

데이터산업 중장기 발전 전략

2016. 12.

주관기관 : 한국데이터진흥원

수행기관 : 투이컨설팅

제 출 문

한국데이터진흥원 원장 귀하

본 보고서를 “데이터산업 중장기 발전 전략”에 관한 보고서로 제출합니다.

2016. 12.

주 관 기 관 : 한국데이터진흥원

총괄 책임자 : 이 창 한 (한국데이터진흥원 정책기획실 실장)

사업 참여자 : 이 중 서 (한국데이터진흥원 정책기획실 차장)

임 태 훈 (한국데이터진흥원 정책기획실 책임)

하 진 희 (한국데이터진흥원 정책기획실 선임)

강 재 혁 (한국데이터진흥원 정책기획실 선임)

김 형 건 (한국데이터진흥원 정책기획실 선임)

김 초 룡 (한국데이터진흥원 정책기획실 연구원)

수 행 기 관 : 투이컨설팅

총괄 책임자 : 김 인 현 (투이컨설팅 대표)

사업 참여자 : 김 찬 수 (투이컨설팅 상무)

김 이 환 (투이컨설팅 선임)

문 보 라 (투이컨설팅 선임)

이 정 수 (투이컨설팅 선임)

한 드 립 (투이컨설팅 선임)

백 새 봄 (투이컨설팅 사원)

목 차

I . 추진배경	1
II . 데이터산업 분석	6
1. 데이터 원칙	6
2. 데이터 제도	16
3. 데이터 기반 비즈니스 모델(Data Driven Business Model)	21
4. 개인 데이터(Personal Data)	38
5. 데이터 전문인력	47
6. 데이터 기술	63
III . 추진과제	72
IV . 추진일정	85
※ 참고문헌	89

〈표 차례〉

표 1 데이터 확보방식의 변화	1
표 2 데이터 원칙 프레임워크	6
표 3 국외 주요 데이터 원칙 비교	14
표 4 국가별 데이터산업 규제	17
표 5 입법이 지연되고 있는 규제 현황	19
표 6 캠브리지 DDBM 분류 결과	22
표 7 A유형 국내사례: 말랑스튜디오	25
표 8 A유형 해외 사례1: 그립	26
표 9 A유형 해외 사례2: 데이터시프트	26
표 10 B유형 국내사례: 노리	27
표 11 B유형 해외사례: 그래너파이	28
표 12 C유형 국내사례: 노매드커넥션	29
표 13 C유형 해외사례1: 스왑니	30
표 14 C유형 해외사례2: 믹스패널	30
표 15 D유형 국내사례: 데일리	31
표 16 D유형 해외사례: 길드	32
표 17 E유형 국내사례: 나인플라바	33
표 18 E유형 해외사례: 액시엄	34
표 19 F유형 국내사례: 파킹스퀘어	35
표 20 F유형 해외사례: 팩추얼	36
표 21 ‘개인 데이터’ 정의	38
표 22 각 국의 개인 데이터 활용 정책	39
표 23 MesInfos 연도별 프로젝트	43
표 24 국내 데이터 직무 구분 비교	47
표 25 국내 정부·공공기관 주도 인력 양성 프로그램	50
표 26 국내 데이터 관련 학부과정 개설 현황	52
표 27 국내 대학원 데이터 전문인력 양성 과정 현황	53
표 28 국내 민간 교육기관의 데이터 인력양성 과정 현황(일부)	54
표 29 데이터 인큐베이터 - 데이터 과학자 양성프로그램	56
표 30 데이터 사이언스스쿨의 온라인 교육과 온라인+오프라인 교육 방법	57

표 31 국내 데이터 전문인력 채용 요구경력 현황	58
표 32 국내외 데이터 관련 자격증	59
표 33 머신러닝 대표 사례	65
표 34 해외 비식별화 지침 특성	67

〈그림 차례〉

그림 1 데이터산업의 부문별 시장 규모	3
그림 2 데이터산업의 6가지 핵심요인	4
그림 3 캠브리지 DDBM의 대표적인 6가지 유형	23
그림 4 국내외 데이터 기반 비즈니스 현황 비교	24
그림 5 2015년 전체 산업 별 데이터 직무 인력 수요	48

I. 추진배경

- 모든 것이 연결되고 데이터화 되어 저장되는 디지털 시대¹⁾이자 데이터가 핵심 자원으로 활용되는 데이터 경제 시대 도래
 - 빅데이터, 사물인터넷, 머신러닝 기술 등 디지털 기술은 데이터 경제 시대 촉진
 - ICT의 주도권이 ‘인프라’에서 ‘데이터’로 이동하면서 데이터가 新성장동력 및 신규 자원으로 인식
- 빅데이터, 오픈데이터뿐만 아니라 개인 데이터(Personal data)를 아우르는 데이터산업 육성 방안 마련 필요
 - 데이터 확보 방식이 공급자 중심에서 사용자 중심으로 바뀌면서, 개인 데이터의 경제적 가치 증대와 개인 사용자 중심의 데이터 수집/확보 필요성이 강조

표 1 데이터 확보방식의 변화

구분	‘공급자’ 중심의 데이터 확보	‘사용자’ 중심의 데이터 확보
주요 내용	특정 서비스 또는 플랫폼 내에서의 사용자 이용 데이터를 확보하여 형태 분석	사용자를 중심으로 다양한 서비스 및 플랫폼에서의 이용데이터를 확보하여 성향 분석

출처: LG 경제연구원(2012.12.05.), 이종근,

‘빅데이터 시대의 스마트 비서 경쟁 시작되었다’ LG Business Insight

- ‘데이터’ 기반(Data-driven) 산업 활성화에 대한 관심 고조
 - OECD 보고서에 따르면 데이터가 ▲신규 제품과 서비스 생성 ▲공정의 최적화 및 자동화 ▲마케팅 개선 ▲새로운 조직 관리방식 혹은 의사결정 방식의 개선 ▲연구개발 개선 등 거의 모든 비즈니스 영역에 영향을 미칠 것이라 분석
 - EU, OECD, 미국, 뉴질랜드 등은 데이터 주도 경제, 데이터 주도 혁신 전략 등을 계획하거나 이미 추진 중

해외 주요 데이터 기반 정책 동향

○ ‘데이터 기반 경제(DDE : Data-driven Economy)’ 전략

▷ EU : Towards a thriving Data-driven Economy

- 2014년 2월 유럽위원회는 유럽각료이사회가 단일시장 아래에서 빅데이터와 클라우드 컴퓨팅을 위해 적절한 프레임워크를 제공할 수 있는 유럽연합 실행계획 수립을 요청함에 따라 ‘데이터 기반 경제 추진을 위한 실행 계획’ 발표
- 데이터혁명 초기단계에서 EU가 데이터경제의 세계적 경쟁력을 확보할 수 있도록 관련 생태계 활성화와 기반 마련을 집중 추진
- 생태계 활성화 : ▲ 데이터분야 공공-민간 파트너십 ▲ 데이터기업 인큐베이팅 ▲ 기초기술 육성 ▲ 데이터시장 모니터링
- 기반 마련 : ▲ 데이터 가용성 및 상호운용성(개방정책 활성화, 데이터 처리도구 및 방법론 보급, 표준 등) ▲ 인프라 개선(클라우드 컴퓨팅, 사물인터넷, 공공데이터 인프라 등) ▲ 법제화(개인정보보호, 데이터 오너십과 이전 등)

○ ‘데이터 기반 혁신(DDI : Data-driven Innovation)’ 전략

▷ OECD : Exploring Data-Driven Innovation : Big Data for Growth and Well-Being

- 2013년 OECD는 데이터 활용 시 혁신과 생산성 향상 가능성이 높은 5대 분야(온라인광고, 보건·의료, 설비, 물류·교통, 공공행정) 선정
- 사회·경제 활동에 대한 마이그레이션 증가와 데이터 수집·전송·저장·분석 비용의 하락, 데이터양의 폭발적 증가 예상
- 2015년 ‘성장과 웰빙을 위한 빅데이터’에서 데이터 가치 사슬에 따라 분석서비스제공자, 데이터기반기업가, 데이터제공자, IT인프라제공자, 인터넷서비스제공자 5가지로 주요 구성원을 선정하고 그 역할을 정의
- 데이터 기반 경제로의 전환을 위해 ‘창조적 파괴’의 부정적 효과에 대한 고려와 개방의 사회적 편익과 개인조직에 대한 권리 침해간의 균형을 언급함

▷ 미국 : Data-driven Innovation : A Guide for policymakers

- 2013년 SW산업협회는 데이터 기반 혁신으로 경제적·사회적 가치를 창출하기 위한 균형 잡힌 데이터관리(stewardship) 정책을 상무부에 제안
- 리스크와 이익의 균형 고려, 개인정보보호 수준의 차등 적용 등 데이터 수집·분석을 원천적으로 제약하는 정책의 지양
- 산업계 기반의 개방형 표준 개발, 공공-민간 파트너십 등 장기적 발전을 감안해 의무 부과보다 자율생태계 중심의 유연하고 개방적인 정책 건의

▷ 뉴질랜드 : Data Driven Innovation in NZ

- 2015년 뉴질랜드 정부는 오픈정부데이터 사용 격려, 데이터의 스마트한 사용에 대한 정부 리더십, DDI 추진을 위한 훈련 및 교육 지원 관련 정책을 수립
- 기업과 개인을 보호하며 해를 끼치지 않는 정책 수립 필요
- 가장 큰 영향을 받을 산업으로 건설, 제조, 유틸리티, 파이낸스&보험 분야 예측

□ 데이터를 축적하고 정보화하는 데 초점을 둔 기존 데이터베이스산업에서 벗어나 데이터를 분석하여 의미를 찾아내고 이를 실행하는 산업으로 빠르게 성장 중

- 인터넷과 모바일 기술의 발전 및 클라우드, 사물인터넷 등과 같이 새로운 분야의 등장으로 데이터 사용량이 폭발적으로 증가하면서 전통적인 데이터베이스산업의 경계가 허물어지고 데이터를 기반으로 하는 신기술 시장 영역까지 포함하면서 기존의 구축과 축적에서 활용과 융합 관점으로 바뀔



□ 데이터산업의 문제 해결과 지속적 발전을 위하여 분석한 결과 한국 데이터진흥원에서 연구 중인 주요 영역을 포함하여 해외에서 중요한 가치 창출로 부각되는 개인 데이터(Personal Data)까지 6가지 핵심 요인이 도출됨

- 6가지 측면: 데이터 원칙, 데이터 제도, 데이터 기반 비즈니스 모델, 개인 데이터, 데이터 전문인력, 데이터 기술
- 각각의 요소를 ▷ 개념, ▷ 주요 현안 제시, ▷ 국내외 현황 분석, ▷ 주요 시사점 분석을 통해 데이터산업에 대한 중장기 발전 과제 도출

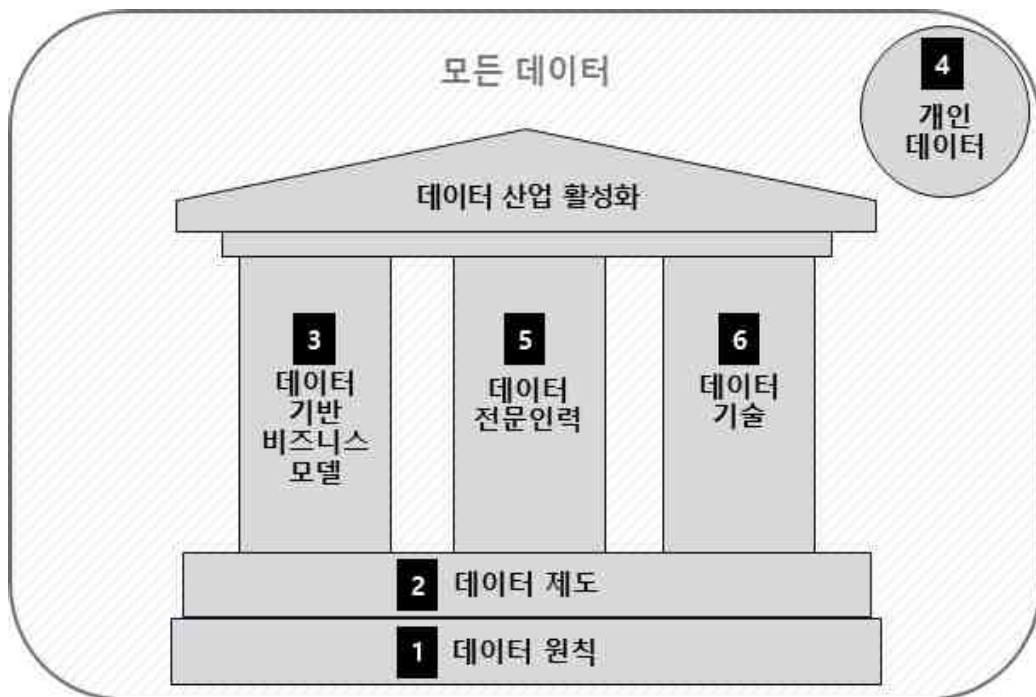


그림 2 데이터산업의 6가지 핵심요인

- ① **데이터 원칙** : 데이터산업의 기반이 되는 주요 원칙으로, 데이터 활성화를 위한 원칙과, 데이터 직간접 활동을 규제하는 원칙으로 구분
- ② **데이터 제도** : 데이터 경제가 원활히 동작할 수 있도록 추진하는 다양한 가이드라인, 지침, 법규 등을 포괄적으로 가리키는 체계

- ③ 데이터 기반 비즈니스 모델 : 데이터의 수집 및 통합부터 데이터 분석을 통한 가치 창출을 위한 다양한 관점의 데이터 기반 비즈니스 모델
- ④ 개인 데이터 : 개인에 관련된 데이터로, 2011년 세계경제포럼에서 빅데이터와 더불어 새로운 가치창출의 수단으로 주목받고 있음
- ⑤ 데이터 전문인력 : 데이터 생성, 운영, 관리, 분석 및 보안, 성능, 품질 향상 등 데이터 관련 전문 스킬/지식을 보유한 인력
- ⑥ 데이터 기술 : 데이터 수집, 저장, 분석, 처리, DBMS 관련 전통적인 데이터 기술부터 머신러닝, 사물인터넷, 오픈소스 기반 플랫폼 등 최신 테크놀로지까지 데이터산업의 기반이 되는 기술

II. 데이터산업 분석

1. 데이터 원칙

□ 개념

- 데이터 원칙은 데이터산업 활성화 원칙, 데이터 직간접 활동을 규제하는 원칙으로 구분²⁾
 - 데이터산업 활성화 원칙: 세부적으로 활성화 지지 원칙 및 규제장벽 제거 원칙으로 구분. 최대한 데이터를 활용하기 위해 규정한 원칙으로 주로 일반적인 데이터 공개 원칙, 공공데이터 활용 원칙 등이 해당
 - 데이터 직간접 활동 규제 원칙: 데이터 직접 활동 규제 원칙과 데이터 간접 활동 규제 원칙으로 구분. 주로 개인정보의 수집, 저장, 조합, 분석, 활용, 교환 시의 문제 발생을 방지하고자 데이터 직간접 활동을 규제하는 내용으로 구성

표 2 데이터 원칙 프레임워크

원칙 유형		세부 내용
데이터산업 활성화 원칙	활성화 지지 원칙	- 직접 편딩 - 생태계 발전 - 오픈 데이터 - 스킬 개발 - 기술 표준 외
	규제 장벽 제거 원칙	- 규제 명확화 - 신뢰기반 데이터 보호 - 국제 상호간 호환 가능한 데이터 운영 외
데이터 직간접 활동 규제 원칙	데이터 직접 활동 규제 원칙	- 데이터 수집 시 - 데이터 저장 시 - 데이터 조합 및 다른 목적으로 사용 시 - 데이터 분석 시 - 데이터 활용 시 외
	데이터 간접 활동 규제 원칙	- 회사 간 데이터 교환 시 - 나라 간 데이터 교환 시 - 데이터 보안 외

□ 주요 현안

- 국내는 해외 대비 공공데이터 활성화 원칙이 구체적이지 못하거나 명문화 되지 않았고, 민간 주도의 데이터 활용/보호를 위한 원칙 전무

□ 현황 분석

○ 영국

- 공공 데이터 원칙(Public Data Principles)³⁾: 2012년 공공부문 투명성 위원회(Public Sector Transparency Board)를 통해 공공 데이터 공개의 기본 전제인 '공공 데이터 원칙'을 공개

영국 '공공 데이터' 주요 원칙

- 공공 데이터는 어떠한 형식으로 언제 배포되는지 명시
- 공공 데이터는 재사용이 가능하도록 기계 판독이 가능한 형태
- 공공 데이터는 오픈 라이선스로 영리 목적으로 이용 가능해야 함
- 공공 데이터는 사용하기 쉽도록 하나의 접근 포인트로 쉽게 접근 및 검색이 가능하여야 함(예: www.data.gov.uk)
- 공공 데이터는 공개 표준을 준수하며, 월드와이드웹 컨소시엄 권장사항을 준수
- 동일한 주제인 서로 다른 부서의 공용 데이터는 동일한 표준 형식 및 정의로 게시되어야 함
- 정부 웹사이트의 공개 데이터는 재사용 가능한 형태로 게시되어야 함
- 데이터 공개는 시기적절하고 세분화되어야 함
- 공공 데이터는 링크를 포함한 개방형 표준 형식으로 빠르게 사용할 수 있어야 함
- 공공 데이터는 합법적인 방법으로 자유롭게 사용 가능하여야 함
- 공공 데이터는 응용 프로그램 또는 등록 절차 없이 사용 가능하며, 사용자의 세부 정보는 필요하지 않아야 함
- 공공 기관은 공공 데이터의 재사용을 적극적으로 장려해야 함
- 공공 기관은 데이터 보유 목록을 게시하고 유지하여야 함
- 공공 기관은 데이터셋에 대한 메타데이터를 게시하여야 하며, 데이터의 형식, 출처, 의미에 대한 설명을 게시하여야 함

- 데이터 과학 윤리 프레임워크(Data Science Ethical F/W)⁴⁾: 2016년 5월 영국의 내각사무처(Cabinet Office)에서 정책 및 의사 결정 과정에서 데이터 사용이 증가함에 따라 고려해야 할 사용자 이익, 개인정보보호 및 보안 등에 관한 법적 측면과 원칙 등을 설명한 '데이터 과학 윤리 프레임워크' 발표

‘데이터 과학 윤리 프레임워크’ 6가지 주요 원칙

- ① 명확한 사용자의 니즈와 공공의 이익에서 시작하라.
 - 데이터 과학은 정책입안자들에게 근거를 만들어주고, 더 빠르고 정확한 운영상의 결정을 할 수 있는 큰 기회를 제공한다. 공공의 이익에 관해 분명히 해두는 것이 데이터의 민감성(원칙2)과 사용하고자 하는 방법(원칙3)을 정당화하는 데 도움을 준다.
- ② 개입을 최소화한 데이터와 도구를 사용하라.
 - 공공의 이익을 달성하기 위해 항상 필요한 최소한의 데이터만을 사용해야 한다. 가끔 민감한 개인정보를 사용해야 하는 상황이 발생할 것이다. 그럴 경우, 개인정보를 식별할 수 없도록 하거나 상위 레벨로 데이터를 집계하고, 데이터셋(datasets)의 조건부 검사나 합성 데이터(synthetic data) 등과 같이 개인정보를 안전하게 사용할 수 있는 단계를 사용하면 된다.
- ③ 강력한 데이터 과학 모델을 개발하라.
 - 좋은 머신러닝(machine learning) 모델은 전통적인 방법보다 훨씬 많은 데이터를 더욱 빠르고 정확하게 분석할 수 있다. 만약 알고리즘이 민족성(ethnicity)과 같이 보호되는 특성들을 사용하는 것이라면 데이터의 품질과 대표성을 통해 생각하고, 의도치 않은 결과들을 고려하라. 복잡한 결정들은 정책에 대한 폭 넓은 지식 또는 운영 전문가(operational experts)를 필요로 할 수도 있다.
- ④ 대중의 인식에 대해 신경 써라.
 - 데이터 보호법(Data Protection Act)은 사람들이 자신의 개인정보가 어떻게 합리적으로 사용되어야 하는지에 대한 이해를 필요로 한다. 대중의 인식이 변하고 있다는 것을 알아야 한다. 소셜 미디어 데이터(social media data), 상업적 데이터(commercial data)와 웹에서 수집된 데이터들은 세상에 대한 이해력을 높일 수 있으나, 다른 약관(terms and conditions)과 동의 수준(levels of consent)이 함께 따라온다.
- ⑤ 가능한 공개적이고 책임감 있게 하라.
 - 공개된다는 것은 데이터 과학의 공공의 이익에 대해 논할 수 있다는 것이다. 목적 자체를 위태롭게 하지 않는 선에서 최대한 사용 도구, 데이터, 알고리즘에 대해 공개하라. 알기 쉬운 영어로 설명을 제공하고, 국민이 잘못된 결정이라 생각하는 경우 되돌릴 수 있도록 하라. 프로젝트는 감독 하에 진행되도록 하고, 전체 수행과정에서 책임 소재가 있는 부분을 명확히 하라.
- ⑥ 데이터 보안을 유지하라.
 - 대중들은 당연히 자신들의 데이터가 소실되거나 도용되지 않을까에 대해 우려한다. 정부는 적절한 보안 도구들을 가지고 개인정보 데이터를 보호해야 하는 법적인 의무를 지닌다.

- DMA(Direct Marketing Association)의 데이터 보호 원칙⁵⁾: 영국 DMA는 2014년 데이터 거래를 통해 얻을 수 있는 가치의 신뢰성을 부여하기 위해 데이터 가이드 발표

영국 DMA의 데이터 가이드 내 8가지 데이터 보호 원칙

- ① 공정하고 합법적으로 처리
 - 고객의 개인정보는 공정하고 합법적으로 처리되어야 하며, 특정 조건이 충족되지 않으면 처리되지 않아야 함
- ② 제한된 목적으로 처리
 - 개인 데이터는 하나 이상의 특정하고 합법적인 목적을 위해서만 획득되어야 하며, 목적과 양립할 수 없는 방식으로 처리되어서는 안 됨
- ③ 필요한 범위 내 처리
 - 개인 정보는 처리 목적에 적절하게 처리되어야 함
- ④ 정확성과 최신성 유지
 - 개인 데이터는 정확하여야 하며, 필요 시 최신으로 유지하여야 함
- ⑤ 필요한 기간에만 보존
 - 어떠한 목적으로 처리된 데이터는 그 목적에 필요한 것보다 오래 보관되어서는 안 됨
- ⑥ 권리 보장
 - 개인 데이터는 데이터 보호법(Data Protection Act, DPA)에 의한 데이터 주체의 권리에 따라 처리되어야 함
- ⑦ 보호
 - 개인 데이터를 불법적인 처리, 손실, 파괴 등에 보호하기 위한 적절한 기술적, 조직적 조치가 취해져야 함
- ⑧ 허가 없이 외국 유출 금지
 - 적절한 수준의 보호를 보장하지 않는 한 개인 데이터는 유럽 경제 지역 이외의 국가나 영토로 반출 될 수 없음

o 미국

- DMA(Data&Marketing Association)의 가이드라인⁶⁾: 1917년 설립된 미국 DMA는 1,400개 이상의 브랜드가 참여하는 미국 최대 데이터 기반 마케팅 관련 커뮤니티임. DMA는 데이터 활용 관련 정부의 법률 규제보다 민간 주도의 자율적 가이드라인이 기술, 경제 및 사회적 상황변화에 보다 적절하다고 판단하고, ‘윤리적 비즈니스 프랙티스를 위한 DMA 가이드라인(DMA Guidelines for Ethical Business Practice)’ 제시(향후 ‘데이터 표준 2.0(Data Standards 2.0)’으로 업그레이드 예정)

미국 DMA의 가이드라인 중 일부 발췌

마케팅 데이터의 수집, 활용, 유지 관련 원칙(31조항)

A. 소비자 선택 권리와 프라이버시 관련 정보 제공

- 마케터는 소비자가 회사 또는 조직과 마케팅 관련 의견교환, 조정, 또는 제거할 수 있도록 중간매체를 제공해야 하며, 관련 기관 또는 조직에 대한 데이터 수집, 활용 및 정보의 교환에 관한 프라이버시 정책을 확보해야 한다. 마케터 연락처 관련 정보(예를 들어 웹사이트, 전화번호 또는 주소)는 마케팅 오퍼 시 또는 소비자의 요구 시 항상 제공되어야 한다.
- 온라인 마케터는 38조항(온라인 정보 및 온라인 행동 광고 관련 조항)에 따라 사전고지를 해야 한다.
- 이메일 마케터는 39조항(무선 기기에 전송되는 모바일 서비스 상업적 요구 메시지 관련 조항) 및 CAN-SPAM Act.(정보의 수집과 인터넷 상에서의 원치 않는 이메일 홍보를 제한하는 법률)에 따라 사전고지를 제공해야 한다.
- 모바일 마케터는 54조항(모바일 기기 접속에 대한 사전동의 획득 관련 조항) 및 55조항(모바일 마케팅 프랙티스 관련 고지 제공에 대한 조항)에 따라 사전동의를 받고 사전고지를 해야 한다.
- 연락처 공지정보는 소비자가 쉽게 찾고, 이해하고, 행동하기 수월하게 표현되어야 한다.
- 마케터는 기존 소비자에게 소비자 관련 데이터의 대여, 맞교환, 또는 전달에 관한 정책을 주기적으로 고지해야 하며, 마케팅 프로세스에서 옵트아웃 할 수 있는 기회를 제공해야 한다. 모든 옵트아웃 요구는 신속히 처리되어야 한다.
- 소비자의 인하우스 정보 활용 자제 요청은 그 개인 정보의 외부 교환 역시 멈추길 원하는 것으로 이해하고 옵트아웃 처리해야 한다.
- 소비자 요구 시, 마케터는 그 소비자에 대한 개인식별정보를 취득한 원천에 대해 밝혀야 한다.

B. 소비자 권리 처리

- 소비자가 향후 마케팅 오퍼에서 제외해 달라는 요청은 처리되어야 한다:
 - 소비자 요구에서 30일 이내 또는 법에서 정한 기간 중 보다 짧은 기간 안에 처리
 - 요구를 받은 시점에서 적어도 3년 이내의 기간 내에 처리
- 계열사, 분사, 자회사가 다른 회사 또는 브랜드 이름으로 마케팅하고 있을 때, 소비자가 이를 서로 다른 조직으로 인식할 경우, 각 관련 회사 또는 브랜드는 소비자의 요청을 별도로 처리해야 한다.
- 마케터는 이 가이드라인을 따르면서 소비자가 마케팅 채널에 무관하게 어떤 비용을 감수할 필요없이 소비자 선호 요구사항을 반영할 수 있도록 내부 정책 및 실례를 수립해야 한다. 관련 정책이 상당 부분 바뀌어야 할 경우가 발생 시, 마케터는 소비자 관련 데이터의 대여, 맞교환, 또는 전달에 관한 변경 정책을 의무적으로 고지해야 하며, 그 시기에 마케팅 프로세스에서 옵트아웃 할 수 있는 기회를 제공해야 한다.

- **썬라이트 재단의 공공 데이터 원칙**⁷⁾: 썬라이트 재단(Sunlight Foundation)은 열린 정부를 지지하고 미국의회, 행정부 및 주·연방 정부의 투명성과 책임성을 높이기 위해 2006년 4월에 설립된 비영리단체로 2007년 10월, 30명의 정부 공개 지지자들이 모여 주정부와 연방 정부의 불일치/불완전 데이터들을 활용하기 위한 10가지 원칙을 정의

썬라이트 재단의 주요 공공 데이터 원칙

- ① 완전성
 - 정부가 발표한 데이터셋은 가능한 완전해야하며, 특정 주제에 대해 기록된 내용 전체를 반영하여야 함
- ② 우의성
 - 정부가 발표한 데이터셋이 주요 소스 데이터야 함
- ③ 적시성
 - 정부가 발표 한 데이터셋은 적시에 대중에게 공개되어야함
- ④ 물리적 및 전자적 접근 용이성
 - 정부가 발표 한 데이터셋은 물리적 또는 전자적 수단을 통해 정보를 얻기 용이하여야 함
- ⑤ 기계 가독성
 - 정보는 널리 사용되는 파일 형식으로 저장되어 기계 처리에 적합하여야 함
- ⑥ 차별 금지
 - 모든 사람이 자신을 식별하지 않고도 언제든지 데이터에 액세스 할 수 있어야 함
- ⑦ 공통적인 표준 사용
 - 데이터가 저장된 파일을 읽을 수 있는 프로그램을 제조하는 회사가 단 한 곳이라면 해당 정보에 대한 액세스는 회사의 처리 프로그램 사용에 따라 다름(ex: Microsoft Excel). 소프트웨어 라이선스 없이도 저장된 데이터에 액세스 할 수 있는 자유로운 대체 형식이 종종 존재하며, 이 비용을 감소시키면 더 많은 잠재적 사용자 풀에서 데이터를 사용할 수 있음
- ⑧ 라이선스
 - 공개 정보를 정부 저작물로 명확하게 표시하고 공개 도메인의 일부로 사용하는 것에 대한 제한 없이 사용할 수 있어야 함
- ⑨ 영구성
 - 정부에 의해 온라인으로 공개 된 정보는 영구적으로 아카이브에 온라인으로 제공되어야함
- ⑩ 사용 비용
 - 사용 요금을 부과하면 정보를 사용 할 의사가 있는지 왜곡됨

o 국내

- 공공데이터의 제공 및 이용 활성화에 관한 법률⁸⁾: 공공데이터를 국민이 최우선적으로 이용할 수 있도록 보장하고, 공공기관에 공공데이터 제공의무를 부여하며, 효과적인 민간제공과 이용 활성화를 지원할 수 있는 법적 근거를 마련하기 위하여 2013년 7월 제정, 2013년 10월 시행. 공공데이터를 이용한 창업에 대한 지원 근거를 마련하고 공공기관의 중복·유사 서비스 개발을 금지함으로써 공공데이터 이용자의 권익을 보호하고 공공데이터를 이용한 창업을 활성화하기 위해 2016년 4월 개정

‘공공데이터의 제공 및 이용 활성화에 관한 법률’ 기본원칙

- ① 공공기관은 누구든지 공공데이터를 편리하게 이용할 수 있도록 노력하여야 하며, 이용권의 보편적 확대를 위하여 필요한 조치를 취하여야 한다.
- ② 공공기관은 공공데이터에 관한 국민의 접근과 이용에 있어서 평등의 원칙을 보장하여야 한다.
- ③ 공공기관은 정보통신망을 통하여 일반에 공개된 공공데이터에 관하여 제28조 제1항 각 호의 경우를 제외하고는 이용자의 접근제한이나 차단 등 이용저해행위를 하여서는 아니 된다.
- ④ 공공기관은 다른 법률에 특별한 규정이 있는 경우 또는 제28조 제1항 각 호의 경우를 제외하고는 공공데이터의 영리적 이용인 경우에도 이를 금지 또는 제한하여서는 아니 된다.
- ⑤ 이용자는 공공데이터를 이용하는 경우 국가안전보장 등 공익이나 타인의 권리를 침해하지 아니하도록 법령이나 이용조건 등에 따른 의무를 준수하여야 하며, 신의에 따라 성실하게 이용하여야 한다.

- 개인정보보호법 : 개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 2011년 09월 제정

‘개인정보보호법’의 개인정보 보호 원칙

제3조(개인정보 보호 원칙)

- ① 개인정보처리자는 개인정보의 처리 목적을 명확하게 하여야 하고 그 목적에 필요한 범위에서 최소한의 개인정보만을 적법하고 정당하게 수집하여야 한다.
- ② 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 적합하게 개인정보를 처리하여야 하며, 그 목적 외의 용도로 활용하여서는 아니 된다.

- ③ 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 한다.
- ④ 개인정보처리자는 개인정보의 처리 방법 및 종류 등에 따라 정보주체의 권리가 침해받을 가능성과 그 위험 정도를 고려하여 개인정보를 안전하게 관리하여야 한다.
- ⑤ 개인정보처리자는 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하여야 하며, 열람청구권 등 정보주체의 권리를 보장하여야 한다.
- ⑥ 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하여야 한다.
- ⑦ 개인정보처리자는 개인정보의 익명처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다.
- ⑧ 개인정보처리자는 이 법 및 관계 법령에서 규정하고 있는 책임과 의무를 준수하고 실천함으로써 정보주체의 신뢰를 얻기 위하여 노력하여야 한다.

o 오픈 데이터 헌장(International Open Data Charter)⁹⁾

- 2013년 영국 북아일랜드에서 진행된 G8 정상회의(독일, 미국, 이탈리아, 영국, 일본, 캐나다, 프랑스, 러시아)에서 발표
- 오픈데이터를 ‘정보시대의 중요한 자원’으로 평가하였으며, 오픈데이터의 활용을 통해 시민 생활의 질적 향상과 혁신, 경제 성장과 고용 창출로 이루어 질 것으로 기대
- 현대 오픈 데이터 헌장은 캐나다, 멕시코, 영국, Centre for Internet & Society(인도), 국제 개발 연구 센터, Latin American Open Data Initiative, ODI(Open Data Institute), 개방지식인터넷내셔널/OD4D, 월드와이드 웹 재단, 오미디아르 네트워크 재단 등에 의해 주도되고 있음
- 우리나라를 포함하여 16개 국가, 25개의 지방정부가 오픈데이터 헌장에 가입 및 참여 중

오픈데이터 헌장 원칙

- ① 기본적으로 공개될 것
- ② 시의적절하고 포괄적일 것
- ③ 접근 가능하며 사용이 가능할 것
- ④ 비교할 수 있고 상호연동이 가능할 것
- ⑤ 거버넌스와 시민의 참여를 증진시킬 것
- ⑥ 포괄적인 개발과 혁신을 위해 공개될 것 등

0 데이터 원칙 프레임워크에 국외 주요 공공데이터 관련 원칙 비교

표 3 국외 주요 데이터 원칙 비교

원칙 유형	세부 내용	영국	미국
데이터 산업 활성화 원칙	활성화 지지 원칙	[공공데이터] - 데이터 형식/배포시기 명시 - 재사용이 가능한 기계 판독가능형태로 제공 - 쉽게 접근 가능하도록 함(예: www.data.uk) - 빠르게 사용가능한 형태로 제공 - 공공기관 공공데이터 사용 적극 권장 - 데이터 보유 목록 제공 - 데이터 형식, 출처, 의미에 대한 설명 게시 [데이터 과학 윤리 Framework] - 사용자 니즈/공공 이익 고려 - 개입 최소화한 데이터와 도구 사용 - 강력한 데이터과학 모델 개발 - 대중 인식 고려 - 책임감 있게 공개적으로 제공	[오픈라이트 공공데이터] - 적시에 대중에게 공개해야함 - 데이터셋은 얻기 용이하여야 함 - 널리 사용 가능한 파일형식으로 저장 - 기계 처리에 적합하도록 함 - 공개된 정보는 영구적으로 아카이브에 온라인으로 제공 - 모든 사람이 자신을 식별하지 않고도 언제든지 데이터에 액세스 할 수 있어야 함
	규제 장벽 제거 원칙	[공공데이터] - 공개 표준 및 월드와이드웹 컨소시엄 권장사항 준수	[오픈라이트 공공데이터] - 공개 정보를 정부 저작물로 명확하게 표시하고 공개 도메인의 일부로 사용하는 것에 대한 제한 없이 사용할 수 있어야 함 - 공통적인 표준 사용
데이터 산업 직간접 활동 규제 원칙	데이터 직접 활동 규제 원칙	[공공데이터] - 데이터 사용은 합법적인 방법으로 자유롭게 사용	[오픈라이트 공공데이터] - 데이터셋은 가능한 완전해야함 - 주요 소스 데이터야 함
	데이터 간접 활동 규제 원칙	[공공데이터] - 사용자 세부정보는 불필요 [데이터 과학 윤리 프레임워크] - 데이터 보안을 유지	[오픈라이트 공공데이터] - 사용 요금을 부과하면 정보를 사용할 의사가 있는지 왜곡됨

□ 주요 시사점

○ 공공데이터 뿐만 아니라 모든 데이터 활용 관점의 데이터 원칙 필요

- 해외 데이터 원칙은 데이터산업 활성화를 위한 데이터 접근 용이성, 사용 용이성, 사용 장려를 위한 원칙*을 구체적으로 제시

* 기계 가독성, 표준 준수, 무료 공개, 재사용 장려, 보유 목록 유지 등

- 이에 반해, 우리나라는 공공데이터 활성화 측면의 데이터 원칙은 규정되어 있으나, 대부분의 데이터는 개인정보보호법에 의해 규제를 받음

○ 민간 중심의 데이터 원칙 명문화 필요

- 민간 주도의 데이터 원칙은 데이터 활성화 측면이 강하고 의무사항이 아니라 기술, 경제, 사회 변화에 보다 빠르게 적용시킬 수 있다는 장점이 있음. 이에 해외 주요 국가에서는 민간 주도의 데이터 원칙*을 장려(특히 마케팅 분야)함

* 미국의 'DMA 가이드라인', 영국의 'DMA 데이터 보호 원칙' 등

- 국내 데이터 관련 원칙은 '공공데이터의 제공 및 이용 활성화에 관한 법률', '개인정보보호법'과 같은 법규를 통해 데이터산업 활성화와 규제를 위한 원칙을 규정하고 있으며, 민간 주도의 데이터 활용 기본 원칙은 전무

2. 데이터 제도

□ 개념

- 데이터산업의 전반적인 운영 및 관리, 개선, 활성화를 위한 법규, 지침, 가이드라인, 정책을 모두 포함하여 가리킴
- 데이터산업은 외형적 성장* 및 높은 잠재력에도 불구하고 질적 고도화가 지연되고 있다는 진단을 받음에 따라 이를 개선하고 발전시킬 수 있는 다양한 정책 추진 중
- * ‘2016 데이터산업 백서’에 의하면 2010년 이후 국내 데이터산업 시장 규모는 연평균 9.3%로 매년 성장세를 유지

국내 데이터산업 활성화를 위한 정책

○ 데이터 유통 활성화 제도

1) 공공데이터 개방을 위한 정책

- 2013년 10월 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’이 시행되면서 공공기관의 공공데이터 개방 노력 가시화
- 정부 3.0의 핵심정책으로 정부와 공공기관이 보유하고 있는 공공데이터 개방 진행 중

○ 빅데이터 활성화 제도

1) 빅데이터산업 발전 전략

- 미래창조과학부는 2013년 관계부처와 합동으로 빅데이터 시장 2배 이상 창출 지원, 글로벌 전문기업 육성, 5천명 이상 고급인력 양성 등을 목표로 하는 ‘빅데이터산업 발전 전략’ 발표
- 데이터 과학자급 전문인력 양성, 현업 재직자 중심의 인력 양성, 대학/대학원에서의 빅데이터 잠재인력 인력양성 추진 중

2) ‘K-ICT 전략’ 재편

- 2016년에는 ICT특성을 고려, K-ICT 2016 전략을 통해 국내 빅데이터 시장규모를 2019년까지 8,797억원 규모로 성장시키고 국가 데이터 허브를 구축하는 등 데이터산업 성장을 주도할 수 있도록 추진 중

3) 빅데이터 활성화를 위한 빅데이터진흥법

- ‘개인정보’가 비식별화 될 경우 민간 또는 공공기업이 이를 자유롭게 사용할 수 있게 데이터 활용을 허용하는 제도로 국회 발의 중

□ 주요 현안

- 국내 데이터산업 제도는 데이터 보호를 위한 규제 측면이 강해, 데이터산업 활성화의 주요 장애 요소임

□ 현황 분석

- 2014년 애널리시스 메이슨(Analysys Mason) 보고서의 분석 결과에 따르면, 국내의 데이터산업 규제는 해외 주요국 대비 엄격한 것으로 나타남(4: 장애 요인, 3: 상당한 수준의 장애 요인, 2: 약간의 장애 요인, 1: 규제가 약한 수준)

표 4 국가별 데이터산업 규제

국가	데이터 수집	데이터 보관	융합 및 타목적 사용	프로파일링 & 차별화	데이터 교환	평균
대한민국	4	4	3	3	4	3.6
유럽 연합	3	2	3	3	3	2.8
미국	2	2	3	3	2	2.4
OECD	1	2	2	3	2	2.0
뉴질랜드	3	2	3	1	2	2.2
싱가포르	2	2	2	2	2	2.0
호주	1	2	2	3	2	2.0

출처: Analysys Mason(2014), 'Data-Driven Innovation in Singapore'

- 호주의 데이터 수집 경우, 민감한 데이터라도 합리적으로 필요하거나 분명한 합의가 있는 경우는 동의 없이 수집 가능
- 뉴질랜드는 차별금지법*에 따라 대부분 데이터의 수집을 허가함

* 인권위원회법(1997)과 인권법(1993)

국가별 데이터 규제 비율 비교¹⁰⁾

1. 대한민국
 - 1) 데이터 수집 : 한정된 수집. 동의 없이 수집 불가
 - 2) 데이터 보관 : 동의를 얻을 시부터 특정기한 제한
 - 3) 융합 및 타목적 사용 : 서비스 제공자와 사용자 간 동의하에 가능
 - 4) 프로파일링 & 차별화 : 특정한 사항 없음. 상호 동의하에 가능
 - 5) 데이터 교환 : 동의하에 가능
2. 유럽 연합(EU)
 - 1) 데이터 수집 : 데이터가 민감하지 않고 서비스 제공자와 사용자 간의 합의가 있지 않은 한 동의 필요
 - 2) 데이터 보관 : 동의를 얻을 시부터 특정기한 제한

- 3) 융합 및 타목적 사용 : 서비스 제공자와 사용자 간 동의하에 가능
- 4) 프로파일링 & 차별화 : 옵트 아웃 필요
- 5) 데이터 교환 : 타 국가의 적절한 보호 하에 제공

3. 미국

- 1) 데이터 수집 : 서비스 제공자와 사용자의 합의 시 수집 가능. 민감하지 않은 데이터 수집 가능
- 2) 데이터 보관 : 동의를 얻을 시부터 특정기한 제한
- 3) 융합 및 타목적 사용 : 서비스 제공자와 사용자 간 동의하에 가능
- 4) 프로파일링 & 차별화 : 서비스 제공자와 사용자 간 동의하에 가능
- 5) 데이터 교환 : 서비스 제공자와 사용자 간 동의하에 가능

4. OECD

- 1) 데이터 수집 : 필요시 동의 필요
- 2) 데이터 보관 : 동의를 얻을 시부터 특정기한 제한
- 3) 융합 및 타목적 사용 : 기존 및 호환되는 목적으로 제한
- 4) 프로파일링 & 차별화 : 없음
- 5) 데이터 교환 : 충분한 보호 하에 가능

5. 뉴질랜드

- 1) 데이터 수집 : 정보는 법적 목적을 위해 수집 가능. 개인은 정보가 수집된 사항에 대해 인지하고 있어야 함
- 2) 데이터 보관 : '법적으로 사용되는 목적'의 데이터는 보유 기간이 한정
- 3) 융합 및 타목적 사용 : 일반적인 규율은 한 가지 목적으로 수집된 개인정보는 허가없이 다른 목적을 위해 사용할 수 없음. 데이터 조합에 대한 특정한 규율은 없음. 데이터 조합 및 타목적 사용 가능.
- 4) 프로파일링 & 차별화 : 차별금지법에 접촉되지 않는 한 대부분 허가함
- 5) 데이터 교환 : 허가없이 주체간 개인 정보 이동은 안 됨.

6. 싱가포르

- 1) 데이터 수집 : 개인의 흥미 및 타당하게 보류 되지 않은 사항은 동의가 필요함.
- 2) 데이터 보관 : 기존 목적에 대해서 보유 기간 한정
- 3) 융합 및 타목적 사용 : 개인 흥미, 동의 된 목적 하에 가능
- 4) 프로파일링 & 차별화 : 개인 흥미, 동의 된 목적 하에 가능
- 5) 데이터 교환 : 개인의 흥미 및 동의하에 가능. 국제적 이동은 제한

7. 호주

- 1) 데이터 수집 : 개인정보간 차이. 합리적으로 필요하거나 분명한 합의가 있는 민감한 데이터(인종, 정치적 성향, 성별, 범죄기록)는 동의 없이 수집 가능. 개인정보 수집은 개인을 통해 직접 받을 수 있으며 제공자는 이에 대해 명시하고 있어야함
- 2) 데이터 보관 : 허가된 목적을 달성 후 무조건 폐기.
- 3) 융합 및 타목적 사용 : 기존 목적과 관련 있는 제2의 사용 가능. 비식별화 정보 리포퍼징 가능.
- 4) 프로파일링 & 차별화 : 인종, 정치성향, 성별, 범죄 기록 등 민감한 정보는 프로파일링 목적으로 사용 불가
- 5) 데이터 교환 : 국제적 이동 가능. 호주 프라이버시 원칙에 필요한 사항 충족 시 이동 가능

o 데이터 수집/보관 제도 현황

- 국내는 일반법인 개인정보보호법에 의해 개인정보 수집이 엄격히 제한됨. 또한, 데이터 수집 체계가 유럽연합과 같은 옵트인(Opt-In) 체계*로 미국의 옵트아웃(Opt-Out) 체계에 비해 데이터 수집 및 보관이 제한적

* 옵트인은 개인이 개인정보 수집을 명시적으로 동의하기 전까지 개인정보의 수집을 금지하는 방식인데 반해, 옵트아웃은 개인이 명시적으로 데이터 수집을 거부하기 전까지 개인정보를 수집할 수 있는 방식

- 이밖에도 위치정보의 보호 및 이용 등에 관한 법률(2005년), 정보통신망 이용 촉진 등에 관한 법률(1999년), 신용정보의 이용 및 보호에 관한 법률(1995년), 공공기관의 개인정보보호에 관한 법률(1994년)을 통해 데이터의 보호를 법률로 정하고 있음

데이터 수집/보관 관련 제도

o 개인정보보호법('11)

- 2011년 9월 30일, 공공부문과 민간부문을 망라하는 개인정보보호에 관한 일반법 제정
- 포괄적인 개인정보 정의 규정, 엄격한 사전 고지·동의 제도, 형사처벌 조항 등 비교적 엄격한 법제 채택
- 개인정보보호법에 의해 개인정보가 담긴 데이터 활용 규제

* 개인정보보호법 제1장 제2조 제1항 따르면 개인정보는 개인을 식별할 수 있는 데이터로 다른 데이터와 결합해서 개인을 식별할 수 있는 경우를 포함함

- 2016년 9월 개인정보 처리 시 안전성 확보를 위한 규정을 강화함

o 위치정보의 보호 및 이용 등에 관한 법률('05)

- 위치정보의 유출 및 오남용으로부터 개인정보의 보호와 건전하고 안전한 정보 통신망 조성

o 정보통신망 이용 촉진 등에 관한 법률('95)

- 정보통신망의 이용을 촉진하고 안정적 관리·운영을 도모하며, 정보통신 서비스를 이용하는 자의 개인정보를 보호하여 정보사회의 기반 조성

o 신용정보의 이용 및 보호에 관한 법률('95)

- 신용정보의 효율적 이용과 체계적 관리를 도모하며 신용정보의 오용·남용으로 부터 사생활의 비밀 등을 보호
- 최근 개정안에 개인정보 수집·이용·제3자 제공 시 사전 동의 예외, 비식별 조치 규정을 추가

o 공공기관의 개인정보보호에 관한 법률('94)

- 공공기관의 컴퓨터에 의해 처리되는 개인정보를 보호하기 위해 그 취급에 관해 필요한 사항을 정한 법률

0 데이터 융합/타목적 사용 제도 현황

- 빅데이터 활성화를 위한 빅데이터진흥법 등 정책적 기반을 마련했지만, 국회에 계류되면서 실제 적용되지 않고 있음

표 5 입법이 지연되고 있는 규제 현황

시기	추진주체	법안명	관련내용
20대 국회	배덕광 의원	빅데이터법 제정안	개인정보가 비식별화 될 경우 민간 또는 공공 기업이 이를 자유롭게 사용할 수 있게 데이터 활용 허용

- 개인생활 로그정보(비정형 데이터)를 유통용 빅데이터로 생성할 수 있도록 개인정보 비식별화 기법인 옵트온(Opt-on)* 방법론을 도입하고자 했지만 정책 연구에만 그침

* 옵트온 방식은 운영계 서비스를 위해 저장/관리하는 개인정보 DB에서 모든 개인 식별 정보를 삭제하고 개인신상정보에 대한 유사 그룹 기반 대표화 과정을 통해 원천적으로 모든 개인정보를 가공/변환함으로써 개별 개인의 활용 동의가 별도로 필요 없는 방식 실제로는 적용되지 못한 연구 사례임

□ 주요 시사점

0 규제 중심의 데이터 제도 개선 필요

- 국내는 개인정보보호법을 포함한 다양한 규제성 법률이 산재해 있으며 해외 대비 데이터 수집/보관/데이터 교환 등 데이터 활용에 대한 규제 수준이 높은 것으로 나타남

0 국내 데이터산업 활성화를 위한 정책연구에 대한 지속적인 연구/지원 필요

- ‘빅데이터진흥법’ 등 국내 실정에 적합한 제도가 제정되고 데이터 활성화 정책이 지속적으로 추진될 수 있도록 해외 사례 연구 및 정책연구에 대한 지속적인 관심과 지원 필요

3. 데이터 기반 비즈니스 모델(DDBM)

□ 개념

- o 데이터의 수집 및 통합부터 데이터 분석을 통한 가치 창출 등 다양한 데이터 및 데이터를 통해 수익을 창출하는 비즈니스 모델을 가리킴
 - ‘데이터 기반(Data-Driven)’이란 어떠한 행동이 직관이나 개인의 경험에 아닌 데이터에 의해 일어나게 되는 과정
 - ‘데이터 기반 비즈니스(Data-Driven Business)’란 데이터를 축적하고 정보화하는 것에서 벗어나 데이터를 분석하여 의미를 찾아내고 이를 실행에 연결시키는 산업

□ 주요 현안

- o 국내는 해외 대비 다양한 데이터 기반 비즈니스 모델(Data-Driven Business Model, DDBM)이 미비

* 데이터 추출/정제/가공을 통한 효과적인 활용으로 수익을 낼 수 있는 새로운 비즈니스 모델 발굴 및 기반 조성 필요

□ 현황 분석

- ① 국내외 DDBM 유형 비교 분석을 위해 캠브리지(Cambridge) 대학에서 연구한 DDBM 프레임워크 적용¹⁾

데이터 기반 비즈니스 모델 연구 방법론

데이터 기반 비즈니스 모델은 7가지 측면(마케팅, 채널, 고객 서비스, 분류 관리, 공급망 관리, 실행·재무 관리, IT)으로 360도 분석한 TCS(TATA Consultancy Services) 모델, 데이터 가치사슬(데이터 수집, 데이터 분석, 비즈니스 인텔리전스)과 데이터 유형(공공, 기업, 개인)으로 분석한 유럽 디지털 포럼 모델, 데이터 가치사슬(생산, 유통, 활용)과 생산물(하드웨어, 소프트웨어, 서비스) 기준으로 국내에 맞는 새로운 분류를 제안한 Kdata 모델 등 다양하게 연구되고 있음

- 캠브리지 DDBM 프레임워크는 스타트업을 대상으로 6가지 측면(데이터 원천, 주요 활동, 공급 형태, 목표 고객, 수익 모델, 특정 비용 측면 이점)으로 분석하였으며, 양적 분석 측면에서 현실적인 모델이라는 평을 받음

캠브리지 DDBM 프레임워크¹²⁾

o 분류 기준 및 결과

- 대부분의 데이터 기반 비즈니스 모델에 공통적인 6가지 측면으로 프레임워크를 구성함
- 100개의 스타트업 대상으로 6가지 측면으로 각각 분석한 내용을 클러스터링 하여 특성 있는 대표적인 비즈니스 유형을 6가지로 추출하였으며, ‘특정비용측면 이점’을 제외한 분류 기준 별 유형 비율은 아래와 같음

(단위 : %)

표 6 캠브리지 DDBM 분류 결과

비즈니스 유형				DDBM 유형 비율						
				A	B	C	D	E	F	
				무료 데이터 수집 및 통합 유형	서비스 분석 유형	데이터 생성 및 분석 유형	무료 데이터 정보 디스커 버리 유형	데이터 통합 유형	멀티 소스 데이터 통합 분석 유형	
분류 기준										
점유율				18	29	17	15	6	15	
데이터 원천	내부	기존재 데이터		0	0	0	0	0	0	
		자가 생성	클라우드소싱	18	0	31	0	17	7	
			트래킹&생성	6	11	88	0	0	21	
	외부	획득데이터		12	11	0	7	0	21	
		고객제공		24	100	13	0	100	79	
		무료	오픈데이터		100	11	19	100	0	93
			소셜 미디어데이터		12	4	0	7	0	7
			웹 크롤링		71	11	13	50	0	57
주요 활동	데이터 생성	크롤링		35	0	6	50	0	43	
		트래킹 & 클라우드소싱		12	4	100	0	17	29	
	데이터 획득(Data acquisition)		24	21	6	21	17	29		
	처리(Processing)		100	18	0	0	100	93		
	통합(Aggregation)		12	82	69	86	0	93		
	분석	현황(Descriptive)		0	50	13	21	0	14	
		예측(Predictive)		0	11	6	7	0	7	
		규범(Prescriptive)		24	39	44	7	33	36	
	시각화(Visualisation)		100	36	19	29	83	50		
	배포(Distribution)		35	0	6	50	0	43		
공급 형태	데이터		12	0	0	0	0	14		
	정보/지식		88	100	88	100	83	93		
	비데이터(제품/서비스)		0	0	13	7	0	0		
목표 고객	B2B		71	96	63	86	83	86		
	B2C		47	18	50	21	33	21		
수익 모델	자산 판매		0	0	19	0	0	0		
	대출/임대/리스		6	0	0	0	0	0		
	라이선스		0	0	0	0	0	0		
	사용료		12	14	13	21	0	0		
	구독료		47	46	44	64	33	36		
	광고료		12	0	6	7	0	7		

o 대표적인 비즈니스 유형

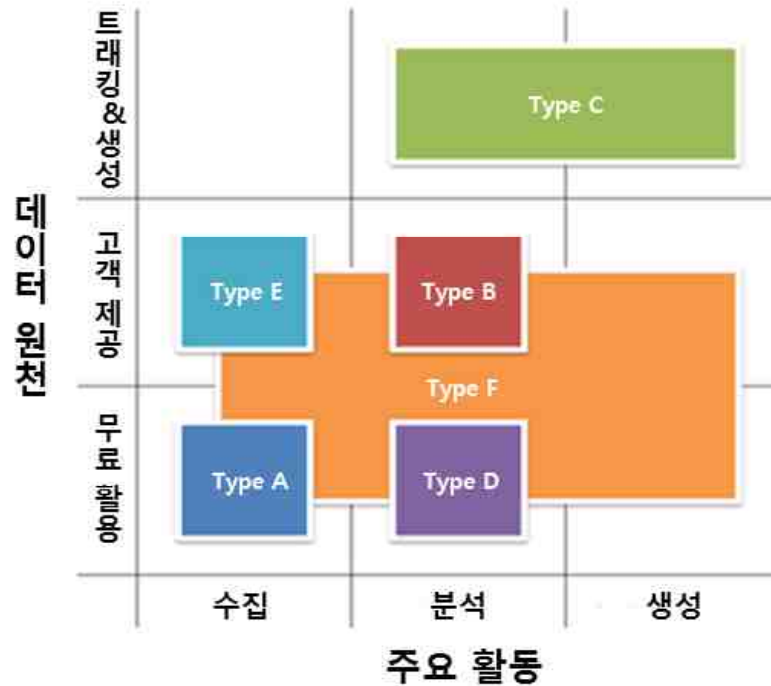


그림 3 캠브리지 DDBM의 대표적인 6가지 유형

- 하트만(Hartmann, 2014)은 100개 스타트업을 데이터 원천(Data Source : Y축)과 주요 활동(Key Activity : X축)으로 기준*으로 분석하여 6가지로 유형화
- * 클러스터링 기준 중 차별성을 줄 수 있는 기준인 '데이터 원천'과 '주요 활동'을 꼽고 나머지는 적합하지 않아 제외함
- 연구에 활용된 기업들 중 36%는 수익모델에 대한 정확한 정보를 얻을 수 없었으며, 83%의 회사가 구독료 또는 사용료 기반의 수익모델을 사용
- 데이터 원천에 따른 구분은 자체 생산이나 기존의 데이터 등을 활용하는 내부 데이터의 활용 또는 사용자 입력 데이터, 공공 데이터 등을 활용한 외부 데이터의 활용으로 분류
- 주요 활동에 따른 구분은 무수히 많은 컴퓨터에 분산 저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술인 크롤링(Crawling), 웹상에서 수집하는 크라우드소싱(Crowdsourcing), 데이터간 통합(Aggregation), 데이터간의 상관관계 등을 통한 분석 및 수집된 정보를 종합적으로 볼 수 있도록 제공하는 기술인 시각화(Visualization), 기술적 현황(descriptive) 분석, 예측(predictive) 분석, 예측최적화(prescriptive) 분석, 데이터를 나열하는 일반적인 방식인 배포 등으로 분류

② 국내외 비즈니스 유형 비교¹³⁾

o 국내외 데이터 기반 비즈니스 현황 비교

- 캠브리지 DDBM의 6가지 분류기준으로 스타트업 대상 분석
- 해외는 기존 캠브리지 DDBM 연구 당시 사용한 Angel List에서 추출한 100개 기업으로 분류했고, 국내는 2014 Startup Top 100에 선정된 기업 대상으로 유형화 하여 비교

* 유형이 명확하게 분류되지 않거나, 데이터 관련성이 낮은 기업 일부 제외

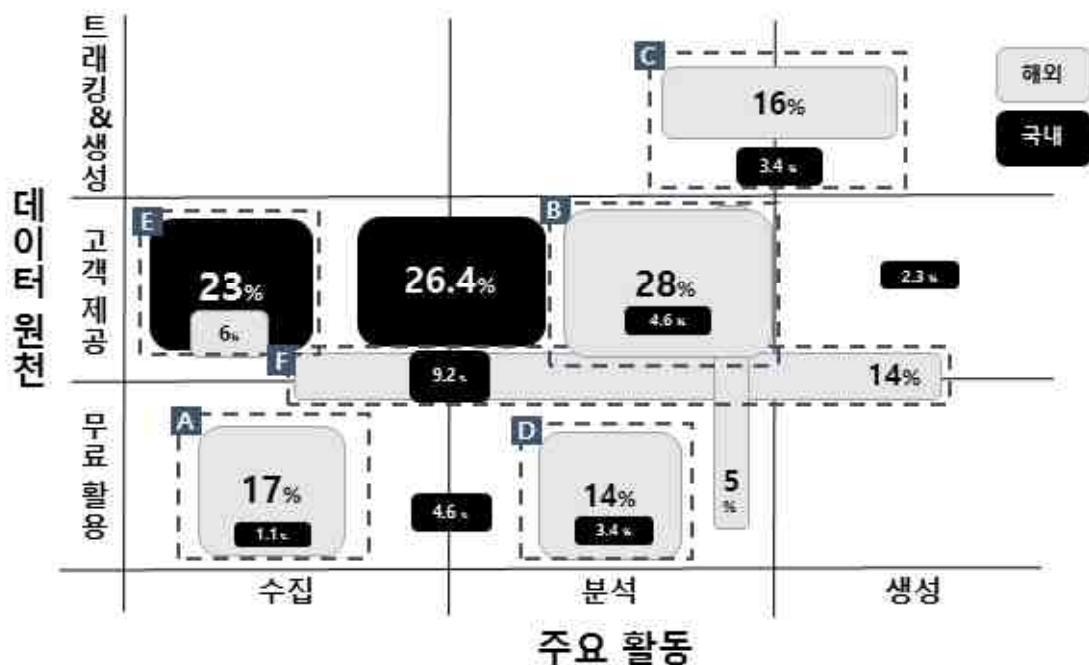


그림 4 국내외 데이터 기반 비즈니스 현황 비교

- 국내는 해외에 비하여 데이터 수집 영역에 많은 비중을 차지하나, 해외는 분석 및 생성 활동 영역에 많은 비중 차지함
- 오픈 데이터와 같은 무료 활용 데이터 및 트래킹을 통한 자체 생성 데이터 활용이 해외에서 가장 높게 나타나는 것에 비해, 국내는 고객 제공 데이터를 분석하여 결과를 제공하는 비즈니스 유형에 치중되어 있음
- 국내외 데이터 기반 비즈니스 유형 차이의 발생 원인을 파악하기 위하여 각 유형에 대한 프레임워크 요소별 차이점 분석

㉠ 유형: 무료 데이터 수집 및 통합 유형(Free data collector and aggregator)

- 특징 : 공개된 무료 API를 활용하여, 고객이 원하는 방식으로 제공하는 형식으로 실시간으로 API 분석을 통한 트렌드 분석 및 조사, 소셜 네트워크의 콘텐츠를 추출하여 사용자가 재가공할 수 있도록 원소스 제공
- 수입원 : 데이터 이용량에 따른 이용료 및 등록비
- 비중 : 국내 1.1%, 해외 17%
- 대표적 기업
 - 국내 : 말랑스튜디오, 쉐어앤케어, 썬크폴, 웹스, 텐커뮤니티
 - 해외 : 그립(Gnip), 데이터시프트(Datashift), 아부시(AVUXI), 에니그마(Enigma), 헬씨큐(healthiQ)

o 국내외 비즈니스 사례

- 국내 사례¹⁴⁾

표 7 A유형 국내사례: 말랑스튜디오

구분	내용
	· 회사이름 : 말랑스튜디오(malangstudio)
	· 설립년도 : 2013년 · 본사위치 : 서울시 강남구 삼성동 · 홈페이지 : www.malangstudio.com
기업소개	· 모바일 콘텐츠의 새로운 방식인 캐릭터 라이선스 기반 비즈니스 · 라이프 스타일의 변화시키는 서비스를 개발하고 전 세계에 퍼블리싱하는 앱 개발&퍼블리싱 회사
서비스	· SNS데이터를 수집하여, 취미 및 소모임 등 위치기반 모임을 개설 가능하게 하는 1Km 어플 서비스
비즈니스 성과	· 앱스토어 무료 앱 전체 2위 및 309만명 다운로드 · 2014년 옐로모바일로부터 20억원 투자 유치 · 중국 'MARS IT 인터넷창업대회' 1위 수상

- 해외 사례

표 8 A유형 해외 사례1: 그립

구분	내용
	· 회사이름 : 그립(Gnip)
	· 설립년도 : 2008년 · 본사위치 : 미국, 콜로라도 · 홈페이지 : https://gnip.com/
기업소개	· 소셜 미디어 API 통합 제공 기업
서비스	· 트위터 데이터 Reseller · 하나의 API로 여러 소셜 미디어 데이터에 쉽고 편리하게 접근할 수 있도록 서비스
특징	· 단일 API를 통해 여러 개의 소셜 미디어 데이터 이용 · ‘소셜웹의 중앙역’으로 불림
비즈니스 성과	· 2010년 트위터와 데이터 라이선싱 협정 · 2014년 트위터에 매각, 현재 서비스를 유지 · 매달 1,000억 회의 실시간 소셜 데이터 활동을 제공 · 2.3조 트위터 데이터를 42개국의 고객에게 제공

표 9 A유형 해외 사례2: 데이터시프트

구분	내용
	· 회사이름 : 데이터시프트(DataShift)
	· 설립년도 : 2007년 · 본사위치 : 미국 샌프란시스코 · 홈페이지 : http://datasift.com/
기업소개	· 수십 개의 소셜 미디어 데이터들로부터 의미 있는 정보를 추출해 수요자의 목적에 맞게 분석 · 트위터 이외에 페이스북, 구글 플러스, 유튜브, 인스타그램, 링크드인, 중국의 최대 SNS인 웨이보 등
서비스	· SNS나 웹 등 공개된 온라인상의 데이터를 고객의 목적이나 선호도에 따라 취합하고 정리, 분석해서 제공하는 플랫폼
특징	· 소셜 미디어 데이터와 이를 이용하려는 이용자를 이어주는 데이터 큐레이션 플랫폼 제공하는 중개형(Agent) 형태 · 자신이 데이터를 보유하지 않고 그 대신 고객들의 외부 소셜 미디어 빅데이터를 쉽게 분석하고 여러 필터를 적용해 원하는 정보만 정제해서 볼 수 있도록 지원
비즈니스 성과	· 소셜미디어의 업데이트 내용에서 키워드를 분석하는 기본 서비스의 한 달 사용료는 3000달러 · 검색하는 데이터 소스에 따라 요금 부과

o 주요 시사점

- 해외 : 오픈 데이터, 소셜 미디어 데이터, 웹 크롤링 데이터 등 무료로 활용 가능한 데이터의 이용률이 높고, 데이터 수집을 통한 처리 영역이 활발함. 대부분 B2B 형태
- 국내 : 상대적으로 무료 활용 데이터의 이용이 낮으며, 원천 데이터를 온라인 화면에 보여주거나 또는 API형식으로 데이터를 사용자에게 제공하는 배포 형태 서비스에 치중됨. 대부분 B2C 형태
- 평가
 - 개인정보로 분류된 소셜 미디어 데이터 활용을 통한 무료데이터 이용의 활성화를 위하여 사전형식 규제 완화가 필요
 - 수집한 데이터의 정제 및 재판매 등 다양한 비즈니스 모델 수립을 위한 지원 체계 강화
 - 데이터를 공유하고 매매 하는 중개자 역할의 비즈니스가 활성화 될 수 있도록 유통에 대한 법적 문제 및 제약사항 완화 필요

㉠ 유형: 분석 서비스 유형(Analytics-as-a-service)

- 특징 : 사용자가 생성한 정보를 통해서 사용자에게 기존 데이터를 기반으로 분석 정보(마케팅, 고객관리, 판매, 침해 감시 등)를 제공하는 형태로 대부분 자체 분석 솔루션으로 분석 서비스 제공
- 수입원 : 등록비 또는 서비스 이용료
- 비중 : 국내 4.6%, 해외 28%
- 대표적 기업
 - 국내 : 노리, 도너스, 마카롱, 인콘텐츠(inContents), 플라이북
 - 해외 : 그래너파이(Granify), 데이터쿱(Datacoup), 듀딜(DueDil), 샌디파이(Sendify), 코어로직(CoreLogic)

o 국내외 비즈니스 사례

- 국내 사례

표 10 B유형 국내 사례: 노리

구분	내용
	· 회사이름 : 노리(knowre)
	· 설립년도 : 2012년 · 본사위치 : 서울 서초구 강남대로 · 홈페이지 : http://www.knowre.com/
기업소개	· 맞춤형 수학교육 플랫폼을 만들고 있는 EdTech 벤처
서비스	· 풍부한 현장 교사경험과 최고 수준의 개발력, 사용자 경험에 대한 통찰력을 바탕으로 디지털 디바이스의 장점을 최대한 끌어낼 수 있는 교육 서비스
비즈니스 성과	· Global K-Startup 2012 대상 수상 · 뉴욕 250 여개 학교에 공급 기회 확보 · Global Startup Showdown 2013 수상

- 해외 사례

표 11 B유형 해외 사례: 그래너파이

구분	내용
	· 회사이름 : 그래너파이(Granify)
	· 설립년도 : 2011년 · 본사위치 : 캐나다, 알버타 · 홈페이지 : http://www.granify.com/
기업소개	· 빅데이터 및 머신러닝을 사용하여 온라인 소매업체의 매출을 극대화 시킬 수 있는 SaaS 솔루션 제공
서비스	· 고객들이 지속적으로 사이트에 다시 유입될 수 있도록 적합한 메시지 주입 · 실시간 빅데이터 활용을 통해 자동으로 구매 확률을 높일 만한 메시지나 할인 코드 전송
특징	· 자체적인 데이터 분석에 따라 온라인 사이트를 방문한 고객들이 구매 이전에 사이트를 빠져나가는 성향을 95%의 정확도로 판단 · 쿠폰 메시지가 팝업창으로 나타나 반품 정책, 무료 배송 같은 메시지 등을 보여주어 사이트 이탈을 방지에 도움
비즈니스 성과	· 기존의 온라인 매장들과 협업하여 한 달에 53억 개 이상의 데이터 처리 · 2013년에는 150만 달러(한화 약 1500만원)의 투자 유치

o 주요 시사점

- 해외 : 해외 비즈니스 모델 중에서 가장 큰 비중을 차지하는 유형이며 주요 데이터 원천으로 무료 데이터 및 고객 제공 데이터를 사용하여 분석함. 분석 기법 중 현황 분석 활동이 가장 활발하며, 예측 분석 및 규범 분석 활동은 낮은 편임
- 국내 : 주요활동으로 수집에 비하여 분석영역은 낮은 수준임
- 평가
 - 해외의 경우 데이터를 기반으로 분석 정보를 제공하는 서비스 비중이 가장 높은데, 이는 해외의 데이터 수집 영역은 이미 성숙 단계에 접어든 것으로 국내에서 분석 영역이 발전하려면 우선적으로 수집 기반 강화 필요

㉞ 유형: 데이터 생성 및 분석 유형(Data generation and analysis)

- 특징 : 데이터를 수집하지 않고 스스로 생산하고 분석하여 결과 제공. 클라우드소싱, 웹 분석, 스마트폰이나 다른 센서 등을 통해 데이터를 생산
- 수입원 : 데이터 수집을 위해 기기를 별도로 판매하기도 함
- 비중 : 국내 3.4%, 해외 16%
- 대표적 기업
 - 국내 : 노매드커넥션, 룩랫(loplat), 에듀스파, 젠포트(Gen Port), 캄스콘(CAMSCON)
 - 해외 : 믹스패널(mixpanel), 스웜니(Swarmly), 스피나커(Spinnaker), 오토매틱(Automatic), 킨사(Kinsa)

o 국내외 비즈니스 사례

- 국내 사례

표 12 C유형 국내사례: 노매드커넥션

구분	내용
	· 회사이름 : 노매드커넥션(Nomad Connection, Inc.)
	· 설립년도 : 2008년 · 본사위치 : 서울 영등포구 · 홈페이지 : http://www.nomadconnection.com/

기업소개	· 기술혁신형(INNO-BIZ) 기업 · 미디어 & 네트워크 플랫폼 관련 솔루션과 서비스를 개발하는 소프트웨어 기술 중심의 회사 · P2P 비디오 스트리밍 및 핀테크 플랫폼을 개발
서비스	· 금융 큐레이션 실시간 빅데이터 수집으로 음악, 영화와 같은 콘텐츠를 언제 어디서나 자신의 휴대폰, 태블릿, 노트북, PC등에서 시청할 수 있도록 서비스
비즈니스 성과	· 짐리(Zimly)라는 안드로이드 미디어 플레이어는 250만 다운로드 · 전체 이용자의 절반 이상이 해외 사용자며 전 세계 세 번째 안에 드는 플레이어 · 2011년 CJ Hellovision TVing Air Media Platform 공급 계약

- 해외 사례

표 13 C유형 해외사례1: 스왐니

구분	내용
	· 회사이름 : 스왐니(Swarmly) · 설립년도 : 2012년 · 본사위치 : 영국, 스코틀랜드 · 홈페이지 : http://swarmly.co/
서비스	· 술집, 음식점이나 클럽 등 가장 인기 있는 장소를 실시간으로 검색 하는 지도 서비스
특징	· 한 지역사회 내에서 핫 플레이스를 쉽게 파악하는데 도움을 주며 개인뿐 아니라 기업체 입장에서 지역 내에서 게릴라 이벤트를 펼치거나 장기적인 마케팅 전략 수립에 도움 · 사용자들은 swam을 태그할 수 있고 새로운 소셜 데이터를 큐레이션의 형태로 더할 수도 있음

표 14 C유형 해외사례2: 믹스패널

구분	내용
	· 회사이름 : 믹스패널(mixpanel) · 설립년도 : 2009년 · 본사위치 : 캘리포니아, 샌프란시스코 · 홈페이지 : www.mixpanel.com
서비스	· 사이트 분석 서비스 · 사용자의 웹사이트에 믹스패널 코드를 심어서 사용자의 정보를 수집하여 자체 분석된 정보를 사용자에게 제공하는 서비스
특징	· 웹 및 모바일 응용 프로그램과의 사용자 상호 작용을 추적 · 대상 응용 프로그램과의 통신을 위한 도구 제공

o 주요 시사점

- 해외 : 데이터를 수집하지 않고 사물인터넷 및 웹 트래킹, 클라우드소싱을 통하여 자체적으로 생성 및 분석하는데 강점을 보이며 분석 영역 중 규범 분석이 가장 많은 부분을 차지하고 있음. 이는 머신 러닝 기법 등 선진화된 분석 기법을 활용하여 예측 모델을 개선하고 예측력을 높이는 방향으로 기술이 앞서 가고 있음을 나타냄. 대부분이 구독료, 사용료 기반이지만, 자산을 판매하거나 데이터 생성 기기 판매로 수익을 창출함
- 국내 : 클라우드소싱이나 사물인터넷, 웹분석 등을 통하여 데이터를 자체 생산하는 활동이 저조함
- 평가
 - 해외의 경우 데이터 생성을 위한 다양한 사물인터넷기기를 직접 제작, 판매하고 있으나, 국내의 경우 데이터 수집을 위한 기기의 생산에 있어 해외에 비하여 기술력이 부족하며 이를 발전시키기 위하여 정부 주도하의 민간 데이터 수집 활성화 지원이 필요

㉔ 유형: 무료 데이터 정보 디스커버리 유형(Free data knowledge discovery)

- 특징 : 무료로 이용 가능한 데이터를 수집, 분석하여 결과제공. 기계가독형 포맷의 무료 데이터뿐만 아니라 웹에서 데이터를 크롤링 함
- 수입원 : 사용료 수입과 함께 광고료, 중개수수료
- 비중 : 국내 3.4%, 해외 14%
- 대표적 기업
 - 국내 : 데일리, 레드테이블(REDTABLE), 세이바이, 유학닷컴, 인포케어옥션
 - 해외 : 길드(guild), 오렐리(Olery), 딜엔젤(DealAngel), 인사이트풀(Insightpool), 트렌드스팟(TrendSpottr)

o 국내외 비즈니스 사례

- 국내 사례

표 15 D유형 국내사례: 데일리

구분	내용
	· 회사이름 : 데일리(daily)
	· 설립년도 : 2013년 07월 · 본사위치 : 서울시 강남구 테헤란로 · 홈페이지 : http://dailyhotel.co.kr/
기업소개	· ‘당일호텔예약’서비스를 제공하는 회사 · 월 300%의 성장률을 보이며 재 구매율이 5% 이상임
서비스	· 데일리호텔 어플은 저렴한 가격 혹은 급하게 당일 호텔 예약이 필요한 소비자 그리고 오늘 남은 객실을 판매하고자하는 호텔의 수요를 충족시키는 서비스
비즈니스 성과	· 프라이머, 본엔젤스, 세쿼이어캐피탈 등 투자유치 받음

- 해외 사례

표 16 D유형 해외사례: 길드

구분	내용
	· 회사이름 : 길드(gild)
	· 설립년도 : 2012년 · 본사위치 : 샌 프란시스코 · 홈페이지 : http://www.gild.com
서비스	· 기업들의 개발자 채용 지원. · 프로그래머의 능력을 구분하기 위해 GitHub나 구글 코드 등 오픈 소스 사이트에 공개된 코드, Stack Overflow와 같은 Q&A 웹 사이트의 기여도를 자동으로 평가 · 평가 결과에 근거하여 점수를 매긴 후 개발자의 실력을 평가한 정보를 제공하는 서비스

o 주요 시사점

- 해외 : 비즈니스의 대상고객은 대부분이 B2B이며, 완전한 구독료 및 사용료 기반으로 수익을 창출
- 국내 : B2B 및 B2C의 유형이며, 등록된 재산권에 대한 대가를 받는 라이선싱 비중이 큼. 웹 데이터를 탐색·수집(웹 크롤링 등을 활용) 가능한 데이터를 분석하여, 결과를 제공 하는 비즈니스 유형이 해외와 비교하여 매우 낮은 수준에 머무르고 있음

- 평가

- 국내의 경우 범국가 차원에서 공공데이터 오픈을 통해 무료 데이터 제공을 위해 노력을 하고 있으나, 가공되지 않은 Raw 데이터 수준으로 오픈되어 그 활용도가 해외에 비하여 상대적으로 낮음. 오픈 데이터 개방에 따른 활용성 강화를 위한 정부 차원의 관리 및 지원 방안 마련
- 웹상 데이터 탐색·수집 활동의 경우 법적 제약과 국내 데이터 시장의 비활성화로 인하여 해외에 비하여 상당히 낮은 수준에 머무르고 있음. 이에 활성화를 위한 규제완화 필요

웹 크롤링 분쟁 사례15)

사람인HR이 2008년 이전부터 수차례의 합의와 법원 조정 및 판결을 일방적으로 무시하고 지속적으로 잡코리아 채용공고를 무단으로 복제해 이를 이용, 부정 경쟁 행위를 해왔다고 판결

재판부는 사람의 크롤링 행위를 부정경쟁행위로 판단했다. 부정경쟁방지 및 영업비밀보호에 관한 법률 제2조 1호 차목이 규정하고 있는 '타인의 상당한 투자나 노력으로 만들어진 성과 등을 공정한 상거래 관행이나 경쟁질서에 반하는 방법으로 자신의 영업을 위해 무단으로 사용함으로써 타인의 경제적 이익을 침해하는 행위'에 해당된다는 것임

잡코리아에 따르면 법원을 통해 이미 2010년 가처분결정, 2011년 강제조정결정, 2012년 집행문부여판결, 2013년 1월 조정이 성립됐음에도 불구하고, 사람인HR의 불법행위는 계속돼 왔다. 소송이 진행되는 와중에도 무단복제 행위를 멈추지 않아, 최근까지 이와 같은 문제가 지속됐으며, 사람인HR의 무단 채용공고 퍼가기 영업으로 인해 피해를 입는 구직자와 기업이 발생한다는 점을 강조함

㉔ 유형: 데이터 통합 서비스 유형(Data-aggregation-as-a-service)

- 특징 : 데이터를 분석하거나 데이터를 생산하지 않고, 기업 내의 다양한 데이터를 수집하여 웹 등 다양한 경로로 고객에게 서비스함 개인 데이터를 포함한 다양한 종류의 고객 데이터를 수집 또는 특정 영역이나 특정 문제 처리함
- 수입원 : 가입비, 사업자가 주요 고객
- 비중 : 국내 23%, 해외 6%
- 대표적 기업

- 국내 : 나인플라바, 세렌트립, 세틀뱅크, 올하우, 쿠스토
- 해외 : 블루노즈(Bluenose), 알마(ALMA), 액시엄(Acxiom), 올웨이즈프리퍼드(AlwaysPrepped), 후(Who@)

o 국내외 비즈니스 사례

- 국내 사례

표 17 E유형 국내사례: 나인플라바

구분	내용
	· 회사이름 : 나인플라바(9flava)
	· 설립년도 : 2011년 · 본사위치 : 서울 서초구 서초동 · 홈페이지 : http://wepass.co.kr/
서비스	· 위패스는 뷰티, 패션, 컬처, 외식, 쇼핑 등 다양한 라이프 스타일 브랜드의 마일리지를 하나의 앱으로 적립할 수 있는 모바일 마일리지 카드 서비스
비즈니스 성과	· 2012년 본엔젤스 3억원 투자유치 · 다음커뮤니케이션 10억원 투자유치

- 해외 사례

표 18 E유형 해외사례: 액시엄

구분	내용
	· 회사이름 : 액시엄(Acxiom)
	· 설립년도 : 1969년 · 본사위치 : 미국, 아칸소 · 홈페이지 : http://www.acxiom.com/
기업소개	· 인구통계학 서비스 기업으로 미국에서 설립 · 현재 유럽, 아시아 및 남미 등에 지사를 두고 마케팅 및 정보 관리 서비스를 제공
서비스	· 온라인에 게재된 모든 고객 관련 데이터를 수집하고 분석하여 기업들에게 판매 · 차별화된 방법으로 고객 데이터를 수집, 창출하여 가치를 제공하는 서비스
특징	· 수집한 데이터(공공/정보/고객동의데이터)를 업무에 활용하는 것이 아니라 이를 필요로 하는 기업들에 판매하는 것을 주 사업으로 하는 데이터브로커(Data Broker) 형태
비즈니스 성과	· 전 세계적으로 약 5억 명의 개인정보를 보유 · 연간 50조 거래내역을 분석 · 다이렉트 마케팅 서비스 부문 110억 달러 시장에서 12%인 11억 5천만 달러의 매출을 올리고 있음

o 주요 시사점

- 해외 : 기업 내의 다양한 데이터를 수집하여 데이터 변형 및 데이터 정제를 제공해주는 서비스에 많은 비중을 차지
- 국내 : 국내 비즈니스 모델 중 가장 큰 비중을 차지하는 유형
- 평가
 - 해외는 고객데이터를 수집하여 업무에 활용하는데서 더 나아가 판매 까지 하는 데이터브로커 시장이 활성화되어 있으며, 국내 또한 비식별 화법이 통과됨에 따라 활성화 될 것이라 예상되나 가속화시키기 위 하여 편리하고 빠르게 공유 및 거래 할 수 있는 체계 마련 필요

㉔ 유형: 멀티 소스 데이터 통합·분석 유형(Multi-source data mash-up and analysis)

- 특징 : 사용자가 생산한 데이터베이스와 외부의 다양한 무료 데이터를 통합하여 분석 정보를 제공하는 서비스
- 수입원 : 대부분 대상 고객은 기업이며, 등록비 수익이 많음
- 비중 : 국내 9.2%, 해외 14%
- 대표적 기업
 - 국내 : 굿닥(Gooddoc), 비엔투스, 유웨이 중앙교육, 지디에스케이, 파킹스퀘어
 - 해외 : 넥스트빅사운드(Next Big Sound), 위러브로이(welovroi), 팜 로그(FarmLogs), 팩츄얼(Factual), 핀지니어스(FinGenius)

o 국내외 비즈니스 사례

- 국내 사례

표 19 F유형 국내사례: 파킹스퀘어

구분	내용
	· 회사이름 : 파킹스퀘어(Parking Square Inc.)
	· 설립년도 : 2013년 · 본사위치 : 서울시 강남구 테헤란로 · 홈페이지 : https://www.parkhere.co.kr/

서비스	· ‘파크히어’는 정부/지자체 개방 정보로부터 사용자 제공 정보, 직접 수집한 고유 정보까지 방대한 주차장 정보 제공 · 독자적인 시스템으로 고객이 원하는 주차장 자리를 미리 확보하며, 수요/공급에 따라 경쟁력 있는 가격을 제공
비즈니스 성과	· 15억원 투자유치(스마일게이트인베스트먼트, 미래에셋캐피탈, KTB네트워크) · 미래부 선정 ICT분야 유망 기업 ‘K-Global 300’ 선정 · ‘주차 정보 제공 방법 및 장치’에 관한 특허 등록

- 해외 사례

표 20 F유형 해외사례: 팩추얼

구분	내용
	· 회사이름 : 팩추얼(Factual)
	· 설립년도 : 2007년 · 본사위치 : 로스앤젤레스 · 홈페이지 : https://www.factual.com/
기업소개	· 개방형 데이터 플랫폼 운영사로 설립
서비스	· 비즈니스 리스트와 관심지역에 대한 정보를수록 · 누구나 접근 가능하도록 개방 데이터를 공유하고 매시업 할 수 있는 플랫폼을 제공
특징	· 데이터 개선에 기여한 고객에게 할인 정책을 펴 클라우드소싱을 통한 데이터 생성 및 큐레이션의 협력을 유도
비즈니스 성과	· 여행 사이트 TripAdvisor와 제휴하여 전 세계적으로 유명한 식당 정보 등을 공유하기로 협약

o 주요 시사점

- 해외 : 기업 내의 다양한 데이터를 수집하여 데이터 변형 및 데이터 정제를 제공해주는 서비스에 많은 비중을 차지. 주요 데이터 원천이 오픈데이터 및 고객 제공 데이터에 가장 많은 부분을 차지하고, 대부분이 구독료 및 사용료 기반이며, 대부분 B2B형태
- 국내 : 외부의 고객 제공이 주요 데이터 원천이며, 주요 활동으로 처리(Process) 영역이 가장 높은 비율을 차지. 광고를 통한 수익 창출에 대한 비중이 큼

- 평가

- 해외의 경우 고객 보유 데이터와 무료 데이터의 수집 뿐만 아니라 분석을 통한 인사이트 도출을 통해 추가 데이터를 생성하는 데이터 활동의 전반적인 영역을 아우르고 있으나, 국내의 경우 데이터 수집 및 분석의 영역에만 치중해 있으며 데이터 생성 영역은 미비함. 생성기반 강화를 위한 민간 기업 육성 방안 마련 필요

□ 유형별 주요 시사점 종합

- 첫째, 국내외 현황비교를 통해 나타난 국내에 미비한 다양한 데이터 기반 비즈니스 모델 활성화를 위한 민간기업 육성방안 마련 필요
=> Cambridge의 DDBM의 6가지 비즈니스 유형에 따라 국내 데이터 기반 비즈니스 현황을 분석한 결과 다양한 측면에서의 데이터 기반 비즈니스 모델 활성화를 위한 민간 기업 육성을 통한 데이터산업 발전 방안 제시
- 둘째, 데이터 유통 · 연계 · 활용은 공공과 민간의 데이터산업을 활성화 하며 데이터 자산은 국가 경쟁력을 좌우하는 요소로 자리 잡게 되었음. 이에 공공 · 민간데이터 활용 촉진을 위한 데이터 유통 체계 수립 필요
=> 공공 · 민간데이터 연계, 정제, 가격산정 등 다른 기관과 편리하게 빠르고 공유 및 거래할 수 있는 체계 마련
- 셋째, 데이터산업 비즈니스 활성화 및 공공 · 민간 데이터 활용을 촉진하기 위해서는 데이터 수집 · 활용을 위한 사전형식에 대한 규제에 대한 완화 및 재정립 필요

4. 개인 데이터

□ 개념

- o ‘개인 데이터’는 개인과 관련한 데이터를 의미
 - OECD 및 세계경제포럼(WEF)에서는 개인 데이터를 개인과 관련한 모든 데이터로 정의. 특히, 세계경제포럼에서는 개인 데이터를 자발적 데이터, 관찰된 데이터, 추론된 데이터로 3가지 유형으로 구분

표 21 ‘개인 데이터’ 정의

구분	주요 ‘개인 데이터’ 정의
세계경제포럼 (WEF)	2010년 6월 WEF에 의하면, 개인 데이터는 사람에 의해, 사람에 관한 디지털 데이터로 3가지 유형을 포함 - 자발적 데이터 개인에 의해 생성되고 개인의 명시 하에 공유한 데이터 (예 : 소셜 네트워크 프로필) - 관찰된 데이터 개인의 활동에 의해 수집된 데이터 (예 : 핸드폰 사용 시 관측된 개인의 위치 정보) - 추론된 데이터 자발적 개인 데이터 또는 관측된 데이터의 분석을 통해 추측한 개인에 관한 데이터 (예 : 개인 신용평가 점수)
OECD	2013년의 OECD ‘Protection of privacy 및 Transborder Flows of Personal Data’ 가이드라인에 따르면, 개인 데이터는 식별된, 또는 식별 가능한 개인에 관련한 모든 정보* * 개인데이터 유형 - 사용자가 제공한 콘텐츠(블로그, 의견, 사진, 비디오) - 활동 데이터(예: 인터넷으로 검색하는 기록, 온라인으로 구매하는 상품, 지불하는 금액 및 방법) - 소셜 데이터(소셜 네트워킹 사이트의 연락처 및 친구기록) - 위치 데이터(거주지 주소, GPS 및 핸드폰의 지오로케이션, IP 주소) - 인구통계학적 데이터(나이, 성, 인종, 수입, 성적 취향, 정치적 성향) - 공인된 데이터(이름, 금융정보, 계좌번호, 건강정보, 국가건강 또는 사회보장번호, 전과기록)

- 2013년 세계경제포럼에 의해 새로운 가치창출의 수단으로 주목받고 있음
 - 개인 생활 편의성 향상
 - * 진료, 취업, 대출 등에서 데이터를 요구 받을 때 일괄 제출 가능
 - * 건강 진단, 개인 재무관리 등을 효과적으로 수행 가능

- 신 서비스 산업 창출 : 개인에게 제공되는 데이터를 활용하는 서비스 기업 창출
예) 개인 건강가이드, 개인 재무관리 가이드 등
- 서비스 기업 효율화 : 개인 데이터를 통합하여 활용할 수 있으므로 서비스 품질 제고
예) 의료기관은 개인의 질병 진단 등에 과거 이력 통합 활용. 보험회사는 보험료 산정에 개인 과거 데이터 활용

□ 주요 현안

- o 국내는 엄격한 개인정보보호법과 사회 문화적 영향으로 개인 데이터를 활용하는 체계 도입이 해외 대비하여 거의 미비함
 - * 예: 은행 등 금융회사에서는 개인이 원할 경우 엑셀 등의 파일 형태로 다운로드 받을 수 있는 서비스를 제공하지만, 금융회사 내부 방침에 의해 제한된 데이터 형식임

□ 현황 분석

- o 개인 데이터 활용 해외 정책 추진현황
 - 개인 데이터를 통한 가치 창출에 대해 해외 각국에서 다양한 노력 진행 중
 - 영국, 프랑스, 미국, 핀란드 등 주요 국가에서는 2010년 초기부터 개인 데이터의 활용 가치를 인식하고 새로운 가치 창출의 한 수단으로 개인 데이터 활용과 관련한 ‘Mydata’ 정책을 추진 중

표 22 각 국의 개인 데이터 활용 정책

구분	영국	프랑스	미국	핀란드 ¹⁶⁾
추진 년도	2011	2012	2012	2014
추진 정책	Better Choices, Better Deals 전략 중, midata 프로그램 추진	MesInfo	Smart Disclosure	Mydata
관련 주요 내용	소비자의 이용내역 데이터 활용을 목적	개인 데이터 사용 편의의 개인·조직·사회 간 분배	소비자가 현명한 의사결정을 내릴 수 있도록 적시에 필요한 데이터를 제공	MyData를 데이터 경제 실현을 위한 핵심 인프라로 접근

11 영국의 ‘midata’ 정책

o 추진 목적17)

- 소비자가 기업이 보유하고 있는 개인 이력(transaction) 정보를 기계 가독형의 재활용이 가능한 포맷으로 접근할 수 있도록 하는 것
- (소비자 측면) 안전하게 소비자가 자신들의 데이터에 접근 및 사용
- (기업 측면) 소비자가 데이터를 해석 및 사용할 수 있도록 기업의 혁신적 서비스와 어플리케이션 개발 보조
- (경제적 측면) 요청에 따른 개별 소비자의 개인 데이터를 공개하는데 동의한 기업들을 보호

o 추진 주체

- 정부 주도 하에 운영 및 활동위원회로 구성되어 있으며 나이젤 셰드볼트(Nigel Shadbolt) 교수가 운영위원회장으로 진행
- 2011년 4월 이래로 정부, 기업, 소비자, 규제기관, 무역기관에서 자발적 파트너십 아래 운영

o 추진 경과

- 2011년 11월, 국민 대상 ‘midata’ 호응도 및 인식 파악을 위한 설문 조사 진행

국민 대상 설문조사
o midata 전략위원회 주도로 실시, 영국 전역에서 1,002명의 국민 참여 o 설문조사 실시 전, 부문별 활용 시나리오(10개 사례)를 제공하여 일반 국민의 정책 이해도 및 수혜 효과 인식 제고 o 성별, 연령대, 직업 상태, 결혼 여부, 주택 현황, 가계 소득, 인터넷 이용태도 등 표본 통계 구성을 위한 기초 인적사항 조사 실시(12문항) o 구체적인 midata 서비스 모델을 제시하여 선호도 및 예상 문제점 조사(15문항) ① 은행 계좌, 신용카드, 포인트 카드 및 지출·수입 등을 일괄적으로 관리하여 재정 현황을 분석하고 계획을 수립할 수 있도록 도와주는 재정관리 서비스 ② 서비스 이용 내역 등 서비스 제공자(기업)가 보유한 본인의 정보에 쉽게 접근하여, 효율적 의사결정 및 기업이 보유한 정보의 사실 여부 확인에 활용 ③ 신용카드 사용·온라인 구매 시 구매 금액이 큰 항목에 관해 전자 보증서와 영수증을 생성하여 일괄적으로 관리하며, 보증 현황(유효기간 등) 및 A/S 관련 정보 확인 가능 ④ 본인의 건강 정보를 일괄적으로 저장·관리하여 의사, 보호자 등 본인이 선택한 공유 대상과 공유할 수 있으며, 건강 관리를 위한 기타 사항(혈압, 몸무게, 운동량, 혈당수치 등) 기록 ⑤ 서비스 공급자(기업)로부터 제공받은 에너지, 통신 사용 현황 데이터를 기반으로 한 맞춤형 전문가 상담을 제공, 기업간 비교 및 상황에 맞는 최적의 선택에 도움 ⑥ 동의 기반 아이튠즈, 아마존, 티켓 예매 사이트의 이용·구매 내역을 분석하여, 본인의 관심사에서 발생하는 공연 정보 등을 적시에 추천 ⑦ 개인정보 저장소(Personal Data Store)를 활용하여 본인의 개인정보를 일괄적으로 관리(예: 이사를 할 경우 개인정보 저장소에서 주소를 변경하고 허용 범위 내의 관계자에게 알림으로써 정보 갱신)

- 설문 조사를 통해 국민 호응도 파악
 - 조사 결과 : 높은 득점 분야 - 에너지, 휴대전화 분야
 - 낮은 득점 분야 - 나의 이벤트 관리(My Events) 서비스, 금전관리(My Money) 서비스
 - 동의 없이 데이터를 사용하지 않는다는 확신이 있는 경우, midata 활용 서비스 이용에 상대적으로 긍정적임
 - 동의 없는 데이터 공유 및 신분 도용에 관한 우려가 많음
- 2012년 9월, midata 법제화를 위해 기업·기관 대상 기초조사 실시

기업·기관 대상 기초조사
- o midata 제공 의무화, 적용 범위, 전담기관 등에 대한 조사 진행(주관식 22문항) - o 대·중소기업, 소비자 단체, 규제 기관, 협회 등 각 기관 담당자 대상 - o 우편, 팩스, 이메일, 온라인 등을 통해 결과 접수, 총 433명 응답

- midata 취지에 대략 50%의 기업 및 기관이 동의했으며, 자율적 추진 시 전체 기업의 74%가 동의
- 2013년 4월, ‘개인정보 개방·활용의 가치 탐구’ 를 목표로 midata 해커톤 개최
- 2013년 4월, midata 자발적 참여 프로그램 1년 후, 보다 효율적인 정책 추진 및 성과 향상을 위해 기업규제개혁법(Enterprise and Regulatory Reform Act) 개정

기업규제개혁법(Enterprise and Regulatory Reform Act)
- o 제89조 소비자 데이터 제공 관련 - 서비스 이용자의 데이터 제공 요청이 있을 경우, 서비스 제공자가 전자형태의 서비스 이용내역 데이터를 요청한 서비스 이용자에게 제공하도록 조치 가능. 이때, 서비스 제공자는 에너지(가스 및 전기), 휴대전화, 금융(계좌 및 신용카드) 서비스 및 그 외 재화나 서비스를 제공하는 자를 일컬음 - 서비스 제공자는 midata 제공시 비용을 책정 가능하며 금액은 midata 제공을 위해 기업이 부담하는 비용을 초과해서는 안됨 - 총리는 제공되는 midata 내용을 결정함에 있어, 포함되는 기간, 빈도, 금액의 상세 내역, 소비자에게 미치는 효과 등을 고려해야 함 - o 제90조 소비자 데이터의 제공: 집행 관련 - 정보보호위원회(Information Commissioner's Office, ICO) 및 지정된 집행권자의 규정 집행을 위한 조항 신설 가능 - ICO 및 지정된 집행권자는 소비자 데이터의 부적절한 사용 및 유출이 있었는지 조사 가능 - ICO를 제외한 집행권자는 집행시 ICO에 보고 필요 - o 제91조 소비자 데이터의 제공: 부칙 관련 - 제89조 및 제90조에 의거 추가적인 조항 신설 가능

- 2015년 3월, 금융, 에너지, 보험, 차량 등 종합 가격비교 사이트인 GoCompare.com에서 midata를 활용한 분석 서비스 개시

o 활용 사례

- midata를 csv형식 파일로 제공함
 - 에너지 분야: E.on, M&S Energy(SSE), British Gas, EDF Energy, nPower
 - 금융 분야: Barclays, RBS, NatWest, Lloyds, HSBC, Santander, Tesco

FAIR DATA 원칙
- 2013년 1월, 영국내 The Market Research Society(MRS)에서 소비자들이 신뢰할 수 있도록 수립한 원칙으로, MRS에 가입되어 있는 60개국 이상의 회원 기업들이 준수하고 있음 - MyData 활용을 위해 영국에서 FAIR DATA 원칙 수립 - 10가지 Fair Data 원칙 ① 모든 개인 정보가 고객의 동의하에 수집되도록 보장한다. ② 고객의 동의된 데이터를 제공된 목적 이외의 목적으로 사용하지 않는다. ③ 고객이 보유하고 있는 개인 데이터에 액세스 할 수 있는지 확인하고 사용 방법을 알려준다. ④ 개인 데이터를 보호하고 기밀로 유지한다. ⑤ 개인 데이터가 존중받으며 취급되도록 보장한다. ⑥ 취약 계층과 미성년자의 데이터 수집은 사용하는 프로세스에 의해 적절히 보호되도록 할 것이다. ⑦ 데이터 공급사슬의 기업들에게 공급자들과 동일한 표준을 준수하도록 권장한다. ⑧ 개인 데이터의 윤리적인 모범 사례는 구매 프로세스를 통해 사용하는 것이다. ⑨ 우리는 개인 데이터에 접근 할 수 있는 모든 직원에게 사용법에 대한 적절한 교육을 실시한다. ⑩ Fair Data 원칙의 적용 여부가 불확실한 경우 개인 데이터를 사용하지 않는다.

출처: FAIR DATA 홈페이지, <http://www.fairdata.org.uk/10-principles/>

② 프랑스의 ‘MesInfos’ 정책

o 추진 목적

- 개인 데이터 사용을 통해 개인 · 조직 · 사회 간 편익 분배
- 기업은 개인 데이터를 활용하여 상당한 이익을 취하고 있는 반면, 개인은 이와 같은 이익을 받지 못함. 이에, 동일 플랫폼을 통해 데이터의 전송 · 저장 · 관리 · 서비스 연결이 가능한 기술생태계(디지털 환경, Self Data Ecosystem) 강조

o 추진 주체

- 민간 중심의 자발적 프로그램으로 프랑스 싱크탱크인 FING (Foundation Internet Nouvelle Generation)과 민간 비즈니스 클러스터인 캡 디지털(Cap Digital)의 공동 참여로 시작(2012)
- 정부는 제한적인 역할 수행(전문가 자문 지원)

o 추진 경과

- 소비 외 개인의 일상 전반에 보편적 관심을 두고 실제 데이터의 활용과 개인 참여에 동반한 실험적 탐구 강조 및 추진
- 2011년, 15개 기업 및 공공기관 참여 하에 실생활에 적용 가능한 MesInfos 준비 워크숍 진행¹⁸⁾
- 2012년, Mydata 이론 기반 실험 전략 수립 및 환경 설계¹⁹⁾
 - 지식 및 경험 기반으로 주요 이슈 및 과제 확인 및 실험 환경 설계
- 2013년 10월부터 2014년 5월까지 콘셉트 실증 및 프라이버시 작동 원리 탐구를 위해 대기업, 일반인, 디자이너, 개발자 등이 참여하는 실험 수행

MesInfos 파일럿 실험	
· 기간	: 2013년 10월 - 2014년 5월
· 참가자	: 개인 참가자(자발적 참여자) 300여명+8개 기업(데이터 제공기업 6사 및 서비스 지원 2사)
· 제공데이터	: 5백만(은행 기록, 현금 영수증, 통화기록, 지오로케이션(geolocation) 등)
· IT 인프라	: 코지 클라우드(Cozy Cloud)사의 pPaaS(Personal platform as a service) 사용
· 수행 결과	: 50개 이상의 서비스 아이디어 발굴 및 18개 프로토타입 개발
· 특징	1) 개인 인식 변화 조사 프로그램 병행 (실험 전·후 설문 조사 및 인터뷰 실시) 2) 이용내역, 프로필, 유형별 및 제공자별 데이터로 구분하여 상세 관찰 3) 실험 과정 및 결과를 기록한 보고서 발간 ²⁰⁾

- MesInfos 추진을 위해 연도별 다양한 프로젝트 실시

표 23 MesInfos 연도별 프로젝트

연도	프로젝트	주요내용
2012	Exploring Self Data	MesInfos 개념 수립 및 정책소개서 발간
2013	Experiencing Self Data	MesInfos 파일럿 연구 진행
2014	Challenges of Self Data	기술적, 법적 위험성 및 소매·미디어 등 확장 가능성 탐구
2015	Engaging major projects of Self Data	MesInfos Health, MesInfos Energy 추진
2016	Put Data Self in action	파일럿 프로젝트: 'Living lab of the Self Data'

- 2016년 10월, 일반데이터보호법(General Data Protection Regulation, GDPR) 개정 후속조치의 일환으로 디지털 재원의 인프라적 측면을 강조한 ‘Digital Republic Act’ 제정 및 공포
- 2016년부터 2017년 데이터 기반 MesInfos 건강 프로젝트 진행²¹⁾

o 활용 사례²²⁾

- 코지 클라우드 (Cozy Cloud): ‘개인 클라우드 플랫폼(Personal Cloud Platform)’ 제공. 각 개인은 개인별 데이터베이스와 함께 개별화된 서버인 ‘Cozy’ 를 가지고 있음. 코지는 Raspberry Pi2를 사용하여 데이터 제공자 혹은 소비자 본인에 의해 관리됨. 사용자는 데이터를 저장하고 어플리케이션을 통해 사용 가능
- 디기.미 (Digi.Me): 디기.미는 소셜미디어 데이터를 바탕으로 자동적으로 일기를 생성. 삶에서 특별한 순간을 검색 엔진과 달력을 통합해서 확인 가능. 디기.미는 소셜미디어에 올라온 모든 데이터 및 연락처 복사본을 생성. 생성된 데이터는 개인 저장소에 저장. 차후, 건강, 웰빙, 쇼핑, 금융 및 다른 데이터를 추가할 예정

③ 미국의 ‘스마트 공시(Smart Disclosure)’ 정책²³⁾

o 추진 목적

- 개인(소비자)이 더 나은 의사결정을 내릴 수 있도록 데이터 제공 및 데이터 투명성 확보
- 소비자의 데이터 접근 및 사용성 강화

o 추진 주체

- 2011년 7월, 국가과학기술위원회(National Science and Technology Committee) 주도 태스크포스팀 운영을 통해 진행
 - (에너지) 그린 버튼 : 미국 에너지국 주도
 - (의료) 블루 버튼 : 미국 재향군인회 주도
 - (학적) 마이데이터 버튼 : 미국 교육부 주도

o 추진 경과

- 2011년 9월, 백악관 예산관리국은 각 기관에 ‘스마트 공시’ 를 통해 소비자에게 가이드 전달
- 2011년 9월, 열린정부국가정책수립안(Open Government National Action

Plan)을 통해 스마트 공시 사용 촉진

- 2012년 3월, 백악관 및 미국 국가기록원(National Archives and Records Administration)은 스마트 공시 활성을 위한 정상회담 진행
- 2013년 3월, 백악관 내 과학기술정책실은 정부 기금을 받아 진행하는 과학 연구에 대한 접근 증가에 대한 가이드라인을 각 기관에 배포
- 2013년 5월, 백악관은 ‘Making Open and Machine Readable the New Default for Government Information’에 대한 행정명령 발표
- 2016년 백악관 보도에 의하면, 미국 에너지국 내 전기배달 및 에너지 신뢰성 사무국(Office of Electricity Delivery and Energy Reliability)은 스마트 공시에 대한 정부 정책에 따라 파일럿 프로그램인 데이터가드(DataGuard) 에너지 내 구현된 소비자 데이터 개인정보 프로그램에 부응할 수 있도록 민간 부문 Green Button Alliance의 노력을 분석할 계획²⁴⁾

o 활용 사례

- 스마트 공시 정책을 통해 각 개인은 그린/블루/마이데이터 버튼을 통해 각각 에너지/의료/학적 정보를 다운받아 직접 활용하거나 제3자와 쉽게 공유가 가능
 - 2016년, ESP 솔루션 그룹 보고서에 의하면, 현재 미국 내 교육부내 Mydata 버튼(학적 관련)이 존재하지 않음²⁵⁾
- 약 8,800만명의 미국인이 자신의 데이터를 모니터링하고, 1,500만명 이상이 블루버튼을 다운로드
 - 의료데이터를 안전하게 업데이트 해주는 ‘블루버튼 자동화’, 데이터를 활용한 다양한 앱·서비스 개발이 가능한 ‘블루버튼 매시업 도전’, ‘데이터 제공기관의 의미 있는 정보 활용(Meaningful Use: MU) 관리를 돕는’ 블루버튼 플러스 ‘서비스 등 제공
- 특히, 미국은 헬스케어 분야에 개인정보 활용사례가 다수 존재
 - ‘Open-Notes’ 사이트²⁶⁾ : 모든 개인 데이터 이용 환경에 적용 기업 참여를 보장. 환자 본인이 진료기록을 확인할 수 있도록 함
 - ‘에픽시스템즈’ 사²⁷⁾ : 개인의 개인건강기록(PHR: Personal Health Record)을 확인할 수 있는 마이차트(MyChart) 서비스 제공. 2015년 기준 미국의 54% 환자 정보와 전 세계 2.5%의 환자 정보 보유

4 핀란드의 ‘Mydata’ 정책

o 추진 목적

- (개인적 측면) 개인에게 개인 별 데이터셋 접근, 획득, 사용할 수 있

도록 구매기록 데이터, 교통 데이터, 통신 데이터, 건강기록, 재무 정보 및 다양한 온라인 서비스 데이터 등 제공

- (기업적 측면) 개인 데이터를 보유한 기업이 개인에게 데이터를 컨트롤 할 수 있도록 장려

o 추진 주체

- 방송통신 주무부처인 교통통신부(LVM)가 주관하고 비영리법인 OKF(Open Knowledge Finland)가 정책 수행 주도

o 추진 경과

- 기업 참여를 정책 성공 핵심요소로 판단하고, 기업 참여를 위한 기업 의견 수립
- 2014년, MyData 개념, 경제주체별 기대효과, 작동원리 등이 포함된 백서 발행 후 정책 구체화 추진
- 2015년 10월, 디지털 비즈니스 성장 전략 내 세부과제로 MyData 채택
 - 개인 중심 프라이버시 통제 강화 및 데이터 가치 증대와 함께 개방형 혁신을 MyData 핵심가치로 제시
- 2016년 6월, MyData가 포함된 데이터 비즈니스 활성화* 방안 공개
 - 데이터 경제 활성화 및 데이터 활용 규제 개선 및 MyData 서비스 인프라 및 시범사업 추진. 기업 간 데이터 공유를 위한 인터페이스 마련

o 활용 사례

- 각종 파일럿 사례를 통해 MyData 정책 추진 타당성 검토 중

□ 주요 시사점

- o 국내도 데이터산업 활성화 측면에서 개인 데이터를 개인이 쉽게 활용할 수 있도록 한 인프라/제도 연구 및 관련 정책추진 필요
 - 개인 데이터 활용을 위한 정책 타당성 분석은 해외 정책 추진 경과를 봤을 때 최소 2년 이상의 시간 소요 예상
 - 개인 데이터 활용을 위해서는 개인정보보호법의 규제성 조항이 함께 개정되어야 함
- o 민간 중심으로 다양한 개인 정보 거래가 활성화 될 수 있도록 민간 주도의 개인 데이터 원칙 수립 필요

* 예: 영국 MRS의 FAIR Data 원칙

5. 데이터 전문인력

□ 개념

- 기업에서 활용하는 데이터 및 데이터베이스의 보안, 성능, 품질 향상 또는 데이터베이스를 기획하거나 생성된 데이터나 데이터베이스를 관리하고 운영하는 인력을 가리킴
- 국내 주요기관에 따른 데이터 전문인력 직무는 아래와 같음

표 24 국내 데이터 직무 구분 비교

구분		주요 데이터 관련 정부 기관		
		한국데이터진흥원	한국정보화진흥원	고용노동부 NCS
데이터 설계		데이터 아키텍트	하둡/ NoSQL/ MapReduce 엔지니어	DB엔지니어
데이터 개발		데이터 개발자		
데이터 운영관리		데이터 엔지니어		
데이터 활용	데이터 분석	데이터 분석가/ 데이터 과학자	빅데이터 분석가	빅데이터 분석가
	데이터 시각화			
	데이터 연구			
데이터 관리행정	프로젝트 기획관리	데이터 과학자	-	-
	데이터 관리	DBA	-	
데이터 판매		데이터 마케터	빅데이터 기술영업 /마케터	-
데이터 컨설팅		데이터 컨설턴트	빅데이터 컨설턴트	-

□ 주요 현안

- 데이터 전문인력 수요 규모에 맞는 데이터 전문인력 양성이 필요하며, 기업이 요구하는 수준의 데이터 인력과 배출되는 데이터 인력 간 불일치* 심화
- * 데이터 전문인력을 원하는 수요기업들은 단순히 대학이나 교육기관의 전문과정을 통하여 직무능력을 습득한 인력이 아니라, 실무경험이 있는 경력자를 원함

□ 현황 분석

① 데이터 전문인력 수요·공급 현황

- 한국데이터진흥원에서 2015년 전체 산업 별 데이터 직무 인력 수요를 조사한 결과는 다음과 같음

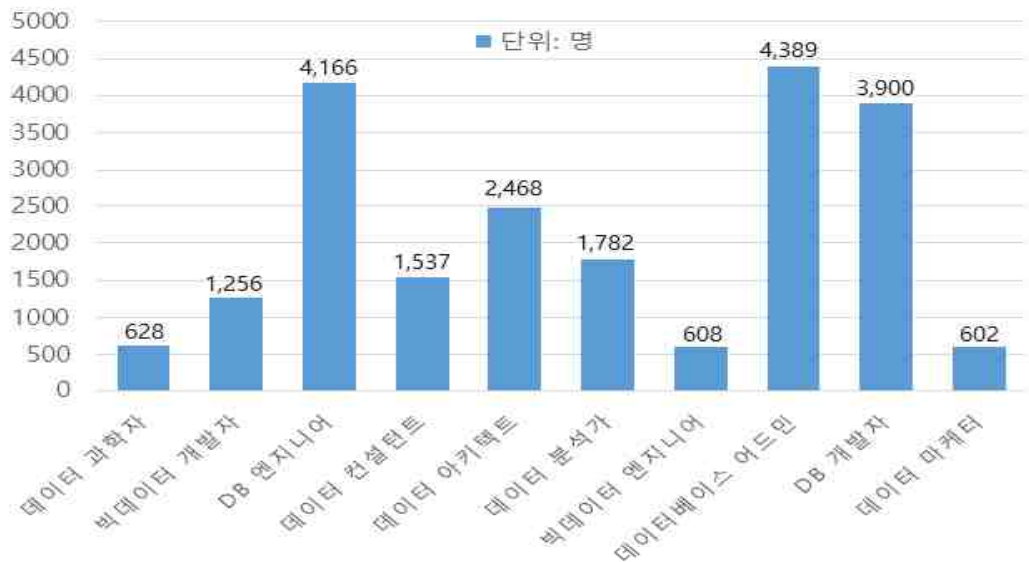


그림 5 2015년 전체 산업 별 데이터 직무 인력 수요

- ‘2016 데이터산업 백서’ 내 인력 대비 수요 비중을 살펴보았을 때 특히, 데이터 과학자 및 빅데이터 개발자 직무에서 수요가 높은 것으로 파악

* 인력 대비 수요 비중 : 데이터 과학자(46.8%), 빅데이터 개발자(39.3%)

○ 국내 전문인력 공급

- 데이터 전문인력에 대한 수요가 증가함에 따라 국내 정부 산하기관, 지방자치단체, 대학 및 대학원, 민간 교육기관 등에서 전문인력을 배출하고 있음
 - ▷ 정부 주도하에 진행되는 한국데이터진흥원의 빅데이터아카데미 및 고용노동부의 다양한 빅데이터 전문가 과정을 통해 2016년 기준 약 800명 이상의 교육생이 배출될 예정
 - ▷ 지방자치단체 주도하에 추진하고 있는 경기 빅데이터 전문인력 양성, 강원창조경제혁신센터 BIGTORY 등의 과정 및 청년희망재단의 빅데이터 기반 인력양성을 통해 약 200명 이상의 교육생이 배

출될 예정

- ▷ 대학 및 대학원의 경우, 최근 1~2년 내에 새롭게 데이터 관련 학과가 신설된 관계로 아직까지 학부과정에서 졸업생이 배출되지 못했음. 현재 대학원에서 연간 약 180명 정도 인력이 배출되는 것으로 보임. 향후 학부 졸업생을 포함하면 학부 및 대학원으로부터 매년 300~400명 정도의 전문인력이 배출될 예정
- ▷ 민간 교육기관의 경우, 삼성 SDS멀티캠퍼스, SAS코리아, 에이콘 아카데미, 한국 EMC, 엔코아 등에서 데이터 전문인력 양성을 통해 1,000명 이상의 교육생 배출

※ 정부, 지자체, 대학 및 대학원, 민간 기관을 통해 배출되는 전문인력은 총 약 2,200~2,500명으로 추정

해외 데이터 전문인력 수급 현황

1. EU

- 'EU & IDC(International Data Corporation)의 유럽 데이터 시장 보고서'에 따르면 2015년 EU의 데이터 근로자 수는 약 6백만 명이며, 2020년에는 최소 2.0% ~ 최대 9.2% 증가 할 것으로 예측²⁸⁾

2. 영국

- e-skills UK는 2017년까지 데이터 전문인력에 대한 수요가 영국 내 대략 6만9천 명으로 예측되지만 공급은 절반도 미치지 못할 것으로 보고 있음²⁹⁾
- 직무 별 인력 수요 현황으로는 개발자(42%), 아키텍처(10%), 분석가(8%), 관리전문가(6%) 순이었으며 데이터 과학자에 대한 수요는 1% 미만으로 나타남

3. 미국

- 맥킨지에 의하면 2018년까지 데이터 분석가에 대한 수요가 공급에 비해 50~60% 높은 것으로 예측³⁰⁾
- 또한 미국 노동 통계국에 의하면 '컴퓨터 조사 및 데이터 과학자'고용은 2014년에서 2024년까지 약 11% 증가할 것으로 전망³¹⁾
- 2013년 미국 대학에서 데이터 과학자 양성 과정을 통해 약 2,432명에게 학위를 수여했지만 산업계 및 연구소에서 필요한 데이터 과학자의 수는 수만 명에 이르는 것으로 조사되어 심각한 수급 불균형 현상을 보여줌을 확인

4. 일본

- 일본 경제산업성에서 IT기업과 일반 기업의 IT부서 대상으로 설문조사를 한 결과 IT인력에 대한 총 수요는 2016년 약 11만 2,000명으로 추정되고 그 중 부족한 IT인재는 약 13%(약 15,190명)인 것으로 조사
- 2020년에는 IT인재 수요가 약 17만 명을 돌파할 것으로 예상되며 약 22%(47,810명) 공급이 부족할 것으로 예측³²⁾

o 국내 전문인력 양성 프로그램

- 데이터 전문인력 양성 프로그램에서는 최근 수요가 가장 높은 ‘데이터 설계, 개발, 운영관리, 활용’ 직무 위주의 국내외 양성 프로그램을 확인

① 정부·공공기관 주도

표 25 국내 정부·공공기관 주도 인력 양성 프로그램

주도	프로그램명	교육 상세	
미래부 빅데이터 아카데미	빅데이터 기술 전문가 과정	교육 목적	빅데이터 수집·저장·처리·운영관리 등 주요 처리기술에 대한 지식 및 기술 습득 과정. 빅데이터 기술 습득을 통한 실무형 빅데이터 기술 전문가 양성
		커리 클	▲빅데이터 기술환경 구축방법 ▲데이터 저장·수집·처리, 기계학습, 기술응용 분석 ▲빅데이터 실무
		교육 대상	개발자·DBA·SE 등 3년 이상 경력자 (대학(원)생 및 미 취업자 참여불가), 빅데이터 프로젝트 수행 혹은 예정 인력
		교육 기간	10주
	빅데이터 분석 전문가 과정	교육 목적	빅데이터 분석 기획, 분석 방법, 분석 도구 활용 지식 기반으로 빅데이터 분석을 통해 새로운 가치 창출을 목표로 통계학 중심의 빅데이터 분석 과정, 처리기술 기반의 빅데이터 분석 과정
		커리 클	▲통계학 중심의 빅데이터 분석과정 ▲처리기술 기반의 빅데이터 분석 과정
		교육 대상	고개분석 및 마케팅, 데이터 분석 등 3년 이상 경력자, 빅데이터 프로젝트 수행 또는 예정 인력
		교육 기간	10주
	빅데이터 기획 전문가 과정	교육 목적	빅데이터에 대한 다양한 시각과 기술 및 분석 방법론, 활용 환경 등에 대한 이해를 바탕으로 빅데이터 활용의 기회 발굴, 사업 기획, 사업 관리가 가능한 빅데이터 기획 전문가 양성
		커리 클	▲데이터 과학자의 현실과 미래 ▲데이터처리분석과 비즈니스인사이트 ▲빅데이터 기획의조건 ▲기획자가 바라보는 빅데이터의 의미 ▲빅데이터 기획 프레임워크
		교육	사업 발굴·관리 업무 3년 이상 경력자

		대상	
		교육 기간	4주
	빅데이터 융합 전문가 과정	교육 목적	빅데이터 기술과 산업간 시너지 창출을 목적으로 '의료빅데이터 전문가 과정'과 '제조 빅데이터 전문가 과정'으로 운영
		커리 클	① 의료 빅데이터 전문가 과정 : 데이터 기반 의료 업무 수행 분석전문가 양성과정 ② 제조 빅데이터 전문가 과정 : 생산량 및 품질향상 제고 자동화 빅데이터 과정
		교육 대상	① 빅데이터 활용에 관심이 있는 의료 관련 재직자 ② 제조업 품질관리 및 생산기술 관련 빅데이터 활용에 관심 있는 재직자
		교육 기간	8주
	지역별 맞춤형 전문가 과정	교육 목적	빅데이터 활용 인력의 전국적 확산 및 시장 활성화 도모를 목적으로 강원, 경기, 대전, 부산 등 지역별 빅데이터 전문인력을 양성하는 프로그램
		커리 클	▲빅데이터 분석 개요 ▲데이터 수집 ▲R을 이용한 데이터 분석 ▲R을 이용한 머신 러닝 ▲고급 데이터 분석기법 ▲실시간 데이터 처리 ▲R을 이용한 분석 시스템 구축
		교육 대상	빅데이터 활용에 관심이 있는 기업 재직자 및 예비창업자
		교육 기간	4일
고용 노동부 빅데이터 전문가 과정	청년취업 아카데미	교육 목적	고용노동부가 한국 생산성본부, 한국표준협회 한국경제 TV등 교육기관에 위탁하여 실시하는 국비지원 프로그램으로 청년들에게 산업현장 수요에 적합한 교육과정을 제공
		커리 클	▲공통역량훈련(멘토링, 기초역량교육) ▲기초프로그래밍 ▲DB설계 ▲SQL 등
		교육 대상	취업을 앞둔 대학 졸업예정자 및 미취업졸업생
		교육 기간	아카데미 내 프로그램에 따라 상이 (2 ~ 6개월)

경기도	빅데이터 전문인력 양성과정	교육 목적	지자체 주도 및 산하 지역 공공기관이 운영하는 국내 최초의 광역자치단체의 데이터 인력양성 모델
		커리큘럼	① 빅데이터 전문가 과정 : 빅데이터 기반·기술·분석 역량을 강화하기 위한 과목으로 구성 ② 빅데이터 재직자 과정 : 빅데이터 분석역량을 강화하기 위한 과목으로 구성
		교육 대상	① 빅데이터 전문가를 희망하는 경기도민, 경기도 소재 대학(원)생, 타지역 소재 대학(원)생 중 경기도민 등 ② 빅데이터 활용에 관심 있는 재직자, 경기도 소재 기업 재직자, 경기도민 중 타 지역 소재 기업 재직자 등
		교육 기간	① 10주 ② 10일

② 대학 주도

1) 학부 과정

- 정규 교과과정은 산업에서 요구하는 빅데이터 전문인력 역량에 맞춰 융합형·실무형 인재육성을 지향
- 대학에서는 기존 데이터 마이닝과 정보통계처리학 등의 부분에서 확장되어 빅데이터와 융합한 학문융합, 경영정보, 마케팅 등 다양한 교육커리큘럼을 개발 중
- 최근 하기 대학에서 데이터 관련 학부 과정을 개설했으며 대부분 기존 학과 내 세부 전공 또는 연계 전공으로 운영 중

표 26 국내 데이터 관련 학부과정 개설 현황

학교명	단과대 및 소속	정원수	학위명	설립년도
국민대학교	경영대학 경영학부 빅데이터 경영통계전공	50	경영학사	2013
강원대학교	IT대학 컴퓨터정보통신공학과 데이터 분석 전공	60	공학사	2014
충북대학교	경영정보학과 빅데이터 전공	미정	경영학사	2015
성균관대학교	문과대학 데이터사이언스학과	미정	미정	2015

출처: DB Issue Report, 2016-11 제85호(국내 데이터 인재 양성 정책 및 교육 현황)

2) 대학원 과정

- 기업의 직무역량이 세분화됨에 따라 빅데이터 전문인력도 프로그래밍·수학·통계·비즈니스 등 모든 역량을 기대하는 것이 아닌 특정 분야의 전문인력을 요구하고 있기 때문에 고용계약형 과정, 재직자 계약학과 등의 과정이 증가하는 추이
- 대학원 과정은 크게 실무경력이 부족한 학부 졸업생을 위한 과정과 현재 관련 분야에 종사하고 있는 재직자 대상 과정으로 구분 가능
- 2012년 설립된 충북대학교 비즈니스 데이터융합학과를 비롯해 현재도 지속적으로 개설되고 있는 현황

표 27 국내 대학원 데이터 전문인력 양성 과정 현황

학교명	단과대 및 소속	정원수	학위명	설립년도
충북대학교	비즈니스 데이터융합학과	11명	공학석사	2012
	일반대학원 빅데이터학과	유동적	공학석·박사	2015
국민대학교	경영대학원 빅데이터경영MBA	15명	경영학석사	2013
	일반대학원 인문사회계열 데이터사이언스학과	5명	경영학석·박사	2014
서울과학기술대학교	일반대학원 공학계열 데이터사이언스학과	10명	공학석사	2014
연세대학교	정보대학원 빅데이터학과 (SMBA)	15명	정보시스템학 석사	2014
이화여자대학교	일반대학원 빅데이터분석학과	5명	데이터분석학 석·박사	2015
고려대학교	정보대학원 빅데이터응용 및 보안학과	20명 내외	공학석사	2015
단국대학교	일반대학원 공학계 데이터사이언스학과	유동적	데이터사이언 스학석사	2015
서울과학종합대학교	특화전공 MBA (빅데이터)	20명 내외	경영학전문석 사	2015
과학기술연합대학교	빅데이터 과학	6명	석·박사	2014

출처: DB Issue Report, 2016-11 제85호(국내 데이터 인재 양성 정책 및 교육 현황)

③ 민간 주도

- 대부분 국내외 IT기업 또는 민간 교육전문 기관에 의해 운영되고 있으며 기관 자체 기획 과정과 고용노동부 등의 위탁교육 과정으로 구분
- 재직자의 실무역량 강화 및 수요기업과 교육 이수자의 교육목적 달성을 위한 커리큘럼과 교육방법, 학습지원 등을 목적

표 28 국내 민간 교육기관의 데이터 인력양성 과정 현황(일부)

프로그램명	과정명	교육기관	대상	기간	방법	역량	수준	협력기관
빅데이터 교육	비즈니스 분석가	SAS Korea 교육팀	대학생 대학원생 일반인	6일	온라인 교육	분석	초급 중급 고급	
MDS 아카데미	Splunk 정규	MDS 아카데미	재직자	1일	이론·실습 교육	기술	고급	
에이콘 아카데미 프로그램	빅데이터 분석을 위한 머신러닝	에이콘 아카데미	재직자	5일	이론·실습 교육	분석	고급	연구소/ 강사협회/ 타 교육기관/ 출판사

해외 데이터 전문인력 양성 프로그램

① EU

1) 유럽 데이터사이언스 아카데미(EDSA)

- EDSA(European Data Science-Academy)는 EC가 Horizon 2020 프로그램을 통해 공공 자금을 지원받아 2015년 2월 1일부터 36개월 간 진행하는 프로그램
- 이러닝(E-learning) 기반의 범유럽 포털이자 데이터 분석과 데이터 과학 연구를 위한 열린 교육 채널로서, 유럽의 데이터 과학 관련 커리큘럼을 정립하고 교육 주제 및 교육 방식을 결정하고, 교재를 개발하는 것을 목표
- 영국 개방대학(Open University)와 오픈 데이터 연구소(ODI: the Open Data Institute) 주도하에 유럽 6개국에서 대학, 연구소, 중소기업, 비영리 기관 등 총 9개 기관이 컨소시엄에 참여
- EDSA 교육프로그램의 주요테마 : ▲데이터 사이언스의 기초 ▲데이터 저장 및 프로세싱 ▲데이터 분석 ▲데이터 해석 및 응용
- 교육과정 유형 : ▲셀프학습(Self Study Course) ▲온라인 대중공개 강좌(MOOC) ▲혼합강좌 ▲오프라인강좌 등
- MOOC는 코세라에서 제공하고 있으며 온라인을 활용해 언제, 어디서든 양질의 대학 강의를 들을 수 있게 한 새로운 형태의 고등교육 시스템

2) 빅데이터 민관협력 파트너십 프로젝트

- 주로 대학과 연구기관에 소속된 대학생들에게 실제 데이터셋과 사용자 친화적인 오픈리소스 데이터 분석 툴을 제공
- 데이터 직무 종사자에 대한 직무훈련과 다른 수준의 데이터 기술들을 습득하기 위한 특별과정을 지원하고, 엔지니어나 다른 인적자원을 위한 신규 자격증을 제공

② 영국

1) 정부·공공기관 주도

주도	프로그램명	교육 상세	
영국 데이터 센터	Bootcamp	교육 목적	2012년부터 미취업 졸업자들이 데이터산업 분야에서 일할 수 있도록 취업역량 강화를 목적
		커리 클럼	▲기업, 방송매체, 정부 및 공공에 데이터 센터 산업에 대한 인식 고취 ▲데이터센터 기준의 발전 및 조화 도모 ▲산업적 연구 및 발전안전 선도 ▲데이터센터 분야 기술적 차이 환기 ▲산업적 이해당사자들과의 협력체계 구축 ▲환경 지속성을 위한 프로그램 개발 ▲새로운 시장 및 성장 시장에 대한 참여 확대 등
		교육 대상	미취업 졸업자
		교육 기간	10일
영국 정부	데이터 과학촉진 프로그램	교육 목적	범정부적 훈련프로그램의 하나로 공공기관 데이터 기반 업무 능력 향상이 목적
		커리 클럼	본인 업무와 관련된 주제 직접 선정 후 새로운 프로그램을 도입시켜 데이터를 구축
		교육 대상	공공기관 임직원
		교육 기간	3개월

2) 대학 주도

대학	교육명	교육 상세	
유니버시티 칼리지 런던	기계학습 석사	과정	컴퓨터과학·수학·전기공학·물리과학 학위 졸업생 대상으로 1년간 part-time으로 진행
		커리클럼	▲감독학습 ▲프로그래밍 및 기계 학습을 위한 수학적 방법 ▲그래픽 모델 ▲적용 기계학습 ▲생물 정보학 ▲기계 학습의 고급 주제 ▲Machine Vision ▲정보 검색 및 데이터 마이닝 ▲추정과 확률 모델에서의 학습 ▲통계적 자연언어 처리 ▲이미지의 역문제 ▲바이오메디컬 이미징을 위한 전산 모델링 ▲Individual Project 등

		과정	-
		커리큘럼	▲CLOUD 기술과 시스템 ▲고성능 컴퓨팅 ▲기계학습 및 데이터마이닝 ▲자연언어 공학 ▲전문 연습과 연구 방법론 ▲텍스트 분석 ▲고급 웹기술 ▲컴퓨터보안 ▲컴퓨터비전 ▲IP Networking and Application ▲모바일 및 소셜 애플리케이션 프로그래밍 ▲네트워킹 원칙 등

③ 미국

1) 정부·공공기관 주도

※ 미국 국립과학재단(NSF)

- 학부생 및 대학원생을 대상으로 한 'IGERT(Integrated Education & Research Training)' (통합교육연구 훈련 프로그램)은 기술 분야의 대학원 과정(주로 박사과정)에서 인접 학문을 결합해 연구하는 '학제간 연구 인력'을 양성하는 프로그램
- IGERT 프로그램 중 'DIF21 Track for IGERT'는 핵심기술, 사회문제, 사이버 인프라와 관련된 빅데이터의 도전과제를 해결해 나갈 수 있는 미래 혁신인재 양성을 목적으로 운영
- 교육 : ▲빅데이터 과학의 핵심기술 습득 ▲컴퓨터 과학 및 데이터 관련 도전 과제의 해결 및 분석 역량 강화 ▲사이버 인프라 활용능력 학습으로 구성

※ OSTF - TechHire 프로그램

- 청년고용 촉진프로그램으로 '코딩 부트캠프'(coding bootcamp) 및 혁신적인 온라인 훈련을 통해 미래의 IT분야 인력을 양성하는 프로그램으로 독립적으로 운영되거나 지역 커뮤니티 칼리지 또는 대학을 통해 운영
- 교육 : 데이터 분석 교육 18개 과정, 데이터 과학 교육 5개 과정을 무료로 제공

2) 대학 주도

※ Data Incubator Bootcamp

- 코넬 대학에서 투자한 '데이터 인큐베이터 부트캠프'는 데이터 과학 교육과 고용지원을 위해 설립된 전문 조직으로 재직자 대상 맞춤형 교육 및 고용 지원과 전공학생 대상 기술교육을 통해 데이터 과학자를 양성

표 29 데이터 인큐베이터 - 데이터 과학자 양성프로그램

프로그램명	세부내용
데이터 분석가	· 재직자 대상의 4주 온라인 과정으로 기초 데이터 과학 기술의 실제 사례를 체험할 수 있는 프로그램 제공 · 커리큘럼은 수백 명의 산업별 파트너의 피드백을 거쳤으며 펠로우십 프로그램과 동일한 엄격한 방법론에 따라 개발

데이터 사이언스 펠로우십	· 뉴욕, 워싱턴, 샌프란시스코의 지정된 교육장에서 혹은 온라인을 통해 프로그램 참가 가능 · 동문회와 시니어 데이터 과학자로부터 멘토링을 받거나, 전문가들과 네트워크 형성 기회 제공
데이터 과학자 고용지원	· 데이터 사이언스 펠로우십 졸업 후 고용을 지원하는 프로그램 · 관심 있는 기업에게 이력서 열람, 코드 리뷰, 프로젝트 검토, 이벤트 참석, 펠로우와의 만남 및 인터뷰 등을 무료로 지원
데이터 사이언스 트레이닝	· 기업체 재직자 대상 프로그램으로 관심 있는 기업들은 직원을 8주간의 데이터 사이언스 펠로우십 프로그램에 등록시키거나 기업체 맞춤형 교육훈련 프로그램을 요청할 수 있음 · 데이터 인큐베이터에서는 수 주간의 커리큘럼 모듈을 2-3일간의 워크숍을 통해 콤팩트하게 제공하면서 초보 데이터 분석가를 유능한 실무자로 양성
데이터 사이언스 트레이닝	· 기계학습부터 하둡과 스파크까지 최신 기술을 학습하며, 학습도구를 실무에서 바로 사용할 수 있도록 프로그램 설치를 통해 클라우드 환경 제공 · 기업들은 재능 있는 데이터 사이언스 펠로우 중에서 데이터 과학자를 고용할 수도 있음

출처: 데이터 인큐베이터 홈페이지

④ 일본

1) 정부·공공기관 주도

※ 데이터 사이언스 스쿨

- 총무성 통계국, 통계연구소, 일본통계학회의 협력을 통해 통계 청 홈페이지에 ‘데이터 사이언스 스쿨’이 개설됨. 본 프로그램은 온라인 공개수업(MOOC) 방식으로 ‘온라인 수업’과 ‘온라인+오프라인 집체수업’으로 운영

표 30 데이터 사이언스스쿨의 온라인 교육과 온라인+오프라인 교육 방법

구분	학습방법
온라인 교육	· 1주일을 기본 학습 단위로 하루에 5-10개 강의 수강 (각 강의는 10분 정도의 동영상으로 구성)하고 강의 후 평가를 위한 쪽지 시험 실시 · 학습이 끝나면 과제가 제시되며 과제는 기한 내 제출 · 1개월 과정이라면 이를 4주 지속되면서 마지막으로 종합 과제 제출 · 매주 제시되는 과제와 마지막 주 제시되는 종합과제에 대한 전체 평가를 통해 수료기준을 채우면 수료증 수여
온라인+ 오프라인 집체 교육	· 온라인 수업에서 강의 동영상 시청과 과제 제출을 위한 기본적인 내용 학습 · 이후 오프라인 집체수업에서 강사와 수강생들 간의 논의를 통해서 이해를 돈독히 하고 응용력을 키우는 학습

	· 제공하는 강좌의 수준이나 기간은 대학 또는 대학원 수업 수준 상정 (교육생은 중학생부터 80대까지 폭넓게 이용하고 있지만, 대학 수준의 강의를 요구하기 때문에 그 수준으로 제공) · 고교 수준의 지식 등 수강 전에 필요한 '전제 조건'이 있는 경우는 수강 접수 시 모집 개요에 명시
출처 : gacco 홈페이지 - 데이터 사이언스 입문과정과 데이터 사이언스 훈련과정(실전편)으로 구분하여 진행 * 데이터 사이언스 입문과정 : 통계학의 기초 및 데이터의 분석 등 기본적인 지식을 배울 수 있는 입문 과정으로 4주간 진행 : ▲통계데이터의 활용 ▲통계학의 기초 ▲데이터 보는 방법 ▲공공데이터 입수 및 코스 정리 등의 과목으로 구성 * 데이터 사이언스 훈련과정 : 사례를 중심으로 현장에서 사용되는 실전적 데이터 분석 (통계분석) 기법을 실전해 보는 과정으로 5주간 진행 : ▲데이터 사이언스란? ▲분석의 개념과 사례 ▲분석의 구체적인 방법 ▲비즈니스에서의 예측과 분석 결과의 보고 ▲비즈니스에서의 데이터 사이언스 실현 등의 과목으로 구성	

② 기업이 원하는 데이터 전문인력과 배출되는 인력 간의 갭 분석

o 기업이 원하는 데이터 전문인력 수준

- ‘2016 데이터 전문인력 양성 방안 연구 결과 보고서’ 내 사람인 데이터 인력 채용공고 분석 결과, 수요기업 중 실무경력을 필요로 하지 않는 곳은 19.1%에 불과 했으며 요구 경력이 3~5년(중,고급)인 기업이 40.2%로 대다수 차지

표 31 국내 데이터 전문인력 채용 요구경력 현황

신입 또는 무관	1-2년	3-5년	6-10년	11년 이상
19.1%	15.6%	40.2%	18.6%	5.5%

출처 : 취업포털 사람인 빅데이터 관련 채용정보 활용

- 향후 교육 수준 향상 및 데이터 전문인력에 대한 다양한 요구 수준을 감안했을 때 고급 인력에 대한 수요 증가 예상

0 배출되고 있는 인력의 수준

- 국내 정부(미래부) 및 민간 기업(에이콘 아카데미, MDS 아카데미) 외 대다수의 데이터 전문인력 양성 프로그램은 현재 실무경험이 전무한 취업 준비자 대상으로 인력의 데이터 활용 수준이 초,중급에 머물러 있음
- ‘2016 데이터 전문인력 양성 방안 연구 결과 보고서’에 의하면 국내 데이터 전문인력 양성 프로그램의 교육특징을 기술, 기반, 분석, 사업 분야로 세분화하였을 때, 현재 모두 초, 중급 수준에 머물러 있음을 확인
- 고급 인력 자격 검증을 위한 국내 ‘데이터 관련 자격시험’은 해외와 달리 종류가 한정되어 있으며 직무별 세분화 미흡

표 32 국내외 데이터 관련 자격증

구분	주관	자격증	상세 내용
국내	한국 데이터 진흥원	데이터 분석 전문가	데이터 이해 및 처리 기술에 대한 기본 지식을 바탕으로 데이터 분석 기획, 데이터 분석, 데이터 시각화 업무를 수행하고 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 의사결정을 지원하는 직무를 수행하는 전문가
		데이터 분석 준전문가	데이터 이해에 대한 기본 지식을 바탕으로 데이터 분석 기획 및 데이터 분석 등의 직무를 수행하는 실무자
		SQL 전문가	데이터를 조작하고 추출하는데 있어 정확하고 최적의 성능을 발휘하는 SQL을 작성할 수 있고, 이를 토대로 SQL을 내포하는 데이터베이스 프로그램이나 응용 소프트웨어의 성능을 최적화하거나, 이러한 성능 최적화를 지원할 수 있는 데이터베이스 개체(뷰, 인덱스 등)의 설계와 구현 등의 직무를 수행하는 전문가
		SQL 개발자	데이터베이스와 데이터 모델링에 대한 지식을 바탕으로 응용소프트웨어를 개발하면서 데이터를 조작하고 추출하는데 있어서 정확하고 최적의 성능을 발휘하는 SQL을 작성할 수 있는 개발자
		데이터 아키텍처 (DA) 전문가	효과적인 데이터아키텍처 구축을 위해 전사아키텍처와 데이터 품질관리에 대한 지식을 바탕으로 데이터 요건 분석, 데이터 표준화, 데이터 모델링, 데이터 베이스 설계와 이용 등의 직무를 수행하는 전문가

		데이터 아키텍처(DA) 준전문가	효과적인 데이터아키텍처 구축을 위해 전사아키텍처에 대한 지식을 바탕으로 데이터 요건 분석, 데이터 표준화, 데이터 모델링 등의 직무를 수행하는 실무자
	한국 경제	경영 빅데이터 분석사 (기본, 중급)	빅데이터 분석 전문가란 빅데이터 이해 및 처리 기술에 대한 기본지식을 바탕으로 빅데이터 분석 기획, 빅데이터 분석, 빅데이터 시각화 업무를 수행하고 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 과학적 의사결정을 지원하는 직무를 수행하는 전문가
	한국 CRM 협회	CRM 전문가	고객 전략을 수행하는데 필요한 마케팅, 경영전략, 통계, IT의 기본지식을 바탕으로 체계적이고 과학적인 경영전략을 기획 및 실행할 수 있는 전문가
		마케팅 통계분석가	각종 통계분석 기법에 대한 이해를 바탕으로 비즈니스 상황에 따라 적절한 분석 기법을 능동적으로 활용하고 분석결과에 대한 종합적인 해석이 가능하며, 이를 마케팅 전략 기획에 반영할 수 있는 전문가
		데이터 마이닝 전문가	대용량의 데이터에 내재된 패턴, 규칙, 연관성 등을 찾아내 문제해결을 위해 적용할 수 있으며, 자료의 추출, 적재, 변환, 분석, 리포팅 및 전략화에 이르는 일련의 과정을 수행할 수 있는 전문가
고객DB 분석가		데이터베이스에 대한 이해와 SQL에 대한 운영 능력을 바탕으로, 고객데이터 분석결과에 대한 종합적인 해석이 가능하며, 전략적인 마케팅 통찰력을 도출할 수 있는 전문가	
해외	인폼즈 (INFORMS)	CAP (Certified Anaytics Professional)	공공 기관에서 운영되는 전방위적 분석 자격증으로 비즈니스 및 애널리틱스 문제점을 잡아내는 것에서부터 데이터 취득, 방법론, 모형 건설, 전개, 모형, 라이프사이클 관리에 이르기까지 다양한 스킬 측정 가능
	클라우드라	데이터 과학 전문가 자격증	빅데이터 관련 자격증 중 명성이 자자하며 빅데이터 고급기술을 보유하고 있는 엘리트 수준의 데이터 평가하는 서면 시험 및 생산 가능한 데이터 과학 솔루션을 개발해 실제 상황에서의 문제 해결 능력도 증명해야 함

	인소프트 (INSOFE)	빅데이터 애널리틱스, 최적화, 엔지니어링 자격증	국제 엔지니어링 학교에서 진행되는 CPEE(Certificate in Engineering Excellence Big Data Analytis and Optimization)은 R과 하둡에 중점을 두고 통계 모형, 데이터 분석, 머신러닝, 텍스트 마이닝, 최적화 등의 과정으로 구성
	컬럼비아 대학교	데이터 과학 전문성 인증시험	컬럼비아 일반대학원과 공과대학이 함께 실시하고 있는 데이터 전문성 인증시험은 데이터 사이언스 알고리즘, 확률과 통계, 데이터 과학 머신러닝, 데이터 분석 및 시각화 탐구에 이르는 과정으로 구성
	스탠포드 대학교	대용량 데이터셋 마이닝 자격증	소프트웨어 엔지니어, 통계전문가, 예측 모형 전문가, 시장 조사원, 분석 전문가 및 데이터 마이너를 위한 자격증으로 소셜 네트워크 그래프, 대규모 문서 저장소 등에서 효율적으로 정보를 추출해내는 기술과 알고리즘을 증명해야 보여야 함
	EMC	EMC 데이터 과학자 협회 자격증	한 개인이 데이터 과학팀의 일원으로서 빅데이터 프로젝트에 참여하고 기여할 수 있는 능력을 측정. 데이터 분석 라이프사이클, 비즈니스 과제의 분석 과제로의 재구성, 분석 기술 적용 및 빅데이터 분석 툴, 통계 모형 제작, 적합한 데이터 시각화 선정 등의 과정
	레볼루션 애널리틱스	레볼루션 R 엔터프라이즈 프로페셔널	고급 분석 프로젝트에서 R 통계언어 사용 능력 측정. 여기에는 실용적, 전략적 측면에서의 빅데이터 분석, 데이터 분석 라이프사이클, 고급 분석의 이론 및 방법론, 통계 모형 등 포함

□ 주요 시사점

o 전문인력에 대한 수급 불균형 현상 해결 필요

- 2010년 이후 국내 데이터산업 시장 규모는 매년 증가 추세에 있으며, 그에 따른 데이터 전문인력의 수요는 급증할 것으로 전망
- 현재 운영 중인 산학협동 프로그램의 수로는 수요측이 원하는 실무 위주의 데이터 전문인력 수급 현상을 해소하기는 힘든 실정
- 2016년 한국정보화진흥원의 보도 자료에 따르면, 중소기업의 경우 자금과 인력부족으로 인한 데이터 활용 부재로 인하여 대외 경쟁력이 약화될 가능성 존재

=> 현재 운영 중인 산학협동 프로그램 지원 및 고도화로 2020년까지 데이터 전문가 20,000명 배출을 통해 수급 불균형을 완화 할 수 있는 방안 마련

o 기업 요구 수준에 맞는 고급 인력 양성 필요

- 국내 데이터 전문가 양성 프로그램의 명확한 커리큘럼 부재로 인한 인력양성의 질적 측면 미흡
- 2016년 데이터 전문인력 양성 방안 연구 결과 보고서에 따르면, 수요기업의 데이터 전문인력 수준 니즈는 중, 고급 인력이나 현재 배출 인력의 수준은 초, 중급 수준에 머물러 있음
- 해외 대비 국내 데이터 자격시험은 데이터 전반으로 직무별 자격시험 불일치 및 직무별 세분화 미흡
- 산재 되어 있는 데이터 교육 커리큘럼으로 인하여 정확한 데이터 전문가 교육 진행에 있어 수준에 대한 불일치 현상이 나타남

=> 국가 공인 데이터 전문가 자격시험 확대를 통한 인력 수준 향상

6. 데이터 기술

□ 개념

- 전통적인 데이터 수집, 저장, 분석, 처리, DBMS 관련 기술과 머신러닝, 사물인터넷, 오픈소스 기반 플랫폼, 인공지능 등 데이터산업 주요 기반이 되는 최신 테크놀로지를 가리킴

전통적인 데이터 기술 동향

○ 오픈 API

- HTTP 프로토콜 통해 오픈 API를 제공하는 사이트의 수집 요구가 증가하고 있으며, 웹 크롤러 범용 제품에서 오픈 API 설정까지 할 수 있는 제품 등이 개발되고 있는 추세

○ DBMS

- 기존 온프레미스 기반에서 장기적으로 클라우드 기반으로 재편될 것으로 보이며, 온프레미스 기반에서 운영되던 시스템과 새롭게 구축되어지는 클라우드 시스템 간의 효과적인 운영이 새로운 시대 IT 조직의 고민이 될 것으로 보임

○ 구조적 데이터 관리 기술

- 구조적 데이터의 관리 기술은 우선 데이터를 물리적으로 적재함에 따른 효율적 알고리즘과 관련된 기술의 발전이 주류를 이루었으나, 최근에는 데이터 관리와 관련된 사용자 편의성 증진을 위한 여러 기능의 발전이 주류를 이루고 있음

○ 비구조적 데이터 관리 기술

- 비구조적 데이터의 관리 기술은 구글 GFS(Google File System)와 맵리듀스(MapReduce)기술에서 파생된 오픈 소스인 하둡 생태계에 기반을 두고 관련 기술이 발달되고 있음

○ 빅데이터 처리 기술

- 빅데이터 처리를 위한 기술 제공업체들의 다양한 시도들이 확대되고 있는데, 대표적인 것이 대량 데이터의 빠른 처리를 위한 인메모리 기술 활용, 많은 데이터를 처리하기 위한 MPP 아키텍처도입, 기존 DBMS 성능을 개선하고자 하드웨어와 결합된 DB 어플라이언스, 적은 비용으로 대량 데이터들을 관리하기 위한 하둡 활용 등

○ 빅데이터 분석 기술

- 빅데이터 분석 기술에 대한 관심은 비정형 데이터뿐만 아니라 정형 데이터로 확대되고

있으며, 특히 장기간 의무적으로 데이터 저장이 법제화 되는 경우나 금융권을 중심으로 주목받고 있는 이상거래 탐지시스템(FDS, Fraud Detection System) 등에서 보다 폭넓게 빅데이터 분석 기술들이 확산되고 있음

- 빅데이터 분석을 위해 하둡 진영에서는 기존 맵리듀스 기법을 개선하기 위해 스파크가 등장하여 하둡 환경을 빠르게 재편하고 있음
- 하둡 HDFS에 저장된 데이터에 SQL 질의를 쓸 수 있도록 한 SQL-on-Hadoop 제품인 아파치 타조(Apache Tajo), 클라우데라 임팔라(Cloudera Impala) 등이 등장하여 빅데이터 분석 시장으로의 접근을 모색하는 중

o NoSQL

- 몽고DB로 대표되는 NoSQL 기술은 복수의 저가 서버들을 클러스터링, 샤딩(Sharding) 등의 방법을 통해 데이터를 분리·처리하며 대량의 데이터를 빠르게 처리할 수 있는 기술로 평가받고 있음

o 데이터 보안

- DB접근제어(DAP, Database Audit and Protection)에서의 데이터베이스에 대한 접근 패턴을 학습하고 이상 징후 시 이를 차단하는 기술
- 다양한 가상화 환경과 클라우드 환경이 기존 인프라에 연계되고 있으며, 수많은 기업내 IDC에 존재하던 데이터베이스들이 가상화/클라우드 공간으로 이동하면서 데이터베이스 보안에 대한 기술도 진화하고 있음

o 데이터 품질

- 최근 데이터 품질관리솔루션과 모델링 도구, 애플리케이션 영향도 분석 도구, 메타데이터 관리 도구 등의 솔루션을 함께 도입하여 상호 연계를 통해 프로파일 자동 연계 등록, 메타데이터 등과 싱글뷰(single view) 구현, 변경관리 시점에서의 품질 점검 등을 수행함. 또한 영향도 분석 솔루션과 연계를 통해 오류 데이터 원인 추적 등 데이터 품질관리를 위한 활용성을 극대화하려는 사례가 늘고 있음
- 데이터 품질관리 관련 기술의 경우 데이터의 정확성을 위하여 업무규칙이 세분화되고 전문화된 방향으로 발전하고 있는데, 향후에는 AI를 사용하여 데이터의 정확성, 일관성 등 데이터 품질과 관련된 진단을 하는 방향으로 발전될 가능성이 높음

□ 주요 현안

- o 머신러닝과 같은 인공지능 기술이 발전함에 따라 데이터 확보 필요성이 증대되고, 사물인터넷을 포함한 다양한 데이터 수집을 위한 데이터 표준화, 개인정보 비식별화 이슈 부각

□ 현황 분석

① 머신러닝 등 데이터 처리 및 분석 기술 진화

o 머신 러닝에 대한 관심 급증

- 급진적인 연산 능력, 무한대에 가까운 데이터 양, 심층 신경망(Deep Neural Network, DNN)에 대한 발전으로 활용 방향성 증대
- 인공지능 및 고급 머신러닝은 계속 발전할 것으로 예측
 - 가트너(Gartner)가 발표한 2017년 10대 전략 기술 트렌드에 인공지능과 고급 머신러닝이 포함

o 해외 기업들의 머신러닝 기술 공개 증가

- 머신러닝 기술을 선도하고 있는 글로벌 기업들은 개발된 기술을 경쟁적으로 공개하여 기술주도권을 선점, 시장 영향력을 확대, 다양한 응용을 발굴하며, 사용자 확장을 통해 우수한 인재를 확보하고, 외부 연구자들의 도움으로 기술개선 효과 도모

o 머신러닝 기술의 실사용 확대

표 33 머신러닝 대표 사례

사례명	회사명	적용 분야	주요 내용
Watson Health	IBM	헬스케어	뉴욕암센터, MD앤더슨 암센터와 협업하여 진단 및 치료법 제안, 보험사 웰포인트는 치료적절성에 왓슨 권장
자율운행차	구글	제조	주변상황과 표지판 등을 알아보고, 자율운행차가 판단하고 인식하여 운행
Siri	애플	생활 서비스	음성인식기반 개인비서 서비스 ‘시리’를 통한 자연어 대화 기술을 스마트폰에 최초 탑재
켄쇼	켄쇼 테크놀로지	금융	방대한 양의 금융 데이터(기업 정보, 주요 경제수치, 주가 등)를 분석하여 사용자가 원하는 질문의 답을 제공. 50만 달러의 연봉을 받는 전문 애널리스트가 40시간에 걸쳐 하는 일을 단 몇 분 내에 처리

- IBM의 왓슨을 활용한 ‘IBM 왓슨 인공지능 암센터’ 국내 진료 실시
- 환자 정보를 입력하면 왓슨이 환자 상태를 분석하고 치료 방법 및 근거 제시

- IBM의 경우 익스플로리스(Explorys)와 같은 의료 데이터 분석 업체를 인수하고, 각종 대학교 및 병원 등을 제휴하는 등 수 많은 데이터를 확보하여 학습효과를 증대시켰음
- 미국중양학회에 따르면 왓슨의 진단 정확도는 자궁경부암 100%, 대장암 98%, 직장암 96%, 난소암 95%, 췌장암 94%, 방광암 91%, 신장암 91%로 발표

② 사물인터넷(Internet of Things, IoT) 시대 도래

- o 인터넷, 센서 기술 등의 발달로 사물들끼리도 정보의 공유가 가능한 사물인터넷 시대 도래
 - 교환해야 하는 데이터의 양이 많지 않은 사물들을 소물(small thing)이라 하며, 소물들로 구성된 네트워크를 소물 인터넷(IoST)이라 하고, 최근 실생활에 밀접하게 접근하고 있음
 - 2016년 11월 9일 IDC의 연구보고서(G20 국가와 사물인터넷: 비교 평가)에 사물인터넷로 수익을 거둘 준비가 가장 잘 된 국가 2위로 한국을 선정함
- o 사물인터넷 표준화 기관 중 가장 규모가 큰 OCF(Open Connectivity Foundation)가 지난 2016년 10월 또 다른 표준화 기구인 올신얼라이언스(AllSeen Alliance)와 합병하면서 활발한 표준화 작업이 진행 중

사물인터넷 관련 국내외 정책 추진 경과

- o (미국) 경제 성장과 소비자 보호를 위해 사물인터넷 개발지원을 위한 국가 전략 수립을 촉구하는 결의안 통과('15.03)
 - 국가 전략을 수립해 공공 기관과 민간 기업에 적절한 가이드 제시
 - 사물인터넷 개발과 활용 촉진을 우선시하는 정책 수립
 - 다양한 이해 관련자 간 합의 기반의 우수 사례와 커뮤니케이션의 중요성 강조
 - 사물인터넷 사용 의무화
 - 차세대를 위한 삶의 질 증진에 기여
 - 스마트 시티, 스마트 인프라 등 지속적인 혁신 기술 개발 및 세계 주도적 역할 추구
 - 사물인터넷 공청회 개최
- o (중국) 감지중국 전략 추진
 - 사물인터넷과 클라우드 등을 활용하여 사물인터넷 기반의 스마트 도시 건설, 사물인터넷 산업시범단지 등 추진

- (일본) 첨단기술에 대응한 정책 로드맵 및 목표 수립('16.7)
 - 2025년 모든 생활 영역에 사물인터넷 접목 목표
 - 활용 모델 발굴, 인재 육성, 인프라 정비, 기술표준화 등을 기준으로 방향성 수립
 - 최근 일본에서는 사물인터넷과 인공지능을 결합하려는 노력 시도
 - NCE는 센서 등으로부터 수집된 데이터를 데이터베이스에 축적하고, 각 AI 엔진에 맞게 분석하는 솔루션 제공
 - 후지쯔는 AI '진라이'를 활용해 제품과 업무 공정별로 학습 모델을 달리하는 AI 시스템인 '제조용 AI 프레임워크' 제공
- (국내) K-ICT 전략
 - SW·신산업 분야의 전략산업 육성 목표 중 하나로 사물인터넷 선정
 - 세계에서 가장 손쉽고 다양한 사물인터넷 제품 생산, 서비스 활용이란 목표로 차세대 기술 표준화를 선도하기 위한 전략 수립

③ 데이터 비식별화

- 개인정보 보호 강화에 대한 사회적 요구는 증가하고 있으나 빅데이터 등 데이터 활용 확산으로 인한 데이터 활용가치가 증대됨에 따라 개인정보 보호와 활용을 동시에 모색하기 위한 비식별화 기법³³⁾에 관심을 기울이고 있음

표 34 해외 비식별화 지침 특성

	영국	호주	미국
명칭	익명화 : 데이터 보호 위험 관리 지침	데이터 및 개인정보의 비식별화	개인정보 비식별화
발행 기관	영국 독립 정보감독청 (ICO, Information Commissioner's Office)	호주 독립 정보감독청 (OAIC, Office of the Australian Information Commissioner)	미국 국립표준기술원 (NIST, National Institute of Standards and Technology)
발행 시기	2012년 11월	2014년 4월	2015년 10월
개인정보 정의	자체로 식별되는 개인 또는 개인정보처리자가 입수 가능한 다른 데이터와 결합해 식별 가능한 개인에 관한 데이터	식별된 개인 또는 합리적으로 식별 가능한 개인에 관한 정보	식별되는 개인과 식별 가능한 개인과 연관되는 정보

	영국	호주	미국
용어	익명화 (Anonymisation)	비식별화 (De-identification)	비식별화 (De-identification)
비식별화 처리 원칙	정보주체를 식별할 수 없도록 익명화해서 유럽 데이터 보호 원칙에 적용되지 않도록 함	개인 식별정보 삭제 독특한 정보와 결합이나 특징으로 개인이 식별될 수 있는 정보 삭제, 변경	정보주체와 데이터셋 연관성 제거

o 국내, 빅데이터 비식별화 가이드라인 발표

- 도입배경 : 빅데이터 활용을 위한 비식별화를 구체적으로 안내하여 안전한 빅데이터 활용기반 마련과 개인정보 보호 강화
- 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부 등 개인정보보호 관련 유관 정부부처 합동으로 2016년 6월 발표

빅데이터 비식별화 추진 절차
- 1단계(사전 검토) : 개인정보에 해당하는지 여부를 검토 후, 개인정보가 아닌 것이 명백한 경우 법적 규제 없이 자유롭게 활용 - 2단계(비식별 조치): 정보집합물(데이터셋)에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체하는 등의 방법을 활용, 개인을 알아볼 수 없도록 하는 조치 - 3단계(적정성 평가): 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는지를 「비식별 조치 적정성 평가단」을 통해 평가 - 4단계(사후관리): 비식별 정보 안전조치, 재식별 가능성 모니터링 등 비식별 정보 활용 과정에서 재식별 방지를 위해 필요한 조치 수행

- 개인정보 비식별 조치 지원을 위해 각 부처 소관 분야별로 중앙행정기관이 전문기관 지정
- 전문기관에서는 비식별 조치 적정성 평가단 풀 구성 및 운영, 산업별 비식별 조치 수준 이행 권고(k-익명성 값 등), 비식별 조치 적정성 실태 점검, 비식별조치 교육 및 컨설팅, 정보집합물 결합 지원, 국내외 비식별조치 관련 정책·제도·기술 동향 조사 및 연구 등을 수행
- 추가로 비식별화 관련 법제화 노력이 필요하며, 기술 및 프로세스에 대한 국제 표준화 적극 참여 필요

④ 오픈소스 환경의 부각

o 공공기관의 오픈소스 사용 증가

- 공공데이터 개방을 선도하고 있는 영국은 데이터 플랫폼까지 오픈소스로 개방해, 이를 활용한 다른 플랫폼과 데이터를 더욱 자유롭게 긴밀히 연동할 수 있게 지원(2010년 ‘CKAN’ 공개)

CKAN³⁴⁾

o 특징

- 비영리단체인 Open Knowledge Foundation(OKF)에 의해 설립된 공공데이터 개발을 지원하는 오픈 소스 데이터 플랫폼
- 현재 영국, 미국, 캐나다 등 40여 개국에서 CKAN을 기반으로 데이터 포털을 운영 중에 있음
- CKAN은 데이터 게재, 공유, 조사, 사용을 위한 도구를 제공하여 데이터 접근을 용이하게 만드는 데이터 관리 시스템

o CKAN 해외사례

- 영국, ‘Data.gov.uk’ : 2009년 사이트 개설을 위해 Drupal을, 후위 처리(back-end)를 위해 CKAN을 사용하여 개설된 최초 정부 데이터 포털임. 2012년 CKAN에 의해 즉시 데이터 검색이 가능하게 조정 후 재개설
- 미국, ‘Data.gov’ : 미국 정부는 2013년 1월 오픈 데이터 카탈로그, 기존 다양한 장소에서 발행된 데이터셋(Data set)을 융합하기 위해 CKAN을 도입하기로 발표하였으며 도입 후 2013년 5월 새로이 개방
- 미국, ‘Energy Data eXchange’ : 에너지부(Department of Energy, DOE)와 국가에너지기술노동부(National Energy Technology Laboratory)에 의해 개발되고 관리되고 있음. 본 사이트는 내·외부적 조정 및 화석 에너지 조사 지원을 위해 개설됨. EDX를 통해 내·외부 데이터셋(Data set)에 접근할 수 있을 뿐만 아니라 데이터 상호작용을 위한 도구로도 제공됨. 다른 CKAN 사이트(예. Data.gov)와의 협업을 조성하기 위해 CKAN 플랫폼을 선택
- 캐나다, 오타와 ‘City of Ottawa Open Data Catalog’ : 오타와 시는 2013년 4월 CKAN을 도입하여 오픈 데이터 포털을 새로이 런칭함. CKAN 도입을 통해 검색 측면, 데이터 어플리케이션 구축을 위한 API 프로그래머, 알림 및 RSS 피드 향상이 가능해짐
- 일본, ‘Data for Japan’ : CKAN Japanese로 유명한 일본 내 오픈 데이터 수집을 목적으로 한 커뮤니티 프로젝트임. 본 사이트는 링크드 오픈 데이터 이니셔티브(Linked Open Data Initiative), Code for Japan, 일본 OKF 주도로 진행

- 데이터의 원활한 원스톱 서비스를 위해 데이터 업로드, 다운로드, 검색, 시각화 등 오픈데이터 플랫폼의 중요성 대두

- 국내에서는 데이터진흥원의 데이터스토어를 개방형 플랫폼으로 변경 개발('18)할 계획을 준비하고 있음

o 민간기업의 오픈소스 사용 증가

- 최근 들어 오픈소스 R과 파이썬을 활용하여 데이터 분석 환경을 구축하는 사례가 급격하게 증가하고 있음
- 대부분의 공공사업은 오픈소스 활용을 주요 사업 규격으로 제시하고 있고, 특히 머신러닝 분야의 경우 오픈소스 기반의 솔루션들이 기존 상업용 솔루션과의 경쟁에서 우위를 점하는 현상이 발생하는 중
- 오픈소스 관리 업체 블랙덕 소프트웨어와 벤처 캐피털 업체 노스 브릿지가 공동으로 조사한 결과에 따르면, 온라인 응답자 전체 1,300명 중 3분의 2가 소속 기업이 개발자의 오픈소스 프로젝트를 장려하고 있다고 답했으며, 응답자의 60%가 오픈소스 제품 도입이 경쟁력 요인이라고 응답함

오픈소스 프로젝트 기대주

- 1) 로켓챗(Rocket.Chat) : 자체 채팅 서비스를 관리하고 싶어 하는 커뮤니티와 기업을 위해 개발된 오픈소스 웹 채팅 플랫폼
- 2) 매터모스트(Mattermost) : 어느 곳에서든 검색 및 접근이 가능한 직관적인 인터페이스를 통해 모든 팀 의사소통을 한 곳으로 통합한 협업 솔루션
- 3) 허블인(Hubl.in) : 무료 오픈소스 화상회의 솔루션
- 4) MXNet : DMLC가 개발한 가벼운 딥 러닝 라이브러리
- 5) 바젤(Bazel) : 내부 소프트웨어 개발 시스템의 하위 집합으로, 소스로부터 모든 소프트웨어를 개발하는 공유 코드 저장소를 통해 개발 과정을 가속화하고 신뢰성을 증가시킬 수 있음
- 6) 리액트 네이티브(React Native) : iOS와 안드로이드 모두에서 사용 가능한 애플리케이션을 두 번 개발이 아닌 한 번에 구축
- 7) 콘테나(Kontena) : 기업이 컨테이너화된 시스템을 개발, 배치, 모니터링 하는 데 필요한 모든 것을 제공
- 8) 글루코시오(Glucosio) : 최초의 오픈소스 당뇨 모니터링 애플리케이션. 사용자들은 혈당 수준을 입력하고 추적하며 사용자에게 수분을 보충하거나 혈당을 확인하도록 알리는 앱 내 지원을 통해 유용한 정보를 얻음

□ 주요 시사점

- 데이터 학습량 증진을 위한 기반 조성 필요
 - 머신러닝을 적용한 구글, IBM, 애플과 같은 해외 기업의 경우, 방대한 양의 데이터 학습을 통해 머신러닝 서비스 품질 향상
- 사물인터넷 기술 표준화를 위한 표준업체간의 지속적인 노력 필요
 - 표준화를 통해 데이터 접근성을 향상시키고, 이를 기반으로 다양한 기술을 접목시킨 활용 모델 개발
- 해외 CKAN과 같은 오픈 소스 데이터 개방형 플랫폼을 통해 데이터산업 활성화에 일조
 - ckan.org에 의하면 CKAN을 통해 사람들은 다양한 데이터 접근이 가능하며 기존에 비해 많은 지식을 쌓을 수 있음

III. 추진과제

1. 데이터 원칙 수립 및 제도 확립

가. 현안

- 미국, 영국 등 선진국은 데이터 원칙을 수립하고 있음. 우리나라는 데이터 원칙이 수립되어 있지 않으며, 관련 법규에 일부 포함되어 있는 형태임. 데이터 원칙이 없이 관련 법규들이 제정되어서 중복 규제 등 제도 통합성 미흡
- 데이터 원칙은 데이터 활성화 원칙과 데이터 규제 원칙의 균형을 이루어야 함. 우리나라 법규 및 제도는 데이터 규제 원칙 중심으로, 데이터 이용 활성화를 위한 원칙은 미흡. 선진국의 경우, 기계 판독성, IoT 표준 준수, 상세 데이터 공개 등 활성화를 위한 기준을 데이터 원칙에 표현하고 있음
- 미국, 영국 등 선진국은 데이터 활용기업들의 단체가 중심이 되어 데이터 원칙 제정에 적극 참여. 또한, 자율 규제 및 자체 인증 등을 통해 소비자들의 데이터 우려 불식에 기여. 우리나라는 소비자 단체 중심으로 데이터 제도화에 참여함으로써 상대적으로 개인 프라이버시 보호 입장이 강조되고 있음

나. (과제 1-1) 데이터위원회 구성

- 목표 : 산업 육성과 데이터 보호의 균형을 이루기 위한 범국가 차원의 공감대 형성을 달성하기 위한 조직 구성 및 거버넌스 확립
- 추진 방안
 - 1) 국회와 행정부, 산업계와 학계, 소비자보호 단체와 데이터 활용 기업 등을 대표하는 전문위원으로 구성
 - 국회 : 여당과 야당 의원 포함
 - 행정부 : 범정부 차원에서 폭넓게 구성
 - 산업계, 학계, 소비자보호단체, 데이터활용 기업 등 : 개인정보보호와 데이터활용의 관점이 균형을 이루도록 구성
 - 데이터 활용 기업들을 대표하는 단체: 외국의 경우 DMA가 역할을 수행하고 있으나, 우리나라에는 현재 설립되어 있지 않음. 데이터위

원회 활동을 위해서는 외국의 DMA와 같은 단체 설립 필요

- 2) 데이터 관련 현안을 식별하고, 연구 및 조사 활동을 통해 대안 수립
- 3) 전문가 리뷰, 공청회, 여론 조사 등을 통해서 국가 차원의 공감대 형성
- 4) 데이터 관련 법규와 제도, 정부 사업 기획, 예산 편성, 사업 평가 등에 의견 제시

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
준비위원회 구성								
데이터위원회 발족								
운영계획 수립								
데이터위원회 활동								

다. (과제 1-2) 데이터 원칙 제정

- 0 목표 : 데이터 규제와 활용의 균형을 이루는 원칙을 제정하고, 국가 차원의 공감대 확보를 통해 데이터산업 육성 및 개인정보보호를 위한 기반 마련

0 추진 방안

- 1) ‘(과제 1-1) 데이터위원회’가 구성되면, 주체가 되어 데이터 원칙 제정 착수
- 2) 국외 사례 조사 및 국내 현황 파악을 통해 관련 법규의 중복 규제 등 개선점과 소비자피해 규제 등 보완점 도출
- 3) 데이터 활성화 원칙을 구체적으로 수립
- 4) 데이터 원칙 초안을 제정하고 공청회 등을 통해서 의견 수렴 후 확정. 이후 데이터 원칙은 매년 필요에 따라 지속적으로 개정
- 5) 데이터 원칙에 맞도록 관련 법규 및 제도 보완 추진

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
데이터위원회 발족								
데이터원칙 수립								
데이터원칙 적용/보완								

라. (과제 1-3) 데이터 규제 패러다임 전환

- 0 목표 : 국가 경쟁력 차원에서 높은 수준의 데이터 규제 패러다임을 보통 수준으로 전환함으로써 데이터 활용을 통한 기업 혁신 및 신성장 산업 창출 기반 확보

0 추진 방안

- 1) ‘(과제 1-2) 데이터원칙수립’ 과 연계하여 추진
- 2) 데이터 규제 관련 연구활동 수행을 통하여 공감대 기반 형성
 - 선진국과 우리나라의 데이터 규제 수준 비교 연구
 - 데이터규제가 국가 경쟁력에 미치는 영향 연구
 - 사전 규제와 사후 책임 관리, 옵트인과 옵트아웃, 포지티브시스템과 네거티브시스템 등의 대안 비교 연구
 - 소비자 관점에서 데이터 활용 편의성과 소비자피해 보상 제도 연구
- 3) 개인데이터 활용 촉진 및 검증을 위한 규제 프리존 운영
- 4) 데이터 제도 패러다임 설정 및 법규 및 제도에 반영

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
데이터연구활동 수행								
규제프리존 운영								
데이터원칙 수립								
법규 및 제도 적용								

② 데이터산업 생태계 육성

가. 현안

- 데이터 활용이 활성화되기 위해서는 데이터 생산, 가공, 유통, 활용 등의 데이터산업 생태계가 조성되어야 하나, 우리나라는 고객이 승인한 데이터 생산에만 치우쳐 있어서 산업 전반에 데이터 활용 환경이 선진국 대비 열악한 수준
- 다양한 데이터 비즈니스 모델은 스타트업을 발굴하고 양성하는 환경이 갖추어져 있어야 하나, 정책 및 제도적 뒷받침이 부족한 상태임

나. (과제 2-1) 데이터 중개 비즈니스 모델 육성

- 목표 : 데이터 가공, 데이터 분석 등 미흡한 분야의 데이터 기업 발굴. 산업별로 데이터 중개를 담당하는 민간 주도 데이터 중개 기업 육성 추진

○ 추진 방안

- 1) 데이터기반 비즈니스모델 프레임워크(Data Driven Business Model Framework)을 정의하고, 이를 토대로 국내 데이터산업의 강점과 약점 분석
- 2) 금융, 의료, 관광, 교통, 교육 등 주요 산업별로 데이터기반 비즈니스모델 프레임워크 분석을 통해 산업별 목표 모델 및 육성 전략 수립
- 3) 산업별 데이터 중개 및 공유 활성화를 위한 대표 기업 선정 및 중점 지원을 통해 민간 주도 데이터산업 육성
- 4) 산업간 데이터 융합 기업 발굴 및 육성

○ 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
DDBM 정의								
국내 DDBM 현황분석								
산업별 전문기업 육성								
융합 전문기업 육성								

다. (과제 2-2) 데이터 스타트업 발굴 및 지원

0 목표 : 우버와 에어비엔비 등과 같은 창의적 데이터 비즈니스모델 발굴 및 사업화 정착을 통해 데이터 기반 산업 창출

0 추진 방안

1) 데이터 분석 등 기술력 강화를 위한 산업계와 학교 및 연구기관 등의 협력 네트워크 구성

- 산업별 기능별 역할 모델을 수립하고, 협업 체계 마련

예) 기술개발은 연구소, 데이터 전처리는 학교, 데이터 활용은 산업계

2) 공공 부문에 ‘데이터랩(전담 조직)’ 신설

- 데이터 기반 비즈니스 모델 발굴을 위한 ‘데이터해커톤’ 개최

- 공공데이터 활용 모델 개발을 위한 ‘데이터베이션데이’ 개최

- 데이터 스타트업 인프라 지원과 멘토링을 위한 ‘데이터이노베이션랩’ 운영

* 참고: DB-Stars (데이터활용 창업 및 사업화 지원사업), 2015년 18개 팀 선정 및 지원을 통해 8개 기업 57억원 투자 유치 달성

3) 데이터비즈니스 기업 네트워크 구성으로 아이디어 교환, 기술 교류, 인력 공유, 서비스 융합 등 추진

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
산학연 네트워크 구성								
데이터랩 신설								
데이터랩 활동								
데이터비즈니스기업 네트워크 구성								

③ 데이터 유통 체계 마련

가. 현안

0 데이터 유통이 활성화되기 위해서는 데이터 유통서비스 확립, 데이터 기술 표준 정비, 데이터 매매 결제 수단 확보 등이 필요하나, 산업계를 아우르는 제도 미비

- 데이터 수요 특성에 따라 맞춤형 데이터 서비스 정립 필요. 예를 들면 분석용 DB, 앱개발용 API상품, 특정 상황에 따른 정보 알림 등
- 데이터 전달 및 활용 기술 표준화를 통해서 제공 기업과 사용 기업의 편의성이 확보되어야 함. 데이터 매매 거래의 조건과 지급 결제 방법 등이 정의되어야 함

나. (과제 3-1) 데이터 유통 서비스 확립

- 목표 : 데이터 유통 서비스 패턴을 정의하고 거래 방식 제시 등을 통해 데이터 유통을 위한 기본 원칙 수립
- 추진 방안
 - 1) ‘과제 2-1 데이터중개 비즈니스모델’에서 정의한 데이터기반 비즈니스모델 프레임워크(Data Driven Business Model Framework)을 토대로 데이터 유통 서비스 패턴 정의
 - 2) 데이터 유통서비스 패턴 별로 국내외 사례 조사를 통해 수익모델 등 거래 방식 제시
 - 3) 신설 데이터기업의 비즈니스모델 진단 및 개선 방안 수립에 활용

○ 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
DDBM 정의								
데이터서비스패턴 정의								
수익모델 제시								
데이터기업 진단 활용								

다. (과제 3-2) 데이터 유통 기술표준 정의

- 목표 : 데이터 유통 표준을 정의함으로써, 데이터 수요를 확대하고 데이터 활용 기회 및 편의성 제고
- 추진 방안
 - 1) 산업별 API 플랫폼 확대, API 활용 가이드 제공
 - 대표 산업별로 API 플랫폼 선정
 - API플랫폼을 중심으로 산업별 데이터 연계 API 개발 및 공유
 - API플랫폼 기반 데이터 활용 촉진

2) 오픈소스 기반의 데이터 공유 표준 검토

- 데이터 게재, 공유, 조사, 사용을 위한 오픈 소스 표준화 방안 조사
- 국제 사실 표준 기반의 개방형 플랫폼(CKAN)으로 전환

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
산업별 API플랫폼개발								
API플랫폼 기반 공유								
오픈소스 표준 검토								
CKAN 전환								

라. (과제 3-3) 데이터스토어 결제 기능 추가

- 0 목표 : 데이터진흥원이 운영하는 데이터스토어는 국내 유일한 민간부문 데이터 중개 기능을 수행하고 있으나, 결제 기능이 미비하여 소개 수준에 그치고 있음. 향후 데이터 유통을 확산하기 위해서는 결제 기능 확보 필수. 장기적으로 공공과 민간의 다양한 부문에서 데이터브로커 기업 성장 촉발 기능 수행

0 추진 방안

- 1) 데이터스토어에 결제 기능 추가. 데이터 유통 가격의 다양한 패턴 반영 필요
- 2) 데이터스토어는 공공과 민간 데이터 유통 플랫폼으로 성장 달성
 - * 참고: 한국데이터진흥원이 운영하는 '데이터스토어'의 데이터이용신청(2012년~2015년) 추이를 살펴보면 연평균 74.3%가 증가하는 추세
- 3) 데이터스토어 경험과 노하우를 기반으로 민간 부문의 다양한 데이터브로커 및 데이터유통플랫폼 발굴 및 육성 지원

* 참고

- 데이터브로커(Data Broker) : 자신이 직접 수집하거나 또는 제3자로부터 데이터를 구매하여, 데이터서비스 형태로 가공하여 제3자에게 판매하는 기업. 예) 블룸버그, 지그나이트 등. 데이터브로커는 선진국에서는 활발하게 비즈니스가 이루어지고 있음
- 데이터플랫폼(Data Platform) : 다수의 데이터 생산자와 다수의 데이터 소비자가 마켓에 모여서 수요와 공급을 매칭시켜주는 역할을 하는 기업. 데이터 가격은 생산자가 정하는 것이 일반적. 경매 또는 역경매 방식도 가능함. 데이터브로커가

활성화되면 이후에 데이터플랫폼이 등장할 것으로 전망됨. 아직은 데이터플랫폼 비즈니스모델이 활성화되어 있지는 않음. 우리나라는 데이터브로커가 등장하기에는 제도적으로 제약이 있어서, 현재는 공공 부문이 주도하는 것이 바람직함

- **데이터거래소(Data Exchange)** : 다수의 데이터 수요자와 데이터 공급자가 모여서 데이터를 거래하도록 하는 기업. 데이터 가격은 매수자가 매수 가격을 내고 매도자가 매도 가격을 호가하여 매도가격과 매수가격이 교차하는 점에서 매매가 성립되는 방법. 거래소는 보통 회원제로 운영되며 매도자와 매수자는 회원사를 통하여 거래해야 한다는 제한을 두고 있음. 거래소는 주식, 채권 등과 같이 동질의 자산을 다수가 보유하고 있는 경우에 성립할 수 있는 비즈니스 모델. 데이터의 경우 수요와 공급 특성에 따라 상태가 다양하게 제공될 수 있기 때문에 거래소 모델은 적합하지 않음

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
데이터스토어 결제 기능								
데이터플랫폼 성장								
민간 플랫폼 육성								

4 개인 데이터 활용 추진

가. 현안

- o 데이터 시대는 빅데이터(Big Data), 오픈데이터(Open Data), 개인데이터(Personal Data) 등을 3대 축으로 하고 있음. 빅데이터는 센서데이터, 소셜데이터, 로그데이터 등 주로 민간 기업에서 발생하는 데이터로서 기업 프로세스 혁신을 위해 주로 사용되고 있음. 오픈데이터는 GPS, 기상, 관광, 정책 등 주로 공공 부문에서 발생하는 데이터로서, 새로운 서비스기업 창출, 민간 기업 프로세스 혁신, 개인 소비자 편의성 제고 등을 위해 사용되고 있음. 개인데이터는 개인과 관련하여 생성된 데이터로서 현재는 기업 및 기관에 흩어져 있으나 이를 개인 중심으로 통합할 경우 새로운 성장 동력으로서 활용가치가 매우 큰 분야임
- o 우리나라는 오픈데이터는 데이터공개 차원에서 OECD 국가 중에서 1위를 달성하고 있는 등 개방 측면에서는 앞서가고 있음. 다만, 데이터 활용 측면에서는 제도 보완 및 비즈니스 창출 노력이 뒷받침되어야 함. 빅데이터는 개인정보보호의 과도한 규제에 의한 제약이 있으나,

기업 차원에서 대규모 투자가 진행되고 있고 일부 성과도 달성하고 있음. 개인데이터 부문은 선진국은 2011년 무렵부터 제도 준비, 개인 데이터 확산, 비즈니스 모델 발굴 등의 노력이 진행되었지만, 우리나라는 상대적으로 매우 미흡한 상태임

- o 개인데이터를 활용하게 될 경우 크게 세가지 효익이 있음. 첫째, 개인 데이터를 활용한 다양한 비즈니스모델이 등장함으로써 일자리 창출 등 신성장 동력이 됨. 둘째, 개인이 자신의 데이터를 통합하여 분석하고 서비스를 받을 수 있게 됨으로써, 개인의 금융활동과 소비활동, 여가활동 등에서 편의성의 획기적 향상과 경제적 개선을 기대할 수 있음. 셋째, 기업들은 과거 개인데이터를 통합하여 분석하고 이를 토대로 최적의 서비스를 제공함으로써 기업 프로세스 효율성과 효과성을 획기적으로 혁신할 수 있음
- o 우리나라는 금융위원회가 주관하여 은행과 증권 등 금융회사가 보유하고 있는 개인 계좌 데이터를 개인이 이용하는 앱으로 제공하는 오픈API플랫폼을 2016년에 가동하였음. 데이터 공개 폭은 금융회사들이 협의에 의하여 정하고 있으며, API사용에 따른 수수료를 부담해야 하는 등 개인데이터가 충분하게 공개되고 있지는 않음
- o 개인데이터 기반의 서비스 개발 및 스타트업 육성은 선진국 대비 5년 이상 뒤져 있는 수준으로서, 향후 집중적인 투자가 필요한 것으로 판단됨

나. (과제 4-1) 개인데이터 국가 전략 수립

- o **목표** : 개인데이터를 정의하고 필요성과 제약사항을 공유함으로써, 국민적 공감대를 형성하고 향후 개인데이터 제도화 및 사업화 기반 준비
- o **추진 방안** : 프로젝트 형태로 수행하는 것이 바람직함. 연구소 또는 전담기관에 의뢰하거나 외부 컨설팅회사에 용역 발주 형태 검토
 - 1) 개인데이터 선진국 추진 현황 및 제도 조사
 - 2) 개인데이터 이용, 거래, 제공 현황, 관련 법규 등에 대한 실태 조사
 - 3) 개인데이터 활용을 위한 국민 호응도 및 산업별 모델과 효과 분석. 산업별 잠재적 소비자 수요 파악. 우선 적용 분야 선정
 - 4) 개인데이터 활용을 위해 필요한 제도 연구
 - 5) 개인데이터 활용 과제 도출 및 로드맵 설정

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
프로젝트 기획								
프로젝트 수행								
전략과 로드맵 공유								
국가정책 반영								
제도 준비								
개인데이터 시행								

다. (과제 4-2) 개인데이터 제도화

- 0 목표 : 개인데이터 보유 실체가 개인에게 데이터를 제공하는데 관련된 의무와 기준을 명확하게 함으로써, 개인데이터 활용 기반 조성. 개인데이터 보유 기업 및 기관의 역할과 책임 규정, 개인 데이터 제공 방법 정의, 활용 조건 및 기준 등을 정의함으로써 개인데이터 활용 기반 구축

0 추진 방안

- 1) ‘과제 4-1 개인데이터 국가전략 수립’에서 정의한 제도정비 방안을 토대로 추진
- 2) 산업별 우선 순위를 선정하고, 이에 따라 순차적으로 적용 확산
- 3) 시범 프로젝트를 통해 적용의 문제점과 효과를 평가하고 제도화에 반영

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
산업별 영향 평가								
시범 프로젝트 수행								
제도 준비								
개인데이터 시행								

라. (과제 4-3) 개인데이터 비즈니스모델 발굴

0 목표 : 개인데이터를 활용한 비즈니스모델은 지금까지는 존재하지 않은 새로운 영역으로 창의적인 비즈니스 모델 발굴을 통해, 개인데이터 활용 성과 최대화

0 추진 방안

- 1) ‘과제 4-2 개인데이터제도화’ 이후 비즈니스모델 발굴 수행
- 2) 실제 비즈니스 모델 발굴은 ‘과제 2-2 데이터 스타트업 발굴 및 지원’을 통해서 수행

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
제도 준비								
비즈니스모델 발굴								

⑤ 데이터 인력 공급 부족 해소

가. 현안

- 0 우리나라 데이터 전문인력 수요 대비 공급 현황은 향후 연간 약 1,700 ~ 2,000 명 정도 데이터 전문인력이 부족할 것으로 예상
- 0 빅데이터 및 인공지능 등 데이터 활용 고도화를 위해서는 고급 데이터사이언티스트가 필요하지만, 민간 기업들은 기존 데이터 분석 인력의 수준을 뛰어 넘는 글로벌 경쟁력을 갖춘 데이터 전문인력 확보에 어려움을 겪고 있음
- 0 데이터전문가들의 수요가 공급을 초과하면서, 데이터활용이 필요한 중견기업들은 자체적으로 데이터전문인력을 확보하지 못함으로 인해 데이터 활용 기회를 살리지 못하고 있음. 또한, 필요한 데이터 인력의 전문성이 다양한데 비하여 직무별 데이터 전문가 확보가 쉽지 않음
- 0 민간 기업 및 학교 등 데이터전문인력 양성을 전담하고 있는 교육기관은 데이터교육과정에 필요한 실습용 데이터세트를 각자 구축하는 노력을 보이고 있으며, 실제 일정한 수준이 갖추어진 데이터세트 제공 기반을 필요로 하고 있음

나. (과제 5-1) 데이터 전문인력 교육과정 지원

0 목표 : 현재, 정부, 지자체, 대학(원), 민간 데이터 전문인력 양성과정의 정원 수 및 프로그램 수를 확대하고, 교육 과정 운영에 필요한 데이터세트 지원 등을 통해 교육 과정 품질 제고

0 추진 방안

- 1) 데이터전문가 프로파일 표준 정의. 전문가 프로파일 별 교육 과정 매칭을 통한 양성 대상 정원 최적화
- 2) 프로파일 별 표준 교육과정 커리큘럼 개발 및 공유. 데이터전문가 커리어맵 개발 보급 추진
- 3) 산업계와 연계하여 데이터전문인력 양성과정에 현장 실습 기회 제공
 - * 참고: 미래부는 인턴십 및 산학연계 프로젝트 등과 같은 과정을 통한 전문인력 수준 강화 체계 수행 중. 2014년 35개 대학 지원, 2017년 70개 대학으로 확대
- 4) 산업별 데이터교육을 위한 실습용 데이터세트 개발 및 공유
- 5) 산업계 데이터 우수기업의 교육 참여 인센티브 제공

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
프로파일 정의								
커리큘럼 정의								
데이터세트 개발 지원								
산업계 연계 지원								

다. (과제 5-2) 데이터 전문가 지원센터 운영

0 목표 : 중견 및 중소기업의 데이터 활용 수요 충족

0 추진 방안

- 1) ‘데이터활용지원센터’ 설치
- 2) 데이터전문가 및 전문기업 풀 구축
- 3) 중견 및 중소기업 데이터 활용 지원 프로그램 개발 및 예산 확보
- 4) 중견 및 중소기업 데이터 활용 지원 프로그램 운영

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
지원센터 설치								
전문가 풀 구축								
프로그램 개발								
지원 가동								

라. (과제 5-3) 데이터 전문가 육성 체계 고도화

0 목표 : 고급 데이터 전문가 육성 및 기존 데이터 인력 고도화

0 추진 방안

1) 산업별 데이터전문가 육성 프로그램 개발 및 확대

* 참고: 한국데이터진흥원은 ‘융합과정’을 통해 산업에 특화된 데이터전문가 육성 추진 중임

2) 산업별 교육용 콘텐츠 개발 및 배포, 평가 지표 마련

3) 기업 자체 데이터전문가 교육과정 고도화 및 지속적 업그레이드 지원

4) 데이터전문가 자격증 제도 고도화. 데이터사이언티스트 직무 세분화 추세를 반영하여 데이터전문가 유형을 다양화하고 이에 맞는 자격증 제도로 발전

5) 데이터전문가 네트워크 구성을 통해 자체적으로 지식 공유 및 협업을 통한 문제 해결 능력 개발 추진

0 추진 일정

추진 업무	2017				2018		2019	
	1/4	2/4	3/4	4/4	상	하	상	하
산업별 육성프로그램								
기업 교육과정 지원								
자격증 제도 고도화								
전문가 네트워크 구성								

IV. 추진일정

세부 추진과제		'17				'18		'19	
		1/4	2/4	3/4	4/4	상	하	상	하
1. 데이터 원칙 수립 및 제도 확립									
	1-1 데이터위원회 구성								
	준비위원회 구성								
	데이터위원회 발족								
	운영계획 수립								
	데이터위원회 활동								
	1-2 데이터 원칙 제정								
	데이터위원회 발족(1-1)								
	데이터원칙 수립								
	데이터원칙 적용/보완								
	1-3 데이터 규제 패러다임 전환								
	데이터연구활동 수행								
	규제프리존 운영								
	데이터원칙 수립(1-2)								
	법규 및 제도 적용								
2. 데이터산업 생태계 육성									
	2-1 데이터 중개 비즈니스 모델 육성								
	DDBM 정의								
	국내 DDBM 현황분석								
	산업별 전문기업 육성								
	융합 전문기업 육성								
2-2 데이터 스타트업 발굴 및 지원									

세부 추진 과제			'17				'18		'19	
			1/4	2/4	3/4	4/4	상	하	상	하
		산학연 네트워크 구성								
		데이터랩 신설								
		데이터랩 활동								
		데이터비즈니스기업 네트워크 구성								
3. 데이터 유통 체계 마련										
	3-1 데이터 유통 서비스 확립									
		DDBM 정의(2-1)								
		데이터서비스패턴 정의								
		수익모델 제시								
		데이터기업 진단 활용								
	3-2 데이터 유통 기술표준 정의									
		산업별 API플랫폼개발								
		API플랫폼 기반 공유								
		오픈소스 표준 검토								
		CKAN 전환								
	3-3 데이터스토어 결제 기능 추가									
		데이터스토어 결제 기능								
		데이터플랫폼 성장								
		민간 플랫폼 육성								
4. 개인 데이터 활용 추진										
	4-1 개인데이터 국가 전략 수립									
		프로젝트 기획								
		프로젝트 수행								
		전략과 로드맵 공유								

세부 추진 과제			'17				'18		'19	
			1/4	2/4	3/4	4/4	상	하	상	하
		국가정책 반영								
		제 도 준비								
		개 인데이 터 시 행								
	4-2 개인데이터 제도화									
		산업별 영향 평가								
		시범 프로젝트 수행								
		제 도 준비(4-1)								
		개 인데이 터 시 행(4-1)								
	4-3 개인데이터 비즈니스모델 발굴									
		제 도 준비(4-1)								
		비즈니스모델 발굴								
5. 데이터 인력 공급 부족 해소										
	5-1 데이터전문인력교육과정지원									
		프로파일 정의								
		커리큘럼 정의								
		데이터세트 개발 지원								
		산업계 연계 지원								
	5-2 데이터 전문가 지원센터 운영									
		지원센터 설치								
		전문가 풀 구축								
		프로그램 개발								
		지원 가동								
	5-3 데이터 전문가 육성 체계 고도화									
	산업별 육성프로그램									

세부 추진 과제			'17				'18		'19	
			1/4	2/4	3/4	4/4	상	하	상	하
		기업 교육과정 지원								
		자격증 제도 고도화								
		전문가 네트워크 구성								

※ 참고문헌

- 1) 세계경제포럼(World Economic Forum, WEF), <http://www.weforum.org>
- 2) Analysys Mason(2014), *Data-Driven Innovation in Singapore*
- 3) 영국정부(2012), Public data - statement of principles,
<https://data.blog.gov.uk/2012/10/04/public-data-statement-of-principles>, 검색일 2016. 11. 21.
- 4) 영국정부(2016), Data Science Ethical Framework,
<https://www.gov.uk/government/publications/data-science-ethical-framework>, 검색일 2016. 11. 22.
- 5) 영국 Direct Marketing Association(2014), Data Guide, <http://dma.org.uk/guide/data-guide>, 검색일 2016. 11. 23.
- 6) 미국 Data & Marketing Association(2017), PROVIDING CONSUMER CHOICE & THE COLLECTION, USE, AND TRANSFER OF PERSONALLY IDENTIFIABLE DATA,
<https://thedma.org/accountability/ethics-and-compliance/dma-ethical-guidelines/collection-use-and-maintenance-of-marketing-data/#providing-consumer-choice>, 검색일 2017. 1. 5.
- 7) 썬라이트재단(2010), <https://sunlightfoundation.com/policy/documents/ten-open-data-principles>, 검색일 2016. 11. 23.
- 8) 국가법령정보센터(2016), 공공데이터의 제공 및 이용 활성화에 관한 법률,
<http://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=142444#0000>, 검색일 2016. 12. 11
- 9) 오픈데이터헌장, <http://opendatacharter.net/>
- 10) New Zealand(2015), *Data Driven Innovation in New Zealand*
- 11) Cambridge Service Alliance(2014), ‘Big Data for Big Business?, A Taxonomy of Data-driven Business Models used by Star-up Firms’, University of Cambridge
- 12) 주석 11 참조
- 13) 한국데이터진흥원(2016), 주요국 데이터 주도 비즈니스 현황 및 비즈니스 유형 비교·분석
- 14) 데모데이, <http://www.demoday.co.kr>
- 15) 법률신문(2016), 웹사이트 무단 크롤링’ 소송… 잡코리아, 사람인에 승소, 검색일 2016. 11. 24.
- 16) 핀란드 정부부처(2014), *MyData : a Nordic Model for human-centered personal data management & processing*
- 17) 영국정부(2013), Better choices, better deals,
<https://www.gov.uk/government/news/better-choices-better-deals>, 검색일 2016. 12. 21.
- 18) FING(2012), *MesInfos: Experimenting the sharing of personal data between business and consumers*, Retrieved Dec 22, 2016, from http://doc.openfing.org/MesInfos/MesInfos_Presentation_V5EN.pdf
- 19) FING(2013), *MesInfos Explorer’s notebook*
- 20) FING(2015), *The MesInfos Pilot Study*
- 21) FING(2016), <http://www.fing.org/?MesInfos-un-monde-de-Self-Data&lang=fr>, 검색일 2016. 12. 22
- 22) FING(2016), What is the value of Self Data for organizations?, Retrieved Dec 22, 2016, from http://mesinfos.fing.org/wp-content/uploads/2015/05/MesInfos_SelfData_business_models_English.pdf
- 23) Executive Office of the President National Science and Technology Council(2013), *Smart Disclosure and consumer decision making: Report of the task force on Smart Disclosure*
- 24) The White House(2016), *FACT SHEET: Obama Administration Announces Federal and Private Sector Actions on Scaling Renewable Energy and Storage with Smart Markets*, Office of the Press Secretary
- 25) Glynn D. Ligon, Ph.D(2016), *The Optimal Reference Guide: MyData Button: Button, Button, Who’s God the Button?, ESP Solution*

- 26) OpenNotes, <http://www.opennotes.org>
- 27) Epic, <http://www.epic.com>
- 28) 현대경제연구원(2016), '데이터, 빅 이코노미 실현', 16-28
- 29) SAS(2013), Big data analytics: An assessment of demand for labour and skills, 2013-2020
- 30) McKinsey Global Institute(2011), Big data: The next frontier for innovation, competition, and productivity
- 31) DeZyre(2016), Data Scientist Salary Report of 100 Top Tech Companies, Retrieved Dec 1 2016, from <https://www.dezyre.com/article/data-scientist-salary-report-of-100-top-tech-companies-/218>
- 32) 한국과학기술기획평가원(2016), 과학기술&ICT 정책기술동향 73호
- 33) 한국인터넷자율정책기구(2016), '빅데이터 관련 개인정보 비식별화 정부 가이드라인 수립', KISO저널 제24호
- 34) CKAN, <http://ckan.org/>

1. 본 보고서는 미래창조과학부의 출연금으로 수행한 DB산업 육성 사업의 결과입니다.
2. 본 보고서의 내용을 발표할 때에는 반드시 미래창조과학부 DB산업 육성 사업의 결과임을 밝혀야 합니다.
3. 본 보고서 내용과 관련한 문의는 한국데이터진흥원 정책기획실 (02-3708-5364)로 연락해 주시기 바랍니다.