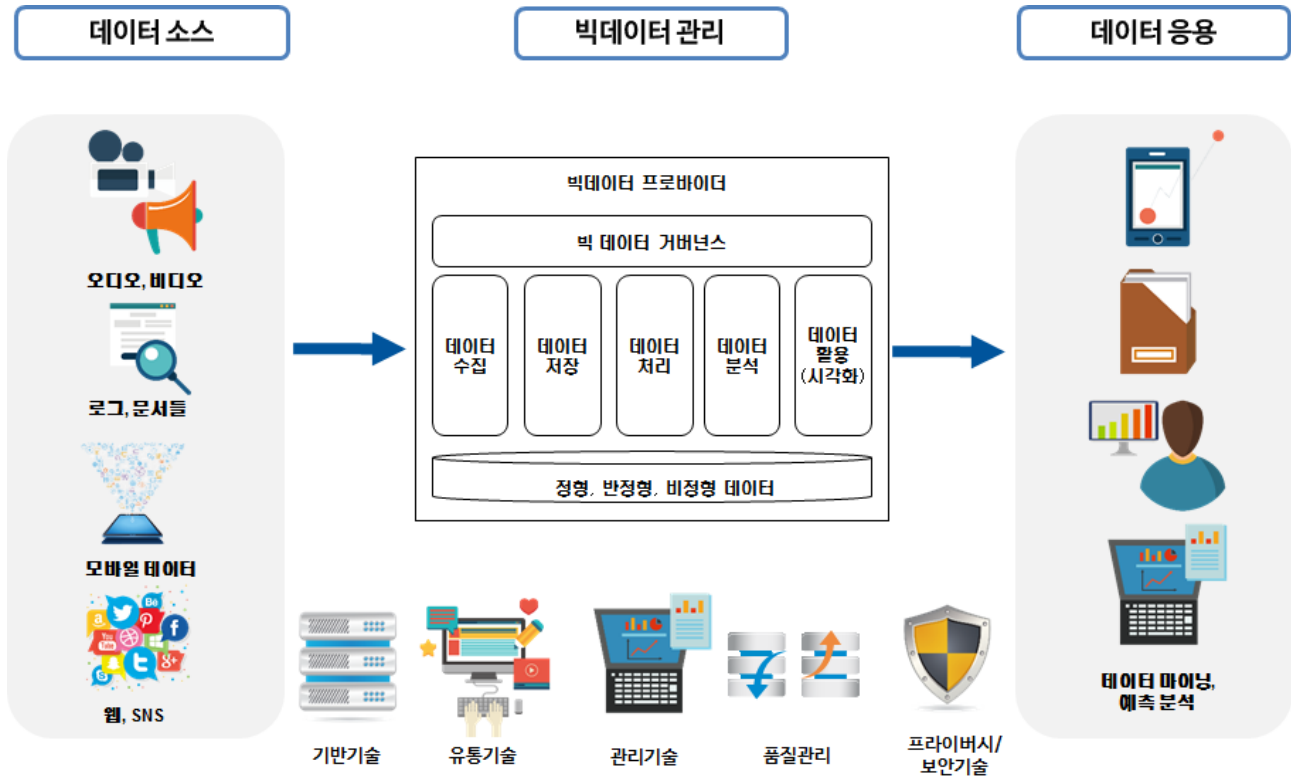
1.1. 기술 개요

빅데이터는 기존의 데이터 아키텍처로 효과적으로 처리하기 어려운 데이터를 지칭하는 용어로써, 대용량 데이터(volume), 다양한 분야별 데이터 유형의 다양성(variety), 데이터 생성 시간의 가속화(velocity) 및 빠른 변화(variability) 등 특성을 지닌 데이터를 효과적으로 처리하기 위한 기술

- 빅데이터는 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처로서, 당초 수십~수천 테라바이트에 달하는 거대한 데이터 집합 자체만을 지칭하였으나, 점차 관련 도구, 플랫폼, 분석기법까지 포괄하는 용어로 변화

※ 매킨지 세계 연구소(2011), 삼성전자경제연구소(2011), 가트너(2012), Oracle(2012) 등에서 빅데이터에 대한 정의를 다양하게 내리고 있으며, 현재 표준화 기구에서 빅데이터에 대한 정의를 개발 중. 이들의 공통된 정의를 규정짓는 특징은 양(Volume), 속도(Velocity), 다양성(Variety), 가치(Value) 등

- 빅데이터는 “새로운 가치 추출을 위해 기존의 기술 또는 기법으로 처리하기 어려운 특징들(규모, 신속성, 가변성, 다양성, 진정성)을 갖는 데이터 모음”으로 정의
- 규모(Volume): 전통적으로 데이터 분석 서비스에 필요한 데이터의 규모는 테라바이트(terabyte) 규모가 통용되었으나 최근 테라바이트에서 페타바이트(petabyte)까지 그 규모가 커짐으로 인해 기존의 정보 기술 구조로는 효과적인 처리가 어렵게 됨. 이에 따라 수평적으로 연계 가능한 자원들간의 확장과 분산 질의 처리 방법들로 비 관계형 데이터 모델, NoSQL, 빅 테이블, 그래프 모델 등이 사용
- 다양성(Variety): 기존의 데이터 분석에는 개별 도메인별로 구조화된 단일 데이터 유형을 사용. 하지만 최근 데이터 응용들은 정형, 반정형, 비정형 등 다양한 형태의 데이터들을 사용하고 있으며, 또한 이들 데이터들을 서로 조합함으로써 새로운 지식을 추출하는데 활용. 따라서 반정형/비정형의 데이터들을 보다 효과적으로 처리하고, 다양한 데이터 유형들을 여러 데이터 저장소에 효과적으로 저장하고 연계 활용하기 위한 방법들이 사용
- 신속성(Velocity): 신속성은 데이터가 생성되어, 저장, 분석, 시각화되는 과정에 필요한 시간으로 표현. 기존에는 트랜잭션 처리와 분석을 분리하여, 미리 저장된 데이터를 일괄 처리 방식으로 분석하여 사용. 그러나 실시간, 스트리밍 데이터 기반의 분석 요구가 증가하고 있으며, 이를 지원하기 위한 방법들이 모색되고 있음. 이는 기존에 이미 연구가 이루어지고 있다는 점에서 새로운 접근은 아니며, 빅데이터 관점에서는 이러한 데이터들을 효과적으로 다루기 위한 수평적 확장성에 초점이 맞추어져 있음
- 가변성(Variability): 가변성은 데이터의 포맷/구조, 의미, 품질 등과 같이 응용 또는 분석 유형에 따라 변하는 데이터의 특징을 의미. 이는 빅데이터의 다른 특징들이 변하는 것을 의미하며, 아키텍처, 인터페이스, 처리 알고리즘, 데이터 융합, 저장, 데이터 사용 등의 요인으로 인하여 발생. 예를 들어 빅데이터 아키텍처를 채용한 시스템일 경우에도 데이터를 용도에 따라 재가공하여 사용하거나 외부의 다양한 데이터와 융합하여 분석할 필요가 있으며, 따라서 데이터의 변화에 대한 신속한 대응이 필요
- 진정성(Veracity): 진정성 데이터에 대한 진실성, 사용 가능성, 왜곡 정도 등 데이터의 품질을 나타내는 것으로, 데이터의 융복합에 따른 원시 데이터에 대한 추적, 처리 방법 등에 대한 투명성을 보장하기 위한 기술적 장치가 필요



<빅데이터 기술 개요도>

1.2. 중점 표준화 항목

○ 중점 표준화 항목 범위의 설정

- (중분류 범위 설정) 빅데이터의 중분류는 ITU-T SG13, JTC1 SC42 및 SC32, 기타 표준화 기구에서의 개발하고 있는 표준화 항목을 기준으로 빅데이터 자체에 대한 “기반 기술”, 서로 다른 기관과 데이터 교환을 위한 “유통기술”, 메타데이터 및 데이터 관리 측면에서의 “관리 기술”, 마지막으로 개인정보 및 보안을 위한 “프라이버시/보안기술”, 이렇게 네가지 분류로 범위를 설정
- (중점 표준화 항목 선정 이유) 표준화전략맵 빅데이터 분과에서는 과기정통부와 관련 없는 항목인 “제조 운영에 대한 데이터 품질 프로세스 평가 표준”, 국제적으로 표준화가 완료 단계인 “빅데이터 개념 및 정의 표준”, “빅데이터 교환을 위한 프레임워크 및 요구사항 표준”, 현 시점에서 명확한 전략을 세울 수 없는 “빅데이터 프리저베이션 요구사항 표준”, 국제 표준화는 진행 중이나 국내 참여가 없는 “빅데이터 프라이버시 및 보안 관리 표준”을 제외하였으며, 그리고 표준 개발이 본격적으로 착수하지 않은 “빅데이터 참조 구조를 위한 인터페이스 표준”, “메타데이터의 의미적 상호운용성 개념 및 요구사항 표준”, “빅데이터 분석을 위한 데이터 거버넌스 표준” 등을 제외하여 아래와 같이 중점 표준화 항목을 선정

| 표준화 항목 | | 표준화 내용 | Target SDOs | 표준화 특성 | 중점 항목 |
|--------|-------------------------------|---|-----------------------|--------|-------|
| 기반기술 | 빅데이터 참조 구조 표준 | 빅데이터 생태계를 구성하는 각 요소(데이터 제공자, 빅데이터 서비스 제공자, 빅데이터 소비자)들 대한 역할 정의 및 이들 간의 관계, 그리고 이들을 구현하기 위한 개념적 기능 컴포넌트들을 정의 | JTC1 SC42, ITU-T SG13 | ③ | ○ |
| | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준 | 클라우드 기반에서 빅데이터 서비스를 제공에 필요한 요구사항을 기반으로 기능적인 요소 및 구조를 개발하는 표준 - 클라우드 기반의 빅데이터 제공을 위한 기본 요소 기능 정의 - 클라우드 기반의 빅데이터 제공에 필요한 클라우드 구조(CCRA)의 확장 내용 정의 - Bigdata as a service(BDaaS) 구성 요소 들 간의 인터페이스 정의 표준 | ITU-T SG13 | ③ | ○ |
| | 빅데이터 참조 구조를 위한 인터페이스 표준 | ISO/IEC 20547-3 빅데이터 참조 아키텍처에 정의된 인터페이스들에 대한 특징(characteristics) 및 이들을 구현하는데 요구되는 디자인 가이드 라인, 참조 가능한 베스트 프랙티스 등을 제공 | JTC1 SC42 | ③ | X |
| | 빅데이터 개념 및 정의 표준 | 빅데이터의 일반 개념과 빅데이터 분야에서 공통적으로 사용되고 있는 용어들을 식별하고 그 의미를 정의 | JTC1 SC42 | ① | X |

| 표준화 항목 | | 표준화 내용 | Target SDOs | 표준화 특성 | 중점 항목 |
|--------|------------------------------------|---|-------------|--------|-------|
| 유통기술 | 빅데이터 프로비넌스 요구사항 표준 | 빅데이터 환경에서 데이터의 융합 및 변경 등에 대한 이력 및 출처에 대한 추적을 지원하기 위한 기능 요구사항을 정의하기 위한 표준 - 빅데이터 프로비넌스의 특징, 활용분야, 기능적 프레임워크를 포함하는 빅데이터 프로비넌스 개념을 정의 - 빅데이터 프로비넌스 유즈케이스 및 이를 바탕으로 도출된 요구사항을 정의 | ITU-T SG13 | ② | ○ |
| | 빅데이터 교환을 위한 프레임워크 및 요구사항 표준 | 빅데이터 교환을 데이터 수집(import) 및 배포(export)를 위한 기능 - 데이터 교환과 관련된 빅데이터 생태계 모델(big data ecosystem) 상에서의 역할을 식별하는 프레임워크 - 각 역할들에게 요구되는 기능적 요구사항 | ITU-T SG13 | ② | X |
| 관리기술 | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 | 빅데이터 환경에서 저장되는 도메인별 메타데이터의 상호운용성을 확보하기 위한 개념, 등록, 관리, 거래, 활용에 대한 표준 - 빅데이터 환경에서의 메타데이터 개념, 이용 상황(context), 일반 요구사항을 포함하는 메타데이터 프레임워크를 정의 - 메타데이터를 구성하는 컴포넌트 클래스 및 이들의 속성을 정의 - 메타데이터 교환을 위한 인코딩 포맷을 제공 | ITU-T SG13 | ④ | ○ |
| | 메타데이터의 의미적 상호운용성 개념 및 요구사항 표준 | 빅데이터 환경에서 저장되는 도메인별 메타데이터의 의미적 상호운용성을 확보하기 위한 요구사항, 변환, 표현에 대한 표준 - 메타데이터 의미적 상호운용성 요구사항 개념 및 요구사항 표준 - 빅데이터 환경에서 메타데이터 등록, 변환, 표현에 대한 참조모델 표준 | JTC1 SC32 | ② | X |
| | 빅데이터 분석을 위한 데이터 거버넌스 표준 | 빅데이터 분석 라이프사이클에 따른 데이터 활용 및 관리를 위한 성숙도 모델 및 가이드라인을 정의 - 빅데이터 분석을 위한 데이터 라이프사이클 정의 - 데이터 관리를 위한 데이터 활용 성숙도 모델 정의 - 데이터 관리 요구사항 정의 | JTC1 SC42 | ③ | X |
| | 빅데이터 프리저베이션 요구사항 표준 | 빅데이터 저장소의 효과적/효율적 관리를 위한 관리 정책 및 기능 요구사항 정의 - 빅데이터 프리저베이션 개념 정의 - 빅데이터 프리저베이션 관련 유즈케이스 및 이를 통해 도출된 기능적 요구사항 정의 | ITU-T SG13 | ② | X |

| 표준화 항목 | | 표준화 내용 | Target SDOs | 표준화 특성 | 중점 항목 |
|------------|-----------------------------|---|-----------------------|--------|-------|
| 품질관리 | 빅데이터 품질측정 및 관리 모델 표준 | 빅데이터 품질관리를 위한 품질측정, 관리 모델 및 품질요소에 대한 정의 - 빅데이터 품질관리를 위한 요구사항의 개념 정의 표준 - 빅데이터 처리과정에 따른 이력관리 및 품질요소 측정 기준 표준 - 활용용도 기반의 빅데이터 품질요소 정의 표준 - 빅데이터 품질 관리 요구사항 정의 표준 | JTC1 SC32, ISO TC184 | ① | O |
| | 데이터 프로파일링 기반의 빅데이터 품질진단 표준 | 데이터 프로파일링 기반의 데이터 품질진단을 위한 메타데이터 분석, 프로파일링 방법 등에 대한 정의 - 데이터 품질진단 기본 개념 및 구조 정의 표준 - 데이터 품질진단을 위한 메타데이터 분석, 데이터 프로파일링 및 테스트 방법 - 데이터 품질진단 방법에 따른 가이드라인 정의 표준 - 데이터 품질진단 기능 설명을 위한 유스 케이스 및 양식 정의 표준 | JTC1 SC32, ISO TC184 | ① | O |
| | IoT 기반 빅데이터 품질관리 프레임워크 표준 | IoT 기반의 빅데이터 품질을 관리하기 위해 프레임워크에 대한 정의 표준 - IoT 기반의 SCP(smart connected product)에서 운영되는 데이터에 대한 품질 관리 표준 - IoT 데이터 품질 특성 표준 - IoT 데이터 프로파일링 표준 | JTC1 SC32, ISO TC184 | ① | O |
| | 제조 운영에 대한 데이터 품질 프로세스 평가 표준 | 제조 운영 등의 산업에서 생산·활용하는 데이터의 품질을 평가하기 프로세스에 대한 정의 표준 - IEC 62264에서 정의된 제조 운영의 데이터에 대해 ISO 8000-61의 데이터 처리 프로세스를 적용 - ISO/IEC 33001에 준하는 평가 지표 적용 - 제조 활동에서 발생하는 데이터의 품질관리 성숙도 평가 | ISO TC184 | ③ | X |
| 프라이버시/보안기술 | 빅데이터 프라이버시 및 보안 관리 표준 | 빅데이터 참조 모델에서의 보안 역할 및 책임, 보안 및 프라이버시 패브릭 - 모바일 인터넷서비스에서의 보안챌린지, 보안위협 및 보안 요구사항 - 서비스로서의 빅데이터 환경에서의 보안 챌린지, 역할 및 책임, 보안대책 | JTC1 SC27, ITU-T SG17 | ③ | X |

<표준화 특성>

① : 개념, 정의 표준

② : 유즈케이스 및 요구사항 표준

③ : 기능 도출 및 참조구조 표준

④ : 데이터포맷, 스키마 표준

⑤ : 프로토콜, 인터페이스 표준

○ 추진경과

- Ver.2017(2016년)에서는 신규 표준화 항목으로 빅데이터의 연역 관리를 위한 빅데이터 프로비넌스가 추가
- Ver.2018(2017년)에서는 Ver.2017에서 클라우드컴퓨팅과 빅데이터를 분리하였으며, 항목 변경은 없으나, 빅데이터 프라이버시 및 보안 관리, 빅데이터 환경을 위한 메타데이터 등의 항목은 보다 구체화된 이름으로 변경
- Ver.2019(2018년)에서는 기존에 진행되는 표준화 항목 중 국제 표준화가 완료된 것은 종료하고, 품질관리 분야에 신규로 2건의 표준화 항목을 추가

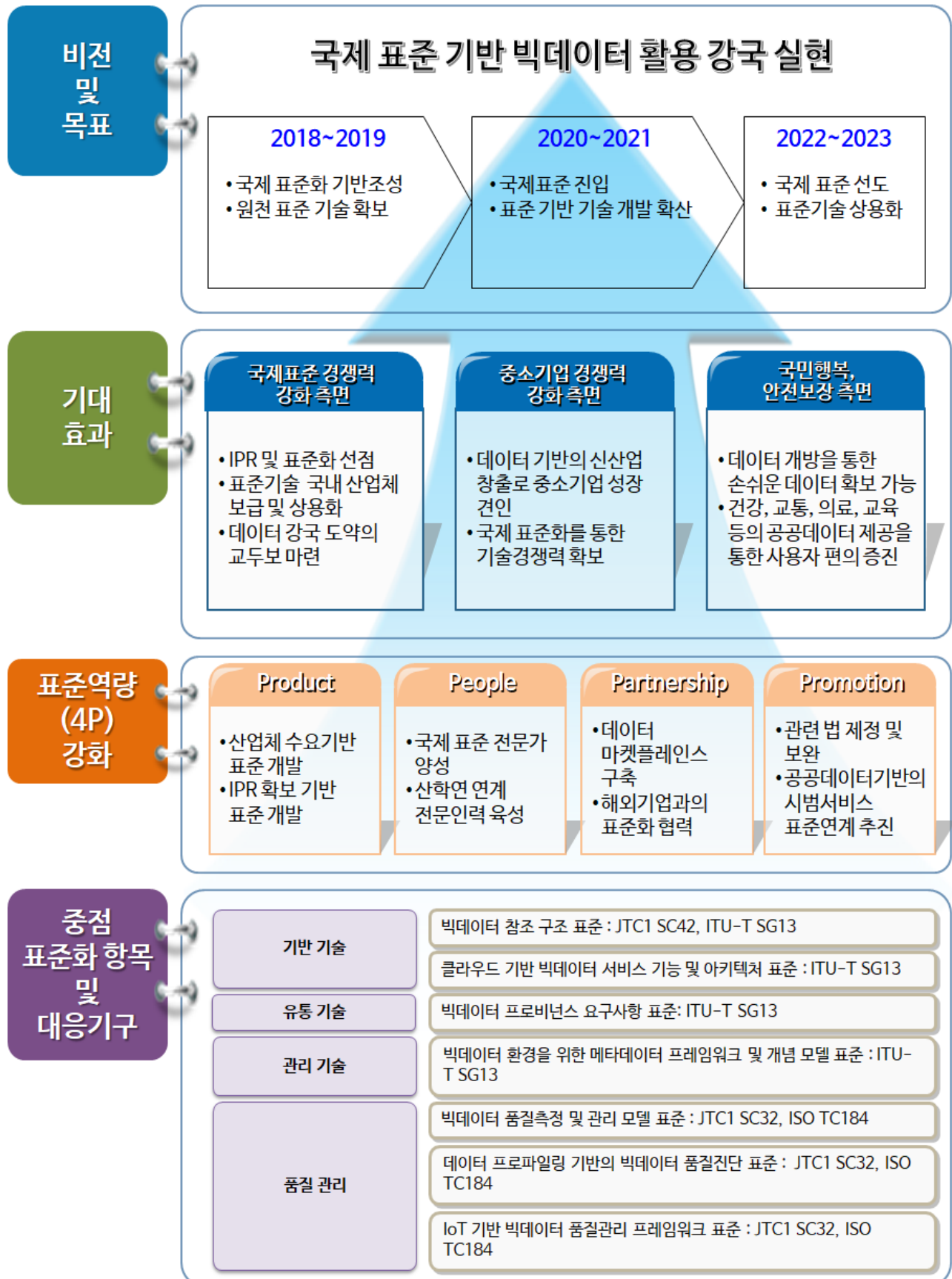
<버전별 중점 표준화 항목 비교표>

* Ver.2019 신규 항목

| 구분 | Ver.2017 | Ver.2018 | Ver.2019 |
|------------|----------------------------|----------------------------|------------------------------------|
| 기반기술 | 빅데이터 개념 및 정의 | 빅데이터 개념 및 정의 | - |
| | 빅데이터 참조 모델 | 빅데이터 참조 구조 | 빅데이터 참조 구조 표준 |
| | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준 |
| 유통기술 | 빅데이터 교환을 위한 프레임워크 및 요구사항 | 빅데이터 교환을 위한 프레임워크 및 요구사항 | - |
| | 빅데이터 프로비넌스 요구사항 | 빅데이터 프로비넌스 요구사항 | 빅데이터 프로비넌스 요구사항 표준 |
| 관리기술 | 빅데이터를 위한 메타데이터 | 빅데이터 환경을 위한 메타데이터 | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 |
| 품질관리 | 빅데이터 품질 측정 및 관리 모델 | 빅데이터 품질측정 및 관리 모델 | 빅데이터 품질측정 및 관리 모델 표준 |
| | - | - | 데이터 프로파일링 기반의 빅데이터 품질진단 표준* |
| | - | - | IoT 기반 빅데이터 품질관리 프레임워크 표준* |
| 프라이버시/보안기술 | 빅데이터 보안 및 개인정보 보호를 위한 기능 | 빅데이터 프라이버시 및 보안 관리 | - |

1.3. 표준화 비전 및 기대효과

○ 표준화 비전



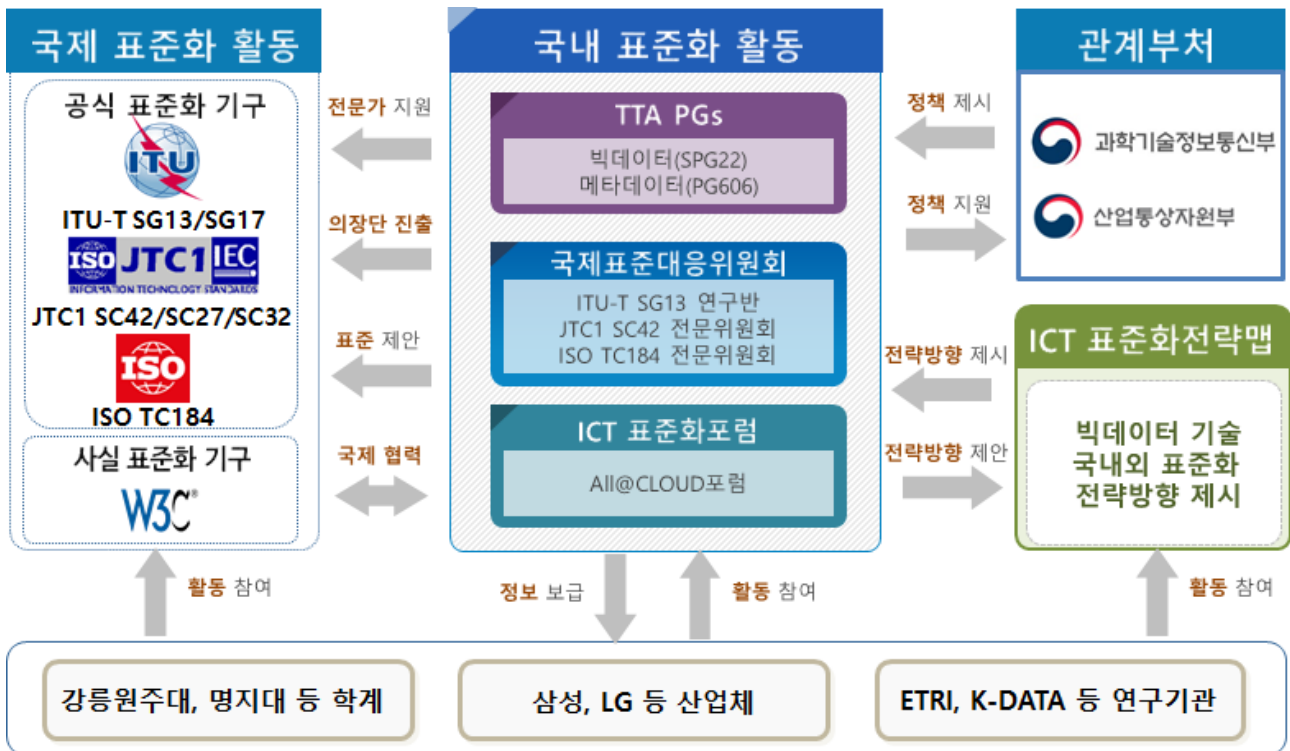
○ 표준화 목표

- 국내 빅데이터 관련 산·학·연은 지속적인 주도권 및 경쟁력 확보를 유지하기 위하여 다음과 같은 표준화 목표를 설정
 - (2019년경까지), 국내 공공 및 민간 데이터 개방에 따라 확보된 데이터를 유통 및 활용하는데 필요한 표준의 기획, 표준 과제 발굴, 표준(안) 개발을 통해 국내 표준 기반 조성 및 국제 표준화를 추진하기 위한 JTC1 및 ITU-T 등의 빅데이터 관련 표준화 그룹의 의장단 및 에디터십 확보 추진
 - (2021년경까지), 빅데이터 유통 활성화 기술 개발에 필수적인 국제 표준 기고를 통해 국가 기술 경쟁력을 확보하며, 국제 표준화를 선점하여 국제 표준을 선도. 또한 빅데이터와 관련 기술과의 연계 표준 개발 노력. 현재 JTC1은 빅데이터 표준 아이টে에 대해서 SC42(인공지능)으로 이관한 상태
 - (2023년경까지), 2단계 활동으로 개발된 표준 기술에 대한 상용화를 추진함과 동시에 국제 표준 기고를 통해 국제 표준화를 선점하여 국제 표준을 선도. 특히 빅데이터 관련 인터페이스, 스키마 등의 표준에 대한 국제 표준 선도

○ 표준화 기대효과

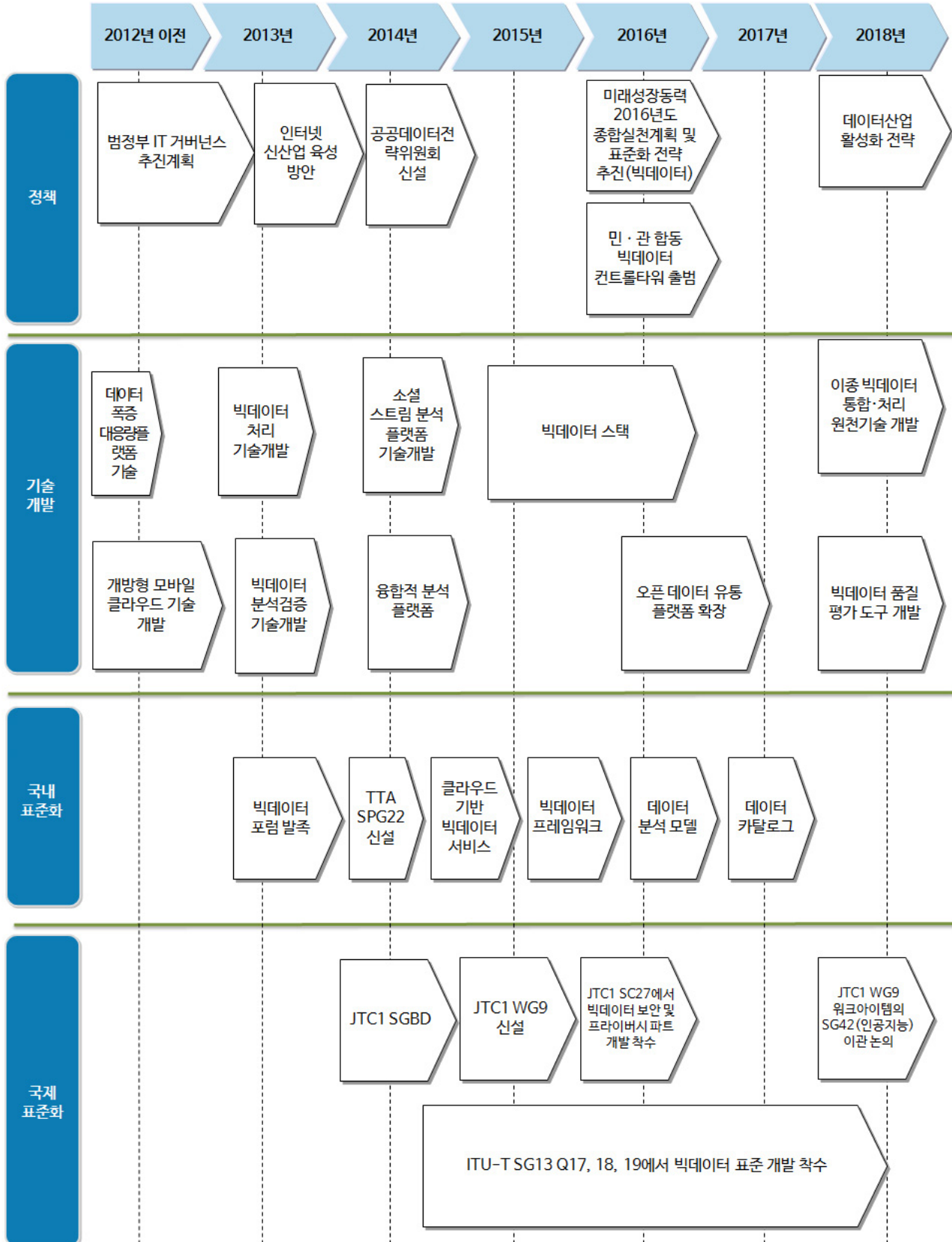
- 국제표준 경쟁력 강화 측면
 - 빅데이터의 IPR 및 표준화를 선점하고 국내 산업체 보급 및 상용화를 통하여 도메인별 클라우드 서비스를 포함한 SW 서비스 및 플랫폼 분야의 세계시장 경쟁력을 제고시키며, 차세대 SW 산업 강국 도약의 교두보 마련
 - 빅데이터의 수요자 중심의 기술 표준화 추진을 위하여 기술·제품에 대한 개방형 표준규격 개발 및 제품 간(HW, SW 솔루션 등) 호환성 제공으로 국내 빅데이터 활성화와 국제 표준화를 통한 기술경쟁력을 확보
- 중소기업 경쟁력 강화 측면
 - 글로벌 기업의 기술-서비스 독점 및 종속성 극복을 통한 시장 공정경쟁 환경 마련 및 이용자 편의 증진
 - 빅데이터 산업을 통한 신생 기업 활성화로 고용 창출 효과
- 국민행복·안전보장 측면
 - 국내의 빅데이터를 활용한 인력수요가 증가할 것으로 예상되며 빅데이터 관련 일자리 창출 효과
 - 건강, 교통, 의료, 교육 등의 공공 데이터 제공을 통한 사용자의 편의 증진

○ 표준화 추진체계



II. 국내외 현황분석

2.1. 연도별 주요 현황 및 이슈



2.2. 정책 현황 및 전망

| 구분 | 주요 현황 |
|----|--|
| 한국 | <ul style="list-style-type: none"> - 빅데이터 산업 활성화 도모를 위해 31억 규모의 '2018년 빅데이터 플래그십 선도사업(실증확산)' 7개 과제를 선정하여 추진[2018.4, 과학기술정보통신부] - 빅데이터 산업 생태계 강화를 위해 12억 규모의 '2018 빅데이터 전문센터 구축 지원사업' 3개 센터 선정 추진 [2018.4, 과학기술정보통신부] - 빅데이터 산업 발전 등 4차 산업혁명 기술혁신을 위한 I-KOREA 4.0 : ICT R&D 혁신전략 발표 [2018, 과학기술정보통신부] - 빅데이터 확산을 위한 39억 규모의 '2017년 빅데이터 플래그십' 프로젝트 12개 과제 선정해 시범사업 추진 [2017.6, 과학기술정보통신부] - 빅데이터 분야 개인정보 규제혁신 방안 마련 및 빅데이터 분석 활용을 위한 '개인정보 비식별 조치 가이드라인' 발표 [2016, 정부합동] - TTA는 2014년부터 빅데이터 SPG(SPG22)를 신설하고 빅데이터 플랫폼에 대한 요구사항, 표준개발을 착수하였으며, 빅데이터 참조 아키텍처, 데이터 서비스 개념, 정의 및 요구사항 등 관련 표준 개발을 추진 중 [2014~ 현재] - '빅데이터 산업 발전 전략' 수립 및 추진, '정부 3.0'으로 공공데이터 개방 추진 등 [2013, 舊 미래창조과학부] - 4차산업혁명위원회는 마이데이터 선도입, 원시데이터 전방위 구축·개방, 데이터 안심존 구축을 위한 데이터 산업 활성화 전략 발표 [2018.6, 4차산업혁명위원회] |
| 미국 | <ul style="list-style-type: none"> - "빅데이터 R&D 전략 계획"(2016)을 수립하여 범부처 차원에서 빅데이터 7대 R&D 전략과 18개 세부과제를 제시하여 미래 빅데이터 환경 변화에 대응 - "빅데이터 R&D 전략 계획"의 후속조치로 빅데이터 지역 혁신 허브화를 추진하여 사회문제 해결 연구에 활용 |
| 일본 | <ul style="list-style-type: none"> - 일본 정부는 "미래투자전략"(2017)을 수립하여 빅데이터를 ICT 융합 활성화 촉매로 활용하기 위해 이노베이션·벤처를 탄생시키는 선순환 시스템 구축하여 빅데이터 산업에 집중 투자 지원 추진 - 경제 균형발전에 빅데이터를 활용하기 위하여 일본 경제산업성에서 비장 빅데이터 체계 구축사업 개시 [2016] - 일본 정부는 2012년 빅데이터 활용특별위원회를 운영하여 빅데이터 활용을 위한 정책의 기본방향을 검토하여 '빅데이터 활용 방안'을 발표 |
| 유럽 | <ul style="list-style-type: none"> - 2017년 1월부터 30개월간 유럽 내 47개 주요 교통, 물류, IT 관련기업 및 기관으로 구성된 컨소시엄이 1,870억 유로의 예산으로 Transforming Transport 프로젝트를 추진 - 유럽연합은 2015년부터 빅데이터에 대해 향후 5년 동안 산업 컨소시엄의 20억 유로와 공적 자금 5억 유로를 합쳐 총 25억 유로(3조 3,870억원 상당)를 투자 - 2015년부터 아토스(Atos), IBM, 노키아, 솔루션앤네트웍스(Nokia Solutions and Networks, Microsoft), 오렌지(Orange Telecom), SAP, Siemens 등의 기업들과 연구기관들이 주도로 민관협력(public-private partnership, PPP)을 통해 진행 중 |
| 중국 | <ul style="list-style-type: none"> - 2017년에는 '빅데이터 산업 발전계획'을 수립하여 구체적인 육성정책을 제시하였으며, 10개 이상 글로벌 빅데이터 기업, 500개 응용서비스 기업을 육성하기 위해 빅데이터를 핵심산업으로 추진 - 중국 정부는 13·5 계획(2016~2020년)에서 14대 전략 중 하나로 국가 빅데이터 전략을 제시하고 있으며, 빅데이터 산업 육성정책을 본격화 - 2016년 베이징은 '베이징시 빅데이터 및 클라우드 컴퓨팅 발전 행동계획'을 발표하고 향후 5년간 정부 데이터의 민간 개방 확대, 핵심기술 개발 및 응용 확대 등을 통해 빅데이터 산업을 육성할 계획이며, 특히 경제·사회의 광범위한 영역에서의 빅데이터 응용 계획을 제시 |

2.3. 기술개발 현황 및 전망

| | | | | |
|---------|----|---|-------|------------------|
| 기술개발 단계 | 국내 | <input type="checkbox"/> 기초연구 → <input type="checkbox"/> 실험 → <input checked="" type="checkbox"/> 시작품 → <input type="checkbox"/> 제품화 → <input type="checkbox"/> 사업화 | 기술 수준 | 80% (선도국가 대비) |
| | 국외 | <input type="checkbox"/> 기초연구 → <input type="checkbox"/> 실험 → <input type="checkbox"/> 시작품 → <input type="checkbox"/> 제품화 → <input checked="" type="checkbox"/> 사업화 | | |

2.3.1. 국내 기술개발 현황 및 전망

[기반기술]

- (빅데이터 참조 구조 표준) 보편적으로 통용되고 있는 하둡, 아파치 등 기존의 오픈소스 플랫폼을 기반으로 비정형 데이터 수집, 분석, 시각화 및 거버넌스 기능을 제공하는 기술에 대한 개발에 집중
 - (ETRI) 클라우드컴퓨팅 환경에서 데이터 접근 제어 및 효율적 시스템 관리, 자원 할당이 가능하도록 멀티테넌시 기능을 확장한 하둡 플랫폼 기술 개발
 - (데이터스트림즈) 빅데이터와 RDBMicrosoft를 동시에 처리할 수 있는 아키텍처로, SQL 및 CDC, OLAP 기술에 실시간 빅데이터 및 거버넌스 기술을 종합적으로 구성한 TeraONE 플랫폼 제공
 - (IMC) 하둡 기반의 데이터 수집 및 전처리, 텍스트 마이닝 및 시각화에 특화된 솔루션인 텍스트툼 출시
- (클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준) 빅데이터 분석을 활용하고자 하는 업계의 수요가 지속적으로 상승 중이며, 클라우드를 활용하여 빅데이터를 분석 개발 중
 - (솔트룩스) 자연언어처리(NLP)와 기계학습, 분산병렬 처리와 같은 핵심에 기반된 빅데이터 분석 플랫폼 BigO 제공
 - (신테카바이오) 네이버(Naver)와 클라우드 플랫폼을 활용해 유전체 빅데이터를 분석하는 MAP(유전체 빅데이터 분석) 플랫폼을 올리는 작업 진행 중
 - (LF) 클라우드 서비스 플랫폼(네이버 지원) LF의 온라인 쇼핑몰의 고객의 빅데이터 분석 시스템을 개발하고 구축
 - (모비젠) 정부의 GCS(Global Creative SW)사업의 일환으로 퍼블릭 클라우드 기반의 초당 100만 레코드 이상 실시간 인덱싱을 보장하는 멀티테넌시 분산 빅데이터 분석 플랫폼 개발을 진행 중이며, 빅데이터DB 및 분석 플랫폼을 상용 퍼블릭 클라우드 상에서 구현 및 관리할 수 있는 IoT 서비스 인프라의 개발이 목표
 - (인브레인) 정부의 GCS(Global Creative SW)사업의 일환으로 클라우드 환경에서 빅데이터 기반 통합 데이터 분석 스위트 개발

[유통기술]

- (빅데이터 프로비넌스 요구사항 표준) 학계, 출연연 등을 중심으로 데이터의 매쉬업 및 변경 등에 따른 데이터 신뢰성 확보 방법으로서 빅데이터 프로비넌스에 대한 연구가 추진 중
 - (알투웨어) 데이터 변경 요청에 대한 SQL을 분석하여 데이터 변경 이력을 관리하는 솔루션인 'SQL 캔버스 DI(Data Image)를 출시

[관리기술]

- (빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준) 빅데이터 환경에서 데이터의 수집, 등록, 유통, 거래, 관리를 위한 프레임워크 및 개념 모델에 대한 연구가 ETRI를 중심으로 연구를 시작
 - (한국데이터진흥원) 데이터 스토어를 개방형 플랫폼으로 전환하여 공식 거래소를 개장하였으며, 데이터 거래를 위한 기술 개발을 추진 중
 - (KB국민카드) 빅데이터 수요자와 공급자를 연결하는 빅데이터 중개·거래 플랫폼 개발 추진
 - (바이오 빅데이터 플랫폼 사업 추진단) 대학병원 등 40개 기관이 참여하여, 30개 이상의 병원이 제공하는 환자 진료 정보를 거래함으로써 기업들이 바이오 빅데이터 통계를 이용할 수 있게 하기 위한 사업을 추진

[품질관리]

- (빅데이터 품질측정 및 관리 모델 표준) 빅데이터 환경에서 발생하는 대용량 데이터, 실시간 데이터, IoT 데이터 등에서 발생하는 데이터 품질 오류를 자동으로 측정하여 분류하는 기술 개발을 추진 중
 - (한국데이터진흥원) 데이터 품질 지침, 데이터 품질관리 지침, 데이터 품질관리 성과 측정 기법 등 데이터 품질과 관련한 지침 및 가이드라인을 개발·배포하고 있으며, 데이터 품질 관리 인증 실시 중
 - (한국정보화진흥원) 공공정보 품질관리 매뉴얼(2014)을 개발하여 공공기관을 대상으로 공공 데이터 품질관리 개선과 품질진단을 실시 중
 - (위세아이텍) 빅데이터 기반의 지능형 데이터 품질평가 솔루션 개발 중
 - (지티원) 데이터 프로파일링, 데이터 오디팅, 데이터 톨 관리, 데이터 품질 분석, 데이터 품질 분석 결과 보고 등을 제공하는 DQMiner 솔루션을 개발
- (데이터 프로파일링 기반의 빅데이터 품질진단 표준) 비정형 데이터, 대용량 데이터, 센싱 데이터 등 다양한 형태의 빅데이터를 데이터 프로파일링 기반으로 품질을 진단하여 데이터의 활용성을 증대시킬 수 있는 기술 개발을 진행 중
 - (한국데이터진흥원) 정형 데이터 기반의 품질관리 방법 및 성과 측정기법 등 데이터 품질 관련 가이드를 개발·배포하고 있으며, 데이터 품질관리 인증 실시 중
 - (와이즈넷) 인공지능 기반의 빅데이터 분석, 수집 및 검색 솔루션을 개발
 - (엔코아) 모델링 도구 및 메타데이터 관리시스템과 통합 리포지터리가 구성된 솔루션 개발
- (IoT 기반 빅데이터 품질관리 프레임워크 표준) IoT 환경에서는 대량의 데이터가 실시간으로 발생. 이러한 IoT 데이터를 처리를 위해 인공지능을 적용하며 기존 플랫폼에 관련 기술을 추가하는 중
 - (지티원) DQMiner의 데이터 처리, 데이터 품질관리 기능에 IoT 데이터 처리 기능을 추가하는 중

<국내 주요 사업자 서비스 동향>

| 사업자 | 주요 현황 |
|-------|--|
| 위세아이텍 | <ul style="list-style-type: none"> - 인공지능 기반의 빅데이터 분석과 머신러닝 프로세스, 데이터 품질관리 및 빅데이터 분석 플랫폼 사업을 수행 · 빅데이터를 세부적으로 탐색하는 시각화 도구, 머신러닝 프로세스 자동화 도구, 데이터 품질관리 등 인공지능 기반의 빅데이터 분석, 관리를 위한 다양한 제품군을 보유(WISE Prophet, WISE Intelligence, WISE DQ, WISE Meta, WISE META) |

2.3.2. 국외 기술개발 현황 및 전망

[기반기술]

- (빅데이터 참조 구조 표준) 기존의 데이터 분석 및 관리 플랫폼을 빅데이터 환경에 적용 가능하도록 기능을 확장하고, 소프트웨어와 함께 하드웨어까지 결합된 솔루션들을 출시
 - (Oracle) 빅데이터 환경에 요구되는 시스템의 개념적 아키텍처와 함께 이를 자사의 솔루션들로 제공하기 위한 빅데이터 분석 참조 아키텍처를 발표
 - (IBM) 빅데이터 참조 아키텍처를 바탕으로 기업에서 빅데이터 도입 시 요구되는 소프트웨어와 하드웨어를 패키지화하여 출시
 - (기타) Google, 아파치, 클라우데라 등 오픈소스 진영을 중심으로 빅데이터 생태계를 구성하고, 개별 솔루션들 간의 상호운용성을 확보하는 비즈니스 모델 추진
- (클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준) 클라우드 서비스 제공 업체들은 빅데이터 데이터 저장소뿐만 아니라, 대용량 데이터의 분석 및 관리를 위한 기능들을 개발 제공하는 추세
 - (트레저데이터) 클라우드 기반 빅데이터 관리 서비스 기업으로, 데이터 수집, 저장, 실시간 예측 분석, BI툴과의 연동을 제공 중
 - (Google) 클라우드 플랫폼의 빅데이터 분석 기술인 '클라우드 데이터플로우(Cloud Dataflow)'를 서비스 중이며, 페타 바이트급의 데이터 저장 및 분석용 클라우드 서비스 개발하여 제공 중
 - (Amazon) EMR(Elastic MapReduce)를 개발하여 AWS 기반 대량의 데이터를 신속하고 비용 효율적으로 처리할 수 있도록 관리형 하둡 프레임워크를 제공 중이며, EMR은 Apache Spark, HBase, Presto 및 Flink와 같은 오픈 소스 프레임워크를 실행할 수 있도록 제공. 대화식 쿼리 서비스로 저장된 데이터를 분석하는 Amazon Athena 등 다양한 서비스를 제공 중
 - (페이스북) 대용량 빅데이터 분석처리 엔진 '프레스토'를 오픈소스 소프트웨어로 개발

[유통기술]

- (빅데이터 프로비넌스 요구사항 표준) 데이터 프로비넌스를 지원하는 솔루션들이 몇몇 발표된바 있으며, 학계 및 연구소를 중심으로 빅데이터 환경에서의 데이터 프로비넌스 지원과 관련된 학술 연구들이 진행 중
 - (Intel & SPARKL) 금융 거래 환경에서의 데이터 프로비넌스를 지원하는 솔루션 발표
 - (Oracle) W3C의 데이터 프로비넌스 표준(PROV)을 만족하는 Oracle Fusion Advanced Controls 발표

[관리기술]

- (빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준) EU를 중심으로 W3C의 DCAT과 DCAT-AP를 확장하여 EU 역내 국가 간 메타데이터 교환을 위해, StatDCAT-AP, TransportDCAT-AP 등의 다양한 DCAT-AP 확장(extension)이 개발 중
 - (Apache Foundation) 2018년 2월 현재 하둡에서의 데이터 거버넌스와 자동화된 메타데이터 솔루션을 제공하는 Atlas(버전 1.0-alpha)를 배포
 - (EU Data Portal) 빅데이터를 포함한 데이터의 유통을 위하여 DCAT-AP 확장과 시맨틱 기반 메타데이터를 지원하는 데이터 포털을 운영 중

[품질관리]

- (빅데이터 품질측정 및 관리 모델 표준) 빅데이터 중 비정형 데이터는 조직에서 보유한 데이터의 70 ~ 80% 이상을 차지하는 것으로 추산되며 이는 데이터 품질 도구 시장의 중요한 성장 영역으로 각광받고 있으며, 특히 텍스트 데이터, 사물인터넷 데이터 등 비정형 데이터를 대상으로 데이터품질 도구를 적용하는데 관심이 증가
 - (인포매티카) 사물인터넷 등 대용량 데이터 분석, 데이터 거버넌스 및 콘텐츠 데이터 분석, 기계학습, 알고리즘 및 예측 분석을 적용한 기술 개발
 - (IBM) Apache Kafka 오픈 소스를 활용하여 사물인터넷 데이터 지원, 기계학습 지원 기술 개발
- (데이터 프로파일링 기반의 빅데이터 품질진단 표준) 비정형, 대용량의 빅데이터 품질을 진단·관리하는데 있어 빅데이터의 특성이 반영된 데이터 프로파일링 방법을 기반으로 빅데이터 품질진단 수행이 데이터 자본화 시장에서 중요한 영역으로 확대될 것으로 판단되고 있으며, 빅데이터(IoT 데이터 등)의 품질을 측정하기 위한 전처리 단계에 기능을 추가하고 있는 상황
 - (SAS) In-Stream 분석 지원과 기계학습 및 데이터 준비 기능을 제공
 - (인포매티카) 인텔리전스 디스커버리 기능을 추가하여, 기계학습에 의해 생성된 로그성 IoT 데이터를 파싱하는 작업을 시각화할 수 있는 기능을 제공
 - (Trillum Sofeware) 클라우드 기반 배치 기능 제공(SaaS) 및 마이그레이션 기능 제공
- (IoT 기반 빅데이터 품질관리 프레임워크 표준) 센서로부터 수집되는 데이터의 처리와 분석, 딥 러닝 기반의 인공지능 알고리즘을 적용하여 IoT 데이터의 패턴을 인식하여 데이터 품질을 높이는 기술이 개발 중
 - (Intelligent Environment) ISO 8000-62(데이터 품질 프로세스 성숙도 평가) 표준을 기반으로 데이터 품질관리 모델의 현장 적용 기술 제공
 - (LUCENTIA LAB SL) 센서 데이터 분석 및 알고리즘 설계를 통해 IoT 데이터 분석 기술 제공

2.4. IPR 현황 및 전망

○ 특허분석 개요

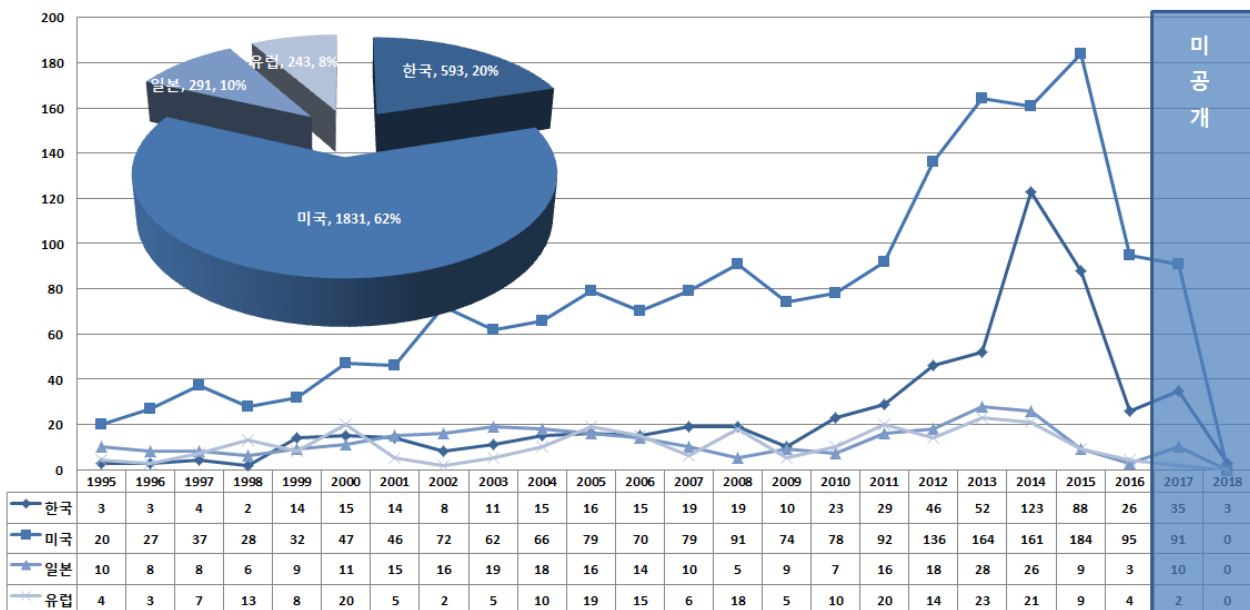
- 빅데이터 분야에 있어서, 2018년 4월 현재까지 한국, 미국, 일본 및 유럽에 공개된 특허들을 대상으로 2017년도 보다 166건이 증가한 2,958건의 특허들을 대상으로 분석을 수행함

○ 연도별 출원동향

| 연도 | '96 | '97 | '98 | '99 | '00 | '01 | '02 | '03 | '04 | '05 | '06 | '07 | '08 | '09 | '10 | '11 | '12 | '13 | '14 | '15 | '16 | '17 | '18 | 합계 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 출원건수 | 41 | 56 | 49 | 63 | 93 | 80 | 98 | 97 | 109 | 130 | 114 | 114 | 133 | 98 | 118 | 157 | 214 | 267 | 331 | 290 | 128 | 138 | 3 | 2958 |

- 빅데이터 기술과 관련된 출원은 2004년부터 점차적으로 증가하기 시작하였고, 2013년을 기점으로 급증하고 있는 것으로 조사되었으며, 2015년 이후 출원량이 감소하는 모습을 보이고 있으나 미공개구간인 2017년도에 다시 증가하는 것으로 조사되어 급격한 증감 없이 꾸준한 특허출원이 예상됨

○ 특허 출원년도별 특허공보별 동향



- 빅데이터 기술과 관련하여 출원국가 DB(한국, 미국, 일본, 유럽)별로 공개 및 등록된 특허는 미국특허가 1,831건(62%)으로 가장 많은 특허가 출원되고 있으며, 그 다음으로 한국특허 593건(20%), 일본특허 291건(10%), 유럽특허 243건(8%) 순으로 특허출원이 이루어지고 있음
- 미국의 경우 1990년대 중반부터 특허출원이 이루어져 꾸준한 특허출원량을 보이고 있으며, 2012년 들어 특허출원량이 증가하는 양상을 보이고 있음, 한국 역시 2012년 들어 출원량이 증가하기 시작하였고 2014년 급격한 출원 증가 양상을 보임, 미공개구간이 2017년의 경우 미국은 다소 감소하였으나 한국, 일본은 증가하는 것으로 조사되어 향후 특허 출원량이 다시 증가할 것으로 예상됨

○ 한국특허에서의 주요 출원인별 출원 현황

| 순위 | 출원인 | 출원건수 |
|----|------------|------|
| 1 | ETRI | 39 |
| 2 | KISTI | 23 |
| 3 | KT | 18 |
| 4 | SAMSUNG | 17 |
| 5 | SK Telecom | 11 |
| 6 | KAIST | 9 |
| 7 | IBM | 8 |
| 8 | 주시엠아이코리아 | 8 |
| 9 | 고려대학교 | 7 |
| 10 | 경희대학교 | 6 |
| 11 | 주비아이매트릭스 | 6 |
| 12 | 기타 | 441 |
| 합계 | | 593 |

- 한국 특허청으로는 ETRI가 39건으로 가장 많은 특허출원을 수행하고 있는 것으로 조사되었으며, 그 뒤를 KISTI, KT, SAMSUNG, SK Telecom 등 다수의 특허출원을 수행한 것으로 조사 되었고, 대학교로는 고려대학교, 경희대학교가 7건, 6건의 특허를 출원하고 있음
- 해외 기업으로는 IBM이 유일하고 국내에 8건의 특허출원을 진행하였음

○ 해외특허에서의 주요 출원인별 출원 현황

| 순위 | 출원인 | 출원건수 |
|----|--------------------------|------|
| 1 | IBM | 193 |
| 2 | HITACHI | 138 |
| 3 | MICROSOFT | 66 |
| 4 | NEC | 42 |
| 5 | SPLUNK | 40 |
| 6 | FUJITSU | 38 |
| 7 | SIEMENS | 34 |
| 8 | NTT | 30 |
| 9 | EMC | 29 |
| 10 | FISHER-ROSEMOUNT SYSTEMS | 24 |
| 11 | GE | 23 |
| 12 | HEWLETT-PACKARD | 23 |
| 13 | TOSHIBA | 23 |
| 14 | GOOGLE | 20 |
| 15 | ORACLE INTERNATIONAL | 20 |
| 16 | 기타 | 1622 |
| 합계 | | 2365 |

- 미국, 유럽, 일본 등 해외의 주요 시장에서는 IBM, HITACHI, Microsoft, NEC, SPLUNK 등 기업이 가장 많은 출원을 하고 있으며, 특히 IBM과 HITACHI는 100건 이상의 특허출원을 진행한 것으로 나타났음
- 가장 많은 특허를 출원 중인 IBM은 국내에서도 다수의 특허를 출원 중에 있으며, HITACHI, NEC, NTT, FUJITSU 등 일본 기업이 빅데이터 관련 기술에 많은 투자를 진행하고 있는 것으로 조사되었음

2.5. 표준화 현황 및 전망

| | | | |
|-----------|--|----------|----------------------|
| 표준화 특성 | <input type="checkbox"/> 개념/정의, <input checked="" type="checkbox"/> 유즈케이스/요구사항, <input type="checkbox"/> 기능/참조구조, <input type="checkbox"/> 데이터포맷/스키마, <input type="checkbox"/> 프로토콜/인터페이스 | 표준 수준 | 100% (선도국가 대비) |
|-----------|--|----------|----------------------|

| 구분 | 표준화 기구 | | 표준화 현황 |
|------------|-------------|-----------|--|
| 국제 (공적) | JTC1 | SC32 | (WG1-eBusiness, WG2-MetaData, WG3-Database language) 데이터 베이스 언어, 멀티미디어 데이터, 메타데이터 등 데이터 관리 및 교환 관련 표준 개발 중 |
| | | SC42 | (Artificial intelligence) 2017년까지 JTC1 WG9에서 참조 아키텍처에 대해서 다섯 개의 개별 문서로 분리(개요 및 적용 절차, 유즈케이스 및 요구사항, 참조 아키텍처, 보안 요구사항, 표준 로드맵)되어 각각 개발 진행하였으나, 2018년 SC42로 이관되어 진행 중 |
| | ISO | TC184 SC4 | (WG13-Industrial Data Quality) 데이터 품질관리 평가 프로세스 참조 모델 표준, 데이터 품질관리 평가 프레임워크 표준, 데이터 품질관리 성숙 모델 표준 개발 중 |
| | ITU-T | SG13 | (Q19-End-to-end cloud computing management, cloud security and big data governance) 클라우드 에코시스템, 인터클라우드와 요구사항, 클라우드 기반의 빅데이터 서비스 요구사항, 클라우드 기능 아키텍처, 인프라와 네트워크, 종단 간 클라우드 자원 관리 요구사항 등의 표준 개발 |
| 국제 (사실) | W3C | | - DCAT, DCAT-AP의 데이터 교환을 위한 메타데이터 모델 및 유즈케이스 표준이 개발되고 있으며, EU를 중심으로 다양한 DCAT-AP 확장 개발 |
| 국내 | TTA | SPG22 | (빅데이터) 빅데이터 참조모형, 유즈케이스 및 요구사항 등 20여건의 표준을 제정하였으며, 관련 기술에 대한 국내 표준화 추진 중 |
| | | PG606 | (메타데이터) 정보식별 및 정보교환을 위한 메타데이터 표준, 데이터의 효율적 관리를 위한 데이터 거버넌스 및 빅데이터 수집, 저장, 분석, 관리와 관련한 메타데이터 표준을 국내 표준화로 제정 추진 |
| | All@CLOUD포럼 | | - 2018년 기존 클라우드컴퓨팅표준화포럼을 All@CLOUD포럼으로 명칭을 변경하고 클라우드 뿐만 아니라 빅데이터 분야 표준개발 추진 |

2.5.1. 국내 표준화 현황 및 전망

[기반기술]

- (빅데이터 참조 구조 표준) 빅데이터 생태계 모델을 기반으로 빅데이터 서비스 구현을 위한 기능 요소들을 정의하는 빅데이터 참조 구조에 대한 표준을 TTA 빅데이터 SPG(SPG22)를 통해 2016년 개발 완료. ITU-T 및 JTC1 SC42에서 관련 국제 표준이 개발 중으로 그 결과에 따른 국내 표준의 개정 논의가 필요
- (TTA 빅데이터 SPG(SPG22)) 2014년 착수하여 2016년 표준 제정

<국내 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|--|------|---------------|
| TTA SPG22 | TTAK.KO-10.0973, 빅데이터 프레임워크 - 제2부: 참조 아키텍처 | 2016 | 빅데이터 참조 구조 표준 |
| | TTAK.KO-10.0705, 클라우드 기반 빅데이터 서비스를 위한 프레임워크 정의 | 2013 | |
| | TTAK.KO-10.0706, 클라우드 기반 빅데이터 서비스 참조 모델 | 2013 | |

- (클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준) ETRI에서 주도하여 국제표준과 연동된 국내표준화 작업을 진행 중. 세부적으로는 공적표준화 기구인 ITU-T와 연동된 표준을 작업 중에 있으며, 국내 산학의 의견을 포럼 및 TTA 빅데이터 SPG(SPG22)을 통해 수용 중
- (TTA 빅데이터 SPG(SPG22)) 2015년 11월에 ITU-T에서 Y.3600표준을 제정 완료됨에 따라 Y.3600을 준용하는 국내 단체표준 개발하여 2016년 12월에 제정완료 하였으며, 이를 KS 표준으로 추진 예정
- (ETRI) 국제표준화 기구(ITU-T)의 Q18/13에서 관련 표준화 작업(Y.BDaaS-arch)에 참여 중에 있으며, TTA 빅데이터 SPG(SPG22)에 참여하여, 개발 현황을 공유 중

<국내 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|---|------|-------------------------------|
| TTA SPG22 | TTAE.IT-Y.3600, “빅데이터 - 클라우드 컴퓨팅 기반의 요구사항 및 기능” | 2016 | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준 |

[유통기술]

- (빅데이터 프로비넌스 요구사항 표준) ITU-T SG13에서 통해 빅데이터를 위한 데이터 프로비넌스 요구사항 표준(ITU-T Y.bdp-reqts)이 ETRI 주도로 개발 중이며, 2018년 개발 완료를 목표로 함에 따라 2019년에 국내 단체 표준 도입을 검토

- (TTA 빅데이터 SPG(SPG22)) 프로비넌스의 구성 항목 중 데이터 분석 모델에 대한 XML 기반의 서술 언어인 PMML V.4.2.1dp 대한 국제표준을 2016년 TTA 단체 표준으로 수용
- (ETRI) 국제 표준화 기구(ITU-T)의 Q17/13에서 관련 표준화 작업(Y.bdp-reqts)에 참여 중

<국내 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|---|------|--------------------|
| TTA SPG22 | TTAE.OT-10.0418, 데이터 분석 모델 공유를 위한 크업 언어 V.4.2.1 | 2016 | 빅데이터 프로비넌스 요구사항 표준 |

[관리기술]

- (빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준) 빅데이터의 상호운용성을 지원하기 위한 메타데이터 프레임워크 및 개념모델에 대한 표준화 요구가 증대되고 있음
 - (TTA 메타데이터 PG(PG606)) 2015년부터 빅데이터 수집, 저장, 분석, 관리 관련 메타데이터 표준을 주요 과제로 하고 있으며, 2018년도에는 구체적인 표준화 작업이 진행될 것으로 예상
 - (TTA 빅데이터 SPG(SPG22)) 웹 환경에서 테이블 데이터를 위한 메타데이터 스키마 및 데이터 카탈로그 작성 및 교환을 위한 개념 모델 표준화를 추진해왔으며, 국내 빅데이터 유통 사업 등 관련 사업과 연계한 표준화 논의가 진행 중

<국내 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|--|------|------------------------------------|
| TTA PG606 | TTAK.KO-10.0976, 연구데이터 관리 및 공유를 위한 메타데이터 | 2017 | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 |
| | TTAK.KO-11.0224, 과학 실험 데이터(HDF5) 관리와 교환을 위한 메타데이터 | 2016 | |
| TTA SPG22 | TTAE.OT-10.0427, 데이터 카탈로그 어휘 | 2017 | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 |
| | TTAE.OT-10.0417, 웹 환경에서 테이블 데이터를 위한 메타데이터 어휘 | 2016 | |
| | TTAE.OT-10.0416, 웹 환경에서의 테이블형 데이터를 위한 데이터 모델 및 메타데이터 | 2016 | |

[품질관리]

- (빅데이터 품질측정 및 관리 모델 표준) 정형·비정형 등의 데이터 품질관리를 위한 데이터 기반 요소를 정의, TTA 메타데이터 PG(PG606)를 통해 표준을 개발 중. 빅데이터, 인공지능, IoT 등의 이슈로 인해 빅데이터 품질을 정의하기 위해 메타데이터 표준화에 대한 요구가 증대. 국가기술표준원 산업데이터 전문위원회에서는 산업데이터 전반에 걸친 표준화를 모니터링하며 산업데이터 관련 표준을 KS로 제정

- (TTA 메타데이터 PG(PG606)) 2009년부터 데이터품질 분야 표준을 제정

<국내 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|---------------------------------|------|-------------------------|
| TTA PG606 | TTAK.KO-10.0475, 데이터 품질 특성 | 2010 | 빅데이터 품질측정 및 관리 모델 표준 |
| | TTAK.KO-10.0474, 데이터 품질 관리 프로세스 | 2010 | |
| | TTAK.KO-10.0473, 데이터 관리 프로세스 | 2010 | |

- (데이터 프로파일링 기반의 빅데이터 품질진단 표준) 데이터 관리 프로세스 및 요구사항이 명확하지 않으며 실시간으로 생산되는 비정형, 텍스트 및 대용량 빅데이터의 품질을 진단하기 위한 방법에 대해 학계 및 산업계 등에서 논의되고 있으나 현재 표준화는 진행되고 있지 않음
- (IoT 기반 빅데이터 품질관리 프레임워크 표준) 센서 디바이스에서 발행하는 다양한 유형의 데이터 오류를 분류하며 데이터 품질에 영향을 미치는 특성을 규명하고, 대용량 데이터처리에 적합한 프로세스를 정립
 - (지티원, 2e컨설팅, 명지대, 강릉원주대) 산학 공동으로 데이터 품질 특성 및 데이터 품질 관리 프로세스의 이론을 데이터 품질관리 플랫폼으로 구현하는 기술을 개발 중

2.5.2. 국제 표준화 현황 및 전망

[기반기술]

- (빅데이터 참조 구조 표준) JTC1 SC42/SC27 및 ITU-T SG13은 각각 빅데이터 참조 구조에 대한 개발을 시작하여 진행 중
 - (JTC1 SC42) 2014년에 JTC1 WG9을 통해 ISO/IEC 20547 빅데이터 - 참조 아키텍처 개발을 시작하였으며, 2016년에는 다섯 개의 개별 문서로 분리(개요 및 적용 절차, 유즈케이스 및 요구사항, 참조 아키텍처, 보안 요구사항, 표준 로드맵)되어 각각 개발 진행, 2018년 SC42로 진행 중인 2건의 과제가 이관 완료
 - (JTC1 SC27) 2016년 JTC1 WG9에서 ISO/IEC 20547-4 보안 요구사항 항목이 이관되어 문서 개발 진행 완료
 - (ITU-T SG13) 2017년 2월 회의를 통해 개발 중인 클라우드 기반의 빅데이터 아키텍처를 빅데이터 참조 아키텍처와 클라우드 기반 아키텍처의 두 가지 문서로 분리하여 개발을 진행할 것이 논의되어, ITU-T Y.bd-arch 과제 시작

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|------------|--|-------------|---------------|
| JTC1 SC27 | ISO/IEC 20547-4, Information technology - Big data - Reference architecture - Privacy and security fabric | 진행 중 (2020) | 빅데이터 참조 구조 표준 |
| JTC1 SC42 | ISO/IEC 20547-3, Information technology - Big data - Reference architecture - Reference architecture | 진행 중 (2019) | 빅데이터 참조 구조 표준 |
| | ISO/IEC 20547-1, Information technology - Big data - Reference architecture - Overview and application process | 진행 중 (2019) | |
| | ISO/IEC 20547-5, Information technology - Big data - Reference architecture - Standards roadmap | 2017 | |
| | ISO/IEC 20547-2, Information technology - Big data - Reference architecture - Use cases and derived technical considerations | 2017 | |
| ITU-T SG13 | ITU-T Y.bd-arch, Big data - Reference architecture | 진행 중 (2019) | 빅데이터 참조 구조 표준 |

- (클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준) ITU-T에서 클라우드컴퓨팅 기반 빅데이터 요구사항(Y.3600)을 기반으로 참조 구조 표준(Y.BDaaS-arch) 개발 중. 현재 한국, 중국, 프랑스, 폴란드에서 개발을 주도 중
- (ITU-T SG13 Q17) 2015년 11월에 빅데이터 기반 서비스 기능 및 아키텍처의 기반이 되는 Y.3600(Big data - Cloud computing based requirements and capabilities) 표준을 제정 완료
 - (ITU-T SG13 Q18) 빅데이터 기능 및 구조 표준(Y.BDaaS-arch, “Big data - Functional architecture of Big Data as a Service”) 표준 권고안 개발 작업을 진행 중

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|------------|---|-------------|-------------------------------|
| ITU-T SG13 | Y.BDaaS-arch, “Big data - Functional architecture of Big Data as a Service” | 진행 중 (2019) | 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준 |
| | Y.3600, “Big data - Cloud computing based requirements and capabilities” | 2015 | |

[유통기술]

- (빅데이터 프로비넌스 요구사항 표준) 빅데이터 교환 시, 데이터 출처 및 데이터 처리에 사용된 기능 정보 등을 추적하여 그 이력을 관리하기 위한 표준화가 ITU-T SG13을 통해 진행 중
- (ITU-T SG13) 2016년 6월 회의를 통해 2018년 12월 표준 제정을 목표로 ‘빅데이터 프로비넌스 요구사항(Y.bdp-reqts, Big data - Requirements for data provenance)’ 표준 개발이 시작

- 되었으며, 빅데이터 환경에서의 프로비넌스 개념 정의와 유즈케이스 기반의 요구사항들을 정의
- (W3C) 단일 도메인 내에서의 프로비넌스 정보 공유를 위한 표준인 PROV 모델을 개발
- 4종의 표준 모델에 대한 개발을 완료

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|------------|--|-------------|--------------------|
| ITU-T SG13 | Y.bdp-reqts, Big data - Requirements for data provenance | 진행 중 (2019) | 빅데이터 프로비넌스 요구사항 표준 |
| W3C | PROV-O: The PROV Ontology | 2013 | 빅데이터 프로비넌스 요구사항 표준 |
| | PROV-DM: The PROV Data Model | 2013 | |
| | PROV-N: The Provenance Notation | 2013 | |
| | Constraints of the PROV Data Model | 2013 | |

[관리기술]

- (빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준) W3C는 웹상에서의 데이터 카탈로그 교환 표준에 대한 확장 프로파일 및 개정안 개발을 진행 중이며, ITU-T SG13은 빅데이터 유통을 위한 메타데이터 정의를 위한 표준 개발을 시작
- (W3C) DCAT, DCAT-AP의 데이터 교환을 위한 메타데이터 모델 및 유즈케이스 표준이 개발되고 있으며, EU를 중심으로 다양한 DCAT-AP 확장이 개발
- (ITU-T SG13) 2017년부터 이통사들을 중심으로 빅데이터 유통에 필요한 메타데이터 프레임워크 및 개념 스키마 개발을 추진 중

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|------------|---|-------------|------------------------------------|
| ITU-T SG13 | Y.bdm-sch, Big data - Metadata framework and conceptual model | 진행 중 (2020) | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 |
| W3C | Dataset Exchange Use Cases and Requirements | 2017 | 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준 |
| | Data Catalog Vocabulary (DCAT) | 2014 | |

[품질관리]

- (빅데이터 품질측정 및 관리 모델 표준) 빅데이터 활용에 있어서 데이터 품질 및 관리가 중요한 요인으로 부각됨에 따라, 빅데이터의 품질을 측정하기 위한 빅데이터 품질요인, 품질진단 방안 등에 대한 표준 개발이 진행 중

- (JTC1 SC32 WG2) 메타데이터 개념(Information technology - Concepts and usage of metadata - Metadata concepts) 표준 제정
- (ISO TC184 SC4 WG13) WG13은 데이터 품질 표준(ISO 8000)을 만들고 있으며 최근에는 데이터 품질 관리 프로세스 평가 모델 표준이 제정 중

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|--------------------|---|------|----------------------|
| JTC1 SC32 | ISO/IEC TR 19583-1, Information technology - Concepts and usage of metadata - Metadata concepts | 2018 | 빅데이터 품질측정 및 관리 모델 표준 |
| ISO TC184 SC4 WG13 | ISO 8000-61, 데이터 품질관리 프로세스 참조모델 | 2016 | |

- (데이터 프로파일링 기반의 빅데이터 품질진단 표준) 실시간으로 생산되는 비정형, 텍스트 및 대용량 빅데이터의 활용성이 높아짐에 따라 생산되는 데이터의 품질을 진단하기 위해 빅데이터 품질요인, 관리 방안 등에 대한 표준을 개발 중
- (JTC1 SC32 WG2) 메타데이터 작업반에서 빅데이터 품질 AHG(Ad-Hoc Group)을 만들어 빅데이터 품질을 정의하는 메타데이터 요소에 대한 기념 개념 표준을 개발 중
- (ISO TC184 SC4 WG13) WG13은 데이터 품질 표준(ISO 8000)을 만들고 있으며 최근에는 데이터 품질진단 모델 표준을 개발 중

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|---|-------------|----------------------------|
| JTC1 SC32 | Data quality assessment based on data profiling | 진행 중 (2020) | 데이터 프로파일링 기반의 빅데이터 품질진단 표준 |

- (IoT 기반 빅데이터 품질관리 프레임워크 표준) IoT 기반의 제조환경에서 발생하는 데이터를 처리하기 위한 표준화 진행 중. 특히 디지털 트랜스포메이션을 중심으로 한 스마트 제조에서의 데이터 품질 기술의 수요가 증가할 전망
- (ISO TC184 SC4 WG13) 데이터 품질 기술을 스마트 제조에 적용하기 위한 표준 개발 진행 중

<국제 표준화 현황>

| 개발기구 | 표준(안)명 | 개발연도 | 관련 중점 표준화 항목 |
|-----------|--|-------------|----------------------------|
| ISO TC184 | Data quality management framework based on IoT | 진행 중 (2020) | IoT 기반의 빅데이터 품질관리 프레임워크 표준 |

2.6. 오픈소스 현황 및 전망

○ 하둡(Hadoop)

- 대용량의 데이터를 저장하고 처리할 수 있는 분산 시스템으로 HDFS, 맵리듀스, YARN으로 구성
- 하둡-분산파일시스템(Hadoop-DFS: HDFS): HDFS는 데이터를 다수의 노드에 분산해서 저장
 - 스케일아웃에 용이하고 전통적인 스토리지의 저장 용량의 한계 초과 가능
 - HDFS는 대용량 파일의 스트리밍 읽기와 쓰기에 뛰어난 성능
- 하둡-맵리듀스(Hadoop-MapReduce): 맵리듀스는 프로그래밍 모델과 프레임워크 구현 환경으로 구성
 - 맵리듀스는 대용량, 분산, 내고장성 데이터 처리 프로그램을 개발자가 쉽게 개발하도록 설계한 최적의 애플리케이션 기술
 - 개발자가 맵리듀스에서 맵 함수와 리듀스 함수를 조합하여 코드를 작성하면, 프레임워크는 작업의 병렬처리, 워커 머신의 태스크 스케줄링, 모니터링, 장애 복구 등의 기능 제공
- 하둡-얀(Hadoop-YARN): 리소스 관리 플랫폼
 - YARN은 하둡 클러스터의 각 어플리케이션에 필요한 리소스를 할당하고 모니터링 하는 업무에 집중함으로써 다양한 어플리케이션이 하둡 클러스터의 리소스를 공유할 수 있도록 하는 핵심 요소

○ 하둡 에코시스템 - 수집 기술

- 하둡 기반의 데이터 수집 기술로 대용량의 데이터를 저장하고 처리할 수 있는 소프트웨어로는 플룸, 스쿱, 카프카 등이 있음
- 플룸(Flume): 플룸은 분산로그 관리 시스템
 - 대용량 로그데이터를 중앙집중화된 데이터 저장소로 효율적으로 수집, 통합, 저장하기 위한 시스템
 - 에이전트가 설치된 머신에서 데이터 소스를 수집하여 하둡 분산 환경의 콜렉터에 전송하고 이를 로컬파일시스템, 하둡-분산파일시스템, 하이브(Hive), HBase 등에 저장하는 데이터 수집 엔진
 - 다수의 웹 서버에서 로그파일을 수집하고 해당 파일의 로그 이벤트를 처리하기 위해 HDFS에 위치한 새로운 통합 파일로 옮기는 용도로 사용
- 스쿱(Sqoop): 스쿱은 구조화된 데이터 저장소에서 데이터를 추출해서 하둡으로 보내 처리할 수 있도록 해주는 오픈 소스 도구
 - 데이터 추출 시에 맵리듀스 프로그램이나 하이브 같은 다른 고차원 도구로도 사용 가능
 - 또한 관계형 데이터베이스의 데이터를 HBASE로 옮기는 데 스쿱을 사용할 수 있으며, 분석 파이프라인을 거쳐 최종 산출물이 나오면 스쿱은 이 산출물을 다시 데이터 저장소로 보내 다른 사용자가 사용 가능
- 카프카(Kafka): 카프카는 확장성이 좋고 처리량이 높은 분산 메시지 시스템
 - 우수한 메시지 전달 성능을 보장하여 실시간 데이터나 오프라인 데이터를 처리하고 수집하는데 사용
 - 카프카는 높은 처리량을 유지하면서 병렬처리에 초점을 맞춘 구조이며 다량의 클라이언트 생성이 가능한 유연성을 제공하고 분산처리를 지원

○ 하둡 에코시스템 - 저장 및 처리 기술

- 하둡 기반의 데이터 저장 및 처리 기술로는 피그, 하이브가 있음
- HDFS에 저장된 데이터를 SQL 혹은 SQL과 유사한 형태로 처리를 요청하고 분산 처리하는 시스템을 통칭하는 것이 SQL On Hadoop가 있으며, 국내에서 개발한 타조와 클라우데라의 임팔라 등 다양한 도구가 있음
- 피그(Pig): 피그는 고수준 처리 요청 언어
 - 피그는 스크립트 언어로 코드를 작성하면 피그 엔진은 이를 맵리듀스 프로그램으로 변환하고 하둡 상에서 실행하고 그 결과를 알려줌
- 하이브(Hive): 하이브는 하둡 기반의 데이터 웨어하우스 프레임워크
 - 빠른 속도로 성장하는 소셜 네트워크에서 매일같이 생산되는 대량의 데이터를 관리하고 학습하기 위해 개발
 - 하이브는 데이터의 스키마를 관리할 수 있으며, 스크립트 언어가 아닌 SQL 언어를 지원하므로 기존 데이터베이스 프로그래머들이 쉽게 하이브 쿼리를 익혀서 빨리 개발 가능
 - 이러한 특징 때문에 하이브는 하둡 기반의 데이터웨어하우스로 불림
- 테즈(Tez): 하이브를 발전시키고 성능을 강화한 분석 엔진
 - 하둡 2.0부터 맵리듀스를 테즈를 기반으로 개발
- 드릴(Drill): Google의 Dremel(일명 BigQuery)의 오픈소스 버전으로 높은 확장성과 하둡/NoSQL과의 쿼리 인터페이스 지원
- SparkSQL(Shark): 스파크(Spark)는 인메모리 기반의 초고속 분산병렬처리 프레임워크로 SparkSQL은 하이브에 저장된 데이터에 SQL 질의 가능
- 임팔라(Impala): 하이브 사용자를 고려한 빠른 SQL 쿼리 엔진
 - HDFS나 HBase에 저장된 데이터를 대상으로 분석, 문법은 HiveQL과 동일
- 타조(Tajo): 타조는 하둡 기반의 대용량 데이터웨어 하우스 시스템
 - 빅데이터 분석 엔진도 제공
- 프레스토(Presto): 페이스북에서 개발했으며, 빠른 속도의 검색과 집계 가능

○ 하둡 에코시스템 - 분석 기술

- 하둡 기반의 분석 기술로는 머하웃과 스파크가 있음
- 스파크(SPARK): 스파크는 대용량 데이터 처리를 위한 클러스터 컴퓨팅 프레임워크
 - 스파크는 실행 엔진으로 맵리듀스를 사용하지 않고, 대신 클러스터 기반으로 작업을 실행하는 자체 분산 런타임 엔진이 있음
 - 스파크는 하둡과 밀접하게 통합되어 있어서 YARN기반으로 실행할 수 있고, 하둡 파일 포맷과 HDFS 같은 기반 저장소를 지원
- 머하웃(Mahout): 머하웃은 추천시스템, 기계 학습 등 데이터 마이닝 알고리즘을 쉽게 구현하도록 지원하는 솔루션으로 하둡을 기반으로 구현

○ 하둡 클러스터 관리 및 운영 기술

- 하둡 클러스터를 관리하고 운영하는 기술로는 주키퍼, 우지 등이 있음
- 주키퍼(Zookeeper): 주키퍼는 분산 환경에서 다수의 머신과 서비스간의 상호 조정 역할을 제공
 - HBase의 필수 구성요소이며, 하둡 2.0 버전의 네임노드 이중화에도 사용

- 주키퍼의 핵심은 분산 데이터 저장소로 다수의 머신에 동일한 데이터를 저장하며 고가용성과 빠른 응답속도를 제공
- 주키퍼의 데이터 동기화와 클라이언트와의 세션관리 및 이벤트 감시자 기능을 활용하면 분산 시스템에서 다양한 머신과 서비스의 상호조정이 가능
- 에이브로(Avro): 에이브로는 특정 언어에 종속되지 않는 언어 중립적 데이터 직렬화 시스템
 - 하둡의 창시자인 더그 커팅이 하둡 Writable(직렬화 방식)의 주요 단점인 언어 이식성을 해결하기 위한 시스템
 - 에이브로는 현재 C, C++, C#, 자바, 자바스크립트, 펄, PHP, 파이썬, 루비 등의 언어로 처리할 수 있는 새로운 데이터 포맷을 제공하기 때문에 단일 언어에 얽매이지 않고 다양한 언어로 데이터셋을 쉽게 공유 가능
- 우지(Oozie): 우지는 하둡의 workflow 스케줄러 시스템
 - 하둡과 하둡 에코시스템의 작업을 관리하는 워크 플로우 엔진이자 코디네이터 시스템으로 다양한 하둡 에코시스템의 작업을 주기적으로 실행하고 그 결과를 통보
- 크런치(Crunch): 크런치는 맵리듀스 파이프라인을 작성하는 고수준 API
 - 맵리듀스 대신 크런치를 사용하면 String이나 POJO(plain old Java object)와 같은 프로그래머에게 익숙한 자바 자료형, 풍부한 데이터 변환 기능, 여러 단계의 파이프라인에 집중할 수 있다는 장점

○ R - 분석 기술

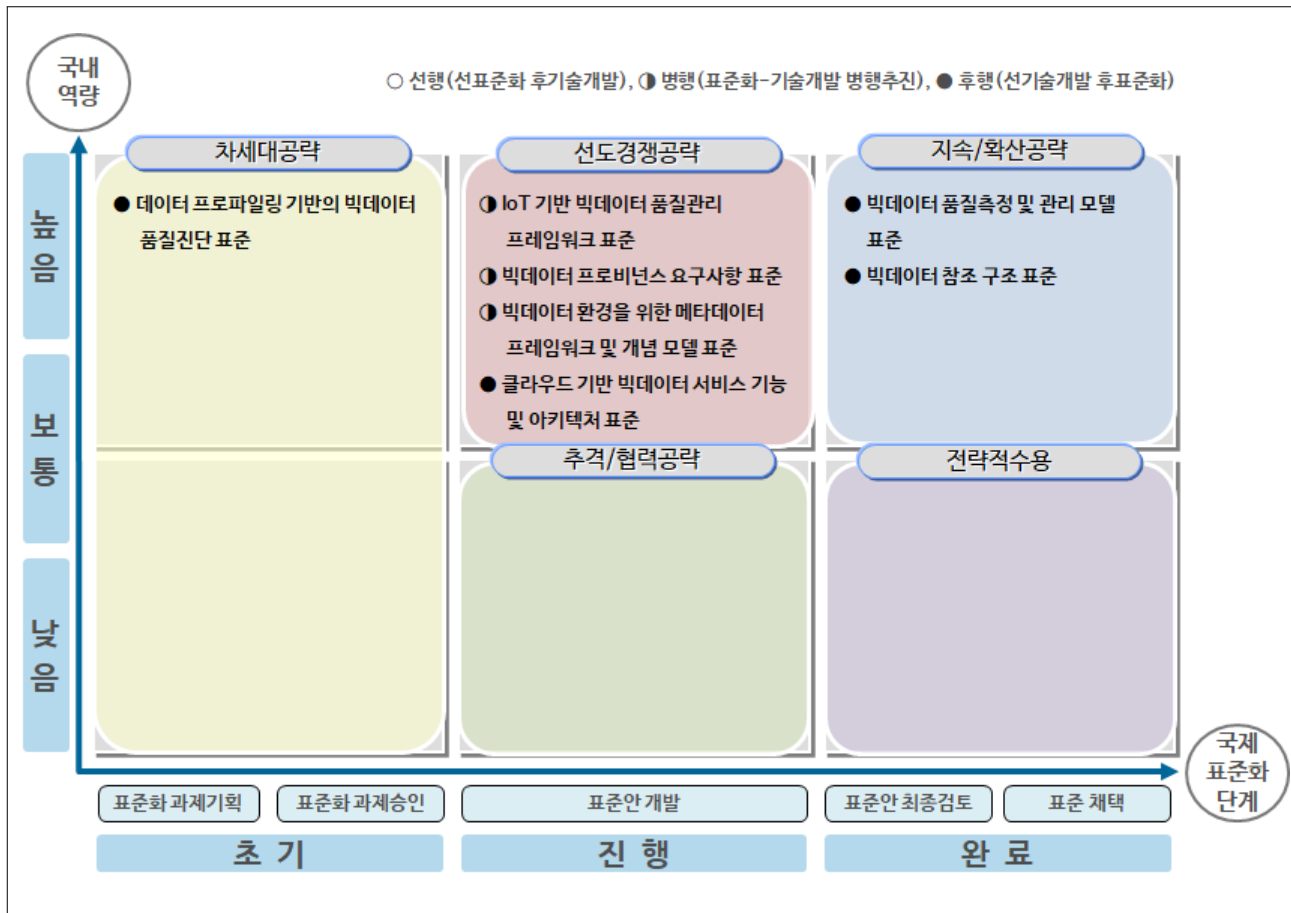
- R은 프로그래밍 언어이자 통계적 컴퓨팅을 지원하는 소프트웨어 환경이며, 빅데이터의 등장과 함께 R은 데이터들을 통계적으로 분석하기 위한 도구로 활용
- R의 상용 버전: R은 레드햇이 리눅스를 지원하는 방식과 유사한 서비스와 지원 모델을 추구하는 레볼루션 애널리틱스(Revolution Analytics)로 상용 버전이 배포
- R의 공개 버전: GNU 제너럴 퍼블릭 라이선스(GNU General Public license)에서 공개 버전은 오픈 소스로 쉽게 사용이 가능
- R의 진화: R언어가 빅데이터 분야에 활용되면서 단순 통계계산 영역에서 벗어나, 머신러닝과 딥러닝을 포함한 기능도 제공하며, 또한, 빅데이터의 활용을 지원하기 위한 시각화 기술도 제공

Ⅲ. 국내외 표준화 추진전략

3.1. 표준화 SWOT 분석

| 국내역량요인 국외환경요인 | | | 강점요인 (S) | | 약점요인 (W) | |
|--|----|--|---|--|--|---|
| | | | 시장 | -IoT, 모바일 등의 다양한 기기를 통한 데이터 생성 증가 -네트워크 및 단말 환경이 최고 수준이라 이에 따른 데이터 생성이 용이 | 시장 | -데이터 자체에 대해서 비용을 지불하지 않고 획득하려는 인식 |
| | | | 기술 | -정부의 지속적인 빅데이터 관련 기술 개발 투자 | 기술 | -개인정보보호에 대한 강력한 규제에 인하여 사용 가치가 있는 데이터 확보의 어려움 |
| | | | 표준 | -국제 표준화 주도(의장, 에디터 등의 의장단 확보) | 표준 | -국내 기업의 표준화 미참여 |
| 기획요인 (O) | 시장 | -클라우드, 인공지능 등의 관련 기술과 연계한 빅데이터 활용 가치 증대 | 【SO전략】 | | 【WO전략】 | |
| | 기술 | -다양한 빅데이터 관련 오픈소스 기술 장벽이 낮음 | -(시장) 데이터 증가 및 데이터 활용 증가로 인한 데이터 확보 용이 -(기술) 오픈 소스 기반의 빅데이터 관련 기술 확보 용이 -(표준) 국제 표준화 의장단 확보 및 표준화 기반 조성 | | -(시장) 활용할 데이터는 다수 확보할 수 있으나 고품질의 데이터 확보는 애로사항이 있음 -(기술) 빅데이터 기반 기술 개발에 주력하고 있으며, 데이터 확보에 대한 기술 개발이 상대적으로 미약 -(표준) 표준 참여를 위한 더 많은 인력이 요구되고 있으며, 이를 통한 표준 경쟁력 강화 | |
| | 표준 | -공적 표준화 기구, 사실 표준화 기구 등에서 빅데이터 표준 개발 중 | | | | |
| 위협요인 (T) | 시장 | -특정 솔루션이 지배적이라 솔루션에 대한 시장적 Lock-in 현상 발생 우려 | 【ST전략】 | | 【WT전략】 | |
| | 기술 | -특정 지배적인 솔루션에 의해 특정 솔루션에 대한 기술적 Lock-in 현상 발생 우려 | -(시장) 빅데이터 유통을 위한 규제 완화 필요 -(기술) 하둡에 의존적이지 않도록 빅데이터 개발에 대한 다양한 생태계 구축 필요 -(표준) 데이터 유통 관련 표준 개발 필요 | | -(시장) 관련 기술과 연계한 융합 산업으로 빅데이터의 시장 확산 노력 필요 -(기술) 개인정보에 대한 안전한 빅데이터 기술 개발 필요 -(표준) 표준전문가 육성 및 핵심 표준 기술 보유기관과의 전략적 공조로 지속적인 국제표준 확보 | |
| | 표준 | -1~2위 글로벌 벤더의 표준화 미참여 | | | | |
| 표준화 추진상의 문제점 및 현안 사항 | | | | | | |
| <div>- 공적 표준화 기구와 사실 표준화 기구에서 데이터 관련 표준화는 추진되고 있으나, 국내외적으로 이에 대한 참여는 미미한편</div> <div>- 특히, 국내의 참여는 매우 저조하며, 연구계와 학계 일부가 표준 개발을 주도하고 있으며, 산업계는 표준보다는 기술 개발 및 서비스 개발에 중점을 두고 있음</div> <div>- 또한, JTC1의 경우 빅데이터 그룹(WG9)이 인공지능 그룹(SC42)로 이관된바, 국내의 대응체계도 인공지능과 빅데이터가 연계되는 모습이 필요</div> <div>- 빅데이터가 다양한 산업군에서 활용되기 때문에 각 산업군에서 빅데이터를 활용하는 형태의 표준은 파악이 제대로 되지 않고 있음</div> | | | | | | |

3.2. 중점 표준화 항목별 국내외 추진전략

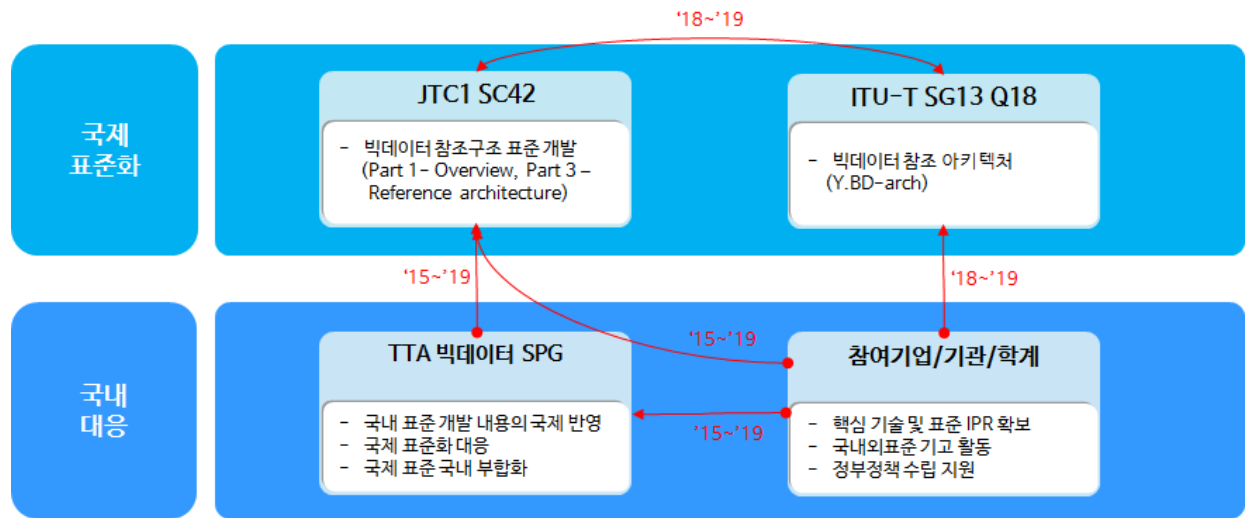


○ 영역별 특징 및 대응전략

- **차세대공략** : 미래 핵심기술 및 유망서비스 신규 표준 제안을 통해 표준화를 선점할 수 있는 분야
: 국제 표준 기획 단계부터 주도적 참여를 통해 국제표준화 선도 기반 확보
: 관련 표준화기구에서의 적극적인 제안으로 국내 핵심 기술의 국제표준화를 위한 발판 마련
- **선도경쟁공략** : 표준화 경쟁이 치열하지만 국내역량이 높아 국제표준 선도가 가능한 분야
: 국내 기술의 국제표준 반영을 위한 관련 표준화기구에서의 적극적인 표준화활동 추진
- **추격/협력공략** : 국제표준화가 활발히 진행 중인 분야 중 국내 진입시기가 다소 늦어졌지만 타 국가의 표준화 수준에 도달하기 위해 후발주자로서 추격하거나 다각화된 협력이 필요한 분야
: 국제 공식 및 사실표준화기구, 포럼, 컨소시엄에서의 다각적인 대응 방안 모색
: 전략적 대외협력 강화 및 제휴를 통한 기술/표준의 Catch-up 전략 추진
- **지속/확산공략** : 국제표준화가 거의 완료단계이나 국내역량이 높아 후속/개정 표준화에서의 선도가 예상되며, 표준 기반 서비스 및 시장 확산에 집중이 필요한 분야
: 높은 국내 역량을 바탕으로 한 후속/개정 표준화 주도 및 추가적인 틈새표준 발굴을 모색
: 표준기반 킬러 애플리케이션 개발 및 서비스 적용을 통한 표준 활용 촉진
- **전략적수용** : 국제표준화가 거의 완료된 분야 중 국내역량은 낮지만 전략적으로 수용이 필요한 분야
: 국제 표준의 수용 및 적용을 통한 국제 호환성 확보와 국내 시장 확산

(지속/확산공략 | 후행) 빅데이터 참조 구조 표준

| | | | | | |
|---|---|----------------------------|------------------|-----------------------|-------------------------------------|
| 전략적 중요도 / 국내 역량 | <p>국외대비 국내 표준화 역량</p> <p>국외대비 국내 기술개발 수준</p> <p>정책 부합성</p> <p>국제표준화 국내 기여도</p> <p>시장/기술적 파급효과</p> <p>IPR 확보 가능성</p> | | 표준화 기구/ 단체 | 국내 | TTA 빅데이터 SPG |
| | | | | 국제 | JTC1 SC42, ITU-T SG13 WP2/Q18 |
| | | | | 국내 참여 업체/ 기관 | ETRI |
| 기술 개발 단계 | 국내 | □기초연구→□실험→■시작품→□제품화→□사업화 | 기술 수준 | 70% (선도국가대비) | |
| | 국외 | □기초연구→□실험→□시작품→□제품화→■사업화 | | | |
| | 선도국가/ 기업 | 미국/Oracle, IBM, Google | | | |
| 표준화 단계 | 국내 | □과제기획→□과제승인→□개발→□검토→■표준채택 | 표준 수준 | 100% (선도국가대비) | |
| | 국제 | □과제기획→□과제승인→□개발→■검토→□표준채택 | | | |
| | 선도국가/ 기업 | 한국/ETRI 미국/NIST, Oracle | | | |
| <p>- Trace Tracking : 적극공략(Ver.2018) → 지속/확산공략(Ver.2019)</p> <p>빅데이터 참조 구조는 JTC1 WG42(舊 JTC1 WG9)가 중심이 되어 2015년부터 개발이 진행 중이며, 우리나라가 에디터로 참여하여 국제 표준 개발을 주도하고 있으나, 국내 기술은 아직 국외 시장을 추격하고 있는 상태로 국제 표준화 주도권을 지속하여 유지하며 국내 기술을 국제 표준에 반영해 나가는 방향으로 Ver.2019에서는 지속/확산공략 항목으로 분류</p> | | | | | |

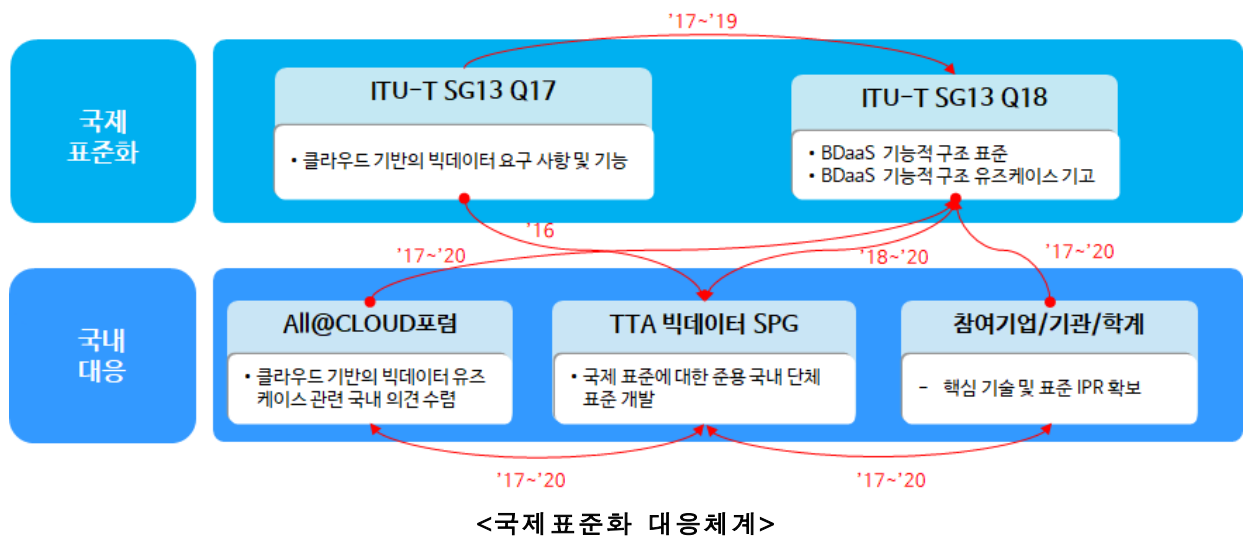


<국제 표준화 대응체계>

| | |
|----------------------|---|
| 국제 표준화 대응 방안 | <p><현황></p> <ul style="list-style-type: none"> 2015년 4월 JTC1 WG42(舊 JTC1 WG9) 1차 회의를 통해 빅데이터 참조 구조 표준 개발이 시작되었으며, 2016년 5개의 세부 파트로 분리되어 개발이 진행 중. 현재 20547-3 Reference Architecture가 핵심 표준으로 2018년 6월부터 DIS 투표가 추진 ITU-T SG13은 개발 중인 클라우드 기반의 빅데이터 서비스 아키텍처에서 빅데이터 참조 아키텍처를 별도로 분리하여 개발을 진행 <p><대응방안></p> <ul style="list-style-type: none"> (공식표준화 대응전략 : 국제표준화기구 활동(적극대응), 국제표준화기구 의장단 수입) JTC1 및 ITU-T 양 기구에서 에디터십 및 Liaison 라포처를 담당하고 있으며, 이를 활용하여 2018년까지 양 기구 간의 표준 간 충돌 방지 및 주도적 표준 개발 추진 |
| 국내 표준화 추진 계획 | <p><현황></p> <ul style="list-style-type: none"> 현재 국책기관인 ETRI를 중심으로 2016년 참조 아키텍처 표준에 대한 제정 완료 <p><추진계획></p> <ul style="list-style-type: none"> (표준화위원회 PG 활동, 국제표준 준용) JTC1 및 ITU-T의 국제 표준화에 대한 주도권을 바탕으로 국내 표준 내용을 적극 반영하고, 국제 표준의 완료 후 TTA 빅데이터 SPG를 통하여 국제 표준의 국내 부합화를 위한 국내 표준 개정 추진 |
| 표준특허 전략 | <ul style="list-style-type: none"> 표준 및 R&D 중후기 전략 : 표준 정합성 확보를 위한 특허 재설계 전략 참조 아키텍처와 관련된 특허 현황 분석 및 정합성 검토를 통하여 누락되었거나 권리 주장이 가능한 기능 요소를 식별하고, 기 출원/등록된 특허를 바탕으로 수정/반영 |
| 기술개발 -표준화 -IPR 연계 방안 | <ul style="list-style-type: none"> 선기술개발 후표준화 국내 기술 개발 중이거나 기존 상용 제품에 대한 공통 기능 요소를 식별, 표준에 반영하되 국내에서 반영 가능한 하둡 확장 기술 또는 기존 데이터 플랫폼을 확장한 데이터 거버넌스 기술 등을 타깃으로한 IPR 연계 |

(선도경쟁공략 | 후행) 클라우드 기반 빅데이터 서비스 기능 및 아키텍처 표준

| | | | | | |
|---|---|--------------------------------|------------------|-----------------------|---------------------------------|
| 전략적 중요도 / 국내 역량 | <p>국외대비 국내 표준화 역량</p> <p>국외대비 국내 기술개발 수준</p> <p>정책 부합성</p> <p>국제표준화 국내 기여도</p> <p>시장/기술적 파급효과</p> <p>IPR 확보 가능성</p> | | 표준화 기구/ 단체 | 국내 | TTA 빅데이터 SPG, All@CLOUD포럼 |
| | | | | 국제 | ITU-T SG13 WP2 |
| | | | | 국내 참여 업체/ 기관 | ETRI |
| | | | | | |
| 기술 개발 단계 | 국내 | □기초연구→□실험→□시작품→■제품화→□사업화 | 기술 수준 | 80% (선도국가대비) | |
| | 국외 | □기초연구→□실험→□시작품→□제품화→■사업화 | | | |
| | 선도국가/ 기업 | 미국/Google, Amazon, 페이스북 | | | |
| 표준화 단계 | 국내 | ■과제기획→□과제승인→□개발→□검토→□표준채택 | 표준 수준 | 80% (선도국가대비) | |
| | 국제 | □과제기획→□과제승인→■개발→□검토→□표준채택 | | | |
| | 선도국가/ 기업 | 중국/차이나텔레콤 폴란드/Orange polska | | | |
| <p>- Trace Tracking : 적극공략(Ver.2018) → 선도경쟁공략(Ver.2019)</p> <p>클라우드 기반의 빅데이터 서비스 기능 및 아키텍처 분야는 개발이 한참 진행 중이며, 우리나라에서 국제 표준 작업에 직접적으로 참여 중. 또한 현실적으로 우리나라의 산업계에서도 클라우드 기반으로 빅데이터 서비스를 제공하고자 하는 움직임이 많으므로, 선도경쟁공략 항목으로 분류</p> | | | | | |



| | |
|---------------------|--|
| 국제표준화 대응 방안 | <p><현황></p> <ul style="list-style-type: none"> ITU-T SG13에서 WP2 Q17에서 클라우드 기반의 빅데이터 서비스 기능에 대한 요구사항 표준 Y.3600(Requirements and capabilities for cloud computing based big data)이 2015년 7월 폴란드, 중국, 한국을 중심으로 제정되었으며, 후속으로 '클라우드 기반의 빅데이터 기능 및 구조 표준(Y.BDaaS-arch)'을 ITU-T SG13 Q18에서 개발 중 <p><대응방안></p> <ul style="list-style-type: none"> (공식표준화 대응전략 : 국제표준화기구 활동(적극대응)) ITU-T에서 개발되고 있는 Y.BDaaS-arch 권고안에 주 에디터로 국내에서 참석하고 있으며, 지속적인 기고 활동을 통해 국내 관련 포럼 및 TTA 빅데이터 SPG에서 제시된 의견을 최대한 반영 필요 |
| 국내표준화 추진 계획 | <p><현황></p> <ul style="list-style-type: none"> 현재 ETRI를 중심으로 ITU-T SG13 Q18에서 국제 표준 개발(클라우드 기반의 빅데이터 기능 및 구조 표준)에 적극 참여하고 있음. 그러나 산업계의 참여는 미진한 상태 <p><추진계획></p> <ul style="list-style-type: none"> (연구개발 표준화 연계 개발) 진행 중인 국제표준 개발에 한국에서 주 에디터로 참여 중. 현재 표준 제정을 위해서 부 에디터를 맡고 있는 중국 및 폴란드와 긴밀한 협력이 필요. 또한 국내 산·학·연의 의견을 국내 관련 포럼 및 TTA 빅데이터 SPG에서 논의하고 제시된 의견을 표준에 반영 필요 (국제표준 준용) ITU-T SG13에서 개발 중에 있는 표준과 관련하여 요구사항 표준은 (ITU-T Y.3600)은 이미 국내 단체표준으로 제정 완료한 상태이며, 현재 진행 중인 국제표준에 대해서는 TTA 빅데이터 SPG에서 국내 단체표준으로 제정이 필요 |
| 표준특허 전략 | <ul style="list-style-type: none"> 표준 초중기 및 R&D 중후기 전략 : 표준 필수특허 설계전략 클라우드 기반에서 빅데이터의 분석에 활용 가능한 자원 연동 방법에 대한 세부적인 사용 예를 제시하고, 표준 제정 이전에 이에 대한 출원과 기고를 제시 |
| 기술개발 -표준화 -IPR 연계방안 | <ul style="list-style-type: none"> 선기술개발 후표준화 현재 개발되고 있는 기술을 진행 중에 표준화에 반영하고 표준이 제정되도록 함 |

(선도경쟁공략 | 병행) 빅데이터 프로비넌스 요구사항 표준

| | | | | | |
|---|---|--|------------------|-----------------------|-------------------|
| 전략적 중요도 / 국내 역량 | <p>국외대비 국내 표준화 역량</p> <p>국외대비 국내 기술개발 수준</p> <p>정책 부합성</p> <p>국제표준화 국내 기여도</p> <p>시장/기술적 파급효과</p> <p>IPR 확보 가능성</p> | | 표준화 기구/ 단체 | 국내 | TTA 빅데이터 SPG |
| | | | | 국제 | ITU-T SG13 Q17 |
| | | | | 국내 참여 업체/ 기관 | ETRI |
| | | | | | |
| 기술 개발 단계 | 국내 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input checked="" type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | 기술 수준 | 80% (선도국가대비) | |
| | 국외 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input checked="" type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | | |
| | 선도국가/ 기업 | 미국/Intel&SPARKL, Oracle | | | |
| 표준화 단계 | 국내 | <input checked="" type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | 표준 수준 | 100% (선도국가대비) | |
| | 국제 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input checked="" type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | | | |
| | 선도국가/ 기업 | 한국/ETRI 미국/Oracle | | | |
| <p>- Trace Tracking : 차세대공략(Ver.2018) → 선도경쟁공략(Ver.2019)</p> <p>빅데이터 분석이 본격화되면서, 다양한 도메인에서의 데이터 출처 관리를 지원하기 위한 표준화 및 프로비넌스 정보를 바탕으로한 분석 방법 재활용 등의 기술 개발이 진행 중으로 선도경쟁공략 항목으로 분류함. 국내 표준화 단계는 기획 단계이나 국제 표준의 개발을 국내에서 주도하고 있어 표준 수준은 선도국가 대비 100%로 표기</p> | | | | | |



(선도경쟁공략 | 병행) 빅데이터 환경을 위한 메타데이터 프레임워크 및 개념 모델 표준

| | | | | | |
|-----------------------------|-------------|--|------------------|-----------------------|-------------------|
| 전략적 중요도 / 국내 역량 | | | 표준화 기구/ 단체 | 국내 | TTA 빅데이터 SPG |
| | | | | 국제 | ITU-T SG13 Q17 |
| | | | | 국내 참여 업체/ 기관 | ETRI |
| 기술 개발 단계 | 국내 | <input type="checkbox"/> 기초연구→ <input checked="" type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | 기술 수준 | 80% (선도국가대비) | |
| | 국외 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input checked="" type="checkbox"/> 시작품→ <input type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | | |
| | 선도국가/ 기업 | 미국/Apache Foundation EU/Data portal | | | |
| 표준화 단계 | 국내 | <input checked="" type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | 표준 수준 | 100% (선도국가대비) | |
| | 국제 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input checked="" type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | | | |
| | 선도국가/ 기업 | 한국/ETRI 프랑스/Orange Telecom 중국/China telecom | | | |

- Trace Tracking : 차세대공략(Ver.2018) → 선도경쟁공략(Ver.2019)

빅데이터의 확산에 따른 유통과 통합 관리가 핵심 이슈로 대두되면서 이의 실현기술로 메타데이터 관리에 대한 중요성이 증대되고 있음. 또한 데이터 거래소 등 데이터 유통 관련 기술 개발을 통한 산업 활성화가 기대되며, 한국이 표준화를 주도하고 있는 상황이므로 Ver.2019에는 선도경쟁공략 항목으로 분류함. 국내 표준화 단계는 기획이나 국제 표준의 개발을 국내에서 주도하고 있어 표준 수준은 선도국가 대비 100%로 표기



(지속/확산공략 | 후행) 빅데이터 품질측정 및 관리 모델 표준

| | | | | | |
|---|-------------|--|------------------|-----------------------|---|
| 전략적 중요도 / 국내 역량 | | | 표준화 기구/ 단체 | 국내 | TTA 메타데이터 PG, 스마트제조 표준화포럼, ISO TC184 SC4 전문위원회 |
| | | | | 국제 | JTC1 SC32 WG2, ISO TC184 SC4 WG13 |
| | | | | 국내 참여 업체/ 기관 | K-DATA, 2e컨설팅, 지티원, 명지대, 강릉원주대 |
| 기술 개발 단계 | 국내 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input checked="" type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | 기술 수준 | 100% (선도국가대비) |
| | 국외 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input checked="" type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | | |
| | 선도국가/ 기업 | 미국/Google, Amazon, IBM, Oracle 한국/다음소프트, SKT | | | |
| 표준화 단계 | 국내 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input checked="" type="checkbox"/> 표준채택 | | 표준 수준 | 100% (선도국가대비) |
| | 국제 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input checked="" type="checkbox"/> 표준채택 | | | |
| | 선도국가/ 기업 | 한국/2e컨설팅 미국/ECCMA 독일/Höhn & Partner PartG 스페인/UCLM | | | |
| <p>- Trace Tracking : 적극공략(Ver.2018) → 지속/확산공략(Ver.2019)</p> <p>국내에서는 데이터 품질 분야에서 한국데이터진흥원이 주도하고 있으며, 공공 데이터 부분에서는 한국정보화진흥원이 참여하고 있으며, 전문 기업들이 솔루션 제공과 컨설팅을 수행 중. 해외 솔루션은 데이터 품질 측정과 관리 솔루션 위주로 보급되어 있는 상황. 한국 연구진이 데이터 품질 관련 국제 표준 개발/제정을 선도하고 있으며 국내외에서 국제 표준의 활용도가 높아서 스마트 제조, IoT 등의 전문 분야에 적용되고 있으므로 지속/확산공략 항목으로 분류</p> | | | | | |



(차세대공략 | 후행) 데이터 프로파일링 기반의 빅데이터 품질진단 표준

| | | | | | |
|--|---|--|------------------|-----------------------|---|
| 전략적 중요도 / 국내 역량 | <p>정책 부합성 국제표준화 국내 기여도</p> <p>국외대비 국내 표준화 역량 국외대비 국내 기술개발 수준</p> <p>시장/기술적 파급효과 IPR 확보 가능성</p> | | 표준화 기구/ 단체 | 국내 | TTA 메타데이터 PG, JTC1 SC32 전문위원회, ISO TC184 SC4 전문위원회 |
| | | | | 국제 | JTC1 SC32 WG2, ISO TC184 SC4 WG13 |
| | | | | 국내 참여 업체/ 기관 | K-DATA, 위세아이텍 2e컨설팅, 강릉원주대 |
| | | | | | |
| 기술 개발 단계 | 국내 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input checked="" type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | 기술 수준 | 90% (선도국가대비) |
| | 국외 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input type="checkbox"/> 시작품→ <input type="checkbox"/> 제품화→ <input checked="" type="checkbox"/> 사업화 | | | |
| | 선도국가/ 기업 | 미국/SAS, 인포매티카, Trillum Software | | | |
| 표준화 단계 | 국내 | <input checked="" type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | | 표준 수준 | 80% (선도국가대비) |
| | 국제 | <input checked="" type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | | | |
| | 선도국가/ 기업 | 미국/ECCMA 스페인/UCLM | | | |
| <p>- Trace Tracking : 차세대공략(Ver.2019 신규)</p> <p>국내에는 한국데이터진흥원이 데이터 품질 관리, 측정 및 진단 등 제반분야를 주도하고 있으며, 공공 데이터 부분에서는 한국정보화진흥원이 참여하고 있는 상황. 데이터 품질 전문 기업들이 솔루션 제공과 컨설팅에 참여 중. 한국의 표준개발 연구진이 데이터 품질 측정 및 진단 관련 국제 표준 개발/제정을 선도하고 있으며 국내외에서 국제 표준의 활용도가 높으며, 사업화가 가능한 분야로 차세대공략 항목으로 분류</p> | | | | | |



(선도경쟁공략 | 병행) IoT 기반 빅데이터 품질관리 프레임워크 표준

| | | | | | |
|---|-------------|--|------------------|-----------------------|---|
| 전략적 중요도 / 국내 역량 | | | 표준화 기구/ 단체 | 국내 | ISO TC184 SC4 전문위원회, TTA 메타데이터 PG |
| | | | | 국제 | JTC1 SC32, ISO TC184 |
| | | | | 국내 참여 업체/ 기관 | 지티원, 명지대, 강릉원주대 |
| 기술 개발 단계 | 국내 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input checked="" type="checkbox"/> 시작품→ <input type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | 기술 수준 | 100% (선도국가대비) | |
| | 국외 | <input type="checkbox"/> 기초연구→ <input type="checkbox"/> 실험→ <input checked="" type="checkbox"/> 시작품→ <input type="checkbox"/> 제품화→ <input type="checkbox"/> 사업화 | | | |
| | 선도국가/ 기업 | 한국/지티원 스페인/Lucendia Lab 스웨덴/Siemens | | | |
| 표준화 단계 | 국내 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input checked="" type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | 표준 수준 | 100% (선도국가대비) | |
| | 국제 | <input type="checkbox"/> 과제기획→ <input type="checkbox"/> 과제승인→ <input checked="" type="checkbox"/> 개발→ <input type="checkbox"/> 검토→ <input type="checkbox"/> 표준채택 | | | |
| | 선도국가/ 기업 | 한국/명지대, 강릉원주대 스페인/Intelligent Environment | | | |
| <p>- Trace Tracking :선도경쟁공략(Ver.2019 신규)</p> <p>IoT의 등장으로 데이터의 실시간 전송이 폭발적으로 증가되면서 IoT 데이터에 대한 품질 수준을 확보하려는 노력의 일환으로 IoT 데이터 품질관리 기술 개발 및 표준화 요구. IoT 데이터 품질 특성 표준, IoT 데이터 프로파일링 표준 등을 중심으로 IoT 기반 빅데이터 품질관리 프레임워크 기술이 기존 기술에 추가되어 글로벌 기업의 IoT 실증 데이터를 대상으로 프로젝트가 진행 중. 국내외에서 국제 표준의 활용도가 높으며, 사업화가 가능한 분야로 선도경쟁공략 항목으로 분류</p> | | | | | |

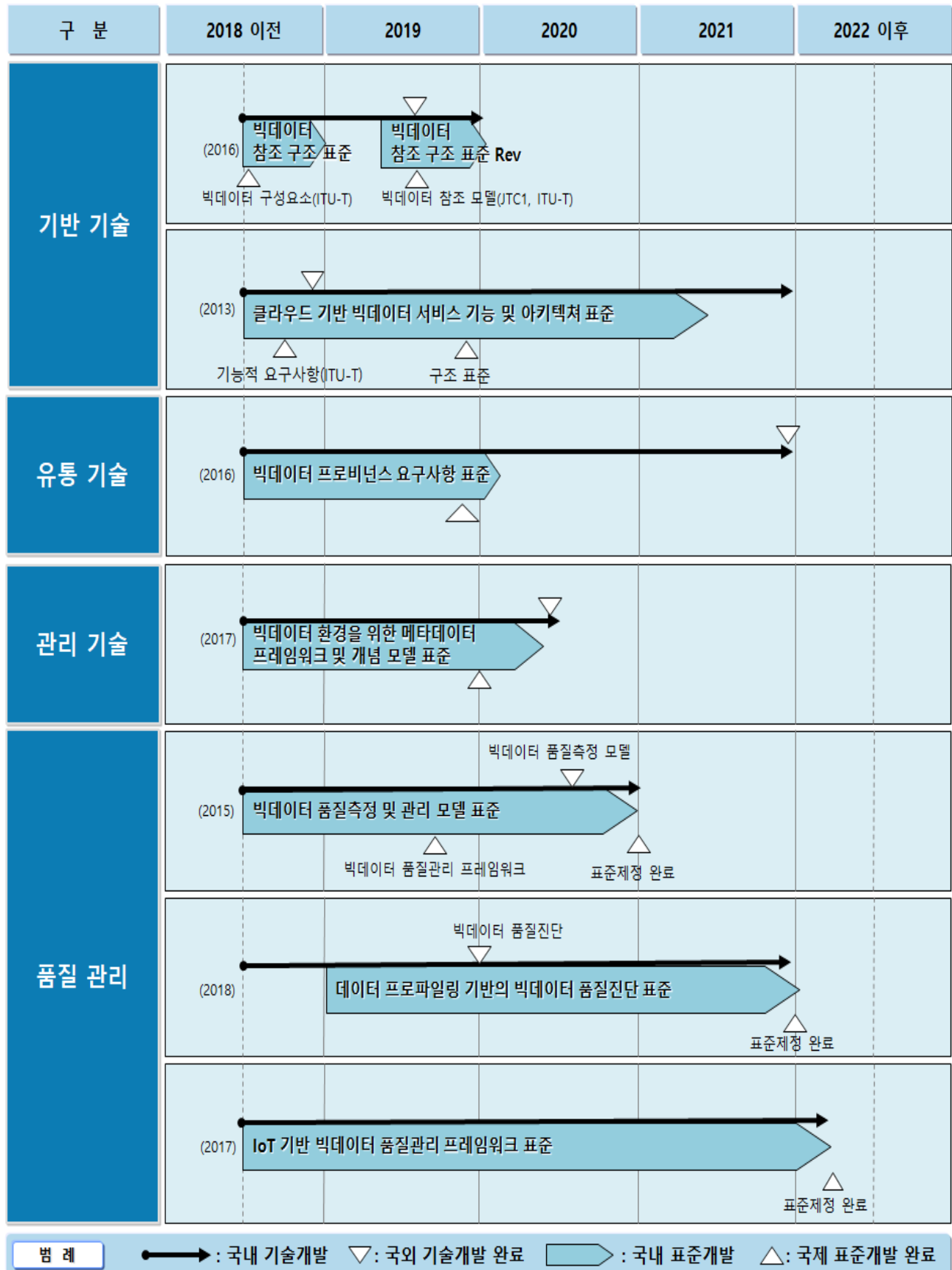


3.3. 오픈소스 국내의 추진전략

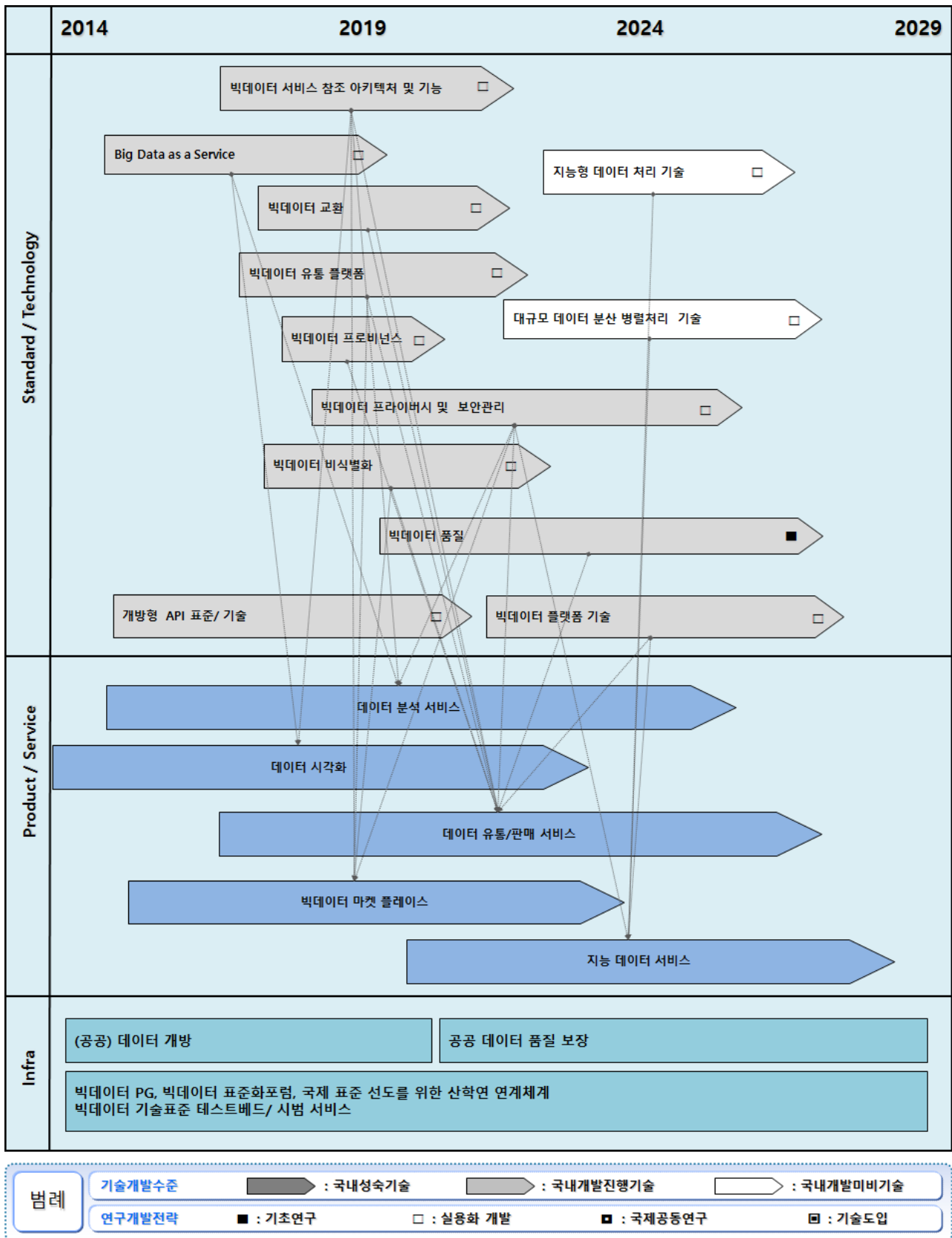


3.4. 중기(3개년) 및 장기(10개년) 표준화 계획

○ 중기(2019~2021) 표준화 계획



○ 장기(~2029) 표준화 계획



[작성위원]

| 구분 | 소속 | 성명 | 직위 | 국내외 표준화활동 |
|----------|---------|-----|----|---|
| 총괄 | IITP | 김형철 | PM | ▶ 과기정통부 기반SW·컴퓨팅 PM |
| 분과장 | 숙명여대 | 윤용익 | 교수 | ▶ TTA 빅데이터 SPG(SPG22) 부의장 |
| 위원 | K-DATA | 박천웅 | 수석 | ▶ JTC1 SC32 활동 중 ▶ TTA 메타데이터 PG(PG606) 위원 |
| 위원 | TCA 서비스 | 오경희 | 대표 | ▶ ITU-T SG17 에디터 ▶ TTA 응용보안/평가인증 PG(PG504) 특별위원 |
| 위원 | ETRI | 이강찬 | 책임 | ▶ ITU-T SG13 Q17 라포처, JTC1 SC38 에디터, ITU-T SG13 에디터 ▶ TTA 클라우드컴퓨팅 SPG(SPG21) 부의장, TTA 빅데이터 SPG(SPG22) 의장 |
| 위원 | 강릉원주대 | 이창수 | 교수 | ▶ ISO TC184 SC4 WG13 ISO 8000-66 프로젝트 리더 ▶ ISO TC184 SC4 Quality Committee 부의장 ▶ ISO TC184 SC4 Implementation Forum 부의장 |
| 위원 | ETRI | 인민교 | 책임 | ▶ ITU-T SG13 Q17/Q18 에디터 |
| 위원 | KERIS | 정의석 | 책임 | ▶ JTC1 SC32, SC34 활동 중 ▶ TTA 메타데이터 PG(PG606), TTA e-퍼블리싱 PG(PG608) 위원 |
| 위원 | 안양대 | 정의현 | 교수 | ▶ TTA IPv6/인터넷주소자원 PG(PG222) 부의장 |
| 위원 | ETRI | 하수욱 | 책임 | ▶ ITU-T SG13 에디터, JTC1 빅데이터 표준 에디터 ▶ TTA 빅데이터 SPG(SPG22) 간사 |
| 특허분석 | KISTA | 박성혁 | 선임 | ▶ TTA 표준화전략맵 빅데이터 특허분석 |
| TTA PG담당 | TTA | 박준환 | 전임 | ▶ TTA 클라우드컴퓨팅 SPG(SPG21), TTA 빅데이터 SPG(SPG22) 담당 |
| 간사 | TTA | 고준호 | 책임 | ▶ TTA 표준화전략맵 빅데이터 분야 간사 |

[참고문헌]

1. Market Watch, "Big Data Equals Big Business opportunity Say Global IT and Business Professionals", 2012.5.14.
2. Informatica, "Balancing Opportunity and Risk in Big Data", 2012.5.15.
3. Y.3600, Initial draft Recommendation Y.BigData-reqts, Requirements and capabilities for cloud computing based big data. ITU-T, 2015.7
4. 정보통신정책연구원, “빅데이터 시장의 현황 및 전망”, 2013.3
5. ISO/IEC JTC 1 SC 32 홈페이지 <http://JTC 1SC 32.org/>
6. 창조경제 및 정부3.0 지원을 위한 빅데이터 산업 발전전략, 미래부 등, 2013.12.
7. 빅데이터 시장·정책·표준화 현황, ETRI 표준연구센터, 2014.2.
8. 한국콘텐츠진흥원, 빅데이터 시장 현황과 콘텐츠산업 분야에 대한 시사점, 2014.1
9. 한국과학기술정보연구원, 빅데이터 산업의 현황과 전망, 2013.4
10. 한국정보화진흥원, 빅데이터 활용 단계별 업무절차 및 기술활용 매뉴얼, 2014.5.
11. 한국정보화진흥원, 알기 쉬운 공공부문 빅데이터 분석·활용 가이드, 2012.12.
12. 방송통신위원회, 빅데이터 개인정보보호 가이드라인, 2014.12
13. 박현아, 빅데이터 시장 현황과 콘텐츠산업 분야에 대한 시사점, 코카포커스, 2013-11호, 한국콘텐츠 진흥원, 2014.1
14. Draft supplement Y.Suppl.BigData-RoadMap, “ITU-T Y.BigData-reqts - Supplement on big data standardization roadmap”, ITU-T, 2015.7
15. 빅데이터 관련 오픈소스 개발동향, 한국통신학회 종합 학술 발표회 논문집, 680-681 (2013)
16. “빅데이터산업의 부상 및 시사점”, 김광섭, 산은조사월보, 2016.10.
17. “IDC 2017년 세계 빅데이터 분석 시장 1500억달러 넘어설 전망”, 조선비즈, 2017. 4. 17, http://biz.chosun.com/site/data/html_dir/2017/04/17/2017041701261.html
18. Recommendation ITU-T Y.3600 (2015), Big data - cloud computing based requirements and capabilities
19. Initial draft Recommendation Y.BDaas-arch, Big data - Functional architecture of Big Data as a Service, http://www.itu.int/itu-t/workprog/wp_item.aspx?isn=10548
20. “한국IBM, 기업용 정보보안 신제품군 대거 발표”, <https://www-03.ibm.com/press/kr/ko/pressrelease/39559.wss>, 2012. 11. 19
21. 양경아 외, “빅데이터 시스템의 보안 위협 및 보안 요구사항 분석”, 보안공학연구논문지, Vol.13, No.6, pp 501-514, 2016. 12. 08
22. 정동원 등, 빅데이터 처리를 위한 메타데이터 표준화에 관한 연구, 한국정보과학회 동계학술발표, 2015.
23. Initial draft Recommendation Y.bdm-sch, Big data - Metadata framework and conceptual model, 2017.
24. W3C Recommendation, Data Catalog Vocabulary (DCAT), 2014.
25. 한국정보화진흥원, 알기 쉬운 Linked Open Data, 2014.

[약어]

| | |
|-------|--|
| AI | Artificial Intelligence |
| API | Application Programming Interface |
| ASP | Application Service Provider |
| BDaaS | Big Data as a Service |
| CAGR | Compound Annual Growth Rate |
| CDC | Cloud Data Center |
| CDG | Cloud Data Governance |
| CDMI | Cloud Data Management Interface |
| CG | Community Group |
| CIF | Cloud Industry Forum |
| CIMI | Cloud Infrastructure Management Interface |
| CSA | Cloud Security Alliance |
| CSB | Cloud Service Brokerage |
| CSV | Comma-Separated Values |
| CTP | Cloud Trust Protocol |
| DaaS | Desktop as a Service |
| DAP | Device API |
| DMTF | Distributed Management Task Force |
| DQ | Data Quality |
| DQM | Data Quality Management |
| EC2 | Elastic Computer Cloud |
| FGCC | Focus group Cloud Computing |
| GLD | Government Linked Data |
| IaaS | Infrastructure as a Service |
| IETF | Internet Engineering Task Force |
| ICT | Information and Communications Technologies |
| IoT | Internet of Things |
| ISO | International Organization for Standardization |
| KIG | Korea Interest Group |
| KMIP | Key Management Interoperability Protocol |
| LDP | Linked Data Platform |
| LOD | Linked Open Data |
| M2M | Machine-to-Machine |
| MVNO | Mobile Virtual Network Operator |

| | |
|-------|--|
| NaaS | Network as a Service |
| NAS | Network Attached Storage |
| NIST | National Institute of Standards and Technology |
| OASIS | Organization for the Advancement of Structured Information Standards |
| OCCI | Open Cloud Computing Interface |
| OCI | Open Container Initiative |
| OCM | Open Cloud Manifesto |
| ODCA | Open Data Center Alliance |
| OGF | Open Grid Forum |
| OMG | Object Management Group |
| OSS | Operation support systems |
| OVF | Open Virtualization Format |
| OWL | Web Ontology Language |
| PaaS | Platform as a Service |
| QoS | Quality of Service |
| RDF | Resource Description Framework |
| REST | REpresentational State Transfer |
| S3 | Simple Storage Service |
| SBC | Server Based Computing |
| SCP | Smart Connected Product |
| SDO | Standard Developing Organization |
| SGBD | Study Group on Big Data |
| SLA | Service Level Agreement |
| SMI | Storage Management Initiative |
| SNIA | Storage Networking Industry Association |
| SOA | Service Oriented Architecture |
| SPG | Special Project Group |
| SQL | SQL Query Language |
| ToR | Term of Reference |
| VDI | Virtual Desktop Infrastructure |
| W3C | World Wide Web Consortium |
| WG | Working Group |
| WP | Working Party |

