



청소년 빅데이터 체계 구축 및 활용방안 연구

서 정 아 선임연구위원

요약¹⁾

- 본 연구에서는 청소년 대상 빅데이터 체계 구축의 가능성을 파악하고 이를 바탕으로 제언점을 도출함으로써 청소년 빅데이터 체계 구축을 위한 기초자료를 제시함. 또한, 정형 빅데이터와 비정형 소셜 빅데이터를 활용한 분석을 시도함으로써 청소년 관련 이슈에 대해 어떻게 예측이 가능한지를 확인해 보고 청소년 빅데이터 체계가 추후 청소년 관련 미래변화나 정책수요를 예측하고 위험요소를 조기 발견·예방하는 데 어떻게 활용될 수 있는지 등에 관한 방안을 도출·제시함.
- 연구방법으로 국내·외 문헌 고찰, 빅데이터 분석방법론에 관한 콜로키움, 외국의 빅데이터 체계 구축 및 활용 사례 검토, 청소년 핵심 온톨로지 개발, 빅데이터 체계 구축, 활용 방안 등에 대한 전문가 자문, 전문가 심층면접, 전문가 설문조사, 정책협의회 등을 활용함.
- 전문가 설문조사 결과 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항으로 개인정보 보호로 인한 접근의 제한성이 37.5%로 가장 많았음. 다음으로 데이터 형식 간의 통일성 부재로 인한 분석 상의 어려움, 데이터 품질 및 개선사항에 대한 의견 수렴 부족, 홍보 부족으로 인한 데이터 관련 정보 부재, 데이터 세부 사항, 범례 등에 대한 설명 부족이 각 12.5%, 데이터 항목의 누락이나 중첩 등 부정확성 존재, 데이터 간의 연계 가능성 희박, 필요한 데이터의 부재, 자료의 복잡성, 데이터 표집의 한계성(예: 편의 추출 등), 신뢰도 및 타당도에 대한 확인의 어려움, 분석과 정책, 서비스 간 연계 프로세스 부재가 각 6.3%였음. 연구결과를 바탕으로 청소년 빅데이터 체계 구축 방안 및 활용방안을 제시함.

1) 본고는 한국청소년정책연구원 2020년 고유과제 '청소년 빅데이터 체계 구축 및 활용방안 연구'를 발체·요약한 것임.

1. 연구의 필요성 및 목적

▶ 4차 산업혁명 시대에 국가 경쟁력의 원천이자 급격한 사회변화에 대응하고 주요 현안을 해결하기 위한 방안으로 빅데이터의 활용가치가 더욱 높아지고 있음.

- '데이터 경제(Data Economy)'의 확산으로 가능할 수 없을 정도로 방대하고 다양한 데이터가 빠르게 생산되고 있으며 이러한 빅데이터의 생산은 앞으로도 폭발적으로 증가할 것으로 예측됨.
- OECD와 EU 등 국제기구는 물론(문유경·전기택·김영택·정성미·배호중·정희태·김유정, 2018) 미국, 영국, 호주, 싱가포르 등 세계 각국은 국가 경쟁력의 핵심 기반으로 빅데이터의 중요성을 인식하고 있으며 국가 차원의 빅데이터 산업 육성 정책을 추진 중에 있음(이슈 & NDSL, 미국 연방정부, 빅데이터 연구개발 전략계획 공개, 2020.2.10. 검색).
- 증거기반 정책수립과 미래예견적 국정관리 측면에서도 빅데이터 구축 및 활용의 필요성이 강조되고 있음.

▶ 최근 정부는 데이터 산업 활성화 전략을 마련하고 2019년부터 빅데이터 플랫폼 및 센터 구축 사업을 추진 중에 있음(대통령직속4차산업혁명위원회, 데이터 산업 활성화 전략, 2020.2.10. 검색)

- 통계청과 행정안전부를 중심으로 공공데이터를 확보하고 개방해 나가는 추세에 있음. 과학기술정보통신부는 공공기관과 민간분야에서 분야별 빅데이터 플랫폼을 확대해 나가고 있음.
- 특히 청소년은 인터넷이나 모바일에 기반한 디지털 기기 사용이 일상화 되어 있고, 디지털 생태계에서 살아가는 청소년의 삶의 양식이 이전 세대와 다른 양상을 나타내는 만큼 청소년에 대한 정확하고 적시성 있는 정책 마련을 위해 특히 소셜 빅데이터를 활용한 청소년 연구가 필수적임.
- 사회 환경의 급속한 변화와 청소년 현안의 복잡성 증가에 따라 다양한 정형, 비정형 빅데이터에 기반한 청소년 정책수립과 미래예견적 의사결정의 중요성이 강조되고 있는 것임.

▶ 빅데이터의 중요성에 대한 전지구적 관심으로 다양한 빅데이터 관련 국·내외 연구(강희정, 2015; 송태민, 2013; 안명숙, 2018; 한윤선·김하영·송주영·송태민, 2019; Fields, Quirke, Amely & Maughan, 2016; Gharabaghi & Anderson-Nathe, 2014; Katal, Wazid & Goudar, 2013; Watson, & Christensen, 2017)들이 축적되고 있으며 빅데이터 분석방법을 소개하는 저서도 다수 나타나고 있고(노만 매트로프 저, 권정민 역, 2016; 벤 웨이버 저, 배충호 역, 2015; 후나오 노부오 저, 김성재 역, 2014) 빅데이터 활용을 위한 체계 마련 노력이 이루어지고 있음(강희정, 2015).

- 그러나 청소년 분야에서 정형, 비정형 빅데이터 체계를 파악하고 이를 구축, 활용하고자 하는 노력은 아직까지 적극적으로 이루어지지 못하고 있음. 일부 연구들이 비정형 빅데이터를 활용하여 청소년의 학업성취(고수정, 2018), 자살(송태민, 2013), 인터넷 중독 위험(송태민·송주영·진달래, 2014), 혐오표현(안명숙, 2018) 등 복지관련 이슈에 영향을 미치는 요인들을 분석·제시한 사례가 있으나 이를 기반으로 청소년 빅데이터 전체 체계를 파악하고 활용방안을 도출하기에는 한계가 있음.
- 이에 본 연구에서는 국내·외 문헌고찰, 빅데이터 분석방법론에 관한 콜로키움, 외국의 빅데이터 체계 구축 및 활용 사례 검토, 청소년 핵심 영역 온톨로지 개발, 빅데이터 체계 구축, 활용 방안 등에 대한 전문가 자문, 전문가 심층 면접, 전문가 설문조사 등을 통해 청소년 대상 빅데이터 체계 구축의 가능성을 파악하고 이를 바탕으로 제언점을 도출함으로써 청소년 빅데이터 체계 구축을 위한 기초자료를 제시함. 또한, 정형 빅데이터와 비정형 소셜 빅데이터를 활용한 분석을 시도함으로써 중요한 청소년관련 이슈에 대해 어떻게 예측이 가능한지를 확인해 보고 청소년 빅데이터 체계가 추후 청소년관련 미래변화나 정책수요를 예측하고 위험요소를 조기 발견·예방하는데 어떻게 활용될 수 있는 지 등에 관한 가능성을 타진함.

- 단, 본 연구에서는 청소년 연령 규정의 다양성과 모호성, 청소년 관련 이슈에 대한 위기중심·잔여적 접근이 가지는 한계, 공공영역과 민간영역 빅데이터의 태생적·목적적 차이와 이질성, 적용가능한 정형·비정형 빅데이터의 광범위성 등을 감안하여 연구의 범위를 (ㄱ) 청소년기본법상의 청소년 연령인 9세~24세 청소년을 주된 대상으로 하는 (ㄴ) 공공영역 정형 빅데이터와 (ㄷ) 소셜 비정형 빅데이터에 초점을 맞춤.

2. 연구결과

▶ 청소년 빅데이터 분류 체계

- 빅데이터 체계를 구축하기 위해서는 무엇보다 청소년 빅데이터의 전체 내용이 어떠한 지형을 이루고 있는지 파악할 필요가 있음.
- 아동에 관한 대표적 지표체계로 UNICEF(2007)는 아동 삶의 질을 웰빙(well-being)으로 측정하면서 하위 요소로 물질적 웰빙, 건강과 안전, 교육적 웰빙, 또래와 가족 관계, 행동과 위험요소(risks), 주관적 복지감(subjective well-being) 등을 산정하고 있음.
- 김기현·김창환(2011)은 아동·청소년 핵심지표 영역체계로 인구 및 가족, 교육 및 역량, 건강 및 심리, 보호 및 안전, 활동 및 문화, 복지 및 권리, 비행 및 범죄, 진로 및 직업, 정책 및 인프라 등을 제시하고 있음.
- 유민상 등(2019)은 국가승인통계 검토와 아동·청소년 심층면접, 삶의 질 관련 이론 및 선행연구를 바탕으로 한국 아동·청소년 삶의 질 지표를 사회적 배경 지표 목록, 아동·청소년 삶의 질 지표 목록으로 제시한 바 있음.
- 국제적 추세와 한국적 상황을 모두 고려하기 위해 국제기구의 지표체계와 한국의 지표체계를 조합하여 한국형 지표체계를 확정하는 과정이 선행되어야 할 것으로 보이며 이렇게 확정된 지표체계를 바탕으로 각 지표영역별로 어떠한 활용가능한 빅데이터가 있는지 파악하고 적절한 빅데이터가 부재한 영역도 파악할 필요가 있음.
- 본 연구에서는 김기현, 김창환(2011)의 자료를 기준으로 전문가 조사 등을 거쳐 해당 지표의 세부 내용에 해당하는 공공 정형데이터와 비정형 소셜 데이터에 대한 빅데이터 온톨로지 수정·보완 방안을 제시함.

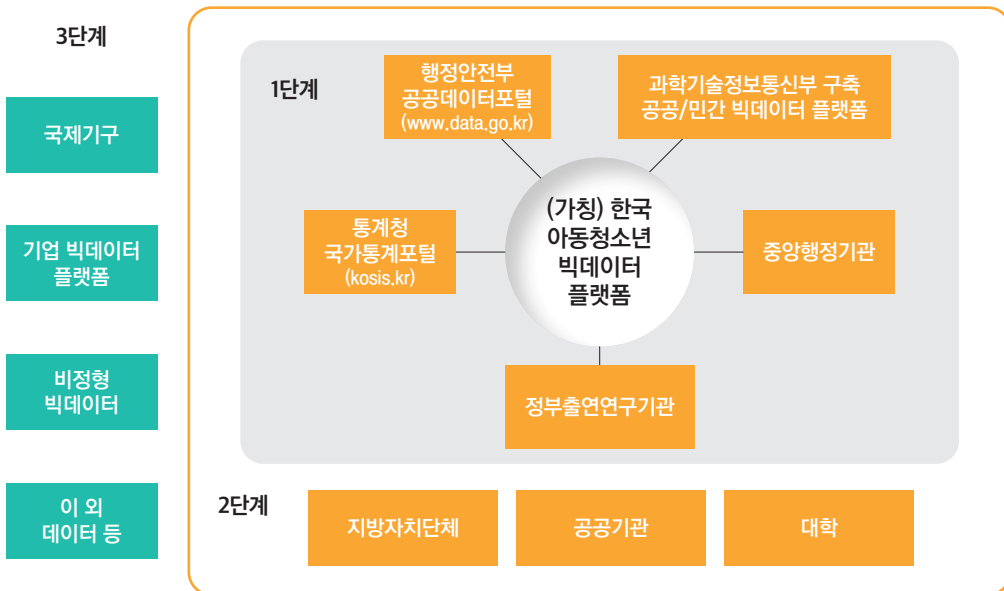
▶ 전문가 조사

- 공공 영역 빅데이터 사용 경험 유무를 살펴보기 위해 “중앙행정기관, 지방자치단체, 공공기관, 공공연구기관 등이 생성한 공공 영역의 빅데이터를 사용하여 연구 또는 분석을 실시한 경험”이 있는지 질문한 결과 전체적으로 51.5%가 있다, 48.6%가 없다고 응답함. 전문가별로는 현장전문가, 공무원은 66.7%, 교수, 연구원은 46.4%가 사용 경험이 있다고 응답함.
- 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 전체적으로 개인정보보호로 인한 접근의 제한성이 37.5%로 가장 많았음. 다음으로 데이터 형식 간의 통일성 부재로 인한 분석 상의 어려움, 데이터 품질 및 개선사항에 대한 의견 수렴 부족, 홍보 부족으로 인한 데이터 관련 정보 부재, 데이터 세부 사항, 범례 등에 대한 설명 부족, 원데이터를 분석 가능한 데이터로 변환하는 과정 불편이 각 12.5%, 데이터 항목의 누락이나 중첩 등 부정확성 존재, 데이터 간의 연계 가능성 희박, 필요한 데이터의 부재(예: 청년), 자료의 복잡성, 데이터 표집의 한계성(예: 편의추출 등), 신뢰도 및 타당도에 대한 확인의 어려움, 분석과 정책, 서비스간 연계 프로세스 부재가 각 6.3%였음.
- 민간 영역 빅데이터를 사용한 경험이 있는 경우 사용한 민간 영역 빅데이터의 유형을 살펴보면 전체적으로 국내 일간지/온라인 뉴스채널, SK텔레콤+통계청 유동생활 인구분석, 복합데이터(뉴스+게시판+자료실+SNS), 다음/네이버카페가 27.3%로 가장 많았음. 다음으로 금융빅데이터플랫폼(BC 카드) 18.2%, 네이버 지식iN, 부동산114, 네이버랩, 트위터, 유튜브, 다음 SNS, 빅카인즈, 구글 트렌드, 인스타그램, 구글 크롬 로그파일(log file), 네이버 블로그, 페이스북, 각종 온라인 게시판/댓글, 네이버판이 각 9.1%로 나타남.

- 민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 전체적으로 민간데이터 구매방법이 어렵거나 가격이 높음이 33.3%로 가장 많았음. 다음으로 공개 데이터 양, 기간의 한계, 공공데이터에 비해 데이터의 신뢰성, 타당성 확인 어려움이 각 25.0%, 개인정보보호로 인한 데이터 활용, 결합의 한계, 나에 대한 정보가 이미 많이 파악되고 있음에 대한 놀라움과 걱정, 연령 등 세부사항을 특정하기 어려움, 분석의 한계(예: 트렌드 분석은 가능하나 궤적 연구는 어려운 것 등)가 각 8.3%였음. 전문가별로는 현장전문가, 공무원의 경우 민간데이터 구매 방법이 어렵거나 가격이 높음이 75.0%로 가장 많았으며 다음으로 개인정보보호로 인한 데이터 활용과 결합의 한계, 공개 데이터 양과 기간의 한계, 연령 등 세부사항을 특정하기 어려움이 각 25.0%였음. 교수, 연구원의 경우 수집 및 분석에 큰 어려움이 없었음, 공공데이터에 비해 데이터의 신뢰성, 타당성 확인이 어려움이 각 37.5%였으며 다음으로 공개데이터 양과 기간의 한계 25.0%, 민간데이터 구매방법이 어렵거나 가격이 높음, 나에 대한 정보가 이미 많이 파악되고 있음에 대한 놀라움과 걱정, 분석의 한계가 각 12.5%였음.

3. 정책제언

- ▶ 연구내용을 바탕으로 청소년 빅데이터 체계 구축을 위한 단계와 비전 및 전략을 제시하면 다음과 같음.



[그림 1] 청소년 빅데이터 체계 구축을 위한 단계

- ▶ 단기적으로 국가통계포털과 공공데이터포털에 한국청소년정책연구원 (현)청소년통계 DB의 데이터를 적극 탑재·확대 제공하되, 장기적으로는 「(가칭)아동·청소년 빅데이터 플랫폼」으로 확대·개편하여 양질의 아동·청소년 관련 데이터를 축적하고 필요시 관련 데이터 제공 주체를 연계하여 원스톱 서비스화 하는 것을 제안함.

▶ 빅데이터 체계 구축을 위한 추진 전략은 다음과 같음(www.bigdata.go.kr, 공공 빅데이터 적용 가이드를 수정·보완함).

- 사전 준비단계 : 추진체계 구성, 데이터 확보 방안 검토, 과제 추진계획 수립
- 사업 추진단계 : 데이터 수집, 데이터 전처리, 시각화, 검증 및 안정화
- 활용 및 고도화 : 업무 적용, 분석 결과 홍보, 유지 관리, 교육, 분석 고도화

▶ 현재 아동·청소년과 관련하여 가장 공신력 있는 데이터를 다수 확보하고 있는 중앙행정기관은 ‘통계청’과 ‘행정안전부’라고 할 수 있음.

- 그러나 데이터 지표의 정의, 측정대상(연령 등), 측정방법이 상이하여 비교 분석이 쉽지 않을 뿐 아니라 데이터를 활용 가능한 형태로 만드는 데에도 상당한 수고와 노력이 뒤따름.
- 이에 반해 다수의 국가승인 통계를 확보하고 있으며, 데이터 풀셋을 제공하고 있는 다양한 국책연구기관의 데이터 허브는 통계청에 비해 광범위한 자료를 확보하고 있지는 않으나 다양한 방식으로 분석 가능한 질 높고 공신력 있는 자체 자료를 다수 확보하여 공개하고 있으며 이미 상당수의 데이터들이 축적되어 있어 비교적 빅데이터 허브로서의 역할로 전환하는 데 큰 무리가 없는 것으로 보임.
- 청소년 분야의 유일한 국책연구기관인 한국청소년정책연구원의 한국 아동·청소년 데이터 아카이브의 경우 한국아동·청소년패널조사, 다문화청소년패널조사 등 다양한 종단, 횡단 통계 자료를 확보하고 있으며 데이터 풀셋을 다양한 방식으로 분석할 수 있도록 제공하여 다양한 미래예측, 상관성 분석, 현황 파악 및 추이 분석이 가능함.
- 이에 기존의 한국아동·청소년 데이터 아카이브를 ‘(가칭) 한국아동청소년 빅데이터플랫폼’으로 확대·개편하여 아동·청소년 분야 데이터 플랫폼으로 활용하는 것을 제안함. 이 과정에서 1단계로 국가통계포털, 공공데이터포털, 공공/민간 빅데이터 플랫폼, 중앙행정기관 통계 부문, 정부출연연구기관 데이터 아카이브가 콘소시엄을 구성하여 상호 연계를 통해 자료를 공유하고 상호 링크를 통해 손쉽게 연관된 자료를 검색할 수 있도록 하되 전체 데이터의 지형을 보여주는 맵을 작성, 빅데이터 허브 내에 공개하여 이용자가 손쉽게 전체 자료의 지형을 파악하고 이용할 수 있도록 하는 것이 필요할 것으로 보임.
- 2단계로는 지방자치단체, 공공기관, 대학의 데이터 허브로 데이터 공유, 연계 체계를 확대하고 3단계로 국제기구, 기업 빅데이터 플랫폼, 비정형 빅데이터 등으로의 확대를 제안함.

▶ 연구결과를 바탕으로 세부 정책과제를 제시하면 다음과 같음.

[표 1] 영역별 주요 정책과제

영역	세부 정책과제
빅데이터 활용 향상을 위한 세부 정책과제	<ul style="list-style-type: none"> - 4차 산업혁명의 기초자원이자 결정적 요인으로 빅데이터의 개념, 활용에 대한 인식 제고 - 빅데이터에 대한 이해를 제고하고 활용성을 향상하기 위한 다양한 교육 기회 제공 - 다양한 빅데이터 분석방법론(다변량 분석, 구조방정식, 미래신허탐색, 머신러닝, 인공지능 등)에 대한 교육, 확산 - 활용가능한 기관, 내용에 대한 정보와 구체적인 활용방법에 대한 홍보 확대 - 공공 정형 데이터의 경우 가급적 원시데이터를 분석할 수 있는 환경 구축 - 개인식별정보를 제외한 공공빅데이터에 대한 적극적인 공개, 접근, 활용 필요. 이와 관련하여 정보관리와 정보보호를 위한 법제도 보강 - 데이터의 책임 있는 생산과 관리 수준 향상, 전문성 강화 - 빅데이터 연구자 간의 워크숍, 연구방법론 공유 등 - 비정형 빅데이터 활용 비용 감소 방안 모색 - 청소년 분야에서 활용가능한 빅데이터가 무엇이 있는지에 대한 종합적인 정보 제공 필요 - 데이터 표준화를 통한 데이터 융합, 가공 용이성 확대 필요. 공통으로 활용할 수 있는 key 확보 등 - 주기적 업데이트 의무화 - 데이터 세부정보 상세 제공 필요(데이터 제공시 데이터 메타 정보 함께 제공) - 정책연구와 학술연구 등 연구 분야 외에 청소년분야 현장에서 빅데이터 활용방안 사례 개발 및 보급 필요 - 데이터에 대한 이해, 분석을 바탕으로 정책의사 결정하는 풍토 조성
아동·청소년 분야 빅데이터 체계 구축을 위한 세부 정책과제	<ul style="list-style-type: none"> - 아동·청소년을 바라보는 시각을 조정: 전반적으로 통제와 보호의 대상으로 아동·청소년을 바라보는 시각에서 시민권을 가진 인격적 주체로 바라보는 시각으로 전환하여 지표 체계를 구성하고 자료를 수집, 관련하여 보편적 복지 관점에서 접근 필요 - 아동·청소년 분야 빅데이터를 생산하는 정부 부처 간의 다기관 협력체계 구축을 통해 관련 데이터를 상호연계할 수 있어야 함. - 증거기반 정책, 서비스 개발 및 제공에 대한 인식 강화. 데이터 기반 문제해결 프로세스 운영을 위한 조직 구성, 규칙 (법률 등), 예산 확보 - 청소년분야 데이터 과학자 전문인력 양성 및 확보 - 청소년 분야의 빅데이터 이용과 관련하여 어떤 해결 문제들이 있고 연구할 내용들이 있는지 선제적으로 사전에 리스트업한 후 그에 맞는 데이터를 수집해주는 것이 향후 효과적인 데이터 이용과 문제 해결을 위해 필요 - 청소년 빅데이터의 대상 범위를 정의하고 그 영역 범위를 규정할 필요. 아동과 청년의 연령 범주와의 관계를 분명히 할 필요 - 청소년의 고민과 사건사고 발생의 원인에 대한 이해를 바탕으로 관련된 데이터 수집 방법을 찾고 수집, 분석한 후 제도개선 및 서비스 구현을 위한 프로세스 구현 필요 - 구글, 네이버, 이동통신회사 등 민간 빅데이터 제공 주체와 공공 주체와의 민간 협력체계 구축 필요 - 데이터 주기별 생산량 확대와 이를 위한 정부 생산, 유통 비용 지원 강화 필요 - 빅데이터별 세부 데이터 가공 및 활용 확대를 위한 분류체계의 세분화 - 소수 청소년(학교밖, 탈북, 다문화 등)에 대한 관심 제고 - 수요자가 필요한 데이터를 맞춤형으로 제공해주고 분석을 지원해 줄 수 있는 체계 마련 필요 - 빅데이터 자료의 체계적이고 정기적인 수집, 관리와 이를 모니터링하는 시스템 필요 - 핵심지표를 활용한 데이터 관리 외에도 청소년백서나 기타 다양한 자료를 바탕으로 데이터 범위 재구조화 - 데이터는 사용자 입장에서 영역별 검색보다 키워드 검색, 푸시방식 데이터 제공 필요 - 인공지능 기능을 활용하여 사용자에게 필요한 데이터 제공 등 ICT 기반시스템에 적합한 방향으로 데이터 범위나 관리 방법 개발 필요 - 데이터 관리 전담요원을 양성하고 전문화하는 과정을 마련하여 청소년 분야의 직무의 하나로 자리매김 - 유튜브 등 아동·청소년들의 새로운 미디어 변화에 맞는 새로운 관심 분야 집중 필요. 관련하여 분류의 재정립 필요 - 비정형데이터에 대한 관심과 수요가 증가하고 있으나 이는 자율적으로 생산되는 상황에 의존하는 것이 일반적인 바, 정책적으로 생산하고 관리할 수 있는 것은 정형데이터에 대한 것이 중심이 되어야 할 것으로 보임 - 아동·청소년 빅데이터 분석을 촉진하기 위한 다학제적 접근, 연구 필요 - 국제비교가 가능하도록 청소년 분류 체계의 국제표준화 필요 - 보건, 교육, 문화 분야 등 타부서의 데이터와 연계성 강화: 수집된 데이터를 공유하여 행정효율과 협업을 촉진할 수 있도록 부처간 협력 강화

영역	세부 정책과제
아동·청소년 분야 빅데이터 체계 구축을 위한 세부 정책과제	<ul style="list-style-type: none"> - 아동·청소년 분야의 빅데이터가 효과적으로 구축되기 위해서는 플랫폼을 만들고 각 기관에서 생산되는 여러 가지 데이터를 연계하여 분석할 수 있도록 하는 시스템을 마련하는 것이 보다 효과적인 것임. 즉, 데이터 연동, match 활용 시스템 구축이 단기적으로 보다 현실적 방안 - 빅데이터의 필요성과 사용방법에 대해 공감대를 형성할 수 있는 논의 활성화 - 청소년사업의 빅데이터 생산기반 구축을 위해 청소년시설의 전자화(디지털화) 명문화 필요. 청소년활동진흥법 개정시 청소년활동디지털화 기반 조성을 위한 전자화 규정 명시 필요 - 현장활동과 관련하여 시설에 방치된 데이터 확보 방안 강구 필요 - 지속적이고 체계적인 발전을 위한 중장기 계획 수립 필요 - 사례 데이터의 체계적 수집과 활용을 위한 전자문서 서식화 필요 - 데이터 생산과 관리의 책무성에 대한 인식 강화 필요

▶ 연구결과를 바탕으로 청소년 분야 온톨로지(안)과 관련하여 김기현, 김창환(2011)에 추가되어야 하는 사항을 제시하면 다음과 같음.

[표 2] 청소년 빅데이터 온톨로지(안) 추가 사항 제시

영역	세부 정책과제
인구 및 가족 분야	인구이동, 결혼/출산 인식, 다문화 아동·청소년 수, 중도입국청소년 수, 다문화 아동청소년 지표(이주배경 아동청소년 현황), 부모자녀관계, 부모의 학력, 가족관계(가족간 응집성, 부모자녀의사소통, 조부모여부, 부모의 한부모여부, 부모양육태도, 부모자녀간 애착), 지역별 인구동태, 교급별 학령인구, 행복, 가정폭력, 부부 경제활동, 부부 가사분담, 부모에 대한 태도(온정적, 반항적), 가출유무, 가정경제수준, 부모와 함께 하는 활동, 대화시간, 함께 식사하는 횟수, 주거형태, 코로나19 이후 생활변화 등
교육 및 역량 분야	학교폭력피해율, 가해율, 학교생활, 코로나19 이후 학교생활 변화, 이성친구, 청소년 재능진단, 자기개발 지표 보완, 대안학교 유형별 청소년 수, 학업성취도, 정서역량, 도덕적 역량, 지역별 사교육비, 학교폭력여부, 아동·청소년 돌봄 형태, 학교유형, 학습시간, 진로 및 직업관련 역량, 방과후 활동 참여여부, 미디어 및 SNS 활용관련 역량, 또래간 폭력, 언어적 역량, 교내 단체활동 참여여부, 문화예술능력, 문화감수성, 사교육 분야, 리더쉽
건강 및 심리 분야	공동체意識(학교, 동아리, 지역사회 등), 흡연중독, 경계성 장애 여부, 심리상담 경험 여부, 도박중독, 자살시도, 자해시도, 자주 사용하는 미디어의 종류, 정기적인 운동 여부, 봉사활동, 인성교육, 지도력, 문화활동, 대인관계, 도덕성, 준법정신, 병원방문 경험(치과, 대형병원, 중소병원 등), 그릿(Grit), 코로나19 등 재난 이후 건강 및 심리 분야 변화, 자기가해(자기조절, 충동조절)
보호 및 안전 분야	인터넷 생활 영역(온라인콘텐츠 생산과 소비, SNS 상의 친구관계, SNS 활용 등), 청소년복지시설, 청소년보호시설 관련 경험, 사고율, 안전교육 이수여부, 코로나 등 질병여부 또는 사고 여부, 웰빙, 미디어리터러시, 유해매체 노출경험과 빈도, 성교육, 안전교육, 학교안전관리와 청소년활동분야 안전관리 데이터, 학교 및 가정의 안전교육 실시여부
진로 및 자립 분야	창업, 교육훈련, 경제활동(취업, 아르바이트 등), 부모의 진로관련, 진로희망, 진로정체감, 직업체험활동 경험, 진로 복지서비스 이용 경험 등
활동 및 문화 분야	뉴미디어, SNS를 활용한 문화생활, 동아리 종류, 문화생산자로서 활동한 경험, 사회 참여(정치참여, 환경운동 및 시민운동 참여, 불매운동, 시위참여 등), 동아리 활동, 봉사활동, 청소년 단체활동의 경우 참여율, 유형, 문화예술 교육 참여 경험, 참여율, 프로그램, 예술활동, 청소년활동 참여동기, 청소년참여기구 현황, 여가활동 유형, 인증수련활동 현황, 청소년활동종사자, 코로나19 이후 청소년활동 및 문화생활 경험 등

영역	세부 정책과제
복지 및 권리/참여 분야	정치 참여, 의식 및 가치(시민의식 등), 이주배경 청소년, 인터넷·게임 과몰입, 도박중독 청소년, 청소년자립지원시설 이용, 청소년참여활동여부(청소년운영위원회, 청소년특별회의, 참여위원회), 청소년 노동권, 청소년증 현황, 청소년증 사용현황, 청소년 액티비티, 문화활동, 학교밖 청소년 여부, 청소년 참여에 대한 인식, 가정에서의 청소년권리 보장 정도, 가정내 주요 의사결정 참여 정도 등
일상생활 분야	소비(용돈 취득 및 사용, 용돈을 주는 사람), 청소년 생활시간, 사회관계망
비행 및 범죄 분야	소년범죄 경험, 재범 비율, 성폭력,데이트폭력, 촉법소년 범죄현황, 사이버 범죄 경험, 온라인 상의 비행 경험, 비행친구, 성교육, 전자담배경험률, 범죄내용, 소년사법 각 단계별 처분인원, 인권침해시 대처방법, 피해 종류, 피해후 감정, 상담, 치료여부 등

참고문헌

- 강희정(2015). 보건의료 빅데이터 활용을 위한 기본계획 수립. 세종: 한국보건사회연구원.
- 고수정(2018). 빅데이터 군집 분석을 이용한 학습성취도 예측: 종단연구를 중심으로. 한국디지털콘텐츠학회, 19(9), 1769-1778.
- 김기현, 김창환(2011). 한국 아동·청소년 종합통계체계 구축 연구. 서울: 한국청소년정책연구원.
- 노만 매토르프 지음, 권정민 역(2016). 데이터 고급 분석과 통계 프로그래밍을 위한 빅데이터 분석 도구 R 프로그래밍. 서울: 갑우문화사.
- 문유경, 전기택, 김영택, 정성미, 배호중, 정희대, 김유정(2018). 여성가족 빅데이터의 활용을 통한 정책 개선과 발굴(Ⅱ). 서울: 한국여성정책연구원.
- 벤 웨이버 저, 배철호 역(2015). 구글은 빅데이터를 어떻게 활용했는가. 서울: 북카라반.
- 송태민(2013). 소셜 빅데이터 분석을 통한 자살 검색 예측모형 개발. 보건복지포럼, 202, 74-86.
- 송태민, 송주영, 진달래(2014). 소셜 빅데이터를 활용한 인터넷 중독 위험예측 모형. 보건사회연구, 34(3), 106-134.
- 안명숙(2018). 신조어 청소년 혐오에 관한 소셜 미디어 빅데이터 분석. 한국청소년시설환경학회, 16(1), 통권 제55호, 167-176.
- 유민상, 유성렬, 이희길, 정지은, 노남희, 구자연, 류정희, 김혜자(2019). 한국 아동·청소년 삶의 질 지표 구축 방안 연구. 세종: 경제·인문사회연구회 협동연구총서.
- 한윤선, 김하영, 송주영, 송태민(2019). 소셜 빅데이터 수집 및 분석을 위한 아동청소년 학교폭력 온톨로지 개발. 한국콘텐츠학회, 19(6), 10-23.
- 후나오 노부오 저, 김성재 역(2014). R로 배우는 데이터 분석 기본기 데이터 시각화. 서울: 한빛미디어.
- Fields, D. A., Quirke, L., Amely, J., & Maughan, J. (2016). Combining big data and thick data analyses for understanding youth learning trajectories in a summer coding camp. In Proceedings of the 47th ACM technical symposium on computing science education (pp. 150-155).
- Gharabaghi, K., & Anderson-Nathe, B. (2014). Big data for child and youth services?. Child & Youth Services, 35(3), 193-195.
- Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In 2013 Sixth international conference on contemporary computing (IC3) (pp. 404-409). IEEE.
- Watson, R. J., & Christensen, J. L. (2017). Big data and student engagement among vulnerable youth: A review. Current Opinion in Behavioral Sciences, 18, 23-27.
- 공공 빅데이터 업무 적용 가이드. https://www.bigdata.go.kr/guide/guidebook_2.pdf, 2021.8.1. 검색.
- 대통령 직속 4차 산업혁명위원회, 데이터 산업 활성화 전략. <https://www.4th-ir.go.kr/>, 2020.2.10. 검색.
- 이슈 & NDSL, 미국 연방정부, 빅데이터 연구개발 전략계획 공개. <http://www.ndsl.kr/ndsl/issueNdsl/detail.do?techSq=121>, 2020.2.10. 검색.