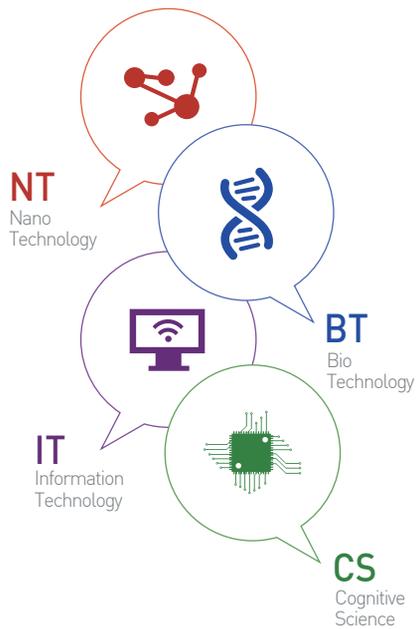

Issue Paper



[융합] 빅데이터 기술 현황 및 시사점

※ 출처 : 융합연구정책센터

□ 선정 사유

- (경제적 측면) 다수의 글로벌 전문기관에 따르면, 빅데이터는 20세기 산업 혁명에 이어 향후 우리 경제사회의 핵심으로 부각될 것으로 전망

- Economist : 데이터를 자본이나 노동력과 비슷한 경제적 투입자본으로 높이 평가하고 있으며, 비즈니스의 새로운 원자재 역할을 할 것으로 전망
- Gartner Group : 데이터는 '21세기 원유'이며, 데이터가 미래 경쟁우위를 좌우 하는데 핵심요소가 될 것으로 전망

- (사회적 측면) 최근 고령화, 건강문제 뿐만 아니라 국가적 재난상황이 급증하고 있어, 이를 극복하기 위해 빅데이터를 활용한 융합기술의 필요성 대두

* 소셜, 모니터링 등의 빅데이터 패턴분석을 통해 국가재난 상황에서 위험징후, 이상 신호 등을 빠르게 포착하여 신속한 사고접수 및 의사결정에 활용

〈 세월호 사건에서 살펴본 빅데이터 활용가능 기술 〉

- 텍스트분석기술 : 카카오톡 등 SNS를 통해 '배가 기울고 있다', '구명조끼를 입었다' 등 사고관련 메시지 자동감지 ⇒ 조기에 사고접수 가능

- 출처 : 매일경제, "한국, 재난 빅데이터가 없다", 2014. 4. 29 -

□ 개요

- (기술개념) 기존의 관리 및 분석 체계로 감당하기 어려운 대규모 정형 및 비정형 데이터를 효과적으로 저장·관리하고, 새로운 가치 창출을 위해 초고속으로 결과를 분석하여 서비스를 제공하는 기술(미래부, '2012년도 기술영향평가 결과(안)', 2013)

* 빅데이터 기술은 McKinsey, SERI, 국가정보화전략위원회 등 기관별로 다양하게 정의

- (핵심기술) 빅데이터 시장 및 기술은 크게 ① 인프라, ② 소프트웨어, ③ 응용서비스로 구분(IDC, 'Big Data technology and services forecast', 2012)

- ① 인프라(HW) : 스토리지, 서버, 네트워크 관련 기술(서버 및 스토리지 지원)

- ② 소프트웨어(SW) : 데이터 수집, 분석·예측, 데이터 처리 자동화 기술

- ③ 응용서비스 : HW와 SW 관리 및 컨설팅·교육 등 응용서비스 기술

□ 국내·외 정책동향

- (국외) '12년을 전후로 국가차원의 빅데이터 전략 발표, 공공데이터 개방은 물론 이를 정부가 직접 활용하여 서비스 촉진 및 정부서비스 혁신 도모
 - (미국) 빅데이터 활용 기술개발 및 인력양성을 위해 빅데이터 R&D 계획(Big Data R&D Initiative, '12년.3월) 발표, 2억 달러 예산으로 84개 프로그램 지원
 - (EU) 공공데이터 활용을 위해 데이터개방전략(Open Data Strategy, '11년.5월) 발표, 3년간('11~'13년) 1억 유로 예산으로 관련 R&D 사업 지원
 - (일본) 액티브 재팬(Active Japan ICT, '12년.5월) 발표, 주요전략으로 빅데이터 활용정책(Active Data) 포함, 89억3천 엔 예산('13년 기준)으로 R&D 지원
 - (국내) 그간 우리나라는 빅데이터 마스터플랜('12년.11월), 빅데이터 산업 발전 전략('13년.12월) 등 종합 육성정책 발표, '17년까지 97개 사업 추진('14년.1월)
 - 안행부·미래부는 국민생활·안전, 일자리 창출 및 국정과제 관련 23개 사업을 중점지원 과제로 선정, 우선적으로 예산을 지원할 계획
 - '14년에는 25개 신규 사업을 추진하고, 전년도에 완료된 '위기 청소년 조기 경보'(여가부), '일일 물가분석'(통계청) 등 주요 과제의 서비스 제공 본격화
- * 정부예산 : '13년 360억(중앙 345억, 지자체 15억) ⇒ '14년 460억(중앙400억, 지방60억)

<빅데이터 관련 국내 추진과제>

구 분	계	'13년	'14년	'15년	'16년	'17년
합 계 (과제수)	97	25	25	26	15	6
중앙부처	61	19	15	12	12	3
지자체	36	6	10	14	3	3
(중점과제)*	(23)	(8)	(4)	(4)	(3)	(4)

※ 원 출처 : 부처합동, '빅데이터 활용 본격화, 복지·치안·물가 등 해결사', 2014. 1

* 중점과제는 국가전략, 파급효과, 실현가능성 등을 고려하여 중요도가 높은 과제로 구성

<시사점>

- ① 미국 등 주요국은 공공기관의 다양한 데이터를 개방하고 공유하는 정책을 추진함으로써 빅데이터의 제도적·기술적 기반 마련
- ② 우리나라 정부는 기술연구개발, 법·제도 정비 등 공공 및 민간 활성화 방안을 종합적으로 제시, 미성숙한 시장의 진흥까지 고려

□ 국내·외 R&D 지원 현황

- (미국) 국립과학재단(NSF), 국립보건원(NIH), 국방부(DoD) 등 6개 정부기관을 중심으로 보건, 국방, 에너지 등의 분야에서 빅데이터 활용 과제 진행
 - 국립과학재단(NSF)은 다량의 빅데이터를 효율적으로 분석·활용하기 위해 빅데이터 전문인력 양성, 클라우드 개발 등과 관련된 장기 R&D전략 수립
 - 국립보건원(NIH)은 국민건강 향상을 위해 세포, 분자, 의학, 질병 등 관련 데이터 개발 및 관리에 중점적으로 투자
 - 국방부(DoD)는 암호해독, 지형분석 등 정보戰의 기본이 되는 정보수집기술 외 자율 의사결정 시스템 등 최신 빅데이터 기술 개발 및 활용에 투자 강화
 - * 특히, 국방부 연구·개발 부문을 담당하는 방위고등연구계획국(DARPA)은 'XDATA', 'ADAMS' 등 최첨단 데이터분석 컴퓨팅 및 SW 개발 관련 R&D 기획 및 투자 진행
 - 그밖에 에너지부(DoE), 미국지질조사원(USGS)이 각각 에너지, 지질조사에 빅데이터를 활용하고 있으며, 최근 다른 부처·기관도 추가로 사업 추진
 - * 미국항공우주국(NASA)은 첨단정보시스템기술(Advanced Information System Technology), 지구과학데이터 및 정보시스템(Earth Science Data and Information System) 등 R&D 사업 추진
 - * 미국해양대기관리처(NOAA)와 국립기상청(NWS)은 대기, 해양 등 해당분야에서 1950년대부터 데이터를 축적해온 기술력을 바탕으로 빅데이터 R&D 계획에 참여

<미국 빅데이터 주요 지원과제 현황>

연방 부처 및 기관 (영문명)	주요 지원과제	정부연구비 (달러)
국립과학재단 (NSF)	컴퓨팅 탐험대 프로젝트(Expeditions in Computing)	1,000만
	학부생 훈련 프로그램: 복잡한 데이터 도표화 및 시각화 연구	200만
	생물학 연구그룹의 단백질 구조 및 생물학 경로 연구	140만
국립보건원 (NIH)	뇌신경 기능 및 활동 경로의 데이터 수집(Human Connectome)	-
	생리학 관련 데이터 스토리지 서비스 연구(PhysioBank)	
	심혈관 공동연구를 위한 데이터 저장·공유·분석 툴 개발(CVRG)	
국방부 (DoD)	자율 의사결정 시스템 구축 및 상황인식능력 개선 연구	2억 5,000만(매년) + 6,000만 (신규)
	국방 관련 빅데이터 기술연구 경진 대회	
	세계의 다양한 언어 및 문자로부터 정보를 추출·분석하는 연구	
방위고등연구 계획국 (DARPA)	대용량 데이터에서 특정 정보만을 탐지하는 기술 개발(ADAMS)	3,500만
	분산 데이터 처리 알고리즘 및 오픈소스 소프트웨어 연구(XDATA)	2,500만(매년)
	데이터 암호화 관련 프로그래밍 언어 개발(PROCEED)	-
에너지부 (DoE)	확장형 데이터 관리·분석 및 시각화 연구소(SDAVI) 설립	2,500만
	중성자 연구에서의 대용량데이터 수집 및 분석 시스템연구(ADARA)	-
미국지질조사원 (USGS)	지구시스템 과학 분야에서의 빅데이터 활용 연구	-

※ 원 출처 : US Government, 'Big Data R&D Initiative', 2012. 3

- (국내) 범부처 차원에서 버스노선, 질병예보, 뉴스검색 등 빅데이터 초기시장 구축 및 사회문제 해결을 위해 다양한 R&D 사업 추진
 - (각 부처) '13년 중점과제로 선정된 8개 과제를 추진하기 위해서 각 과제별 관련 중앙부처(안행부 등) 및 지자체(서울시 등)를 중심으로 사업 추진
 - (미래부) 2012년부터 2017년까지 349억 원을 투자하여 트위터 등 SNS 이용 활성화로 급증한 자연어 정보 및 이미지, 위치정보 등 센서데이터의 자료를 마케팅에 활용하는 기술 개발을 지원(3개 과제)

<2013년 국내 빅데이터 중점지원 과제 현황>

주관	과제명	분석 데이터	연구 내용
안행부	빅데이터 로그분석으로 보안사고 및 장애 예방 강화	2만여 정보시스템의 Log (6개월 동안 Log 데이터) 및 작업승인 내역	악성위협IP의 접근기록 탐지 및 활동흔적 추적 및 장애의 패턴을 인지하여 동일 패턴의 Log 발생 시 담당자 경보
복지부	국민건강정보 DB를 활용한 맞춤형 건강서비스 개발	국민건강공단의 질병정보 및 민간 SNS 자료 등	특정질병에 대한 위험을 예보하고, 개인 건강기록 및 건강나이, 외종증 위험도 예측 등 서비스 제공
고용부	일자리 현황분석을 통한 고용 수급 예측	경제활동인구조사, 국민계정, 고용노동통계조사, 패널조사 데이터, 고용보험 등	정형적·비정형 데이터 융합을 통해 미래 주요산업(직종) 트렌드, 이슈 등 발굴
통계청	빅데이터를 활용한 국민체감 통계생산	온오프라인 병행 사이트 등으로부터 수집한 가격 정보 및 비정형 데이터	일일물가지수, 감성지수(물가변동에 대한 국민정소 영향지수) 추세분석 및 물가예측
교육부	빅데이터 활용 스마트 뉴스 모바일 앱 개발	대량의 기사DB에 대한 중요도, 관계도 등	뉴스기사 시간별 매체별 요약, 의미관계 시각화, 전문가 식별 등을 통해 신속한 지능형 뉴스 검색 서비스 제공
식약처	빅데이터 기반의 의약품 안전성 조기경보 서비스	포털, SNS 상 의약품 부작용 정보, 병원 환자진료기록 정보 등	포털, SNS 상의 비정형화된 의약품 부작용 정보를 분석하여 의약품 안전성을 사전에 예측하고 소비자 등에 알림서비스 제공
중기청	소상공인 창업성공률 제고를 위한 점포평가 서비스	상가 인·허가, 부동산 가격 정보, 신용카드거래 데이터	점포 개·폐업 이력, 창업 아이템별 매출 추이 등을 분석하여 소상공인 성공률 제고 및 '묻지마 창업' 예방으로 과잉 진입 방지
서울시	유동인구 빅데이터를 활용한 심야버스 노선 수립	민간이통사(KT) 통화량 및 거주지데이터, 서울시교통 데이터 등	심야시간 KT통화량으로 유동인구 밀집도를 분석하여 심야버스 노선의 배차간격 등 교통노선의 검증·보완
미래부	차세대 메모리 기반의 빅데이터 분석·관리 소프트웨어 원천기술 개발		빅데이터 실시간 기본 플랫폼을 설계하고, 분산 DB 관리기술 프로토타입 개발
	초소형·고신뢰(99.999%) OS와 고성능 멀티코어 OS를 동시 실행하는 듀얼 운영체제 원천기술 개발		차세대 네트워크 장비 및 자동차 인포테인먼트 등 차세대 임베디드 운영을 위한 듀얼 운영체제 기술 개발
	빌딩내 기기들을 웹을 통해 연동하여 사용자 맞춤형 최적제어·모니터링 서비스를 제공하는 소프트웨어 개발		가상화 기반의 임베디드 제어기 등을 개발함으로써 입주자 위치 등에 따른 다양한 맞춤형 서비스 개발

※ 원 출처 : 부처합동, 「빅데이터 활용 본격화, 복지·치안·물가 등 '해결사'」, 2014. 1. 7; 국가과학기술지식정보서비스(NTIS)

□ 국내·외 기술수준분석

- (국외) 해외 주요국 가운데서는 미국의 기술수준이 가장 높고, 그 다음으로 일본(86.2%), EU(83.5%), 중국(65.6%) 순으로 나타남
 - 세부적으로는 데이터 분산처리 시스템기술(2.3년)이 지식기반 빅데이터 활용 기술(2.8년)에 비해 국가별 기술수준 격차가 크지 않았음(우리나라 포함)
- (국내) 우리나라 빅데이터 기술수준은 최고 기술 보유국인 미국 대비 78.0% 정도이며, 기술격차는 3.5년으로 나타남
 - 세부적으로는 데이터 분산처리 시스템기술(79.3%)이 지식기반 빅데이터 활용 기술(76.7%)에 비해 최고기술국 대비 기술수준이 상대적으로 높았음

<빅데이터 분야의 국·내 외 기술수준>

기술분야	기술수준 (최고 수준 국가 대비 00%)						기술격차 (최고 수준 국가 대비 00년)					
	미국	EU	일본	중국	한국	5개국 평균	미국	EU	일본	중국	한국	5개국 평균
지식기반 빅데이터 활용기술	100.0	84.6	88.1	61.6	76.7	82.2	0.0	24	1.5	6.2	4.0	2.8
데이터 분산처리 시스템기술	100.0	82.3	84.2	69.5	79.3	83.1	0.0	2.2	2.1	4.2	3.0	2.3
2개 세부분야 평균	100.0	83.5	86.2	65.6	78.0	82.7	0.0	2.3	1.8	5.2	3.5	2.6

※ 원 출처 : KISTEP, '2012년 기술수준평가: 120대 국가전략기술', 2013. 2

□ 시장 동향 및 기대 효과

- (적용분야) 빅데이터 기술은 IT산업에 국한되지 않고, 국민경제, 복지, 안전, 고령화 등 거의 모든 범위에 적용이 가능
 - (관련제품) 상관분석 서비스(SK텔레콤), 민원정보분석시스템(국민권익위원회) 등 민간과 공공 부문에서 이미 빅데이터를 활용한 제품이 다수 출시



<상관분석 서비스(SK텔레콤)>

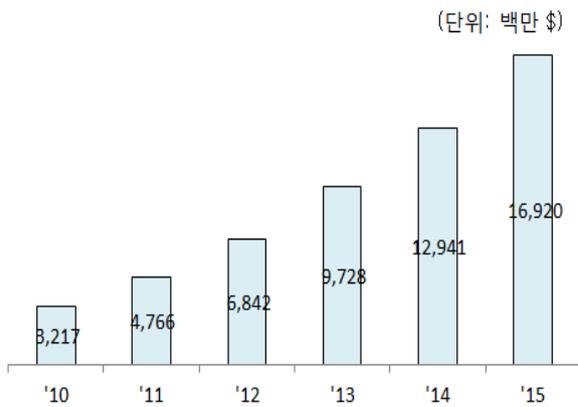


<민원정보분석시스템(국민권익위원회)>

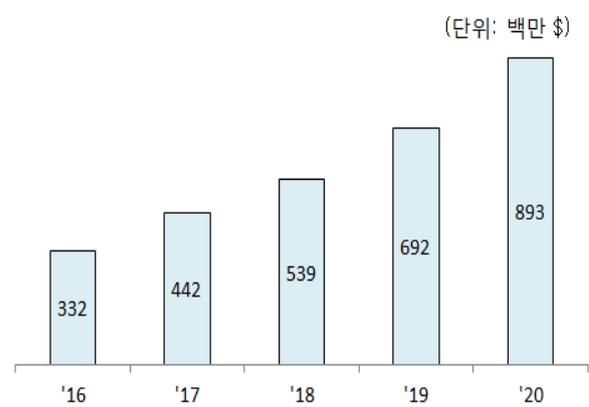
※ 원 출처 : 부처합동, '빅데이터 마스터플랜', 2012. 11

- **(시장규모)** 21세기 정보화 시대를 맞아 다양한 데이터가 폭발적으로 증가함에 따라 이러한 정보를 체계적으로 분석·활용하는 시장도 지속적으로 성장
 - **(세계)** 2010년 32억 1천7백만 달러에서 2015년 169억 2천만 달러로 성장하여 5년간 약 5.3배 성장할 것으로 전망
 - **(국내)** 국내 빅데이터 시장은 2016년 3억 3천2백만 달러에서 2020년 8억 9천3백만 달러로 성장하여 4년간 약 2.6배 성장(세계 대비 시장성장률 1/2수준)

<세계 빅데이터 시장규모 전망>



<국내 빅데이터 시장규모 전망>



※ 원 출처 : KISDI, '빅데이터 동향 및 정책 시사점', 2013. 6 재인용

- **(사회경제적 파급효과)** 공공 및 민간 부문에서 축적된 빅데이터를 활용하여 생산성 향상, 기술 경쟁력 확보, 정부 혁신 등 사회경제적 파급효과 달성
 - **(EU)** 빅데이터 활용으로 유럽 제조부문은 개발·조립비용의 50%, 운전자본의 7% 절감, 공공부문은 GDP의 0.5%의 생산성 증가 효과 기대
 - * 출처: McKinsey, 「Big Data: The next frontier for innovation, competition, and productivity」, 2012. 6
 - **(영국)** '12년부터 '17년까지 빅데이터가 영국 산업에 미치는 경제적 효과(사업 효율성 제고, 新산업 창출 등)가 2,160억 파운드(약 395조원)에 달할 것으로 전망
 - * 출처: CEBR, 「Data equity: Unlocking the value of big data」, 2012. 4
 - **(국내)** 국내 공공부문 빅데이터 활용으로 행정효율성 제고, 세수증대 등을 달성하여 최대 4조 2천억 원(GDP의 0.4%)의 부가가치 창출
 - * 출처: 부처합동, '빅데이터 마스터플랜', 2012. 11

□ 결론 및 정책적 시사점

- **(인식전환)** 빅데이터는 고부가가치를 창출하는 ‘황금알을 낳는 거위’로 주목받고 있으나, 개인정보 유출 및 낮은 인지도에 대한 문제점 내포
 - 대형 개인정보 유출 사고로 인해 정보공개에 대해 국민이 거부반응 표출
 - * (대표사례) 3개 신용카드(농협, 국민, 롯데)는 1억 400만 건의 개인정보 유출(‘14.1월)
 - 이를 위해 금융기관 책임강화, 불필요한 개인정보 삭제 등 대응체계 마련
 - 또, 정부기관을 포함한 민간기업의 빅데이터에 대한 낮은 인지도*를 개선하기 위해 공익광고 등을 통해 빅데이터 활용에 대한 국민적 공감대 형성 필요
 - * 240개 기업·공공기관을 대상으로 설문조사한 결과, 빅데이터를 전혀 모르거나 이름만 들어봤다는 응답이 76.7% 차지(한국정보통신진흥협회, ‘빅데이터 산업 실태조사’, 2012년)
- **(정보개방)** 행정·공공기관은 2,194개 웹사이트를 통해 다양한 데이터를 공개하고 있으나, 민간부문의 데이터 공개는 미흡
 - * (예시) 실시간 버스정보, 공공취업 정보 등 국민생활과 밀접한 33개 분야의 데이터를 공유자원포털(data.go.kr)을 통해 민간에 개방하고 있음
 - 기업이 보유한 대부분의 민간데이터는 기관별 자체 활용으로 국한된 용도로 한정되어 사실상 외부기관의 접근이 용이하지 않는 형태임
 - 정보공개에 상응하는 인센티브를 제공함으로써 필수정보 외에 제한적 범위로 민간 데이터의 외부 공개를 촉진할 수 있는 방안모색이 필요
- **(기술연구개발)** IBM, 오라클 등 글로벌 기업이 이미 빅데이터 초기 시장을 선점하여 우리나라는 해외 우수기술을 차용하여 단순히 활용하는 수준임
 - ※ SK텔레콤은 오라클의 ‘엑사데이터(Exadata)’를 도입하여 요금 분석 및 검증에 활용
 - 그간 서버 등 인프라(HW) 분야는 어느 정도의 경험 및 역량이 갖춰졌으나, 데이터 분석 등 소프트웨어(SW) 및 응용서비스 분야는 여전히 부족한 실정
 - 이를 위해, 구글 ‘빅쿼리(Big Query)’ 등과 같은 데이터 저장·분석 솔루션은 물론, 빅데이터 미래가치 창출을 위해 ‘인공지능(AI)’에 활용하는 기술개발 지원 필요
 - ※ 빅데이터를 가장 잘 활용하고 있는 구글사의 임직원의 상당수가 AI전문가로, 특히 데이터마이닝, 자연어처리, 자동추론 등 빅데이터 기술 대부분은 AI에 기초하고 있음
 - 또한, 활용을 촉진하기 위해 클라우드 서비스*에 대한 지원도 계속 강화 필요
 - * 클라우드 서비스는 영화, 음악 등 사용자 콘텐츠를 서버에 저장하여 스마트폰 등 외부기기에서 실시간 이용하는 기술로, 빅데이터의 他 분야 활용에 크게 기여