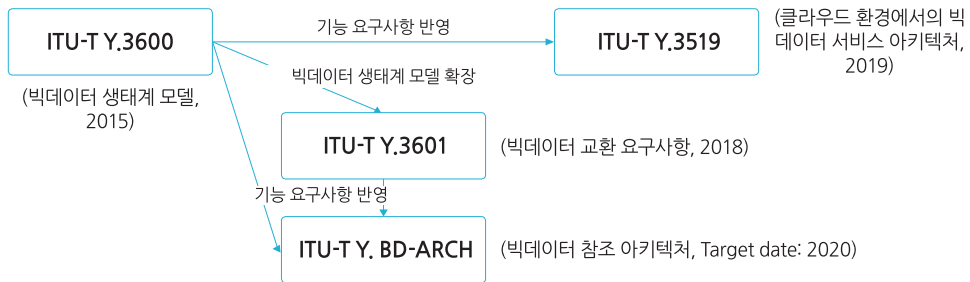




<표 1> 공적 표준화 기구 빅데이터 표준화 활동 범위

표준화 기구	작업그룹	표준화 범위
ITU-T SG 13	Q17	빅데이터 생태계 모델 및 기능적 요구사항 개발
	Q18	빅데이터 관련 기능 아키텍처 표준 개발
	Q19	빅데이터 거버넌스 및 관리 요구사항 표준 개발
ISO/IEC JTC 1/ SC 42	WG 2	빅데이터 고유 표준 개발(용어 및 참조 아키텍처)



[그림 1] ITU-T SG 13 빅데이터 기반 개념 표준 개발 현황

## 2.2 빅데이터 기반 개념 및 참조 아키텍처

ITU-T 및 JTC 1은 동일하게 클라우드 컴퓨팅 참조 아키텍처(ITU-T Y.3502, ISO/IEC 17789) 개발에 사용했던 아키텍처 방법론[2]을 동일하게 적용하여 사용자 뷰(user view), 기능 뷰(functional view) 그리고 크로스 커팅 이슈(cross-cutting aspect) 개발을 추진 중이다. ITU-T SG 13은 ‘ITU-T Y.3600 빅데이터 - 클라우드 기반 기능 및 요구사항 표준(2015)’을 통하여 빅데이터 생태계를 정의하고 이를 클라우드 환경에 적용하기 위한 기술 요구사항을 정의하였다. 제시된 빅데이터 생태계 모델은 데이터 제공자, 빅데이터 서비스 제공자, 빅데이터 서비스 소비자의 역할(role)과 각 역할들이 수행하는 활동(activity)들을 정의한다. 이후 우리나라는 데이터의 공유 및 유통 관점에서의 빅데이터 생태계로의 확장을 위한 표준을 제안, ‘ITU-T Y.3601 빅데이터 - 데이터 교환을 위한 프레임워크 및 요구사항’ 문서를 제정 완

료(2018)함으로써 빅데이터 생태계 모델에 대한 작업을 완료하였다. 또한 Y.3600을 기초로 ‘ITU-T Y.3519 클라우드 컴퓨팅 - 빅데이터 서비스 아키텍처 표준(2018)’ 개발을 완료하고, 빅데이터 참조 아키텍처 개발을 진행 중이다.

JTC 1에서는 2014년 말 제안된 2건의 표준화 항목을 기초로 빅데이터 작업그룹을 설립하여 2015년 표준 개발에 착수하였으며, ISO/IEC 20546 빅데이터 - 개요 및 용어 문서는 FDIS(Final Draft of International Standard)로 개발 완료 단계에 접어들었다. ISO/IEC 20547 빅데이터 - 참조 아키텍처 문서는 이후 5개의 하위 프로젝트로 분리되어 개발이 진행 중이다(<표 2>).

## 2.3 빅데이터를 위한 핵심 공통 기능 요구사항

ITU-T Y.3601에 기술된 확장된 빅데이터 생태계 모델이 제시하는 빅데이터 서비스 제공자의 기본적

**<표 2> 공적 표준화 기구(ITU-T, JTC 1) 빅데이터 기반 개념 및 아키텍처 표준화 현황**

문서번호	표준명	표준화 단계	내용
ITU-T Y.3600	빅데이터 - 클라우드 기반 빅데이터 기능 및 요구사항	Recommendation (2015)	빅데이터 생태계 모델(사용자부) 정의
ITU-T Y.3601	빅데이터 - 데이터 교환 프레임워크 및 기능 요구사항	Recommendation (2018)	빅데이터 생태계 모델 확장
ITU-T Y.3519	클라우드 컴퓨팅 - 빅데이터 서비스 아키텍처	Recommendation (2019)	클라우드 환경에서의 빅데이터 서비스 아키텍처(기능 뷰, 크로스 컷팅 이슈 정의)
ITU-T Y.BD-ARCH	빅데이터 - 참조 아키텍처	Draft Recommendation	빅데이터 기능 뷰, 크로스 컷팅 이슈 정의
ISO/IEC 20546	빅데이터 - 개요 및 용어	Final Draft International Standard (FDIS)	용어 및 관련 기술분야 정의
ISO/IEC 20547-1	빅데이터 참조 아키텍처 - 개요 및 적용 절차	Working Draft (WD)	빅데이터 참조 아키텍처 문서 구조 및 적용 절차 정의
ISO/IEC 20547-2	빅데이터 참조 아키텍처 - 유스케이스 및 기술적 고려사항	Technical Report (TR)	유스케이스 및 관련 기술 고려사항 조사 모음
ISO/IEC 20547-3	빅데이터 참조 아키텍처 - 참조 아키텍처	Second Draft of International Standard (2nd DIS)	빅데이터 참조구조 (사용자 뷰, 기능 뷰, 크로스 컷팅 이슈)
ISO/IEC 20547-4	빅데이터 참조 아키텍처 - 보안 및 개인정보 보호	Committee Draft (CD)	빅데이터를 위한 보안 아키텍처
ISO/IEC 20547-5	빅데이터 참조 아키텍처 - 표준 로드맵	Technical Report (TR)	빅데이터 관련 기존 표준 현황 모음

**<표 3> 빅데이터를 위한 공통 기능 요구사항**

문서번호	표준명	표준화 단계	내용
ITU-T Y.3601	빅데이터 - 데이터 교환 프레임워크 및 기능 요구사항	Recommendation (2018)	빅데이터 생태계에서 역할들 간의 데이터 교환·공유를 위한 기능적 요구사항
ITU-T Y.3602	빅데이터 - 데이터 이력관리 기능 요구사항	Recommendation (2019)	다양한 데이터의 융·복합, 유통과 관련된 이력 관리를 위한 개념 모델 및 기능적 요구사항
ITU-T Y.bdi-reqts	빅데이터 - 데이터 융합 요구사항	Draft Recommendation	빅데이터의 특성(크기, 속도, 다양성 등)에 따른 데이터 융합 시 고려해야 할 기능적 요구사항
ITU-T Y.BDD-reqts	빅데이터 - 데이터 보존 요구사항	Draft Recommendation	빅데이터 환경에서의 데이터 보존·관리를 위한 기능적 요구사항

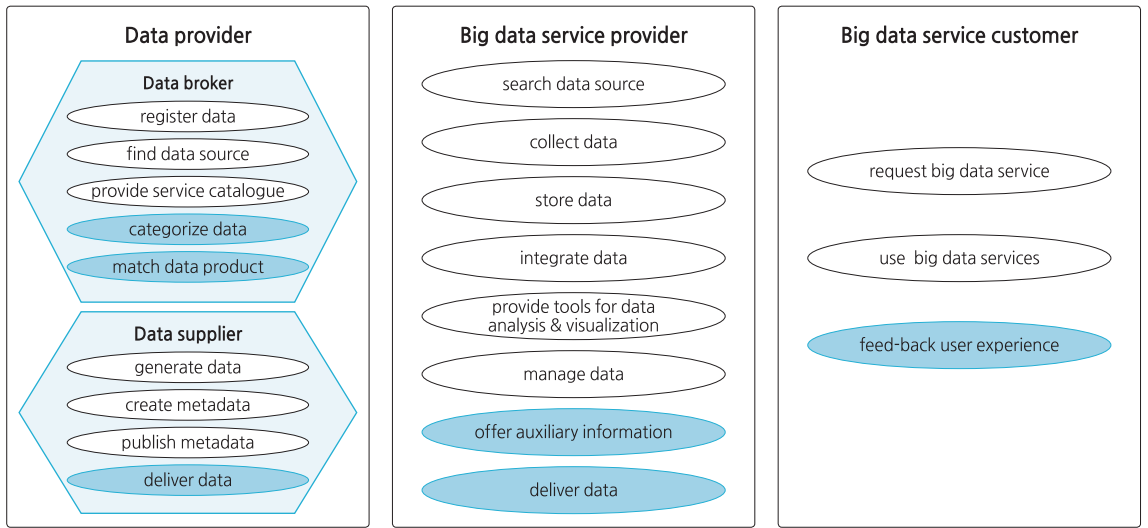
인 역할은 데이터 자원에 대한 검색, 데이터 수집, 데이터 저장, 데이터 융합(integration), 데이터 분석 및 시각화, 데이터 관리, 데이터에 대한 활용정보 제공 및 분석 결과의 전달로 구성되어 있다[3] (그림 2).

빅데이터 아키텍처의 기능 뷰는 이러한 생태계 모델의 역할들로 표현된 유스케이스들을 통하여 도출된 기능적인 요구사항들을 정제하여 공통으로 활용될 수 있는 기능 컴포넌트들을 도출하는 데 있으며, 이와 관련 개발 중이거나 개발이 완료된 표준 목록은 <표 3>과 같다.

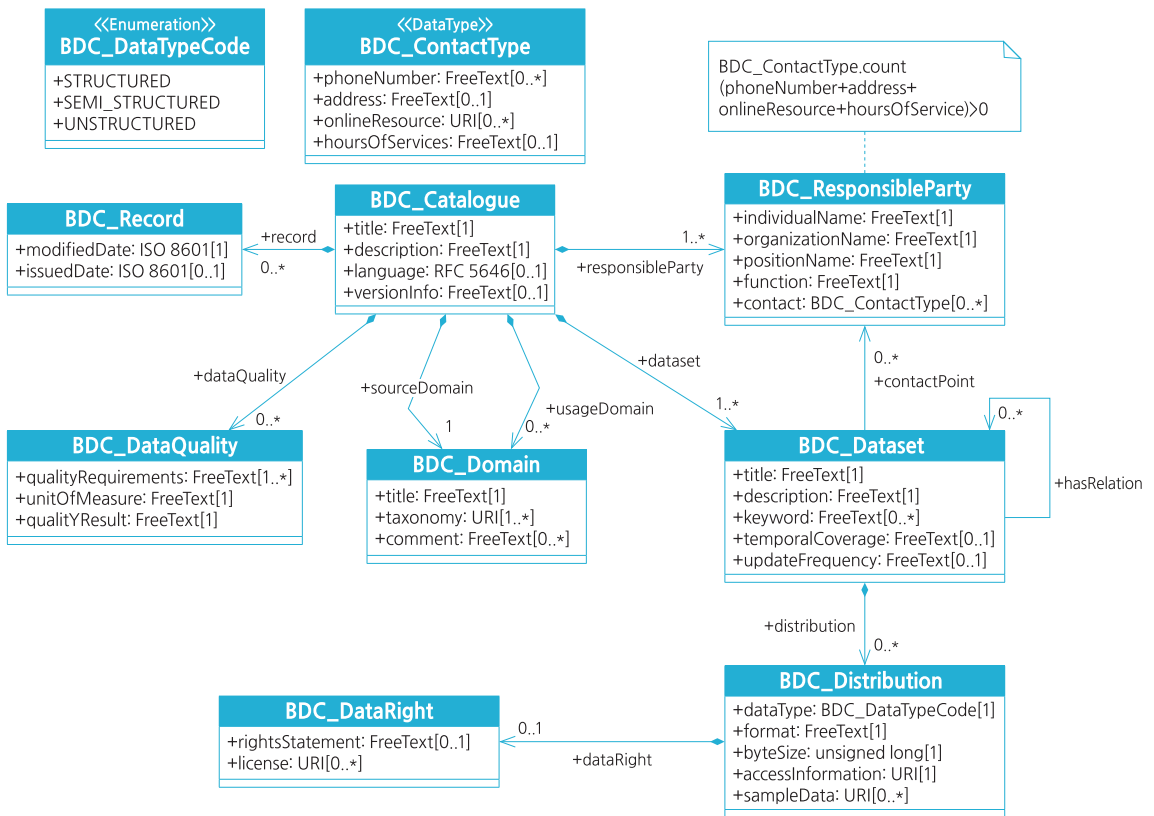
이 외에도 ITU-T SG 13에서는 빅데이터 유통을

위한 핵심 표준으로써 데이터 카탈로그를 정의하기 위한 ‘빅데이터 - 메타데이터 프레임워크 및 개념 모델’이 2019년 완료를 목표로 개발 중이다(그림 3).

JTC 1 SC 42에서는 개발 중인 빅데이터 참조 아키텍처(ISO/IEC 20547-3)의 기능 뷰를 확장한 빅데이터 인터페이스 디자인 가이드라인 및 활용 사례에 대한 개발을 제안하였으며[5, 6], 인도는 조직 수준에서 빅데이터 분석에 대한 성숙도 측정을 위한 관리모델의 개발을 각각 제안하여[7] 2019년 4월 제3차 SC 42 총회를 통해 본격적으로 논의할 예정이다.



[그림 2] 빅데이터 생태계 확장[ITU-T Y.3601]



[그림 3] 빅데이터를 위한 데이터 카탈로그 개념 모델[4]

### 3. 사실 표준화 추진 동향

#### 3.1 하둡 에코시스템

초창기 빅데이터 분야는 하둡(hadoop)을 중심으로 독립적인 소프트웨어들 간의 상호운용을 지원하는 방향으로 발전해 왔으며, 하둡 생태계 내에서의 다양한 오픈 소스 프로젝트들을 깃허브(github)를 통하여 공개하고 있다. 하둡은 여전히 빅데이터를 운용하는 데 있어서 유용한 기능들을 제공하고 있으나, 지속적으로 증가한 에코시스템 내의 각종 소프트웨어들과 이들 간의 상호운용성 테스트 프로젝트들의 수는 기하급수적으로 증가하고 있으며, 프로그램들 간의 기능적 상호 복제로 인하여 소비자들은 어떤 소프트웨어를 사용해야 할 것인지를 판단하는 것은 매우 어려운 일이 되어가고 있다.

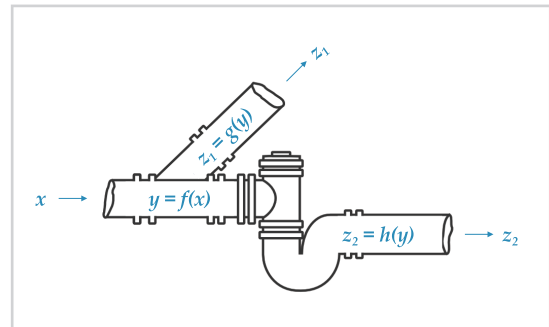
#### 3.2 데이터 유통 및 공유를 위한 상호운용성

W3C는 웹상에서 데이터 자원들의 공유를 위한 데이터 카탈로그 모델인 DCAT 표준을 2014년 제정하였으며 2018년부터 데이터 교환 작업그룹(DXGWG)을 통하여 이를 확장한 차기 버전을 개발하고 있다. 여기에 데이터의 활용정보, 데이터의 이력정보, 데이터 카탈로그들 간의 관계 정보들을 추가해 나갈 계획이다.

웹 플랫폼들을 중심으로 수많은 사용자들이 생산하는 정보들을 상호 이동 가능하게 함으로써 기존의 데이터 비즈니스 생태계를 유지·확장해 나가기 위한 방편으로 페이스북, 구글, MS, 트위터를 중심으로 한 데이터 교환 프로젝트(Data Transfer Project)가 2017년부터 시작되었으며, 데이터들을 수직계층으로 분류하기 위한 메타데이터와 개별 데이터 포맷들 간의 이동성 제공을 위한 어댑터 개발을 진행 중이다[9].

### 3.3 데이터 분석 상호운용성

데이터 마이닝 그룹(Data mining group)은 데이터 마이닝 분야에서 사용되는 각종 알고리즘들을 XML 기반의 스키마를 통해 서로 다른 시스템들과 공유하기 위한 모델인 PMML(Predictive Markup Model Language)과 데이터의 분석·처리 프로세스를 파이프라인으로 정의하여 공유하기 위한 JSON 기반의 교환 포맷인 PFA(Portable Format of Analytics) 표준을 개발, 지속적으로 업그레이드된 버전을 개발 중이다.




[그림 4] PFA 데이터 파이프라인 개념[8]

### 4. 차세대 표준화 핵심 이슈

ITU-T는 FG-DPM(Focus Group of Data Processing and Management to support IoT and Smart Cities & Communication)을 통하여 데이터의 공유 및 상호운용성 등에 대한 향후 표준화 방향을 모색하고 있다. 이와 별도로 ITU-T SG 13을 중심으로 트러스트 표준화가 진행되고는 있으나 빅데이터 생태계에 적용하기 위한 보다 구체적인 수준의 표준 개발이 요구된다. 한편 JTC 1/SC 42에서는 인공지능에 사용되는 데이터에 대한 신뢰성, 분석 결과의 편향성, 인공지능에 대한 위험 관리 등의 이슈들이 지속해서 제기되고 있으며, 이는 빅데이터 기술이

데이터의 효과적 관리 및 활용에서 점차 빅데이터 분석 결과에 대한 신뢰성을 식별하기 위한 기술적, 정책적인 안전장치를 찾기 위한 방향으로 이동하고 있음을 볼 수 있다. 이 외에도 최근 논의가 시작된 온라인 상에서 개인이 생산한 데이터를 플랫폼 제공자가 아닌 개인이 소유권을 가지고 관리할 수 있도록 하는 개인 데이터(personal data) 개념의 등장은 보안과는 또 다른 관점에서 개인정보의 유통 문제를 해결하기 위한 방향을 제공하고 있다.

## 5. 맺음말

공적 표준화 기구들을 중심으로 2014년부터 시작된 빅데이터 분야의 표준화는 IoT, 클라우드 컴퓨팅 등 선제적이며, 발 빠르게 진행된 표준화 양상과는 다르게 비교적 더딘 걸음으로 표준 개발이 진행되고 있다. 이는 그만큼 빅데이터 관련 기술들이 이미 넓게 사용되고 있고, 이에 따라 표준화에 참여하는 국가 및 기업들이 특정 사안에 대하여 기술적 합의가 쉽지 않음을 반증한다고 할 수 있다. 우리나라는 빅데이터 표준화와 관련 시작부터 에디터십을 바탕으로 표준화를 주도하고 있으며, 특히 데이터 유통과 관련된 표준화에 집중하고 있다. TTA 빅데이터 프로젝트그룹을 통하여 개발된 국내 표준들을 국제 표준화에 반영해 왔으며, 빅데이터 참조 아키텍처 표준이 본격적으로 개발되고 있는 시점에서 국제 표준화 방향과 국내 표준들 간의 상호운용성 검토 등 기 개발된 표준들에 대한 정비와 함께 데이터 신뢰성 등 새로운 표준화 이슈들에 대한 국내에서의 기술적 검토 및 대응이 요구되는 시점이다. 

※ 본 연구는 '빅데이터 상호운용성 지원 표준 개발(2017-0-00455)' 과제의 일환으로 수행됨.

## [참고문헌]

- [1] TD 94(GEN/13) Initial draft Recommendation Y.BigData-reqts, Requirements and capabilities for cloud computing based big data(2013.6., ITU-T SG13)
- [2] ITU-T Y.3502 Information technology - Cloud computing - Reference architecture(2014.8., ITU-T SG13)
- [3] ITU-T Y.3601 Big data - framework and requirements for data exchange(2018.5., ITU-T SG13)
- [4] SG13-TD390/WP2 Draft revised Recommendation ITU-T Y.bdm-sch: Big data - Metadata framework and conceptual model(2019.3., ITU-T SG13)
- [5] ISO/IEC JTC 1/SC42 N0144 Potential NP from SAC for Discussion - Big data reference architecture interfaces - Part 1 Characteristics and capabilities(2018.8., National Body of China)
- [6] ISO/IEC JTC 1/SC42 N0145 Potential NP from SAC for Discussion - Big data reference architecture interfaces Part 2 Best practices(2018.8., National Body of China)
- [7] ISO/IEC JTC 1/SC42 N0128 Potential NP from BIS for Discussion - Business Process Management - Data Analytics(2018.8., National Body of India)
- [8] <http://dmg.org/pfa/docs/motivation/>
- [9] <https://datatransferproject.dev/>